The Geometry of Sameness

Riemannian Equivalence of Translation and Distance for Semantic Detection

Expanded Second Edition

Bee Rosa Davis

© 2026 Bee Rosa Davis. All rights reserved.

The Geometry of Sameness. Expanded second edition.

Mathematical constructions, worked examples and computational studies are identified in their respective contexts. Source notes and references give the basis for the reported experiments.

https://www.davisgeometric.com

For the kid in the library who knew things could be measured,
even before she had any of the equations.

Contents

I The object and its descriptions **1**

1 One thing, several observations **2**

2 What the comparison must preserve **9**

3 Two routes to sameness **16**

II The Riemannian foundations **30**

4 Coordinates without confusing the object **31**

5 The metric is the rule of measurement **44**

6 Length, distance and allowable paths **71**

7 Curvature and local approximation **89**

8 Pullbacks and information loss **99**

III Translators and manifolds **113**

9 A translator system becomes a path space **114**

10 Agreement on chart overlaps **138**

11 From an atlas to translators **160**

12 From translators to a geometric realization **167**

13 What equivalence preserves **187**

IV Double cover and transported sameness **219**

14 Symmetry, quotients and two sheets **220**

15 One orientation, two quaternion descriptions **238**

16 What happens around a loop **256**

17 What a single score can conceal **268**

V The Davis manifold and defensible decisions **280**

18 Constructing a Davis manifold **281**

19 Comparing with a changing metric **290**

20 Margins and uncertainty at the boundary **311**

21 What the guarantee actually assumes **330**

22 Worked realizations and reproducible experiments **352**

23 Preserving observations through computation **368**

24 The geometry of a measured distribution **380**

25 The questions that follow **419**

A Working foundations and further derivations **434**

B Solutions and routes through the exercises **438**

C How to train your detector **442**

D The retained operator certificate **451**

E Laboratory solutions and diagnostic readings **454**

F Worked constructions: solutions and checks **458**

G From a geometric specification to a working comparison **464**

H The complete measured-field register **472**

Notation guide **524**

Framework glossary **526**

Sources and reproducibility **532**

References **533**

Closing reflections **536**

About the Author **538**

Index **540**

Preface

Why a Geometry of Sameness?

I went to Valley High School in South Sacramento. When I skipped class, I crossed the street to the library at Cosumnes River College. That's how my 4.0 became a 3.8.

Across the street I could wander the shelves and let whatever I found decide what I read next. I could pull a book down, open it in the middle, and follow a diagram until I wanted to know what the author was trying to explain. Then I could put it back and pick up a book on a subject I barely knew. I collected facts that way. What kept sending me back to the shelves was the moment I'd recognize, in a completely different subject, the way another author had taken a problem apart.

Sometimes I lost the thread of the explanation halfway down the page and still recognized the shape of the argument. I wonder how much of that was my neurodivergence and how much came from having the freedom to read across subjects. Either way, I kept wanting to know what I'd recognized. Had two authors found the same relationship in different things, or did their explanations just happen to look alike?

I didn't have the mathematics to test that in high school. Decades later, I make a living asking a version of the question I carried out of that library: when do two different descriptions tell us about the same underlying thing?

Modern systems are full of this problem. A single person might show up as an image in a face-recognition model, as an embedding vector in another system, and as a short video clip somewhere else. A patient might appear as a CT scan, an MRI volume, and a set of lab results, all living in different formats and databases. A transaction might be represented one way on a card network, another way on a bank ledger, and yet another inside a reconciliation tool. In all these cases, we are trying to decide whether two heterogeneous observations refer to the same underlying entity or event.

Engineers have built two broad styles of solutions around this question.

One style starts from *translations*. You build feature spaces for each format or modality, train translators between them, and declare two things "the same" when all the translations agree within a drift budget. This is how many telemetry, payments, and format-switching systems end up looking once they grow large enough: a graph of embeddings and translators that has to stay coherent as the world changes.

The other style starts from *geometry*. You posit or learn an underlying space of "semantic states" and treat distance on that space as a measure of change. If two observations map to points that are close along the right paths, you treat them as the same identity; if they separate beyond a margin, you treat them as different. This is how a lot of high-end detection work behaves when you strip away the implementation details: you are really working with a manifold, even if nobody calls it that.

For a long time, I treated these as two different traditions: translation-first systems on one side, geometry-first systems on the other. I built and analyzed both kinds. In one part of my life I was wrestling with graphs of translators over telemetry formats and payment rails. In another part, I was formalizing Riemannian manifolds for viral evolution and identity trajectories. It took me an embarrassingly long time to admit that, mathematically, I was circling the same problem from two directions.

This book is about that realization.

I love the geometric culture I encounter in Russian analytic geometry: the insistence that we should be able to see what a mathematical statement means in a construction. In the Russian and Soviet textbooks I read, I keep returning to that movement from a line, an angle or a motion to an equation that expresses its relations [36]. Reading and translating these books is a pleasure, but it also sharpens the question I bring to my own work: what does this mathematics let us see?

Put a bar on a turntable and mark one end. After half a turn, its axis has returned to its starting position; its marked end points the other way. We can perform the motion and compare the photographs before introducing a covering map. We have something specific to explain: one description has returned while another has changed. The mathematics must preserve that distinction. This is the kind of grounding I want from geometry.

In my own research, when a claim concerns the physical world, I want the empirical question asked first. What can we build, move, compare or measure? What does the evidence establish, and under what conditions? Measurement tests the physical claim. A geometric construction makes the relationship inspectable. Proof makes the assumptions explicit and establishes what follows whenever they hold. I want to make that whole passage visible, from the thing we can investigate to the statement we can justify.

Abstraction then lets that understanding travel. Once we know which relationships matter, we can ask where else they occur and how far they survive a change of description. A survey map, an aircraft's attitude and a measured distribution can share a structure whose meaning we have earned through those particulars. That is the order I want this book to follow: meet the object, work the construction, establish the relationship, then develop the language that carries it further. The map follows the globe. I want the world to teach the mathematics what it has to preserve.

At the core of my recent work is a framework I call a *semantic sameness structure*. It is a way to model the hidden world of underlying entities and their benign evolutions, together with all the different ways those entities get rendered into observations. On top of that hidden structure, you can build translation-based systems or manifold-based systems. The main claim of this book is that, under explicit and testable regularity assumptions, those two families of systems are not really separate worlds at all. They are two ways of realizing the same underlying structure.

I start with translator graphs whose local maps behave like smooth, well-conditioned coordinate changes on the part of the space we actually use. I call this *smooth chartability*. Compatible atlas data give a Riemannian construction, and the chart transitions of a manifold give canonical translators. In this book, *Riemannian equivalence* refers to comparisons on specified regions. Chosen invertible realizations give exact constructions. Approximate comparisons require explicit distance, semantic and round-trip bounds. Chapter 13 uses functors to describe how the maps compose. Chapter 20 shows which decision guarantees carry across under the stated comparison and population conditions. The quaternion double cover takes the comparison further. Two representatives describe one orientation, and the lifted path can retain a turn that its physical endpoint cannot tell us about.

This preface is not the place for those details; they will come in the chapters that follow, along with concrete examples from telemetry format switching, aircraft orientation and GIGI's measured distributions. But I want to be clear about two things from the beginning.

First, this book is not a marketing document for a particular system. It is an attempt to write down, in a reasonably self-contained way, the mathematical structure that sits underneath a whole family of systems I have either built or studied. The goal is to make the geometry of sameness visible as a *mathematical object*, not as a collection of unconnected tricks.

Second, I wrote this for readers who may not see themselves in traditional mathematical texts—women, people of color, self-taught engineers, late bloomers, anyone who has ever felt that mathematics was not "for them." You do not need to be a differential geometer or a category theorist to get something real out of what follows. If you are comfortable with first-year university mathematics and you are willing to meet some new ideas slowly, this book is meant to be within reach. Worked constructions, application records and the implementation workshop connect those ideas to calculations you can check. Chapter 21 develops the conditions and tests that support their use.

A note on how to read this book. Follow the maps and the globe through the five parts. Part I establishes the question; Parts II and III construct the metric and translation accounts; Part IV develops double covers and transported histories; Part V carries the geometry into decisions and reproducible examples. The appendices supply proof details, exercise guidance and the detector workshop.

When I was that kid in the library, I wanted proof that the patterns I sensed in the world could be written down, that there was a place in mathematics for the kinds of questions I cared about. This book is my way of answering that younger version of myself. There *is* a geometry of sameness. It lives in metric spaces, in Riemannian manifolds, in translator graphs, and in error budgets. My hope is that, by the time you reach the end, you will be able to see it too.

Working with the drawings and laboratories

The detailed studies follow a common sequence: identify the object, construct its descriptions, calculate what survives the translation and inspect a case where an assumption matters. A plotted curve is accompanied by its model. A recorded experiment is accompanied by its data convention. The drawings are part of the argument: their grids, fibers, tangent arrows and error regions show what the equations are comparing.

Six laboratories appear beside the mathematics they use. Appendix E collects their diagnostic readings.

Laboratory	The question it makes concrete
1	Can two maps give the same point, tangent length and route length?
2	What must change when a survey map changes coordinates?
3	Can local translators agree individually and fail around a loop?
4	Can a stored sign jump while the aircraft's attitude stays continuous?
5	Does a decision survive a transformed metric and bounded uncertainty?
6	What do the retained GIGI observations determine about a withheld value?

Appendices G and F cover implementation and construction solutions. Appendix H records all seventeen measured fields.

Part I

The object and its descriptions

Chapter 1

One thing, several observations

A technician checks a room-temperature probe labeled T-001 and copies its reading from the display to a clipboard, then into a spreadsheet and a message to a monitoring program. One measurement now appears in four places. How can the program tell that all four reports describe the same event?

Follow one observation through the room

Each copy of the room probe's reading needs enough context to be traced back to the instrument. The probe identifier and acquisition time identify a reading if the instrument logs at most one reading per timestamp. Its unit and calibration history tell us what the number means.

Start with the retained T-001 record at 22.4 °C. Its Fahrenheit conversion, 72.32 °F, describes the same acquisition. The conversion $F = 9C/5 + 32$ makes their values agree. Now suppose T-001 measures the room again a minute later and reports 22.4 °C. That hypothetical repeat would be a separate observation. A fourth record rounds the first value to 22 °C. It may be adequate for a room-comfort display while being inadequate for a calibration calculation.

If I'm counting measurements for the morning report, I log one event for the Celsius reading and its Fahrenheit copy. I log another if T-001 actually takes a second reading. If I'm checking whether the room stayed within a comfort band, the rounded display may be enough.

SINGULARS / INSTRUMENT, VALUE, PROVENANCE

The anatomy of one observation

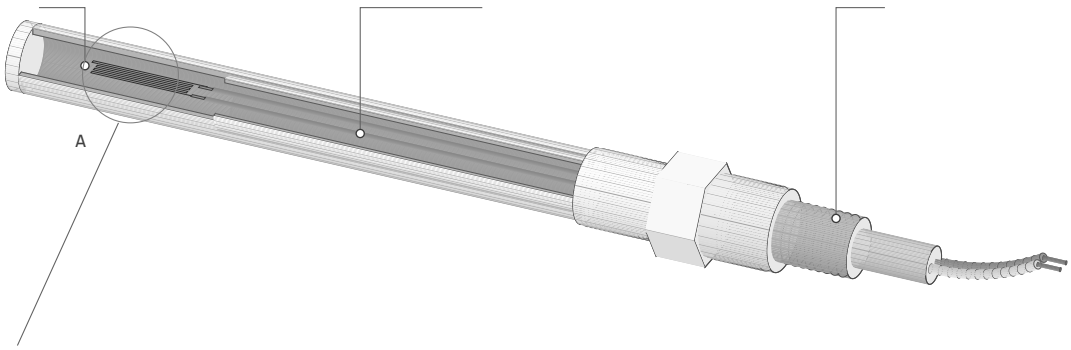
The number begins in an instrument. Its identity begins in an event.

01 A physical origin

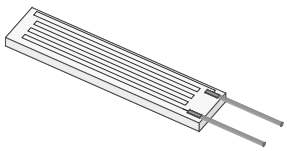
SENSING ELEMENT
The local temperature changes an electrical resistance.

PROTECTIVE SHEATH
A physical package surrounds the sensing region.

LEAD-OUT
Carries the response to the measuring electronics.



A / ELEMENT DETAIL



Ceramic carrier + continuous resistive path

THE RECORD LEAVES WITH ITS ORIGIN

T-001	22.4 °C
ACQUISITION	1710000000
UNIT celsius	STATUS normal

The timestamp belongs to the observation, not its later copy.

Illustrative resistive probe; the recorded fixture does not specify sensor hardware.

02 Three retained records

The Celsius entries and timestamps are retained; the Fahrenheit column is calculated.

PROBE	ACQUISITION TIME	CELSIUS	FAHRENHEIT	STATUS
T-001	1710000000	22.4	72.32	normal
T-002	1710000060	23.1	73.58	normal
T-003	1710000120	45.8	114.44	alert

A repeated value is not enough to identify an event.
Instrument, acquisition time and calibration provenance determine what may be compared.

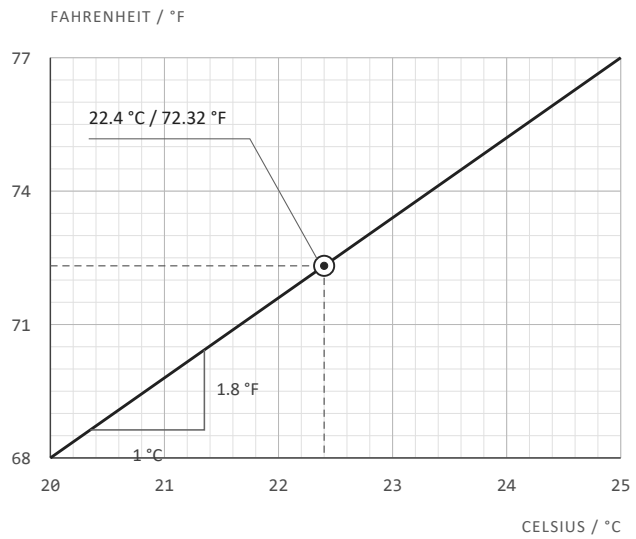
Figure 1.1: *The individual observation survives several descriptions.* The study continues on the following figure page.

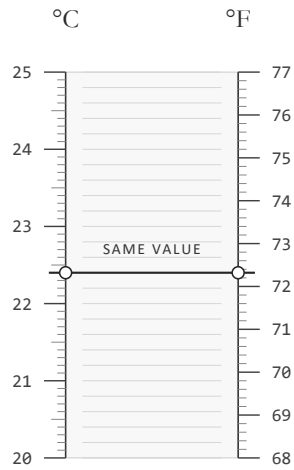
SINGULARS / CONVERSION, IDENTITY, INFORMATION LOSS

What survives a new description?

Changing units, changing time and discarding precision are different operations.

03 Different coordinates, the same quantity





The tick spacing changes.
The measured quantity does not.

$$F = 9C/5 + 32 \qquad C = 5(F - 32)/9$$

04 Equal values; distinct events

Constructed repeat reading of the same probe

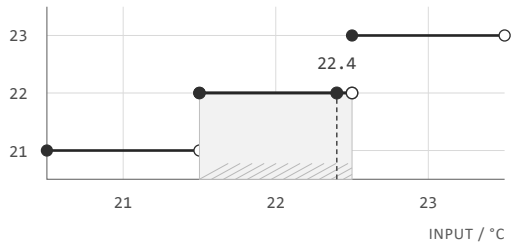


$$\text{value}(r_0) = \text{value}(r_1)$$
$$E(r_0) \neq E(r_1)$$

E assigns the event described by a complete record.
Equal values alone cannot establish event identity.

05 Rounding erases a distinction

Q(C) = floor(C + 1/2), in Celsius degrees



$$[21.5, 22.5) \rightarrow 22$$

The inverse is an interval, not a recovered reading.
Displaying 22 °C does not retain the original 22.4 °C.

Figure 1.1: *The individual observation survives several descriptions (continued).* The register contains three retained readings. Separate panels take the T-001 record, convert it to Fahrenheit, compare it with a hypothetical repeat, and round it to a whole-degree display. Invertible unit conversion preserves a physical value; event identity also needs the probe, acquisition time and calibration. Rounding loses the exact original reading.

An equivalence class needs an explicit rule

Let $\mathcal{R}$ be a set of complete records and let $E : \mathcal{R} \to \mathcal{O}$ assign the underlying event described by a record. Then $r \sim s$ when $E(r) = E(s)$ is an equivalence relation. Reflexivity, symmetry and transitivity follow directly from equality in $\mathcal{O}$.

In practice, E may not be known. A matching algorithm estimates whether two records share an event. That estimated relation need not be transitive. For instance, declaring two temperatures equivalent when their difference is at most one degree makes 20 equivalent to 21 and 21 equivalent to 22, but not 20 equivalent to 22. A tolerance neighborhood is useful; it is not automatically an equivalence class.

Suppose I link consecutive readings whenever they're a degree or less apart. Enough little steps can put a cold morning and a warm afternoon in the same group. That's transitive closure here. To check how much the room warmed, I compare the latest reading directly with the first.

The globe in this example is the event model

Think of several maps marking the same buoy. Here one acquisition event takes the place of the buoy, and record formats supply different descriptions of it. Some are smooth coordinate charts on a state manifold; others are discrete encodings. Their common role is to let us describe the same object in several ways. Their mathematical properties must still be checked individually.

The event model tells us which reports refer to one observation. We can now ask whether their numerical descriptions measure that observation consistently. That next requirement needs an explicit conversion and a ruler.

Give the comparison a numerical test

The probe identifier and acquisition time tell us which event a record claims to describe. The numbers still need a conversion rule. To see what that rule must preserve, we will compare two scales along a straight inspection rail and measure the same movement on both.

Suppose a bench instrument reports the position of an edge as $x = 2.00$ metres. Another uses centimetres and a zero point five centimetres behind the first, so its reading is $y = 205$. The conversion is

$$y = T(x) = 100x + 5.$$

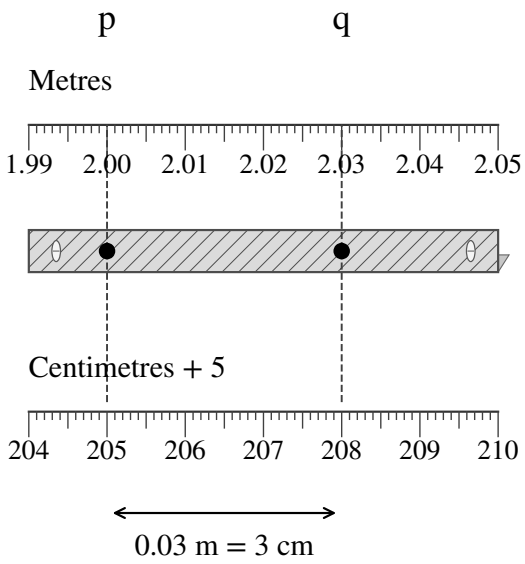
The movement from $x = 2.00$ to $x = 2.03$ becomes a movement from $y = 205$ to $y = 208$. It is three centimetres in either account. The equality of these descriptions requires both the conversion and the ruler. With $dx = dy/100$,

$$ds^2 = dx^2 = 10^{-4} dy^2.$$

The second display goes up by three ticks, each one hundredth of a metre. Our first metric tensor records that scale, with the conversion factor squared because the equation measures squared length.

There is a second question. Two different parts can occupy that same position on different days. Correct coordinates do not establish the identity of the part. We need the record's context, its time and its lineage. Where a part is and which part it is are two different questions.

A *One rail has two coordinate reports*



B *Converting units preserves length*

Quantity	*Metres*	*Shifted cm*
p	2.00	205
q	2.03	208
Difference	0.03	3
Metric	1	1/10,000
Length (m)	0.03	0.03

$$y = 100x + 5, \qquad ds^2 = dx^2 = 10^{-4} dy^2$$

Figure 1.2: One rail described in metres and shifted centimetres. The marked positions remain fixed. Transforming the metric as well as the coordinates gives the same 0.03-metre separation in either description.

The same harbor on two charts

Lay two coastal charts on a table. They cover the same harbor, but one uses a different projection, a different scale and a different place for its coordinate origin. The red mark for a buoy sits near the corner of one sheet and near the center of the other. No one has moved the buoy. What moved was its address on the paper.

A navigator still needs a way to translate that address. Recognizing that two charts depict one harbor is the beginning of the problem, not its solution. The navigator must know which coordinates correspond, how the scale changes, and whether a route that looks short on the sheet is short on the water. A bad translation can put a real boat on a real rock.

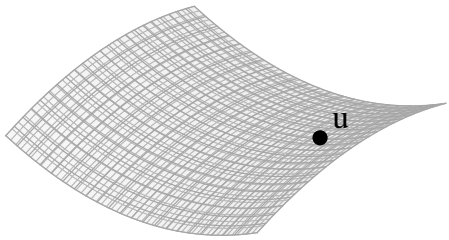
A map is a local description. The globe is the shared geometric object being described. The globe need not literally be spherical: it might be the space of an aircraft's orientations, a line of temperature states, or a curved family of probability distributions. What matters is that the description and the object have different jobs. We can change one without necessarily changing the other.

The engineering version arrives as a telemetry record. One copy was written by the instrument; another has passed through a storage migration. The identifier and time agree. The numerical fields have changed format. Before calling it a new event, the engineer must ask what the conversion preserved. AMC puts that question to work in telemetry (Chapter 22).

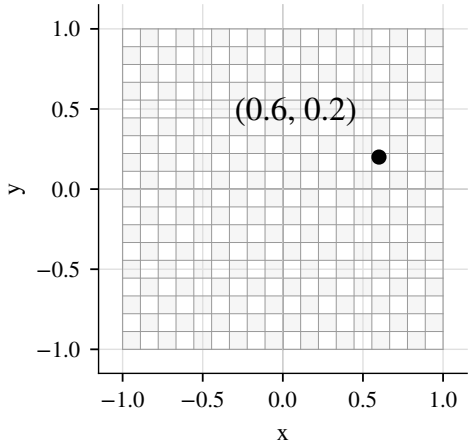
Observation channels describe the same event

The calibration calculation gives a general way to organize the problem. Write u for the underlying event or state. An observation channel produces a report about u, and an encoder supplies the representation a program uses. A record format, an image and a numerical vector can therefore describe one object while carrying different information about it.

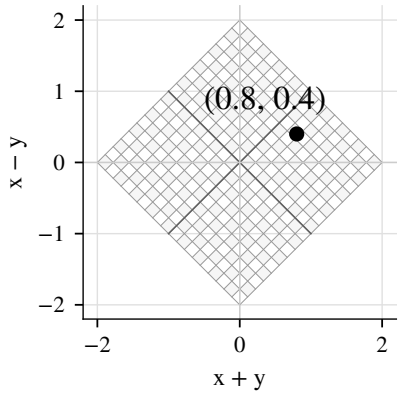
A *One physical point is fixed*



B *One channel reports its coordinates*



C *Another channel mixes the coordinates*



D *The record names its coordinate system*

Field	*Value*
object	u
x	0.6
y	0.2
schema	(x, y), same frame

Constructed channels. Their values differ; the marked point is fixed.

Figure 1.3: A constructed observation example: one marked surface point, an image-plane coordinate pair, a second numerical encoding and a structured record. The second encoding is $(x + y, x - y)$. The values and formats differ while the underlying marked point remains fixed; these are declared channels, not learned-model performance results.

In the drawing, the object is one marked surface point. The image coordinates, transformed coordinates and structured record let us inspect the correspondence directly. An application must establish its own correspondence. A photo and a video may show one person; a laboratory report and a scan may concern one patient; a bank statement and a receipt may refer to one payment. Each requires evidence identifying the common object and a rule for comparing what its observations retain.

In the AMC telemetry project, changing formats can preserve every recorded field or introduce a declared approximation. The ICARUS flight-control work will give us a precise orientation

comparison. The GIGI data system will let us inspect how records are decoded and distributions compared. Each task supplies its own object and measurement rule.

We have established two requirements for a faithful comparison. Its reports must remain attached to the intended event, and their numerical values must be measured with compatible rulers. The rail conversion supplied our first metric coefficient. Its success leaves the next difficulty exposed. A second reading from the same probe is a new event, even when its number repeats. Chapter 2 asks which of those identities a particular task means to preserve.

Practice with the construction

The first calculation repeats the rail's conversion with a new scale and origin. The second checks the event identity that the calculation alone cannot supply.

1.1. An instrument uses $y = 1000x - 12$, with x in metres. Find the reading corresponding to $x = 0.025$ and the metric coefficient in the y coordinate.

1.2. Two records both report $22.4\,^{\circ}\mathrm{C}$. Give one situation in which they describe the same measurement and one in which they do not. State the extra evidence needed to distinguish the situations.

Chapter 2

What the comparison must preserve

The room-temperature probe T-001 has a retained reading of 22.4 °C in Chapter 1. Suppose it measures the room again a minute later and reports the same value. A duplicate-message filter must keep both measurements. A maintenance program puts both in the same probe history so it can check for drift. Each program needs a different test for sameness.

I have spent years building systems that make sameness decisions across very different domains: spacecraft telemetry (keeping track of system state across format translations), viral sequences (tracking antigenic identity across mutations), payment rails (reconciling transactions across heterogeneous ledgers), and face recognition (matching identities across images and videos). In every case, I started with some ad hoc similarity heuristic—match these fields, threshold this score, cluster these embeddings—and refined it until the error rates looked acceptable on test data. It often worked well enough in the narrow setting it was designed for. But it always felt like duct tape. The problem was not that the heuristics failed catastrophically. The problem was that they did not *compose*. Every new format, every new sensor, every new embedding model meant starting over.

The difficulty becomes precise when we ask what a permitted change leaves invariant. Begin with an object we can turn in our hands, then carry that question into the instrument network. We will test three proposed rules. Physical identity must survive a changed appearance. A grouping of stations must satisfy a communication rule. A predictive state must retain the information required for future observations. Each rule specifies what its mathematical comparison must preserve.

A question has to choose its invariants

Photograph a mug with its handle to the right. Turn it and take a second photograph. Then swap in a second mug of the same design, put its handle to the right, and photograph it. The first two images show one mug from different sides. The first and third can look alike while showing different mugs.

The task therefore needs an *invariant*: a property the permitted change leaves unchanged. A warehouse may care about the serial number of the mug's package. A manufacturer may care only about model and dimensions. A photographer may care about the image itself. These are legitimate comparisons of different objects of study. They are not interchangeable uses of the word same.

Write O for the set of observations and $I : O \to U$ for the ideal identity assignment in a specified task. Exact identity in this model means $I(o) = I(o')$. If the identity assignment is unavailable, a detector estimates that statement. If I'm tracking my mug, the detector's yes is a guess about which mug it saw. That guess doesn't turn the second mug into mine.

A detector often substitutes numerical closeness for exact identity. Test that rule on three readings from a position sensor. The relation $|x - y| \le \tau$ is symmetric and reflexive but generally not transitive. With $\tau = 1$, the readings $0, 0.75, 1.5$ satisfy the first two adjacent comparisons but fail the end-to-end comparison. We cannot take its equivalence classes without supplying another

rule. Connected-component clustering supplies transitivity by chaining, and can thereby group points much farther apart than τ.

Proposition 2.1 (Tolerance chaining)**.** *If a sequence $x_0, \ldots, x_k$ in a metric space satisfies $d(x_{j-1}, x_j) \le \tau$, then $d(x_0, x_k) \le k\tau$. For fixed $\tau > 0$, no bound independent of k follows from these hypotheses alone.*

Proof. Apply the triangle inequality along the sequence. The real-line example $x_j = j\tau$ attains the bound. □

A part on the rail advances 0.75 twice. Each move passes a one-unit check between readings, though the part is now 1.5 from its starting mark. The tracker can keep following the same part. A travel alarm needs that starting mark.

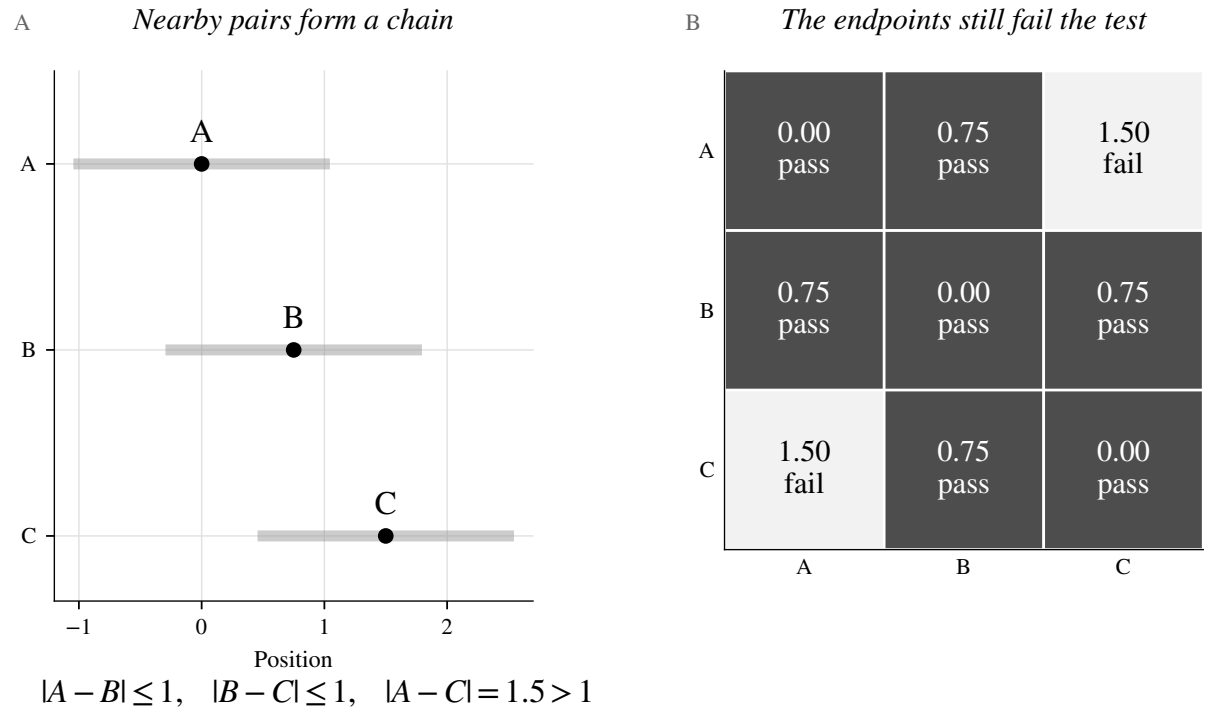


Figure 2.1: Three positions under a threshold of one unit. Each interval shows positions admitted by one observation. The complete pairwise table exposes the failure of transitivity: the two neighboring comparisons pass while the endpoints fail.

The mug showed why a similar appearance cannot supply individual identity. A network operator faces a related design choice when several stations are processed as one unit. The operator must first say what makes such a group admissible. We will require its members to exchange records without routing through stations outside the group. This gives a concrete rule whose consequences we can count.

Which groups can exchange a record?

Put four labeled monitoring stations on a desk. Keep the stations and their labels fixed. First connect them in a chain, then add a link joining the endpoints into a cycle, and finally connect every pair. The individual objects have not changed. In the chain, stations 1 and 3 can exchange

a record through station 2. If we put 1 and 3 in one group and leave 2 and 4 in separate groups, 2 is outside the first group. That partition becomes admissible when the complete graph adds the direct link from 1 to 3.

A connected partition divides the stations into nonempty blocks, each inducing a connected subgraph. For the chain $1-2-3-4$, every block is a contiguous segment. Choosing which of the three chain edges to cut gives $2^3 = 8$ partitions. On the four-cycle, there is one all-station block. Cutting two, three or four edges gives six, four or one additional partitions, for twelve in all. For the complete four-vertex graph every set partition is allowed, giving the Bell number $B_4 = 15$.

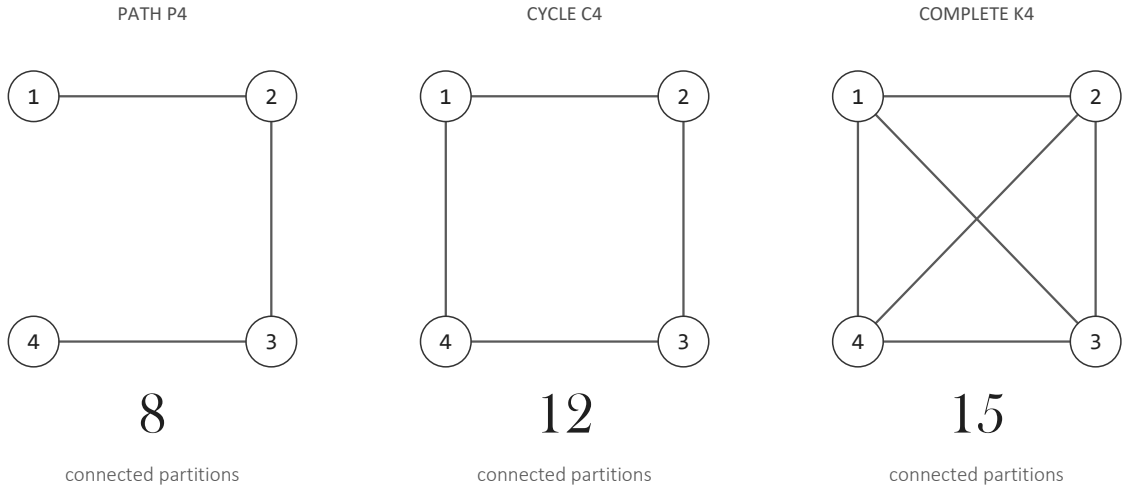
This finite example, developed in the later geometric-partition work [30], puts a limit on what a list of objects alone can explain. Relations constrain which collections count as coherent units. A sameness criterion that ignores the relation graph can merge groups that have no allowed internal connection.

Proposition 2.2 (Adding connections preserves existing connected partitions)**.** *If two graphs G, H have the same vertex set and every edge of G is an edge of H, every connected partition of G is a connected partition of H.*

Proof. Each block connected by edges of G remains connected after more edges are available. The vertices and blocks are unchanged, so the same partition satisfies the condition in H. □

Connectivity determines which groups are allowed

The same four vertices form a path, a cycle and a complete graph.



Adding edges preserves every previously connected block and may admit new ones.

A Connected groups: partitions 1 through 8

PARTITION	P_4	C_4	K_4
1 | 2 | 3 | 4	yes	yes	yes
12 | 3 | 4	yes	yes	yes
2 | 13 | 4	—	—	yes
2 | 3 | 14	—	yes	yes
1 | 23 | 4	yes	yes	yes
123 | 4	yes	yes	yes
23 | 14	—	yes	yes
1 | 3 | 24	—	—	yes

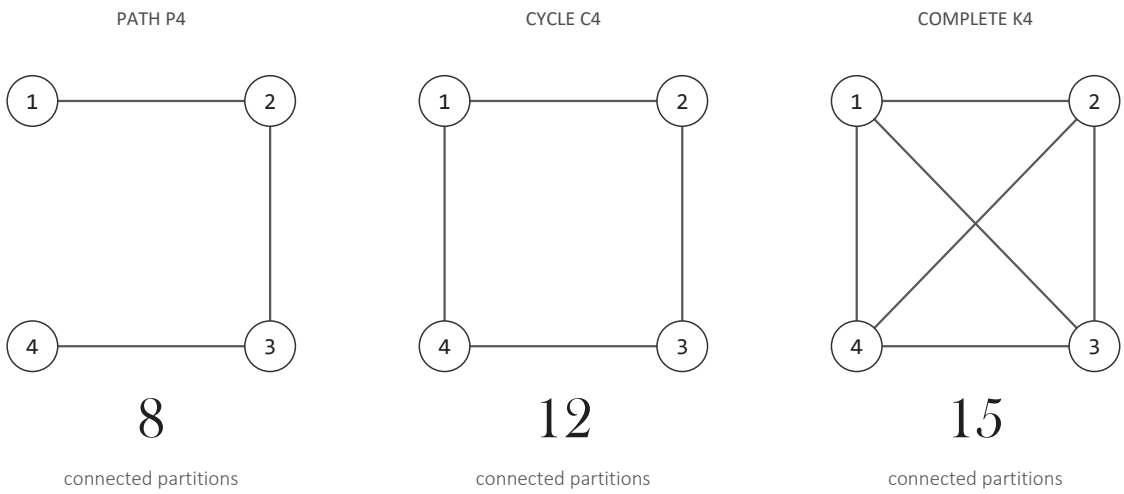
B Connected groups: partitions 9 through 15

PARTITION	P_4	C_4	K_4
13 | 24	—	—	yes
3 | 124	—	yes	yes
1 | 2 | 34	yes	yes	yes
12 | 34	yes	yes	yes
2 | 134	—	yes	yes
1 | 234	yes	yes	yes
1234	yes	yes	yes

Figure 2.2: *The objects stay fixed while their permitted groupings change.* The study continues on the following figure page.

Connectivity determines which groups are allowed

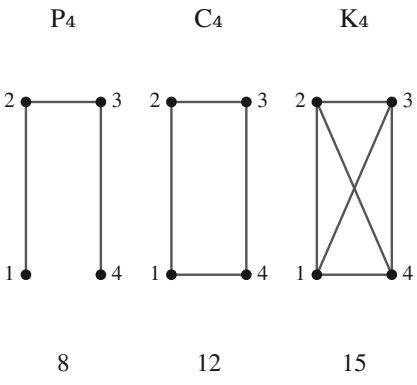
The same four vertices form a path, a cycle and a complete graph.



Adding edges preserves every previously connected block and may admit new ones.

C More edges permit more connected groupings

D Adding edges cannot undo an existing grouping



Connected-block partitions: $8 \leq 12 \leq 15$.

Keep the same vertices and partition.

An existing connected block stays connected after adding edges.

Some previously disconnected blocks become allowed.

Figure 2.2: *The objects stay fixed while their permitted groupings change (continued).* All fifteen partitions of four labeled vertices are classified against a path, a cycle and a complete graph. The eight, twelve and fifteen accepted partitions are obtained by explicit connectivity checks. The drawing makes the relation-dependent definition inspectable.

What history must a station retain?

Each station can now exchange records within its group. Suppose one station monitors a machine and its maintenance log is getting too large to keep online. The operator wants a compact record that preserves the forecasts made from that log. Under a specified stochastic model, two histories can be stored as the same predictive state if they give the same distribution for future machine readings.

For two recorded histories h and h' up to time t with conditional forecasts specified by the model, write $h \sim h'$ when

$$\Pr(X_{t+1:t+k} \in A \mid h) = \Pr(X_{t+1:t+k} \in A \mid h')$$

for every finite future length k and every event A in that future observation space. Suppose two forecasts each give a 50 percent alert chance at both of the next two times. One predicts that alerts always occur together. The other predicts independent alerts. For the event A of two alerts, their probabilities are $1/2$ and $1/4$. The histories therefore need distinct predictive states.

This predictive equivalence is the organizing question in the later causal-state work [25]. In a hidden Markov model, a posterior distribution over hidden states often supplies a convenient predictive representation. Distinct posterior vectors may still have the same future law if the model has observationally indistinguishable hidden states. The vector representation and the predictive equivalence classes should therefore not be identified without checking the model.

We can now state what the compact predictive representation must preserve. Every finite future observation event must receive the same conditional probability after replacement. The graph preserved admissible communication groups; this construction preserves a future law. Both began with a concrete operation that the replacement had to support. The invariant follows from that operation.

Think of it like taking over the night shift. Somebody hands me a one-page note instead of the whole maintenance log. It says there's a fifty-fifty chance the machine will throw an alert at the next check. Okay, but if it does, what should I expect at the check after that? In the first forecast, an alert now means another one is coming. In the second, the next check is still a coin toss. That's a difference the handoff needs to keep. For forecasting, two logs can share the same note when, under our model, they give the same odds for every sequence of readings I might see on my shift.

Specify the common object before building its representations

The three constructions tell us what belongs in the specification of a sameness problem. We need the underlying entities, the changes the task permits, the observations available, and the question we want those observations to answer. Together these form a *semantic sameness structure* S.

For the temperature-monitoring system, the comparison might concern a single reading, the instrument that took it, or the machine state inferred from its history. A permitted temperature drift can preserve an instrument's identity while changing its state. A storage conversion can preserve an acquisition event while changing every byte used to describe it. The observation channels tell us which evidence each comparison actually receives.

A realization of S is a system built to answer the specified question. One implementation translates between observation formats. Another places the observations in a geometric model and measures their separation. The task remains fixed while we compare these implementations. This is the central connection developed in the following chapters, including the exact double-cover constructions in which two representatives describe one physical state.

The next chapter makes the two implementations meet in a calculation. The invariant will be one measured displacement, and we will be able to inspect both the conversion and the ruler that preserve it.

Practice with the construction

The first question tests the chaining failure shown in the tolerance drawing. The second returns to the mug and asks you to declare the invariant yourself. In the station drawing, find the partition $\{1, 3\} \mid \{2\} \mid \{4\}$. Explain what the direct link between stations 1 and 3 changes. These are applications of the same discipline, choosing a question before accepting its proposed grouping.

2.1. For $x_j = 0.6j$ and threshold 1, find the smallest chain whose endpoints fail the threshold although each adjacent pair passes. Explain what connected-component clustering would do.

2.2. Choose a change allowed in the mug example. State an invariant, an observable that changes, and a possible measurement that would test the claimed invariant.

Chapter 3

Two routes to sameness

Two engineers are asked how far a part moved along an inspection rail. One converts the two position readings to a common unit and subtracts. The other measures the length between the marked positions using the rail's metric. They have the same assignment, so we can check exactly when their answers agree.

Make the two routes meet at a number

Use the two scales from Chapter 1. Let x and x' be the rail positions in metres before and after the part moves. The centimetre scale has a shifted zero, so its readings are $T(x) = 100x + 5$ and $T(x') = 100x' + 5$. The translator route converts the difference back to metres:

$$d_T(x, x') = \frac{|T(x) - T(x')|}{100}.$$

The geometric route integrates the ruler on the position line:

$$d_g(x, x') = \inf_{\gamma : x \to x'} \int_0^1 |\dot{\gamma}(t)| \, dt = |x - x'|.$$

Substitution proves $d_T = d_g$. On the second scale I subtract two readings, so their shared offset cancels. I divide by a hundred to get metres. The geometric route builds that scale change into the ruler.

In more dimensions, the ruler is a positive-definite matrix that may vary with position. A translation $y = T(x)$ changes a small displacement v to $DT_x v$. Preserving its squared length requires

$$v^{\mathsf{T}} G_x v = (DT_x v)^{\mathsf{T}} G_y(T(x))(DT_x v),$$

hence $G_x = DT_x^{\mathsf{T}} G_y(T(x)) DT_x$. We will derive this carefully, integrate it along curves and establish exactly when it preserves distances.

Two complementary descriptions are not automatically a two-sheeted covering space. A double cover makes a more specific statement: every base point has precisely two representatives, locally arranged so that each representative describes a separate copy of the same neighborhood. The physical orientation of a small aircraft provides that structure. Its unit quaternions q and $-q$ describe one rotation. The maps and the globe tell us how to keep descriptions separate from objects; the covering map tells us exactly how these two descriptions identify one object.

The translation/geometric thesis concerns the same comparison carried out through compatible constructions. The double-cover examples then reveal why faithful descriptions can require more than one representative, and why some information belongs to the route taken between them. Those are connected mathematical questions, each with an explicit object, map and test.

The bench conversion gave an exact agreement in one coordinate. We now make the same distinction visible in two dimensions, where a map can stretch or shear the displayed grid.

The translation and the journey must meet at a calculation

Take a location on the globe, express it on one map and translate those coordinates onto another. Now take two locations and measure their separation by following the globe's metric along a route. A meaningful equivalence between these descriptions requires more than correct endpoint names. The transition must also carry the measurement rule.

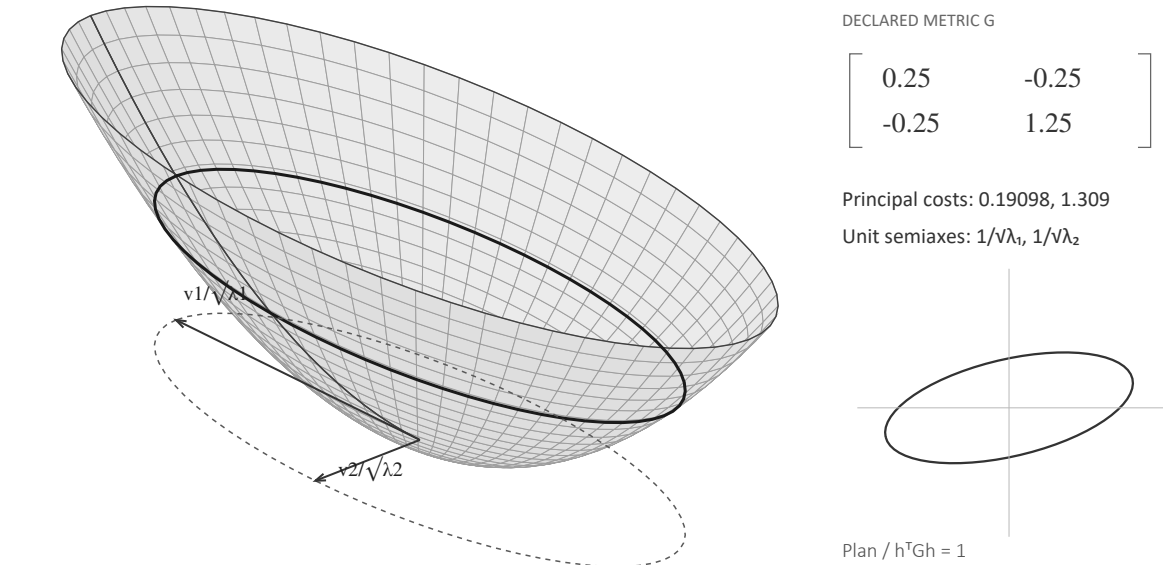
For charts x and $y = T(x)$, a curve has velocities related by $\dot{y} = DT\,\dot{x}$. Its two length calculations agree when

$$g_x = (DT)^T g_y (DT).$$

Then the integrands agree point by point:

$$\sqrt{\dot{x}^T g_x \dot{x}} = \sqrt{\dot{y}^T g_y \dot{y}}.$$

The shear tilts the survey grid. Each small piece of a route gets new coordinates and keeps its length under the new ruler. When the matrix check holds throughout the overlap, each route there keeps its total.



A A shear changes the coordinate grid

B The ruler must transform with the grid

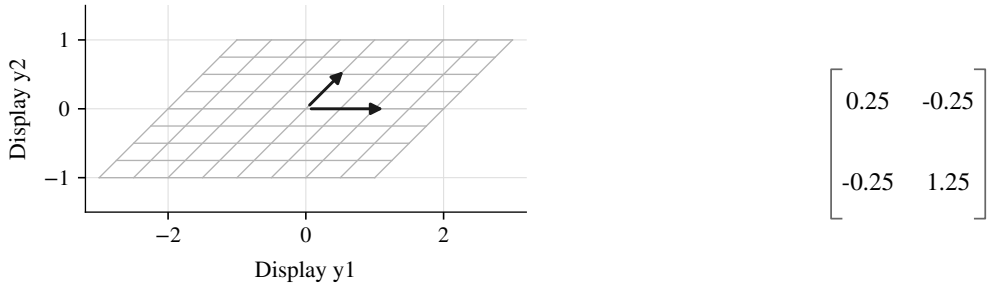
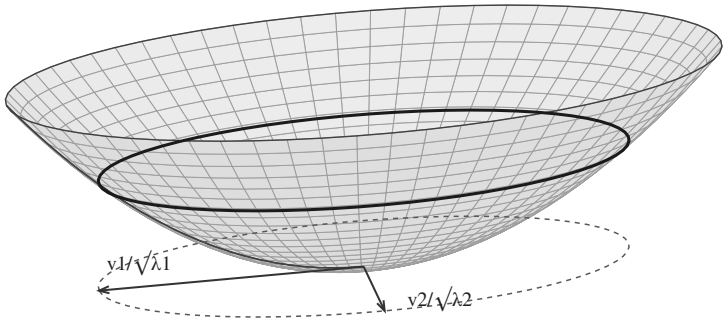


Figure 3.1: *A translator must preserve the quantity the task measures.* The study continues on the following figure page.

The ruler is visible in its level sets

Constructed cost surface: height = $h^T G h$. The bold section has unit cost.

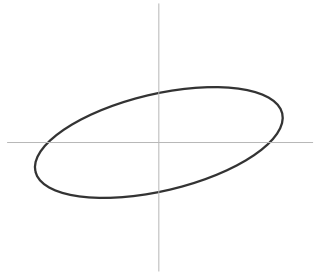


DECLARED METRIC G

$$\begin{bmatrix} 0.25 & -0.25 \\ -0.25 & 1.25 \end{bmatrix}$$

Principal costs: 0.19098, 1.309

Unit semiaxes: $1/\sqrt{\lambda_1}$, $1/\sqrt{\lambda_2}$



Plan / $h^T G h = 1$

C The quadratic form stays invariant

D Approximate translation needs explicit bounds

A=[[2,1],[0,1]]	Compare quadratic forms in every direction.
$G_y = A^{-T} A^{-1}$	Integrate the resulting speed bounds.
dy=A dx	Distance bounds require compatible path families.
$dy^T G_y dy = dx^T dx$	Decision agreement also needs a margin.

Figure 3.1: *A translator must preserve the quantity the task measures (continued).* The sheared grid and its transformed metric give an exact quadratic-form identity. The calculation specifies how vector components and the ruler change together. The final register states the extra domain, path-family and margin requirements for an approximate translator.

When agreement has a tolerance

An implementation may only satisfy

$$(1-\epsilon)^2 g_x \preceq (DT)^T g_y (DT) \preceq (1+\epsilon)^2 g_x, \qquad 0 \le \epsilon < 1.$$

Here $A \preceq B$ means $v^T A v \le v^T B v$ for every vector v. Taking square roots and integrating yields

$$(1-\epsilon) L_x(\gamma) \le L_y(T\gamma) \le (1+\epsilon) L_x(\gamma).$$

If the translator and its inverse identify the relevant path families, the same inequalities pass to their infimum distances. If it maps only some permitted paths, one cannot freely infer both distance bounds. The admissible routes belong in the contract.

Say I need a walking route under a hundred metres. If the map conversion can change its length by two percent, ninety-nine metres could turn into more than a hundred. Ninety-five still fits. Later we'll count how often a reading lands close enough to make us hold the decision.

This is how the two descriptive routes remain one research program. Translation provides the transformation; geometry supplies the invariant or the controlled discrepancy; the application states which decision must survive. A genuine double cover gives a particularly important family of such transformations by separating an underlying object from its paired representatives.

On the rail, both engineers get the same distance because they know the conversion and the physical scale. Consider a less controlled setting: a photograph and a video clip may show the same person. We need both a way to convert what each camera reports into numbers and evidence that comparing those numbers answers the identity question.

In Chapter 1 we treated semantic sameness as a property of a latent world of entities and benign evolutions, together with a family of observation channels and a decision problem. That gave us the abstract structure: the hidden variable u, its trajectories $u(t)$, the observation channels that produce images, embeddings, logs, and so on, and the question of whether particular observations could have come from the same u.

This chapter is about what happens when we actually try to *build* systems that answer that question.

Engineers and scientists have developed two broad styles of realization:

- **Translation-based realizations**, which live in the world of observable feature spaces and learned mappings between them.
- **Geometry-based realizations**, which live in the world of an underlying space of semantic states equipped with a notion of distance.

Both styles can be instantiated inside the same semantic sameness structure S. In the illustrative identity comparison, both are different ways of building a system that decides whether a still photograph and a short video clip show the same person.

The bench gives an exact reference. The following engineering descriptions identify what a more general implementation must supply before it can claim comparable agreement.

Character A: The Translator Builder

Imagine an engineer who thinks in terms of data formats, feature spaces, and functions that turn one representation into another. We will call this person the *translator builder*.

Many spaces, many encoders

The translator builder starts from the observation that different channels produce very different kinds of objects. In the face example, there might be:

- A still-image pipeline that takes raw pixels and produces a face embedding vector in some space we will call V_{img}.
- A video pipeline that takes a short clip, runs it through a temporal model, and produces a video embedding in a space V_{vid}.
- A logging pipeline that writes down a structured record with device IDs, timestamps, and some coarse metadata in a feature space V_{log}.

Each of these spaces V_i comes with its own encoder, which we will denote E_i:

$$\begin{aligned} E_{\text{img}} &: \text{image} \to V_{\text{img}}, \\ E_{\text{vid}} &: \text{video clip} \to V_{\text{vid}}, \\ E_{\text{log}} &: \text{log record} \to V_{\text{log}}. \end{aligned}$$

The translator builder thinks of these encoders as black boxes that take messy, heterogeneous observations and put them into reasonably well-behaved vector spaces.

Translators between spaces

Once the encoders are in place, the translator builder tackles the sameness problem by learning *mappings between feature spaces*. For any two spaces V_i and V_j that seem related, they train a translator

$$T_{ij} : V_i \to V_j$$

that tries to map feature vectors from channel i into the feature space of channel j in a way that preserves semantic identity.

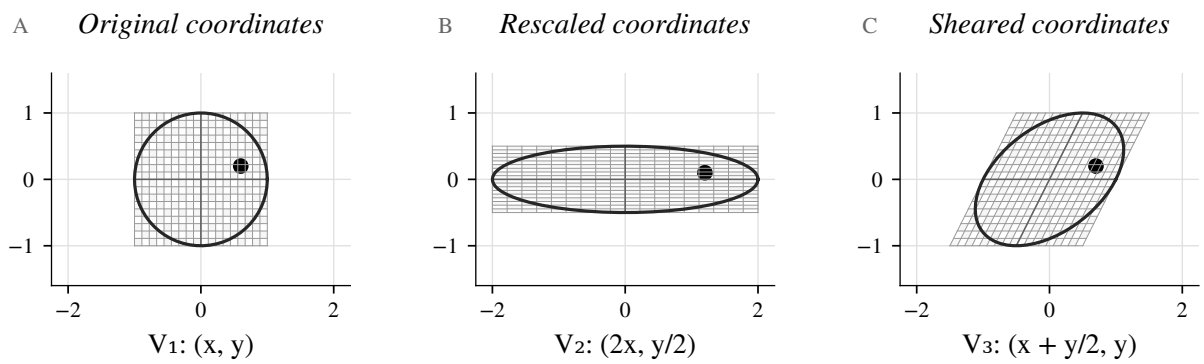
In the face example:

- $T_{\text{img}\to\text{vid}}$ might take an image embedding and predict what the corresponding video embedding would have been, had we seen a short clip of the same person.
- $T_{\text{vid}\to\text{img}}$ might go the other way, mapping video embeddings back into the image-embedding space.
- $T_{\text{img}\to\text{log}}$ might predict which kinds of device or network context usually accompany a given face, based on past data.

These translators are typically trained on pairs or tuples of observations that are known (or strongly believed) to be about the same latent entity: same person, same account, same system state. The details of how this training is done do not matter yet. What matters is that the translator builder ends up with a graph whose nodes are representation spaces V_i and whose edges are learned translators T_{ij}.

Drift and path consistency

The figure lets us trace one report. The direct route sends $(0.6, 0.2)$ to $(0.7, 0.2)$. The other route passes through $(1.2, 0.1)$ and reaches the same point. Agreement checks whether both routes carry that input consistently. A constant encoder would pass this check while erasing every



Transition	*Formula*	*Selected point*
T12	(x, y) → (2x, y/2)	(0.6, 0.2) → (1.2, 0.1)
T13	(x, y) → (x + y/2, y)	(0.6, 0.2) → (0.7, 0.2)
T23	(u, v) → (u/2 + v, 2v)	(1.2, 0.1) → (0.7, 0.2)

Figure 3.2: Three computed representation spaces connected by explicit affine translators. Grids, a reference circle and one marked observation reveal the rescaling and shear. The transition register checks the common point by both the direct and two-step routes. This exact example illustrates the structure to be tested in a learned translator graph.

identity. Comparing two reports still requires an identity test on the translated outputs. Imperfect translators also introduce *drift*: small errors that accumulate as we move along edges.

The translator builder is keenly aware of this. They care about questions like:

- How far is $T_{\text{vid}\to\text{img}}(T_{\text{img}\to\text{vid}}(v))$ from v when v is an image embedding we consider in-regime?
- If we go from V_{img} to V_{vid} directly, and also via $V_{\log}$ (image → log → video), do those two paths agree up to a small tolerance?
- How does drift grow as we traverse longer and longer paths through the graph?

In the illustrative identity comparison, suppose we encode a photo of someone into V_{img}, translate it to video space via $T_{\text{img}\to\text{vid}}$, and then translate it back via $T_{\text{vid}\to\text{img}}$. If the translators are well-behaved and the photo is in-regime, the round-trip embedding $T_{\text{vid}\to\text{img}}(T_{\text{img}\to\text{vid}}(v))$ should land close to the original v. If the round-trip error is large, that tells us something is wrong: either the translators are poorly trained, or we are operating outside the regime where they are reliable, or a translation discards information the original report retains.

To keep things sane, translator builders design and audit translators so that:

- Round trips $V_i \to V_j \to V_i$ move in-regime vectors only a small amount.
- Different routes between the same pair of spaces produce similar results.
- The size of the drift can be summarized as a function of path length or path complexity—a *drift profile*.

In Chapter 9, we will see how to turn this intuitive notion of cost and drift along paths into an actual metric on a glued-together space. For now, the important thing is the picture: a graph of spaces and translators, with an implicit sense of how much semantic error we incur as we move around.

Sameness decisions in translator space

How does the translator builder actually decide if a still image and a video clip show the same person?

One reasonable strategy is:

1. Encode the image into $v_{\text{img}} = E_{\text{img}}(\text{photo})$.
2. Encode the video into $v_{\text{vid}} = E_{\text{vid}}(\text{clip})$.
3. Translate v_{img} into the video space via $T_{\text{img}\to\text{vid}}$ and measure the distance between $T_{\text{img}\to\text{vid}}(v_{\text{img}})$ and v_{vid}.
4. Optionally, translate v_{vid} into the image space and compare it with v_{img}. To test a round trip, send the same vector forward and back.
5. Combine these distances into a score and compare it to a threshold.

If the translators are well-behaved and the thresholds are chosen sensibly, this procedure can work quite well. From the translator builder's perspective, sameness decisions live in *translator space*: they are about whether the observed vectors can be made to agree across spaces, up to known drift.

A learned comparison in an existing system. FaceNet learns a shared 128-component representation and compares face images by their embedding distances. Its triplet training separates matching and nonmatching examples, and its outputs are normalized to unit length. The paper reports 99.63% on Labeled Faces in the Wild and 95.12% on YouTube Faces. Its harmonic-embedding construction also trains compatibility between network versions, a direct instance of the migration question raised here.[1]

This supplies an engineering connection to the translator account. Establishing a particular cross-channel translator graph still requires its actual maps and tests. The geometric structure of the normalized output also matters, as the next comparison makes explicit.

Character B: The Geometer

Now imagine a different kind of person reading the same problem statement. This person has less patience for format-specific details and a strong instinct to look for an underlying geometric object. We will call this person the *geometer*.

An underlying space of semantic states

The geometer suspects that, underneath all the embedding spaces and logs and formats, there is an underlying space of *semantic states*. For identity problems, that might be a space of identities; for viral evolution, a space of antigenic phenotypes; for system monitoring, a space of system states.

Let us call this space M. The geometer wants M to be more than just an abstract set. They want it to be a space where we can measure distances and speak sensibly about paths and curvature.

In the illustrative identity comparison, the geometer might imagine an *identity manifold* M whose points model facial states under different ages and expressions. One person may occupy several states over time, and different people may have similar states. The individual identity is a separate label that the system must infer from observations.

[1] F. Schroff, D. Kalenichenko and J. Philbin, "FaceNet: A Unified Embedding for Face Recognition and Clustering," CVPR 2015, Sections 3, 5 and Appendix 7, https://arxiv.org/abs/1503.03832.

Distance as a measure of change

Once the geometer has specified M, they need a *metric* on it: a way to measure the distance between states. In the Riemannian setting we will eventually adopt, this distance is defined in terms of the lengths of shortest paths, or *geodesics*, between points.

In the illustrative face comparison, a path can represent a change of expression, hairstyle or age. The metric measures changes in modeled facial state. A detector may use that distance as evidence about identity if its evaluation shows that the comparison separates same-person and different-person observations in the intended setting.

Now sameness decisions can be expressed in terms of distance and paths:

- If two observations map to points that are very close in M, along plausible paths, we treat them as the same identity.
- If they map to points that are far apart, or for which all connecting paths would require non-benign changes, we treat them as different.

The geometer thinks in terms of margins, balls, and geodesic neighborhoods. The details of the measurement process are complicated, but the core idea is simple: the geometry of M encodes what counts as small, benign change.

Charts and coordinates

Of course, we do not work directly with abstract manifolds in code. We work with coordinates: charts that map pieces of M into Euclidean spaces where we can do linear algebra.

The geometer imagines that there is an atlas of charts $\varphi_\alpha : U_\alpha \to \mathbb{R}^d$ covering M, each with a well-behaved inverse on its domain. Points in M are described in practice by their coordinates in some chart, and the geometry is encoded in how distances and angles look in those coordinates.

In the face example, two encoders might give coordinate descriptions of a region of the same modeled state space. That interpretation requires each description to be locally invertible on the region and their transitions to satisfy the chart conditions. The raw image and video encoders also have a separate job: estimating which state was observed.

When these conditions hold, the translator builder's maps between representations can agree with the geometer's chart transitions.

Sameness decisions in geometric space

How does the geometer decide if a still image and a video clip show the same person?

One reasonable strategy is:

1. Use the image encoder to estimate a point $p \in M$, represented in a chart.
2. Use the video encoder to estimate a point $q \in M$, possibly represented in a different chart.
3. Compute or approximate the geodesic distance $d_M(p, q)$ between p and q.
4. Compare $d_M(p, q)$ to a threshold or to a pair of thresholds that define a "same," "different," and "ambiguous" region.

If the geometry of M has been chosen and learned appropriately, this procedure can also work quite well. From the geometer's perspective, sameness decisions live in *geometric space*: they are about whether the two points can be connected by a short, benign path in M.

Real-world example: ArcFace and spherical geometry. ArcFace makes an angular margin on normalized features explicit in its training objective. This provides a concrete example of choosing the geometry of a learned comparison. Its 2019 account develops the margin as an arc on a hypersphere.

The key geometric insight: on a sphere, the natural notion of distance is *angular distance* (the angle between two points), which corresponds to geodesic distance along the sphere's surface. ArcFace:

1. Maps face images to unit vectors in $\mathbb{R}^{512}$, which lie on the 511-dimensional sphere $\mathbb{S}^{511}$
2. Computes the angular distance (geodesic) between two face embeddings using

$$d_{\text{geodesic}} = \arccos\left(\langle p, q\rangle\right),$$

   where $\langle p, q\rangle$ is the cosine similarity between unit-norm vectors p and q
3. Declares "same person" if the angle is below a threshold (embeddings point in nearly the same direction)

The reported ArcFace configuration reaches 99.83% on Labeled Faces in the Wild. That benchmark result belongs to its specified training and evaluation setup. Both FaceNet and ArcFace use normalized representations; their objectives and implementations differ. To interpret the result, specify the representation, the comparison rule and the conditions under which they were tested.[2]

The bench calculation can be checked from its conversion and ruler alone. These learned systems additionally require training and evaluation records. We retain that distinction while comparing the two implementation strategies.

Two Stories About the Same Job

So far, it might sound as though the translator builder and the geometer are living in different worlds. One cares about graphs of feature spaces and learned maps. The other cares about manifolds, geodesics, and charts.

In reality, they are trying to do the *same job*. Both are building systems that:

- Take heterogeneous observations drawn from multiple channels,
- Encode them into some internal spaces,
- And decide whether those encoded observations should be treated as the same latent entity evolving benignly, or as different entities.

If you need a concrete picture, you can think of the translator builder as someone who has a collection of flat maps of the same terrain, each drawn from a different perspective, and who checks sameness by seeing whether the maps can be reconciled along their edges. The geometer, by contrast, imagines there is an actual globe—a single, curved object—and checks sameness by measuring distances on that globe. Both are trying to describe the same underlying world. Exact overlap maps and compatible rulers recover the same geometry in either direction. With learned maps, the task is to construct a geometric model and bound how much it changes the comparisons we care about. Several models can satisfy the same bound.

To see the contrast more clearly, it helps to put their perspectives side by side.

[2] J. Deng, J. Guo, N. Xue and S. Zafeiriou, "ArcFace: Additive Angular Margin Loss for Deep Face Recognition," CVPR 2019, Section 2 and Table 4, https://arxiv.org/abs/1801.07698v3.

Translator builder vs. geometer

Table 3.1 sketches the contrast.

	Translator builder	**Geometer**
Basic objects	Feature spaces V_i, encoders E_i, translators T_{ij}	Latent manifold M, distance d_M, charts φ_α
View of channels	Different V_i for each channel; must be stitched together	Different charts or encoders into the same M
Main tool	Paths in a graph of spaces and translators	Paths (geodesics) in a curved space
Error notion	Drift along paths, round-trip error, cocycle inconsistencies	Distortion of charts, curvature, approximation error in d_M
Sameness test	Can we translate observations into a common space and make them agree, within drift budgets?	Are the mapped points close in M, along benign paths, within geometric margins?
Visual intuition	Graph of overlapping patches stitched along shared edges	Single curved surface with multiple coordinate systems

Table 3.1: Two engineering stories for the same job: translation-based and geometry-based realizations of semantic sameness.

In the illustrative identity comparison:

- The translator builder thinks about V_{img}, V_{vid}, $V_{\log}$, the encoders E_{img}, E_{vid}, $E_{\log}$, and the translators $T_{\text{img}\to\text{vid}}$, $T_{\text{vid}\to\text{img}}$, etc. They care about how much these translators drift and how consistent different paths are.
- The geometer thinks about an identity manifold M, about distances $d_M(p, q)$ between points on that manifold, and about charts that describe parts of M in coordinates. They care about how curved M is, how distorted the charts are, and what distance thresholds correspond to acceptable benign evolution.

Both are trying to operationalize the same semantic notion: "these observations come from the same underlying identity, evolving in benign ways" versus "these observations come from different identities."

Advantages and blind spots

Each perspective brings real strengths and real blind spots.

The translator builder's approach is often easier to deploy incrementally:

- It fits naturally into existing pipelines: you can add new translators between new and old feature spaces without redesigning everything.
- It works even when you are not sure that a single global geometry exists: you just need pairwise or local translators where you have data.
- It aligns with the way many organizations think about systems: as collections of services that translate between formats.

But the translator view can obscure global structure:

- It is not always obvious whether the patchwork of translators you have built is consistent with any underlying latent geometry.
- It can be hard to reason globally about paths, curvature, and what counts as "far" versus "near" in a semantic sense.

The geometer's approach, in contrast, makes global structure explicit:

- It gives you a single space M in which to reason about benign evolution, margins, and detection boundaries.
- It makes it possible, in principle, to transfer guarantees across different coordinate systems and implementations, because they are all supposed to be talking about the same geometry.

But the geometric view can be harder to fit to messy reality:

- It assumes that a reasonably smooth manifold structure actually exists and is a good model of the domain, which is not always obviously true.
- It often requires more up-front modeling work to relate real encoders and models to charts on M.

The central thesis of this book is that, under the right conditions, these two views are not competitors. They are two sides of the same coin: two ways of realizing the same semantic sameness structure S.

Realizations of a Semantic Sameness Structure

Let us connect the engineering accounts to the comparison structure specified in Chapter 2.

Recall that a semantic sameness structure S consists, in words, of:

- A latent world U of entities and benign trajectories $u(t)$.
- A family of observation channels that render U into different representation spaces.
- A decision problem about whether finite sets of observations could have come from the same u moving benignly.

A *realization* of S is a concrete system that tries to answer that decision problem in a way that respects the semantics encoded in S.

The inspection bench already gives a realization of S. The physical positions, permitted displacements and comparison task specify the problem. A converter and its distance rule implement one account; integrating the ruler implements the other. Different implementations may use different intermediate coordinates while answering the same measured question.

A musical score offers a secondary instance. Several performances can preserve a work's melody while differing in timing or instrumentation. Before assigning a distance between performances, we must choose which of those differences the task measures. In our bench problem that choice is explicit, the displacement in metres. We can therefore specify the components of each realization without losing the quantity they are meant to preserve.

Translation-based realizations

In a **translation-based realization**, which we will later organize into a category we call SamTrans(S), the key ingredients are:

- A collection of feature spaces V_i associated with the observation channels of S.

- Encoders E_i from raw observations in channel i into V_i.
- Translators $T_{ij} : V_i \to V_j$ that attempt to preserve semantic identity across channels.
- A way of measuring distances and costs along paths in the resulting graph of spaces and translators.
- A decision rule that declares observations to be "the same" if they can be connected, through this graph, in a way that is consistent with the benign evolution allowed by S and with the drift budgets of the translators.

Different translation-based systems for the same S may choose different feature spaces, different encoders, different families of translators, and different decision rules. They are distinct objects in SamTrans(S), but they are all trying to solve the same underlying sameness decision problem using observable, format-level machinery.

In the illustrative identity comparison, one translator-based realization might use a particular pair of image and video encoders and a single cross-modal translator. Another might use a richer graph of intermediate embedding spaces and multiple specialized translators. Both live in the same family.

Geometry-based realizations

In a **geometry-based realization**, which we will later organize into a category we call SamGeom(S), the key ingredients are:

- A latent space M that models the semantic states relevant to S, often endowed with a metric or Riemannian structure.
- Charts or encoders that map observations into (or near) points of M.
- A notion of benign paths in M that reflects the domain-specific understanding of benign evolution.
- A decision rule that declares observations to be "the same" if their images in M lie close to each other along benign paths, according to the geometry of M and the margins we set.

Different geometry-based systems for the same S may choose different latent spaces M, different metrics or Riemannian structures, different atlases of charts, and different margins and decision rules. They are distinct objects in SamGeom(S), but they are all trying to solve the same underlying sameness decision problem using a geometric model of the latent world.

In the illustrative identity comparison, one geometric realization might model identity as points on a high-dimensional sphere with a particular metric. Another might use a more complicated curved manifold learned from data. Both can live in the same family as long as they are faithful to the same semantic structure.

Same structure, different coordinates

A semantic sameness structure S specifies the common task. If the realizations satisfy the compatibility conditions developed later, their measurements can agree across different coordinate descriptions.

- Translation-based realizations express everything in terms of observable spaces and the transformations between them.
- Geometry-based realizations express everything in terms of an underlying space and the geometry on that space.

Both are valid ways to encode the same semantics, provided they satisfy certain regularity conditions and agree on the decision problem.

One of the main tasks of the rest of this book is to make that "provided" precise. We will:

- Construct the translator metric space (Chapter 9), examine overlap compatibility (Chapter 10), and establish the exact or specified approximate realization (Chapter 12).
- Show how, given a geometric realization with well-behaved charts, we can extract canonical translators between those charts and thus build a translation-based realization (Chapter 11).
- Show that, when this correspondence exists, the two realizations share essentially the same detection guarantees, up to small, explicit slack (Chapters 13–20).

Chapters 4–6 will build the geometric language we need to talk about manifolds, paths, and distances. Chapter 9 will revisit the translator builder's world and show how to turn it into a metric object. Chapters 10–12 will spell out the compatibility and Riemannian realizability assumptions that make the bridge between the two stories possible. Chapter 13 will wrap everything in the language of categories and functors, and Chapter 20 will connect those structures back to detection error budgets.

For now, the main thing to keep in view is that the translator builder and the geometer are not solving different problems. They are building different realizations of the *same* semantic sameness structure.

What the construction preserves

The two teams kept the same observation in view. The translator builder changed its coordinates; the geometer supplied a ruler that made the changed coordinates measure the same displacement. Their realizations of S agreed because the conversion, metric and comparison were specified together. The families $\mathrm{SamTrans}(S)$ and $\mathrm{SamGeom}(S)$ will organize those choices without replacing the observation they must explain.

The bench calculation makes the two routes agree through an explicit conversion and ruler. Extending that agreement beyond one coordinate line requires compatible local descriptions. We build those maps next, keeping the object distinct from every address assigned to it.

Practice with the construction

Back to the maps and the globe. A second chart does not add a second harbor. Nor does the mere existence of two charts make the harbor a double cover. We must specify which points represent the same place, how neighborhoods correspond, and which distances are preserved.

3.1. Use both routes to compare the bench readings $x = 2$ and $x' = 2.03$. Show every conversion, including the metric coefficient.

3.2. What additional statement would be needed before describing two representations of a document as a mathematical double cover?

Part II

The Riemannian foundations

Chapter 4

Coordinates without confusing the object

Two scales can mark every point on a straight rail. Try the same idea on a globe: a longitude scale becomes ambiguous at the poles. We will build two flat maps of the sphere and show how to identify the same point wherever both maps apply.

Locate a point near the pole

Put a mark near the north pole of a globe. Longitude becomes a poor address there, and at the pole every longitude names the same point. The mark itself has no such ambiguity. We need another local map, together with a conversion on the region where both maps apply.

The task tells us what the word manifold contributes. Each local address must identify points uniquely on an open region, neighboring addresses must represent neighboring points, and the regions must cover the object. Compatible smooth coordinates then allow us to differentiate the conversions. The definition records those requirements for the local descriptions we will construct.

Definition 4.1 (Manifold (informal))**.** An n-dimensional *manifold* is a space that is locally similar to $\mathbb{R}^n$. Around each point p, there is a neighborhood that can be mapped smoothly and invertibly to an open set in $\mathbb{R}^n$.

The dimension counts the independent coordinates needed locally. A circle needs one, a sphere two, and a torus also two. Their global arrangements differ. A tangent plane provides a local linear model; measuring an extended route requires the metric developed in the following chapter. We will keep the sphere on the desk while constructing its actual addresses and checking their overlap.

For a learned observation space, these requirements become modeling questions. Normalizing an embedding places it on a sphere, but observations need not fill that sphere or form a manifold of the same dimension. An encoder can serve as a chart on a modeled region when its domain, dimension and local invertibility support that role. Our explicit sphere construction gives a reference against which that proposed role can be assessed.

Why a globe asks for more than one map

To give the north pole an unambiguous address, we can look down on the globe from the south pole. To give the south pole an address, we reverse the setup. Each projection covers the point the other one misses.

For an explicit construction, use the unit sphere $S^2 \subset \mathbb{R}^3$. Remove the north pole and project from that pole onto the equatorial plane. A point $p = (X, Y, Z)$ gets coordinates

$$u = \frac{X}{1 - Z}, \qquad v = \frac{Y}{1 - Z}.$$

The inverse map is

$$p(u, v) = \frac{(2u, 2v, u^2 + v^2 - 1)}{1 + u^2 + v^2}.$$

A second projection from the south pole covers the missing point. Where both charts apply, their coordinates satisfy

$$(\tilde{u}, \tilde{v}) = \frac{(u, v)}{u^2 + v^2}.$$

This rule is smooth away from $(0, 0)$, exactly the point that maps to the pole missing from the other chart. We can inspect the domain rather than vaguely asserting that the maps agree.

At $(u, v) = (1, 0)$ the sphere point is $(1, 0, 0)$ and the second chart also reports $(1, 0)$. At $(2, 0)$ the second chart reports $(1/2, 0)$. Applying the transition twice returns $(2, 0)$. The two coordinate arrays differ, while the point on the globe is fixed.

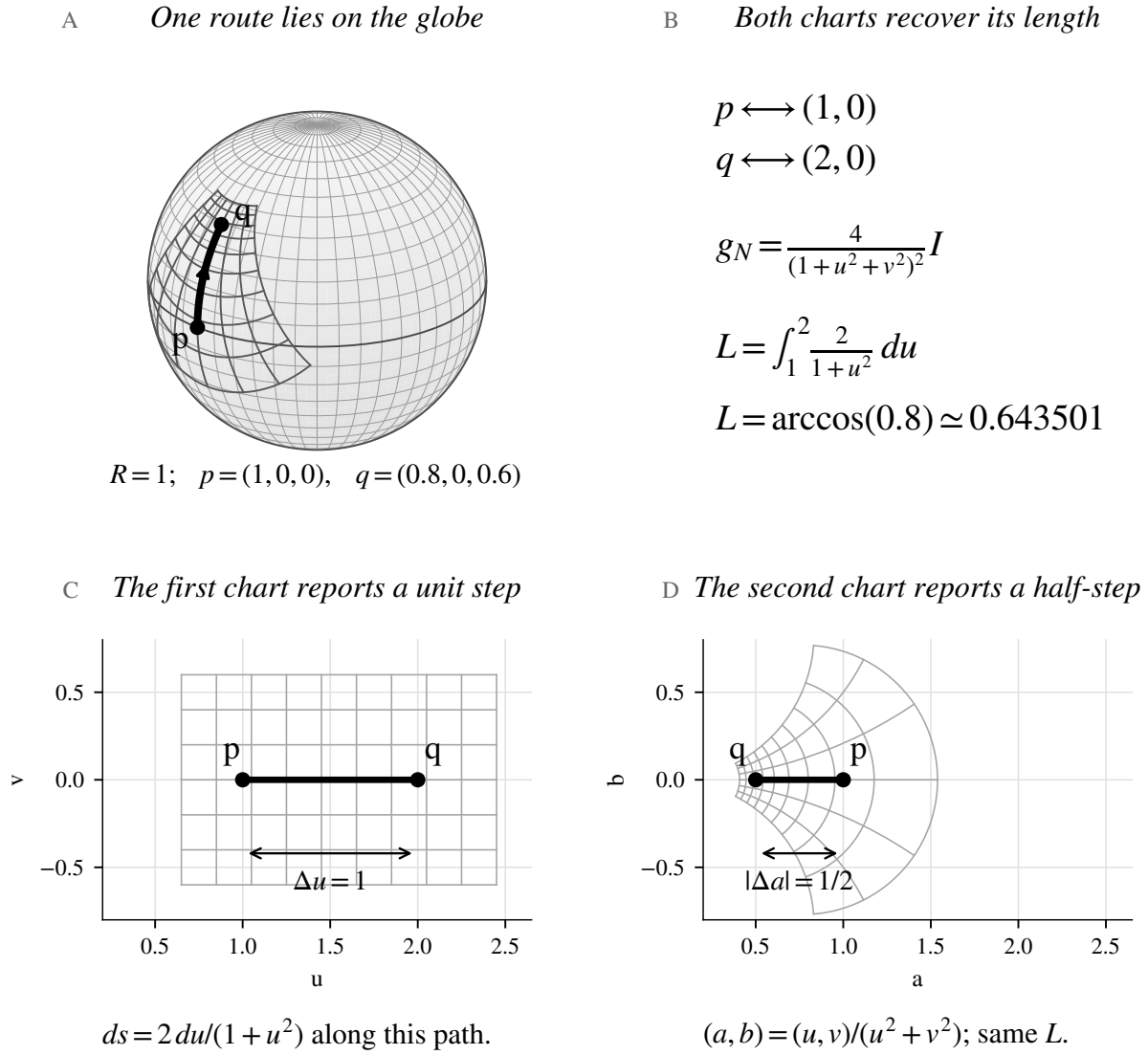


Figure 4.1: A computed path on the unit sphere and its two stereographic descriptions. The coordinate intervals have different lengths, but integrating their transformed rulers gives the same spherical length, $\arccos(0.8)$. The inversion $(a, b) = (u, v)/(u^2 + v^2)$ carries the entire coordinate grid between the charts.

The definition of a smooth manifold requires more than equations that look differentiable. The underlying topological space must be Hausdorff and second countable, locally homeomorphic to Euclidean open sets, with compatible smooth transitions. Hausdorffness separates distinct points by neighborhoods. Second countability gives a countable collection of neighborhoods from which

all open sets can be built. These hypotheses support a workable global space; local formulas alone need not supply them. Standard foundations are developed by Lee [3].

This chapter's globe is an actual sphere because we chose it. Later, the same atlas logic will describe orientation spaces and information-geometric models. The common feature is compatibility of local descriptions, not a claim that every model has the topology of Earth.

The stereographic formulas have identified the overlap. We now follow actual points and coordinate lines through it, so the same surface can be recognized on both maps.

Build the map before trusting its coordinates

A printed globe is curved, but a road atlas is made from flat pages. Each page must specify how its ink corresponds to locations on the globe. Stereographic projection gives that correspondence exactly. Its inverse shows how a flat coordinate patch can describe part of a curved surface.

Use the unit sphere $X^2 + Y^2 + Z^2 = 1$. Put a lamp at its north pole $N = (0, 0, 1)$ and project each other point onto the plane $Z = 0$ along the line from the lamp. The ray through (X, Y, Z) meets that plane at

$$(u, v) = \left(\frac{X}{1 - Z}, \frac{Y}{1 - Z} \right).$$

The north pole has no finite image. That is a missing point of this map, not a missing point of the globe. Near it, a tiny physical displacement can require a very large change of map coordinates.

Solve the ray and sphere equations in the other direction. The inverse map is

$$F_N(u, v) = \frac{1}{1 + u^2 + v^2}(2u, 2v, u^2 + v^2 - 1).$$

Substitution verifies that the image lies on the sphere. At $(0, 0)$ it gives the south pole. At $(1, 0)$ it gives $(1, 0, 0)$. As $u^2 + v^2$ becomes large, it approaches the north pole without reaching it at a finite coordinate value.

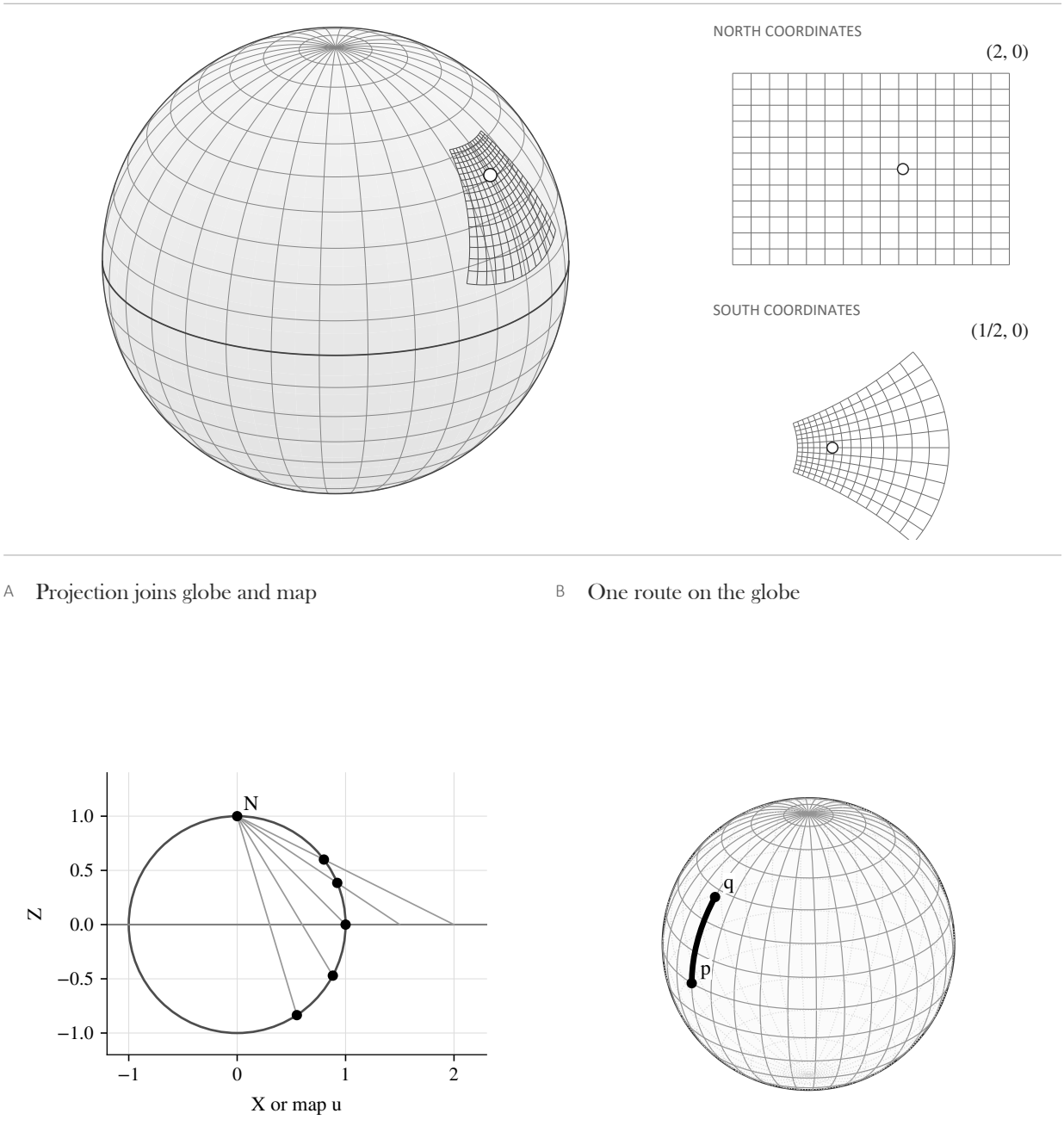
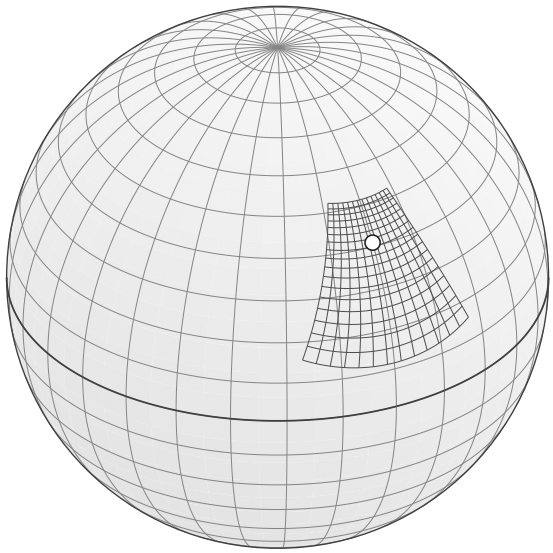
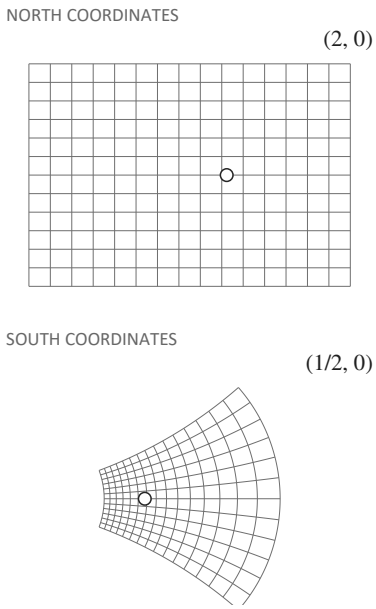


Figure 4.2: *A map is a construction, not a flattened globe.* The study continues on the following figure page.

One surface, two coordinate sheets

Projection changes the map; the material patch stays fixed.

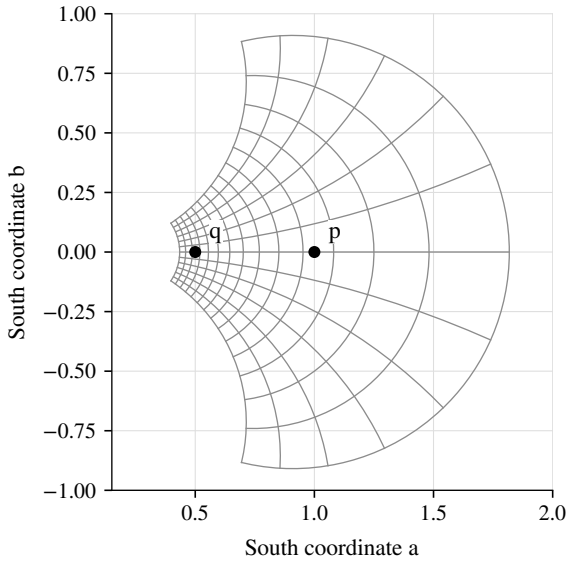




Same material patch / inspect the corresponding coordinates and rulers below.

C The second chart bends the grid

D The same points in both charts



North (1, 0) → (1, 0, 0)

North (2, 0) → (4/5, 0, 3/5)

South coordinates: (1,0), (1/2,0)

The north pole has no finite north-chart coordinate.

Figure 4.2: *A map is a construction, not a flattened globe (continued).* Projection rays, the omitted pole, the map plane and the inverse construction are drawn from the same sphere equations. Enlarged coordinate grids show how the projection amplifies small position errors near the omitted pole.

Make a second map using the south pole as the lamp:

$$(a, b) = \left(\frac{X}{1+Z}, \frac{Y}{1+Z}\right), \qquad F_S(a, b) = \frac{(2a, 2b, 1-a^2-b^2)}{1+a^2+b^2}.$$

On the overlap, where neither pole is present, the translator is

$$(a, b) = \frac{(u, v)}{u^2+v^2}.$$

It is its own inverse. The map sends a point near the center of one sheet to a point far from the center of the other. Both still name the same point on the sphere.

For the concrete point $p = (0.8, 0, 0.6)$, the north map reports $(2, 0)$ and the south map reports $(1/2, 0)$. Their Euclidean coordinate distances from the maps' origins are different because the origins are different physical poles. Comparing those distances as though they were measurements from the same place would be a mistake about the experiment, not a subtle theorem.

The overlap also translates a ruler

Let $r^2 = u^2 + v^2$. The transition Jacobian is

$$J = \frac{1}{r^4}\begin{pmatrix} v^2-u^2 & -2uv \\ -2uv & u^2-v^2 \end{pmatrix}.$$

It reverses the radial direction and rescales both orthogonal directions by $1/r^2$. Thus $J^T J = r^{-4} I$. The negative determinant is consistent with the two particular chart orientations chosen here; the transition remains a smooth invertible map wherever $r > 0$.

The sphere's metric in the north map is

$$g_N = \frac{4}{(1+r^2)^2} I.$$

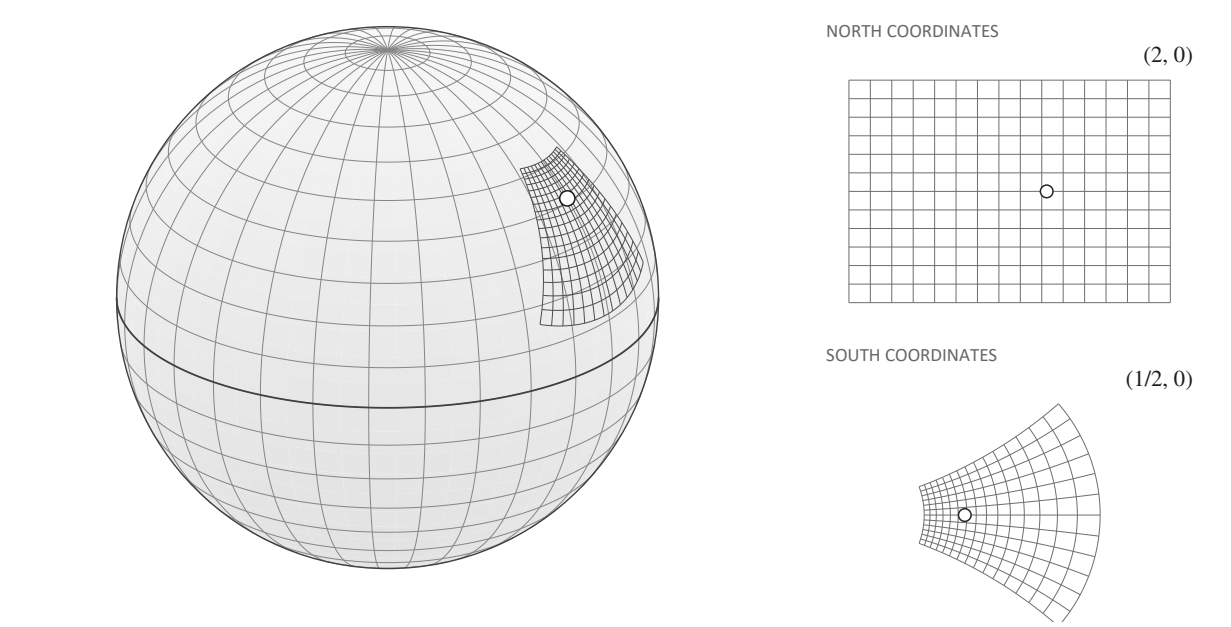
In the south map it is $g_S = 4I/(1+a^2+b^2)^2$. Since $a^2 + b^2 = 1/r^2$,

$$J^T g_S J = \frac{4}{(1+r^2)^2} I = g_N.$$

This equation is the ruler's compatibility check. Translate a small displacement by J, then measure it with g_S: the result equals measuring the original displacement with g_N. The coordinates changed; the length did not.

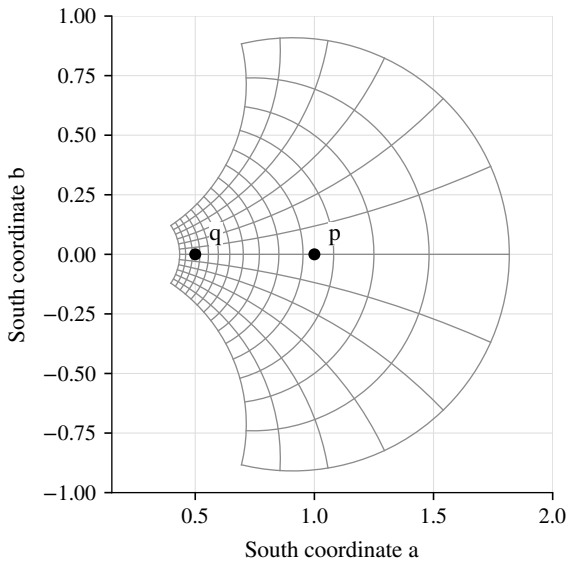
Carry the ruler through the map

A common patch exposes the change of coordinate scale.



A One overlap, two coordinate grids

B Equal physical steps look unequal



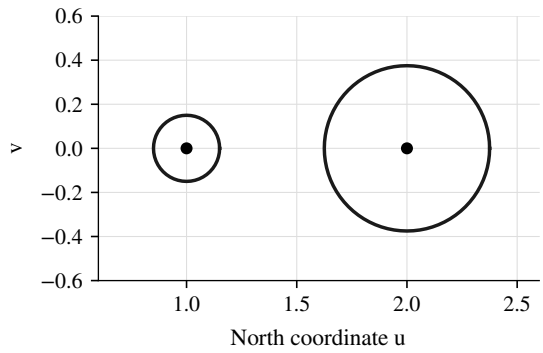
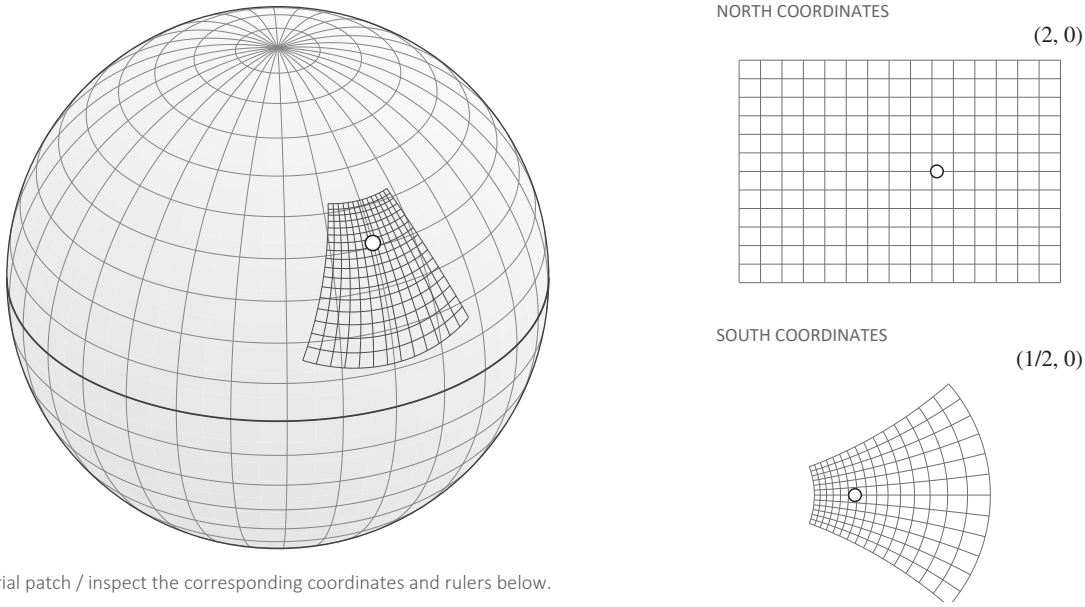


Figure 4.3: *The overlap must translate the ruler too.* The study continues on the following figure page.

Carry the ruler through the map

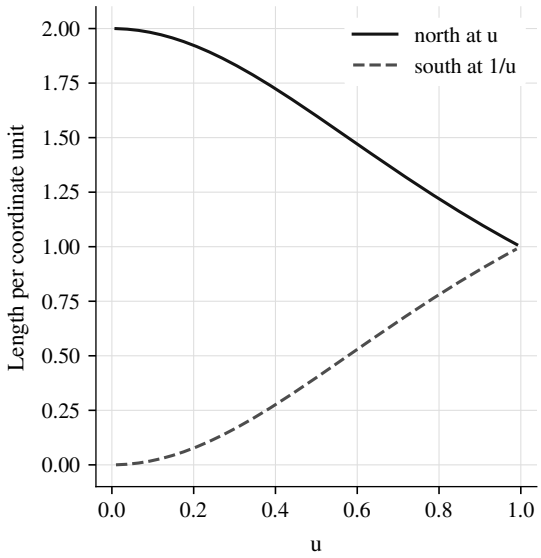
A common patch exposes the change of coordinate scale.



Same material patch / inspect the corresponding coordinates and rulers below.

C Each chart supplies its own scale

D Both rulers give the same length



J = diag(-1/4, 1/4)

North metric = (4/25) I

South metric = (64/25) I

J^T gS J = gN

The common route length is 0.643501.

Figure 4.3: *The overlap must translate the ruler too (continued).* Coordinate grids and equal-length ellipses are matched across stereographic charts. The numerical register evaluates the transition and both quadratic forms at the same marked point. The geometry is preserved by the Jacobian transformation, not by comparing raw coordinate differences.

One route, two integrals, one answer

Follow the meridian segment represented by $(u, 0)$ with $1 \le u \le 2$. Its sphere endpoints are $(1, 0, 0)$ and $(0.8, 0, 0.6)$. In the north chart,

$$L = \int_1^2 \frac{2}{1+u^2}\,du = 2\arctan(2) - \frac{\pi}{2} = \arccos(0.8).$$

The unit-sphere length is approximately 0.643501. In the south chart the route runs from $a = 1$ to $a = 1/2$, and

$$L = \int_{1/2}^1 \frac{2}{1+a^2}\,da = \frac{\pi}{2} - 2\arctan(1/2).$$

The two expressions agree because $\arctan(2) + \arctan(1/2) = \pi/2$. Their coordinate interval lengths, one and one half, do not agree. Multiplying by a physical sphere radius R gives a length in the units of R.

If I pass this route to someone using the south sheet, their coordinate runs from one down to a half. Mine runs from one up to two. The transition tells us which marks to join, and each sheet's ruler gives us the same distance to put in the survey log.

Where the chart should hand over

A numerical implementation can avoid the far end of either map. For example, use the north chart while $Z \le 1/2$ and hand over to the south chart before approaching the north pole. The overlap must be wide enough to make the handover smooth and to leave room for measurement uncertainty. A sharp change of stored coordinates is harmless if it is accompanied by the correct transition of vectors, metrics and uncertainty.

If a program silently keeps the old covariance matrix after changing charts, it has changed the uncertainty model. If it keeps old vector components, it has changed the physical vector. An atlas is therefore a set of compatibility rules for every quantity the calculation carries, not just a dictionary of point names.

Laboratory 1: build two maps of one globe

The aim is to make the map/globe analogy into an executable object. At the end of the laboratory, the same points, tangent vectors and route lengths must agree through two independent chart descriptions. A picture is one output; a table of residuals is another. Neither replaces the other.

Use a unit sphere and the north and south stereographic formulas in this chapter. Work first with rational coordinates so that several checks can be done exactly. Choose north-chart points $(1, 0)$, $(2, 0)$, $(1, 1)$ and $(1/2, 1/2)$. For each point calculate its sphere coordinates, its south-chart coordinates and the sphere point reconstructed from those south coordinates.

North coordinates	Sphere point	South coordinates
$(1, 0)$	$(1, 0, 0)$	$(1, 0)$
$(2, 0)$	$(4/5, 0, 3/5)$	$(1/2, 0)$
$(1, 1)$	$(2/3, 2/3, 1/3)$	$(1/2, 1/2)$
$(1/2, 1/2)$	$(2/3, 2/3, -1/3)$	$(1, 1)$

The tangent experiment. At $(u, v) = (1, 1)$ use a displacement $h = (0.02, -0.01)$. The transition derivative is

$$J = \begin{pmatrix} 0 & -1/2 \\ -1/2 & 0 \end{pmatrix}, \qquad Jh = (0.005, -0.01).$$

The north metric is $4I/9$; the south metric is $16I/9$. Both assign squared length $1/4500$ to the corresponding displacement. Check the equality by direct multiplication. Then compare the exact finite difference $T(x + h) - T(x)$ with Jh. Their difference is a nonlinear remainder, not a failure of the metric transformation law.

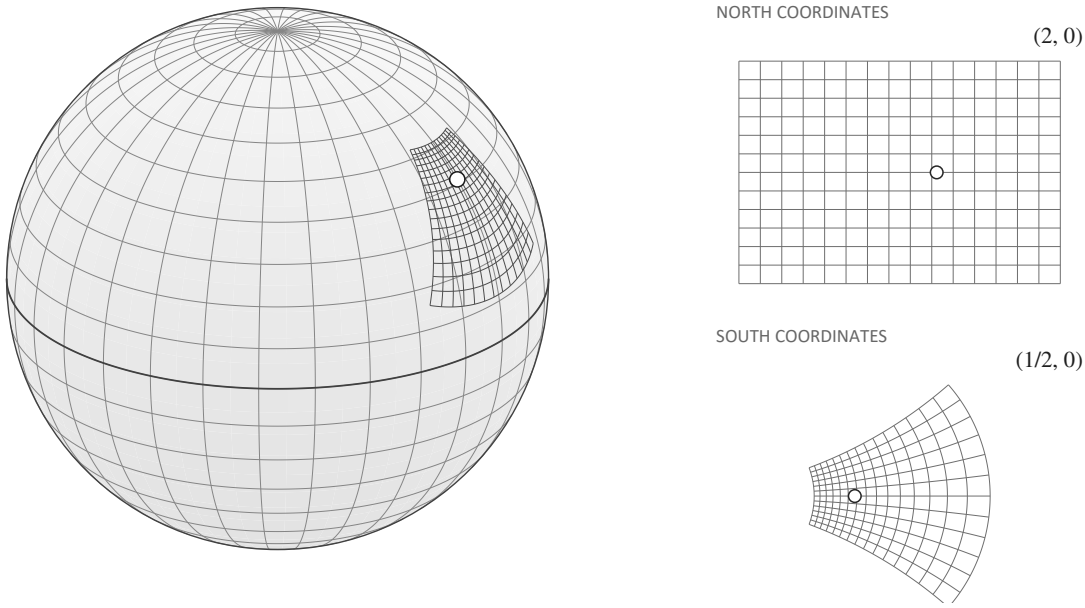
The route experiment. Integrate the meridian route $u \in [1, 2]$, $v = 0$ using the north ruler. Repeat in the south chart, where a runs backward from one to one half. Compare both numerical integrals with $\arccos(4/5)$. Refine the numerical integration until the residual stabilizes at the precision of the chosen method.

The deliberate fault. Repeat the calculation after setting both chart metrics to the identity. The coordinate lengths become one and one half. Preserve this failed result in the laboratory record. It demonstrates why a point translator alone does not preserve a geometric measurement.

The chart-boundary experiment. Choose points whose sphere heights approach $Z = 1$. Record the north coordinate radius $\sqrt{(1 + Z)/(1 - Z)}$ and the south radius, its reciprocal. At $Z = 0.99$, the radii are $\sqrt{199}$ and $1/\sqrt{199}$. Plot the same small spherical neighborhood in both charts. The stretched north-map picture shows why we switch charts before small physical errors become large coordinate errors.

A reproducible globe-and-map construction

The point register and local rulers can be checked independently.



A One route on the globe

B The second chart gives a different map

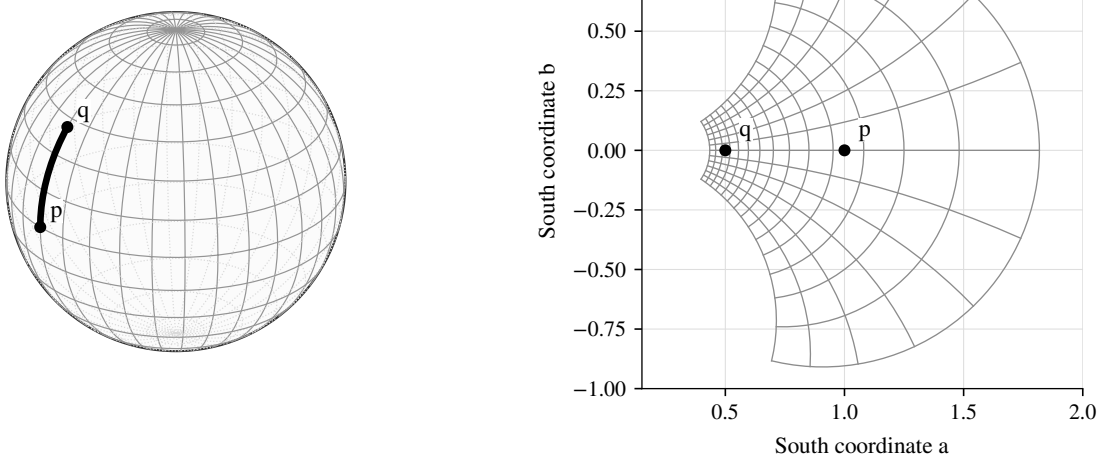
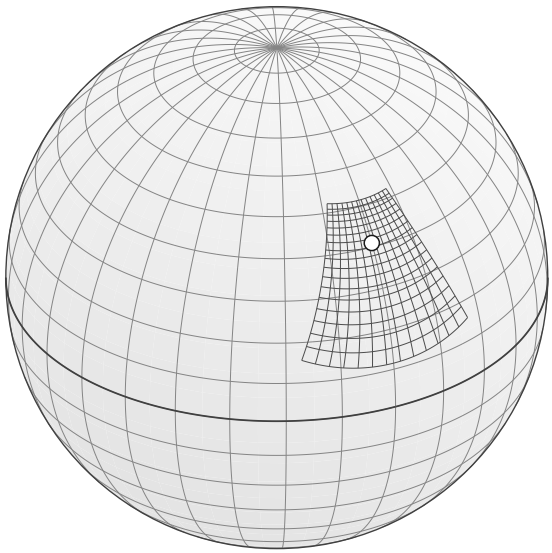


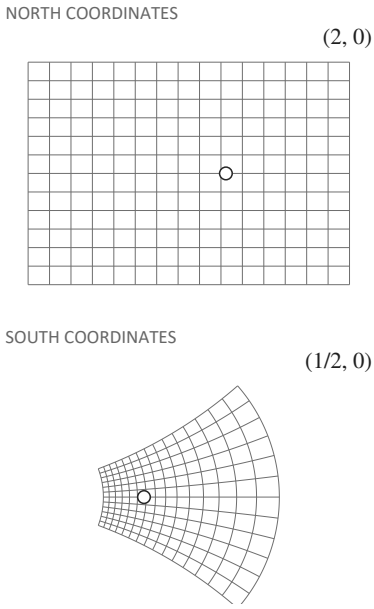
Figure 4.4: *Laboratory record: points, vectors and routes.* The study continues on the following figure page.

A reproducible globe-and-map construction

The point register and local rulers can be checked independently.



Same material patch / inspect the corresponding coordinates and rulers below.



C Four point translations agree exactly

(1,0) ↔ (1,0,0) ↔ (1,0)

(2,0) ↔ (4/5,0,3/5) ↔ (1/2,0)

(1,1) ↔ (2/3,2/3,1/3) ↔ (1/2,1/2)

(1/2,1/2) ↔ (2/3,2/3,-1/3) ↔ (1,1)

D Tangent and route checks test the same ruler

At (1,1), h=(.02,-.01)

Jh=(.005,-.01)

Both squared lengths = 1/4500

Both route lengths = arccos .8

Wrong identity rulers give 1 and 1/2.

Figure 4.4: *Laboratory record: points, vectors and routes (continued).* Four rational sphere points provide exact point and overlap checks. A tangent displacement and a complete route test the transformed ruler. The final panel records the intentionally wrong identity-metric calculation so that the consequence of omitting the ruler is visible.

What to submit. Keep the formulas, the four-row point table, the tangent quadratic forms, both route integrals and the chart-boundary plot. Report residuals with the numerical precision used. A successful record establishes this atlas construction and its tested arithmetic. It does not establish that every arbitrary data encoding is a smooth chart.

Further investigation. Replace the unit sphere by radius $R = 3$. Keep dimensionless chart coordinates and multiply the metric by R^2. Predict the new route length before calculating it. Then use length-valued plane coordinates instead and derive their metric. Explain why the two coordinate conventions give different coefficients and the same physical length.

The stereographic maps give the same route length because their rulers transform with their coordinates. Chapter 5 develops the general rule and shows how to choose a ruler when we measure uncertainty as well as physical length.

Practice with the construction

4.1. Compute both stereographic coordinate pairs of $(0, 1, 0)$ and verify their transition.

4.2. Explain why the undefined longitude of a pole does not make the pole a singular point of the sphere.

Chapter 5

The metric is the rule of measurement

Print a map of a survey site, then stretch a second copy horizontally. The distance between two survey stakes on the ground stays fixed even though the printed separation changes. What ruler should we use on each map to recover the same distance?

Distances and Metric Spaces

The surveyor needs both the coordinates of the stakes and a rule for measuring their separation. In a detector, the same distinction applies to two observations: the numerical difference in an encoder's output is useful only when we know what that difference measures.

We begin with the rules a distance must satisfy. They let us combine several comparisons without changing the meaning of the total. A chain of small changes can accumulate, but its endpoint distance cannot exceed the sum of the steps. That is the triangle inequality. It will later let us measure a route assembled from several maps or converters.

A metric space supplies these rules for complete distances. A Riemannian metric supplies a smoothly varying ruler for small displacements, from which route lengths can be calculated. We will first construct that ruler for the surveyor, then derive another from a specified statistical experiment.

One survey, several routes

Place three stakes A, B, C on the survey site. Any route from A to B followed by a route from B to C is also a route from A to C. The cheapest allowed direct comparison can therefore cost no more than that concatenation:

$$\text{distance}(A, C) \leq \text{distance}(A, B) + \text{distance}(B, C).$$

This triangle inequality is the rule that makes a chain of measurements useful. It bounds the separation of the endpoints using the steps we actually measured. The other metric axioms specify when a distance vanishes and whether reversing the endpoints changes it.

Metric spaces

A *metric space* is one of the simplest abstractions in mathematics that still has a meaningful notion of distance.

Definition 5.1 (Metric space)**.** A *metric space* is a pair (X, d) where X is a set and $d : X \times X \to [0, \infty)$ is a function, called a *metric*, such that for all $x, y, z \in X$:

1. **Non-negativity and identity:** $d(x, y) \geq 0$, and $d(x, y) = 0$ if and only if $x = y$.
2. **Symmetry:** $d(x, y) = d(y, x)$.
3. **Triangle inequality:** $d(x, z) \leq d(x, y) + d(y, z)$.

If you like, you can read this definition as saying: d behaves like a reasonable idea of distance.

Examples

It is worth seeing a few concrete examples.

- **Euclidean space.** Take $X = \mathbb{R}^n$ and define
$$d(x, y) = \sqrt{(x_1 - y_1)^2 + \cdots + (x_n - y_n)^2}.$$
This is the usual straight-line distance between points in n-dimensional space.
- **Manhattan distance.** Still on $X = \mathbb{R}^n$, define
$$d(x, y) = |x_1 - y_1| + \cdots + |x_n - y_n|.$$
This is the distance you would travel on a grid of city blocks if you can only move horizontally and vertically.
- **Graphs.** Let X be the set of vertices of a connected graph. Define $d(x, y)$ to be the length of the shortest path (fewest edges) between x and y. This satisfies all three metric axioms.
- **Embeddings.** In a modern ML system, an embedding space V with a norm $\|\cdot\|$ becomes a metric space by defining $d(v, w) = \|v - w\|$. For example, if we use the Euclidean norm in a face embedding space V_{img}, defining $d(v, w) = \|v - w\|$, then (V_{img}, d) is a metric space. This is exactly the setting the translator builder works in.

Later, we will construct a metric space by *gluing* together several feature spaces along translator edges and defining distances as shortest path lengths in the resulting graph. That construction lives entirely within the language of metric spaces.

Metric spaces as a unifying language

Metric spaces are simple enough to be very general, but structured enough to be useful. They cover:

- The local distances that encoders define in individual feature spaces.
- The global distances we will derive from translator graphs (Chapter 9).
- The geodesic distances we will use on manifolds (later in this chapter).

The key point is that once we have a metric, we can talk about *paths*, *lengths*, and *shortest paths* without committing to any particular coordinate system. That is exactly what we need to connect translator graphs and manifolds.

Riemannian Manifolds: Metrics on Tangent Spaces

Manifolds tell us about local Euclidean structure, but they do not yet tell us how to measure distances and lengths. For that, we need a *Riemannian metric*.

Tangent spaces and measured velocities

At each point p on a smooth manifold M, there is a *tangent space* T_pM. You can think of T_pM as the space of all possible directions you can move in if you start at p and take an infinitesimally small step.

A useful picture is the Earth again. Think of standing on the surface of the Earth at point p. The tangent space T_pM is like the flat plane that just barely touches the Earth at your feet. Any direction you can walk (north, south, east, west, or any combination) corresponds to a tangent

vector in that plane. As you walk a tiny amount in some direction, you are following a tangent vector. The tangent space is where velocity vectors live.

A curve through p gives a tangent vector through its velocity there. All such vectors form the tangent space T_pM, a vector space attached to p. Equivalent definitions use derivations or classes of curves.

Riemannian metrics as inner products

A surveyor can locate a stake by its distance from the instrument and its bearing. A small change in bearing moves a distant target farther than a nearby one. The local ruler must account for the position and direction of the displacement. More generally, an instrument model may weight changes across a road differently from changes along it. A positive-definite inner product supplies that directional ruler while assigning equal length to a displacement and its reverse at the same point.

Definition 5.2 (Riemannian metric (informal))**.** A *Riemannian metric* on a smooth manifold M assigns to each point $p \in M$ an inner product

$$g_p : T_pM \times T_pM \to \mathbb{R}$$

such that g_p varies smoothly with p.

An inner product is just a way to measure angles and lengths in a vector space. Given an inner product g_p on T_pM, we can define the length of a tangent vector $v \in T_pM$ by

$$\|v\|_p = \sqrt{g_p(v, v)}.$$

For an observation model, choosing the metric declares which changes count strongly in the comparison. A permitted change in viewing conditions might receive less weight than a change needed to distinguish the objects. The application must justify those weights and specify any forbidden transitions separately. The resulting ruler then has a precise job: measure the changes the task has chosen to compare.

On a Riemannian manifold (M, g), we therefore have:

- A notion of length for tangent vectors at each point.
- A way to measure the length of curves by integrating the lengths of their velocity vectors.
- A notion of geodesic distance defined as the infimum of curve lengths between points.

Length of curves on a Riemannian manifold

Let (M, g) be a Riemannian manifold. Suppose $\gamma : [0, 1] \to M$ is a smooth curve. At each time t, the velocity $\dot{\gamma}(t)$ is a tangent vector in $T_{\gamma(t)}M$, and we can measure its length using $g_{\gamma(t)}$.

The *Riemannian length* of γ is defined as

$$L(\gamma) = \int_0^1 \|\dot{\gamma}(t)\|_{\gamma(t)}\, dt = \int_0^1 \sqrt{g_{\gamma(t)}(\dot{\gamma}(t), \dot{\gamma}(t))}\, dt.$$

This looks more complicated than the metric-space definition of path length we gave earlier, but conceptually it is the same idea: cut the curve into small pieces, measure the length of each piece, and sum them up, with the metric g telling us how long each infinitesimal piece is.

The velocity $\dot{\gamma}(t)$ describes the survey route in its current coordinates. The point $\gamma(t)$ identifies which local ruler $g_{\gamma(t)}$ to use. Feeding the velocity into that quadratic form gives squared measured speed. The square root gives speed, and integration gives the traveled length. A change of coordinates affects the velocity components and the matrix together.

Measure the survey vehicle's journey. Follow a survey vehicle along the measured route. At each instant t, it has:

- A *position* $\gamma(t)$ (where you are on the road)
- A *velocity* $\dot{\gamma}(t)$ (your speed and direction at that instant)

The speedometer shows $\|\dot{\gamma}(t)\|$. For a one-hour trip, measure t in hours and speed in miles per hour.

To find the total length of the journey, you integrate your speed over time:

$$\text{total distance} = \int_0^1 (\text{speedometer reading})\, dt.$$

The metric converts coordinate velocity into measured speed. A constant rate of change in the displayed numbers need not mean constant physical speed: the coordinates may stretch differently along the route. This happens even on a flat surface described by a distorted grid. Integrating the metric speed keeps the answer attached to the journey.

Geodesic distance on a Riemannian manifold

Once we know how to measure the length of curves, we can define the *geodesic distance* between two points $p, q \in M$ by

$$d_M(p, q) = \inf_\gamma L(\gamma),$$

where the infimum is taken over all smooth curves γ with $\gamma(0) = p$ and $\gamma(1) = q$.

On each connected Riemannian component, this d_M is a metric. A smooth length-minimizing curve, parametrized with constant speed, is a geodesic; a geodesic need not minimize length over an arbitrarily long interval. Chapter 6 examines the route and existence questions.

Examples.

- On a flat plane with the standard Euclidean metric, geodesics are straight lines, and geodesic distance is just straight-line distance.
- On the surface of a sphere (like the Earth) with the standard round metric, geodesics are great-circle arcs, and geodesic distance is the length of the shortest great-circle path between two points.
- On our identity manifold with an appropriate Riemannian metric, geodesic distance between two identity states measures the minimum semantic change needed to move from one to the other along benign directions: small values for "same person over time", large values for genuinely different people.

In our semantic setting, the identity manifold M (or its analogues in other domains) will be treated as a Riemannian manifold. Benign evolution corresponds to moving along curves of small geodesic length, and sameness decisions correspond to asking whether two points are close in d_M.

The infimum asks for the greatest lower bound of the lengths of the allowed curves with those endpoints. It can exist even when no curve attains it. Chapter 6 will use this distinction to examine restricted routes and missing boundaries. For now, the definition tells us what must be held fixed when two coordinate systems claim the same endpoint distance.

For the surveyor, the distinction is concrete. A chord through a globe and a route along its surface connect the same endpoints but belong to different path families. The ruler and the permitted routes determine which length answers the question. On a plane with its Euclidean ruler, a straight segment minimizes length; on a round sphere, the shorter great-circle arc does so.

For an observation model, the metric likewise measures declared changes along admissible paths. Predicting the path a real system actually follows requires its dynamics as well. We will keep those roles separate while calculating how two maps can measure one route identically.

The ruler changes when the grid changes

A surveyor can locate a stake by east/north coordinates or by range and bearing. A one-metre change in range is always a one-metre radial displacement. A one-radian change in bearing sweeps a distance that grows with range. Treating the two coordinate increments as equally long would misread the instrument.

Let $x = r\cos\theta$, $y = r\sin\theta$, with $r > 0$ on a region where the angle has a continuous branch. The derivative is

$$J = \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix}, \qquad J^{\mathsf{T}}J = \begin{pmatrix} 1 & 0 \\ 0 & r^2 \end{pmatrix}.$$

Thus the ordinary Euclidean ruler becomes

$$ds^2 = dr^2 + r^2 d\theta^2.$$

At $r = 2$ metres, changing the bearing by 0.01 radians along the circle travels 0.02 metres. At $r = 20$ metres, the same angular change travels 0.2 metres. The coordinate increment stayed the same; the physical displacement did not.

Proposition 5.3 (Coordinate invariance of length)**.** *Let $T : U \to V$ be a diffeomorphism of coordinate domains, and let their positive-definite metric matrices satisfy $G_U(x) = DT_x^{\mathsf{T}} G_V(T(x)) DT_x$. Then every piecewise C^1 curve γ in U has the same length as $T \circ \gamma$ in V. The corresponding intrinsic distances agree when the same allowable curves are transported by T.*

Proof. The chain rule gives $(T \circ \gamma)' = DT_\gamma \dot\gamma$. Substituting into the squared speed gives the same quadratic form on both sides. Taking square roots and integrating proves equality of curve lengths. Since T and T^{-1} carry the allowable curve families onto each other, their length infima agree. □

The final condition matters. A chart may describe only a small region. A shorter route can leave it. Equality of distances computed with curves confined to that region does not automatically establish a claim about every path on the entire globe.

Different units can also produce misleading matrix condition numbers. A range measured in metres and a bearing measured in radians require the metric before their lengths can be compared. Rescaling either coordinate changes the numerical matrix. When reporting conditioning, state the coordinate units or compare one metric against another through generalized eigenvalues. That comparison has a geometric meaning independent of a simultaneous coordinate change.

The surveyor's ellipse

Two sensors on a survey truck estimate its position along and across a road. Suppose the cross-road reading is noisier. An equal-confidence region can extend farther across the road than along it. An ellipse lets us specify those two scales and the directions they apply to.

For a positive-definite 2×2 matrix G, the equation $h^T G h = 1$ is an ellipse. If $G = Q\,\mathrm{diag}(\lambda_1, \lambda_2) Q^T$, its principal directions are the columns of Q and its semiaxes are $1/\sqrt{\lambda_1}$ and $1/\sqrt{\lambda_2}$. A large metric eigenvalue makes motion in that direction expensive and the corresponding unit-ball axis short.

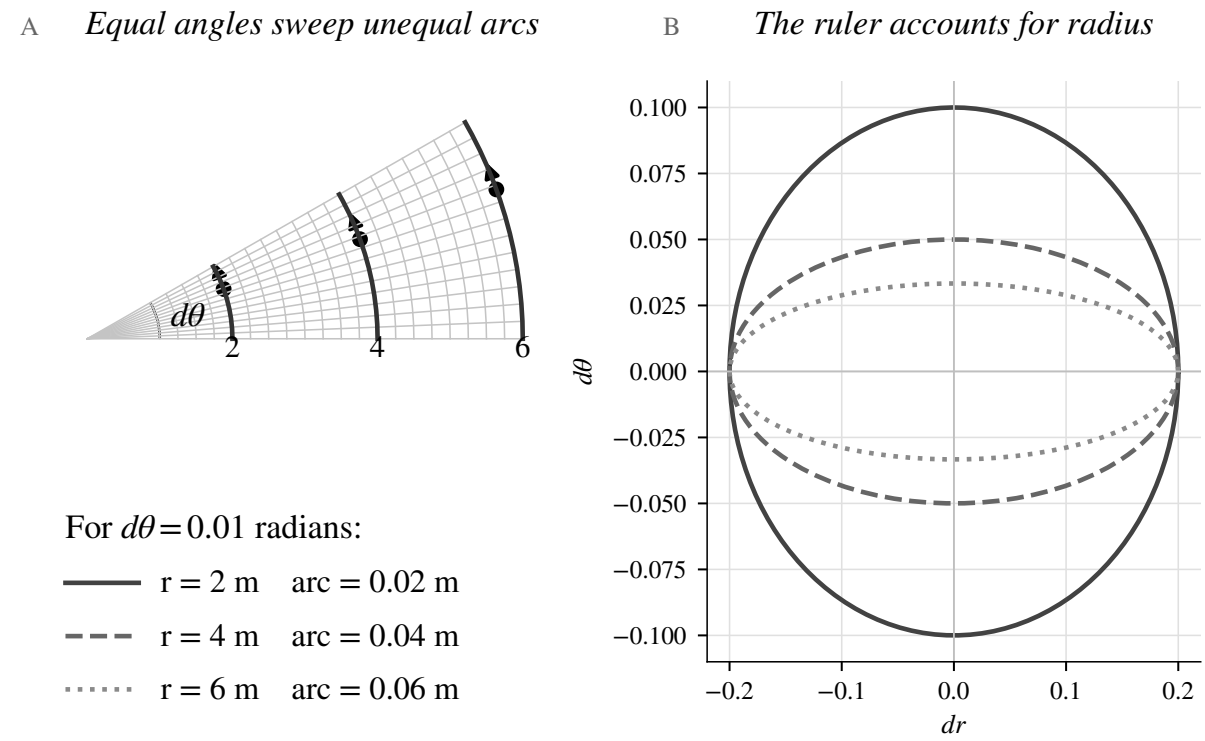
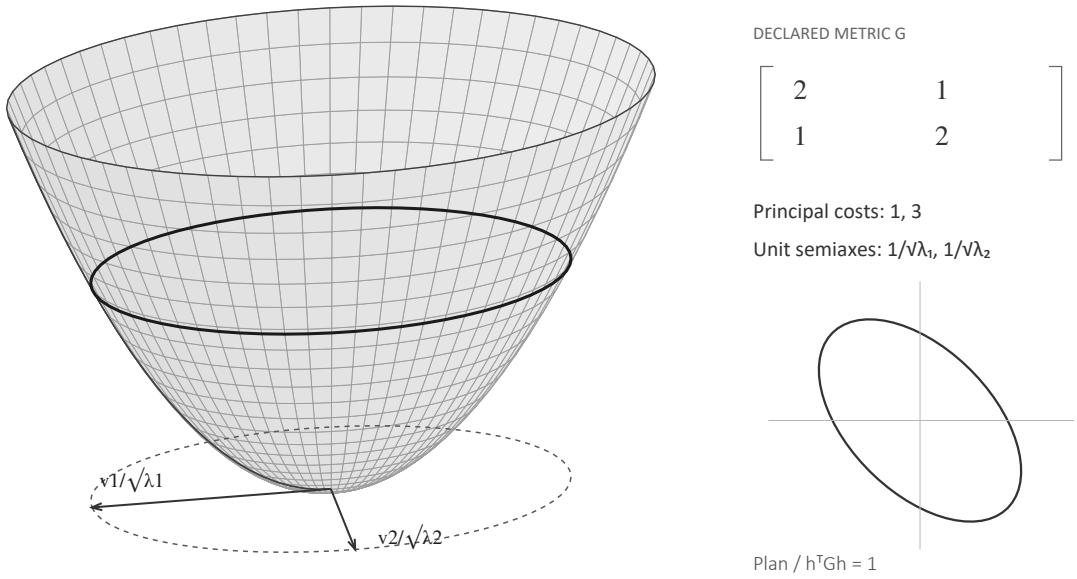


Figure 5.1: Equal angular steps at different ranges represent different physical lengths. The tangent arrows show direction; the circular grids show the range dependence. At right, equal-length ellipses in the coordinate increments satisfy $dr^2 + r^2 d\theta^2 = 0.2^2$. The metric coefficient restores the physical ruler.

The matrix alone does not say why the direction is expensive. It might measure physical length after a coordinate transformation. It might be inverse covariance in a statistical comparison. It might be a task-defined penalty. Those interpretations require different evidence even when the displayed ellipse looks identical.

The ruler is visible in its level sets

Constructed cost surface: height = $h^T G h$. The bold section has unit cost.



A Equal cost forms an ellipse

B The same step costs more in some directions

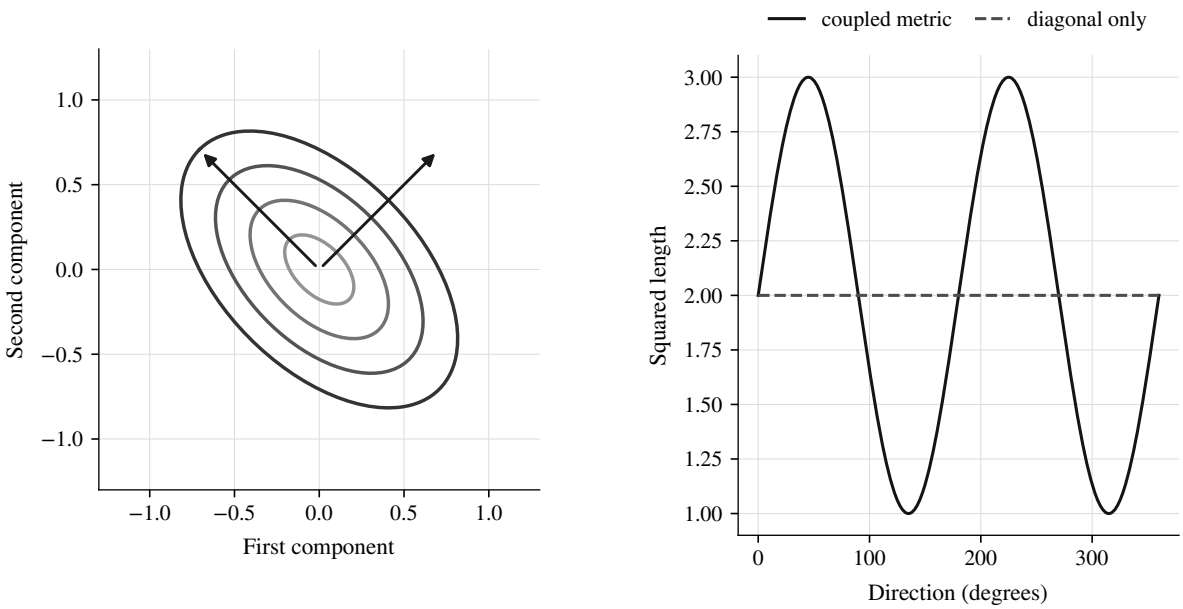
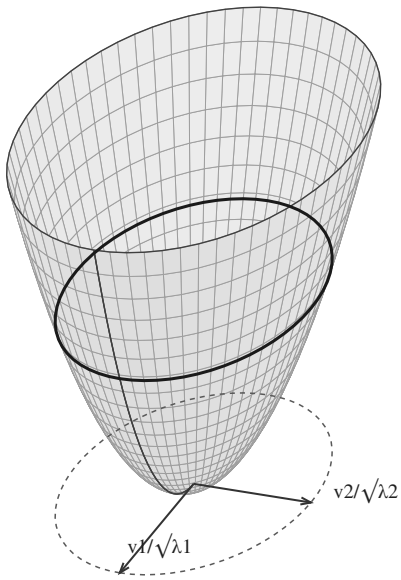


Figure 5.2: *What the local ellipse measures.* The study continues on the following figure page.

The ruler is visible in its level sets

Constructed cost surface: height = $h^T G h$. The bold section has unit cost.

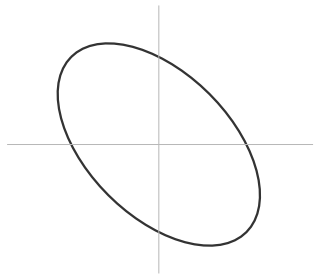


DECLARED METRIC G

$$\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

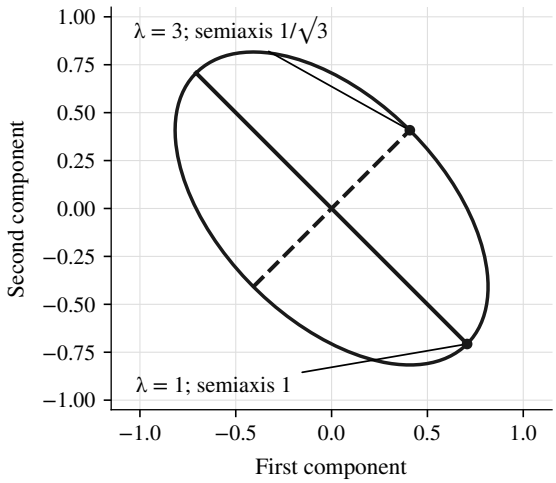
Principal costs: 1, 3

Unit semiaxes: $1/\sqrt{\lambda_1}$, $1/\sqrt{\lambda_2}$



Plan / $h^T G h = 1$

C Principal axes reveal the two costs



D The ruler needs a physical interpretation

RULER	*WHAT IT MEASURES*
Physical pullback	Length after an observation map
Inverse covariance	Discrepancy relative to noise
Task metric	A declared comparison cost

$v = (1, 1), \quad G = [2\ 1;\ 1\ 2], \quad v^T G v = 6$

The number 6 alone does not choose the interpretation.

Figure 5.2: *What the local ellipse measures (continued).* For $G = \left[\begin{smallmatrix} 2 & 1 \\ 1 & 2 \end{smallmatrix}\right]$, equal-cost contours and directional costs are computed from the same quadratic form. The principal semiaxes have lengths 1 and $1/\sqrt{3}$. The final table separates a physical length, a noise-scaled discrepancy and a declared task cost; the shape of an ellipse alone does not choose its interpretation.

Changing units without changing the experiment

Suppose an observation contains a length x in meters and a time t in seconds. A simple dimensionless comparison is

$$ds^2 = \frac{dx^2}{\ell^2} + \frac{dt^2}{T^2},$$

where ℓ and T are declared scales. Switching from meters to centimeters gives $x' = 100x$. The metric coefficient must become $1/(100\ell)^2$ in the new coordinate. Then $dx'^2/(100\ell)^2 = dx^2/\ell^2$.

Keeping the old coefficient would multiply the spatial term in ds^2 by ten thousand. A normalization chooses a ruler. Its coefficients need a physical or statistical basis and must transform with the units.

For a general invertible linear change $y = Jx$, the metric in the new coordinates is $G_y = J^{-T} G_x J^{-1}$. The point displacement changes by $dy = J dx$, while the quadratic form remains $dy^T G_y dy = dx^T G_x dx$. Covariance transforms in the opposite tensor pattern, $\Sigma_y = J \Sigma_x J^T$. When $G = \Sigma^{-1}$, the two rules agree.

The surveyor can now change units and display grids without changing the measured route. Laboratory 2 makes that claim executable before we ask how a ruler can be derived from uncertainty itself.

Laboratory 2: measure a route with transformed rulers

Build a simple survey map with a deliberate shear. Physical coordinates are $x = (x_1, x_2)$ with the Euclidean ruler. Display coordinates are $y = Ax$, where

$$A = \begin{pmatrix} 2 & 1 \\ 0 & 1 \end{pmatrix}, \qquad A^{-1} = \begin{pmatrix} 1/2 & -1/2 \\ 0 & 1 \end{pmatrix}.$$

The display stretches and tilts the grid. Derive its metric rather than assigning the identity:

$$G_y = A^{-T} A^{-1} = \begin{pmatrix} 1/4 & -1/4 \\ -1/4 & 5/4 \end{pmatrix}.$$

Measure three displacements. Use $h_1 = (1, 0)$, $h_2 = (0, 1)$ and $h_3 = (1, 1)$. Their displayed vectors are $(2, 0)$, $(1, 1)$ and $(3, 1)$. The physical squared lengths are $1, 1, 2$; the metric in display coordinates must reproduce all three. A check on only the first coordinate direction would miss the shear term.

Measure a curve. Use the quarter circle $x(t) = (\cos t, \sin t)$, $0 \le t \le \pi/2$. Its physical length is $\pi/2$. Compute $y(t)$ and $\dot{y}(t)$, then integrate $\sqrt{\dot{y}^T G_y \dot{y}}$. The integrand should simplify to one. The ordinary Euclidean length of the displayed curve is different; calculate it as a separate diagnostic.

Translate uncertainty. Start with $\Sigma_x = \mathrm{diag}(0.01, 0.04)$. Compute

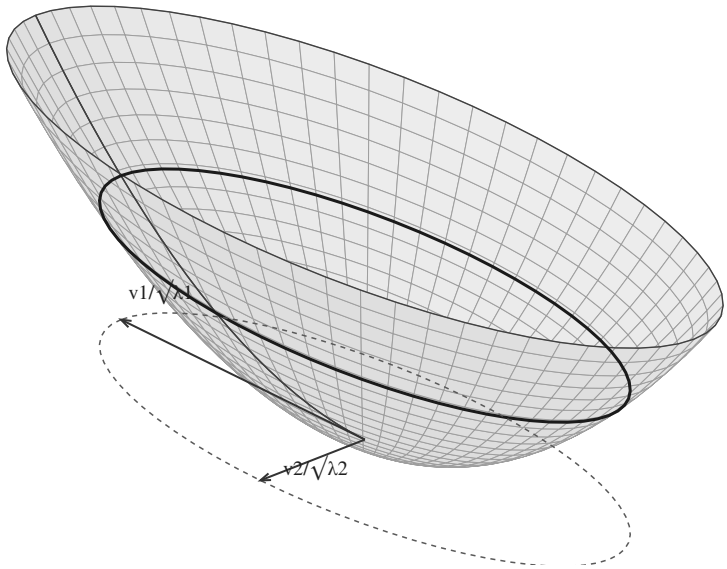
$$\Sigma_y = A \Sigma_x A^T = \begin{pmatrix} 0.08 & 0.04 \\ 0.04 & 0.04 \end{pmatrix}.$$

Verify $\mathrm{tr}(G_y \Sigma_y) = 0.05 = \mathrm{tr}\, \Sigma_x$. Draw both covariance ellipses with their respective coordinate grids. The displayed ellipse has changed shape; the metric-weighted mean squared physical error has not.

Replace a calibration with a task change. Now keep the same coordinates but intentionally double the cost of physical motion along x_1. The physical metric becomes $\mathrm{diag}(4, 1)$. Derive its new display matrix and calculate the same three displacements. Explain why this change should alter the answers, whereas the earlier coordinate shear should not.

The ruler is visible in its level sets

Constructed cost surface: height = $h^{T}Gh$. The bold section has unit cost.

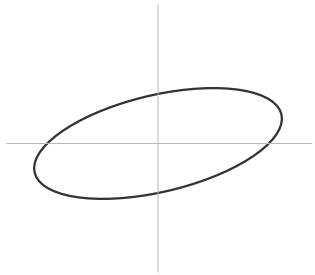


DECLARED METRIC G

$$\begin{bmatrix} 0.25 & -0.25 \\ -0.25 & 1.25 \end{bmatrix}$$

Principal costs: 0.19098, 1.309

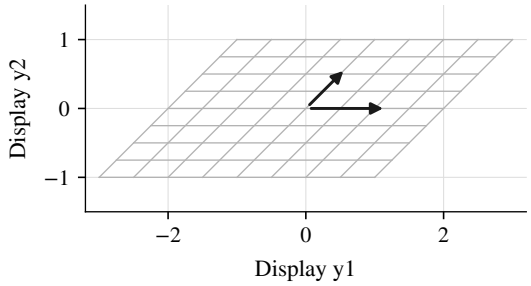
Unit semiaxes: $1/\sqrt{\lambda_1}$, $1/\sqrt{\lambda_2}$



Plan / $h^{T}Gh = 1$

A The shear changes the displayed displacement

B The transformed ruler preserves length

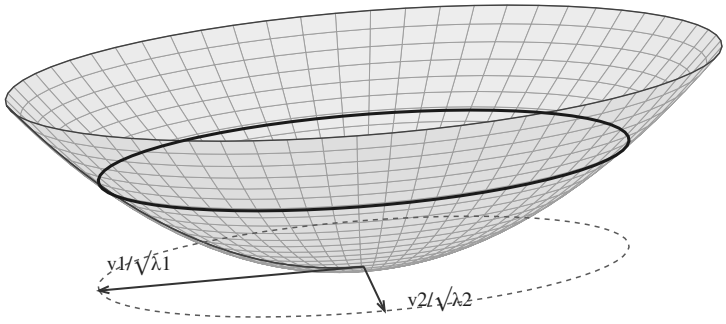


$$\begin{bmatrix} 0.25 & -0.25 \\ -0.25 & 1.25 \end{bmatrix}$$

Figure 5.3: *Laboratory record: the shear changes the map.* The study continues on the following figure page.

The ruler is visible in its level sets

Constructed cost surface: height = $h^T G h$. The bold section has unit cost.

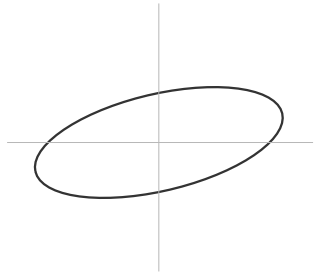


DECLARED METRIC G

$$\begin{bmatrix} 0.25 & -0.25 \\ -0.25 & 1.25 \end{bmatrix}$$

Principal costs: 0.19098, 1.309

Unit semiaxes: $1/\sqrt{\lambda_1}$, $1/\sqrt{\lambda_2}$



Plan / $h^T G h = 1$

C Three displacement checks agree

(1,0) → (2,0): squared length 1

(0,1) → (1,1): squared length 1

(1,1) → (3,1): squared length 2

Quarter-circle length stays π/2.

D Weighted uncertainty survives the translation

Σx=diag(.01,.04)

Σy=[[.08,.04],[.04,.04]]

tr(Gy Σy)=.05=tr(Σx)

Changing the task metric intentionally changes physical costs.

Figure 5.3: *Laboratory record: the shear changes the map (continued).* The sheared display grid and transformed metric accompany three exact displacement checks. The registers also give the quarter-circle length and the covariance trace comparison. These numerical checks distinguish a coordinate transformation from an intentional change of task metric.

What to submit. Supply the two transformed metric matrices, the three-vector table, both curve-length calculations and the covariance identity. Include a short explanation of why an inverse covariance is a plausible ruler for one statistical task but is not automatically a physical distance. The laboratory succeeds when the declared interpretation survives the translation, not merely when a matrix multiplication runs.

Changing the display has left the physical length intact. Now change the question: how distinguishable are two possible settings of an uncertain instrument? A repeated zero-or-one observation is the smallest experiment from which we can derive that statistical ruler.

Derive a statistical ruler from one observed bit

A quality-control station checks each part and records $X = 1$ for a pass or $X = 0$ for a fail. Let $p \in (0, 1)$ be the probability of a pass. Its log likelihood for a single result X is

$$\log f_p(X) = X \log p + (1 - X) \log(1 - p).$$

Differentiating with respect to p gives the score

$$\partial_p \log f_p(X) = \frac{X - p}{p(1 - p)}.$$

Its mean is zero and its mean square is

$$g(p) = \mathbb{E}_p[(\partial_p \log f_p)^2] = \frac{1}{p(1 - p)}.$$

This is the Fisher information metric for the Bernoulli family, per independent observation. It is derived from a specified probability model. Calling an arbitrary distance "Fisher" without that model would omit the construction.

A small change from 0.50 to 0.51 has leading squared length $0.01^2/(0.5 \cdot 0.5) = 0.0004$. A change of the same numerical size from 0.01 to 0.02 starts where the local coefficient is about 101.01. Equal probability increments are not equally distinguishable under this model.

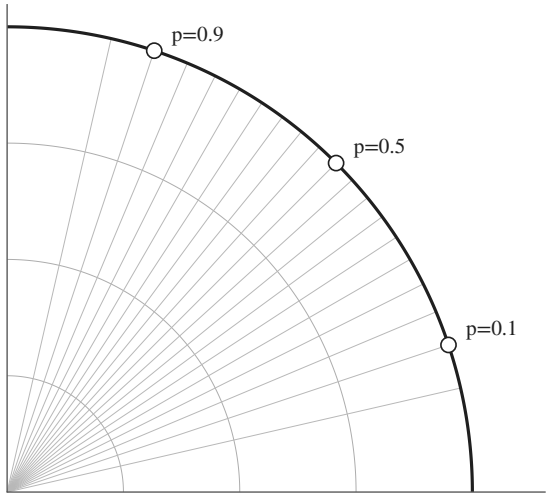
The natural coordinate $\eta = \log(p/(1 - p))$ satisfies $dp/d\eta = p(1 - p)$. In that coordinate the same metric is

$$g_\eta = p(1 - p), \qquad g(p)\, dp^2 = g_\eta\, d\eta^2.$$

Switch the display to log odds and both the coefficient and tick spacing change. Multiply the new coefficient by the squared step on that scale and the new display gives the same local comparison.

One probability traces a curved family

Fisher embedding: $2\psi(p) = (2\sqrt{(1-p)}, 2\sqrt{p})$. Panel A uses ψ at radius 1.



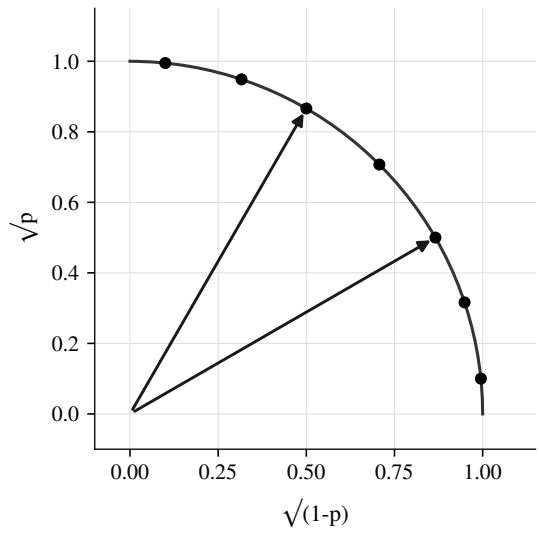
Bernoulli probabilities: (p, 1−p)

Fisher form: $dp^2/[p(1-p)]$

A fixed change in p costs more near either endpoint.

Fisher length uses the radius-two arc. The unit-radius arc is half as long.

A Probabilities become points on an arc



B Small probability changes have unequal cost

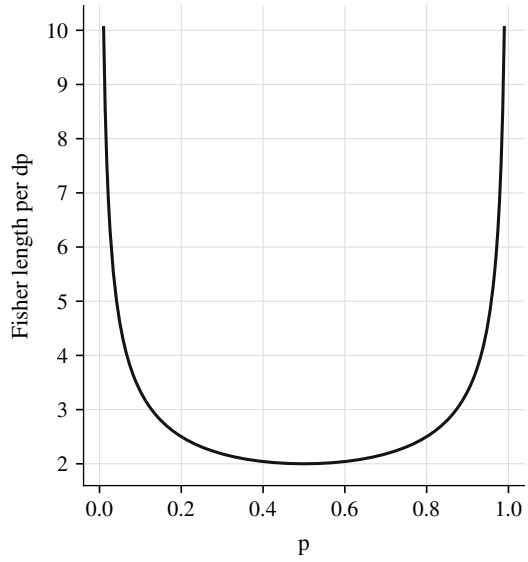
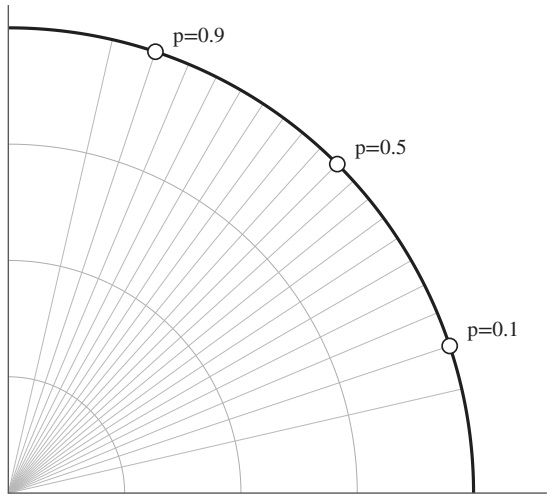


Figure 5.4: *One statistical family, three coordinate rulers.* The study continues on the following figure page.

One probability traces a curved family

Fisher embedding: 2ψ(p) = (2√(1 − p), 2√p). Panel A uses ψ at radius 1.



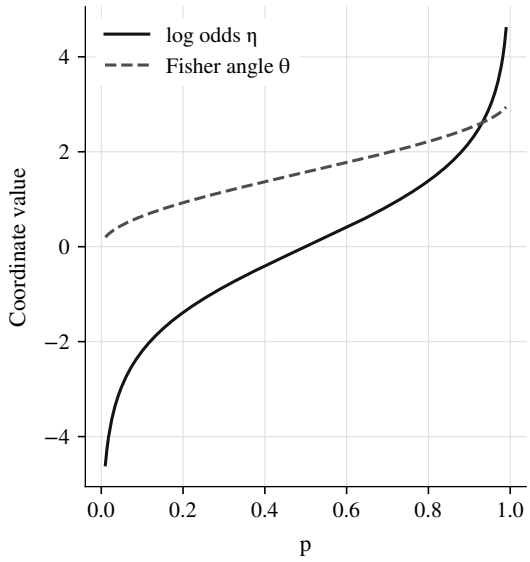
Bernoulli probabilities: (p, 1−p)

Fisher form: $dp^2/[p(1-p)]$

A fixed change in p costs more near either endpoint.

Fisher length uses the radius-two arc. The unit-radius arc is half as long.

C Two coordinates describe one family



D One pair, four different comparisons

Probability difference = 1/2

Fisher distance (radius 2) = π/3

Square-root angle = π/6

Unit-radius chord = 2 sin(π/12)

The regular family has $0 < p < 1$.

Figure 5.4: *One statistical family, three coordinate rulers (continued).* The unit-radius square-root plot and its radius-two scaling describe the same Bernoulli family. Ordinary arc length at radius two equals Fisher distance. Probability, log odds and Fisher angle give three coordinates for that family. The register compares the pair $p = 1/4$, $q = 3/4$. The endpoints $p = 0, 1$ are outside the regular coordinate interval.

The square-root map turns the integral into an arc

To measure the separation between two pass probabilities, set

$$\psi(p) = (\sqrt{1-p}, \sqrt{p}), \qquad \theta = 2 \arcsin \sqrt{p}.$$

The vector $\psi(p)$ lies on the first-quadrant unit circle. Direct differentiation gives $4\|d\psi\|^2 = dp^2/[p(1-p)] = d\theta^2$. Therefore the intrinsic Fisher distance within this one-dimensional family is

$$d_F(p, q) = 2\,|\arcsin \sqrt{q} - \arcsin \sqrt{p}|\,.$$

Between $p = 1/4$ and $q = 3/4$, it is $\pi/3$. The ordinary probability difference is $1/2$. Both numbers are correct answers to their own measurement questions.

The square-root representation also supplies a chord, $\|\psi(p) - \psi(q)\|$. A commonly normalized Hellinger distance is that chord divided by $\sqrt{2}$. Its relationship to the Fisher arc follows from elementary circle geometry, but the quantities remain distinct. The normalization must travel with a reported value.

What survives a sufficient statistic

For n independent Bernoulli observations, the total number of ones $K = \sum X_i$ is sufficient for p: the likelihood factors into a term depending on p through K and a term independent of p. The Fisher information is $n/[p(1-p)]$. Recording K preserves the information about this parameter in this model, although it discards the ordering of the observed bits.

Pass, pass, fail, fail and pass, fail, pass, fail both give two passes. That count suffices for a fixed pass probability under the independent model. If I'm looking for failures in a row, I need the order.

The later information-geometric research develops this connection between statistical structure and geometric comparison [28, 9]. The finite examples here establish the actual transformation rules. General uniqueness claims for Fisher geometry require their own invariance assumptions; they are not a consequence of drawing a curved probability surface.

The Bernoulli calculation already gave one ruler in probability and log-odds coordinates. Keeping that same experiment, we can now derive the relation between those coordinates from a convex potential and see what it does to interpolation.

Two coordinate systems for one statistical state

A quality-control station records whether each part passes a specified test. One display reports the estimated pass probability p. Another reports log odds, $\theta = \log(p/(1-p))$. The displays can show very different numbers while describing the same Bernoulli model. Their relation is an exact translation on $0 < p < 1$, and its derivative determines how their rulers must agree.

The probability mass function can be written

$$P_\theta(X = x) = \exp(x\theta - \psi(\theta)), \qquad x \in \{0, 1\},$$

where $\psi(\theta) = \log(1 + e^\theta)$. Differentiating gives

$$\psi'(\theta) = \frac{e^\theta}{1 + e^\theta} = p, \qquad \psi''(\theta) = p(1-p) > 0.$$

The natural coordinate is θ; the expectation coordinate is $p = E[X]$. Convexity makes the translation one-to-one. This is one precise meaning of dual descriptions in information geometry, and a useful connection to the later semantic-field research [28].

Derive both rulers from the same experiment

The score in natural coordinates is $X - p$. Its variance is $p(1-p)$, so the Fisher line element is

$$ds^2 = p(1-p)d\theta^2.$$

Since $dp = p(1-p)d\theta$, this becomes

$$ds^2 = \frac{dp^2}{p(1-p)}.$$

The log-odds ruler shrinks near very large positive or negative θ, while the probability-coordinate ruler grows near zero and one. The two expressions measure the same tangent displacement. Neither array of coefficients can be copied unchanged into the other coordinate system.

For example, at $p = 1/4$, $\theta = -\log 3$. The derivative $dp/d\theta = 3/16$. An infinitesimal log-odds displacement $d\theta$ has squared length $(3/16)d\theta^2$. Substituting $dp = (3/16)d\theta$ into the probability ruler returns that exact expression.

The finite increment $\theta \mapsto \theta + h$ changes the probability by the logistic formula, not exactly by its derivative. The derivative is a first-order description. A finite update should either use the exact translation or include the corresponding Taylor remainder.

The inspection station still reports one pass probability in two coordinate languages. The next calculation finds a single function whose derivatives generate both the conversion and the ruler. This explains why the two descriptions agree, and gives an implementation quantities it can check together.

The convex potential has a second description

Define the Legendre transform

$$\psi^*(p) = \sup_\theta \{p\theta - \psi(\theta)\}.$$

The maximum occurs where $p = \psi'(\theta)$. Substituting the log-odds expression gives

$$\psi^*(p) = p\log p + (1-p)\log(1-p).$$

Its derivative is θ, and its second derivative is $1/[p(1-p)]$. The Hessians of the two potentials are reciprocals at corresponding points. The two coordinate descriptions therefore carry not only inverse maps but also explicitly related local metrics.

This is a different construction from a topological double cover. Every finite θ corresponds to exactly one interior probability p. There are two useful coordinate languages, with a one-to-one transition. The bar and quaternion examples instead have two representatives over each base point.

Now let the inspection station record three possible outcomes. Two software versions can agree on the starting and ending distributions yet choose different intermediate distributions. Following that update tests a stronger kind of preservation than matching endpoint labels: the route itself needs a declared rule.

Three outcomes reveal different interpolation rules

For a three-outcome distribution $p = (p_1, p_2, p_3)$ with every $p_i > 0$ and $\sum p_i = 1$, choose natural coordinates

$$\theta_1 = \log(p_1/p_3), \qquad \theta_2 = \log(p_2/p_3).$$

The potential is $\psi(\theta) = \log(1 + e^{\theta_1} + e^{\theta_2})$. Its gradient gives (p_1, p_2), and $p_3 = 1 - p_1 - p_2$. Its Hessian has entries

$$(\nabla^2\psi)_{ij} = p_i\delta_{ij} - p_ip_j, \qquad i, j \in \{1, 2\}.$$

This positive-definite matrix is the Fisher metric in natural coordinates. In probability coordinates, with $dp_3 = -dp_1 - dp_2$, the line element is

$$ds^2 = \sum_{i=1}^{3} \frac{dp_i^2}{p_i}.$$

The two forms are related by the Hessian Jacobian. They describe the same statistical state space.

There are now several distinct ways to connect two distributions. A straight path in probability coordinates is the mixture

$$p_i(t) = (1 - t)p_i + tq_i.$$

A straight path in natural coordinates gives the normalized geometric interpolation

$$p_i(t) = \frac{p_i^{1-t}q_i^t}{\sum_j p_j^{1-t}q_j^t}.$$

Neither is generally the shortest path for the Fisher metric. A straight line in a coordinate display is not automatically a Riemannian geodesic.

Calculate the Fisher route

Set $s_i = \sqrt{p_i}$. Because $\sum s_i^2 = 1$ and $dp_i = 2s_i ds_i$, the line element becomes $4\sum ds_i^2$ on the positive part of the unit sphere. The statistical geometry is a spherical geometry with lengths multiplied by two.

Let $a = (\sqrt{p_1}, \sqrt{p_2}, \sqrt{p_3})$, $b = (\sqrt{q_1}, \sqrt{q_2}, \sqrt{q_3})$ and $\alpha = \arccos(a \cdot b)$. For distinct interior distributions,

$$s(t) = \frac{\sin((1 - t)\alpha)}{\sin\alpha}a + \frac{\sin(t\alpha)}{\sin\alpha}b$$

is the short great-circle interpolation. Squaring its components gives the Fisher geodesic in probability coordinates. Its distance is

$$d_F(p, q) = 2\arccos\left(\sum_i \sqrt{p_iq_i}\right).$$

The positive coefficients in the short arc keep the interpolation inside the positive orthant. The coordinate construction has supplied a concrete globe and an exact route.

Inspect three different midpoints

Take $p = (0.8, 0.1, 0.1)$ and $q = (0.1, 0.8, 0.1)$. The mixture midpoint is $(0.45, 0.45, 0.1)$. The natural-coordinate midpoint has components proportional to $(\sqrt{0.08}, \sqrt{0.08}, 0.1)$. Its third component is approximately 0.150221.

The Fisher midpoint is obtained by normalizing $a + b$ and squaring. Its third component is

$$\frac{0.4}{2 + 2(2\sqrt{0.08} + 0.1)} \simeq 0.120071.$$

The two equal remaining components are approximately 0.439965. All three midpoints are legitimate distributions. They answer different interpolation questions. A program that reports "the midpoint" without naming its geometry leaves the main modeling choice unstated.

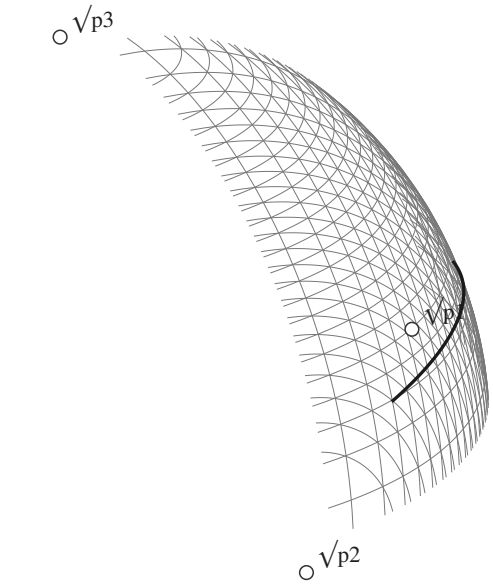
Flat connections need not mean a flat Riemannian metric

Natural and expectation coordinates support the familiar dual affine structures of a regular exponential family. Their straight-coordinate paths are associated with those affine connections. The Levi–Civita connection of the Fisher metric is another connection. In the three-outcome example, the Fisher surface has positive curvature 1/4, as the radius-two sphere calculation shows.

Thus a claim of dual flatness does not say that the Fisher metric is Euclidean. It identifies a pair of affine structures. The map/globe distinction is doing real work here: one must say which maps are straightening which connection, and which ruler measures the route.

A distribution becomes a geometric point

Square-root coordinates send three probabilities to the positive unit sphere.



$p_1 + p_2 + p_3 = 1$

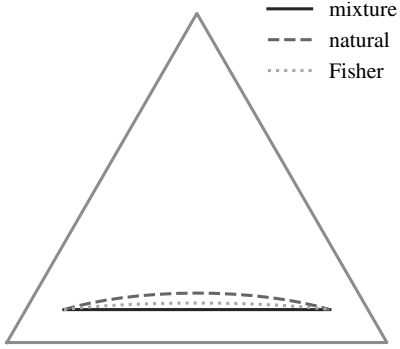
$x_i = \sqrt{p_i}$; $\Sigma\, x_i^2 = 1$

Fisher length = 2 × sphere arc

The ruler and the path both matter.

A mixture path and an exponential path need not follow this arc. The panels below name the precise family used in each comparison.

A Three geometries give three distribution paths



B The third probability separates the paths

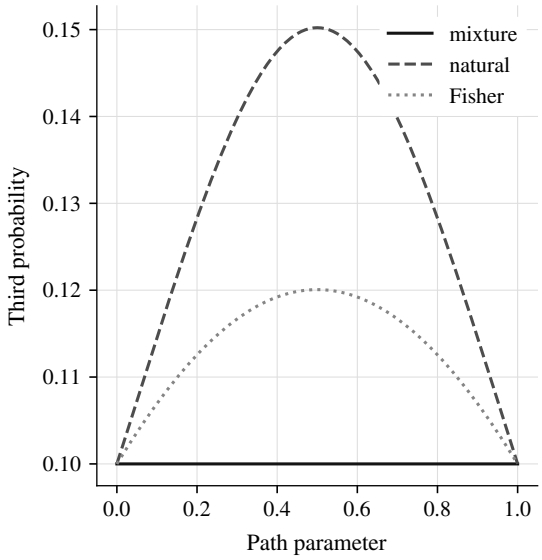
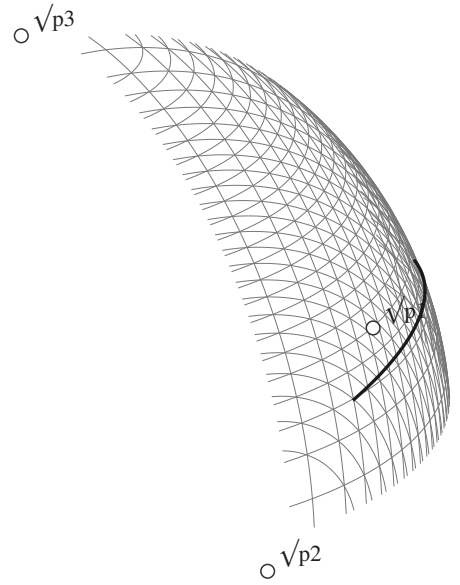


Figure 5.5: *One statistical state has two exact coordinate languages.* The study continues on the following figure page.

A distribution becomes a geometric point

Square-root coordinates send three probabilities to the positive unit sphere.



$p_1 + p_2 + p_3 = 1$

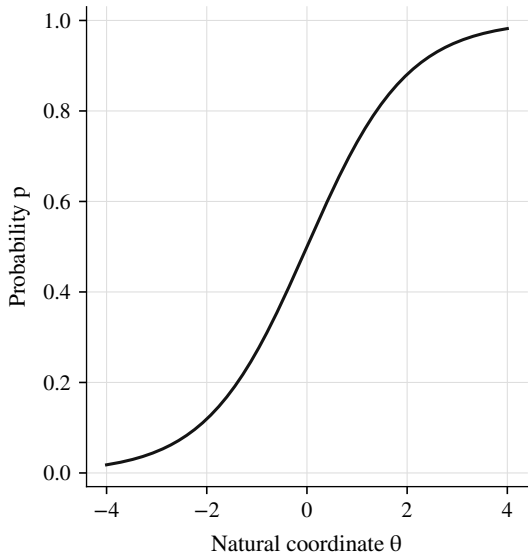
$x_i = \sqrt{p_i}$; $\Sigma\, x_i^2 = 1$

Fisher length = 2 × sphere arc

The ruler and the path both matter.

A mixture path and an exponential path need not follow this arc.
The panels below name the precise family used in each comparison.

C Log odds translate probability coordinates



D The midpoints reveal the different constructions

Mixture: p3 = 0.100000

Natural: p3 = 0.150221

Fisher: p3 = 0.120071

Dual coordinates are one-to-one.

They are not a two-sheeted cover.

Figure 5.5: *One statistical state has two exact coordinate languages (continued).* Mixture, natural-coordinate and Fisher-geodesic interpolation connect the same distributions by different paths. Their midpoints differ. The Bernoulli log-odds translation is one-to-one and preserves the Fisher ruler by the coordinate rule.

Three midpoints. Calculate all three midpoint vectors to six decimal places, verify that each sums to one, and compare its third component. Then set $p = q$ and show that every construction returns that same distribution without dividing by $\sin 0$.

Return to Laboratory 2: the statistical ruler. Repeat the principle on the Bernoulli family. For $p = 1/4$ and $q = 3/4$, integrate $dp/\sqrt{p(1-p)}$. Repeat using log odds and its metric $p(1-p)d\eta^2$. Both give $\pi/3$. A uniform step size in p is not a uniform step size in η or in Fisher arc length. Plot all three sequences for ten equal increments and identify which sequence is uniform in the intrinsic ruler.

The zero-or-one experiment had one varying probability. A calibration run can vary in both its mean and its spread. We keep the same score-based method and let that additional degree of freedom determine the next statistical surface.

Comparing two calibrations, including their uncertainty

A laboratory calibrates the same instrument twice. The first run has a mean reading of zero and standard deviation one. The second has a shifted mean or a different spread. Comparing the means alone would miss the second kind of change. Comparing two columns of numbers with a Euclidean ruler would depend on the units chosen for the instrument.

As a declared model, suppose each run is represented by a univariate normal distribution $N(\mu, \sigma^2)$, with $\sigma > 0$. This model summarizes a mean and a scale; it does not assert that every calibration error is Gaussian. Its parameter space is the upper half-plane. We will derive a ruler from how the probability density changes.

Derive the metric from the score

The log density at a measurement x is

$$\ell(x; \mu, \sigma) = -\log \sigma - \tfrac{1}{2}\log(2\pi) - \frac{(x-\mu)^2}{2\sigma^2}.$$

Its parameter derivatives are

$$\partial_\mu \ell = \frac{x-\mu}{\sigma^2}, \qquad \partial_\sigma \ell = -\frac{1}{\sigma} + \frac{(x-\mu)^2}{\sigma^3}.$$

The Fisher metric is the expected outer product of these two derivatives. Writing $Z = (X-\mu)/\sigma$, the standard normal moments $E[Z] = E[Z^3] = 0$, $E[Z^2] = 1$, and $E[Z^4] = 3$ give

$$I(\mu, \sigma) = \begin{pmatrix} \sigma^{-2} & 0 \\ 0 & 2\sigma^{-2} \end{pmatrix}.$$

Thus

$$ds_F^2 = \frac{d\mu^2 + 2d\sigma^2}{\sigma^2}.$$

The same absolute shift of the mean is geometrically larger when the distribution is narrow. A one-unit shift in a precise calibration is easier to distinguish than a one-unit shift buried inside a much broader distribution. That interpretation follows from the model's expected score; it is not an arbitrary task weight.

Recognize the underlying geometry

Set $x = \mu/\sqrt{2}$ and $y = \sigma$. Then

$$ds_F^2 = 2\frac{dx^2 + dy^2}{y^2}.$$

This is twice the usual hyperbolic upper-half-plane metric. Lengths acquire a factor $\sqrt{2}$. The normal-family distance is consequently

$$d_F = \sqrt{2}\,\operatorname{arcosh}\left(1 + \frac{(\mu_1 - \mu_2)^2 + 2(\sigma_1 - \sigma_2)^2}{4\sigma_1\sigma_2}\right).$$

This closed form is specific to the univariate family. General multivariate normal models require more care; a corresponding closed formula is not supplied simply by replacing the scalar variance with a matrix [35].

Calculate two physically different changes

If the mean stays zero and the standard deviation doubles from one to two, the vertical path gives

$$d_F = \sqrt{2}\left|\int_1^2 \frac{d\sigma}{\sigma}\right| = \sqrt{2}\log 2 \simeq 0.980258.$$

The natural comparison depends on the ratio of scales. Doubling from ten to twenty has the same Fisher distance as doubling from one to two. Those pairs represent the same relative change of uncertainty within this family.

Now compare $N(0,1)$ with $N(2,1)$. The distance formula gives

$$d_F = \sqrt{2}\,\operatorname{arcosh} 2 \simeq 1.862460.$$

The horizontal path that changes only the mean has length two. The shorter full-family geodesic increases the standard deviation along the way and then decreases it. In the (x, y) half-plane it is a circle orthogonal to the boundary, centered at $(1/\sqrt{2}, 0)$ with radius $\sqrt{3/2}$.

This does not mean the physical instrument must become noisier to change its bias. The geodesic is a shortest comparison through the chosen family of distributions. A physical calibration process restricted to constant noise follows a different allowed path family. The distinction is exactly the same as the workbench route: a shortest route in a larger space can leave a constraint imposed by the actual experiment.

Change units without changing the comparison

Report measurements in new units $X' = aX + b$, with $a > 0$. The parameters become $\mu' = a\mu + b$ and $\sigma' = a\sigma$. Their differentials satisfy $d\mu' = a\,d\mu$ and $d\sigma' = a\,d\sigma$, so

$$\frac{d\mu'^2 + 2d\sigma'^2}{\sigma'^2} = \frac{d\mu^2 + 2d\sigma^2}{\sigma^2}.$$

The Fisher length is unchanged. The coordinates of the two calibration models move and scale on the printed half-plane, while their intrinsic comparison remains the same. This is a map/globe example with no literal Earth: the common object is the family of probability distributions with its information metric.

Independent repetitions change information

For n independent readings from one parameter value, the joint log likelihood is a sum of n single-reading log likelihoods. The score is their sum. Under the model the scores have mean zero, so independence makes the cross terms vanish. The joint Fisher metric is

$$I_n = nI_1.$$

Corresponding parameter distances scale by $\sqrt{n}$ when this joint-experiment metric is used. This is a change in the experiment's information, not a coordinate change. A hundred independent readings distinguish nearby parameter values more strongly than one reading.

Repeated copies of the same reading do not justify that factor. Their scores are dependent. The information calculation must use the actual joint model. This is one reason the GIGI power-plant experiment does not count shuffled worksheets as new independent records.

What a lossy summary does to the ruler

For a regular statistical model with score $S(X)$ and a deterministic summary $Y = T(X)$, the summary's score is the conditional expectation $E[S(X) \mid Y]$, under the usual differentiation and integration conditions. The conditional covariance identity then gives

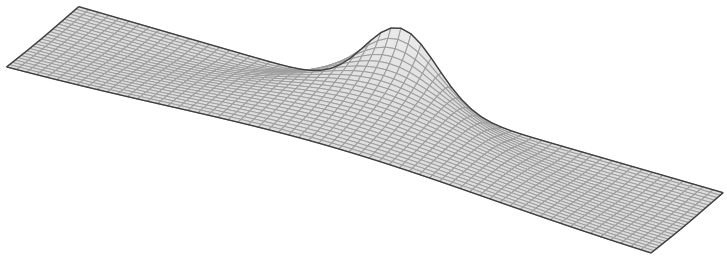
$$I_X - I_Y = E[\mathrm{Cov}(S(X) \mid Y)] \succeq 0.$$

For any parameter direction v, the lost information is the expected conditional variance of $v^T S(X)$ given the summary. It is nonnegative. Equality in every direction means the score is retained by the summary, subject to the model's regularity assumptions [9].

The equation supplies a direct test of what "preserving the geometry" means for this statistical construction. A summary can preserve a task-relevant quantity without preserving the full Fisher metric. A lossless encoding preserves the original observation, but its byte count is not itself a Fisher distance.

Mean and spread specify a family of shapes

Normal densities are drawn for declared parameter values, not as measured samples.



Horizontal: measured value

Depth: standard deviation σ

Height: probability density

Mean is fixed at zero in this view.

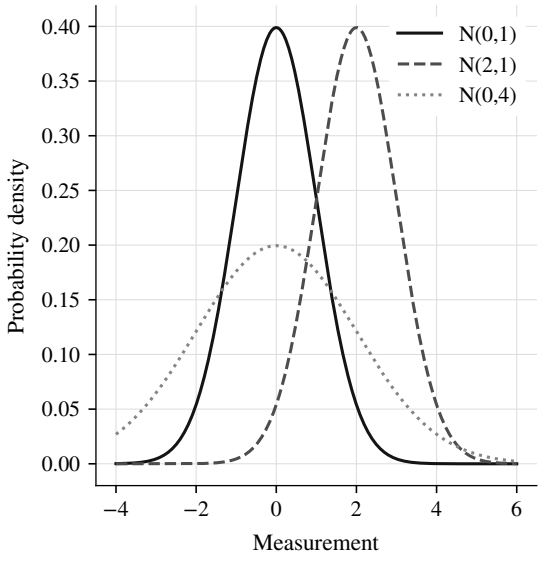
Changing σ changes the ruler.

Fisher metric at (μ, σ):

$\mathrm{diag}(1/\sigma^2, 2/\sigma^2)$

The full family permits both mean and spread to move.

A Mean and spread identify the normal model



B The shortest full-family route can change spread

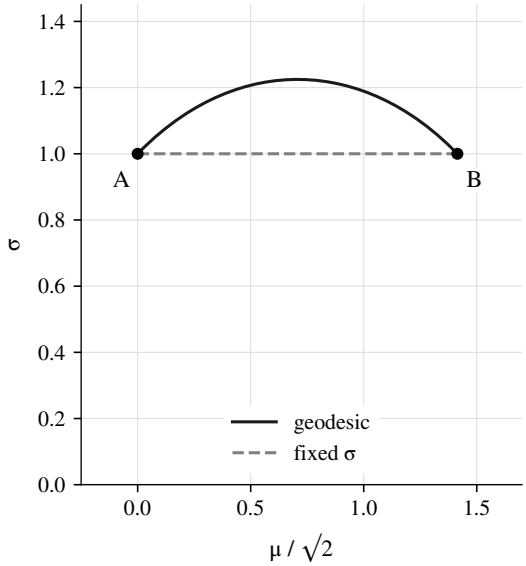
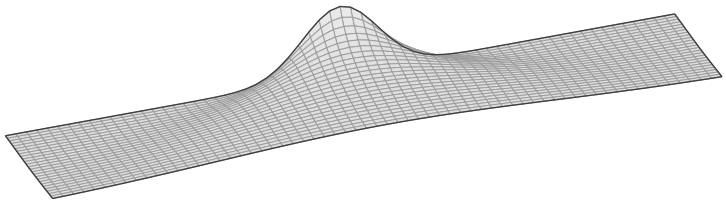


Figure 5.6: *The geometry of two calibration parameters.* The study continues on the following figure page.

Mean and spread specify a family of shapes

Normal densities are drawn for declared parameter values, not as measured samples.



Horizontal: measured value

Depth: standard deviation σ

Height: probability density

Mean is fixed at zero in this view.

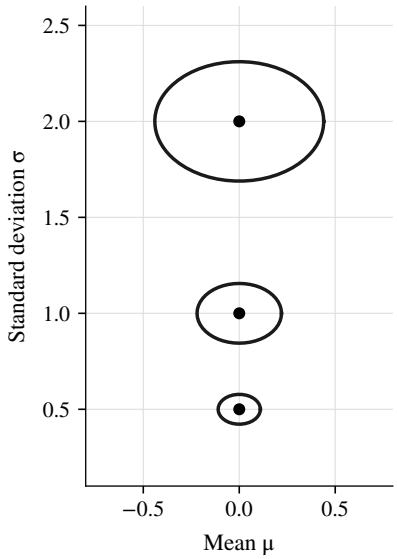
Changing σ changes the ruler.

Fisher metric at (μ, σ):

$\mathrm{diag}(1/\sigma^2, 2/\sigma^2)$

The full family permits both mean and spread to move.

C Equal Fisher steps look different in coordinates



D Exact distances depend on the permitted family

N(0,1) to N(0,4):

distance $\sqrt{2}\,\log 2 = 0.980258$

N(0,1) to N(2,1):

distance $\sqrt{2}\,\mathrm{acosh}\,2 = 1.862460$

The full-family route may leave a fixed-noise constraint.

Figure 5.6: *The geometry of two calibration parameters (continued).* The probability densities, full-family geodesic, local Fisher rulers and exact distances describe a specified normal-family model. The geodesic may vary spread even when its endpoints have equal spread.

Calibration comparison. Compute the distance between $N(5, 4)$ and $N(7, 4)$, remembering that the second argument denotes variance. Express the same pair after multiplying all readings by 100. Verify both the transformed parameters and the unchanged distance.

The surveyor can now switch maps without changing the measured displacement. The inspection station can switch between probability and log odds without changing its Fisher comparison. A change in the experiment itself, such as collecting more independent readings, changes the information available and may change that statistical ruler. These examples establish the same discipline: specify what is measured, derive the ruler, and carry it through every change of description. The next chapter follows that ruler along an entire route.

Practice with the construction

5.1. At $(r, \theta) = (3, \pi/6)$, compute the length of the tangent displacement $(dr, d\theta) = (0.04, 0.02)$. Check it by multiplying by J.

5.2. If the angular coordinate is measured in degrees, derive the angular coefficient in the metric.

Chapter 6

Length, distance and allowable paths

Two doors face each other across a courtyard. A fence goes up through the middle, and anyone walking between the doors must go around it. The positions of the doors and the metre on our ruler stay the same. What changes is which routes we can use to measure the walking distance.

Paths and Geodesics in Metric Spaces

To calculate the walking distance, we need a description of the permitted routes and a way to measure each one. We start by writing a path as a sequence of positions indexed continuously from departure to arrival.

Paths and their lengths

In a metric space (X, d), a *path* from x to y is just a continuous map

$$\gamma : [0, 1] \to X$$

with $\gamma(0) = x$ and $\gamma(1) = y$.

We can think of $\gamma(t)$ as the state of the system at "time" t as it moves from x to y.

To talk about shortest paths, we need to define the *length* of a path. In a general metric space, one standard way to do this is:

Definition 6.1 (Length of a path)**.** Given a path $\gamma : [0, 1] \to X$ in a metric space (X, d), its *length* $L(\gamma)$ is defined as

$$L(\gamma) = \sup_{\mathcal{P}} \sum_{k=1}^{n} d\big(\gamma(t_k), \gamma(t_{k-1})\big),$$

where the supremum is taken over all partitions $\mathcal{P} : 0 = t_0 < t_1 < \cdots < t_n = 1$ of the interval $[0, 1]$.

Measure the route by successive observations. Let a survey vehicle record its position at the times $t_0, \ldots, t_n$. The partition sum adds the metric distances between consecutive reports. The triangle inequality makes each sum a lower bound on the traveled length. Adding sample points refines the partition and cannot decrease that sum. Taking the supremum over all partitions gives the path's metric length.

A vehicle can turn between recorded positions. Its sparse position log therefore need not determine the traveled length exactly. More frequent samples reveal some of the omitted turns, while the supremum defines the quantity the sampling is trying to approximate. On a smooth Riemannian route, the metric-speed integral in Length of curves on a Riemannian manifold computes that same length.

Now hold the endpoints fixed. Measuring the vehicle's actual route and finding the least possible route are two questions. The geodesic definition identifies the case in which the traveled length attains the endpoint distance.

Geodesics as shortest paths

A natural question now is: given two points x and y in a metric space (X, d), does there exist a path from x to y that has the smallest possible length among all paths connecting them?

A path whose length equals the endpoint distance is a *minimizing metric geodesic*. In a general metric space, even the shortest available path can have length greater than $d(x, y)$. The given distance may count shortcuts that no continuous path can follow.

Definition 6.2 (Minimizing metric geodesic)**.** A path $\gamma : [0, 1] \to X$ with $\gamma(0) = x$ and $\gamma(1) = y$ is called a *minimizing metric geodesic* from x to y if its length satisfies

$$L(\gamma) = d(x, y).$$

Not every metric space has a minimizing geodesic between every pair of points, and a minimizer need not be unique. A connected complete Riemannian manifold does have minimizing geodesics by the Hopf–Rinow theorem. On an incomplete region or a general translator quotient, the distance may be an infimum that no path attains. When the admissible path set is nonempty and its infimum is finite, paths can approach that infimum arbitrarily closely; this alone does not produce an exact minimizer.

Some familiar examples:

- In Euclidean space, the geodesics are straight line segments.
- On the round sphere, the shortest great-circle arcs are minimizing metric geodesics.
- In a graph whose positive-length edges are intervals, a shortest route along the edges is a minimizing metric geodesic. A shortest vertex-to-vertex edge sequence is the discrete counterpart.

Why paths and geodesics matter for sameness

In Chapter 1, we emphasized that sameness is not just a static question about two snapshots. It is also about how an entity can move over time without ceasing to be "the same" for our purposes.

Paths and geodesics give us a way to talk about this in a general metric setting:

- A *benign evolution* of a latent entity can be modeled as a path in an appropriate metric space.
- The *amount* of benign evolution needed to move from one state to another can be measured by the length of the shortest benign path between them.
- If the shortest benign path between two states is very short, we may consider them semantically the same; if it is very long, we may consider them different.

In translator space, paths will be built from sequences of translations and within-space moves. In geometric space, paths will be curves on a manifold. The language of metric spaces and geodesics covers both.

Keep the permitted evolution attached to the route. A machine can change gradually while remaining the same instrument. The trajectory records that change, and its length measures it in the chosen ruler. Whether the trajectory counts as normal operation comes from the task and dynamics. A shortest path becomes relevant only after those choices specify the routes over which we are minimizing.

The fence in the opening example changes the set of routes between the same doors. We can calculate that distinction exactly on a circular track and a sphere, then test the steps a program uses to follow a route.

A short route can leave the map

A circular track gives us a simple version of the fenced courtyard. The track is the only permitted walking surface, while a brace can span its interior. Two positions have angular coordinates θ_1 and θ_2. On a radius-R circle their intrinsic distance is

$$d_{S^1}(\theta_1, \theta_2) = R \min_{k \in \mathbb{Z}} |\theta_1 - \theta_2 + 2\pi k|.$$

The straight line between them through the surrounding plane has length $2R \sin(\Delta/2)$, where $\Delta \in [0, \pi]$ is the shorter angular separation. The first answers how far someone walks around the rim. The second answers how long a straight brace between the positions must be. Neither is a mistake; they solve different path problems.

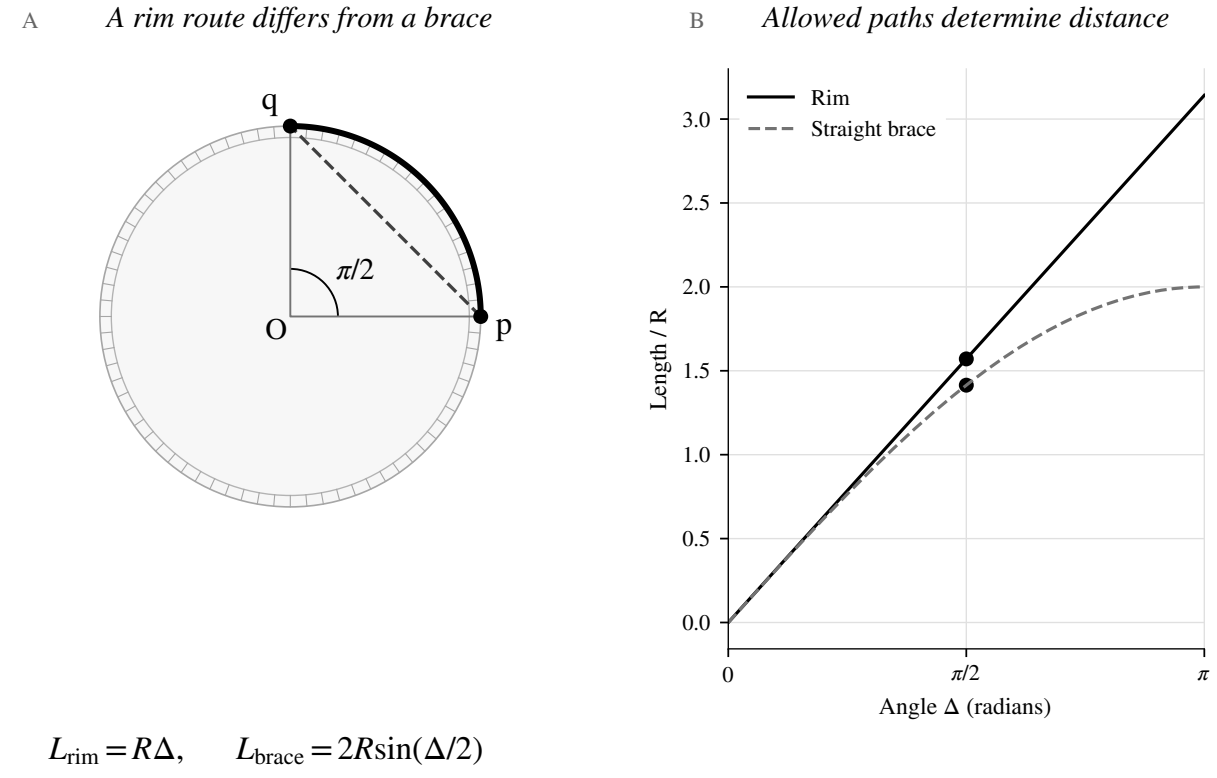


Figure 6.1: A circular rim permits a surface route; a straight brace uses the surrounding plane. The cross-section and length curves compare these two path problems. At a quarter-turn their lengths are $\pi R/2$ and $\sqrt{2}R$, respectively.

On the circle the route was given. On a survey surface we may have only the local ruler and need to determine the route it favors. Vary a connecting curve while holding its endpoints fixed. For a smooth metric g_{ij}, a constant-speed critical point of the energy

$$E(\gamma) = \frac{1}{2} \int_a^b g_{ij}(\gamma) \dot{\gamma}^i \dot{\gamma}^j \, dt$$

satisfies the geodesic equation

$$\ddot{\gamma}^k + \Gamma^k_{ij} \dot{\gamma}^i \dot{\gamma}^j = 0, \qquad \Gamma^k_{ij} = \frac{1}{2} g^{k\ell} (\partial_i g_{j\ell} + \partial_j g_{i\ell} - \partial_\ell g_{ij}).$$

Here and below repeated indices are summed. Differentiating the Lagrangian with respect to $\dot{\gamma}^k$, then applying the Euler–Lagrange equation, gives this expression after multiplying by the inverse metric. These terms compensate for the changing coordinate ruler. They do not imply that a force acts on an otherwise straight physical trajectory.

For polar coordinates, the nonzero coefficients are $\Gamma^r_{\theta\theta} = -r$ and $\Gamma^\theta_{r\theta} = \Gamma^\theta_{\theta r} = 1/r$. A straight Cartesian line can therefore have a nonzero second derivative in polar coordinates while satisfying the geodesic equation. A circle of fixed radius is not a plane geodesic: substituting $\dot{r} = 0$ and $\dot{\theta} \neq 0$ leaves the radial term $-r\dot{\theta}^2$.

A geodesic is locally straight in the chosen metric. It need not be the globally shortest route between distant endpoints. On a sphere a great circle continued past a half-turn is still a geodesic but takes the longer way around. Existence of minimizing geodesics also needs hypotheses: completeness of a connected finite-dimensional Riemannian manifold is a sufficient condition through Hopf–Rinow [4].

The circular track distinguished the distance walked along the rim from the length of a straight brace. On the globe we will compare two routes that both stay on its surface.

The shortest line on the page can be the wrong route

Lay a flexible tape against a globe between two marked points and pull it along the shortest surface route. For non-antipodal endpoints this is the shorter great-circle arc. Draw the same points on a flat map and join their coordinates with a ruler. That line usually describes a different route on the globe.

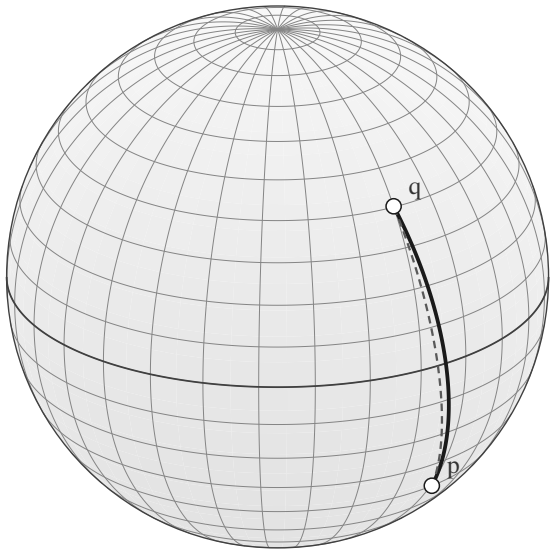
For unit sphere points p, q with $p \neq \pm q$, put $\theta = \arccos(p^T q)$. Their constant-speed shortest arc is

$$\gamma(t) = \frac{\sin((1-t)\theta)}{\sin\theta} p + \frac{\sin(t\theta)}{\sin\theta} q.$$

This lies in the plane spanned by p, q, remains on the sphere and has speed θ on $0 \leq t \leq 1$. At antipodal points there are infinitely many shortest semicircles. An endpoint distance alone cannot choose among them.

Two routes across one curved surface

The solid geodesic and dashed map-straight route share endpoints.



UNIT SPHERE

Solid: shortest great-circle arc

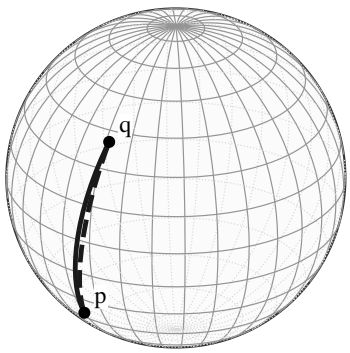
Dashed: straight in map coordinates

Arc length 1.275610

Shared endpoints

A route is more than its endpoints. The speed under the surface metric must be integrated along each path.

A Straight on the map need not be shortest



B The ruler determines speed along each route

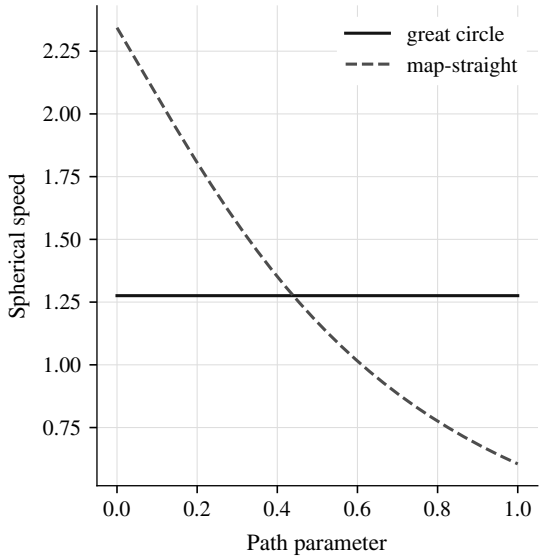
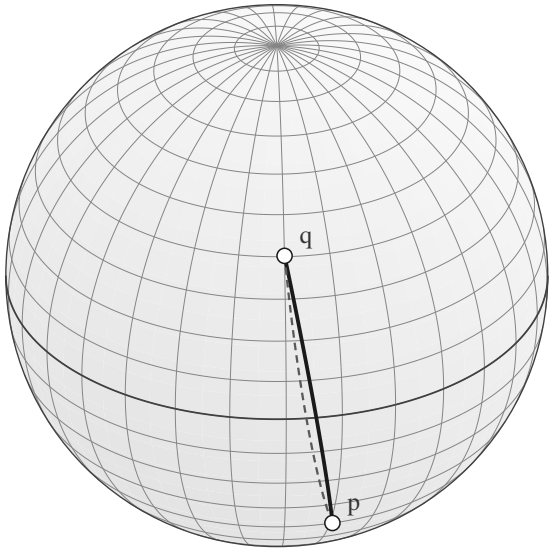


Figure 6.2: *The route on the globe determines the calculation.* The study continues on the following figure page.

Two routes across one curved surface

The solid geodesic and dashed map-straight route share endpoints.



UNIT SPHERE

Solid: shortest great-circle arc

Dashed: straight in map coordinates

Arc length 1.275610

Shared endpoints

A route is more than its endpoints. The speed under the surface metric must be integrated along each path.

C Common endpoints, unequal lengths

Great-circle length = 1.275610	Both routes reach the same q.
Map-straight length = 1.277916	Only one is the shortest spherical arc.
Sphere radius R = 1	Coordinate straightness belongs to the chosen map.
Lengths multiply by R on a sphere of radius R.	The metric determines route speed.

D The surface determines the shortest route

Figure 6.2: *The route on the globe determines the calculation (continued).* A great-circle segment and a coordinate-straight alternative are drawn on the same globe. Their local speeds are evaluated with the spherical metric. The endpoint agreement does not imply equal route length; both computed lengths are given for unit radius.

The exponential map starts with a measured tangent step

The shortest arc supplies the route we want. An implementation often stores a tangent increment instead of a complete curve. We now compare the exact update that follows that arc with a cheaper normalized update, so that a numerical shortcut has a measurable cost in the same route problem.

At a unit sphere point p, a tangent vector v satisfies $p^T v = 0$. For $r = \|v\|$, the exponential map is

$$\operatorname{Exp}_p(v) = \cos r\, p + \frac{\sin r}{r} v,$$

with the continuous value p at $v = 0$. The prescribed geodesic has length r. On the unit sphere this equals the shortest distance for $r \le \pi$, with multiple shortest arcs at $r = \pi$.

A common cheap update instead normalizes $p + v$:

$$\operatorname{Retr}_p(v) = \frac{p + v}{\sqrt{1 + r^2}}.$$

This lies on the same great circle but travels through angle $\arctan r$, not r. The difference is

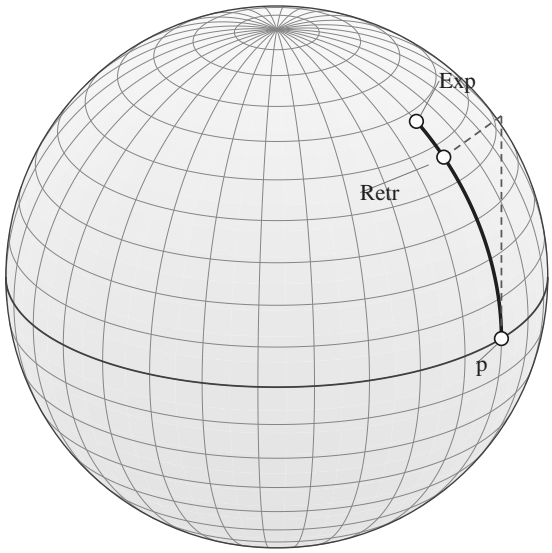
$$r - \arctan r = \frac{r^3}{3} + O(r^5).$$

At $r = 0.1$ the difference is about 0.000331 radians; at $r = 1$ it is about 0.214602 radians. One cheap update leaves that one-radian trip more than a fifth of a radian short.

For repeated steps along one great circle, n normalized steps of size r/n travel $n \arctan(r/n)$ radians. Expanding gives a total deficit $r^3/(3n^2) + O(n^{-4})$. On that circle, ten requests for one tenth of a radian miss by about 0.00331 radians altogether with the cheap update. Dividing up the trip changes what I can promise about the final position. If the heading changes at each step, I need a separate error calculation.

The tangent step has two continuations

A spherical arc and a normalized straight step have different endpoints.



Tangent norm r = 0.9

Exponential: angle r

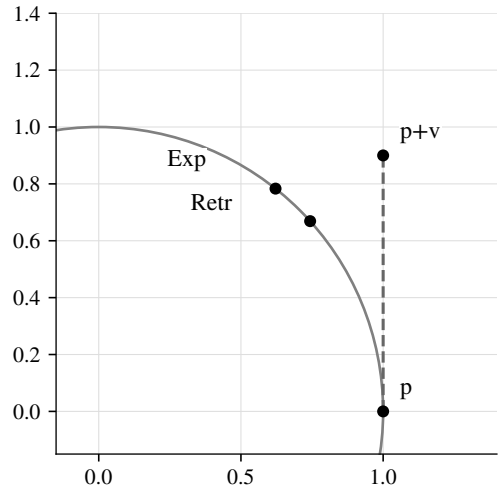
Normalize: angle arctan r

Deficit: r − arctan r

The construction determines the endpoint.

A smaller integration step reduces the discrepancy; it does not make the two operations identical.

A Projecting a tangent step changes its endpoint



B Projection falls short of the geodesic step

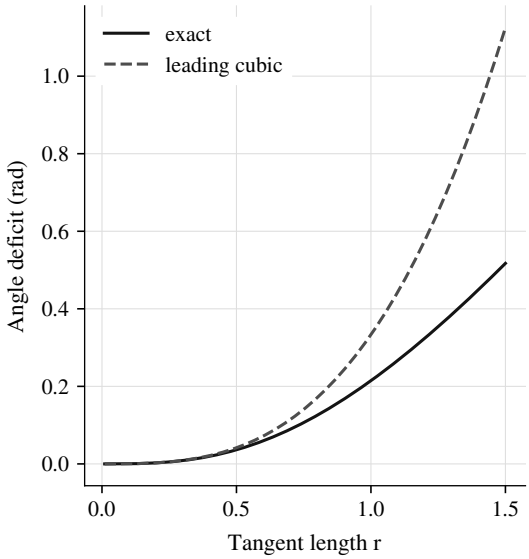
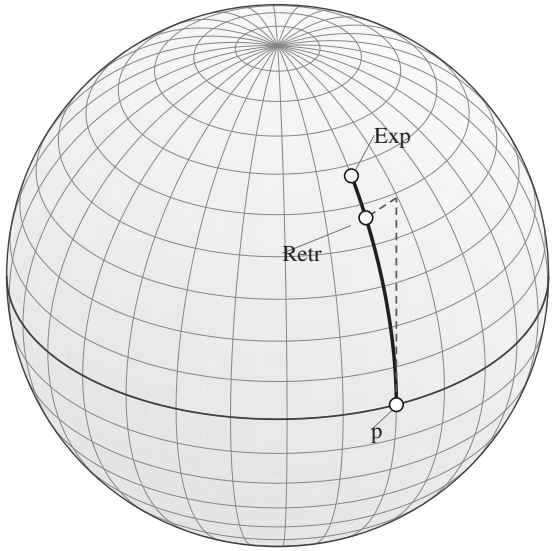


Figure 6.3: *A tangent step and a geodesic step are not the same operation.* The study continues on the following figure page.

The tangent step has two continuations

A spherical arc and a normalized straight step have different endpoints.



Tangent norm r = 0.9

Exponential: angle r

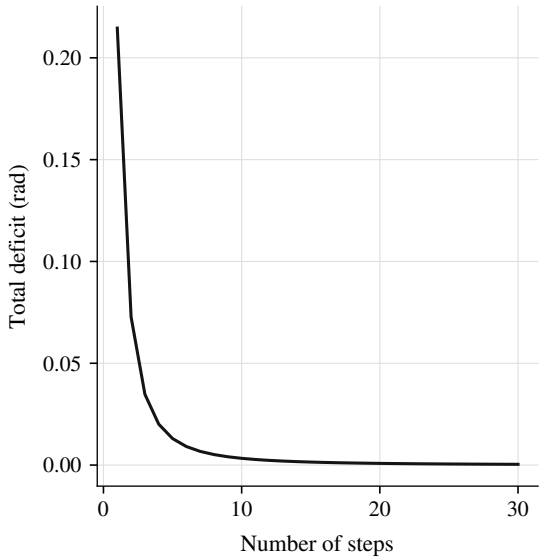
Normalize: angle arctan r

Deficit: r − arctan r

The construction determines the endpoint.

A smaller integration step reduces the discrepancy; it does not make the two operations identical.

C Smaller steps reduce the accumulated deficit



D Two step sizes, two visible errors

r=0.1: deficit 0.000331 rad

r=1.0: deficit 0.214602 rad

Exp uses sin(r), cos(r).

Normalization travels arctan(r).

Figure 6.3: *A tangent step and a geodesic step are not the same operation (continued).* The tangent construction, exponential endpoint and normalized endpoint are drawn to scale. The lower computation shows the cubic one-step error and the quadratic improvement obtained by subdividing a fixed great-circle journey.

A closed road changes the distance problem

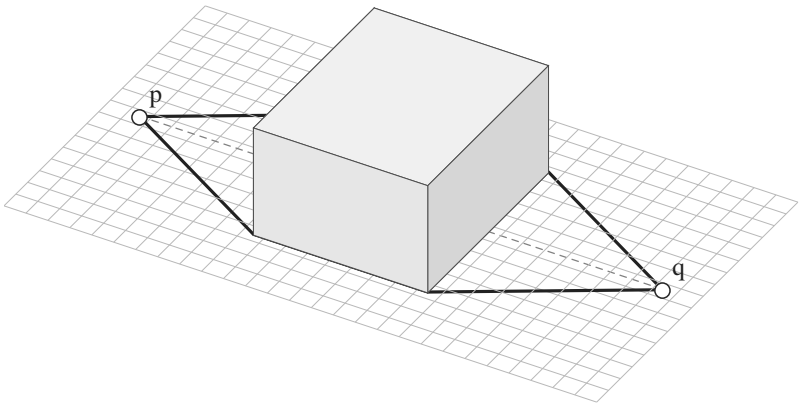
A truck detouring around a lake can drive only where a road runs. The same issue occurs in a configuration manifold when some states are forbidden, when a translator only accepts certain inputs or when a physical actuator cannot move in every tangent direction.

Consider a planar corridor with a rectangular obstacle. The ordinary Euclidean distance between two endpoints ignores the obstacle. The intrinsic path distance in the allowed region takes the infimum over curves that stay outside it. A route touching a forbidden boundary may be an unattained limit if that boundary is excluded; the infimum can still be well defined. A claim that a shortest path exists needs compactness or other appropriate hypotheses.

Walk closer to the edge of the rectangular obstacle and the detour can get shorter, even if the route right along that edge is off limits. A translator program can likewise find an allowed route and bound how much cheaper any other route could be.

The permitted region determines the journey

Plan geometry is extruded only to make the forbidden footprint visible.



Direct segment: 3

Boundary route: 1 + 2√1.36

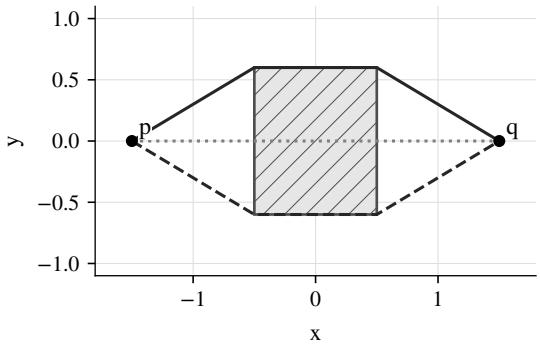
Length: 3.332381

The footprint is 1 by 1.2.

If the boundary itself is excluded, this length is an infimum. Allowed paths approach it without necessarily attaining it.

A An obstacle rules out the direct route

B The detour has a computable length



Direct segment = 3

Boundary route = $1 + 2\sqrt{1.36}$

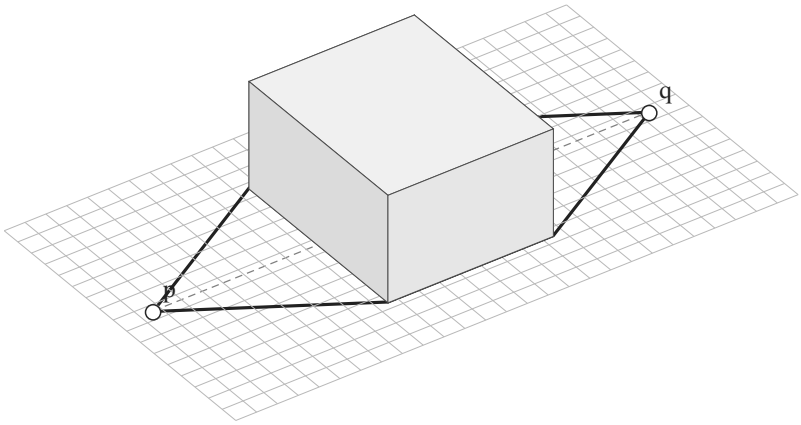
Boundary route ≈ 3.332381

If the boundary is forbidden, this is an infimum approached by allowed routes.

Figure 6.4: *The allowed region changes which route counts.* The study continues on the following figure page.

The permitted region determines the journey

Plan geometry is extruded only to make the forbidden footprint visible.



Direct segment: 3

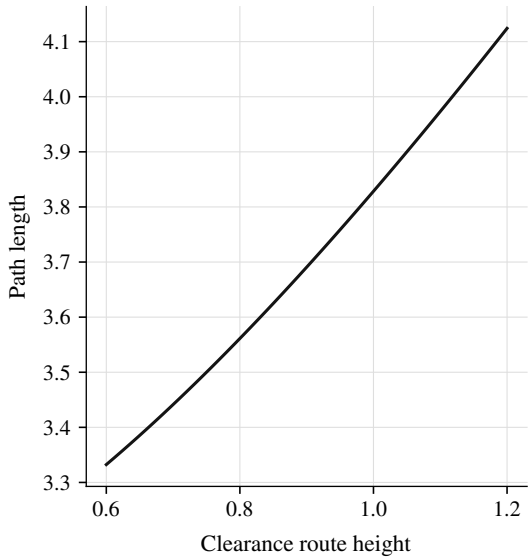
Boundary route: 1 + 2√1.36

Length: 3.332381

The footprint is 1 by 1.2.

If the boundary itself is excluded, this length is an infimum. Allowed paths approach it without necessarily attaining it.

C More clearance requires a longer route



D Distance depends on the allowed paths

Endpoint distance uses allowed curves.

Travel time also needs speed.

A shortest path need not be attained in an open region.

An edge-cost graph is a separate discrete model.

Figure 6.4: *The allowed region changes which route counts (continued).* The Euclidean segment, obstacle-avoiding alternatives and the boundary-limiting route share the same endpoints. The calculation labels the path family before comparing costs. This is a geometric model of a constrained translation task, not a measured vehicle trajectory.

Distance and travel time need not agree

A vehicle traveling at speed $v(x) > 0$ along a route takes time $\int ds/v(x)$. When v is a speed cap, this integral is a lower bound. It gives the modeled minimum if the vehicle can attain the cap throughout the route. A longer route can take less time at higher speeds. On a line split into two regions, this is already enough to separate shortest length from shortest time. In a smooth isotropic medium, the time metric is the physical metric multiplied by $1/v(x)^2$.

If costs depend on direction, the model may leave the Riemannian class and require a Finsler or more general path-cost description. To identify a translator cost with Riemannian length, check that both measure the same allowed routes with the same local rule.

The obstacle changed the allowed routes. Even with the region fixed, the local cost can vary from place to place. The following field keeps that cost explicit while we calculate the route it favors.

A survey route, from the ruler to the equation of motion

A survey team wants to connect two marked locations on a spherical model of Earth. They have latitude and longitude readings, but a straight segment on that rectangular display is not generally the shortest surface route. We can derive the route equation from the ruler rather than inserting a familiar curve by intuition.

Use colatitude θ, measured from the north pole, and longitude φ. On a sphere of radius R,

$$X(\theta, \varphi) = R(\sin\theta\cos\varphi, \sin\theta\sin\varphi, \cos\theta).$$

The two coordinate tangent vectors are

$$X_\theta = R(\cos\theta\cos\varphi, \cos\theta\sin\varphi, -\sin\theta),$$
$$X_\varphi = R(-\sin\theta\sin\varphi, \sin\theta\cos\varphi, 0).$$

Their squared lengths are R^2 and $R^2\sin^2\theta$; their inner product is zero. Thus

$$ds^2 = R^2(d\theta^2 + \sin^2\theta\, d\varphi^2).$$

This is a calculation about the surface, not a correction applied to an otherwise physical longitude ruler. A degree of longitude really spans a shorter route near the pole than at the equator.

Energy makes the derivation manageable

For a curve with a fixed parameter interval, define

$$E = \frac{R^2}{2}\int (\dot\theta^2 + \sin^2\theta\,\dot\varphi^2)\, dt.$$

Stationary curves satisfy the Euler–Lagrange equations. Varying θ gives

$$\ddot\theta - \sin\theta\cos\theta\,\dot\varphi^2 = 0.$$

Varying φ gives

$$\frac{d}{dt}(\sin^2\theta\,\dot\varphi) = 0,$$

or, away from the poles,

$$\ddot\varphi + 2\cot\theta\,\dot\theta\dot\varphi = 0.$$

The first equation tells us why most latitude circles are not geodesics. If θ is constant and the team actually moves, the equation requires $\sin\theta\cos\theta = 0$. Among nondegenerate latitude circles, only the equator satisfies it. A constant-longitude meridian, on the other hand, satisfies the equations with $\dot{\varphi} = 0$ and constant $\dot{\theta}$.

Energy and length are related but not identical functionals. Cauchy–Schwarz gives $L^2 \le 2E$ on a unit parameter interval, with equality for constant speed. Minimizing energy over parametrized connecting curves therefore selects a length-minimizing route with a constant-speed parametrization when a minimizer exists. Stationarity alone gives a geodesic; it does not prove that a long segment is the global minimum.

The connection coefficients have an operational meaning

The geodesic equation in coordinates is

$$\ddot{x}^k + \Gamma^k_{ij}\dot{x}^i\dot{x}^j = 0.$$

The nonzero coefficients in these spherical coordinates are

$$\Gamma^\theta_{\varphi\varphi} = -\sin\theta\cos\theta, \qquad \Gamma^\varphi_{\theta\varphi} = \Gamma^\varphi_{\varphi\theta} = \cot\theta.$$

They account for the changing coordinate basis. A route can have nonzero coordinate acceleration while remaining intrinsically straight. The team following a great circle need not keep its displayed longitude rate constant.

The apparent divergence of $\cot\theta$ at a pole belongs to the coordinates. Longitude loses its ability to distinguish directions there. An equation written in a regular stereographic chart, or directly for the position vector in $\mathbb{R}^3$, remains well behaved. This is a practical reason to keep an atlas: the computer should change descriptions before its current one loses numerical meaning.

For a unit-speed spherical route $x(s)$, the room-coordinate equation is especially simple:

$$x''(s) = -\frac{x(s)}{R^2}.$$

The acceleration points normally inward. Its tangent component is zero, which is precisely the geodesic condition for the induced metric. Given $x(0) = p$ and a unit tangent v with $p \cdot v = 0$,

$$x(s) = p\cos(s/R) + Rv\sin(s/R).$$

The solution lies in the plane spanned by p and v. Its intersection with the sphere is a great circle. We have obtained the familiar route from the metric and the constraint.

One numerical route in three descriptions

Take $R = 1$, $p = (1, 0, 0)$ and $q = (4/5, 0, 3/5)$. Their short great-circle angle is

$$\alpha = \arccos(4/5) = 0.6435011088\ldots.$$

The initial unit tangent is $v = (0, 0, 1)$. The route is $(\cos s, 0, \sin s)$ for $0 \le s \le \alpha$. In colatitude coordinates, $\theta = \pi/2 - s$ and $\varphi = 0$; the metric integral is $\int_0^\alpha ds = \alpha$.

In the north stereographic chart used earlier, the route runs from $u = 1$ to $u = 2$ at $v = 0$. Its length is

$$\int_1^2 \frac{2}{1+u^2}\,du = \alpha.$$

The south chart runs from $a = 1$ to $a = 1/2$ and returns the same value. Three coordinate calculations therefore meet at one physical quantity. The coincidence is explained by their common metric, not by a numerical normalization performed after the fact.

The straight chord has length $\sqrt{2/5} \simeq 0.632456$. It is shorter because it travels through the interior. The long great-circle arc has length $2\pi - \alpha$. It is a geodesic route as well, but it is not the shortest connecting segment. The three quantities answer three different questions: shortest surface route, unconstrained ambient separation, and length of a specified long route.

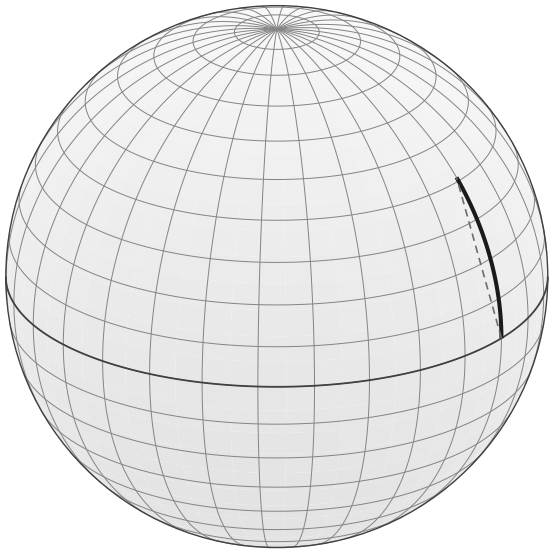
What the map must carry into the calculation

A useful route record contains the chart, metric, parameter convention, and allowed region. If a route is forbidden from crossing a boundary, the shortest admissible path may run along that boundary and be only piecewise smooth. If speed limits or direction-dependent effort are introduced, minimizing length may no longer minimize travel time or energy. The spherical geodesic remains a geometric reference against which those operational changes can be stated.

The map/globe distinction now has a complete chain: a point on the globe gives chart coordinates; the derivative gives tangent components; the metric measures them; and the variational equation selects a locally straight route. Each stage preserves a different kind of information. Confusing the stages is how an innocent-looking coordinate shortcut becomes a different physical problem.

The arc is a physical route

Map interval, chord and surface arc measure different objects.



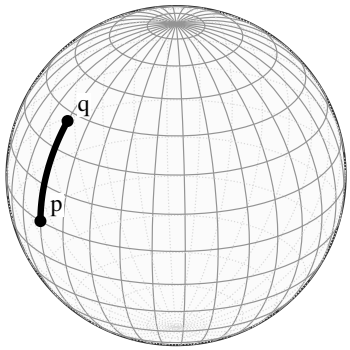
p = (1, 0, 0)

q = (4/5, 0, 3/5)

Arc: arccos(4/5)

Chord: √(2/5)

A Follow one specified spherical route



B Two coordinate reports trace the same route

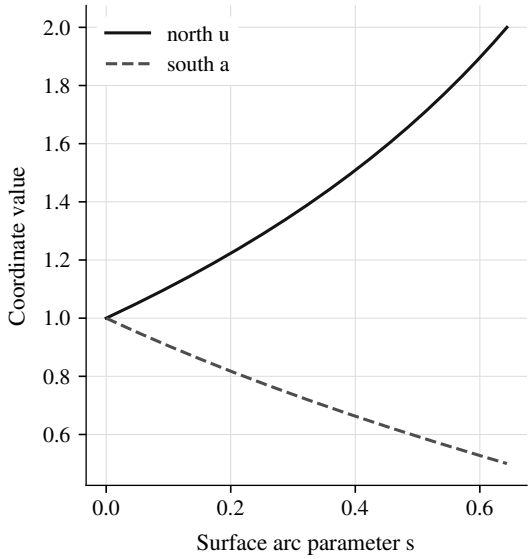
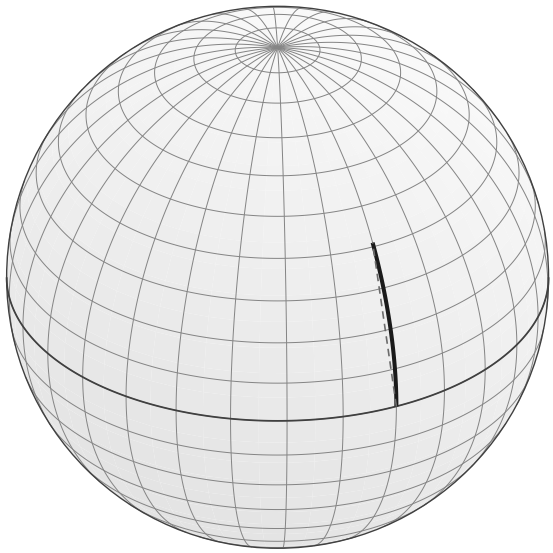


Figure 6.5: *The same surveyed route in three descriptions.* The study continues on the following figure page.

The arc is a physical route

Map interval, chord and surface arc measure different objects.



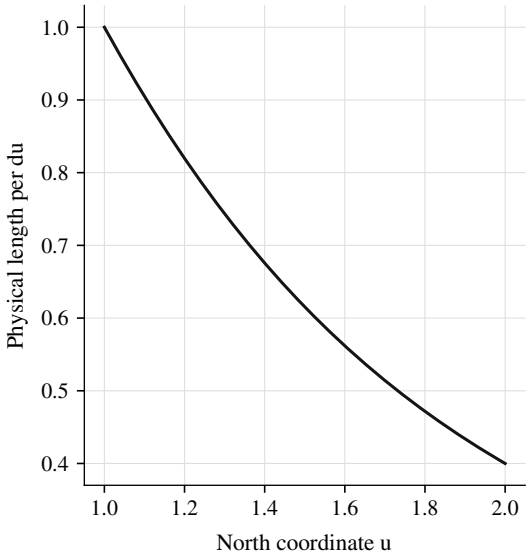
p = (1, 0, 0)

q = (4/5, 0, 3/5)

Arc: arccos(4/5)

Chord: √(2/5)

C The chart ruler recovers physical length



D Arc, chord and winding answer different questions

Shortest surface arc: 0.643501

Interior chord: $\sqrt{(2/5)} = 0.632456$

Long geodesic: $2\pi - 0.643501$

Radius R multiplies all lengths.

Figure 6.5: *The same surveyed route in three descriptions (continued).* The spherical route, two coordinate histories and transformed local ruler give the same intrinsic length. The chord and long great-circle arc remain distinct comparison quantities.

Survey calculation. Repeat the three length calculations at $R = 10$. Keep the chart coordinates dimensionless and identify every quantity that acquires a factor of R. Then use length-valued chart coordinates and explain why the printed metric coefficients change again.

Closing the courtyard changed the available journey while leaving its endpoints fixed. Changing map projections altered the coordinates while leaving a correctly measured surface route unchanged. Changing speed limits changed the travel-time question. Distance therefore belongs to a ruler and a family of permitted paths. To approximate those calculations locally, we next need to know how the error grows as the journey extends beyond the starting neighborhood.

Practice with the construction

6.1. At $R = 2$, compare arc and chord lengths for angular separation $\pi/3$.

6.2. A program estimates a shortest path using only one chart. Describe a test that would reveal whether a shorter route leaves the chart.

Chapter 7

Curvature and local approximation

Two survey teams start a kilometre apart on the equator and walk due north. Their routes are locally straight on the globe, yet they converge toward the pole. A flat model with parallel northbound routes predicts a constant separation. How far can we trust the flat approximation?

Choose the approximation we are testing

A surveyor can make two different approximations. Treating a patch of the globe as flat changes its geometry. Using a derivative to estimate a finite conversion between map coordinates approximates a function. Each shortcut needs an error bound for the quantity it changes.

How far can the tangent map take us?

Touch a flat sheet to a globe. Mark a circle on each surface at the same measured radius from the point of contact. Measure each radius along its own surface. Comparing the two circumferences tests how far the flat ruler departs from the spherical one.

On a sphere of radius R, a geodesic circle at distance s from its center has circumference

$$C(s) = 2\pi R \sin(s/R).$$

Compare the Euclidean prediction $2\pi s$. Expanding the sine gives

$$\frac{C(s)}{2\pi s} = 1 - \frac{s^2}{6R^2} + O(s^4/R^4).$$

Because the sphere's Gaussian curvature is $K = 1/R^2$, the leading relative discrepancy is $Ks^2/6$. Both curvature and the size of the region enter. A curvature value by itself does not tell us whether a particular approximation is adequate.

There is an explicit non-asymptotic statement on this example. For $0 \le u \le 1$, the alternating sine series gives $u - u^3/6 \le \sin u \le u$. Therefore

$$0 \le 1 - \frac{C(s)}{2\pi s} \le \frac{s^2}{6R^2}, \qquad 0 < s \le R.$$

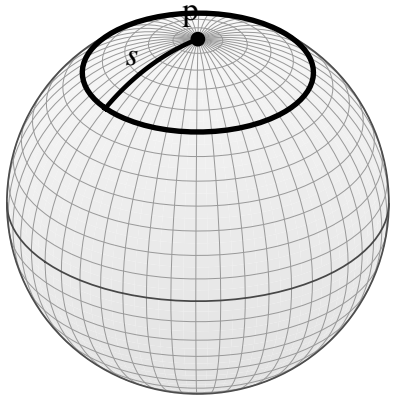
At $s/R = 0.1$, the upper bound is about 0.167%. At $s/R = 0.5$, it is about 4.17%. If I need the circle's circumference to within one percent of the flat prediction, the tenth-radius survey passes this test. At half a radius the bound no longer certifies it. An error budget for a learned surface needs its own calculation.

An encoder also introduces approximation error of its own. For a twice differentiable map f on a convex region, suppose its second derivative has operator norm at most B. Along the line segment from x to $x + h$, Taylor's integral formula gives

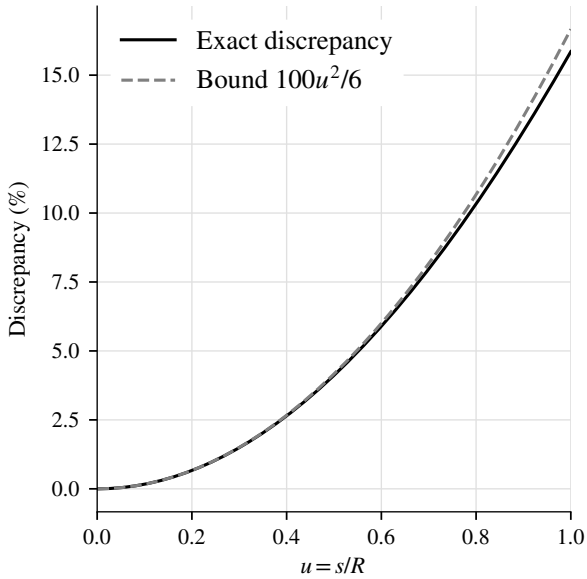
$$f(x + h) - f(x) - Df_x h = \int_0^1 (1 - t) D^2 f_{x+th}[h, h], dt,$$

so the remainder has norm at most $B\|h\|^2/2$. This is a derivative bound for the map in stated coordinates, distinct from the intrinsic curvature tensor of a Riemannian manifold. Both can

A *Measure radius along the sphere*



B *The flat formula overestimates*



s / R	*Sphere circumference / flat*	*Exact loss*	*Upper bound*
0.1	0.998334	0.1666%	0.1667%
0.5	0.958851	4.1149%	4.1667%
1.0	0.841471	15.8529%	16.6667%

Figure 7.1: A geodesic circle on a sphere and the error in its flat circumference approximation. The computed values show how the dimensionless radius s/R controls the discrepancy. The upper bound applies on the stated interval $0 < s \leq R$.

matter in the same system. Naming both quantities curvature would conceal which hypothesis controls which error.

Scalar curvature compresses still more information. In dimensions above two it sums sectional contributions and can hide how different planes behave. Two particular directions can see large curvature even when a scalar summary is small. Later chapters will return to this issue when a single rotation score conceals several independent planes of motion.

The small-circle calculation measured departure from a flat ruler. We now return to the two surveyors and measure how the separation of neighboring routes changes, before comparing that intrinsic effect with an ordinary coordinate approximation.

Two surveyors walk away from the equator

Two surveyors start one kilometer apart on the equator and walk north along meridians. On a round globe their paths converge toward the pole. On a flat map, parallel northbound lines would maintain their separation. The discrepancy is not caused by an error in either surveyor's local compass bearing; it is built into the family of routes on the surface.

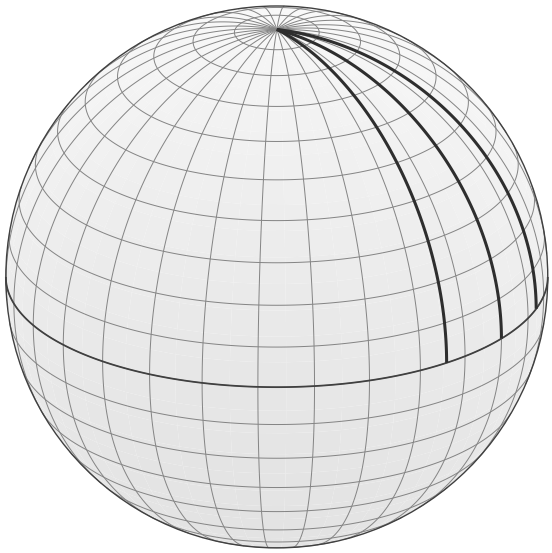
On a sphere of radius R, two meridians separated by longitude $\Delta\varphi$ have parallel-of-latitude separation $R\cos\lambda\,\Delta\varphi$. This is the length along that latitude circle, not generally the shortest

geodesic separation. For infinitesimally neighboring meridians the corresponding transverse Jacobi field has magnitude proportional to $\cos \lambda$. It reaches zero at the pole.

The meridian pair starts apart on the equator with both teams heading north. In the figure's separate curvature test, two teams start at the same point and turn their headings apart. Their initial separation is zero, so $J(0) = 0$. Dividing by the small heading difference makes the initial separation rate $J'(0) = 1$. Those starting conditions explain why the spherical sine and cosine curves describe different experiments.

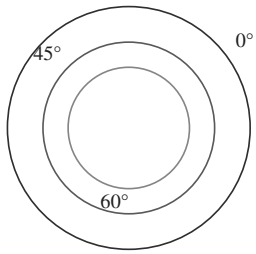
Neighboring routes meet at the pole

Unit sphere / spacing along a latitude differs from shortest surface distance.



$$J(\lambda) = \cos \lambda$$

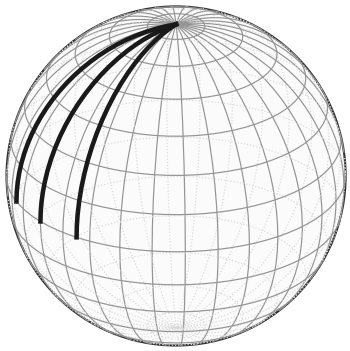
J is relative to equatorial spacing.
Finite latitude arc: cos(λ) Δϕ
At the pole the routes meet.



Plan radii: cos λ

A Northbound meridians converge

B Curvature changes neighboring-route separation



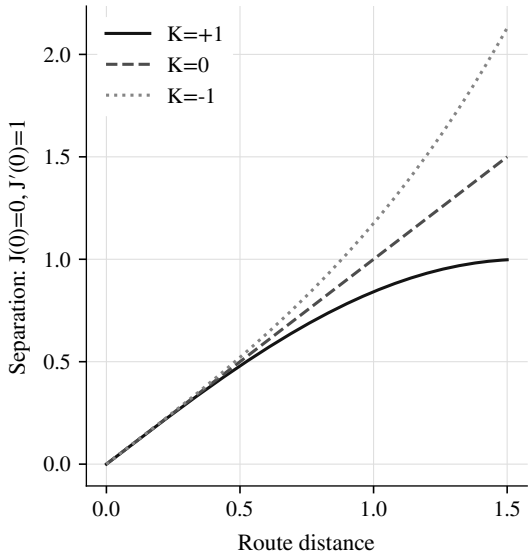
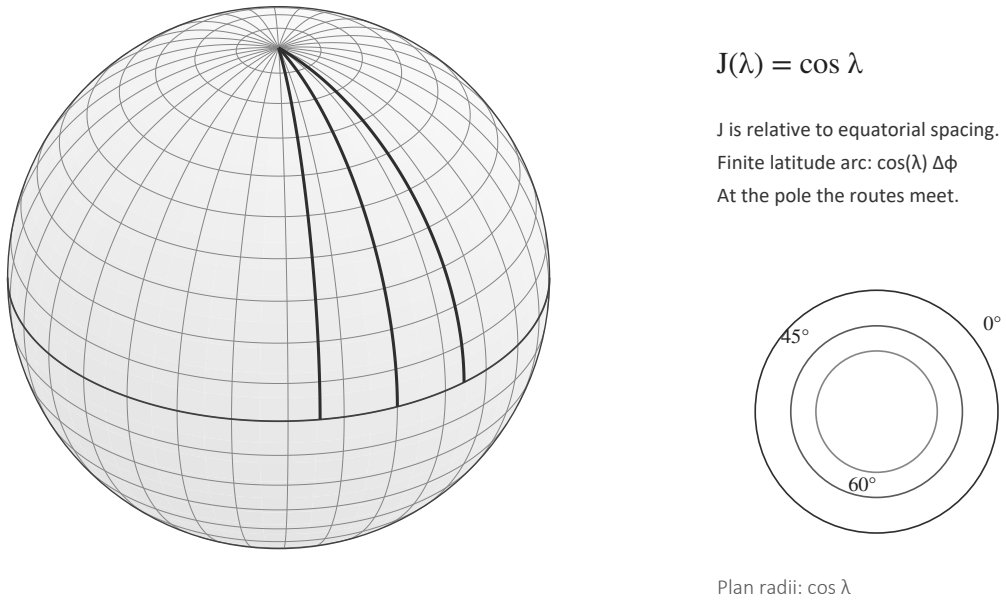
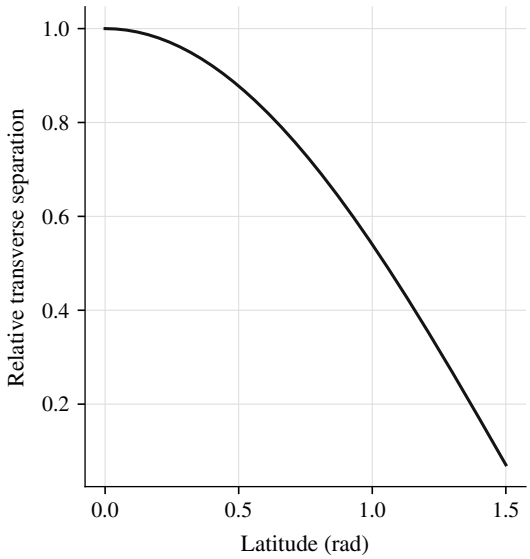


Figure 7.2: *The surface changes the separation of neighboring routes.* The study continues on the following figure page.



C Meridian spacing falls toward the pole

D An arc length and a Jacobi field differ



A finite latitude arc is R cos(λ) Δφ.

The infinitesimal family gives a Jacobi field.

Matched initial data matter when comparing different curvatures.

Figure 7.2: *The surface changes the separation of neighboring routes (continued).* Northbound meridians approach one another as latitude increases. The comparison plot uses matched Jacobi initial data $J(0) = 0$, $J'(0) = 1$: $\sin s$, s and $\sinh s$ for curvatures $+1, 0, -1$. The separate cosine curve describes meridian separation relative to its equatorial value.

Evaluate the circumference at two survey scales

On a sphere, the circle at geodesic radius r from a pole has circumference

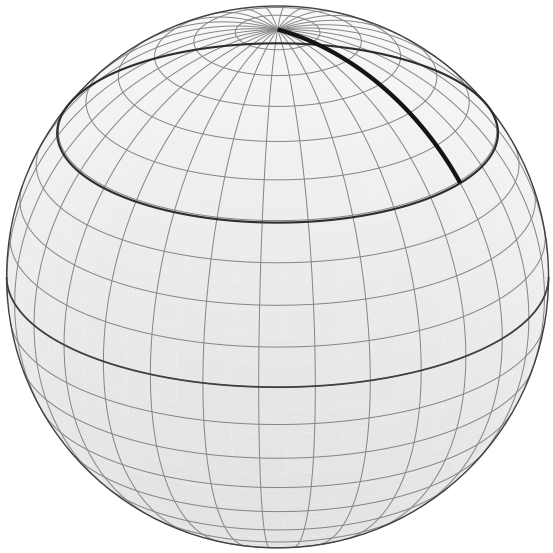
$$C(r) = 2\pi R \sin(r/R) = 2\pi r \left(1 - \frac{r^2}{6R^2} + O(r^4/R^4)\right).$$

Its radius has been measured along the surface, not through the solid globe. Compared with a flat disk of the same measured radius, the circumference is smaller. A constant negative-curvature model with curvature $-1/R^2$ instead has $C(r) = 2\pi R \sinh(r/R)$ and a larger circumference.

At $r/R = 0.2$, the spherical circumference is about 0.993347 times the flat value. At $r/R = 1$, the ratio is about 0.841471. A flat local approximation can therefore be very good at one scale and visibly wrong at another. The statement “locally Euclidean” does not specify an error tolerance or a permitted radius by itself.

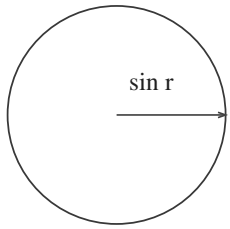
A circle measured from within the surface

Unit sphere / r is measured along a meridian, from the pole to the circle.



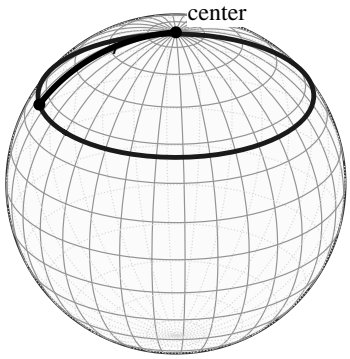
$$C(r) = 2\pi \sin r$$

The bold segment lies on the surface.
Radius is measured along that segment.
The surrounding grid supplies orientation.



Plan: section radius = sin r

A Measure the radius along the surface



B Curvature changes circumference

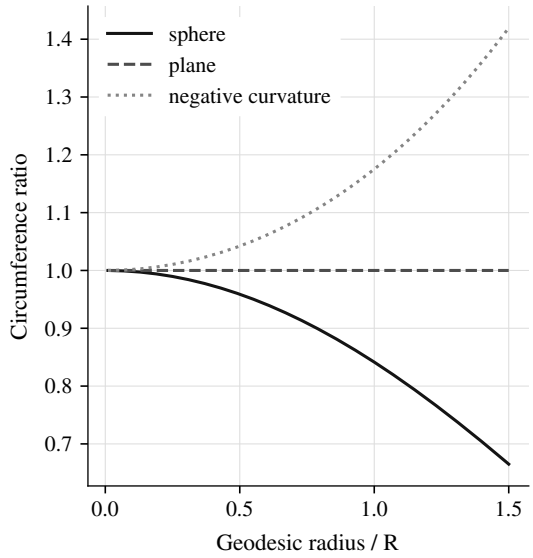
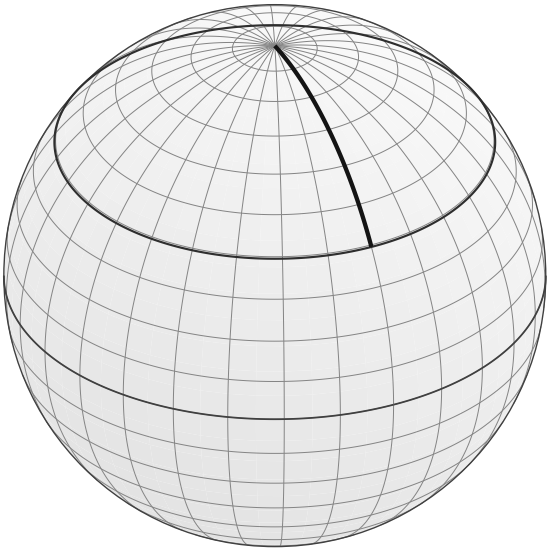


Figure 7.3: *Measure the radius on the surface, then measure the circle.* The study continues on the following figure page.

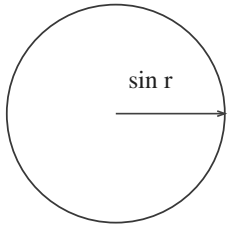
A circle measured from within the surface

Unit sphere / r is measured along a meridian, from the pole to the circle.



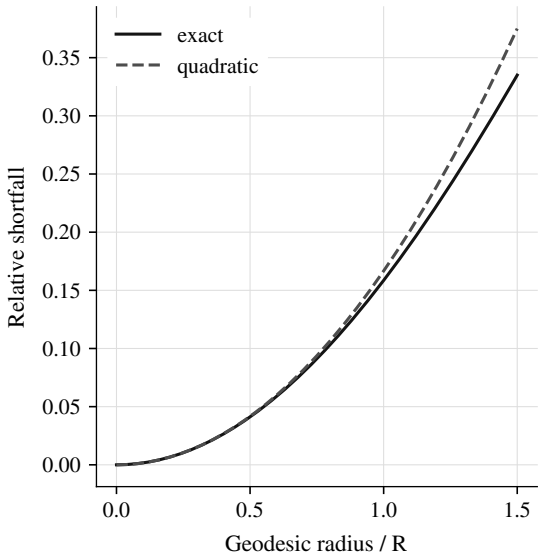
$$C(r) = 2\pi \sin r$$

The bold segment lies on the surface.
Radius is measured along that segment.
The surrounding grid supplies orientation.



Plan: section radius = sin r

C A flat formula overestimates a spherical circle



D Three surfaces, three circumference rules

Sphere: $C = 2\pi R \sin(r/R)$
Plane: $C = 2\pi r$
Negative curvature: $2\pi R \sinh(r/R)$

The radius r is measured along the surface.

Figure 7.3: *Measure the radius on the surface, then measure the circle (continued).* The marked radius follows the sphere from the circle's center to its boundary. The flat and curved circumference comparisons use that intrinsic radius. Exact circumference ratios and their leading approximations are compared over a declared range. The sign of curvature appears in a measured geometric quantity.

Apply the remainder bound to the converter

Let f convert a position on one survey grid to another. For a twice differentiable f on a convex coordinate region, a bound $\|D^2 f\| \le M$ implies

$$\|f(x+h) - f(x) - Df(x)h\| \le \tfrac{1}{2}M\|h\|^2$$

whenever the segment from x to $x+h$ stays in the region. The bound follows by integrating the change in the first derivative along that segment. It is a statement about this coordinate map and these norms.

A curvature bound is related to how geometric quantities depart from flat models, but it is not an interchangeable replacement for every derivative bound on every encoder. An encoder can oscillate rapidly on a flat domain. A curved manifold can be described by a well-conditioned chart on a small region. To control a computational approximation, identify the quantity being approximated and the derivatives or geometric comparison theorem that controls it.

In the translator setting, this means that local agreement of point values is not enough. The derivative controls how small observational errors move, and second derivatives control how quickly that local description changes. For example, Chapter 4's north-to-south map is $f(u) = 1/u$ along $v = 0$. From $u = 2$ to 2.1, its derivative predicts a change of -0.025. The exact change is about -0.023810. Since $|f''(u)| \le 0.25$ on that segment, the error 0.001190 lies below the bound 0.00125. With an allowance of 0.002 in the south chart, I'd accept that shortcut for this step. With an allowance of 0.001, I'd run the exact conversion.

What the translator's derivative contributes

Suppose a survey program converts latitude and longitude into map coordinates. A small change in either input can affect both outputs. The Jacobian collects those first-order responses, which the Taylor bound tells us how far to trust.

When you have a function $f : \mathbb{R}^n \to \mathbb{R}^m$—like an encoder mapping images to embeddings, or a translator converting one feature space to another—the *Jacobian* of f at a point x is the matrix of all its partial derivatives:

$$J_f(x) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}.$$

This matrix tells you how f *distorts space locally* near x. Specifically:

- If you take a tiny step δx away from x, the Jacobian predicts where you'll land: $f(x + \delta x) \approx f(x) + J_f(x) \cdot \delta x$.
- The singular values of $J_f(x)$ tell you how much f stretches or compresses in different directions near x. Large singular values mean f magnifies distances; small singular values mean f collapses them.
- If $n = m$ and $J_f(x)$ is invertible, the inverse function theorem gives a smooth local inverse near x. For a rectangular Jacobian, full rank instead describes an immersion or submersion; it does not automatically give a coordinate change between the entire spaces.

Think of the Jacobian as the "zoom lens" view of a function: if you zoom in close enough to any point, every smooth function looks approximately linear, and the Jacobian is that linear

approximation. For translators, this tells us whether small changes in one embedding space correspond to small, predictable changes in another—or whether the translator magnifies noise, collapses meaningful variation, or behaves erratically.

Riemann's metric construction and Jacobi's tools for coordinate transformations serve complementary roles here. The metric determines local measurement on a chosen smooth space. A derivative tracks how a conversion changes the components of a displacement. Integrating local metric comparisons controls route lengths when the bounds hold along the entire route. Comparing endpoint distances additionally requires the same admissible path problem.

For encoders, this is a concrete test. A small observed residual at one input supplies a value-level check. The Jacobian reveals which nearby distinctions can be magnified or flattened, and a second-derivative bound controls how quickly that local approximation changes. The next chapter takes the limiting case seriously, a direction that the instrument cannot see at all.

From local measurements to a finite comparison

The two surveyors began on neighboring meridians. Their locally straight routes converged because of the sphere's geometry, while the small-circle calculation put a bound on the error of a flat approximation. The Taylor calculation then separated that intrinsic question from the error of replacing a coordinate map by its derivative.

A metric supplies lengths, and compatible Riemannian tensors carry that measurement across charts. Uniform bounds control the specified curves; endpoint distances also require corresponding domains and allowed paths. An encoder supplies a chart where its dimension, smoothness and local invertibility justify the role.

The approximation now has a region and an error allowance. It still assumes the instrument records the directions being compared. We test that assumption next: a better ruler cannot restore an unobserved direction.

Practice with the construction

Back to the maps and the globe. The tangent paper gives a useful local map, with an explicit range of accuracy. The globe determines the geometry; the chosen region and the calculation determine whether the map is good enough for the task.

7.1. What bound on s/R guarantees that the circumference discrepancy in this example is at most one percent?

7.2. For $f(x) = x^2$ on $[-2, 2]$, calculate the linearization error at $x = 1$ exactly and compare it with the stated second-derivative bound.

Chapter 8

Pullbacks and information loss

Slide a marked plate north while its sensor reports only eastward position. The mark moves and the reading stays fixed. We will test what an instrument sees near one position, whether two separated positions can share a reading, and how much measurement error grows when we try to recover the original position.

An instrument can flatten a direction

Use east and north coordinates for the mark's position on the workbench, (x, y) and let the measurement be $f(x, y) = x$. Pulling back the ordinary metric du^2 from the sensor produces

$$f^*(du^2) = dx^2, \qquad G_f = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}.$$

The northward tangent $(0, 1)$ has zero length in this quadratic form. The form is positive semidefinite, not positive definite. It is therefore not a Riemannian metric on the two-dimensional plate. It can still describe exactly which motion the sensor detects.

Proposition 8.1 (When a pullback is a metric)**.** *Let $f : M \to N$ be smooth and let h be a Riemannian metric on N. The pullback f^*h is positive definite at p if and only if Df_p is injective. Consequently it is a Riemannian metric on M exactly when f is an immersion.*

Proof. For $v \in T_pM$, $(f^*h)_p(v, v) = h_{f(p)}(Df_p v, Df_p v)$. Positive definiteness of h makes this positive precisely when $Df_p v \neq 0$. Requiring that for every nonzero v is injectivity of the derivative. Smoothness follows from smoothness of f and h. □

Now turn a pointer on a dial and let t count the accumulated angle in radians. Its normalized tip coordinates are $f(t) = (\cos t, \sin t)$. The derivative is nonzero everywhere, yet t and $t + 2\pi$ put the tip in the same place. A still image shows which way the pointer faces and leaves the full turns unrecorded.

Near loss is measurable too

Suppose we add a northward sensor to the plate setup, but its gain is small. The combined report is $f_\varepsilon(x, y) = (x, \varepsilon y)$ with $0 < \varepsilon \leq 1$. Its pulled-back metric is $\mathrm{diag}(1, \varepsilon^2)$. The smallest singular value of the derivative is ε. A measurement error of size η in the second output coordinate can correspond to an input error η/ε. At $\varepsilon = 0.01$, a small sensor error can be magnified a hundredfold by inversion.

The same amount of noise can move the estimate a hundred times as far north as east. If the job depends on northward motion, I test that channel first. A learned encoder can hide the same weakness.

For an encoder $f : U \subset \mathbb{R}^d \to \mathbb{R}^m$, the matrix $Df_x^{\mathsf{T}} Df_x$ contains this local information. Its eigenvalues are the squared singular values. If $m < d$, full column rank is impossible. If $m \geq d$, it is possible but must be established on the region in question. A sample of full-rank Jacobians is evidence on sampled points, not a proof that rank never fails between them.

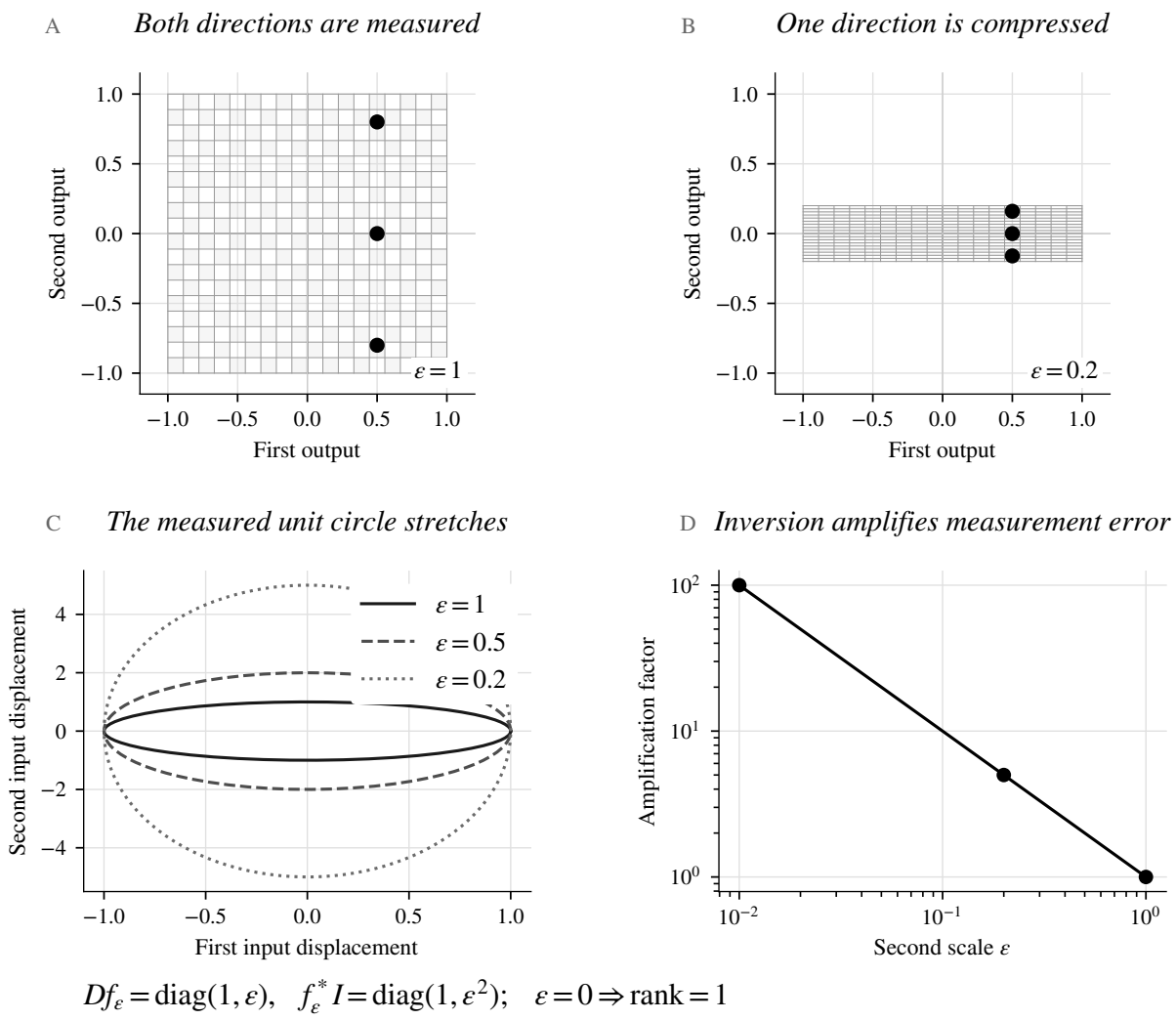


Figure 8.1: The measurement $f_\varepsilon(x, y) = (x, \varepsilon y)$ compresses a coordinate grid. Its equal-measured-length ellipses expand in the insensitive input direction, and inversion amplifies output error by $1/\varepsilon$. At zero scale the derivative loses rank.

What survives a summary

The room-temperature display can use the probe's reading to mark a one-degree band. The band label is enough to count how many readings fall in that band. Recovering the original reading requires finer information. The map has become a quotient of the recorded distinctions. We should write that quotient explicitly instead of asking the summary to recover information it discarded.

GIGI's distribution example will do precisely this. Its polynomial identifies bin proportions with fixed bin edges. Within-bin positions are not part of that particular object. Exact preservation of the polynomial can therefore coexist with loss of individual measurements. The preserved object is more specific than the word data suggests.

The one-coordinate sensor lost a northward displacement. A camera gives a second, inspectable instance of the same rank calculation; we then compare that loss with a periodic report whose derivative remains nonzero.

A camera can lose depth while keeping a sharp picture

Take a camera model that keeps the same magnification at every depth. The viewing direction is the Z axis. This orthographic camera maps (X, Y, Z) to (X, Y). Its Jacobian has rank two. A displacement purely in depth, $(0, 0, dZ)$, disappears. Pulling back the image's Euclidean metric gives

$$g = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

This is a positive semidefinite quadratic form on three-dimensional space, not a Riemannian metric there. A nonzero motion can have zero measured length. The image can still be useful; its geometry simply cannot distinguish everything the three-dimensional object can do.

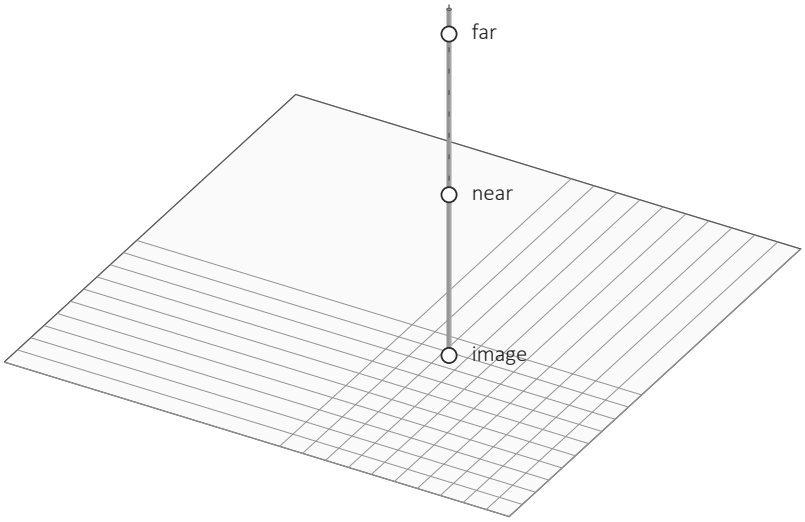
If the object is known to lie on the graph $Z = f(X, Y)$, the same two image coordinates may identify its point uniquely within that constrained model. The induced physical surface metric is then

$$g_{\text{surface}} = I + \nabla f \nabla f^{T}.$$

If the mark is on a ramp of known shape, each image position tells me where on the ramp to look. The slope tells me how much depth changes when the mark moves across the image. That's the extra term in the surface ruler. It comes from knowing the ramp's shape.

A measured image can hide a physical direction

Orthographic observation F(X,Y,Z) = (X,Y); depth lies in the derivative kernel.



DF has two measured directions.

(0, 0, 1) is invisible to the image.

F*I = diag(1, 1, 0)

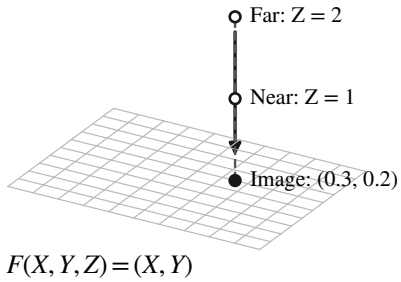
Rank: 2

Extra information must be supplied.

A specified surface Z = f(X,Y)
can restore a depth model.
The model and the measurement
are different sources of information.

A Different depths produce the same image

B The image ruler assigns zero cost to depth



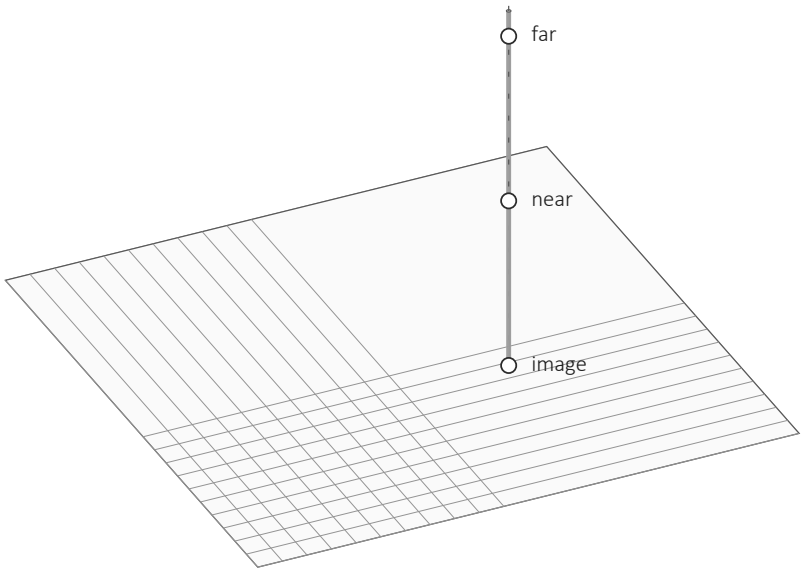
$$F(X,Y,Z) = (X,Y)$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Figure 8.2: *A sharp image can still have an invisible direction.* The study continues on the following figure page.

A measured image can hide a physical direction

Orthographic observation F(X,Y,Z) = (X,Y); depth lies in the derivative kernel.



DF has two measured directions.

(0, 0, 1) is invisible to the image.

F*I = diag(1, 1, 0)

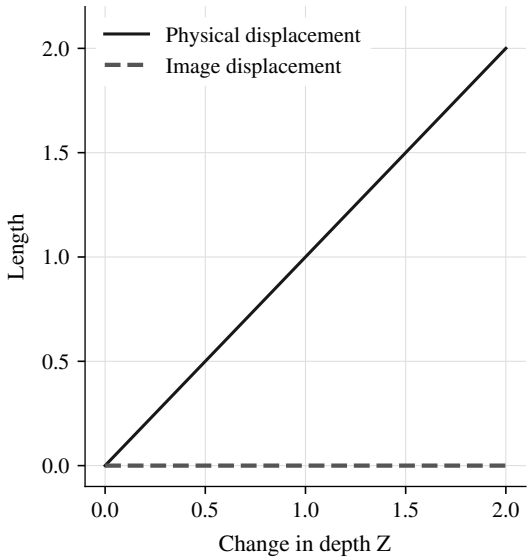
Rank: 2

Extra information must be supplied.

A specified surface Z = f(X,Y)
can restore a depth model.
The model and the measurement
are different sources of information.

C Physical motion can leave the image unchanged

D A surface model supplies missing information



If Z=f(X,Y), a point may be recoverable.

Physical surface metric: $I + \nabla f \, \nabla f^T$

The model supplies the missing depth.

Observation and reconstruction are distinct.

Figure 8.2: *A sharp image can still have an invisible direction (continued).* The orthographic model $F(X, Y, Z) = (X, Y)$ sends two depths to one image location. Its pullback has a zero depth coefficient: motion along the viewing ray changes physical position but not the image. A declared surface equation $Z = f(X, Y)$ supplies additional reconstruction information; it is not supplied by the camera measurement itself.

Full local rank does not make a map globally one-to-one

For the dial, write θ for the accumulated angle and $F(\theta) = (\cos\theta, \sin\theta)$ for the tip position. Its derivative has norm one at every real θ. It preserves infinitesimal length, yet $F(\theta) = F(\theta + 2\pi)$. Local rank has not prevented global identifications.

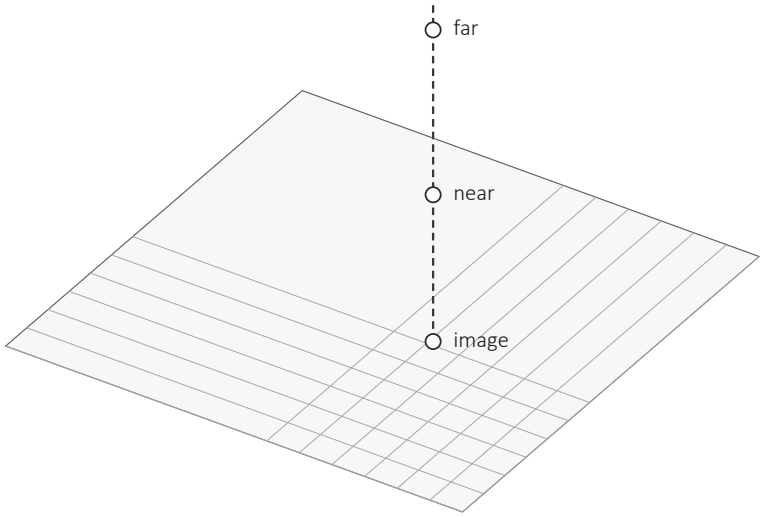
The plate test showed a direction the camera couldn't see even up close. The dial is different. I can watch the pointer move a little and see which way it went. If I look away and later see it at the same mark, I know the angle within a turn. I still don't know how many turns it made. To keep count, I need a starting angle and enough observations to follow the motion. That's what lets us lift the dial reading to a running angle.

Apply Proposition 8.1 to both instruments. The camera derivative kills every depth vector $(0, 0, dZ)$, so its kernel has dimension one. The dial derivative sends a scalar tangent step to that step times the unit vector $(-\sin\theta, \cos\theta)$, so its kernel is zero. The dial preserves local motion while leaving the full-turn count unresolved.

The proposition is local. Global injectivity, preservation of path distances and a stable inverse each require further hypotheses. A faithful comparison system should list those requirements separately rather than hoping that the word "embedding" supplies all of them.

A whole direction disappears in projection

Orthographic image F(X,Y,Z) = (X,Y). The points on one ray share one image coordinate.

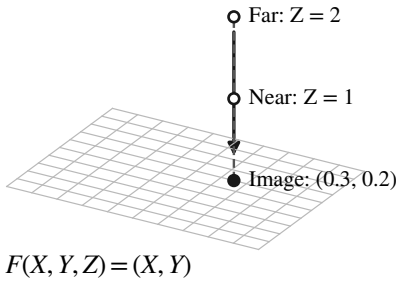


Moving along the viewing ray
does not change either output.
The depth direction is in ker DF.

F*I = diag(1, 1, 0).
Derivative rank = 2.
A depth displacement measures 0.

A Projection loses an entire direction

B A locally regular map can still repeat



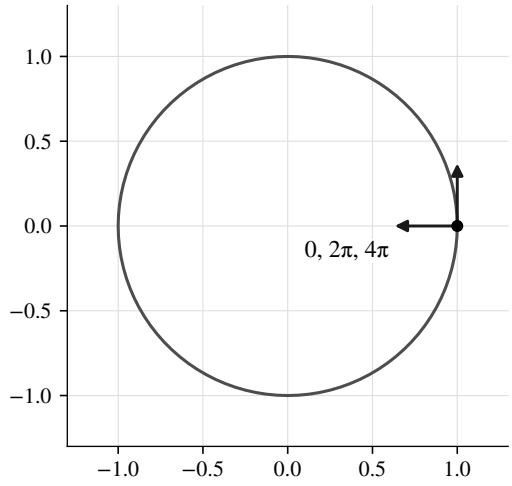
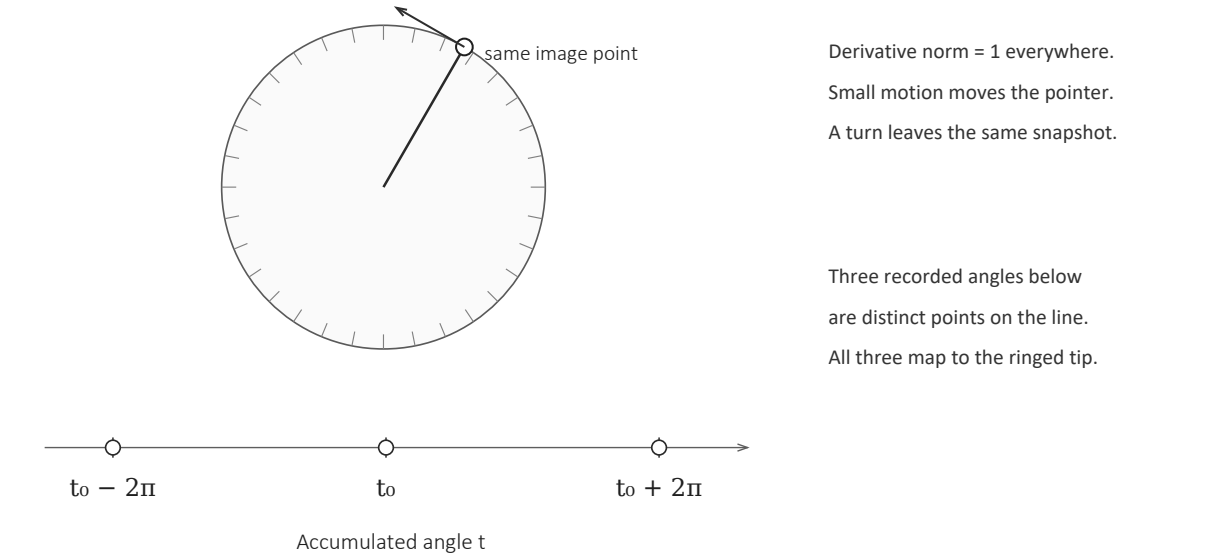


Figure 8.3: *Local rank and global identification answer different questions.* The study continues on the following figure page.



C Rank loss creates invisible motion

D Repeated turns require a lifted history

Rank deficient	Derivative has norm one
Nonzero depth has zero length	Global map is not one-to-one
A semidefinite pullback	Local lengths are preserved
An observation nullspace	Turn history needs a lift

Figure 8.3: *Local rank and global identification answer different questions (continued).* A depth-losing projection is compared with a full-rank periodic parametrization. Their Jacobians and fibers show two distinct sources of ambiguity. The plot separates derivative rank, global one-to-one behavior and task-appropriate quotienting.

A small singular value is a practical warning

Suppose a position sensor takes two position coordinates and returns two numbers. Its 2×2 derivative has singular values 1 and 10^{-3}. Both input directions survive, but recovering a displacement in the weak direction amplifies observation error by a factor of one thousand. An exact theorem about injectivity can coexist with a numerically unreliable inverse.

If $\sigma_{\min}(DF_x) = \inf_{\|v\|=1} \|DF_x v\| \geq s > 0$, the derivative has full column rank and its inverse on its image has norm at most $1/s$. A smooth map with that rank has an inverse on its local image. A global bound also needs control of the region and the map's global behavior. The elementary example $F(\theta)$ already shows why a uniform local singular-value bound cannot rule out periodic wrapping.

Back at the plate with the weak north sensor, I'd move the mark the same distance east and north and compare the readings. The ellipse in that drawing is wider north to south because it takes a bigger move that way to change the reading by the same amount. An average over lots of easy moves could hide that weakness.

For a rotating instrument, ask whether the sensor detects small turns, distinguishes different positions, and permits reconstruction in the presence of noise. Comparing three proposed sensors shows what an extra reading can recover and when a recorded history supplies the missing information.

A rotating instrument seen by three different sensors

Consider a pointer rotating freely about a fixed axle. Its state is an angle θ modulo 2π. There is one pointer and one physical configuration at each angle, but different sensors retain different parts of that configuration. This small instrument lets us distinguish a chart, a covering observation, and a lossy measurement without leaving the workbench.

The first sensor reports the horizontal coordinate of the pointer tip:

$$h_1(\theta) = \cos\theta.$$

The second reports both tip coordinates:

$$h_2(\theta) = (\cos\theta, \sin\theta).$$

The third sees an unmarked bar through the axle and cannot tell its ends apart:

$$h_3(\theta) = (\cos 2\theta, \sin 2\theta).$$

These maps do not solve the same observation problem. The first confuses θ with $-\theta$. The second distinguishes all physical states on the circle. The third identifies θ and $\theta + \pi$ everywhere.

Differentiate what the sensor actually reports

Their derivatives are

$$\begin{aligned} h_1'(\theta) &= -\sin\theta, \\ h_2'(\theta) &= (-\sin\theta, \cos\theta), \\ h_3'(\theta) &= 2(-\sin 2\theta, \cos 2\theta). \end{aligned}$$

With the ordinary Euclidean ruler on the sensor outputs, the pullback quadratic forms are

$$g_1 = \sin^2\theta\, d\theta^2, \qquad g_2 = d\theta^2, \qquad g_3 = 4\, d\theta^2.$$

At the rightmost pointer position, h_1' vanishes. The horizontal sensor cannot detect an infinitesimal angular motion there to first order. Its pullback is degenerate at that point. The two-coordinate tip sensor has no such rank loss.

The bar sensor has full derivative rank at every point, yet it is globally two-to-one. Its local sensitivity is twice that of the tip sensor in these raw coordinates. Dividing its output metric by four recovers the physical angular ruler locally, but it does not restore the missing distinction between the bar's two ends. Metric calibration and global identifiability are separate operations.

A noise calculation exposes the blind direction

Suppose the horizontal output has a small additive error e. Away from the extrema, first-order inversion gives

$$\Delta\theta \simeq -\frac{e}{\sin\theta}.$$

At $\theta = \pi/2$, the angular error has approximately the same numerical magnitude as e. At $\theta = 0.01$, it is amplified by approximately one hundred. Near zero the linear approximation itself becomes unreliable unless e is much smaller than θ^2.

Exactly at zero, use the quadratic expansion

$$1 - \cos\theta = \frac{\theta^2}{2} + O(\theta^4).$$

A small negative additive error e at the reported maximum determines an angular magnitude of order $\sqrt{2|e|}$, with two possible signs. An upward perturbation can leave the physical range $[-1, 1]$ entirely. Clipping that report to one hides the inconsistency; it does not recover the true angle.

The two-coordinate sensor behaves differently. Near a point on the unit circle, the component of output error along the tangent changes the angle to first order; the radial component changes the norm and can serve as a consistency residual. Writing $u = (\cos\theta, \sin\theta)$ and $t = (-\sin\theta, \cos\theta)$,

$$\Delta\theta \simeq t^T e, \qquad \Delta\|u\| \simeq u^T e.$$

If the measured tip moves along the circle, I update the angle. If the reading jumps outward, the tip appears to be farther from the axle than the model allows. That part of the error is a check on the measurement, even if the angle estimate barely changes.

What a second observation can recover

If the first sensor is supplemented by a vertical tip sensor, their combined map is h_2. The added measurement removes both the derivative blind spots and the global reflection ambiguity. If the third sensor is supplemented by another sensor with the same bar symmetry, the two-sheet ambiguity can remain even if the data array becomes larger.

A useful reconstruction question is therefore not "how many features do we have?" but "which states have exactly the same complete observation?" The fiber

$$h^{-1}(y) = \{\theta : h(\theta) = y\}$$

answers that question at a reported value. A local rank calculation describes nearby possibilities. The fiber describes the full set, including separated states that no local derivative can reveal.

Continuous history supplies a different kind of evidence

For the bar sensor, a continuously tracked observation can be lifted to a pointer angle after one initial end has been chosen. If the pointer makes a half-turn, the observed bar returns to its initial

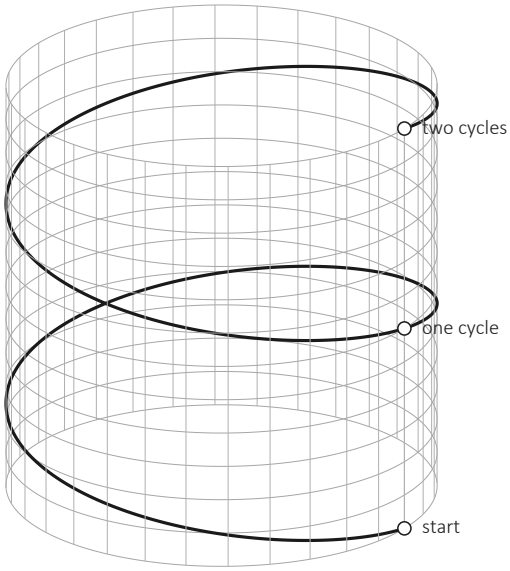
configuration while the lifted pointer reaches the other end. The observation history retains more than the final bar image alone.

With sampled observations, lifting requires a motion bound. If consecutive pointer angles differ by less than $\pi/2$, the nearest bar representative determines the local continuation uniquely except at ties. If the pointer may turn farther between samples, multiple histories fit the same reports. More accurate arithmetic cannot reconstruct a motion that the sampling protocol never distinguished.

This is the double-cover theme in a form one can build: a physical device, an explicitly two-to-one measurement, paired representatives, and a history whose lift depends on an initial choice. It also shows the boundary of the claim. The horizontal-coordinate sensor is not a regular double cover at its extrema; its fibers change size there. A repeated factor of two in a formula does not erase that geometric difference.

A return in the object can retain a history

Height counts completed object cycles; it is a history coordinate, not a physical axis.



Unoriented bar: u and −u

Display: (cos 2α, sin 2α)

After one bar cycle: u → −u

After two: u → u

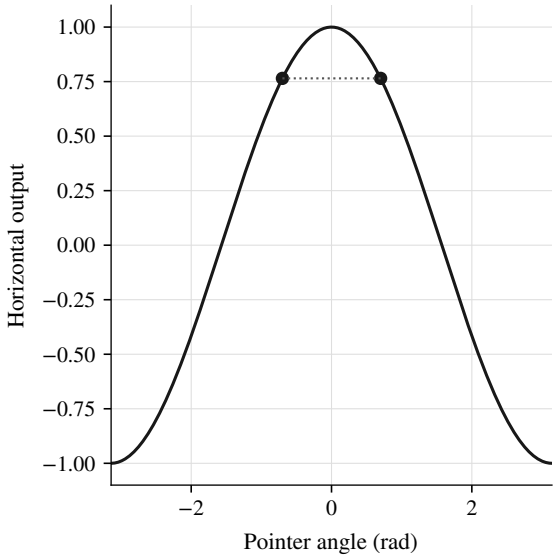


u −u

The physical bar is unchanged.

A Two angles give the same horizontal reading

B A marked tip retains the sign that a bar loses



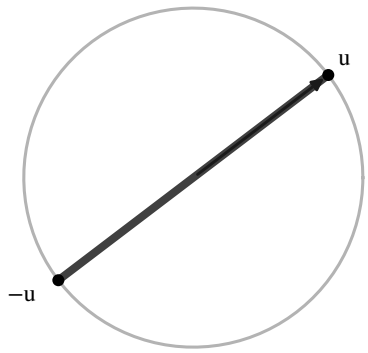
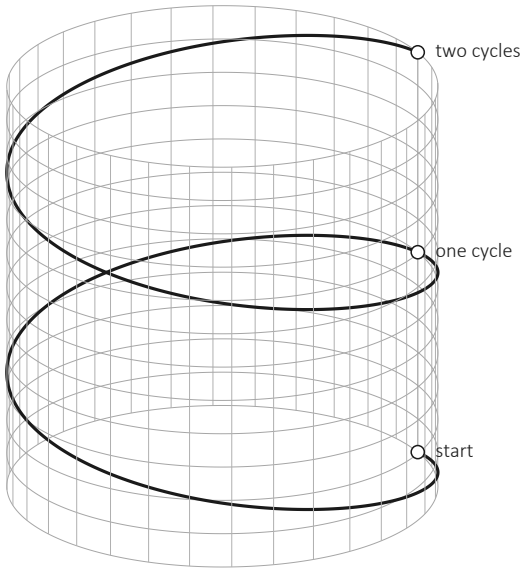


Figure 8.4: *One instrument, three observation geometries.* The study continues on the following figure page.

A return in the object can retain a history

Height counts completed object cycles; it is a history coordinate, not a physical axis.



Unoriented bar: u and −u

Display: (cos 2α, sin 2α)

After one bar cycle: u → −u

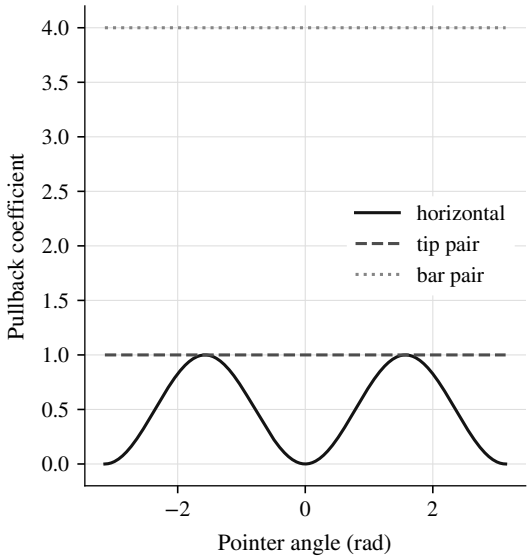
After two: u → u



The physical bar is unchanged.

C The observations have different local sensitivity

D Local rank and global identity are separate



Horizontal: blind at 0 and π

Tip pair: one-to-one on S^1

Bar pair: full rank, two sheets

Rescaling a ruler does not recover a missing end label.

Figure 8.4: *One instrument, three observation geometries (continued).* The horizontal sensor loses derivative rank at its extrema. The marked-tip sensor is globally injective on the circle. The bar observation retains full local rank while identifying two representatives.

Sensor design. Give the pointer a second mark that distinguishes its two ends. Specify an observation that removes the bar ambiguity. Then give an observation that increases the number of recorded coordinates while preserving that ambiguity, and calculate its derivative rank.

We can distinguish a change of coordinates, a loss of local rank, and a global identification with full local rank. Those distinctions must survive when many converters are connected. We now build their path space and inspect what its zero-distance classes identify.

Practice with the construction

Back to the maps and the globe. A map can omit a road. A second map cannot reconstruct it solely from the omission. Before comparing distances, inspect which directions and distinctions the description can represent.

8.1. For $f(x, y) = (2x, 3y)$, find the pullback metric, singular values and the norm of the inverse derivative.

8.2. For $f(t) = (\cos 2t, \sin 2t)$ on a circle, determine the local metric and the number of representatives of each observed point. Explain how this differs from the rank-deficient plate sensor.

Part III

Translators and manifolds

Chapter 9

A translator system becomes a path space

A survey crew wants to measure the distance between two harbor buoys. One buoy appears on the west harbor chart and the other on the east harbor chart. The crew can trace a route through both charts, changing coordinates where they overlap. We will use this route to build a distance for systems that store observations in several representations.

In Chapter 3 we met the translator builder: someone who thinks in terms of feature spaces, encoders, and translators, and who reasons about sameness by moving observations around a graph. In Chapter 5 we developed the language of metric and Riemannian geometry.

This chapter is where those threads meet for the first time. We will show how to treat a translator system as a *metric object*: a single space with a distance function that encodes what the translator graph is doing.

Very roughly, the construction goes like this:

- Start with a collection of feature spaces V_i, each with its own metric d_i.
- Add translators T_{ij} between some of these spaces, with a domain for each conversion.
- Consider paths that consist of continuous movement inside a single V_i, punctuated by discrete jumps along translator edges.
- Define the length of such a path as the sum of the lengths of its within-space segments.
- Take the infimum of path lengths in each direction, then symmetrize the two costs on a mutually reachable component.
- Identify the zero-distance classes of that symmetric comparison.

As the survey boat moves between the harbor buoys, each chart measures part of the distance. Switching charts at a marked location changes the coordinate used to describe it. In the exact atlas model, the switch adds no travel. The charts must use compatible rulers so the crew can add the lengths they measure.

The translator network is built by the same bookkeeping. It records the source of each coordinate, the available conversions, and the cost of movement within each representation. A learned conversion may fail to preserve the intended object, so we must inspect what its zero-cost identifications actually do. We also account for direction before calling the resulting comparison a metric. The construction below makes each of those obligations explicit.

Translator Systems: Spaces, Encoders, Translators

Suppose the crew measures 120 metres from the first buoy to a marker printed on both charts, then 80 metres from the marker to the second buoy. The recorded route is 200 metres. The shortest available route may be cheaper. To find it we need a ruler for each chart and rules for where a route may switch charts.

Each harbor chart supplies coordinates and a ruler. Feature spaces and metrics play those roles in a translator system. To see how the construction can work with unlike records, consider an illustrative identity system that receives still images, video and structured logs.

Feature spaces and encoders

A translator system starts with a finite or countable collection of *channels* indexed by a set I. For each channel $i \in I$, there is:

- A feature space V_i, which we will treat as a metric space (V_i, d_i).
- An encoder E_i that maps raw observations in channel i into V_i.

For now, think of the encoders E_i as learned neural networks or feature-extraction pipelines that map raw observations (images, videos, log records) into vector spaces where distances are meaningful.

- $i = \text{img}$ for still images, with V_{img} a face embedding space and d_{img} a norm-induced metric.
- $i = \text{vid}$ for video clips, with V_{vid} a video embedding space and d_{vid} its metric.
- $i = \log$ for log records, with $V_{\log}$ a structured feature space and $d_{\log}$ an appropriate metric.

The encoders E_i take raw data into the feature spaces V_i. The translator paths begin with those encoded observations.

Translators and the translator graph

Between some pairs of channels (i, j), the system has a translator. It may be an exact coordinate conversion or a learned map defined on an operating domain:

$$T_{ij} : \text{dom}(T_{ij}) \subseteq V_i \longrightarrow V_j.$$

For a learned translator, training aims to preserve semantic identity across the two channels, as in Chapter 3.

The whole system can be summarized by a directed graph G:

- The nodes of G are the channels $i \in I$ (or, more precisely, the spaces V_i).
- There is a directed edge from i to j whenever there is a translator T_{ij} defined on part or all of V_i.

Figure 9.1 shows the three channels of the illustrative identity system. A directed edge records a conversion the system supplies. A conversion in the other direction needs its own map. A route through an intermediate representation is available only if both conversions accept the relevant intermediate inputs. The graph records these possibilities before we assign lengths or identify any points.

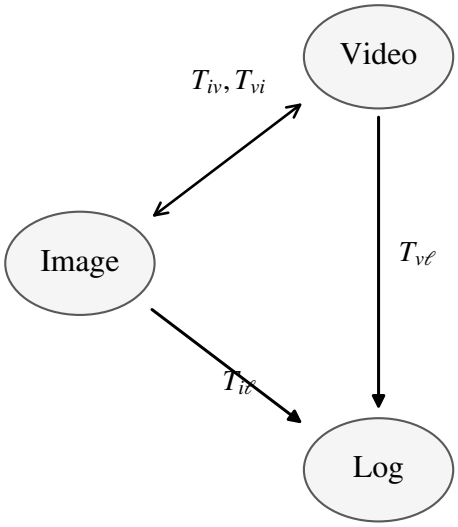
It is sometimes convenient to think of G as a multigraph: there could be multiple translators between the same pair of spaces (for example, learned with different architectures or trained on different regimes). We will ignore that complication for now and treat at most one translator per ordered pair (i, j).

We will refer to a collection $\{(V_i, d_i)\}_{i \in I}$ together with translators $\{T_{ij}\}$ on a graph G as a *translator system*.

Paths in Translator Systems

In Chapter 5, paths and geodesics in a metric space gave us a language for talking about benign evolution: an entity moves smoothly from state x to state y along a path γ, and we measure how much evolution occurred by computing the length of that path. Shortest paths—geodesics—correspond to the most direct, natural transitions.

A *The converters give directed routes*



B *A reachable log need not lead back*

From → to	*Image*	*Video*	*Log*
Image	id	yes	yes
Video	yes	id	yes
Log	—	—	id

A log can be reachable without providing a route back to its source.

Figure 9.1: A directed translator system and its route register. Image and video have translators in both directions; each can reach the log representation. No reverse translator from the log is supplied. Reachability and reversibility are separate properties.

But translator systems are not single metric spaces. They are *collections* of spaces $\{V_i\}$ connected by translators $\{T_{ij}\}$. An observation might start its life encoded in V_{img}, get translated to V_{vid} for comparison with video data, then translated again to V_{log} to check consistency with structured logs. At each stage, we're working in a different coordinate system with its own notion of distance.

The question is: **how do we define a path that moves through this multi-space system?** Intuitively, such a path should allow two kinds of movement:

1. **Within-space movement**: Continuous motion inside a single space V_i, representing benign changes that don't require switching representations. For example, a face aging gradually within the image embedding space, or a transaction amount drifting slightly within a fixed financial schema.
2. **Translator jumps**: Discrete transitions from one space to another via a learned translator T_{ij}. These are the moments when we change coordinate systems—when we convert an image embedding to a video embedding, or translate a log record into a different feature space for comparison.

A translator path follows a coordinate curve within one representation and switches descriptions at a point accepted by the next converter. In the exact atlas case the switch rewrites one location. Its cost is zero because it contributes no additional motion. The within-chart curve lengths still use their own transported rulers.

To measure the length of such a path, we'll sum up the lengths of all the within-space segments, assigning zero length to every translator jump. Exact chart jumps only change coordinates. With

learned converters, a zero assigned cost can hide a change to the point being represented. The shortest such path between two points gives us a notion of distance that respects both the local geometry of each space and the connectivity of the translator graph.

This section makes all of that precise. We'll define what a path in a translator system is, how to measure its length, and how these paths give rise to a distance function on the entire multi-space structure.

The disjoint union of feature spaces

To keep track of which space a vector lives in, it is useful to form the *disjoint union*

$$\mathcal{V} = \bigsqcup_{i \in I} V_i.$$

What this notation means. Let's unpack the symbols:

- I: The channel index set. For a telemetry example, use raw, compact and archival descriptions. Generic image/video examples use $\{\text{img}, \text{vid}, \text{log}\}$ as channel labels, not as a claim about a deployed identity system.
- V_i: The feature space associated with channel i. For example, V_{img} might be $\mathbb{R}^{512}$ (512-dimensional image embeddings), while V_{vid} might be $\mathbb{R}^{256}$ (256-dimensional video embeddings).
- $\bigsqcup_{i \in I} V_i$: The *disjoint union* of all the feature spaces. The square-cup symbol $\sqcup$ means "disjoint union"—we're taking all the spaces V_i and stacking them together, but *keeping track of which space each vector came from.*
- (i, v) **or** $[v]_i$: An element of $\mathcal{V}$. This is a *tagged* vector: the actual vector v plus a label i saying which space it lives in.

Why "disjoint"? A concrete example. Suppose you have a 3-dimensional vector $v = (1.2, 0.5, -0.3)$. This vector might appear in multiple spaces:

- As an image embedding: $[v]_{\text{img}} = (\text{img}, v) \in V_{\text{img}}$
- As a video embedding: $[v]_{\text{vid}} = (\text{vid}, v) \in V_{\text{vid}}$
- As a log embedding: $[v]_{\text{log}} = (\text{log}, v) \in V_{\text{log}}$

Even though the numeric vector $v = (1.2, 0.5, -0.3)$ is the same in all three cases, these are *different points* in $\mathcal{V}$ because they live in different spaces. The tag i prevents us from confusing them.

Think of it like postal addresses. The number "123" could refer to:

- 123 Main Street, Paris
- 123 Elm Avenue, Berlin
- 123 Oak Road, Rome

The street number alone ("123") isn't enough to locate a building—you need the city too. Similarly, a vector v alone isn't enough to locate a point in our multi-space system—you need to know *which feature space* it lives in. The disjoint union $\mathcal{V}$ is like a global addressing system that includes both the street number (v) and the city (i).

Why we need this. Without the disjoint union, we'd have ambiguity: if someone hands you a vector v, which space does it live in? Can you compare it directly to another vector w, or do you need to translate one of them first?

By working in $\mathcal{V}$, we make the space structure explicit. A point $[v]_i$ is clearly in space V_i. To move from $[v]_i$ to another space, you must explicitly use a translator T_{ij}. This bookkeeping will be essential when we define paths and distances: we need to know exactly when we're moving within a space (continuous motion) versus jumping between spaces (translator edges). Elements of $\mathcal{V}$ are pairs (i, v) where $i \in I$ and $v \in V_i$. The disjoint union keeps track of *which* space a vector lives in: the same vector v living in V_i is distinct from v living in V_j if $i \neq j$.

We will write $[v]_i$ for the element (i, v) when we want to emphasize the space.

Allowed moves: within-space segments and translator jumps

In a single metric space, a path is just a continuous curve. In a translator system, we want to allow two kinds of motion that feel natural but do not fit standard continuous paths:

- We might move smoothly within a single feature space V_i (like adjusting a face embedding to account for small lighting changes).
- We might jump from one feature space V_i to another via a translator T_{ij} (like translating an image embedding into a video embedding).

A *translator path* interleaves coordinate curves and converter applications. At a change of chart, start the next segment at the converted endpoint of the preceding segment. When both segments use the same chart, join them at the same coordinate. That condition lets us trace one proposed route without silently substituting an unrelated observation.

We can capture this with the following informal definition.

Definition 9.1 (Translator path (informal))**.** A *translator path* in a translator system is a finite sequence of steps

$$\big((i_0, \gamma_0), (i_1, \gamma_1), \ldots, (i_m, \gamma_m)\big)$$

such that:

- For each k, $i_k \in I$ and $\gamma_k : [t_k, t_{k+1}] \to V_{i_k}$ is a continuous curve in V_{i_k}, with $0 = t_0 < t_1 < \cdots < t_{m+1} = 1$.
- For each $k = 0, \ldots, m-1$, the next curve starts at the previous endpoint expressed in its channel. If $i_{k+1} = i_k$, use $T_{i_k i_{k+1}} = \mathrm{id}_{V_{i_k}}$. Otherwise require a translator $T_{i_k i_{k+1}}$ defined at $\gamma_k(t_{k+1})$. In both cases require

$$\gamma_{k+1}(t_{k+1}) = T_{i_k i_{k+1}}(\gamma_k(t_{k+1})).$$

Unpacking the symbols. This definition packs a lot into compact notation. Let's break it down:

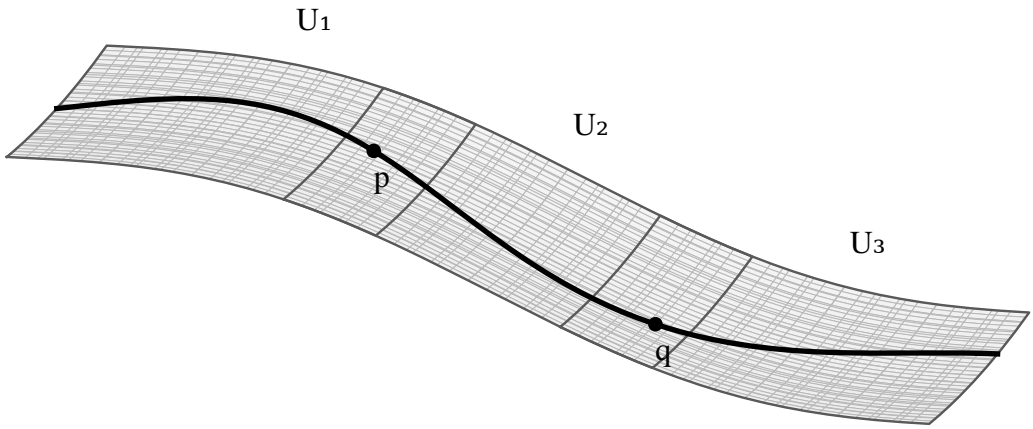
- m: The index of the last segment, so the path has $m+1$ segments. Each segment is a continuous curve in one space, followed (possibly) by a translator jump to another space.
- (i_k, γ_k): The k-th segment of the path. Here $i_k \in I$ is the space we're in during segment k, and γ_k is the curve we trace out in that space.
- $\gamma_k : [t_k, t_{k+1}] \to V_{i_k}$: A continuous curve in space V_{i_k} parametrized by time from t_k to t_{k+1}. Think of this as "how we move through space i_k during the k-th leg of the journey."
- $0 = t_0 < t_1 < \cdots < t_{m+1} = 1$: Time partitions. We chop the unit time interval $[0, 1]$ into $m+1$ pieces, one for each segment. Segment k runs from time t_k to time t_{k+1}.
- **The join at i_k and i_{k+1}**: At the end of segment k (at time t_{k+1}), one of two things happens:

  - We stay in the same space ($i_{k+1} = i_k$) and keep moving continuously (like a sharp turn on the same road)
  - We jump to a different space via translator $T_{i_k i_{k+1}}$
- $\gamma_{k+1}(t_{k+1}) = T_{i_k i_{k+1}}(\gamma_k(t_{k+1}))$: The matching condition. If we jump at time t_{k+1}, the starting point of the next segment γ_{k+1} must equal the translator applied to the ending point of the current segment γ_k. This ensures the path is continuous across the jump: we don't teleport to a random location in the new space; we land exactly where the translator sends us.

The navigator changes charts. Follow a boat through a harbor shown on overlapping survey sheets. While the boat moves within one chart, its coordinate curve has a measured length. At a point covered by both sheets, the navigator rewrites the same location in the second chart. That change of address contributes no physical travel. The next curve begins at the translated coordinate of the same point.

The matching condition states this precisely. It does not describe a train transporting the boat to a different harbor for free. With learned translators, the route is an operational comparison path whose directed costs and residuals must be retained. The next section shows how to combine the two directions into a single comparison.

A *One surface route crosses three charts*



B *Changing charts adds no physical travel*

Piece	*Chart address*	*Contribution*
Within U_1	Curve γ_1	Measured length L_1
At $p \in U_1 \cap U_2$	$\varphi_1(p) \to \varphi_2(p)$	Zero travel
Within U_2	Curve γ_2	Measured length L_2
At $q \in U_2 \cap U_3$	$\varphi_2(q) \to \varphi_3(q)$	Zero travel
Within U_3	Curve γ_3	Measured length L_3

Figure 9.2: A computed curved strip with overlapping chart neighborhoods. The route has three continuous pieces. At each marked overlap point, a chart transition rewrites the same location and adds no physical travel; the total length is the sum of the three curve lengths.

Length of a translator path

We want the length of a translator path to reflect how much benign change happens inside spaces, not how many translators we use. The simplest choice (and the one we will use) is:

Definition 9.2 (Length of a translator path)**.** The *length* of a translator path $\Pi = ((i_k, \gamma_k))_{k=0}^{m}$ is

$$L(\Pi) = \sum_{k=0}^{m} L_{i_k}(\gamma_k),$$

where $L_{i_k}(\gamma_k)$ is the metric-space length of the curve γ_k in (V_{i_k}, d_{i_k}) as defined in Chapter 5.

Translator jumps themselves carry no direct length cost in this definition. They matter indirectly, because they allow us to connect regions of different spaces and thus change which within-space distances we are allowed to use.

Inspect the cost of the switch. The navigator can describe one harbor mark on either of two sheets. Changing sheets adds no traveled length. A learned converter has also been assigned zero jump cost in this construction, so its output must be inspected carefully. A biased converter can create a free displacement in the candidate comparison space. The loop example below will calculate that failure explicitly.

Assemble the endpoint calculation

The navigator's route now supplies four pieces of computational data. First retain the tagged curve segments:

$$((i_0, \gamma_0), \ldots, (i_m, \gamma_m))$$

At each change of chart, join the next segment at the converted endpoint:

$$\gamma_{k+1}(t_{k+1}) = T_{i_k i_{k+1}}(\gamma_k(t_{k+1})).$$

Then add the within-chart lengths. For piecewise smooth segments in Riemannian charts, the norm in the integral is the local Riemannian norm. For a general metric space use the partition-based length in the definition above.

$$L(\Gamma) = \sum_{k=0}^{m} L(\gamma_k) = \sum_{k=0}^{m} \int_{t_k}^{t_{k+1}} \|\dot{\gamma}_k(t)\|\, dt$$

Finally compare the cost of every allowed route between the fixed endpoints:

$$d([v]_i, [w]_j) = \inf_{\Gamma} L(\Gamma),$$

This infimum is initially directed. We still need to compare the reverse route and determine which pairs have zero cost both ways before identifying descriptions. The next step proves that the resulting comparison has the required metric properties.

We have assigned a cost to each permitted journey. We still need one comparison between endpoints, independent of which route happened to be written down. Taking the cheapest available route supplies that comparison; checking both directions and identifying zero-distance copies determines the objects it can distinguish.

The Translator Metric Quotient

We now have enough structure to define a distance between points in $\mathcal{V}$: take the infimum of the lengths of all translator paths connecting them. We first obtain a directed cost, then symmetrize it before taking the zero-distance quotient.

The endpoint comparison has to inspect all routes

Two stored descriptions may represent one location. The candidate distance must therefore permit a zero comparison across their exact conversion, while retaining positive separation between locations the task needs to distinguish. The infimum examines every allowed route. Its zero classes are a result to inspect against the observations.

Pre-distance from path lengths

Fix a translator system. Given two points $x = [v]_i$ and $y = [w]_j$ in $\mathcal{V}$, consider all translator paths Π that start at x and end at y. Define the *pre-distance*

$$D(x, y) = \inf_{\Pi : x \to y} L(\Pi),$$

with the convention that if there is no translator path from x to y (for example, if the graph is disconnected), we set $D(x, y) = +\infty$.

Unpacking the symbols.

- $x = [v]_i$ **and** $y = [w]_j$: Two tagged points in the disjoint union $\mathcal{V}$. Point x is vector v living in space V_i; point y is vector w living in space V_j. These might be in the same space ($i = j$) or different spaces ($i \neq j$).
- $\Pi : x \to y$: A translator path starting at x and ending at y. For the harbor comparison, trace a curve from the first buoy to an overlap point, convert its coordinates, then trace a curve to the second buoy.
- $L(\Pi)$: The length of path Π—the sum of all within-space segments' lengths, as defined above.
- $\inf_{\Pi : x \to y}$: The infimum (greatest lower bound) over all possible translator paths from x to y. Think: "consider every possible route from x to y through the translator graph, and find the shortest one."
- $D(x, y)$: The directed cost is the infimum of lengths of allowed paths from x to y. Reversing the endpoints can change this value. The symmetric pseudometric $\tilde{d}$ is defined next.
- $+\infty$ **convention**: If there's no path from x to y at all (the translator graph is disconnected, so you can't get from space V_i to space V_j using available translators), we say the distance is infinite. This is like saying "Paris and an island with no ferry service are infinitely far apart"—not literally infinite, but unreachable within the system.

From directed cost to a metric

On a component with finite costs in both directions, combine the two directed costs into a symmetric pseudometric

$$\tilde{d}(x, y) = \max\{D(x, y), D(y, x)\}.$$

If all paths reverse with equal length, this is simply D. Otherwise the maximum is an explicit part of our chosen construction. Proposition 9.3 proves its triangle inequality and the metric property of its zero-distance quotient. Small round-trip error alone is not used as an exact symmetry axiom.

Disconnected or one-way-reachable descriptions must be handled on appropriate components or as extended directed costs. A finite metric comparison is made only where its hypotheses hold. Within such a component let X_T consist of the classes $[x]$ under $\tilde{d}(x, y) = 0$. Set $d_T([x], [y]) = \tilde{d}(x, y)$. Proposition 9.3 proves that this value is independent of the representatives and separates distinct classes. The pair (X_T, d_T) is the *translator metric quotient.*

What the quotient identifies. For exact reversible chart transitions, two tags representing one marked point have zero cost in both directions. Identifying them removes the redundant address while retaining the point. For general converters, we check both directed infima before making that identification. The resulting classes describe the distinctions this cost system retains. They must still be checked against the physical observations.

What lives in the translator metric space?

The points of X_T are equivalence classes of feature vectors across all the spaces. Two vectors $v \in V_i$ and $w \in V_j$ represent the same point in X_T when both directed path infima vanish. For every positive allowance, a sufficiently cheap route must exist in each direction. The zero class therefore depends on the full permitted path family.

In the idealized case where translators are perfect and invertible, and encoders are faithful, we can think of each equivalence class as corresponding to a single latent state u in the semantic sameness structure S. In reality, the situation is noisier:

- Equivalence classes in X_T may correspond to small clusters of nearby latent states rather than single points.
- Distances d_T between classes incorporate both meaningful semantic differences and distortions introduced by the translators.

Running example: one identity, many observations. In the illustrative identity comparison, consider two different observations of the same person:

- A still photo taken at age 28 under studio lighting, encoded as $v_{\text{img}} \in V_{\text{img}}$.
- A video clip from age 29 in natural outdoor lighting, encoded as $w_{\text{vid}} \in V_{\text{vid}}$.

These live in different feature spaces, but we can connect them via a translator path:

1. Start at $[v_{\text{img}}]_{\text{img}}$ in the image space.
2. Apply $T_{\text{img}\to\text{vid}}$ to jump to video space.
3. Make small adjustments within V_{vid} to account for the one-year age difference and lighting change, following a benign path.

If the encoders and converter preserve identity across the age and lighting changes in this example, that claim should show up as a short measured path. Once tested on relevant observations, a small positive distance can support a tolerance-based comparison, but it does not place the endpoints in the same zero-distance equivalence class. Identifying a persistent person is an additional task relation, with its own evidence.

By contrast, if w'_{vid} encodes a video of a *different* person, a distance-based system that separates these observations must rule out short translator paths connecting v_{img} and w'_{vid}, so they land in different equivalence classes far apart in d_T.

The translator metric space (X_T, d_T) gives us a disciplined way to treat a translator system as a single space in which such distances are measured.

The quotient now has a mathematical distance. The remaining question is whether its identified points still describe the intended observation. We test that attachment by converting a record out and back, and by comparing two conversion routes. A small residual measures one way the descriptions fail to agree.

Drift, Round-Trip Error, and Cocycle Slack

Imagine you're traveling internationally with a bag of cash that you need to convert between currencies at airport exchange booths. You start with $1000 USD. At the Paris airport, you convert to euros; in Berlin, euros to pounds; in Rome, pounds back to euros; finally in New York, euros back to dollars.

With consistent exchange rates and no transaction fees, you would end up with exactly $1000 when you return. But in reality:

- **Each conversion loses value.** Exchange booths charge fees and use unfavorable rates. Converting $1000 to euros might give you €920 instead of the market rate €950. Converting back to dollars gives you $890, not $1000. You've lost $110 in the round trip, even though you're back to the same currency.
- **Different paths lose different amounts.** One sequence of booths might return $890 from an initial $1000, while another returns $905. The endpoint discrepancy is $15. For a direct conversion and a two-step conversion to the same target currency, this is the kind of discrepancy measured by cocycle slack. For longer routes, it is a path-consistency residual. Exact compatible coordinate conversions agree at their endpoints; fee-charging exchanges need not. Being a well-defined mathematical function does not itself make a conversion path-independent.
- **Exchange rates drift over time.** The booth's rates today are different from yesterday's. If you convert USD→EUR on Monday and EUR→USD on Tuesday, your round-trip error depends not just on fees but on how much the exchange rate moved. The *drift profile* tracks how much the conversion relationship changes as economic conditions shift.

Learned or lossy record converters can produce the same three effects. For instance, converting a photograph embedding to the video channel and back may change its value. Different routes for the same transaction record can produce different representations.

To build a rigorous theory of sameness across representations, we need to measure and control these imperfections. This section introduces three tools for doing so:

- **Round-trip error** $\varepsilon_{\mathrm{rt}}$: How much you lose going from space V_i to space V_j and back. Like measuring the $110 loss when converting USD→EUR→USD.
- **Cocycle residual** Cyc: How far apart a direct conversion and a two-step conversion land. Compare the pounds from a direct USD-to-GBP exchange with the pounds from changing dollars to euros first.
- **Drift profiles**: How translators change over time as data distributions shift, models are retrained, or system configurations evolve. Like tracking how exchange rates move as markets change.

These three quantities will be crucial in Chapter 10 when we ask: can this patchwork of translators be interpreted as a smooth manifold? These diagnostics are useful inputs to that question. The geometry also requires its own construction.

These residuals test agreement between routes. A geometric realization also needs domains, smooth transitions and compatible local rulers.

Let's make these notions precise.

Round-trip error

Given a translator $T_{ij} : V_i \to V_j$ and (ideally) a corresponding translator $T_{ji} : V_j \to V_i$, we can look at how close the composition $T_{ji} \circ T_{ij}$ is to the identity on typical points in V_i.

For a point $v \in V_i$ in the regime of interest, define the *round-trip error*

$$\mathrm{RT}_{i \to j \to i}(v) = d_i(T_{ji}(T_{ij}(v)),\, v).$$

In the illustrative identity comparison, we might:

- Encode a still photo into $v \in V_{\text{img}}$.
- Map it to V_{vid} using $T_{\text{img}\to\text{vid}}$.
- Map it back to V_{img} using $T_{\text{vid}\to\text{img}}$.
- Measure the distance between the original and the round-trip embedding.

A small $\mathrm{RT}(v)$ says the returned embedding is close to the input in the source ruler. A map can stretch distances by a factor of 100 and then undo that stretch perfectly. To compare the rulers we need distances between pairs in both spaces.

Cocycle slack

Round trips are about going out and back along the same edge pair. We also care about consistency when there are multiple ways to go from one space to another.

Suppose we have spaces V_i, V_j, V_k and translators $T_{ij} : V_i \to V_j$, $T_{jk} : V_j \to V_k$, and $T_{ik} : V_i \to V_k$. We can compare:

- Going directly from V_i to V_k using T_{ik}.
- Going indirectly via V_j using $T_{jk} \circ T_{ij}$.

For $v \in V_i$, define the *pointwise cocycle residual*

$$\mathrm{Cyc}_{i \to j \to k}(v) = d_k(T_{jk}(T_{ij}(v)),\, T_{ik}(v)).$$

A small residual means the tested routes nearly agree. The uniform allowance ε_{cyc} bounds this discrepancy at every admissible point on the overlap. An exact chart atlas has zero cocycle residual.

Drift profiles along paths

More generally, we can define error profiles along longer translator paths. Given a path $i_0 \to i_1 \to \cdots \to i_m$ in the translator graph and a starting vector $v_0 \in V_{i_0}$, we can propagate v_0 along the path:

$$v_{k+1} = T_{i_k i_{k+1}}(v_k),$$

and then compare v_m to what we would have gotten via some reference path or direct mapping.

Rather than introducing heavy notation, we will summarize this informally:

- For each family of paths of interest, we can empirically estimate how much drift accumulates as a function of path length and structure.

- These *drift profiles* tell us how trustworthy the translator metric d_T is at different scales and along different routes.

In later chapters, these local error terms (round-trip error, cocycle slack, drift profiles) will show up in two places:

- In *smooth chartability* conditions that ensure (X_T, d_T) behaves like a Riemannian manifold at the scales we care about.
- In *detection error budgets*, where they contribute to the slack between translator-side and geometry-side guarantees.

For now, it is enough to remember that the translator metric quotient is defined purely from within-space metrics and translator connectivity, but that the quality of this metric as a model of semantic sameness depends on how small these error terms are.

Use the distance on an observed pair

The quotient construction has produced a comparison space. Suppose an archive receives a still photo and a video clip and must decide whether they show the same person. Each encoder and source tag locates its observation in the quotient. A threshold turns the distance between those points into a decision. The task must establish that the retained distinctions are the distinctions it needs.

We will trace that use on an observation pair, then test the ways the connection can fail. A faithful encoder can still be paired with a wrong cross-channel correspondence. A locally consistent converter can still create a global shortcut. The mathematical cost and its application therefore need their own checks.

From feature vectors to points in X_T

Given a raw observation in some channel i, we can:

1. Encode it into a feature vector $v \in V_i$ via E_i.
2. Embed it into the disjoint union as $[v]_i \in \mathcal{V}$.
3. Map it to a point $q([v]_i) \in X_T$ via the quotient map.

If two observations in possibly different channels correspond to the same underlying latent entity u, and if the translator system is well-behaved, we expect their images in X_T to lie close together in d_T. If they correspond to different entities, we expect them to be farther apart.

Sameness as a distance test

A simple translator-metric-based sameness test between two observations might be:

1. Map each observation into X_T as above, obtaining points $x, y \in X_T$.
2. Compute or approximate $d_T(x, y)$.
3. Compare $d_T(x, y)$ against one or more thresholds:
   - If $d_T(x, y) \leq \tau_{\text{same}}$, declare them "same".
   - If $d_T(x, y) \geq \tau_{\text{diff}}$, declare them "different".
   - If $\tau_{\text{same}} < d_T(x, y) < \tau_{\text{diff}}$, abstain or send to a higher-fidelity process.

This three-zone structure—same, different, ambiguous—is common in high-stakes applications where false positives and false negatives have different costs, and where we can route uncertain cases to human review or higher-confidence processes.

This is conceptually similar to what we might do in a geometry-based system, except that:

- In the translator case, d_T is induced by the graph of feature spaces and translators, not by a Riemannian metric on a latent manifold.
- The topology and geometry of (X_T, d_T) may be complicated, reflecting both the semantic structure and the idiosyncrasies of the translators.

Apply the same construction to a heterogeneous model. Blood measurements, imaging, symptoms and genetic records illustrate why a common comparison can be difficult. Their observation spaces retain different information. A hypothetical translator model supplies maps between their encoded descriptions, local rulers, allowed paths and a target comparison. The quotient then calculates the distance specified by those inputs.

For such a model to support a decision, matched observations must establish the intended correspondence, and the retained distinctions must serve the target. The same issue arises if a harbor converter sends the coordinates of one buoy to those of another. A mathematically valid route cost can be the wrong cost for the observation. We now inspect that failure before treating the quotient as a realization of S.

When does the translator metric reflect semantic sameness?

We have built the translator metric quotient (X_T, d_T) on a component where the symmetric path cost is finite. The feature spaces, within-space lengths and translator edges specify the construction. Its distance obeys the metric axioms after zero-distance classes are identified.

But having a metric space is not the same as having a *semantically meaningful* metric space. Recall from Chapter 1 that we started with a semantic sameness structure S: an abstract specification that says which observations represent the same underlying entity u evolving benignly. The translator metric quotient is supposed to be a *realization* of S—a concrete system that answers sameness questions in a way that respects S's semantics.

The question is: **when does (X_T, d_T) actually do this? When does distance in translator space correspond to semantic difference?**

The answer is not automatic. You can build a translator system where:

- Encoders drop critical information, collapsing semantically distinct entities into the same feature vector
- Translators introduce large round-trip residuals in a feature space; because translator jumps cost zero, these cycles can instead make distinct feature values indistinguishable in the quotient
- Different paths through the translator graph give wildly different distance estimates for the same pair of points
- Translators behave erratically on out-of-regime data, causing distances to blow up or collapse unpredictably

These failures can make the metric unsuitable for the semantic task without violating its formal metric axioms. Their effect on a particular guarantee depends on its assumptions, domain and margin.

To ensure (X_T, d_T) is a good realization of S, we need *regularity conditions*. Here are the key ones, stated informally. Chapter 10 will make them precise.

Condition 1: Faithful encoders. Each encoder E_i takes raw observations into V_i and must preserve the distinctions required by the task. If two observations o_1 and o_2 represent different latent entities (different people, different viral strains, different transactions), their encodings $E_i(o_1)$ and $E_i(o_2)$ should be far apart in V_i.

What goes wrong if this fails:

If the encoder collapses semantically distinct states, the translator metric will declare them "same" even though they're not. For example:

- A face encoder trained only on frontal lighting might map all poorly-lit faces to nearly the same embedding, regardless of identity. The translator metric d_T would see them as close, even though they represent different people.
- A transaction encoder that ignores merchant category might collapse "$50 at grocery store" and "$50 at casino" into the same feature vector, causing a fraud detector to miss suspicious patterns.

Faithful encoders ensure that semantic distance ("how different are the underlying entities?") is reflected in embedding distance.

Condition 2: Approximately invertible translators with bounded round-trip residual. Translators $T_{ij} : V_i \to V_j$ should be approximately reversible on in-regime data. Specifically, if $T_{ji} : V_j \to V_i$ is the reverse translator, then for typical in-regime vectors $v \in V_i$, we should have:

$$d_i(v, T_{ji}(T_{ij}(v))) \approx 0.$$

In words: translating forward and back should return you close to where you started.

What goes wrong if this fails:

Large round-trip errors mean the translator graph is internally inconsistent. A point $[v]_i$ in space V_i might land at $[T_{ij}(v)]_j$ in space V_j, but when you translate back via T_{ji}, you arrive at a *different point* $[T_{ji}(T_{ij}(v))]_i$ far from where you started. This creates ambiguity: which representation is "correct"?

In the currency exchange metaphor: if you convert $1000 to euros and back and only get $700, the exchange system is losing massive value. In record systems, a large round-trip error says these two converters fail to reconstruct the input under the stated ruler. A change can occur even when each converter is invertible, as in the biased-converter example later in this chapter.

Small round-trip error shows that a conversion and its chosen reverse recover nearby feature values. Semantic agreement is a separate check: matched observations of an entity u must satisfy $T_{ij}(E_i(u)) \approx E_j(u)$, with an allowance small enough for the task's decision margin.

Imagine two cameras looking at numbered lockers. One encoder assigns the signed position $E_A(u) = u$; the other uses the reversed coordinate $E_B(u) = -u$. If both translators are the identity map, every round trip and every translator cycle is perfect. Yet locker u in camera A is compared with locker $-u$ in camera B. Each encoder retained the positions; the correspondence between them is wrong. Known matched landmarks or labeled pairs reveal what internal consistency cannot.

Condition 3: Small cocycle slack on overlaps. When multiple translators connect the same spaces, or when we can reach the same destination via different paths through the graph, those paths should *agree*.

Formally: if we have translators T_{ij}, T_{jk}, and T_{ik}, then for typical $v \in V_i$, the composed path $T_{jk}(T_{ij}(v))$ should land close to the direct path $T_{ik}(v)$:

$$d_k(T_{jk}(T_{ij}(v)), T_{ik}(v)) \approx 0.$$

This is the *cocycle condition*: different routes through the graph should commute approximately.

What goes wrong if this fails:

If cocycle slack is large, different paths give different distance estimates. Going from V_i to V_k directly might suggest two points are close, while going via intermediate space V_j might suggest they're far apart. The translator graph is giving contradictory information.

Compare a printed road map, the car's navigation screen and a survey chart. Translate the coordinates of the same marked junction directly from the road map to the survey chart, then translate them through the navigation screen. Both conversions should identify that junction. Driving to it by different physical roads can legitimately take different distances; the consistency check concerns the same landmark represented through different coordinate systems.

Small cocycle slack measures agreement of the tested coordinate conversions. An exact smooth atlas additionally needs genuine charts, compatible smooth transitions and the topological conditions that permit gluing. Agreement of endpoint values does not, by itself, supply those properties or a shared metric.

Condition 4: Controlled drift profiles. Translators should not behave erratically as we move through the feature spaces. Specifically:

- The Jacobian $J_{T_{ij}}(v)$ should have upper and positive lower singular-value bounds in the chosen tangent metrics. The condition number measures their ratio; it alone cannot detect uniform expansion or collapse of scale
- As we move along paths in V_i, the translator T_{ij} should vary smoothly: small changes in v should produce small, predictable changes in $T_{ij}(v)$

This is the *drift profile*: how much does translator behavior change as we move through the space?

What goes wrong if this fails:

If translators stretch distances wildly in some regions (large Jacobian singular values) or collapse them to zero in others (small singular values), then d_T will be distorted: some semantically small changes will appear huge, others will be invisible.

If translators vary erratically—behaving one way at point v and completely differently at nearby point v'—then paths through the translator graph will accumulate unpredictable errors. The system becomes fragile: tiny perturbations in input cause large swings in distance estimates.

Smoothness together with a nonsingular derivative in equal dimensions gives a local smooth coordinate change. Derivative bounds control local length distortion; global one-to-one behavior and admissible paths still require their own checks.

Putting the checks together: The bridge to geometry. The construction has three distinct obligations. First, the encoders and their cross-channel correspondences must preserve the distinctions required by S. Second, the translator paths must induce a useful metric: unrestricted zero-cost cycles must not collapse distinctions the task needs. Third, a Riemannian realization must be constructed or supplied with a map and quantitative comparison bounds.

For an exact atlas, injective charts from open subsets of a common-dimensional Euclidean space, exact smooth overlap maps, the Hausdorff and second-countability conditions, and compatible

positive-definite tensors provide the geometric construction. Chapter 10 works through these requirements. Learned translators with small residuals instead supply evidence for a proposed approximate realization. The map, domain and distortion bound are additional mathematical objects to establish.

Once those comparisons hold, Chapter 20 relates decisions through a distance allowance, a probability of leaving the good event, and the population mass near the threshold. Internal consistency, semantic alignment, geometric realization and decision transfer are connected stages; each has an inspectable obligation.

The path from construction to correctness. The directed path cost can always be defined, allowing infinite cost for unreachable pairs. Its symmetric finite quotient is taken on a component where the requisite paths exist. This gives a mathematical metric. To use that metric as a ruler for individuals, we then compare matched entities across channels, measure round-trip and cycle residuals, inspect both extremes of the Jacobian spectrum, and test for unintended short routes.

An audit reports the exact region, sampling rule, units and decision margin. A small residual on sampled pairs is useful evidence about those pairs. Extending it to a whole region needs a proof or a coverage argument; extending it to future observations needs a justified population model. Chapter 21 develops those procedures. The audit ties the comparison back to a named individual and bounds how much a change of representation can affect it.

A change of chart is not a journey to another city

A navigator can switch between two charts while the boat stays anchored. That is the physical picture for a zero-cost transition: the description changes at one underlying point. Moving the boat from one port to another is a different operation and has a length. Keeping this distinction visible prevents a translator from silently making travel free.

For general learned translators, begin with a directed cost. Let $D(x, y)$ be the infimum of nonnegative lengths of permitted paths from the tagged description x to the tagged description y. Assume constant paths are allowed and finite paths can be concatenated. Then $D(x, x) = 0$ and

$$D(x, z) \leq D(x, y) + D(y, z).$$

The proof uses paths within any positive tolerance of the two infima, concatenates them, and then lets the tolerance tend to zero. It does not require a shortest path to be attained. Reversing a path, however, requires its own permission.

Proposition 9.3 (A symmetric translator quotient)**.** *On a component where both directions have finite cost, define*

$$d_{\mathrm{sym}}(x, y) = \max\{D(x, y), D(y, x)\}.$$

This is a pseudometric. The relation $x \sim y$ iff $d_{\mathrm{sym}}(x, y) = 0$ is an equivalence relation, and $d_T([x], [y]) = d_{\mathrm{sym}}(x, y)$ is a metric on its quotient.

Proof. Nonnegativity, symmetry and zero diagonal are immediate. Each directed triangle inequality is bounded by $d_{\mathrm{sym}}(x, y) + d_{\mathrm{sym}}(y, z)$, so their maximum is too. Transitivity of the zero-distance relation follows. If $x \sim x'$ and $y \sim y'$, two applications of the triangle inequality give $d_{\mathrm{sym}}(x, y) = d_{\mathrm{sym}}(x', y')$. Thus the quotient distance is well-defined and separates distinct classes. □

When every permitted path has a reverse of the same length, D is already symmetric and this construction returns D itself. Exact atlas transitions with transported local metrics provide that

case. Approximate inverse translators require the directed construction or an explicitly chosen symmetric alternative; their small round-trip residual alone is not exact reversibility.

A small loop can change a global infimum

Take two copies of the real line, A and B. Let $T_{AB}(x) = x + \varepsilon$ and $T_{BA}(x) = x$, with positive ε. Both maps preserve within-line differences. Traversing the two jumps shifts the coordinate in A by ε at zero jump cost. Repeated loops move forward for free. Moving backward by ε requires paid within-line travel, so $D_A(0, \varepsilon) = 0$ and $D_A(\varepsilon, 0) = \varepsilon$.

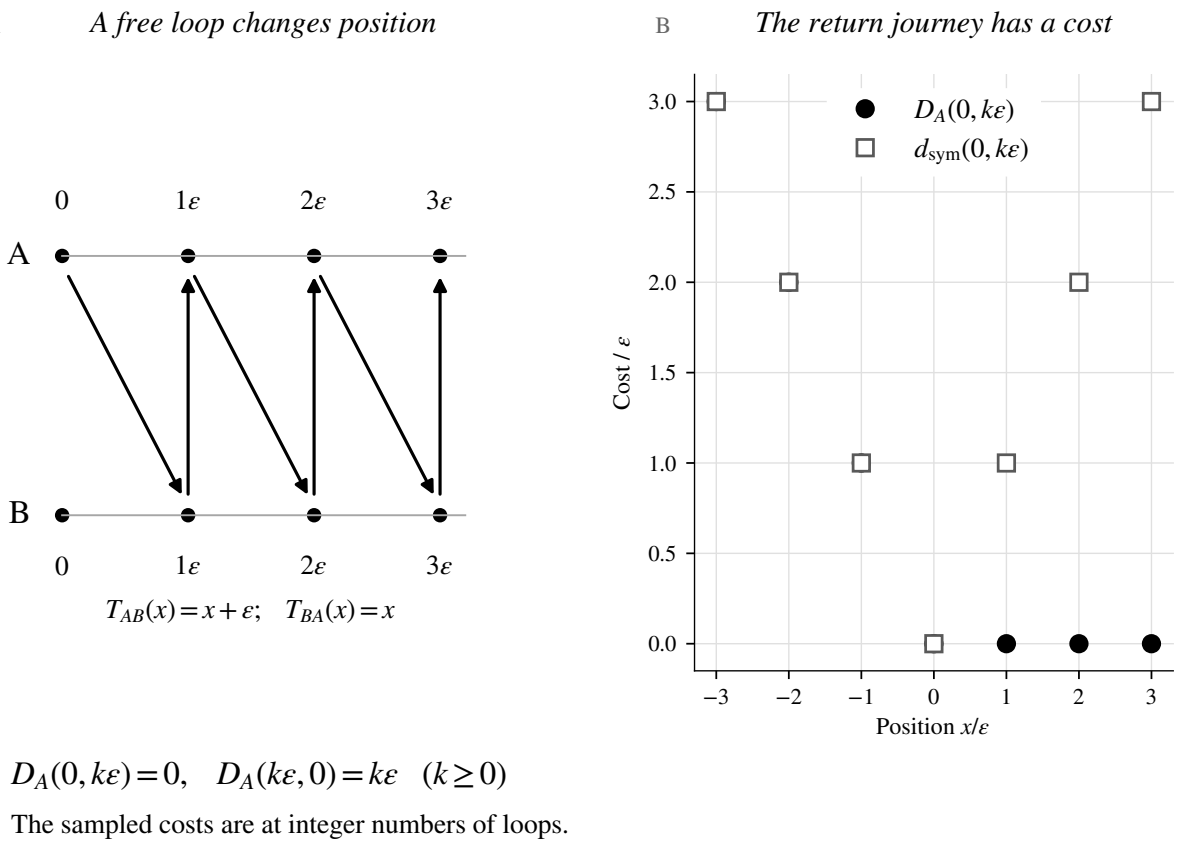


Figure 9.3: Two locally distance-preserving translators accumulate a zero-cost drift when their loop is repeated. The cost plot samples integer numbers of loops; the reverse direction still requires paid motion. Symmetrization retains that cost.

Think of lines A and B as two map scales. I put a pin on A, convert to B, and convert back. The software walks my pin a little to the right and says the trip was free. Fine, ask it to go left for free, too. The return cost is what keeps the pins apart when I count both directions. If I cap the number of conversions or the uncertainty they can build up, two allowed routes may exceed that cap when joined. Then I have to check the triangle inequality again.

The quotient identifies exact zero-distance pairs, not every pair with a small positive distance. A tolerance threshold belongs to a later decision stage. Nor is the metric quotient necessarily a smooth manifold. A tree, a branching graph and a surface can all carry metrics while having different local structures. The next chapters specify the additional atlas and realization hypotheses.

To inspect the infimum operation numerically, use a finite graph with declared nonnegative costs on its converter edges. This is an explicit extension of the zero-jump-cost atlas model, useful

when an operational comparison assigns such costs. It lets us calculate every route, then test direction and zero-cost identification before returning to the continuous path approximation.

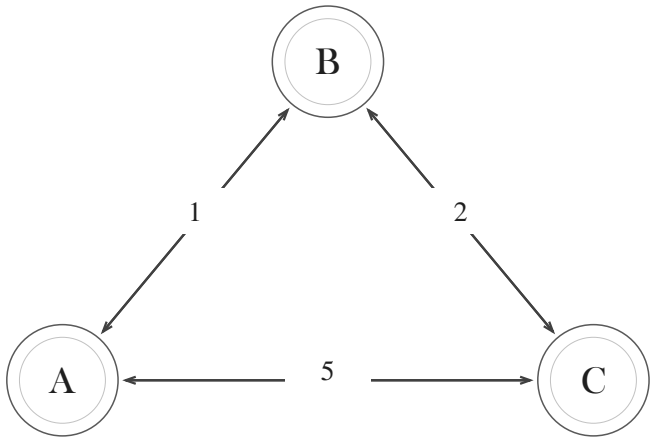
Three converters and a cheapest route

Suppose three record formats are connected by converters. Going directly from A to C has declared cost five. Going from A to B costs one, and from B to C costs two. The cheapest available route from A to C costs three. The edge label is not the path distance.

If every converter is reversible at equal cost, all costs are nonnegative and distinct records cannot be connected by arbitrarily cheap routes unless they are deliberately identified, the shortest-path construction gives a metric on the resulting quotient. Each assumption does work. A zero-cost chain may identify distinct stored records. A one-way converter can make the cost asymmetric. A negative cycle makes repeated circulation cheaper without bound.

Reversible links still require a route

The edge labels specify the calculation; the placement of nodes is schematic.



Direct route A → C: 5

Via B: 1 + 2 = 3

Cheapest route: 3

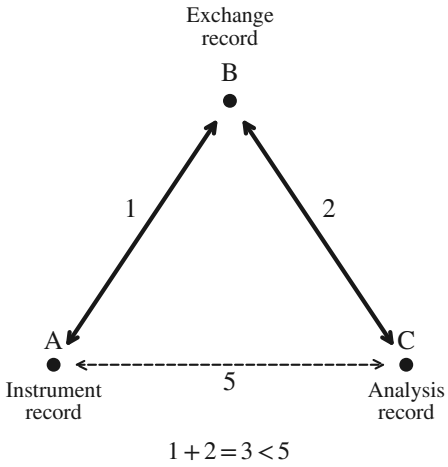
Same cost in reverse.

The path changes the answer.

Direct converter cost and shortest-path distance are different entries in the audit.

The numerical register below specifies the exact graph and its assumptions.

A Two conversions cost less than one



B Direct costs become cheapest-route distances

C: direct links

	A	B	C
A	0	1	5
B	1	0	2
C	5	2	0

D: best routes

	A	B	C
A	0	1	3
B	1	0	2
C	3	2	0

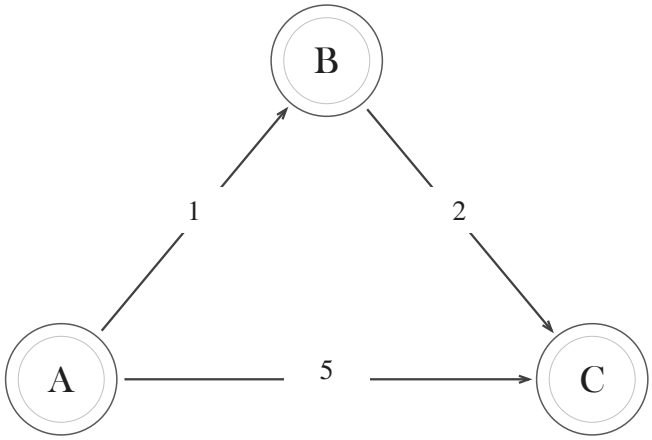
$$D_{AC} = \min(C_{AC},\ C_{AB} + C_{BC})$$

Only the route cost changes; the endpoints do not.

Figure 9.4: *The cheapest translator may be a composition.* The study continues on the following figure page.

A route is a composition with a cost

The edge labels specify the calculation; the placement of nodes is schematic.



Direct route A → C: 5

Via B: 1 + 2 = 3

Cheapest route: 3

Return costs require their own edges.

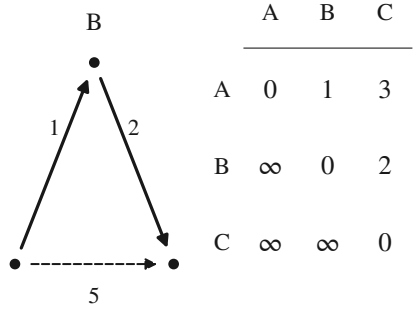
The path changes the answer.

Direct converter cost and shortest-path distance are different entries in the audit.

The numerical register below specifies the exact graph and its assumptions.

C A one-way route may have no return

D Zero distance merges two records into one class



Destination

	A	B	C
A	0	1	3
B	∞	0	2
C	∞	∞	0

$d(A, C) = 3$ but $d(C, A) = \infty$

Separate records

	A	A′	B	C
A	0	0	1	3
A′	0	0	1	3
B	1	1	0	2
C	3	3	2	0

→

Distance classes

	[A]	B	C
[A]	0	1	3
B	1	0	2
C	3	2	0

$[A] = \{A, A'\}, \quad \bar{d}([A], C) = 3$

Both A rows agree; distance cannot distinguish them.

Figure 9.4: *The cheapest translator may be a composition (continued).* The direct-cost and shortest-path matrices distinguish an edge cost of five from a composed cost of three. Removing return links makes some directed distances infinite. A separate zero-cost extension identifies A with A' and gives a metric on the displayed classes. Costs are declared model inputs, not measured execution times.

The zero-cost quotient must be checked

For a symmetric nonnegative path pseudometric d, define $x \sim y$ when $d(x, y) = 0$. The triangle inequality makes this relation transitive. The quotient distance $\bar{d}([x], [y]) = d(x, y)$ is well defined because

$$|d(x, y) - d(x', y')| \le d(x, x') + d(y, y') = 0$$

whenever $x \sim x'$ and $y \sim y'$. It is a metric on equivalence classes.

Suppose a converter drops the timestamp and we give that change zero cost. Then two readings with the same value but taken an hour apart can land in the same class. That class works by the rules we gave it. If the time matters, the rule is what we need to fix.

Path length has a discretization error

A sampled curved route can be approximated by a polygonal chain. On a circle of radius R, an arc of angle Θ split into n equal chords has total chord length

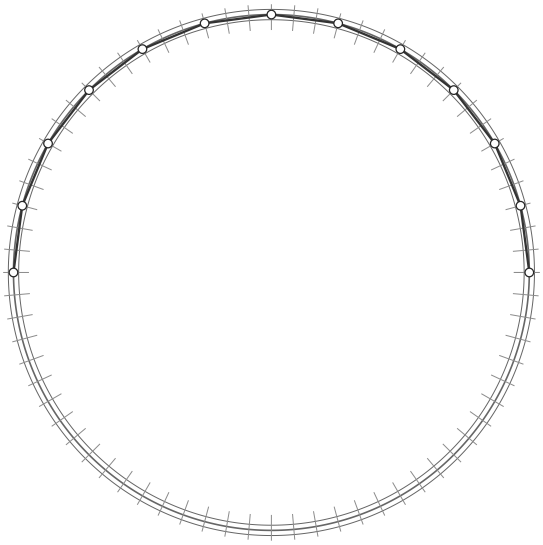
$$L_n = 2nR \sin \frac{\Theta}{2n}, \qquad R\Theta - L_n = \frac{R\Theta^3}{24n^2} + O(n^{-4}).$$

The chain approaches the arc from below. If a graph uses straight chords as its edge costs, its path length may underestimate the surface route until the sampling error is controlled. If it uses exact geodesic edge lengths, a different approximation problem remains: whether the graph offers enough directions and routes.

This matters when a translator graph is used as evidence for an underlying smooth geometry. The graph calculation is exact for the graph. Its relationship to a manifold needs a sampling and approximation argument. A good worked example shows both the discrete answer and the reference geometric answer.

The arc follows the track; the chord cuts across it

Unit-radius construction / station spacing controls the polygonal length error.



12 equal chords on a semicircle

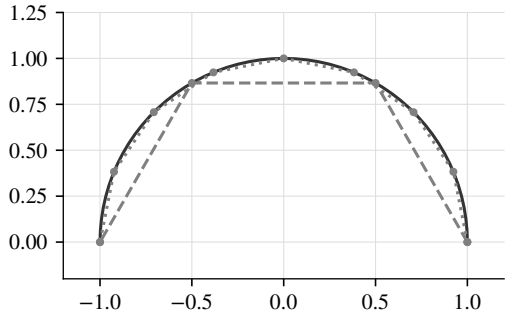
Arc length: π

Polygon length: 3.132629

Shortfall: 0.008964

Refinement improves the approximation.
The edge metric and the permitted routes still need to be specified.

A Straight chords fall short of the arc



B More chords reduce the length deficit

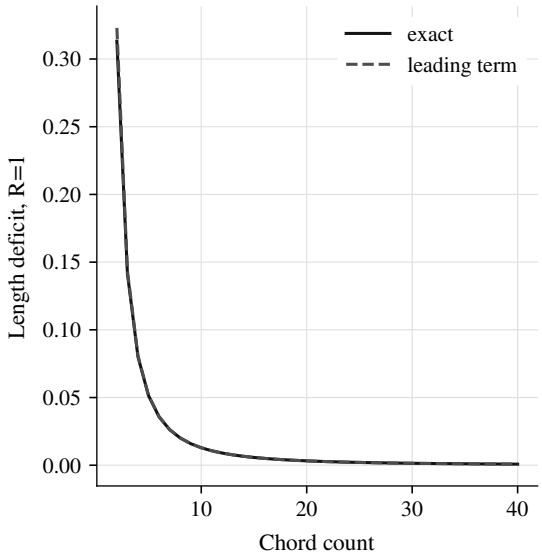
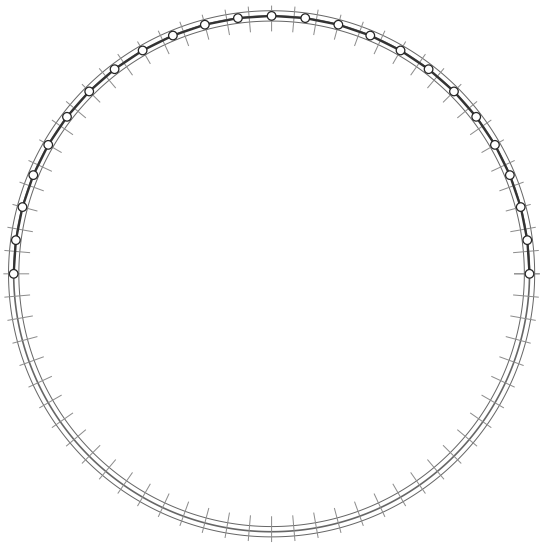


Figure 9.5: *A discrete route approaches a geometric length.* The study continues on the following figure page.

The arc follows the track; the chord cuts across it

Unit-radius construction / station spacing controls the polygonal length error.



24 equal chords on a semicircle

Arc length: π

Polygon length: 3.139350

Shortfall: 0.002242

Refinement improves the approximation.
The edge metric and the permitted routes still need to be specified.

C The computed lengths approach pi

D Edge lengths and allowed routes both matter

n=2: 2.8284271

n=3: 3.0000000

n=4: 3.0614675

n=8: 3.1214452

n=16: 3.1365485

n=32: 3.1403312

Are edge costs arcs or chords?

Does the graph offer the right routes?

Exact graph arithmetic does not alone establish a smooth realization.

Figure 9.5: *A discrete route approaches a geometric length (continued).* Chord and arc costs are compared on one circular route at several sampling densities. The exact error formula and a numerical convergence table distinguish the graph calculation from the continuous reference.

What the network has established

The translator graph has supplied a specific comparison space. We tag each coordinate by its source, accumulate lengths along translator paths, and take the infimum over the permitted routes. Because routes can be directed, the construction first gives $D(x, y)$ and then uses $\tilde{d}(x, y) = \max\{D(x, y), D(y, x)\}$. On a component where these costs are finite, identifying zero-distance classes gives (X_T, d_T).

That sequence matters in the worked network: a one-hop restriction can destroy the triangle inequality, and an unrestricted zero-cost loop can erase a distinction the task needs. The resulting metric therefore has two separate obligations. Its construction must satisfy the metric contract, and its comparisons must remain attached to the intended observations.

Round trips, cycle residuals and both extremes of the Jacobian spectrum help audit that attachment. Their sampled values describe the tested records; coverage and regularity are needed to extend a bound over a region. The path space makes the converter network's implicit ruler explicit. To identify it with a Riemannian surface, we still need the overlap and realization arguments.

The route construction supplies a comparison space. Whether its local descriptions fit one smooth geometric object remains a further question. We turn to overlap consistency, using a closed chain of measured conversions to expose the difference.

Practice with the construction

9.1. For the loop between two copies of the real line with $T_{AB}(x) = x + \varepsilon$ and $T_{BA}(x) = x$, compute the directed and symmetrized distances between 0 and 2ε in copy A.

9.2. Construct three vertices for which a one-hop restriction gives finite costs from the first to the second and the second to the third, but infinite cost from the first to the third. What metric property fails?

Chapter 10

Agreement on chart overlaps

Three laboratories timestamp the same event with separate clocks. A technician can convert the first reading directly into a reading on the third clock or go through the second clock. Those routes should give the same result. We will calculate a disagreement and use the same test to check overlapping survey maps.

Begin with measured clock offsets. Each report translates one laboratory's timestamp into another's. Following all three conversions around a loop will tell us whether they can describe one common set of clock times. We then fit the compatible part and retain the discrepancy as evidence.

Three clocks and one impossible round trip

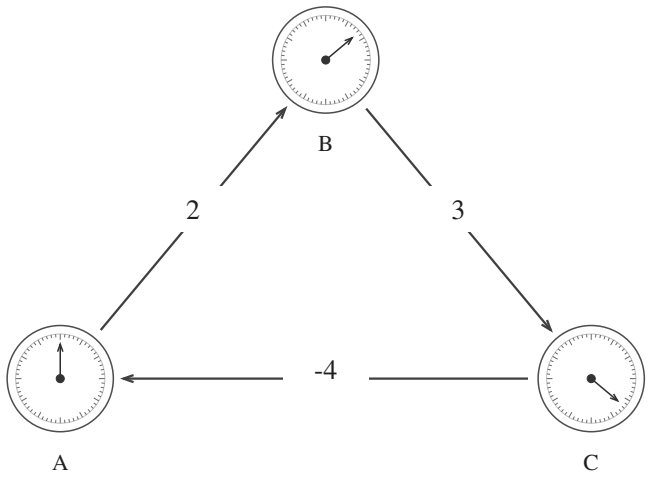
Three laboratories timestamp the same event using clocks with unknown offsets. Let t_A, t_B, t_C be their readings. Suppose the recorded translations are $t_B = t_A + 2$, $t_C = t_B + 3$ and $t_A = t_C - 4$, in seconds. Following the round trip gives $t'_A = t_A + 1$. The discrepancy is one second.

No choice of the three absolute offsets can satisfy all three equations. The inconsistency belongs to the cycle, not to a preferred starting clock. Changing all clock origins leaves the cycle sum unchanged. This is a finite, physical example of overlap compatibility: local relations must agree where they compose.

For an oriented graph with node offsets x and edge observations b, let B be its incidence matrix, so a row for $i \to j$ has -1 at i and $+1$ at j. Exact compatibility is $Bx = b$. The sum of b around every cycle must vanish. On a connected graph this condition is also sufficient: fix one node's offset, integrate along a spanning tree and use zero cycle sums to show path independence.

Local comparisons must fit one global record

The edge labels specify the calculation; the placement of nodes is schematic.



OBSERVATIONS / 2, 3, −4

Cycle sum = 1 second

Fitted offsets: 0, 5/3, 13/3

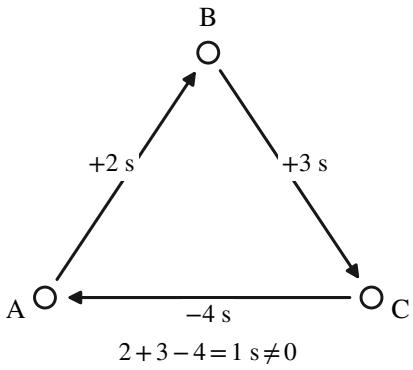
Fitted cycle sum = 0

One common origin

Changing every clock offset by the same amount changes none of the pairwise differences.

Clock dials: one turn = 12 seconds of fitted offset.

A The clock differences leave a one-second gap



B One matrix records all three comparisons

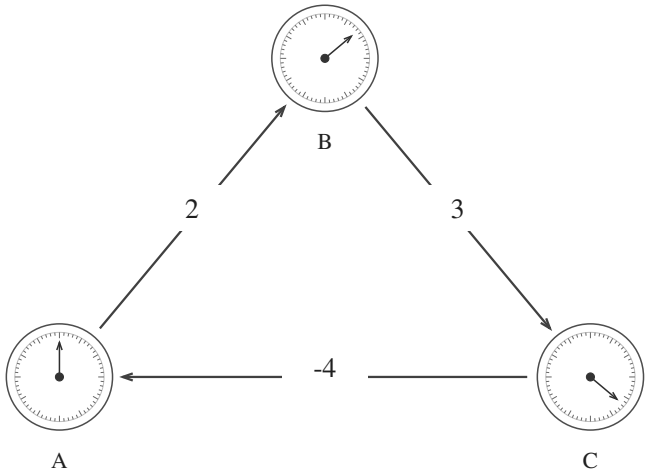
	A	B	C
A → B	-1	1	0
B → C	0	-1	1
C → A	1	0	-1

$$\min_{x_A = 0} \|Bx - b\|_2^2$$

Figure 10.1: *A one-second loop identifies an incompatible set of translations.* The study continues on the following figure page.

Local comparisons must fit one global record

The edge labels specify the calculation; the placement of nodes is schematic.



OBSERVATIONS / 2, 3, −4

Cycle sum = 1 second

Fitted offsets: 0, 5/3, 13/3

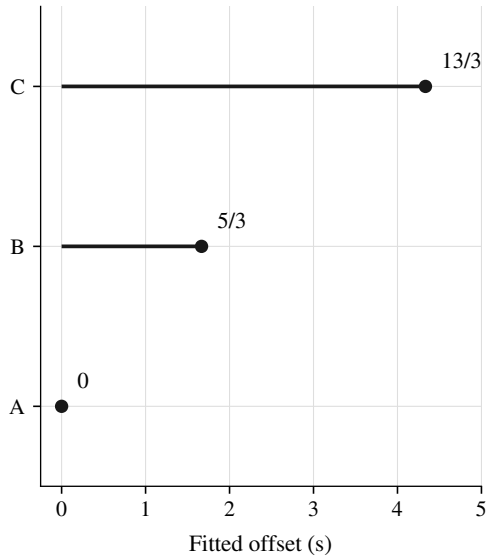
Fitted cycle sum = 0

One common origin

Changing every clock offset by the same amount changes none of the pairwise differences.

Clock dials: one turn = 12 seconds of fitted offset.

C The shared fit fixes relative clock offsets



D The residuals account for the whole gap

EDGE	*OBSERVED*	*FITTED*	*RESIDUAL*
A → B	2	5/3	−1/3
B → C	3	8/3	−1/3
C → A	−4	−13/3	−1/3
Cycle sum	1	0	−1

Seconds; equal weights; clock A fixes the time origin.

Figure 10.1: *A one-second loop identifies an incompatible set of translations (continued).* Three clock offsets are followed around a cycle. The incidence matrix, cycle residual and least-squares repair are printed beside the network. The example shows what is observable without claiming which physical clock caused the error.

A least-squares fit does not erase the evidence

When the translations are noisy, minimize

$$\frac{1}{2}(Bx - b)^T W (Bx - b),$$

where W is a positive-definite weight matrix. Differentiating gives $B^T W B x = B^T W b$. The connected graph has a constant-vector nullspace, reflecting the arbitrary common clock origin. Fixing one clock or imposing zero mean removes that freedom.

With equal weights the fitted increments are $5/3, 8/3, -13/3$ seconds. They sum to zero and each differs from its observation by $-1/3$ second. Fix $x_A = 0$ and the offsets are $x_B = 5/3$ and $x_C = 13/3$. I'd hand the next technician those settings and the residuals. The fit shares the extra second evenly. We still need evidence to tell which reading was wrong.

From overlap equations to a precise algebraic language

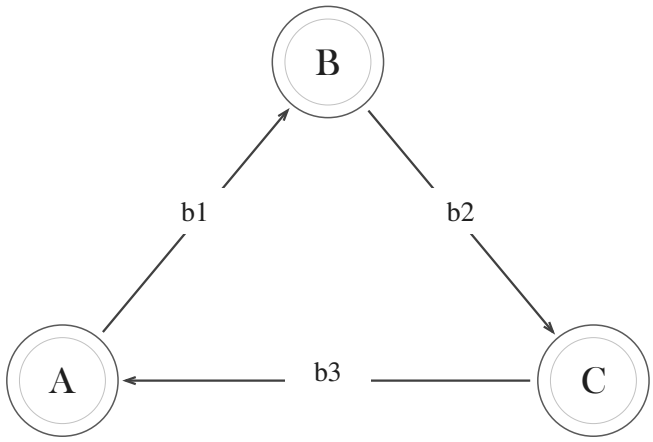
Resetting clocks changes edge differences while leaving each loop sum fixed [29]. For V clocks and E comparisons, offsets form $C^0 = \mathbb{R}^V$, readings form $C^1 = \mathbb{R}^E$, and $d_0 = B$ maps offsets to differences. Identifying readings that differ by a reset leaves $E - V + 1$ independent discrepancies on a connected graph.

Filling a face imposes a zero-sum test on its boundary. The map d_1 sums oriented readings around each face. Clock differences telescope, giving $d_1 d_0 = 0$. The quotient $\ker d_1 / \operatorname{im} d_0$ is the first cohomology. A failed face test gives nonzero $d_1 b$. A closed reading around a hole can still represent a nontrivial class if it is not a gradient.

Go back to that extra second. No clock reset can take it out of the loop. Filling the triangle means I require the three readings around it to add to zero. The extra second now fails that test. That's why I have to say which loops I've filled before I call a residual a cohomology class. Nonlinear constraints can need a different construction.

Closing the face imposes a relation

Arrows orient the comparisons; a filled face imposes the zero-sum boundary relation.



UNFILLED / one independent cycle

FILLED / boundary sum must vanish

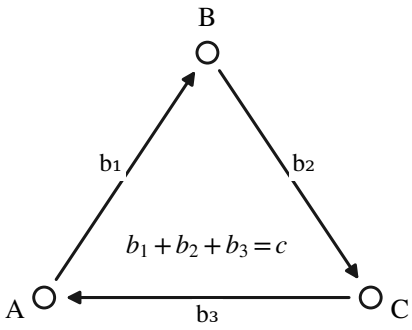
$d_1 d_0 = 0$

The 2-cell imposes a new relation.

The numerical register below specifies the exact graph and its assumptions.

A An unfilled loop can retain a mismatch

B A filled face imposes a zero-sum condition



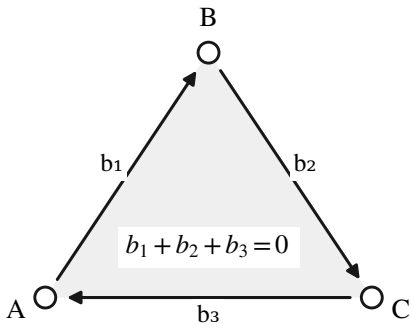
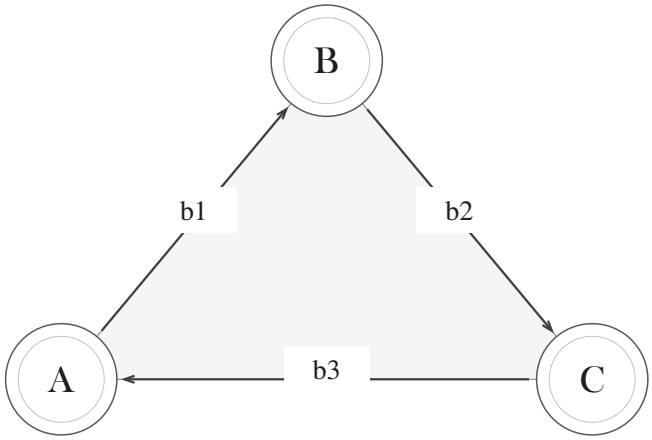


Figure 10.2: *A filled triangle and a hole have different algebra.* The study continues on the following figure page.

Closing the face imposes a relation

Arrows orient the comparisons; a filled face imposes the zero-sum boundary relation.



UNFILLED / one independent cycle

FILLED / boundary sum must vanish

$d_1 d_0 = 0$

The 2-cell imposes a new relation.

The numerical register below specifies the exact graph and its assumptions.

C The unfilled loop leaves one independent cycle

D The filled face removes that cycle

C0 dimension = 3	d1 b = b1+b2+b3
C1 dimension = 3	d1 d0 = 0
rank d0 = 2	Closed assignments have zero sum.
dim(C1 / im d0) = 1	ker d1 / im d0 has dimension 0.

Figure 10.2: *A filled triangle and a hole have different algebra (continued).* The same three oriented edges are shown without and with a filled face. Edge labels are assigned offsets. Filling the face imposes their zero-sum boundary equation; the dimension registers show the resulting change in the cohomology calculation.

Laboratory 3: audit a translator loop

Build a four-clock network with oriented edges $A \to B$, $B \to C$, $C \to D$, $A \to D$ and $A \to C$. The underlying offsets are $x = (0, 2, 5, 4)$ seconds. The exact edge differences are $b = (2, 3, -1, 4, 5)$.

Construct the incidence matrix. Use one row per edge and one column per clock. Each row contains -1 at its source and $+1$ at its destination. Verify $Bx = b$. Calculate its rank and nullspace. The rank is three; the constant vector spans the nullspace. A common clock origin is unobservable from differences alone.

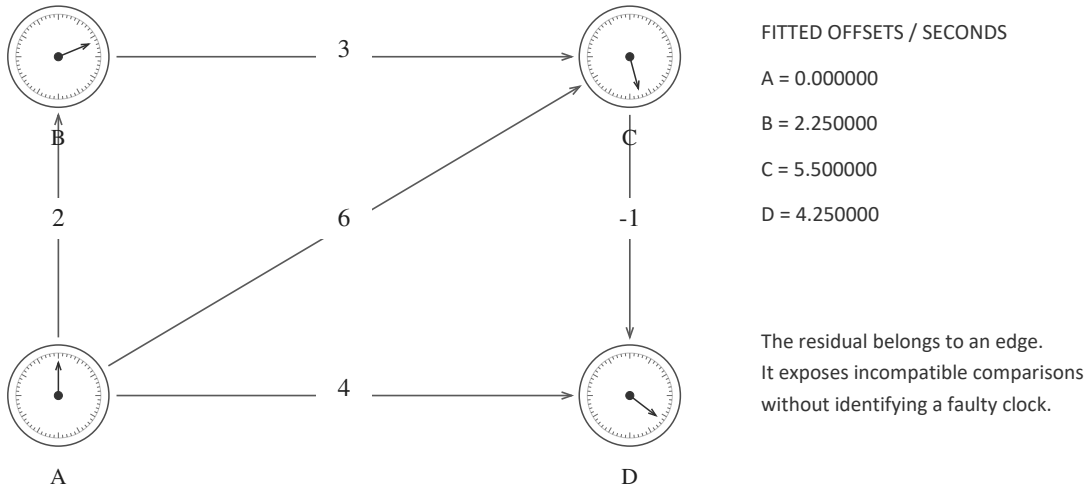
Add one inconsistent observation. Change the last edge measurement from five to six seconds. The two independent cycles now expose residuals. Solve the equal-weight least-squares problem with $x_A = 0$. Keep the fitted offsets and every edge residual, not only the total squared error. A small total can conceal a concentrated defect on one important link.

Change the weights. Give the altered edge weight four while leaving other weights one. Refit. The solution moves toward that observation and distributes more discrepancy elsewhere. The weights express confidence or a loss model; they cannot identify truth by themselves. Repeat with the altered edge given weight $1/4$ and compare.

Change the clock origins. Add arbitrary node shifts u and transform edge values to $b + Bu$. The cycle residuals should remain unchanged, since every cycle sums a gradient to zero. This is the finite version of an invariant under a change of local description.

Five observations couple four clocks

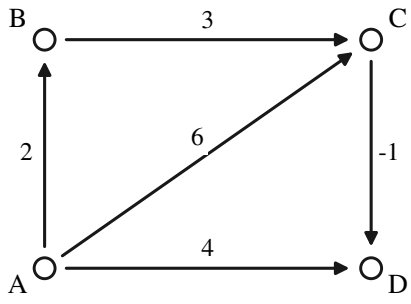
Fixing clock A at zero removes the common-origin freedom before the least-squares fit.



Dials: one turn = 12 seconds of fitted offset.

A Five comparisons couple four clocks

B The incidence matrix records each difference

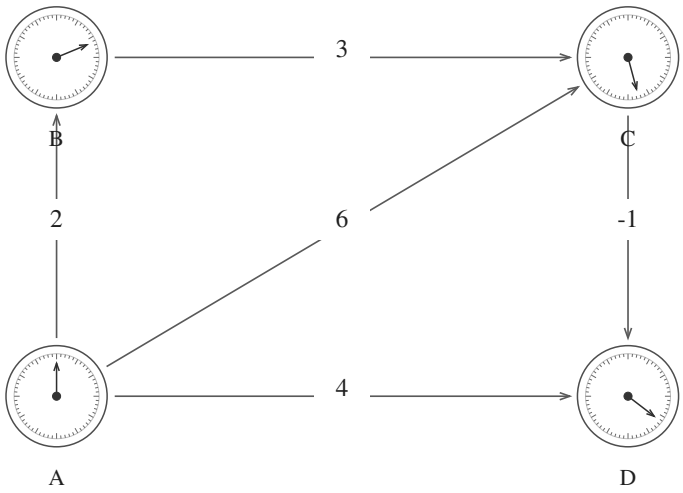


$$\begin{bmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \\ -1 & 0 & 0 & 1 \\ -1 & 0 & 1 & 0 \end{bmatrix}$$

Figure 10.3: *Laboratory record: local offsets and invariant cycles.* The study continues on the following figure page.

Five observations couple four clocks

Fixing clock A at zero removes the common-origin freedom before the least-squares fit.



FITTED OFFSETS / SECONDS

A = 0.000000

B = 2.250000

C = 5.500000

D = 4.250000

The residual belongs to an edge.
It exposes incompatible comparisons
without identifying a faulty clock.

Dials: one turn = 12 seconds of fitted offset.

C One time origin fixes the fitted offsets

D The residuals expose incompatible comparisons

A = 0.000000 seconds
B = 2.250000 seconds
C = 5.500000 seconds
D = 4.250000 seconds

Edge 1: +0.250000 seconds
Edge 2: +0.250000 seconds
Edge 3: -0.250000 seconds
Edge 4: +0.250000 seconds
Edge 5: -0.500000 seconds

Node-origin changes preserve cycle sums.

Figure 10.3: *Laboratory record: local offsets and invariant cycles (continued).* The four-clock, five-observation network is paired with its incidence matrix. Fixing clock A at zero gives the displayed least-squares offsets and five residuals. The source exercise derives the cycle checks and explains why changing time origins does not remove incompatible observations.

An exact nonlinear continuation. Replace clock offsets by rigid planar coordinate frames. A translation from frame i to frame j is a homogeneous matrix

$$T_{ji} = \begin{pmatrix} R_{ji} & t_{ji} \\ 0 & 1 \end{pmatrix}.$$

Choose three absolute frame matrices F_A, F_B, F_C and set $T_{ji} = F_j^{-1} F_i$. Verify $T_{AC} T_{CB} T_{BA} = I$ when the subscripts denote destination first. Introduce a small rotation error into one transition and compute the resulting round-trip matrix. The residual now has both rotational and translational structure; adding edge angles alone would not describe the whole transformation.

What to submit. Keep the matrix convention, incidence matrix, weights, fitted states, residuals and changed-origin check. For the rigid-frame extension, print the product order. A reversed multiplication order can produce a different physical route even when every individual matrix is correct.

What must agree beyond the clock model?

The three laboratory clocks differed by additive offsets. Now suppose two crews map the same courtyard on grids whose axes can rotate, stretch or curve. We need the corresponding compatibility conditions for those more general reports. A familiar projection such as Mercator can preserve angles while distorting areas. The quantity we compare must therefore have a specified ruler on each chart.

Declare the surveyed region

Let $\mathcal{D}$ denote the observation distribution, and let $\mathcal{R} \subseteq X_T$ be the region on which we propose to compare its descriptions. For a survey this could be the instrument's calibrated range and the part of the surface actually covered. For a learned model it is a declared operating population, together with any neighborhood needed for derivative or path calculations.

A region is a hypothesis to audit. A high-density cluster alone cannot establish its boundary, and an unfamiliar input can fall inside a familiar-looking cluster. Record the excluded cases, the rule for accepting an input, and the uncertainty of that rule. The distance comparison and any later probability statement must refer to this same declared region or population.

The local chart test now has a concrete job. It asks whether two marked states remain distinguishable after entering the global comparison space, and how their distances change. The overlap test then asks whether different reports place the same mark consistently. The camera construction at the end will show what remains undetermined even after a joint fit succeeds.

Chartable Regions of Translator Metric Space

Local charts from channels

For each channel $i \in I$, recall the map

$$\iota_i : V_i \to X_T, \qquad \iota_i(v) = q([v]_i),$$

which sends a feature vector $v \in V_i$ to its equivalence class in the translator metric quotient.

Intuitively:

- ι_i tells us where points from V_i land in X_T.

- On regions where V_i provides a good view of X_T, ι_i should behave like a coordinate chart: distances in V_i and distances in X_T should agree up to small distortion, and ι_i should not collapse distinct points that matter.

We formalize this with a local bi-Lipschitz condition.

Definition 10.1 (Local chart region)**.** A set $U_i \subseteq V_i$ is a *local chart region* for X_T if:

- $\iota_i(U_i) \subseteq \mathcal{R}$, and
- there is a constant $0 \le \varepsilon_{\text{loc}} < 1$ such that for all $v, w \in U_i$,

$$(1 - \varepsilon_{\text{loc}})\, d_i(v, w) \;\le\; d_T(\iota_i(v), \iota_i(w)) \;\le\; (1 + \varepsilon_{\text{loc}})\, d_i(v, w).$$

Unpacking the symbols and the Lipschitz condition. This definition is central to smooth chartability, so let's unpack it carefully.

The map $\iota_i : V_i \to X_T$.

Recall from Chapter 9 that X_T is the translator metric quotient: the space we get by stitching together all feature spaces $\{V_i\}$ and identifying points that are zero distance apart. Each feature space V_i has a natural map into X_T:

- Start with a vector $v \in V_i$ (a face embedding, a viral genome embedding, a transaction feature vector)
- Tag it with its space: $[v]_i = (i, v) \in \mathcal{V}$ (the disjoint union)
- Apply the quotient map $q : \mathcal{V} \to X_T$, which identifies points at zero distance
- The result $\iota_i(v) = q([v]_i)$ is an equivalence class in X_T—a point in the stitched-together space

The map ι_i places observations from channel i in the quotient. It is an embedding on a region that satisfies the two-sided bound. The operation asks: "take this image embedding and tell me where it lives in the global metric space."

What does "local chart region" mean?

We're saying that $U_i \subseteq V_i$ is a subset of feature space V_i that acts like a good coordinate chart for part of X_T. Specifically:

- The image $\iota_i(U_i)$ lands in the in-regime region $\mathcal{R} \subseteq X_T$ (we're only charting the part of the space we care about)
- Distances measured in U_i using d_i approximately match distances measured in X_T using d_T, up to small distortion ε_{loc}

This is like saying: "The Portuguese captain's map of the West African coast is a good chart for that part of the world—distances measured on the map correspond accurately to real sailing distances, with at most 5% error."

What is a bi-Lipschitz condition? A brief history.

The term "Lipschitz" honors Rudolf Lipschitz (1832–1903), a German mathematician who studied how functions can change. A function $f : X \to Y$ between metric spaces is *Lipschitz continuous* if there's a constant L such that:

$$d_Y(f(x_1), f(x_2)) \le L \cdot d_X(x_1, x_2)$$

for all $x_1, x_2 \in X$.

In words: f never stretches distances by more than a factor of L. If two points are distance d apart in X, their images under f are at most distance $L \cdot d$ apart in Y.

Think of L as the "maximum magnification" of the function. For example:

- A camera zoom lens with 2× optical zoom is Lipschitz with $L = 2$: objects get at most twice as large
- A function that collapses everything to a single point is Lipschitz with $L = 0$ (degenerate case)
- A function with wild oscillations or infinite slopes is *not* Lipschitz

A function is *bi-Lipschitz* if it's Lipschitz in both directions: it doesn't stretch *or* compress too much. Formally, f is bi-Lipschitz if there exist constants $0 < c < C$ such that:

$$c \cdot d_X(x_1, x_2) \leq d_Y(f(x_1), f(x_2)) \leq C \cdot d_X(x_1, x_2).$$

The lower bound c says: f doesn't collapse distant points into nearby points (no excessive compression). The upper bound C says: f doesn't magnify small distances into huge distances (no excessive stretching).

A bi-Lipschitz map is a *good coordinate change*: it preserves the rough shape and scale of the space, distorting distances only by bounded factors.

Our specific bi-Lipschitz condition.

The definition says that on U_i, the map $\iota_i : V_i \to X_T$ is bi-Lipschitz with constants $c = 1 - \varepsilon_{\text{loc}}$ and $C = 1 + \varepsilon_{\text{loc}}$:

$$(1 - \varepsilon_{\text{loc}})\, d_i(v, w) \;\leq\; d_T\big(\iota_i(v), \iota_i(w)\big) \;\leq\; (1 + \varepsilon_{\text{loc}})\, d_i(v, w).$$

Let's unpack this inequality piece by piece:

- $d_i(v, w)$: The distance between vectors v and w in feature space V_i, measured using V_i's metric d_i. For example, Euclidean distance between face embeddings.
- $d_T(\iota_i(v), \iota_i(w))$: The distance between the same two vectors after lifting them into X_T via ι_i. This is the translator metric distance: the shortest path length through the translator graph.
- ε_{loc}: The *local distortion parameter*. This is a small positive number (ideally $\varepsilon_{\text{loc}} \ll 1$, like 0.05 or 0.1) that bounds how much ι_i distorts distances.
- **Upper bound** $(1 + \varepsilon_{\text{loc}})d_i(v, w)$: The distance in X_T is at most $(1 + \varepsilon_{\text{loc}})$ times the distance in V_i. The map doesn't stretch distances by more than a factor of 1.05 (if $\varepsilon_{\text{loc}} = 0.05$).
- **Lower bound** $(1 - \varepsilon_{\text{loc}})d_i(v, w)$: The distance in X_T is at least $(1 - \varepsilon_{\text{loc}})$ times the distance in V_i. The map doesn't compress distances by more than a factor of 0.95 (if $\varepsilon_{\text{loc}} = 0.05$).

What this condition captures.

The bi-Lipschitz condition says that ι_i is a *nearly isometric embedding*: it preserves distances almost perfectly, up to a small relative error ε_{loc}.

Concrete example: Face embeddings as charts. Suppose $V_{\text{img}} = \mathbb{R}^{512}$ has its Euclidean metric. The image encoder returns unit vectors, and a path can use a straight segment in the ambient feature space. Take two face embeddings:

- v_1: A 25-year-old person, frontal lighting, neutral expression
- v_2: The same person at age 26, slightly different lighting, smiling

In V_{img}, the Euclidean distance is $d_{\text{img}}(v_1, v_2) = 0.08$ (normalized embeddings, so distance is between 0 and 2).

Lift the embeddings into X_T through ι_{img}. Their comparison is

$$d_T\big(\iota_{\text{img}}(v_1), \iota_{\text{img}}(v_2)\big).$$

Use the translator path construction and its stated symmetrization to evaluate this distance. If V_{img} is a good local chart, this distance should be close to 0.08:

- If $\varepsilon_{\text{loc}} = 0.05$, the bi-Lipschitz condition says:
$$0.95 \times 0.08 \leq d_T(\iota_{\text{img}}(v_1), \iota_{\text{img}}(v_2)) \leq 1.05 \times 0.08$$
$$0.076 \leq d_T \leq 0.084.$$
- The translator metric distance should be between 0.076 and 0.084—within 5% of the image-space distance.

If the translator metric gives $d_T = 0.079$, this pair passes the distance test. A bound for the whole region must control every pair in its stated domain. Establishing a smooth chart also requires the topology and transition conditions below.

If $d_T = 0.02$, the pair fails the lower bound. A route through other channels may make these records too cheap to connect in both directions. A result like 0.15 would instead signal a problem with the calculation or its route assumptions: the straight segment costs 0.08 in either direction, so $d_T \leq 0.08$ here.

Why this matters for smooth chartability. A covering by local bi-Lipschitz regions still needs a topology check. Consider three unit rays joined at one endpoint. Each pair of rays, with its path distance, is isometric to an interval, and those pairs cover the junction. Yet every neighborhood of the junction branches into three local pieces. A genuine manifold chart has additional structure:

- Each chart U_i gives coordinates for part of $\mathcal{R}$
- Distances measured in chart coordinates match distances in X_T accurately
- Transitioning between charts (via translators) doesn't introduce wild distortions

For the courtyard survey, the distance bounds measure whether the map's ruler agrees with the surveyed one. An open chart image and compatible smooth transitions let that ruler be part of a manifold atlas.

The constant ε_{loc} is a relative length error: 0.05 means a five-percent allowance. Its acceptability depends on the decision margin. Even a small value does not supply smooth transitions or prove that the chart images are open. Those are separate requirements.

In practice, we estimate ε_{loc} empirically by:

1. Sample pairs (v, w) from in-regime data in V_i
2. Compute $d_i(v, w)$ (distance in feature space)
3. Compute or approximate $d_T(\iota_i(v), \iota_i(w))$ (distance in translator metric space, using shortest paths)
4. Check whether the bi-Lipschitz bounds hold
5. Report the largest sampled relative distortion separately from any certified uniform bound ε_{loc}

Chapter 21 will provide concrete diagnostic procedures for this. For now, the key insight is: **local bi-Lipschitz charts with small ε_{loc} are the foundation of smooth chartability.** To obtain a manifold atlas, also require open chart images, suitable topology and exact compatible smooth transitions. Approximate diagnostics motivate a realization; they do not construct one unaided.

This says that within U_i, the feature space metric d_i and the translator metric d_T agree up to a small multiplicative distortion. In other words, U_i with coordinates in V_i is an almost-isometric local chart for $\mathcal{R}$.

For the illustrative system that converts among still images, video and logs, a chart region $U_{\text{img}} \subseteq V_{\text{img}}$ might correspond to face embeddings produced by a particular encoder on a certain demographic and pose range. Within that region, the encoder's geometry and the translator metric geometry agree closely.

Smooth, well-conditioned chart transitions

Local charts are only the first step. To resemble a Riemannian manifold, $\mathcal{R}$ also needs *smooth* and *well-conditioned* transitions between overlapping charts.

Suppose $U_i \subseteq V_i$ and $U_j \subseteq V_j$ are local chart regions whose images overlap in $\mathcal{R}$:

$$\mathcal{R}_{ij} = \iota_i(U_i) \cap \iota_j(U_j) \neq \emptyset.$$

On this overlap, we can define an approximate *transition map* from U_i to U_j by following the translators:

$$\Phi_{ij} : U_i \supseteq W_{ij} \to V_j,$$

where $W_{ij} = \iota_i^{-1}(\mathcal{R}_{ij})$ is the overlap domain. For injective chart maps, the exact coordinate transition is $\Phi_{ij} = (\iota_j|_{U_j})^{-1} \circ \iota_i$. A learned translator with $T_{ij}(W_{ij}) \subseteq U_j$ is compared with it through

$$\Phi_{ij}(v) \approx T_{ij}(v).$$

We do not insist on exact equalities here; we only need T_{ij} to behave like a smooth, well-conditioned map between overlapping chart regions. Concretely, we require:

- **Regularity:** C^1 means differentiable with continuous Jacobian; this suffices for the local derivative diagnostics. Where a smooth atlas is asserted, require C^∞ transitions and smooth local inverses on the exact overlap domains.
- **Bounded distortion:** There are constants $\lambda_{\min}, \lambda_{\max} > 0$ such that for all tangent directions δv at in-regime points,

$$\lambda_{\min}\|\delta v\| \;\le\; \|JT_{ij}(v)\,\delta v\| \;\le\; \lambda_{\max}\|\delta v\|.$$

- **Small round-trip error:** For in-regime $v \in W_{ij}$,

$$d_i\big(T_{ji}(T_{ij}(v)), v\big) \le \varepsilon_{\text{rt}},$$

  for some small ε_{rt}.
- **Small cocycle slack:** For triples of overlapping charts i, j, k, the pointwise cocycle residual

$$\text{Cyc}_{i\to j\to k}(v) = d_k(T_{jk}(T_{ij}(v)), T_{ik}(v))$$

  is bounded by a small ε_{cyc} on in-regime v.

These four conditions together say that translators between chart regions behave like well-conditioned coordinate changes on a smooth manifold: their required differentiability is specified, they do not blow up or collapse directions (bounded distortion), they are approximately invertible (bounded round-trip residual), and different routes agree (small cocycle slack). These are quantitative checks on proposed chart transitions. The exact gluing construction additionally requires exact overlap compatibility, open chart images and the stated separation and countability conditions.

On the harbor charts of Chapter 9, nearby marks should convert to nearby marks, and the conversion should preserve small displacements within the stated distortion bounds.

Smooth chartability (assumption)

We can now bundle these ideas into a single hypothesis.

Assumption (Smooth chartability). A translator system is *smoothly chartable* on the in-regime region $\mathcal{R} \subseteq X_T$ if there exists:

- A collection of local chart regions $U_i \subseteq V_i$ for a subset of channels $i \in I_{\text{chart}}$.
- Constants $0 \leq \varepsilon_{\text{loc}} < 1$, $\varepsilon_{\text{rt}}, \varepsilon_{\text{cyc}} \geq 0$ and $0 < \lambda_{\min} \leq \lambda_{\max} < \infty$.

such that:

1. **Coverage:** The images of the chart regions cover $\mathcal{R}$:
$$\mathcal{R} \subseteq \bigcup_{i \in I_{\text{chart}}} \iota_i(U_i).$$
2. **Local near-isometry:** For each $i \in I_{\text{chart}}$ and all $v, w \in U_i$,
$$(1 - \varepsilon_{\text{loc}})\, d_i(v, w) \;\leq\; d_T\big(\iota_i(v), \iota_i(w)\big) \;\leq\; (1 + \varepsilon_{\text{loc}})\, d_i(v, w).$$
3. **Smooth, well-conditioned transitions:** On overlaps, translators T_{ij} between chart regions behave like smooth, bi-Lipschitz maps with bounded Jacobians, bounded round-trip residual ε_{rt}, and small cocycle slack ε_{cyc} as described above.

The residuals quantify disagreements between conversions. The two singular-value bounds track absolute scale, and their ratio measures anisotropy. Exact overlap maps between genuine charts obey the coordinate identities. Learned maps can be assessed against them with nonzero residuals.

What smooth chartability looks like in practice. In a hypothetical model comparing blood work, imaging, symptoms and genetic markers, the following chosen values illustrate the diagnostics. They are assumptions for this example, not measured patient results. Smooth chartability would mean that we can find several well-behaved regions in the feature spaces—say, one chart covering typical inflammatory conditions, another for metabolic disorders, another for acute infections—such that:

- These regions cover the clinical presentations and test profiles we care about (the conditions the hospital actually treats, not exotic rare diseases).
- Within each region, the feature space metric closely matches the translator metric. For example, if two blood work vectors are distance $d_{\text{blood}} = 0.10$ apart in blood feature space, the translator metric distance d_T should be between 0.095 and 0.105 (local distortion $\varepsilon_{\text{loc}} \approx 0.05$).
- When we translate a blood work embedding to imaging space and back, we land close to where we started: round-trip error $\varepsilon_{\text{rt}} \approx 0.03$–$0.08$ on in-regime patient data. Going blood→imaging →blood doesn't lose much information.
- If we route through different intermediate spaces, we land in approximately the same place. A route through imaging should give roughly the same final genetic representation as a route through symptoms (cocycle bound $\varepsilon_{\text{cyc}} \approx 0.05$–$0.10$ on overlapping regions).
- The translators have well-conditioned Jacobians: singular values bounded between $\lambda_{\min}$ and $\lambda_{\max}$ with condition number $\kappa = \lambda_{\max}/\lambda_{\min}$ on the order of 2–5 in clinically relevant directions. The translator doesn't wildly magnify noise in some directions while collapsing signal in others.

These diagnostics can be estimated from matched observations, with a stated metric, sampling population and numerical method. Chapter 21 develops that audit. Finite observations establish sampled performance; extending the bounds to every point in a region requires additional regularity and coverage arguments.

The geometric question remains precise: do the local descriptions agree about the same underlying object and its ruler? Exact compatible charts yield the construction proved in Chapters 11 and 12. Approximate residuals inform a proposed realization and its uncertainty; they do not alone certify global topology.

Interpretation. Smooth chartability does not say that the *entire* translator metric space X_T is manifold-like. It says that on the region $\mathcal{R}$ where we have data and care about detection, we can:

- Use some of the feature spaces V_i as local coordinate charts.
- Glue these charts together via translators that behave like smooth, well-conditioned chart transitions.
- Control the amount of drift and inconsistency at seams via explicit small constants.

For those image, video and log converters, Jacobians, round-trip errors and cocycle residuals audit the proposed local descriptions. To call them an atlas on an identity model, establish the exact chart and topology requirements or supply a separately justified approximate realization. Matched-identity checks remain necessary to connect that model to the people observed.

Apply the clock test to three survey maps

The clock model tested an additive translation. Three survey teams now measure the same courtyard. The first uses metres, the second centimetres, and the third millimetres with a shifted origin. Their pairwise conversion rules should agree regardless of the route used to convert a marked point. A discrepancy between the direct and indirect conversions belongs to the conversion system; it is not a new motion of the mark.

For exact charts φ_i, define $T_{ij} = \varphi_j \circ \varphi_i^{-1}$ on the relevant overlap. Then

$$T_{ji}T_{ij} = \mathrm{id}, \qquad T_{jk}T_{ij} = T_{ik}$$

where all terms are defined. These are the inverse and cocycle identities. The second follows by cancelling $\varphi_j^{-1}\varphi_j$ on the common domain. Domain restrictions are part of the equality. A composite cannot be tested at a point that leaves the chart where its next map is defined.

For approximate maps, record the residual

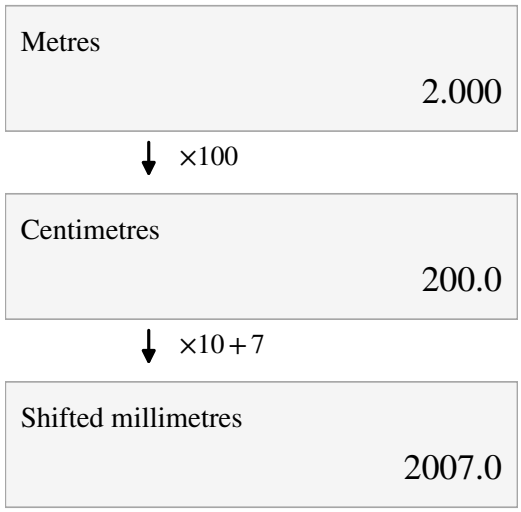
$$r_{ijk}(x) = \|T_{jk}(T_{ij}(x)) - T_{ik}(x)\|.$$

The norm must use the same target ruler for both outputs. If coordinates are rescaled, the numerical residual changes unless the metric changes with them. A report that states only a bare residual without units or normalization does not yet describe geometric distortion.

From a sample to a bound

The survey teams can compare the two conversion routes at a finite set of courtyard marks. How much disagreement could they miss between marks? Suppose the residual r is L-Lipschitz on the compact test region and the marks form an h-net, so every point is within distance h of a tested

A *Three reports locate one point*



B *Compatible routes agree exactly*

Route	*Output (mm)*
Direct: 1000x + 7	2007.0
Via centimetres	2007.0
Exact residual	0.0
Biased direct route	2007.2
Biased residual	0.2

$$T_{23}(T_{12}(x)) = 10(100x) + 7 = T_{13}(x)$$

A 0.2 mm residual is 0.0002 m; the target metric carries the conversion.

Figure 10.4: One marked position converted through three coordinate systems. Both exact routes end at the same millimetre address. An added 0.2-millimetre bias appears as a target-coordinate residual, whose physical size also requires its unit.

mark. Then

$$\sup_x r(x) \le \max_j r(x_j) + Lh.$$

To prove it, choose a tested point within h of each x and use the Lipschitz inequality. Without coverage and a regularity bound, a small maximum on the observed sample is only a statement about that sample. This elementary result gives the audit a precise route from finite testing to a uniform conclusion when the needed information is available.

For a smooth transition in Euclidean coordinates, a uniform derivative bound supplies a Lipschitz bound on a convex domain. For curved regions or nonconvex domains, use paths remaining in the domain and the corresponding intrinsic metric. The geometry determines how the sampling coverage must be measured.

Clock offsets required compatible differences and an absolute reference. Reconstructing a spatial frame asks for the same discipline with directional measurements: identify what the observations determine and what freedom remains.

Reconstructing one workbench from two camera frames

Two calibrated overhead cameras observe the same three marks on a flat workbench. Each camera reports coordinates in its own orthonormal plane. The task is to translate the reports while preserving the workbench distances and the identity of each mark. This is a genuine coordinate problem when the point correspondences and the rigid-frame model are supplied.

Let camera A report

$$p_1 = (0,0), \quad p_2 = (1,0), \quad p_3 = (0,1).$$

Camera B reports

$$q_1 = (3,2), \quad q_2 = (3,3), \quad q_3 = (2,2).$$

The transform $q = Rp + t$ has

$$R = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \qquad t = (3,2).$$

Every mark has moved on the coordinate display; no mark has moved on the bench. The translation affects coordinates of points but disappears from differences. Orthogonality gives $\|R(p_i - p_j)\| = \|p_i - p_j\|$.

Calculate the transform from the observations

Subtract the first mark from the other two. Form

$$P = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad Q = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

The columns are the two observed displacement vectors. Since P is invertible, $R = QP^{-1}$, and $t = q_1 - Rp_1$. This direct construction makes the assumptions visible: correspondence is known, the marks span the plane, and the two coordinates measure the same physical units.

If camera B used centimeters while A used meters, the computed matrix would satisfy $R^T R = 10^4 I$, not I. It would be a scaled coordinate transformation, and the output ruler would need the reciprocal scale. If one image were reflected, the determinant would be negative. Both cases can preserve pairwise relationships while violating the proposed orientation-preserving rigid model.

What happens when the observations are noisy

For more marks and measurement noise, estimate a rotation and translation by minimizing

$$J(R,t) = \sum_i \|q_i - Rp_i - t\|^2, \qquad R^T R = I, \quad \det R = 1.$$

For fixed R, differentiating in t gives

$$t = \bar{q} - R\bar{p}.$$

The remaining problem uses centered coordinates $p'_i = p_i - \bar{p}$ and $q'_i = q_i - \bar{q}$. Expanding the squares shows that minimizing J is equivalent to maximizing

$$\sum_i (q'_i)^T R p'_i.$$

In the plane, write $R(\alpha) = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix}$. Define

$$A = \sum_i (q'_{ix} p'_{ix} + q'_{iy} p'_{iy}),$$
$$B = \sum_i (-q'_{ix} p'_{iy} + q'_{iy} p'_{ix}).$$

The quantity to maximize is $A\cos\alpha + B\sin\alpha$. If $(A, B) \neq (0, 0)$, its maximum occurs at

$$\widehat{\alpha} = \operatorname{atan2}(B, A).$$

This is a complete planar registration algorithm, derived from the stated loss. A different uncertainty model can require weights and change the estimator. The model should be chosen before interpreting its residual as evidence of a bad correspondence.

A residual is more informative when its pattern is retained

Suppose only the third B observation is shifted by a small amount. The fitted transform moves all three predictions to reduce the total squared discrepancy. That is expected: one shared frame must explain all points. Keep the three residual vectors, not only their summed squared length. A systematic radial pattern can suggest a missing scale; an orientation reversal can suggest a handedness mismatch; one isolated large residual can suggest an incorrect correspondence. Those are diagnostic possibilities, not proofs about the cause of an error.

To check the exact model, also inspect the three pairwise distances. They are invariant under a rigid motion. If one distance changes while the others remain fixed, no single exact rigid transform can reconcile the two triangles. This gives an independent geometric check before solving the least-squares problem.

Compose frames without changing their meaning

Use homogeneous matrices

$$T_{BA} = \begin{pmatrix} R_{BA} & t_{BA} \\ 0 & 1 \end{pmatrix}$$

to map A coordinates to B coordinates. If a third camera supplies T_{CB}, then $T_{CA} = T_{CB}T_{BA}$. The translation part is

$$t_{CA} = R_{CB}t_{BA} + t_{CB}.$$

It is not generally $t_{BA} + t_{CB}$, because the two translations were expressed in different frames. The rotation must act before they can be added.

For a three-camera loop, exact compatibility requires

$$T_{AC}T_{CB}T_{BA} = I.$$

A changed global workbench frame does not alter the relative observations. If all world-frame poses are multiplied on the left by the same rigid transform, the relative transforms $F_j^{-1}F_i$ remain unchanged. This is gauge freedom: the data determine relative geometry without determining the absolute coordinate frame.

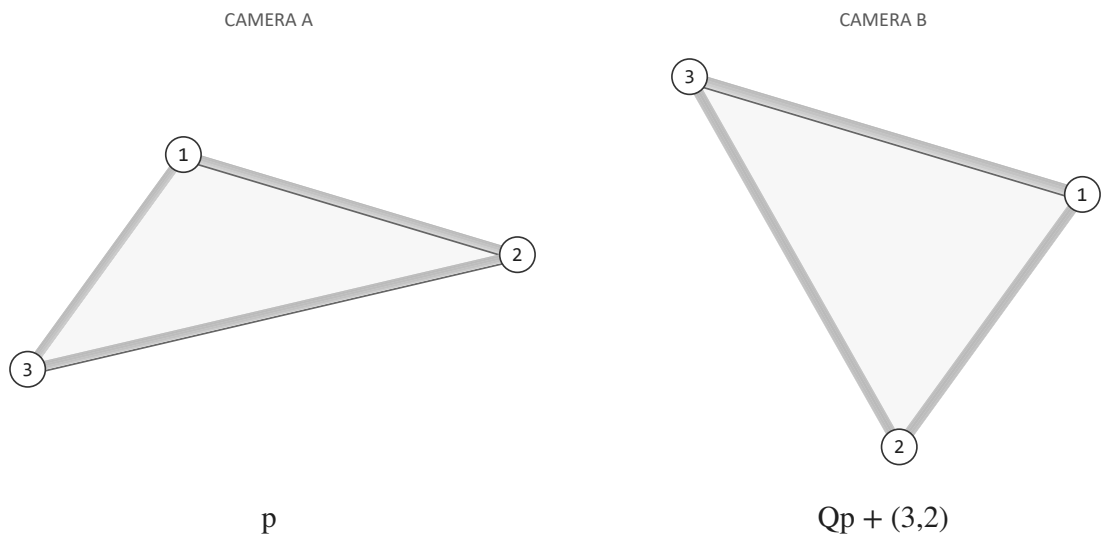
Why a regularizer is not an observation

An optimizer solving for all camera poses must remove that freedom or otherwise handle its nullspace. Fixing the first camera is one legitimate convention. Adding a penalty that favors a particular global pose also makes the numerical system invertible, but it introduces a preference rather than new measured information. The distinction matters when reporting uncertainty.

The same logic applies to the clock network. A common time offset is invisible to differences, just as a common rigid motion is invisible to relative poses. Both systems describe a common object through local reports. Their ambiguity is determined by the observation rule, and their consistency is tested through compositions on overlaps.

A rigid transformation preserves the marked fixture

Three planar marks: rotation +90° and translation (3,2) give the exact second view.



Changing one mark alters a pairwise distance; a common rigid fit cannot erase that inconsistency.

A Three marks define the first camera view

B The same marks appear after a rigid transform

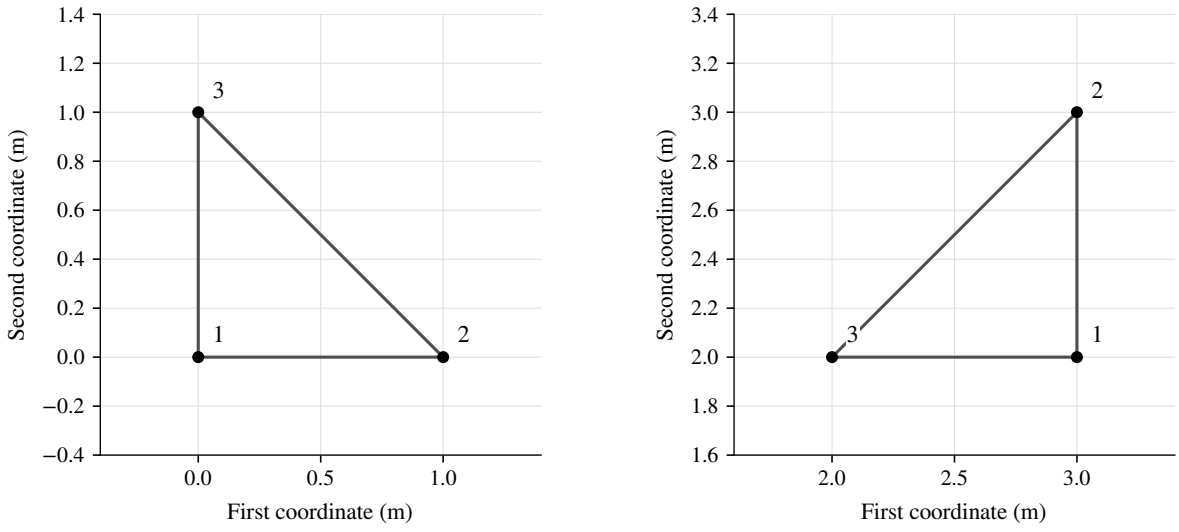
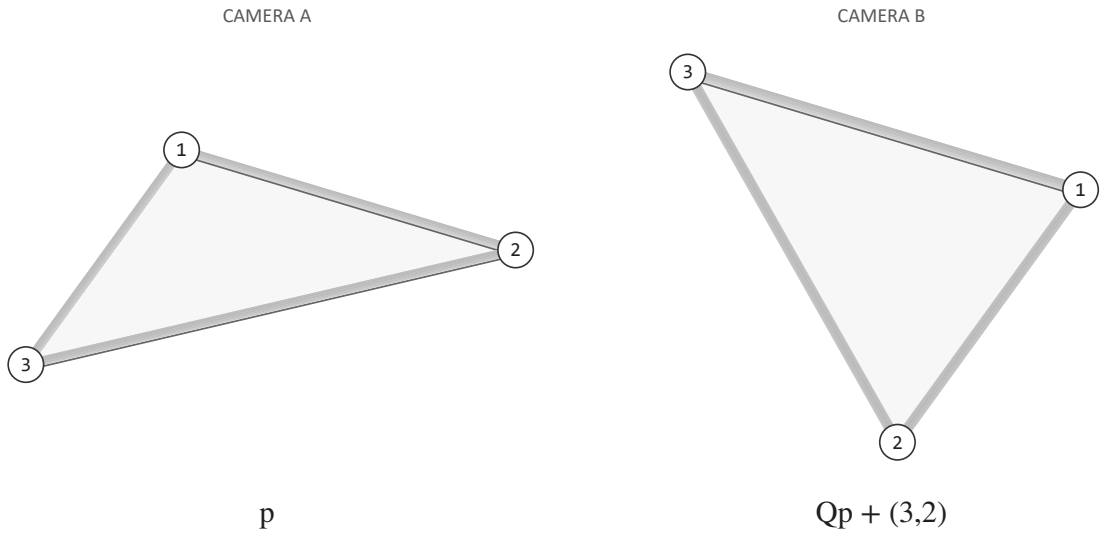


Figure 10.5: *The workbench stays fixed while its coordinates change.* The study continues on the following figure page.

A rigid transformation preserves the marked fixture

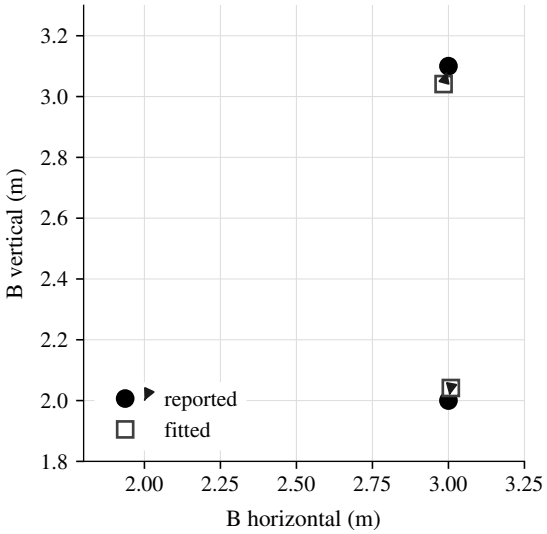
Three planar marks: rotation +90° and translation (3,2) give the exact second view.



Changing one mark alters a pairwise distance; a common rigid fit cannot erase that inconsistency.

C One altered mark leaves a registration residual

D The fitted transform records the perturbation



Exact: rotation +90°

Exact: translation (3, 2) m

Altered fit: 91.36393°

Fitted shift: (3.007840, 2.041362)

The fit cannot erase inconsistent triangle lengths.

Figure 10.5: *The workbench stays fixed while its coordinates change (continued).* The exact triangles are related by a quarter-turn and translation. Altering one observation makes their distance registers inconsistent; the shared least-squares fit distributes the remaining residuals.

A mismatched frame. Change the second B point to $(3, 3.1)$. Calculate the three pairwise-distance discrepancies, then evaluate A, B, the fitted angle, and the translation. Explain why the best fit cannot make every residual zero.

From agreement to a common object

The clock loop made an overlap failure measurable: individually plausible offsets did not describe one common set of clock times. The exact survey conversions supplied the contrasting case, where different coordinate routes reached the same marked point. The reconstruction study then kept the remaining freedom explicit instead of confusing a good fit with a unique frame.

The semantic structure S specifies the entities, allowed evolution and observation channels. The translator quotient supplies a candidate comparison space. Smooth chartability states local compatibility and distortion requirements. An exact atlas additionally supplies genuine charts, smooth transitions, appropriate topology and compatible positive-definite tensors. An approximate realization must supply its map, operating region and quantitative distance bounds.

Ch.	Component	Cartographic analogy
1–2	Semantic sameness structure S	The territory: hidden coastline, latent entities evolving benignly
3	Two realizations: translators & geometry	Two mapping strategies: local surveys with translators vs. global projection on curved surface
4–8	Metric spaces, manifolds, geodesics, Riemannian geometry	Mathematical toolkit: distances, paths, curvature, Jacobians
9	Translator metric quotient (X_T, d_T)	Stitched-together quilt from survey charts; distance = shortest path length through patches
10	Smooth chartability	Quality checks: Do charts overlap smoothly? Small $\varepsilon_{\mathrm{rt}}, \varepsilon_{\mathrm{cyc}}, \varepsilon_{\mathrm{loc}}$?
11–12	Riemannian realizability	Reconstruction: Does a smooth globe exist that the stitched maps approximate?

First derive translators from a known geometric space. Then ask when compatible translators reconstruct such a space. Once we can compare the rulers, Chapter 20 uses the probability of comparison failure and the share of the population near the threshold to transfer a decision guarantee.

Practice with the construction

Back to the maps and the globe. Three charts do not become compatible because they use familiar place names. Their coordinate conversions must bring the same mark to the same location, and the audit must specify the region on which that agreement is supported.

10.1. Take $T_{12}(x) = 100x$, $T_{23}(y) = 10y + 7$ and $T_{13}(x) = 1000x + 7$. Verify the cocycle identity. Then add 0.2 to the direct route and calculate the residual in millimetres.

10.2. The largest tested residual is 0.012, the net radius is 0.01 and $L = 0.4$. What uniform bound follows? Which part of the conclusion fails if the tests are merely convenient examples rather than an h-net?

Chapter 11

From an atlas to translators

Start with a globe whose surface distances are known. A navigator switches between two maps and wants to know whether the switch changes a route's measured length. The globe supplies the map conversions and the surface ruler, giving us an exact reference for testing approximate conversions.

Build the translator system from the globe

Suppose we already know the Riemannian manifold (M, g). The translation account can now be constructed explicitly. Choose charts $\varphi_i : U_i \to V_i \subset \mathbb{R}^d$ covering M. Give each coordinate domain the metric carried from g through φ_i^{-1}. On overlaps, use $T_{ij} = \varphi_j \circ \varphi_i^{-1}$.

The translator remembers which chart a coordinate belongs to. The tagged pair (i, x) represents the point $\varphi_i^{-1}(x)$ on the globe. Changing to $(j, T_{ij}(x))$ changes the address without changing that point. This provides an exact setting for free translator jumps.

Theorem 11.1 (Atlas path realization)**.** *For a connected Riemannian manifold, use all finite piecewise-smooth chart paths with exact overlap transitions and the transported local metrics. The zero-distance quotient of tagged descriptions is isometric to* (M, d_g).

Proof. Every translator path projects to a continuous piecewise-smooth path in M. Each segment has the same length by Proposition 5.3; jumps identify one point. Its total length is therefore at least the Riemannian distance of its endpoints.

Conversely, a piecewise-smooth path in M has compact parameter interval. The inverse images of the chart domains form an open cover of that interval. A finite subdivision can be chosen so that every segment lies in a chart. Describing each segment there gives a translator path of the same length. Taking the infimum proves the reverse inequality. Two tagged descriptions then have zero distance exactly when they represent the same point, since d_g is a metric. □

We began with the globe, wrote down its maps and transitions, and proved that the two methods measure the same routes. They agree because each exact coordinate change carries the local ruler with it.

Different atlases, the same construction

Redraw one big harbor chart as several smaller, overlapping sheets. I may have to turn more pages to trace the route, but every time I change sheets I pick up at the same spot on the water. More pages, same distance. The quotient distance stays the same.

This is why the number of available charts is not itself a physical capacity or a distance. An atlas can be made arbitrarily redundant. Capacity claims must concern a specified computation, resolution, resource or class of distinguishable states rather than the raw count of maps.

If the application forbids some paths, apply the same construction to the specified admissible family. The equality still follows if chart subdivision, concatenation as required, and coordinate transition preserve that family. It then gives the corresponding constrained path cost. Whether this cost is finite, symmetric and Riemannian is a further question about those constraints.

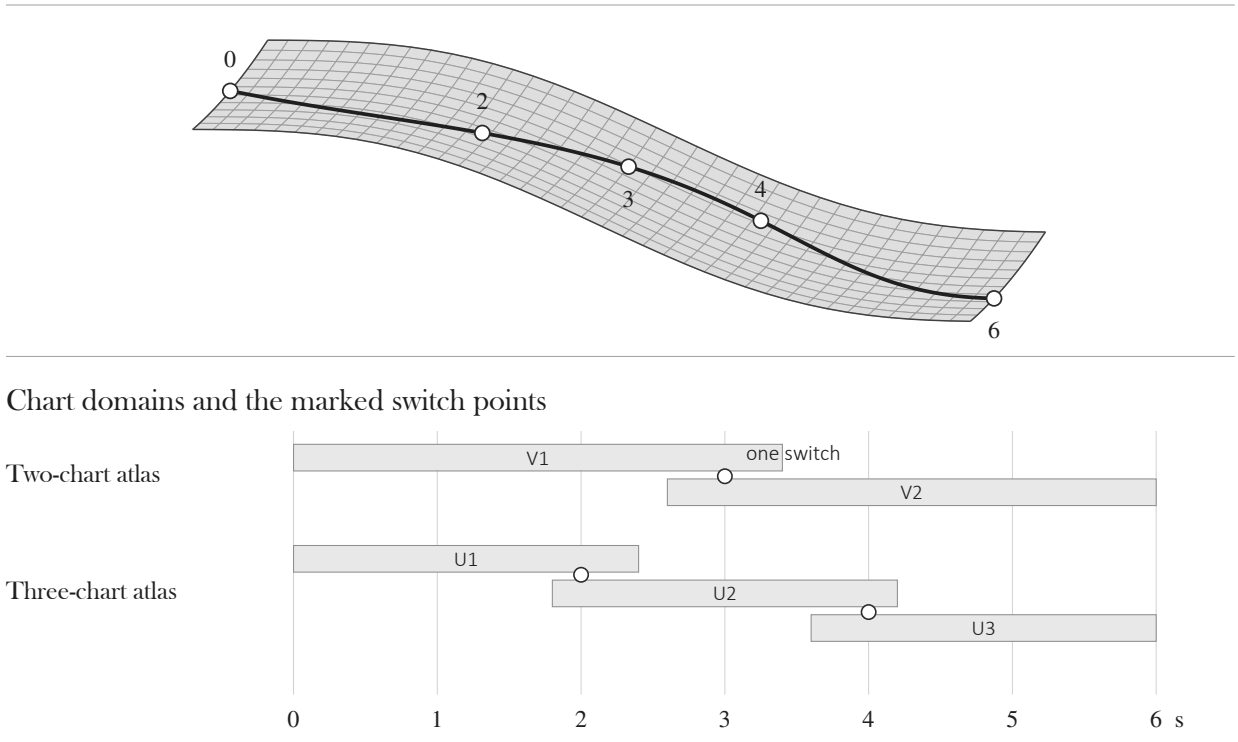


Figure 11.1: One route on a curved strip is measured with two different atlases. The marked chart switches occur at different points. Both assignments give the same total by adding the lengths of the pieces of the same curve.

The exact path construction can now be made into a register of coordinate operations. We keep the point, derivative and uncertainty together as they cross an overlap.

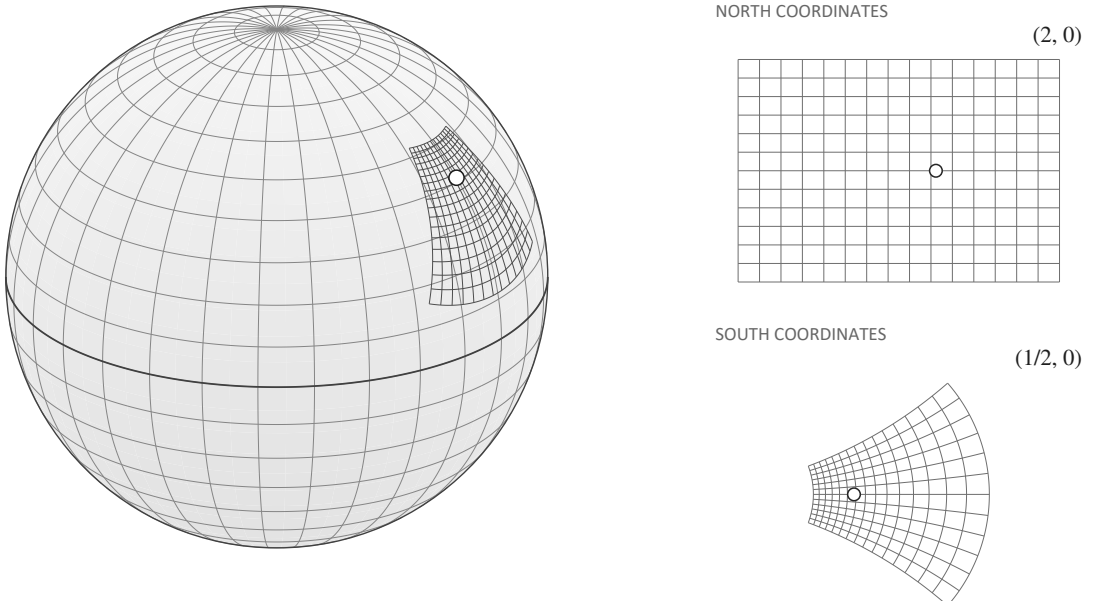
An atlas as an executable translation register

For each chart, store its domain, its coordinate map and its inverse. For each nonempty overlap, store the transition and its derivative. This small register already supports a sequence of concrete checks: points return to the same physical location, round trips return the same coordinates, triple-overlap compositions agree, and metric matrices transform consistently.

Return to the two stereographic maps of a unit globe in Chapter 4. The inverse maps F_N and F_S send planar coordinates back to the sphere. On their overlap T converts between the maps. The point check is $F_S(T(u,v)) = F_N(u,v)$. The round-trip check is $T(T(u,v)) = (u,v)$ for $(u,v) \neq 0$. The derivative check is $DT(T(u,v))DT(u,v) = I$. The metric check is the quadratic-form identity derived in Chapter 4. They test different parts of the construction; none is redundant merely because another passes.

One point survives a change of chart

Match the point, its tangent vectors and its uncertainty.



A One physical route on the globe

B The overlap changes its coordinate description

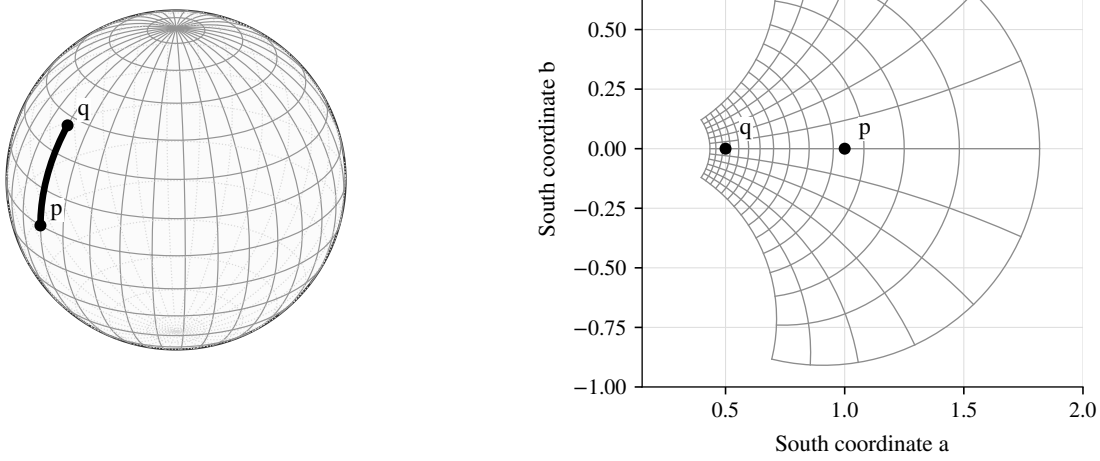
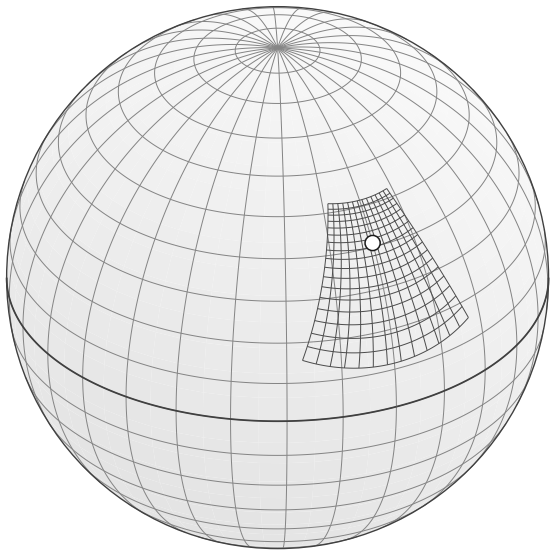
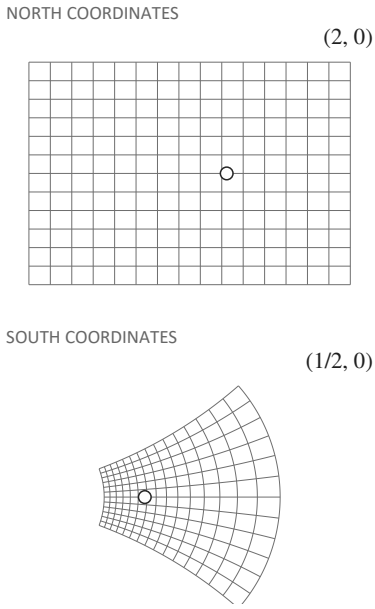


Figure 11.2: *Four checks turn an atlas into a usable translation system.* The study continues on the following figure page.

One point survives a change of chart

Match the point, its tangent vectors and its uncertainty.





Same material patch / inspect the corresponding coordinates and rulers below.

C Points and derivatives must transform together

D Weighted uncertainty agrees in both charts

$FS(T(u,v)) = FN(u,v)$	$J = \mathrm{diag}(-1/4, 1/4)$
$T(T(u,v)) = (u,v)$	$\Sigma N = \sigma^2 I \rightarrow \Sigma S = \sigma^2 I/16$
$DT(T(u,v))\, DT(u,v) = I$	$\mathrm{tr}(gN\, \Sigma N) = 8\sigma^2/25$
Exclude $(u,v) = (0,0)$ from the overlap.	$\mathrm{tr}(gS\, \Sigma S) = 8\sigma^2/25$

Figure 11.2: *Four checks turn an atlas into a usable translation system (continued).* A spherical segment and its overlap coordinates accompany the point, inverse and derivative identities. At north coordinates $(2, 0)$, the covariance and metric calculations give the same weighted trace in both charts. These are exact model calculations, not sampled residual measurements.

Uncertainty crosses the overlap with the point

Let a small coordinate error have covariance Σ_x. To first order, after $y = T(x)$ its covariance is

$$\Sigma_y = DT_x \Sigma_x DT_x^T.$$

At $(u, v) = (2, 0)$ the stereographic transition derivative is $\mathrm{diag}(-1/4, 1/4)$. An isotropic covariance $\sigma^2 I$ becomes $\sigma^2 I/16$. Switch from the north map to the south map and the little patch around our mark looks smaller. To find out how much ground that patch covers, read it with the ruler on its own map.

Multiplying the transformed covariance by the corresponding metric recovers invariant local summaries such as $\mathrm{tr}(g\Sigma)$ under the linear transformation. With $g_N = 4I/25$ and $g_S = 64I/25$, both traces equal $8\sigma^2/25$.

If the error patch covers a wide part of the globe, the scale changes across the patch itself. The conversion can move its mean and reshape it. For that case, compare the local calculation with converted samples or higher-order estimates over the whole patch.

From Geometry to Translators: The Functor G_S

The construction just performed has a formal name: G_S. It takes a geometry-based realization and constructs a canonical translator-based realization from it. Naming the operation lets us ask whether a map between geometric models induces a corresponding map between their translator systems.

What G_S does on objects

Let

$$G = (M, g, \mathcal{R}_M, \pi_M)$$

be a geometry-based realization of S on a connected M. The exact construction uses a full atlas of M and every finite piecewise smooth route through its charts. The operating region tells us which endpoints we compare. To build a translator-based realization $G_S(G)$, we proceed in three conceptual steps:

1. Choose an atlas covering all of M.
2. Treat each chart as a feature space V_i with an induced metric.
3. Define translators T_{ij} on overlaps using the chart-transition maps.

More concretely:

- Select a finite or countable atlas (U_i, φ_i) with $M = \bigcup_i U_i$ and each
$$\varphi_i : U_i \to \mathbb{R}^d$$
is a smooth coordinate map.
- For each i, define a feature space $V_i = \varphi_i(U_i) \subseteq \mathbb{R}^d$ with metric
$$d_i(v, w) = d_M(\varphi_i^{-1}(v), \varphi_i^{-1}(w))$$
induced from the geodesic distance on M.
- On overlaps $U_i \cap U_j \neq \emptyset$, define the translator
$$T_{ij} = \varphi_j \circ \varphi_i^{-1} : \varphi_i(U_i \cap U_j) \to \varphi_j(U_i \cap U_j).$$

The atlas path theorem above supplies the reference for these data. With the same transported local rulers and full corresponding path family, the quotient recovers the geometric distance exactly. If an implementation keeps only a connected open subregion of M and all paths within it, the theorem recovers the intrinsic metric of that subregion. A further route restriction needs its own comparison with the full distance.

This construction is *canonical* in the sense that it depends only on the choice of atlas on M, not on arbitrary design choices. Different atlases will yield translator systems that are related by morphisms in $\mathsf{SamTrans}(S)$.

We therefore define:

Definition 11.2 (Action of G_S on objects)**.** Given a geometry-based realization $G = (M, g, \mathcal{R}_M, \pi_M)$ of S, define $G_S(G)$ to be a translator-based realization

$$T_G = (\{(V_i, d_i)\}, \{T_{ij}\}, X_{T_G}, d_{T_G}, \mathcal{R}_{T_G}, \pi_{T_G})$$

where:

- The spaces V_i, metrics d_i, and translators T_{ij} are constructed from charts as above.
- (X_{T_G}, d_{T_G}) is the translator metric quotient of this system.
- $\mathcal{R}_{T_G}$ is the image of $\mathcal{R}_M$ under the embedding of M into X_{T_G}.
- π_{T_G} is defined so that the diagram

  $$S \xrightarrow{\pi_M} \mathcal{R}_M \hookrightarrow \mathcal{R}_{T_G}$$

  commutes exactly with the induced assignment $\pi_{T_G} = Q_G^{-1} \circ \pi_M$.

In effect, G_S takes the geometer's manifold and builds a hypothetical "ideal translator system" whose feature spaces are chart coordinates and whose translators are exactly the chart-transition maps.

What G_S does on morphisms

Now let $h : G \to G'$ be a morphism in $\mathsf{SamGeom}(S)$, i.e., a smooth, approximately bi-Lipschitz map between in-regime regions that respects semantics.

The chart constructions above embed $\mathcal{R}_M$ into a translator metric space X_{T_G} and $\mathcal{R}_{M'}$ into $X_{T_{G'}}$. On the in-regime regions, we get an induced map

$$f = (\text{chart embedding for } G') \circ h \circ (\text{chart embedding for } G)^{-1}$$

between $\mathcal{R}_{T_G}$ and $\mathcal{R}_{T_{G'}}$, which is again approximately bi-Lipschitz and semantically compatible.

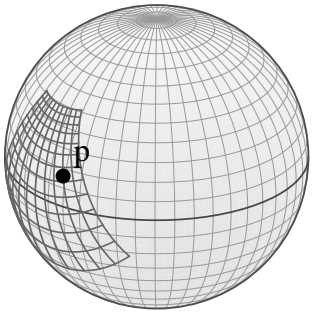
We then define:

Definition 11.3 (Action of G_S on morphisms)**.** For a morphism $h : G \to G'$ in $\mathsf{SamGeom}(S)$, define

$$G_S(h) : \mathcal{R}_{T_G} \to \mathcal{R}_{T_{G'}},$$
$$G_S(h) = Q_{G'}^{-1} \circ h \circ Q_G,$$

where $Q_G : X_{T_G} \to M$ and $Q_{G'} : X_{T_{G'}} \to M'$ are the canonical isometries of Theorem 11.1. The domains are restricted to the stated regions. The distance distortion is exactly that of h on the corresponding points.

A *One point belongs to both charts*



B *The point and ruler transform together*

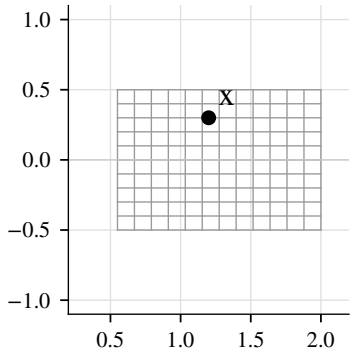
$$x = \varphi_i(p)$$

$$y = \varphi_j(p) = T_{ij}(x)$$

$$T_{ij} = \varphi_j \circ \varphi_i^{-1}$$

$$G_i = (DT_{ij})^T G_j DT_{ij}$$

C *The first chart uses a rectangular grid*



D *The second chart bends that grid*

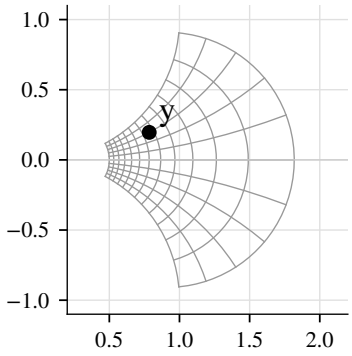


Figure 11.3: An explicit sphere patch, two coordinate grids and their common marked point. The chart transition gives a translator, while its derivative transports the metric. Together they preserve the measured length of corresponding local curves.

This exact chosen-atlas construction respects identities and composition as maps. The reverse construction, F_S, is the next chapter's task; approximate reconstruction will require its separately stated composition bounds.

Starting with the geometric object has supplied exact conversions and preserved its ruler across overlaps. Starting with the conversions alone is harder. The next chapter identifies the gluing, topology and metric conditions needed to reconstruct an object from those descriptions.

Practice with the construction

11.1. Cover a circle with overlapping angular intervals. Describe a shortest path that crosses the $0/2\pi$ coordinate seam and show that the seam adds no length.

11.2. Why does the compactness of a curve's parameter interval help the proof even when the manifold itself is not compact?

Chapter 12

From translators to a geometric realization

Suppose a crew brings back overlapping sheets from a survey of a cylindrical tank. Could an engineer reconstruct the tank's surface and measure routes on it from those sheets? We first use exact overlap and length data, then ask what an approximate survey would need.

An exact cylinder from overlapping strips

Unroll one strip of the tank of radius R. Let θ be angle around the tank and z height. The surface metric is

$$ds^2 = R^2 d\theta^2 + dz^2.$$

On overlapping angular strips use $T(\theta, z) = (\theta + 2\pi k, z)$, with the integer fixed on each connected overlap. Its derivative is the identity and its local metric is unchanged. The gluing records that walking once around the tank returns to the same position while increasing the unrolled angular coordinate by 2π.

For two surface points the intrinsic distance is

$$d^2 = \min_{k \in \mathbb{Z}} \{R^2(\theta_2 - \theta_1 + 2\pi k)^2 + (z_2 - z_1)^2\}.$$

Each candidate is the squared length of a straight line on the corresponding unrolled sheet. The minimum chooses the shortest winding class. For $R = 1$, angular difference $3\pi/2$ and height difference 1, the distance is $\sqrt{\pi^2/4 + 1}$, using the shorter quarter-turn route.

Build a globe from exact compatible maps

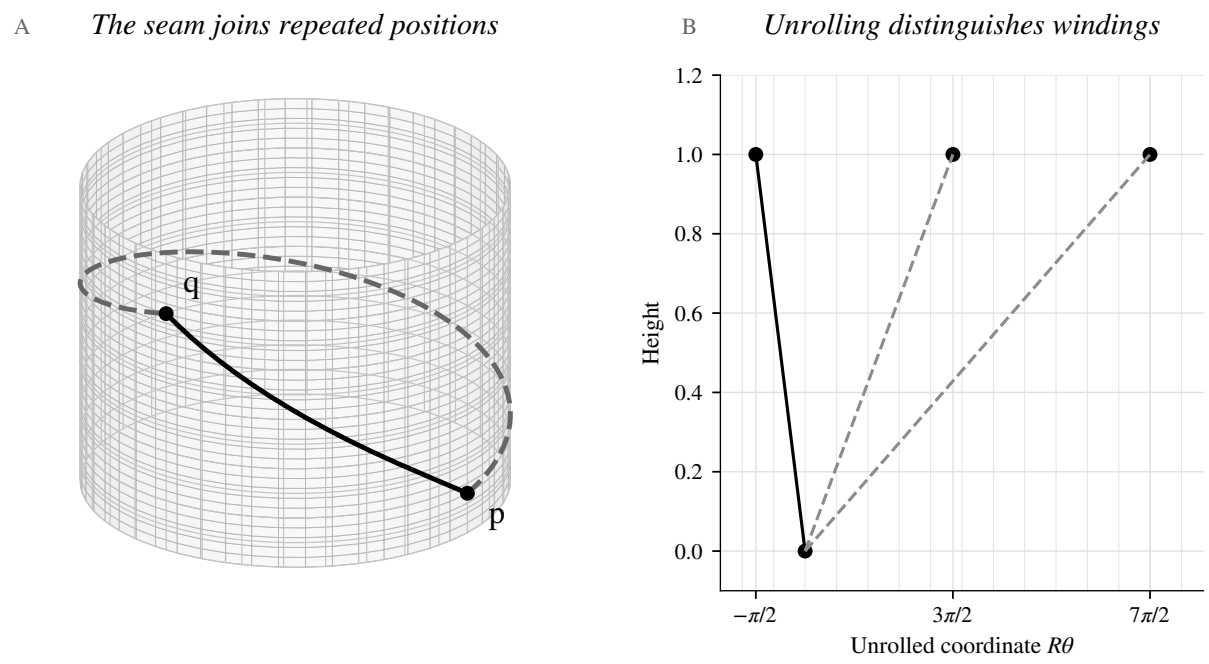
The reverse construction begins with coordinate domains and asks whether they really describe one smooth object. Imagine assembling overlapping survey sheets. Matching the grid locally is necessary, but a crease that identifies unrelated places or a branching junction can prevent the assembly from being a surface. The global gluing conditions belong in the construction.

Take a countable collection of open sets $V_i \subset \mathbb{R}^d$. On open overlap domains supply smooth diffeomorphisms T_{ij} with identity, inverse and cocycle laws. Form the quotient of their disjoint union by the specified identifications. Assume its topology is Hausdorff and that each coordinate domain maps homeomorphically onto an open subset of the quotient, with these images covering it. Under these hypotheses, the inverse coordinate maps give a smooth d-manifold. Where several overlap steps create a chart transition, compose their maps on their common domain. Countability of the domains supplies a countable base.

To recover a particular Riemannian ruler as well, supply smooth positive-definite local matrices G_i satisfying

$$G_i(x) = DT_{ij}(x)^{\mathsf{T}} G_j(T_{ij}(x)) DT_{ij}(x).$$

These matrices define one metric on the glued tangent spaces: different representatives of a tangent vector receive the same squared length by this identity. Smoothness and positive definiteness can be checked in any chart. The translator path theorem then identifies its length distance with the exact chart-path construction.



Winding k	*Angular difference*	*Length (R = 1, Δz = 1)*
−1	−π/2	1.862096 — shortest
0	3π/2	4.817324
1	7π/2	11.040953

Figure 12.1: A cylinder and its unrolled path problem. The same endpoint has several lifts, one for each winding class. The three displayed candidates have distinct lengths; minimizing over all integer windings selects the shorter quarter-turn route in this example.

This is a constructive theorem under explicit hypotheses. A collection of approximate transition residuals is a different input. For that input, the existence of a nearby smooth realization and the size of its distortion must be supplied by a separate construction, theorem or modeling assumption. The practical diagnostics guide that work; they are not a replacement for its global conditions.

The cylindrical seam worked because the identification was specified consistently. We now inspect the assumptions that let other local descriptions assemble into one space, including the ways an apparently reasonable gluing can fail.

Assembling a manifold from compatible local reports

Imagine a team surveying a curved tank. Each technician measures one patch, assigns two coordinates, and records how to convert coordinates where two patches overlap. The collection of files should describe one surface. To reconstruct that surface, we need to say which local records denote the same point, ensure that those identifications are consistent, and check that the assembled space has the topological properties required of a manifold.

The tank is a useful guide, but the following construction applies to any dimension. Let U_i be open subsets of $\mathbb{R}^n$, with overlap domains $U_{ij} \subset U_i$ and smooth diffeomorphisms

$$T_{ji} : U_{ij} \longrightarrow U_{ji}.$$

The subscripts mean that T_{ji} converts coordinates from report i to report j. Require the identity and inverse rules

$$T_{ii} = \mathrm{id}, \qquad T_{ij} = T_{ji}^{-1},$$

on their declared domains. On every legitimate triple overlap, require

$$T_{ki} = T_{kj} \circ T_{ji}.$$

The domain qualifier matters: the intermediate point must belong to the second transition's domain. An equation that composes unavailable reports is not an overlap condition.

Identify records, not nearby numbers

Form the disjoint union $\coprod_i U_i$. A point here is a pair (i, x) containing both its chart label and its coordinate. Declare $(i, x) \sim (j, y)$ when the overlap data identify $y = T_{ji}(x)$, using the equivalence relation generated by these identifications. Let

$$M = \left(\coprod_i U_i \right) / \sim .$$

Write $j_i : U_i \to M$ for the map taking a coordinate to its equivalence class. The charts should remain injective under the identifications: two different points within one local report must not become the same global point merely through an unintended chain of overlaps. This property is supplied by coherent atlas data; it must be checked when assembling a proposed system from arbitrary transformations.

We also require the quotient to be Hausdorff and second countable, and the chart images to be open with j_i homeomorphisms onto their images. The atlas transition $j_j^{-1} j_i$ agrees with each supplied T_{ji} on its declared domain. If a chain of identifications creates an overlap, its transition there is the composition of the maps in that chain. The domains are open, so these compositions are smooth locally. Injectivity of the chart maps makes the result independent of the chain. Thus $\{j_i^{-1}\}$ defines a smooth atlas. A list intended to name every overlap must include these induced domains. The conditions are constructive information: they say what the patch files and their identifications must establish.

The quotient identifies equal points, not points within a numerical tolerance. Replacing exact identification by "closer than ε" does not generally define an equivalence relation because that relation need not be transitive. Taking its transitive closure can instead merge a whole chain whose endpoints are far apart. An approximate atlas needs an explicit model of approximation, not an unqualified tolerance substitution in the exact construction.

A small failure explains the Hausdorff requirement

Take two copies of the real line and identify their points at every nonzero coordinate, while leaving their origins distinct. Each origin has an ordinary-looking interval as a local neighborhood. Yet every neighborhood of one origin intersects every neighborhood of the other through identified nonzero points. The origins cannot be separated by disjoint neighborhoods.

The resulting line with two origins is not Hausdorff. Smooth overlap formulas away from zero did not prevent this global defect. This example explains why local differentiability alone cannot

certify a global manifold realization. A single underlying physical point should not acquire two inseparable identities because the patch assembly omitted a needed identification.

Glue the rulers as well as the coordinates

Suppose chart i carries a smooth positive-definite matrix field $g_i(x)$. For these matrices to describe one Riemannian metric, require

$$g_i(x) = DT_{ji}(x)^T \, g_j(T_{ji}(x)) \, DT_{ji}(x).$$

For a tangent vector with chart-i components v, the right side gives the squared length of its chart-j components $DT_{ji}v$. The equality therefore states exactly that the two reports assign the same length to the same infinitesimal displacement.

Define a bilinear form at the assembled point $p = j_i(x)$ by

$$g_p(dj_i v, dj_i w) = v^T g_i(x) w.$$

If another chart is used, the compatibility equation gives the same value. Thus the definition is independent of the chosen report. Smoothness and positive definiteness hold because they hold in each chart. This proves that the local rulers define a global Riemannian metric.

The coordinate identifications alone would not select this metric. Many different compatible metrics can live on the same smooth manifold. Conversely, a collection of positive matrices that fails the transition rule supplies conflicting rulers, even if the point identifications are perfect. The geometry of sameness needs both parts of the construction.

Prove equality of the two path lengths

Take a piecewise smooth curve $\gamma : [0,1] \to M$. Compactness of its parameter interval lets us subdivide it into finitely many pieces, each contained in a chart. On a piece described by $x_i(t) = j_i^{-1}(\gamma(t))$, its Riemannian length is

$$L_i = \int \sqrt{\dot{x}_i(t)^T g_i(x_i(t)) \dot{x}_i(t)} \, dt.$$

If we change charts on an overlap, the chain rule gives $\dot{x}_j = DT_{ji}\dot{x}_i$. The metric compatibility equation makes the integrand unchanged. Summing the pieces therefore produces a well-defined length independent of the subdivision and chart choices.

Now define a translator route to be exactly such a finite list of local curve pieces, connected by the stated overlap identifications. Its cost is the sum of those same integrals. Every global curve produces a translator route of equal cost, and every compatible translator route assembles into a global piecewise smooth curve of equal cost. Taking infima over corresponding routes proves equality of the two intrinsic distances.

That last sentence contains the central mathematical mechanism. The distance equality follows because the path families correspond and each path has the same length. If an implementation discards available overlap routes, adds shortcuts, or charges a nonzero cost merely to change charts, its computed distance solves a different problem. Such choices may be operationally useful, but their comparison with the intrinsic distance needs its own bound.

Separate consistency from transport around the surface

An exact coordinate round trip on a triple overlap is the identity. A vector parallel-transported around a physical loop can nevertheless return rotated when the connection has holonomy. These

are different operations. One changes labels for the same vector at the same point; the other compares vectors at successive points using a specified transport rule.

On the sphere, stereographic chart transitions satisfy their exact overlap identities. The tangent-sled experiment still accumulates a return angle. There is no contradiction: chart consistency did not claim that the sphere's connection is flat.

For a learned translator, a nonzero coordinate-loop residual can indicate approximation error or incompatible identifications. Calling every such residual curvature would conceal which construction was tested. To relate it to a connection curvature, one must specify the connection, the loop, the limiting process, and how the discrete maps approximate transport.

A tiny fit error does not bound derivatives

Let a proposed one-dimensional translator be

$$T_\varepsilon(x) = x + \varepsilon \sin(x/\varepsilon^2).$$

Its pointwise discrepancy from the identity is at most ε. Its derivative is

$$T_\varepsilon'(x) = 1 + \varepsilon^{-1} \cos(x/\varepsilon^2).$$

For small ε, that derivative varies widely and can change sign. An excellent uniform fit in coordinate values therefore does not imply a well-conditioned local coordinate transformation. The derivative or metric distortion must be controlled separately.

This is why a serious translator audit measures more than reconstruction loss. The overlap residual tests one claim, the Jacobian tests another, and the global assembly tests another. Together these checks show where the descriptions agree and which condition fails when they diverge.

Assemble and measure. Apply this construction to a radius-R circle with angular coordinates in $(-\pi, \pi)$ and $(0, 2\pi)$, as used in Chapter 13. Write both components of their overlap, verify the metric compatibility equation, and recover the shortest arc distance from the translator path infimum. Then charge a fixed positive cost for each chart switch and exhibit a route whose computed cost depends on the chosen atlas.

The realization hypothesis

The exact construction just given establishes a manifold from compatible charts under stated topological and metric conditions. For approximate learned translators, a realization is a separate object to be constructed or assumed and then tested. Smooth local residuals alone are not a general manifold-existence theorem.

Manifold learning offers useful numerical constructions. Isomap estimates geodesic relationships through neighborhood graphs and embeds their distances; it does not certify arbitrary learned translator gluing. Its original development is due to Tenenbaum, de Silva and Langford [6]. The useful common question is whether local measurements support a faithful global model, and on which scales.

Riemannian realizations (assumption)

For the tank, agreement on overlapping strips addresses where the survey marks fit. Realizability asks for a smooth surface that also reproduces distances measured from the assembled survey.

Why look for this structure? A surveyor records lengths and angles on overlapping map sheets. If their positive-definite metric tensors transform by the pullback rule on an exact manifold atlas,

they define one Riemannian ruler g. Smooth transitions alone do not select or reconcile the rulers. After that compatibility is established, the path comparison asks whether integrating g and computing translator-path costs measure the same journeys.

Riemannian realizability assumes the answer is yes, up to small controlled error.

Assumption (Riemannian realizability). Let (X_T, d_T) be the translator metric quotient of a translator system, and let $\mathcal{R} \subseteq X_T$ be an in-regime region on which the system is smoothly chartable with small constants $\varepsilon_{\text{loc}}, \varepsilon_{\text{rt}}, \varepsilon_{\text{cyc}}$ and bounded Jacobians.

We say that $\mathcal{R}$ is *Riemannian realizable* if there exists:

- A d-dimensional Riemannian manifold (M, g), and
- A map $\Phi : \mathcal{R} \to M$

such that:

1. **Bi-Lipschitz up to small slack:** There exist constants $\alpha_{\min}, \alpha_{\max} > 0$ and $\varepsilon_{\text{geom}} \geq 0$ with
$$d_M(\Phi(x), \Phi(y)) \geq \alpha_{\min} d_T(x, y) - \varepsilon_{\text{geom}},$$
$$d_M(\Phi(x), \Phi(y)) \leq \alpha_{\max} d_T(x, y) + \varepsilon_{\text{geom}}.$$
for all $x, y \in \mathcal{R}$ in the scales of interest.
2. **Compatibility with charts:** On each chart image $\iota_i(U_i) \subseteq \mathcal{R}$, the composition $\Phi \circ \iota_i$ behaves like a smooth coordinate map in the atlas of M, with derivatives compatible (up to small slack) with the Jacobians of the translators.

The additive allowance $\varepsilon_{\text{geom}}$ measures error in this distance comparison. Each equivalence class is an exact point of the quotient, even if it contains many original descriptions. For distances at least $r > 0$, the inequality gives lower and upper factors $\alpha_{\min} - \varepsilon_{\text{geom}}/r$ and $\alpha_{\max} + \varepsilon_{\text{geom}}/r$. A positive lower factor separates those pairs; injectivity at smaller scales does not follow.

The constants $\alpha_{\min}, \alpha_{\max}$ and $\varepsilon_{\text{geom}}$ quantify how close Φ is to an isometry between the translator metric and the manifold metric. The compatibility condition ensures that the Riemannian structure g is not arbitrary: it is tied to the way distances look in the feature spaces and translators.

Interpretation. Realizing a general translator metric as a Riemannian space requires a modeling assumption.

> On the chosen operating region, the model supplies a manifold M and a map Φ with the stated bounds on distance distortion and chart compatibility.

For the illustrative identity system that converts among images, video and logs, Riemannian realizability would assert that there exists an "identity manifold" M and a mapping from the region of translator metric space we care about into M such that:

- The distance between two identities in X_T is, up to small scaling and slack, the geodesic distance between the corresponding points in M.
- The chart structure on M induced by Φ matches, up to small errors, the encoders and translators we already have.

Comparing two realizations

If Φ_1 and Φ_2 are injective on the same operating region, their images can be compared through $\Phi_2\Phi_1^{-1}$. Substitute their distance inequalities to derive its multiplicative and additive bounds. Without injectivity, choose reconstruction maps and bound their endpoint errors instead. The existence of two low-distortion descriptions alone does not supply an inverse map. For decisions, the resulting distance bound is combined with the threshold-margin and population boundary terms of Chapter 20.

What an approximate realization must state

For a general translator quotient (X_T, d_T), specify a region $\mathcal{R}$ and a map $\Phi : \mathcal{R} \to M$ with constants $a, b > 0$ and $\eta \geq 0$ such that

$$a\, d_T(x, y) - \eta \leq d_g(\Phi x, \Phi y) \leq b\, d_T(x, y) + \eta.$$

This bounds observed pair distances on the stated region. If $\eta > 0$, it does not establish injectivity: distinct points at distance at most η/a might have the same image. Writing Φ^{-1} therefore needs a further injectivity assumption, or an explicit approximate reconstruction map with its own error bound. For smooth conjugation between geometric realizations, require the chosen maps to be diffeomorphisms onto their specified images.

The tank used the angular coordinate θ. Rebuild the same seam using the length coordinate $s = R\theta$. This will verify the metric conversion directly, then let us test why a seam with a fixed point requires a different local model.

Make a cylinder from a sheet and a seam rule

Take a strip of paper with horizontal coordinate $s \in \mathbb{R}$ and height z. Declare (s, z) equivalent to $(s + 2\pi Rk, z)$ for every integer k. The quotient is a cylinder of radius R. Its realization in ordinary space is

$$F(s, z) = (R\cos(s/R), R\sin(s/R), z).$$

Differentiating gives orthonormal coordinate directions, so the induced metric is $ds^2 + dz^2$. The sheet is locally a faithful ruler. Its seam rule carries the global information.

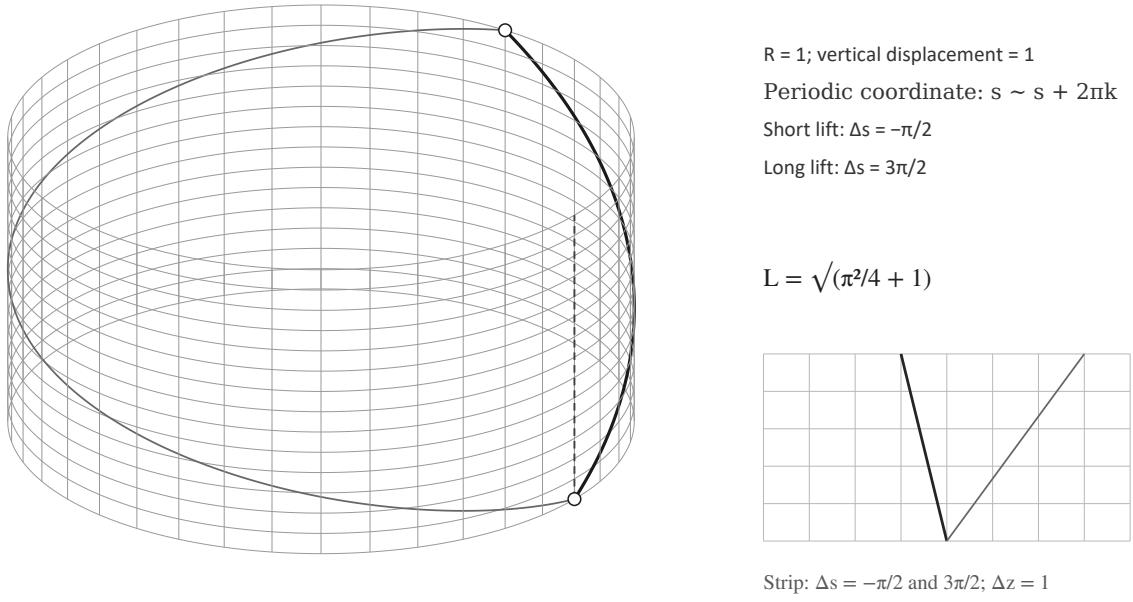
For two cylinder points, choose lifted coordinates (s_1, z_1) and (s_2, z_2). Their shortest intrinsic distance is

$$d = \min_{k \in \mathbb{Z}} \sqrt{(s_2 - s_1 + 2\pi Rk)^2 + (z_2 - z_1)^2}.$$

I'm running a tape along the tank wall between two marks. On the sheet I choose the closest copy of the far mark. If I'm reconstructing a worker's route instead, I need to know how many times it wound around the tank.

The seam changes coordinates, not the track

A periodic position and its unwrapped travel history answer different questions.



A The seam identifies repeated positions

B Unrolling separates routes by winding

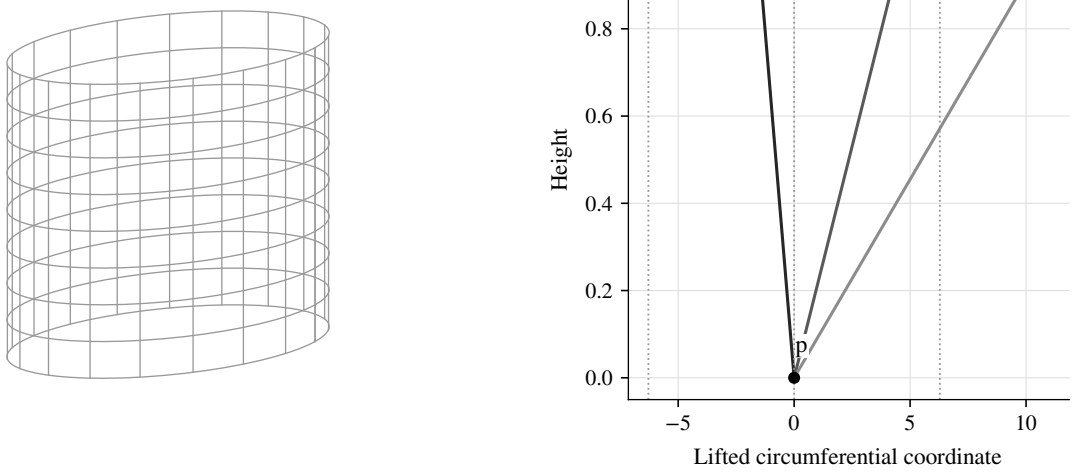
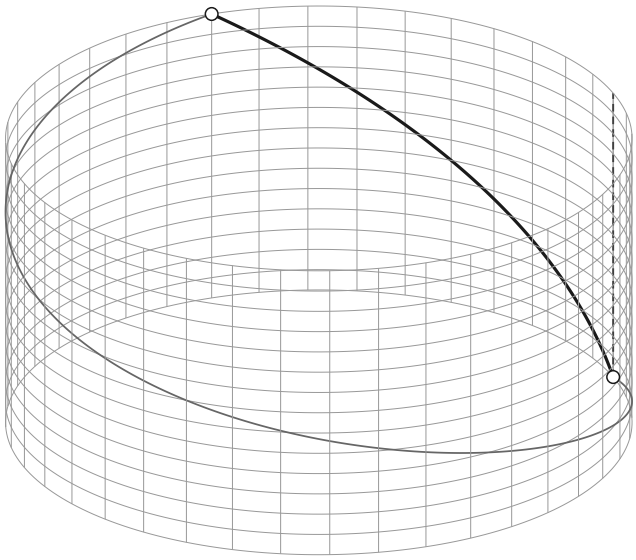


Figure 12.2: *The seam rule creates the global geometry.* The study continues on the following figure page.

The seam changes coordinates, not the track

A periodic position and its unwrapped travel history answer different questions.



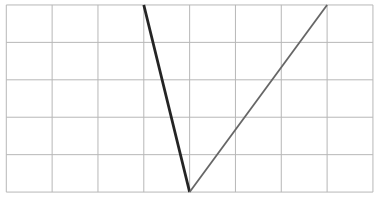
R = 1; vertical displacement = 1

Periodic coordinate: $s \sim s + 2\pi k$

Short lift: $\Delta s = -\pi/2$

Long lift: $\Delta s = 3\pi/2$

$L = \sqrt{(\pi^2/4 + 1)}$



Strip: $\Delta s = -\pi/2$ and $3\pi/2$; $\Delta z = 1$

C The shortest lift selects the distance

$\Delta s = 3\pi/2$
Shortest lift uses k=-1
Length = $\sqrt{(\pi^2/4+1)} = 1.862096$
Other lifts retain extra winding.

D Translation closes the seam, not the history

Seam: $s \sim s+2\pi Rk$
Metric: $ds^2 + dz^2$
A free translation action
Local flatness does not remove the seam.

Figure 12.2: *The seam rule creates the global geometry (continued).* The cylindrical grid and unrolled strip show the geometry of a periodic seam. Three lifted straight routes reach different representatives of the target. The register evaluates the shortest choice for $R = 1$ and a unit height change; the other lifts retain additional winding.

A finite graph cannot supply smoothness by itself

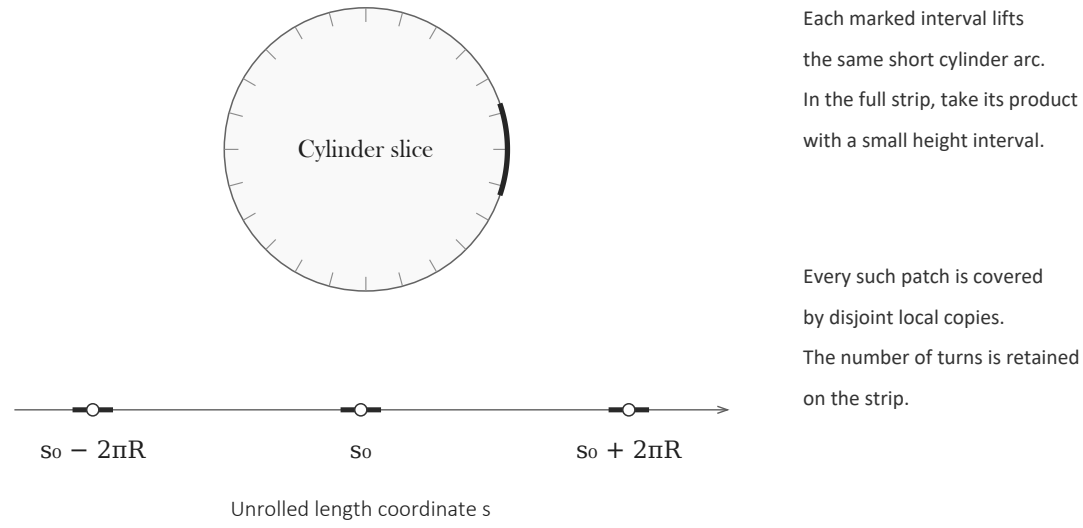
A graph with edge costs supplies a discrete path metric. It does not automatically supply tangent spaces, differentiable transition maps or a smooth Riemannian tensor. One may construct a realization by adding that structure, or show that a sequence of sampled graphs approximates a specified manifold. Those are mathematical steps, not consequences of naming a graph a manifold.

The cylinder succeeds because its construction supplies a smooth free translation action, compatible local coordinates and an invariant metric. A quotient with fixed points can instead produce singular behavior. For example, identifying (x, y) with $(-x, -y)$ fixes the origin. Away from the origin the quotient is locally covered by two separate patches; at the origin the ordinary covering-space picture breaks.

This makes the recurring focus on individual objects sharper. Some descriptions differ by a harmless symmetry. Some collapse an entire direction. Some create singular quotient points. The realization theorem must match the actual construction rather than treating all identifications as the same operation.

A small cylinder patch has separate lifts

Translation by 2πR has no fixed points. This fixed-height slice shows the repeating lift pattern.



A A cylinder joins freely translated points

B Sign identification fixes the origin

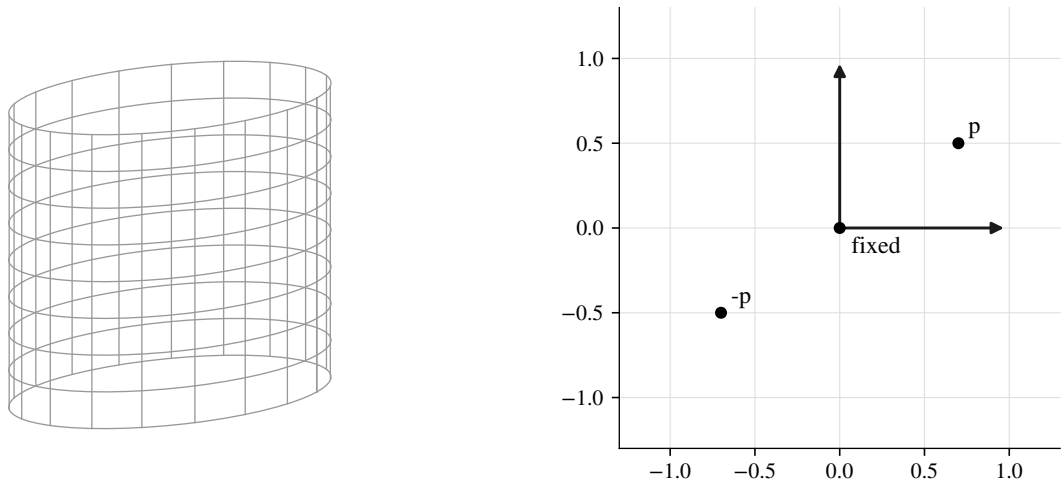
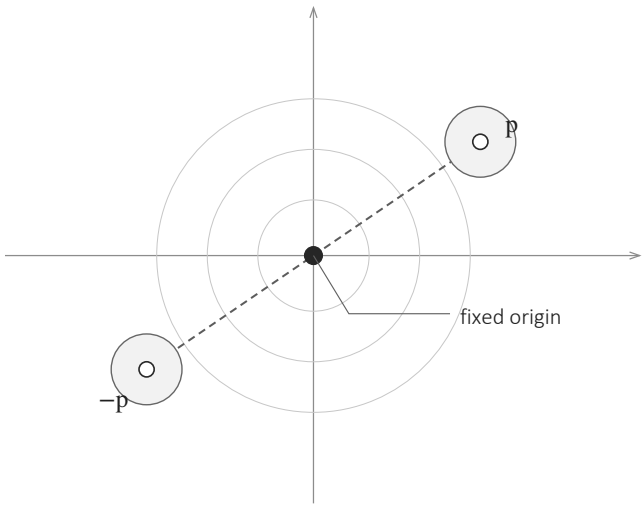


Figure 12.3: *A free seam identification and a fixed-point quotient differ.* The study continues on the following figure page.

Sign identification fixes the origin

The plane identifies p and −p. Every nonzero point has two representatives, while the origin has one.



Distinct neighborhoods of p and −p away from zero

The two small disks project onto the same quotient patch. They do not intersect.

At zero the two representatives coincide. The ordinary local two-sheet pattern fails there.

C Every seam point has an ordinary local lift

Every orbit has no fixed point.
Small patches lift separately.
Smooth local chart transitions
Global winding survives.

D The fixed origin breaks that local pattern

The origin is fixed.
Away from zero: two representatives
At zero: one representative
Ordinary covering neighborhoods fail at the origin.

Figure 12.3: *A free seam identification and a fixed-point quotient differ (continued).* A cylindrical translation seam is compared with the sign identification of the plane. The plane diagram marks its fixed origin and a pair of opposite representatives. The accompanying registers explain why the free seam has ordinary local lifts and the fixed point does not.

From Translators to Geometry: The Functor F_S

The tank supplied an assembled surface and a ruler. We now record that choice as a geometric realization of the translator system. This lets a change of implementation be followed through the geometric model as well. The operation is called F_S because it acts on the systems and on the maps relating them.

The clinical examples below are stipulated model assignments used to trace that operation. They reuse one proposed panel upgrade as a correspondence between specified model images. The numerical marker predictions and distances are chosen inputs. The cylinder supplies the exact reference for the construction; the panel example tests what additional assumptions its application would need.

We now compare a translator realization with a chosen geometric realization under the chart, metric and domain assumptions of Chapter 10.

What F_S does on objects

Input: A translator-based realization T. Take an object

$$T = \big(\{(V_i, d_i)\}, \{T_{ij}\}, X_T, d_T, \mathcal{R}_T, \pi_T\big)$$

in $\mathsf{SamTrans}(S)$, assumed to satisfy smooth chartability and Riemannian realizability on its in-regime region $\mathcal{R}_T$.

The tuple records the feature-space rulers and converters, the quotient they define, its operating region $\mathcal{R}_T$, and the assignment π_T connecting the model's entities to their representations. In the stipulated panel model those channels are blood, imaging, symptoms and genetic records. Each component has the same role as its counterpart in the tank construction, local reports, a common space, and a map specifying what each point represents.

The assumption: T satisfies:

- **Chart data** (Chapter 10): Designated regions of the feature spaces supply coordinates, with stated local metric bounds and transition regularity. Round-trip and cocycle residuals are reported with their domains and units. Numerical cutoffs depend on the task's resolution.
- **Riemannian realizability**: Supply a manifold, a realization map and its distance bounds. The exact construction uses open chart images, smooth compatible overlaps, and a Hausdorff, second-countable quotient. The positive-definite local metrics must agree on overlaps. Small empirical residuals alone do not establish these conditions.

The chosen realization. For this categorical construction, each object comes with a chosen realization Φ_T that is bijective onto its designated image. Require distance bounds for every pair in the designated region so their inverses and compositions can be used on that region. Give the translator region the smooth structure transported by this map, require the image to be an open manifold region or another specified smooth domain, and use smooth arrows in those structures. Thus Φ_T and its inverse are diffeomorphisms between the chosen regions. Their metric distortion is bounded separately. Exact compatible atlas data provide one construction; approximate data need the additional realization assumptions above. A positive additive distance allowance alone establishes neither injectivity nor smoothness.

Properties of the realization map Φ_T. The map $\Phi_T : \mathcal{R}_T \to M_T$ satisfies two crucial properties:

1. Approximate isometry:

For any two points $x, y \in \mathcal{R}_T$ (in translator space), the distance between them measured via d_T (shortest translator path) approximately equals the distance between their images $\Phi_T(x), \Phi_T(y)$ measured via d_{M_T} (geodesic distance on the manifold):

$$d_T(x, y) \approx d_{M_T}(\Phi_T(x), \Phi_T(y)) \pm \varepsilon_\Phi,$$

Here the shorthand means $|d_{M_T}(\Phi_T(x), \Phi_T(y)) - d_T(x, y)| \leq \varepsilon_\Phi$ on the specified pairs, with an established allowance in distance units. A general multiplicative-and-additive comparison is handled by Chapter 20; no universal value of ε_Φ follows from chartability alone.

What "approximate isometry" means: An isometry is a distance-preserving map—distances don't change at all. An *approximate* isometry preserves distances up to small bounded error.

In the survey, compare the route distance calculated from the strips with the surface distance between the corresponding marks. Their difference is bounded by ε_Φ under this allowance.

Concrete example:

- In translator space X_T, active Crohn's disease and Crohn's in remission are distance $d_T = 0.4$ apart (measured by shortest translator path through blood/imaging/symptom spaces)
- After mapping to manifold M_T via Φ_T, the geodesic distance on the manifold is $d_{M_T} \approx 0.38$ to 0.42 (within ± 0.02 of the translator distance)
- The manifold geometry faithfully represents the translator system's distance structure

2. Chart compatibility:

Φ_T is compatible with the chart structure arising from feature spaces V_i and translators T_{ij}.

Informally: if you use feature space V_{blood} as a coordinate system for part of $\mathcal{R}_T$, and you map that region to M_T via Φ_T, the result should be a smooth coordinate chart on M_T. The manifold's charts are the images of the translator's feature spaces, and transitions between charts correspond to translators.

On each chosen open chart region U_i, require $\Phi_T \circ \iota_i|_{U_i}$ to be a smooth parametrization with a smooth inverse onto its open image. Its inverse is the coordinate chart in the usual manifold convention. Exact transition maps agree on overlaps; learned translators are compared with those maps using the declared residual bounds.

On the tank, a strip gives coordinates for an open patch of the surface. Where two patches cover the same painted mark, their coordinates must convert according to the transition on that overlap.

Output: A geometry-based realization G_T. We can now define a geometry-based realization:

$$F_S(T) = G_T = (M_T, g_T, \mathcal{R}_{M_T}, \pi_{M_T}),$$

where:

- (M_T, g_T): The Riemannian manifold and its metric tensor (as constructed above)
- $\mathcal{R}_{M_T} = \Phi_T(\mathcal{R}_T)$: The in-regime region on the manifold, defined as the image of the translator system's in-regime region under Φ_T.
  Why this makes sense: If $\mathcal{R}_T$ is the region of translator space visited by actual patient data, then $\mathcal{R}_{M_T} = \Phi_T(\mathcal{R}_T)$ is the corresponding region on the manifold. Patients who land in $\mathcal{R}_T$ (translator space) map to $\mathcal{R}_{M_T}$ (manifold).
  Concrete example: If $\mathcal{R}_T$ contains representations of adult patients (ages 18-75) with common inflammatory conditions, then $\mathcal{R}_{M_T}$ is the corresponding region of the disease manifold M_T—the part of the manifold representing those same patients and conditions.

- $\pi_{M_T} = \Phi_T \circ \pi_T$: The realization map from the abstract semantic structure S to the manifold M_T, defined by composing:
  - $\pi_T : S \to X_T$ (maps disease states to translator space)
  - $\Phi_T : X_T \to M_T$ (maps translator space to manifold)

  Why composition makes sense: If disease state u (like "active Crohn's") maps to point $\pi_T(u) \in \mathcal{R}_T$ in translator space, and Φ_T maps translator points to manifold points, then $\pi_{M_T}(u) = \Phi_T(\pi_T(u))$ is where disease u lives on the manifold.
  Concrete example:
  - Active Crohn's disease has representation $\pi_T(\text{Crohn's}) \in \mathcal{R}_T$ in translator space (a reference point around which the model places observations)
  - Applying Φ_T gives the corresponding manifold point:

$$\pi_{M_T}(\text{Crohn's}) = \Phi_T(\pi_T(\text{Crohn's})) \in \mathcal{R}_{M_T}.$$

  - Both representations—translator and geometric—describe the same disease state, just in different coordinates

What F_S accomplishes. The functor F_S sends a translator-based realization T to its *geometric shadow* G_T: the manifold model that captures the same semantic structure in Riemannian language.

Key properties:

1. **Same semantic structure**: Disease states have the same identities in both realizations. $\pi_T(\text{Crohn's})$ in translator space corresponds to $\pi_{M_T}(\text{Crohn's})$ on the manifold.
2. **Approximately same distances**: Distance in translator space d_T matches geodesic distance on manifold d_{M_T} up to small error ε_Φ.
3. **Same in-regime region**: Patients in the translator system's regime ($\mathcal{R}_T$) map to the manifold's regime ($\mathcal{R}_{M_T}$).
4. **Smooth chart structure**: The translator's feature spaces V_i become coordinate charts on the manifold M_T via Φ_T.

The tank construction supplies the exact reference for this inventory. We can now apply it to the chosen diagnostic model.

Medical diagnosis summary:

- Input T: Blood/imaging/symptom translator system with metric quotient X_T
- Construction: A chosen manifold M_T, map Φ_T and verified comparison bounds realize the stipulated disease-state model on the declared region
- Map Φ_T: Connects translator representations to manifold points, preserving distances approximately
- Output $G_T = F_S(T)$: The manifold (M_T, g_T) measures the chosen geometric cost; which paths represent benign disease evolution remains part of S

The geometric realization G_T is not a new system—it's the *same semantic structure* S, expressed in differential-geometric language rather than translator language. The functor F_S formalizes the translation between these two languages.

We now compare a translator realization with a chosen geometric realization under the chart, metric and domain assumptions of Chapter 10.

What F_S does on morphisms

We've seen how F_S transforms a translator system T into its geometric shadow $G_T = F_S(T)$. Now we need to see how F_S transforms *relationships between translator systems* into *relationships between geometric realizations.*

The setup: A morphism between translator systems. Let $f : T \to T'$ be a morphism in $\mathsf{SamTrans}(S)$. In the terminology developed in Chapter 13, f is a map

$$f : \mathcal{R}_T \to \mathcal{R}_{T'}$$

between in-regime regions that:

- **Preserves semantics**: Disease state u in system T (represented by $\pi_T(u)$) maps to the same disease state u in system T' (represented by $\pi_{T'}(u)$). Both systems agree on what "Crohn's disease" means.
- **Controls distances**: The map f is bi-Lipschitz with constants $(L_{\min}, L_{\max}, \varepsilon_f)$:

  $$L_{\min} d_T(x, y) - \varepsilon_f \leq d_{T'}(f(x), f(y)) \leq L_{\max} d_T(x, y) + \varepsilon_f.$$

  Distances in system T and system T' are related by bounded distortion.
- **Is smooth in the chosen coordinates**: This is a hypothesis on f, separate from its distance bound. A bi-Lipschitz map can have a corner, as $x \mapsto x + 0.1|x|$ does at zero. Such a map belongs to a metric comparison but not to the smooth arrow class used here.

Concrete example (lab panel upgrade): System T uses a 2018 lab panel, system T' uses a 2024 comprehensive panel. The morphism $f : T \to T'$ tracks how a patient's representation in the old system maps to their representation in the new system. Active Crohn's disease in T (at point $\pi_T(\text{Crohn's})$) maps to active Crohn's disease in T' (at point $\pi_{T'}(\text{Crohn's})$), with distances controlled by $(L_{\min} = 0.9, L_{\max} = 1.3, \varepsilon_f = 0.05)$.

The challenge: Lifting to geometric morphisms. Each translator system has its own geometric realization via F_S:

- System T has geometric realization $G_T = (M_T, g_T, \mathcal{R}_{M_T}, \pi_{M_T})$ with realizability map $\Phi_T : \mathcal{R}_T \to M_T$
- System T' has the corresponding geometric data $G_{T'}$ and a realization map $\Phi_{T'} : \mathcal{R}_{T'} \to M_{T'}$.

We want to construct a geometric morphism $h : G_T \to G_{T'}$—a map between the manifolds—that corresponds to the translator morphism $f : T \to T'$.

The survey version would compare two assemblies of the same tank. Each has a chosen map to its geometric model. A correspondence between their local reports must therefore determine the corresponding map between the models. The three-map composition below follows one point through that change.

The construction: Composing through translator space. The answer is to route through translator space in three steps:

Definition 12.1 (Action of F_S on morphisms)**.** For a morphism $f : T \to T'$ in $\mathsf{SamTrans}(S)$, define

$$F_S(f) : \mathcal{R}_{M_T} \to \mathcal{R}_{M_{T'}}, \qquad F_S(f) = \Phi_{T'} \circ f \circ \Phi_T^{-1},$$

interpreted on the designated images of the in-regime regions.

All three maps are smooth by the chosen arrow and realization hypotheses, so $F_S(f)$ is a smooth geometric arrow. Exact cancellation of $\Phi_T^{-1}\Phi_T$ proves functorial composition. If only an approximate inverse is available, one must instead track a comparison with reconstruction error; the exact functor identity is not supplied by that approximation.

Follow the actual point. Say I've got a pin on the old tank model and I want to place it on the new one. I look up the old survey reading, run it through the converter, and use the converted reading to place the pin. That's the formula, read right to left. Next we'll try it with a made-up lab upgrade. A smaller set of blood tests gives way to a panel with more markers. We'll follow one hypothetical Crohn's case through the switch. The converter has to estimate what the new panel would report for markers the old one never measured.

Check the complete route. The composition must keep the same model state attached to all three descriptions:

$$p \in M_T \quad \xrightarrow{\Phi_T^{-1}} \quad x \in \mathcal{R}_T \quad \xrightarrow{f} \quad x' \in \mathcal{R}_{T'} \quad \xrightarrow{\Phi_{T'}} \quad p' \in M_{T'}.$$

The endpoint is obtained by applying the supplied maps in the displayed order. To test the construction numerically, keep the hypothetical panel upgrade and follow one stipulated state through each of those maps. This adds a concrete trace to the symbolic identity.

A hypothetical panel upgrade lifted to geometry. The values below are constructed to show the composition, not patient data or a clinically validated inference rule. Additional marker values cannot be recovered from a smaller panel without an explicit predictive model and uncertainty. Here f is such an assumed model; the chosen realizations are invertible only on their specified model images.

Translator morphism $f : T \to T'$:

- System T: 2018 lab panel, translator metric quotient X_T
- System T': 2024 comprehensive panel, translator metric quotient $X_{T'}$
- Morphism f: Maps 2018 blood marker embeddings to 2024 embeddings
- Constants: $L_{\min} = 0.9$, $L_{\max} = 1.3$, $\varepsilon_f = 0.05$

Geometric realizations:

- $G_T = F_S(T)$: Disease manifold M_T (3D curved space, metric g_T)
- $G_{T'} = F_S(T')$: Disease manifold $M_{T'}$ (also 3D, but metric $g_{T'}$ slightly different due to additional markers)
- Realizability maps: $\Phi_T : X_T \to M_T$ and $\Phi_{T'} : X_{T'} \to M_{T'}$

Applying $F_S(f)$:
Consider a disease state: "moderate Crohn's disease, stable for 6 months."

1. **Start - geometric representation in old system**: On manifold M_T, this state is at point $p = (0.8, 1.2, 0.5)$ (some coordinate system on M_T). Geodesic distance from healthy baseline is $d_{M_T}(p, \text{healthy}) = 1.1$.
2. **Pull back to translator space**: $\Phi_T^{-1}(p)$ gives the translator representation in 2018 system:

$$x = \Phi_T^{-1}(p) \in \mathcal{R}_T : \quad \text{CRP} = 15 \text{ mg/L}, \text{ ESR} = 28, \\ \text{WBC} = 9.2, \ldots \text{(12 markers total)}.$$

Plus imaging and symptom embeddings.

3. **Apply translator morphism**: $f(x)$ maps to 2024 system representation:

$$x' = f(x) \in \mathcal{R}_{T'}: \quad \text{12 original markers}, \quad \text{calprotectin} = 250, \quad \text{IL-6} = 8.5, \ldots \text{(35 markers total)}.$$

The morphism f predicts: "given these 2018 lab values, the 2024 panel would show these additional markers."

4. **Push forward to new manifold**: $\Phi_{T'}(x')$ gives geometric representation in new system:

$$p' = \Phi_{T'}(f(\Phi_T^{-1}(p))) \in M_{T'}$$

where $p' = (0.82, 1.25, 0.48)$ are coordinates on $M_{T'}$. Geodesic distance from healthy baseline on new manifold: $d_{M_{T'}}(p', \text{healthy}) = 1.15$.

5. **Comparison**: The disease state "moderate stable Crohn's" is:

   - Distance 1.1 from healthy on old manifold M_T
   - Distance 1.15 from healthy on new manifold $M_{T'}$

   The morphism $F_S(f)$ relates the two: $d_{M_{T'}}(p', \text{healthy}) \approx 1.04 \times d_{M_T}(p, \text{healthy})$ (slight distance increase due to finer resolution from additional markers).

Distortion and composition. The distance bounds for the chosen inverse, morphism and target realization compose by substitution, as worked through in Chapter 13. Inverse bounds require the chosen map to be injective, with a positive lower multiplicative constant; any additive slack transforms with those constants.

For actual inverse maps, $\Phi_T^{-1}\Phi_T = \text{id}$ exactly, regardless of whether Φ_T preserves distances exactly. Thus $F_S(hf) = F_S(h)F_S(f)$ and $F_S(\text{id}) = \text{id}$. Metric approximation does not turn an algebraic inverse into an approximate inverse. If the construction uses a reconstruction map instead, its error must be carried through separately.

What F_S accomplishes. Chapter 10 gave tests for compatible local descriptions. Here the functor $F_S : \mathsf{SamTrans}(S) \to \mathsf{SamGeom}(S)$ uses the geometric realization chosen for each system to carry comparisons and maps into that geometry.

- Every object equipped with the stated realization has its specified geometric counterpart
- Every structure-preserving map between translator systems lifts to a map between their shadows
- The lifting procedure is systematic, not ad hoc
- Distortion is controlled and explicit

Chapter 20 uses the map-distortion bounds of F_S to transfer decision guarantees, including through system upgrades. The calculation retains the geometric failure probability and threshold-band mass on the relevant population.

What the realization makes possible

Smooth chartability and Riemannian realizability are the bridge between the translator builder's metric object and the geometer's manifold. Once we accept these hypotheses (in a particular problem domain, on a particular in-regime region), several things become possible.

From translator systems to geometric realizations

First, under smooth chartability and Riemannian realizability, every translator system that realizes a given semantic sameness structure S and satisfies the chartability conditions on its in-regime region gives rise to a Riemannian manifold (M, g) together with a map Φ that is an approximate isometry on that region.

We can think of this as a *geometric shadow* of the translator system:

- (X_T, d_T) captures how the translator system measures distances and drift.
- (M, g) captures the same geometry in the language of Riemannian manifolds.
- The map Φ aligns the two pictures.

Recall from Chapter 3 the translator builder and the geometer: one who thinks in terms of feature spaces and translators, one who thinks in terms of manifolds and geodesics. Smooth chartability and Riemannian realizability are the conditions under which their two perspectives describe the same object. When these conditions hold:

- The translator builder's graph of spaces and translators becomes an atlas of charts on the geometer's manifold.
- The translator metric d_T becomes an approximation to the geodesic distance d_M.
- Detection thresholds in translator space correspond to detection thresholds in geometric space.

The two characters are not solving different problems—they are using different coordinate systems to describe the same semantic sameness structure.

The operation F_S above packages that construction for objects equipped with the stated realization. Chapter 13 develops the categories and compares the round trips, retaining the difference between exact maps and approximate reconstruction.

Transferring guarantees between representations

Second, once we know that (X_T, d_T) and (M, g) agree on distances up to small, explicit slack, we can begin to *transfer detection guarantees* from one representation to the other:

- A theorem about robustness of a detector that uses geodesic distances in (M, g) can be turned into a theorem about robustness of a detector that uses translator metric distances in (X_T, d_T), with error terms adjusted by the distortion constants in the realizability assumption.
- Conversely, empirical guarantees derived from experiments directly on the translator system can be lifted to the manifold model, giving us a theoretical understanding of how curvature and benign paths behave in practice.

The Error Budget Transfer Theorem in Chapter 20 will make this precise. For now, the key point is that smooth chartability and Riemannian realizability endow distances in X_T and distances in M with a controlled relationship, which supplies the geometric part of decision transfer. The threshold-band probability supplies its population part.

Falsifiability and audits

Each hypothesis has a test on measured data.

- Local near-isometry can be probed by sampling pairs of points in each U_i and comparing d_i to d_T.

- Smoothness and well-conditioned Jacobians of translators can be checked by estimating local derivatives on data and examining their singular values.
- Round-trip error and cocycle slack can be measured directly as in Chapter 9.
- Drift profiles along paths can be estimated by composing translators along typical routes.

These diagnostics can be implemented as automated evaluation checks. Passing supports the hypotheses on the tested records. Extending them to a whole region also requires the coverage and regularity arguments already described. A verified violation identifies a condition the tested system does not satisfy there; the corresponding guarantee cannot be invoked on that basis. Chapter 21 develops the practical audit.

The audit keeps each measured result attached to the regime in which it was established.

The tank reports describe one surface when their identifications fit, distinct points remain separable, and their local rulers agree. Under the exact construction, changing survey sheets preserves the measured route. For approximate reports, the realization must instead supply a region and quantitative bounds. That is the answer to the reconstruction question: the common object and its measurement rule must be built together. We next follow what they preserve through several successive conversions.

Practice with the construction

12.1. Calculate the cylindrical distance for $R = 2$, angular difference $\pi/2$ and height difference 3.

12.2. If $\Phi x = \Phi y$ under the displayed approximate bound, what upper bound follows for $d_T(x, y)$? Why is that insufficient to define an inverse?

Chapter 13

What equivalence preserves

The previous two chapters constructed the passage from a geometric space to translators and the return passage from compatible descriptions. We now ask what survives that round trip, and what happens when several such passages are composed. An engineer moving a comparison between implementations needs those answers before reusing a threshold or a guarantee.

A calibration cart runs on a circular track. One software version stores its angular position, another uses a different seam, and a third approximates the track by stations. Before accepting a replacement we need to know which position comparisons survive, what approximation error it introduces, and whether the task also needs the number of completed laps. We can settle each question on this one track.

Build both descriptions of one circular track

A circular calibration track has circumference $2\pi R$. A cart's state is its position on the track, not the number of laps it has completed. One instrument reports an angular coordinate with a seam opposite the starting gate. Another puts its seam at the gate. We can build the translator and geometric descriptions explicitly, including the point where one coordinate stops being usable.

Use a first chart with $-\pi < \theta < \pi$, omitting the point opposite the gate. Use a second with $0 < \eta < 2\pi$, omitting the gate itself. The overlap has two components. On the upper half of the track, $\eta = \theta$. On the lower half, $\eta = \theta + 2\pi$. In both components the derivative is one.

Both charts carry the metric $R^2 d\theta^2$ in their own coordinate. The transition preserves tangent length exactly. The discontinuity one would create by forcing both rules into one unqualified formula belongs to the seam convention, not to the cart's motion.

Construct distance by translating local routes

An admissible chart route consists of finitely many coordinate segments, with chart switches only where both charts describe the same point. Give each segment length $R \int |\dot{\theta}| dt$. Chart switches cost zero because they change the description without moving the cart.

For two track points represented by angles α and β, every lifted route on the real line ends at $\beta + 2\pi k$ for an integer k. Its length is at least

$$R|\beta - \alpha + 2\pi k|.$$

That lower bound is attained by the monotone lifted segment. Taking the best integer gives

$$d(\alpha, \beta) = R \min_{k \in \mathbb{Z}} |\beta - \alpha + 2\pi k|.$$

This is the ordinary shortest arc distance on the circle. The translator path construction and the Riemannian construction therefore give exactly the same result, with no appeal to a fitted embedding.

Take $\alpha = 170^\circ$ and $\beta = -170^\circ$. Subtracting the two displayed numbers gives 340°. The quotient calculation selects $k = 1$ and returns 20°. At $R = 2$ meters, the shortest route has length $2\pi/9$ meters. The large raw-coordinate jump is a seam crossing.

Retaining the lap count changes the object

If the task needs signed travel or completed turns, the state can include a continuously lifted coordinate $s \in \mathbb{R}$. The projection to the track is $s \mapsto e^{is/R}$. Positions differing by $2\pi R$ project to the same cart location but retain different lap histories.

The real line is an infinite-sheeted cover of the circle. It is not the same covering as the two-sheeted bar construction, and it is not a second coordinate chart on the circle. The fiber over one point contains every integer lap offset. Which state is appropriate depends on what the instrument must report.

For position comparison, shortest arc distance is useful. A signed odometer keeps the net unwrapped displacement. If the cart reverses, its total distance also needs the variation of the lifted path, because forward and backward travel cancel in the endpoint displacement. Sameness is therefore determined by the operational object and its projection, not by the most compact representation available.

A finite graph approximates the same track

Place n equally spaced stations on the track and connect neighboring stations in a cycle. If each edge is assigned the arc length $2\pi R/n$, the graph shortest-path distance between any two stations agrees exactly with the circle's shortest arc distance between those stations.

If the edge is instead assigned its straight chord length $2R\sin(\pi/n)$, the graph distance for a shortest k-edge route is

$$d_n = 2kR\sin(\pi/n), \qquad d = 2kR\pi/n.$$

Their ratio is $\sin(\pi/n)/(\pi/n)$, independent of k for these station pairs. Since $\sin x \geq x - x^3/6$ for the relevant positive range,

$$0 \leq 1 - \frac{d_n}{d} \leq \frac{\pi^2}{6n^2}.$$

The graph construction now gives a controlled approximation rather than an exact identity. At $n = 12$, the relative shortfall is about 1.1384 percent; at $n = 24$, about 0.2853 percent. The error comes from the declared edge-length convention, not from the existence of a seam.

Add observation uncertainty separately

Suppose each endpoint is reported within arc distance e of its true location. For any metric,

$$|d(p,q) - d(\widehat{p},\widehat{q})| \leq d(p,\widehat{p}) + d(q,\widehat{q}) \leq 2e.$$

That endpoint allowance is additive. The chord-versus-arc approximation is multiplicative. Keeping them distinct gives a more useful bound than hiding both inside one unexplained "small error."

Suppose my cart alarm goes off when two reported positions are farther apart than τ. If the whole distance interval is above τ, the alarm is settled. If it straddles the cutoff, I need a better measurement. Changing where the display wraps cannot settle that case.

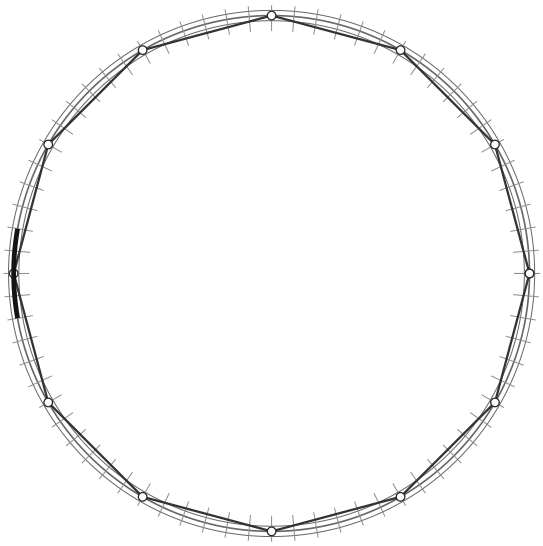
When the two constructions agree

We started with one individual cart position. The atlas supplied two descriptions, the translators specified their overlap, and the metric measured the same physical route in either chart. The path construction recovered the geometric distance. A finite approximation then introduced an explicit, calculable distortion.

The cart supplies a complete equivalence test. A general realization needs the same domain, compatibility and path conditions. A small cycle residual checks only part of that requirement. The construction provides a reference for measuring the distortion of an approximate implementation.

The arc follows the track; the chord cuts across it

Unit-radius construction / station spacing controls the polygonal length error.



The marked trip crosses the seam.

170° → 190° is a 20° trip.

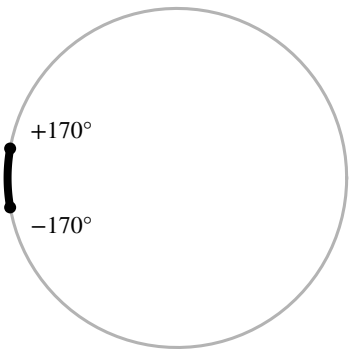
The endpoint may display −170°.

A display jump is not physical travel.

At R = 2 m: length = 2π/9 m

Refinement improves the approximation.
The edge metric and the permitted routes still need to be specified.

A A short physical trip crosses the map seam



B Arc edges and chords join the same stations

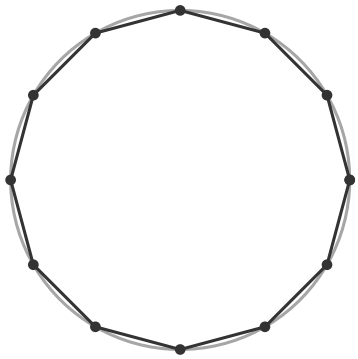
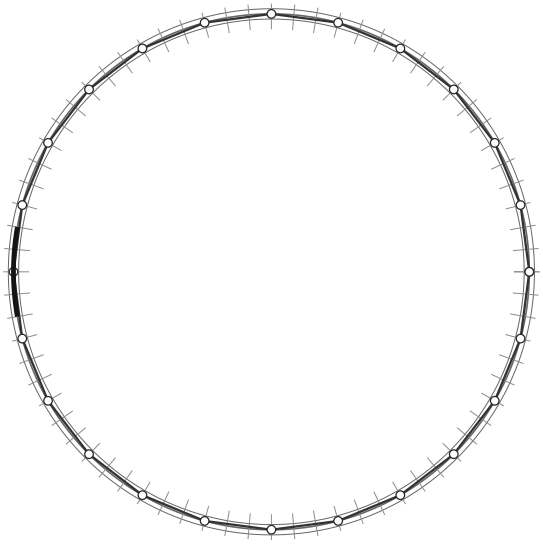


Figure 13.1: *A seam changes the report without lengthening the route.* The study continues on the following figure page.

The arc follows the track; the chord cuts across it

Unit-radius construction / station spacing controls the polygonal length error.



The marked trip crosses the seam.

170° → 190° is a 20° trip.

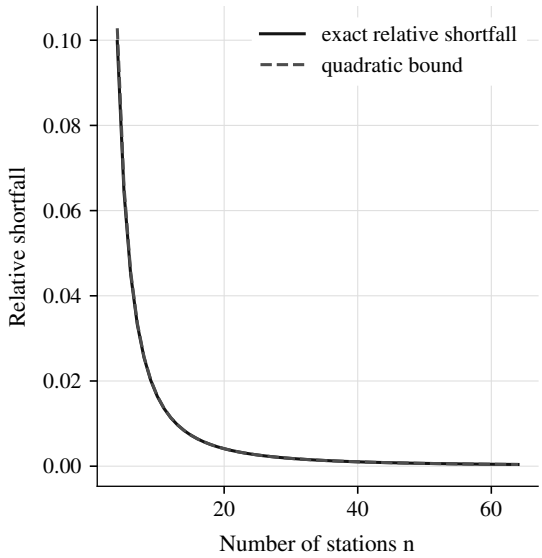
The endpoint may display −170°.

A display jump is not physical travel.

At R = 2 m: length = 2π/9 m

Refinement improves the approximation.
The edge metric and the permitted routes still need to be specified.

C The chord error shrinks with station count



D Position and lap history are different records

$d = R \min_k |\beta - \alpha + 2\pi k|$

At R=2 m: 20° gives 2π/9 m

Arc edges: exact station distance

Chord edges: controlled approximation

Lap count requires a lifted state.

Figure 13.1: *A seam changes the report without lengthening the route (continued).* The exact circle atlas recovers the shortest arc distance. A station graph with chord-valued edges has the explicit approximation error shown; that error decreases quadratically with station count.

Seam and error. Use $n = 36$ stations, radius $R = 2$, and endpoint errors bounded by one centimeter. Compare the chord-graph distance for a ten-edge route with its arc distance. Give a final interval that includes both edge approximation and endpoint uncertainty without counting either error twice.

Keep the conversion rule attached to the claim

An engineer replaces a database representation and wants to reuse a distance threshold. The useful question is not whether both versions are called embeddings. It is how their comparisons are related on the observations the system will actually receive.

Suppose $f : X \to Y$ and $h : Y \to Z$ obey two-sided bounds

$$a_1 d_X - \eta_1 \le d_Y(f\cdot, f\cdot) \le b_1 d_X + \eta_1,$$

$$a_2 d_Y - \eta_2 \le d_Z(h\cdot, h\cdot) \le b_2 d_Y + \eta_2.$$

Substituting the first into the second gives

$$a_2 a_1 d_X - a_2\eta_1 - \eta_2 \le d_Z(hf\cdot, hf\cdot) \le b_2 b_1 d_X + b_2\eta_1 + \eta_2.$$

The error incurred earlier is amplified by the later map. A version history therefore needs composition rules for its bounds, not just a list of individual test scores.

The exact functor and the approximate reconstruction

There is a clean exact categorical construction when the chosen realizations come with bijections $\Phi_T : X_T \to M_T$ onto their designated images. For a map $f : X_T \to X_{T'}$, set

$$F(f) = \Phi_{T'} \circ f \circ \Phi_T^{-1}.$$

Then $F(\mathrm{id}) = \mathrm{id}$ and $F(h \circ f) = F(h) \circ F(f)$ by cancellation. This is exact functoriality of maps. If the maps and their inverses are smooth, it also gives the smooth version. Metric distortion is quantified separately by the inequalities above. A map can obey composition exactly while preserving distances only approximately.

If a representation loses information, choose a reconstruction Ψ_T explicitly and record a bound such as $d_T(\Psi_T \Phi_T x, x) \le \varepsilon_T$. The composition $\Phi_{T'} f \Psi_T$ is then an approximate transported map. Its composition defect depends on the reconstruction errors and the intervening Lipschitz constants. It cannot use exact inverse cancellation because that inverse was not supplied.

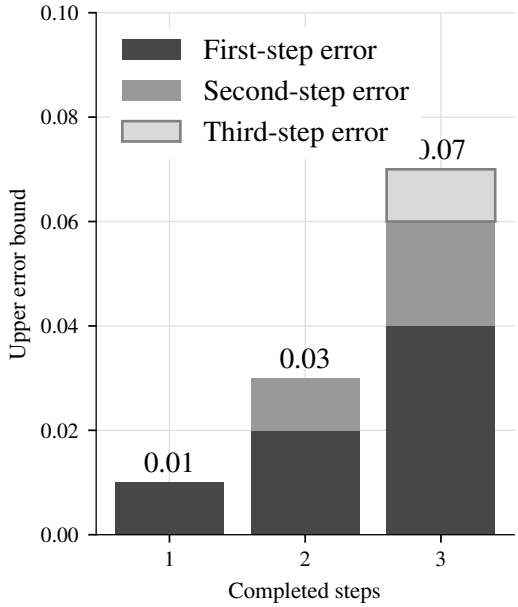
Proposition 13.1 (A round-trip comparison)**.** *Let $\Phi : X \to Y$, $\Psi : Y \to X$ satisfy $d_X(\Psi\Phi x, x) \le \varepsilon$ for all x in a region. Then*

$$|d_X(\Psi\Phi x, \Psi\Phi x') - d_X(x, x')| \le 2\varepsilon.$$

Proof. The triangle inequality bounds the first distance by the second plus the two endpoint errors. Interchanging the two endpoint pairs proves the opposite inequality. □

This is a useful guarantee even when the round trip is not an identity. A later decision threshold still needs enough margin to tolerate the possible 2ε change.

A *Later steps amplify earlier errors*



$e_{j+1} \le L_j e_j + \eta_j; \qquad L_j = 2, \quad \eta_j = 0.01$

An early discrepancy is amplified by every later transformation.

B *Each step carries the previous error*

Step	*Previous × 2*	*New*	*Total*
1	0	.01	.01
2	.02	.01	.03
3	.06	.01	.07

Figure 13.2: An error budget decomposed by the step at which each discrepancy enters. With Lipschitz factor two and a new error of 0.01 at each step, the bounds are 0.01, 0.03 and 0.07. Earlier errors are amplified by later transformations.

A single approximate comparison has a stated cost. An operating system usually chains several of them. We now keep each intermediate domain and error allowance visible while composing the maps.

Carry the error through the entire translation route

A telemetry observation passes through a sensor format, a normalized record, and a geometric representation. Each conversion has a documented error. A useful guarantee must describe the entire route, including the possibility that a later conversion magnifies an earlier error. Adding every number in the log without accounting for its units and amplification can produce a misleading budget.

Let $f : X \to Y$ and $g : Y \to Z$ be maps between metric spaces. Suppose

$$\ell_f d_X(x, x') - \varepsilon_f \le d_Y(fx, fx') \le L_f d_X(x, x') + \varepsilon_f,$$
$$\ell_g d_Y(y, y') - \varepsilon_g \le d_Z(gy, gy') \le L_g d_Y(y, y') + \varepsilon_g,$$

where all constants are stated on domains containing the relevant points and images. The lower factors are nonnegative. Apply the second upper bound to fx, fx', then the first:

$$d_Z(gfx, gfx') \le L_g L_f d_X(x, x') + L_g \varepsilon_f + \varepsilon_g.$$

The lower bound is

$$d_Z(gfx, gfx') \ge \ell_g \ell_f d_X(x, x') - \ell_g \varepsilon_f - \varepsilon_g.$$

The two additive constants need not be equal. If a common symmetric allowance is desired, use the larger one. The calculation states exactly what composition preserves.

An error expressed in one ruler must be transported

Suppose f changes a distance by at most a factor 1.02 plus 0.01 in Y units, and g changes a Y distance by at most 1.05 plus 0.02 in Z units. The upper composite factor is 1.071, and the additive allowance is $1.05(0.01) + 0.02 = 0.0305$ in Z units.

The original 0.01 was not added as though it already lived in Z. It was first passed through the second map's Lipschitz factor. With more stages, an early additive error is multiplied by every later upper factor:

$$d_m \le \left(\prod_{j=1}^{m} L_j \right) d_0 + \sum_{i=1}^{m} \varepsilon_i \prod_{j=i+1}^{m} L_j.$$

An empty product is one. This formula follows by induction and can be evaluated sequentially without storing the entire symbolic expression.

The factors describe the stated comparison model. They are not automatically independent random variables, and the additive terms are not automatically standard deviations. If they are deterministic bounds, the worst-case expression is appropriate. A statistical combination requires a separate dependence model.

A round trip answers a different question

Let $a : X \to Y$ and $b : Y \to X$ satisfy

$$d_X(ba(x), x) \le \eta_X, \qquad d_Y(ab(y), y) \le \eta_Y.$$

These bounds say that each round trip reconstructs its starting point within an allowance. They do not alone give a small distortion of all pairwise distances. For example, $a(x) = 100x$ and $b(y) = y/100$ on the real line have exact round trips, while a magnifies the ordinary coordinate distance by one hundred.

If the Y ruler is transformed to $d_Y(y, y') = |y - y'|/100$, the same maps become exact isometries. The stored arrays and round trips are unchanged. What changed was the comparison geometry. This simple example explains why encoder reconstruction and metric preservation need distinct lines in the specification.

Recover one distance bound from a controlled inverse

Suppose a is L_a-Lipschitz and b is L_b-Lipschitz, and the first round-trip bound holds. The triangle inequality gives

$$\begin{aligned} d_X(x, x') &\le d_X(x, ba(x)) + d_X(ba(x), ba(x')) + d_X(ba(x'), x') \\ &\le 2\eta_X + L_b d_Y(a(x), a(x')). \end{aligned}$$

For $L_b > 0$,

$$\frac{d_X(x, x') - 2\eta_X}{L_b} \le d_Y(a(x), a(x')) \le L_a d_X(x, x').$$

The lower estimate becomes uninformative below scale $2\eta_X$, which is precisely where the reconstruction allowance can overwhelm the separation. This is a scale-dependent statement, not a failure of the triangle inequality.

If the lower bound is negative, replace it by zero when reporting a distance interval. Do not interpret the negative number as a possible negative distance. The inequality simply stops excluding coincidence at that scale.

Compose semantic reconstruction allowances too

Let $r_X(u)$, $r_Y(u)$ and $r_Z(u)$ represent the same underlying entity u in three realizations. Suppose

$$d_Y(f(r_X(u)), r_Y(u)) \le \eta_f, \qquad d_Z(g(r_Y(u)), r_Z(u)) \le \eta_g.$$

If g obeys an upper comparison with factor L_g and additive allowance ε_g, then

$$d_Z(gf(r_X(u)), r_Z(u)) \le L_g\eta_f + \varepsilon_g + \eta_g.$$

This tracks identity compatibility along the route. It is distinct from comparing two different entities, and it uses the same basic principle: express an earlier discrepancy in the later ruler before combining it with the later allowance.

What makes the categorical language precise

Objects in the realization categories carry their domains, representation maps, and metrics. Morphisms carry the actual comparison maps together with finite bounds of the stated kind. The underlying function composition is associative, and the formulas above show that compositions remain in the allowed class when constants and domains are propagated. Identity maps have factor one and zero allowance.

A class restricted to one fixed small tolerance need not be closed under composition. The constants should therefore be carried as bounds that can change along a route, or organized in an explicitly filtered family. The word category does not make accumulated error disappear. It organizes the maps while their numerical certificates record what the organization costs.

Likewise, a geometric realization chosen from approximate data is not automatically a canonical functor. Its action on objects and morphisms must be specified, and the relevant composition identities or comparison maps must be checked. The exact atlas construction provides a concrete case. For an approximate realization, the chosen construction and its distortion are part of the statement.

Choose the route for the quantity that matters

If two translation routes reach the same final format, they may have different comparison budgets. One route might have fewer stages but a poorly conditioned intermediate representation; another might have more stages but preserve the relevant metric more faithfully. The best route cannot be selected merely by counting conversions.

For a distance decision, compare the resulting upper and lower intervals at the operational scale. For a lineage requirement, inspect whether the exact event identifier survives. For a lifted orientation history, inspect the sign and time conventions. These are separate contracts attached to one observation, and the construction makes their shared dependencies visible.

Three-stage budget. The successive upper factors are 1.02, 1.05 and 1.01. The corresponding additive allowances are 0.01, 0.02 and 0.005. Calculate the final upper factor and additive allowance. Then reverse the stages, stating the hypothetical domain assumptions required to make that reversed route meaningful. Explain why the additive budget changes even though the product of scalar factors does not.

The compositions so far translated a fixed comparison. An incoming observation can also update the represented state. We now keep one sensor and reverse the order of two readings to test whether a proposed equivalence preserves the updates as well as their individual values.

A replacement can reproduce today's score yet discard information needed to process tomorrow's reading. We therefore test the conversion on a sequence. If the hidden condition changes between

observations, swapping their order can change the inferred current state. Preserving the update process is part of preserving that predictive task.

Two sensor readings can have the same counts and different consequences

Imagine a pump with two hidden states, 0 for steady and 1 for unstable. Its vibration sensor reports 0 for quiet or 1 for an alert and is wrong with probability β. Between readings the pump switches states with probability α. An engineer updates the probabilities after each reading and then predicts the state at the next reading. Write q for the probability vector before the reading and $U_x(q)$ for the update and prediction after report x:

$$U_x(q) = M^T \frac{E_x \odot q}{\mathbf{1}^T (E_x \odot q)}, \quad M = \begin{pmatrix} 1-\alpha & \alpha \\ \alpha & 1-\alpha \end{pmatrix},$$

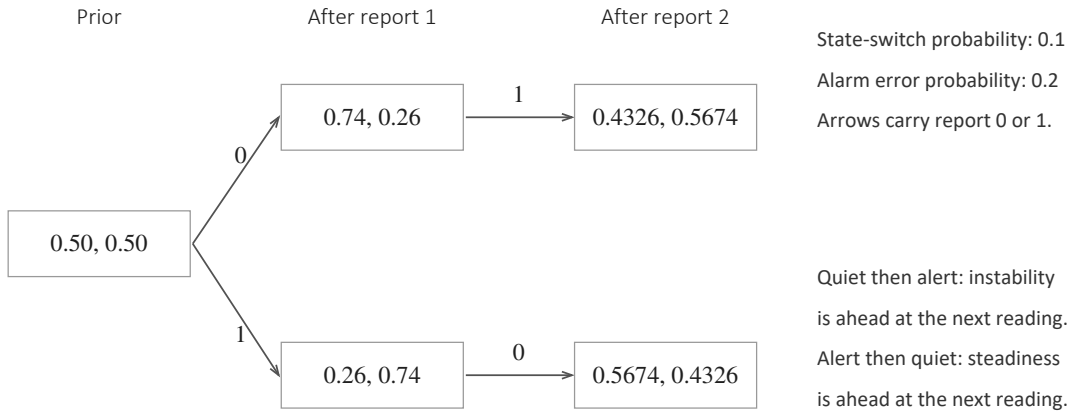
where $E_0 = (1-\beta, \beta)^T$, $E_1 = (\beta, 1-\beta)^T$ and $\odot$ denotes componentwise multiplication. The denominator is the observation probability and must be positive.

Take $\alpha = 0.1$, $\beta = 0.2$ and start at $q = (0.5, 0.5)$. After observing zero, the posterior before transition is $(0.8, 0.2)$ and the predicted state after transition is $(0.74, 0.26)$. Observing one next gives the unnormalized weights $(0.148, 0.208)$, whose sum is 0.356. Normalizing and transitioning gives approximately $(0.432584, 0.567416)$.

Reverse the observations. The two final probabilities exchange places: the probability of hidden state zero is now approximately 0.567416. If I only keep an alarm count, both pump logs say “one alert.” But with the alert last, this model puts instability ahead of steadiness at the next check. Reverse the log and steadiness comes out ahead. The total-variation distance between the final distributions is approximately 0.134831.

The pump sees the same reports in two orders

State 0 is steady and state 1 is unstable. Each report is followed by a prediction to the next reading.



Pairs give probabilities of steady and unstable.

A The hidden state can change between readings

$$\begin{bmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{bmatrix}$$

B Reversing the readings changes the prediction

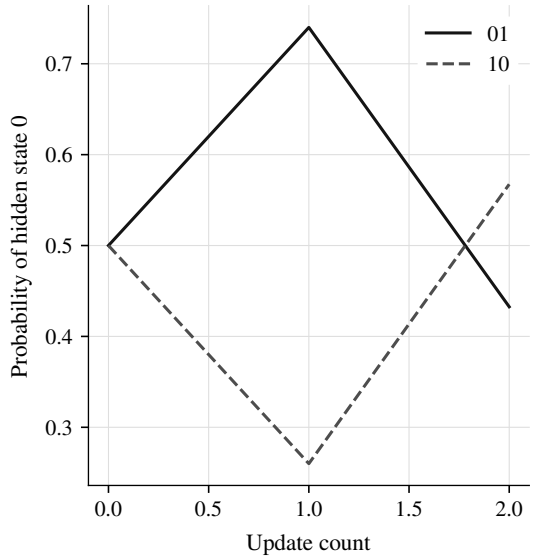
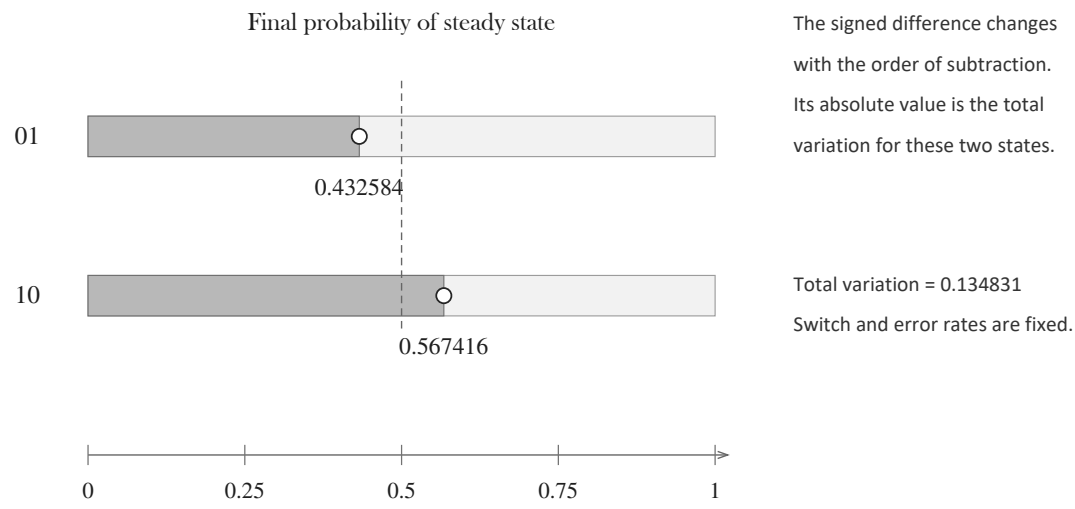


Figure 13.3: *The same two observations in a different order.* The study continues on the following figure page.

Reversing the reports reverses the prediction

Both sequences contain one quiet report and one alert. The final predictions use the same starting prior.



C Follow the update from zero to one

Prior: (0.5, 0.5)

After zero + transition: (.74,.26)

Next weights: (.148,.208)

Normalizer: .356

Final: (0.432584, 0.567416)

D Equal counts can lead to different answers

01: (0.432584, 0.567416)

10: (0.567416, 0.432584)

Total variation = 0.134831

Both histories contain one 0 and one 1.

Figure 13.3: *The same two observations in a different order (continued).* The transition matrix and two prediction trajectories compare orders 01 and 10. The register gives the intermediate weights and normalization for 01, followed by both final predictions and their total-variation difference. The diagram follows the two report orders at the stated switch and error rates.

The update commutator has a precise meaning

The difference $U_1(U_0(q)) - U_0(U_1(q))$ measures order sensitivity for these two nonlinear maps at this starting distribution. It is an update commutator. The causal-state research distinguishes this discrete construction from the holonomy of a differential-geometric connection [25]. That distinction is helpful: it lets us use the finite example without claiming that every order effect is curvature.

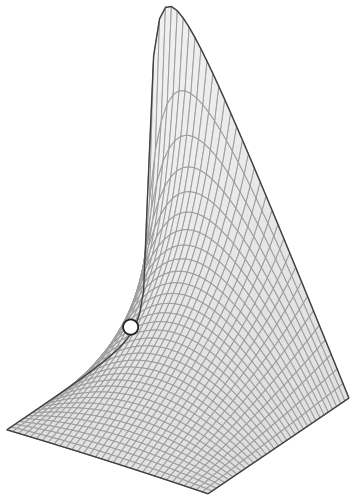
For the symmetric two-state model at the stationary prior, the total-variation difference is

$$\Delta(\alpha, \beta) = \frac{\alpha(1-2\alpha)(1-2\beta)}{\alpha(1-2\beta)^2 + 2\beta(1-\beta)}, \qquad 0 < \alpha, \beta < \tfrac{1}{2}.$$

This is the height plotted on the next plate. At the marked operating point $\alpha = 0.1$, $\beta = 0.2$, it is 0.134831 to six decimal places. It vanishes as the observations become uninformative, $\beta \to 1/2$, and as the transition makes the next hidden state independent of the previous one, $\alpha \to 1/2$. It also tends to zero as $\alpha \to 0$ with fixed positive β: then repeated observations update one fixed hidden condition, and their likelihood factors commute.

The order effect has a computable landscape

Height = total-variation difference between the two observation orders.



α: state-switch probability

β: alarm error probability

Ring: α = .1 and β = .2

Prediction difference = .134831

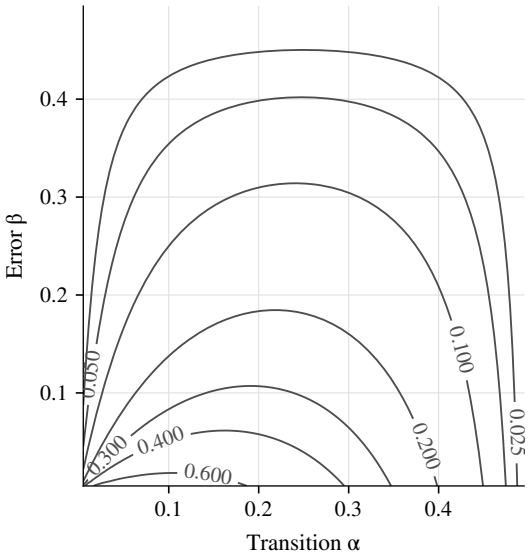
Zero limits: α → 1/2 or β → 1/2.

Also α → 0 with fixed β > 0.

At β → 0 with α > 0 fixed,

the limit is 1 − 2α, not zero.

A Order sensitivity depends on two error rates



B Fixing one rate exposes the dependence

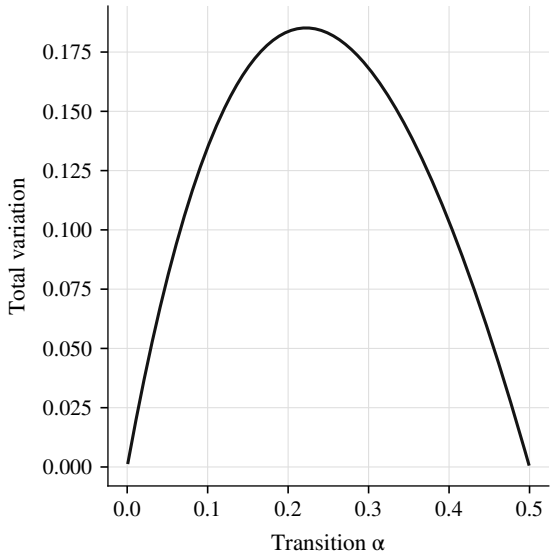
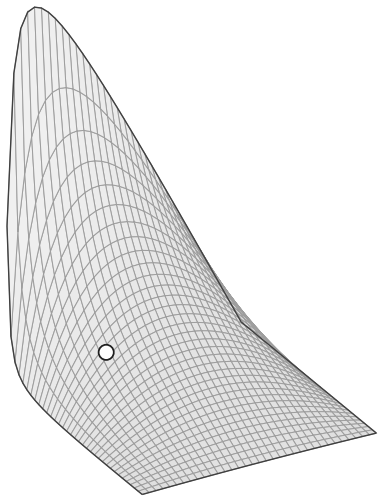


Figure 13.4: *An order effect depends on both dynamics and observation noise.* The study continues on the following figure page.

The order effect has a computable landscape

Height = total-variation difference between the two observation orders.



α: state-switch probability

β: alarm error probability

Ring: α = .1 and β = .2

Prediction difference = .134831

Zero limits: α → 1/2 or β → 1/2.

Also α → 0 with fixed β > 0.

At β → 0 with α > 0 fixed,

the limit is 1 − 2α, not zero.

C Boundary cases remove the order effect

α→0: fixed hidden state

α→1/2: next state is independent

β→1/2: observations uninformative

Each gives zero order difference for the stated limit.

D The calculation compares discrete updates

U1(U0(q)) − U0(U1(q))

A discrete update commutator

Hidden-state beliefs are represented explicitly.

No differential connection is inferred from order sensitivity alone.

Figure 13.4: *An order effect depends on both dynamics and observation noise (continued).* The exact two-state formula is evaluated over the stated parameter domain. Boundary limits are paired with explicit update calculations. The surface is a constructed mathematical example, not an empirical performance map.

What an equivalence must preserve under composition

Suppose a representation map F is meant to preserve predictive updates. The relevant requirement is a commuting relation, $F \circ U_x = \widetilde{U}_x \circ F$, for each observation symbol. It then preserves any finite sequence by repeated substitution. If it only preserves one-step scores, it need not preserve the state needed for the next update.

Ask what happens at the next pump reading. A tally can be enough if the pump stays in one state and the sensor makes independent readings with the same error rate. In our switching example, I need enough of the history to tell whether the alert came first or last. Any translation I use for prediction has to preserve the next update.

Organize the maps the calculations used

Two angular coordinates agreed on the distance between positions of the calibration cart. Replacing arcs by chords introduced a measurable error. Chaining conversions can amplify an earlier error, and the sequential sensor example showed that equal scores can hide differences in what the next update needs. We now organize those requirements for families of implementations. Each family needs specified objects, admissible maps, identities and a composition that stays within its declared domains.

Eilenberg and Mac Lane introduced the categorical language in their 1945 paper "General Theory of Natural Equivalences" [1]. Here it records a concrete requirement already met in our chosen constructions, the result of a sequence of operations must agree with the result obtained by composing their translated descriptions. The numerical distortion bounds accompany those maps.

What is a category? What is a functor? Before we define our categories $\mathsf{SamTrans}(S)$ and $\mathsf{SamGeom}(S)$, let's understand what these words mean.

A category is a collection of:

- **Objects**: Things you're studying (spaces, groups, graphs, models, systems)
- **Morphisms** (or arrows): Structure-preserving transformations between objects
- **Composition**: A way to chain morphisms together, satisfying natural rules (associativity, identity)

The key insight: you don't study objects in isolation. You study them *together with the ways they transform into each other.* The morphisms are as important as the objects.

Examples of categories:

- **Set**: Objects are sets, morphisms are functions $f : A \to B$
- **Top**: Objects are topological spaces, morphisms are continuous maps
- **Grp**: Objects are groups, morphisms are group homomorphisms (structure-preserving maps)
- **Vect**: Objects are vector spaces, morphisms are linear transformations

A functor is a structure-preserving transformation *between categories.* It maps:

- Objects in category $\mathcal{C}$ to objects in category $\mathcal{D}$
- Morphisms in $\mathcal{C}$ to morphisms in $\mathcal{D}$
- While preserving composition and identities

Think of a functor as a "translation device" between two mathematical worlds. It says: "for every object X in world $\mathcal{C}$, here's a corresponding object $F(X)$ in world $\mathcal{D}$. And for every transformation $f : X \to Y$ in world $\mathcal{C}$, here's a corresponding transformation $F(f) : F(X) \to F(Y)$ in world $\mathcal{D}$."

The functor ensures that relationships in $\mathcal{C}$ translate coherently to relationships in $\mathcal{D}$. If $f : X \to Y$ and $g : Y \to Z$ compose in $\mathcal{C}$ to give $g \circ f : X \to Z$, then $F(g) \circ F(f) = F(g \circ f)$ in $\mathcal{D}$. Structure is preserved.

A concrete example: Fundamental group functor. Here's a classic functor from topology:

The fundamental group functor $\pi_1 : \mathbf{Top}_* \to \mathbf{Grp}$ starts with a space and a marked starting point. Imagine drawing routes on a globe with a pin at your home port: every loop must leave that pin and return to it. The pin is the *basepoint* x_0. Objects of $\mathbf{Top}_*$ are pointed spaces (X, x_0); arrows are continuous maps $f : (X, x_0) \to (Y, y_0)$ with $f(x_0) = y_0$.

The group $\pi_1(X, x_0)$ records loops up to continuous deformation that keeps the starting point fixed. Traversing one loop and then another is the group operation. For the familiar connected examples,

- $\pi_1(S^2, x_0) = \{e\}$: every loop on the two-dimensional sphere can contract to the pin.
- $\pi_1(T^2, x_0) \cong \mathbb{Z} \times \mathbb{Z}$: a torus has two independent winding directions.
- The figure-eight has fundamental group $\mathbb{Z} * \mathbb{Z}$: the order of the two basic loops matters.

A basepoint-preserving map sends a loop γ to $f \circ \gamma$ and induces a homomorphism $f_* : \pi_1(X, x_0) \to \pi_1(Y, y_0)$. Applying two maps to a loop in succession agrees with applying their composite, so this is a functor. Nonisomorphic fundamental groups rule out a homeomorphism between path-connected spaces; isomorphic groups alone do not establish one. Hatcher develops the construction and these examples in Chapter 1 [5].

Why composition matters here. The loop example illustrates the feature we need: performing two operations and then translating their description agrees with translating each operation and composing the results. Programs, databases and models of interacting systems offer other settings for that question. A categorical model must specify its objects, maps and composition before this agreement can be checked.

In topological quantum field theory, for example, geometric pieces are assigned algebraic data in a way compatible with gluing.[1] For a software comparison, translating a composite operation must give the same map as translating the steps and composing them in order. Track the distance distortion of that map separately.

Symmetry and a stated invariant. Noether's theorem supplies a related discipline from variational mechanics: under its hypotheses, continuous symmetries of the action yield conservation laws. Time, spatial and rotational symmetries lead to energy, momentum and angular-momentum conservation.[2] The preserved quantity and the allowed transformation must both be specified. Here we must establish our own preservation equations for translated observations, route lengths and decisions. The following constructions state those equations for the systems we actually compare.

[1] Michael Atiyah, "Topological quantum field theory," *Publications Mathématiques de l'IHÉS* 68 (1988), 175–186, https://doi.org/10.1007/BF02698547.

[2] Emmy Noether, "Invariant Variation Problems" (1918), English translation by M. A. Tavel, https://arxiv.org/abs/physics/0503066.

Why we need functors for semantic sameness. Up to now we have mostly talked about a *single* translator system or a *single* manifold-based model at a time. In practice, we rarely have just one:

- We iterate on architectures and training objectives
- We deploy multiple versions of a system in parallel (A/B testing, regional variants)
- We build different models for different regimes (faces in daylight vs. low light; transactions in USD vs. EUR), all realizing the same underlying semantic sameness structure S

On the translator side, we get many different graphs of feature spaces and translators. On the geometric side, we get many different manifolds, metrics, and chart systems. We would like to talk about *families* of such realizations at once, and about the transformations that connect them.

The language for this is category theory. This chapter introduces two categories:

- $\mathsf{SamTrans}(S)$, whose objects are translator-based realizations of a semantic sameness structure S
- $\mathsf{SamGeom}(S)$, whose objects are geometry-based realizations of S

The compatible chart data and chosen realizations of Chapters 11 and 12 supply the following functors on the specified smooth objects and maps

$$F_S : \mathsf{SamTrans}(S) \to \mathsf{SamGeom}(S),$$
$$G_S : \mathsf{SamGeom}(S) \to \mathsf{SamTrans}(S).$$

which relate these categories.

What these functors do:

- F_S takes a translator system and constructs its Riemannian realization (the manifold it approximates)
- G_S takes a Riemannian manifold and constructs a translator system that samples it via charts
- Composing them: $G_S \circ F_S$ takes a translator system, constructs its manifold, then samples it back into translators. A return bound requires the reconstruction conditions stated below.

For chosen realization maps, we call the resulting comparison an **ε-equivalence of realizations** when the distance, semantic and round-trip bounds below hold. The bounds $\varepsilon_{\mathrm{rt}}, \varepsilon_{\mathrm{cyc}}, \varepsilon_{\mathrm{loc}}$ and ε_{Φ} track different sources of error.

The Error Budget Transfer Theorem (Chapter 20) combines the comparison bound with the probability of comparison failure and the mass near the decision threshold. Each direction of transfer needs those quantities for the population and decision rule it uses.

The comparison is useful when it identifies a specific invariant: the same measured distance, the same admissible path, or a decision whose uncertainty remains bounded. We can examine those obligations before introducing the formal families of realizations.

Which claims can a replacement map carry?

The cart replacement had several obligations. A seam conversion preserved position and route length. A finite station graph introduced a calculable length error. Discarding lap count changed the task the representation could answer. The sensor example added another obligation, preserving the state needed by the next update.

The common question is which maps preserve the specified operation with a stated allowance. The realization categories organize those maps. Their distance inequalities can support a threshold

comparison when the decision's margin and population conditions are also supplied. We now state the map conditions that the worked examples have made necessary.

Why "approximate" equivalence is correct

A map can preserve composition exactly while distorting distances. An approximate reconstruction can also move a point on return. We therefore keep algebraic identities, metric comparison bounds and reconstruction errors separate, as in the preceding constructions.

The approximation comes from three sources:

- **Translator imperfections**: Round-trip and cocycle bounds ε_{rt} and ε_{cyc} (Chapter 10)
- **Chart distortion**: Local bi-Lipschitz constants ε_{loc} (Chapter 10)
- **Realizability distortion**: The map Φ from the operating region of X_T into M has its own distance bounds (Chapter 12)

These error bounds are *not nuisances we wish away.* They are the measurable constants that determine how well guarantees transfer. An "approximate equivalence" between translator and geometric realizations means:

- Distances agree up to multiplicative factors $(1 \pm \varepsilon)$ and additive slack δ
- Round-trips return you to within ε of where you started
- Under the chartability assumption, the cocycle bound ε_{cyc} applies on chart overlaps.

Key insight: The ε bounds are:

1. **Measurable**: We can estimate them empirically (Chapter 21)
2. **Auditable**: We can check whether they're in green/yellow/red regimes
3. **Task-dependent**: Whether an allowance of $\varepsilon = 0.05$ or 0.1 is useful depends on the units, decision margin and population; those numbers are illustrative tolerances, not a general empirical guarantee
4. **Explicit**: Error budget transfer will give concrete formulas relating slack in one realization to slack in another

Exact coordinate identities can hold in the mathematical construction, and exact decoding identities can hold for discrete records. Numerical geometry also has approximation errors to bound. The relevant question is which identity or allowance the operation actually supplies.

The category records which maps can be composed. Their quantitative bounds record what that composition costs. Keeping those jobs separate lets us compare different implementations without treating an approximate metric agreement as an exact inverse.

The Category $\mathsf{SamTrans}(S)$ of Translator Realizations

We begin with the translator builder's side.

Return to the cart's implementations. Each has a representation space, a region of operation, and a map attaching physical positions to its records. A replacement map must retain the quantities the application needs. The same inventory describes the telemetry conversions and the predictive sensor state. It is this explicit inventory that the category collects.

The systems and their permitted replacements. An object records one realization of S. Its feature spaces and translators define the computational comparison; its operating region specifies the

claim's domain; its reference assignment connects the represented values to the underlying entities. A map between two such objects must satisfy the semantic, regularity and distance conditions given below.

The following hypothetical diagnostic model applies that inventory to heterogeneous observations. Its channel dimensions, example distances and uncertainty allowances are stipulated for the construction. Each proposed replacement adds a different issue, a panel upgrade, a changed observation quality, or compression. We use them to test the same map conditions rather than treating a distance equation as clinical evidence.

Concrete example: A medical diagnosis translator system. Consider a hospital's diagnostic system for detecting inflammatory conditions. The semantic sameness structure S is:

- **Latent entities** u: actual patient disease states (active inflammatory bowel disease, lupus flare, normal inflammation from infection, healthy baseline)
- **Benign evolution**: disease progression under treatment, normal recovery, lab value fluctuations within physiological ranges
- **Observation channels**: blood work (C_{blood}), medical imaging (C_{imaging}), patient-reported symptoms (C_{symptoms}), genetic markers (C_{genetic})

A translator-based realization T of this structure consists of:

1. Feature spaces $\{(V_i, d_i)\}$:

- $V_{\text{blood}} = \mathbb{R}^{128}$: Blood work encoded into 128-dimensional vectors (white cell counts, inflammatory markers like CRP and ESR, antibody panels, liver enzymes). Distance d_{blood} is normalized Euclidean distance.
- $V_{\text{imaging}} = \mathbb{R}^{256}$: CT and MRI scans encoded by a trained radiology CNN into 256-dimensional embeddings. Distance d_{imaging} measures similarity of imaging presentations.
- $V_{\text{symptoms}} = \mathbb{R}^{64}$: Patient questionnaires and clinical notes processed by NLP into 64-dimensional symptom embeddings (pain scores, fatigue levels, GI symptoms). Distance d_{symptoms} is cosine distance.
- $V_{\text{genetic}} = \mathbb{R}^{512}$: Genomic data (SNP profiles, gene expression) embedded into 512-dimensional space. Distance d_{genetic} is based on genetic similarity.

2. Translators $\{T_{ij}\}$:

- $T_{\text{blood}\to\text{imaging}} : \mathbb{R}^{128} \to \mathbb{R}^{256}$: Neural network trained to predict what imaging features we'd expect given blood work (e.g., high CRP $\to$ predict intestinal wall thickening on CT)
- $T_{\text{symptoms}\to\text{blood}} : \mathbb{R}^{64} \to \mathbb{R}^{128}$: Maps reported symptoms to expected lab value ranges (e.g., severe abdominal pain + fatigue $\to$ elevated inflammatory markers)
- $T_{\text{genetic}\to\text{imaging}} : \mathbb{R}^{512} \to \mathbb{R}^{256}$: Connects genomic risk factors to typical imaging presentations
- Plus reverse translators and other pairs, forming a connected graph

3. Translator metric quotient (X_T, d_T):

The stitched-together space formed by taking the disjoint union of all four feature spaces, defining translator paths (move within blood space, jump via $T_{\text{blood}\to\text{imaging}}$, move within imaging space, etc.), and measuring distance as shortest path length. Identify descriptions when the path infima are zero in both directions. An exact pair of inverse translations suffices, and longer routes can also give zero cost.

4. In-regime region $\mathcal{R}_T \subseteq X_T$:

The subset of X_T visited by patients with conditions the system is trained for:

- Adult patients (ages 18-75)
- Common inflammatory conditions (IBD, lupus, rheumatoid arthritis, celiac disease)
- Standard test protocols (not exotic assays or experimental imaging)
- Compliant patients who can report symptoms clearly

Not in $\mathcal{R}_T$:

- Rare autoimmune diseases (insufficient training data)
- Pediatric patients (different reference ranges)
- Patients with multiple comorbidities creating confounding signals
- Adversarial cases (fabricated symptoms, tampered samples)

For this constructed model, stipulate chartability allowances $\varepsilon_{\text{rt}} \approx 0.05$, $\varepsilon_{\text{cyc}} \approx 0.08$, $\varepsilon_{\text{loc}} \approx 0.06$ on $\mathcal{R}_T$. An application would have to establish them on its actual operating region.

5. Realization map $\pi_T : S \to X_T$:

For each disease state u in the regime of interest, $\pi_T(u)$ is a point (or small region) in X_T representing that state:

- π_T(active Crohn's disease): A cluster in X_T where observations from Crohn's patients land. Typical features: elevated CRP (blood), intestinal wall thickening (imaging), abdominal cramping (symptoms), specific genetic markers (genetic). All four channels, when encoded and lifted to X_T, cluster around this point.
- π_T(lupus flare): A different cluster, distant from Crohn's in d_T. Features: different antibody profile (blood), joint inflammation patterns (imaging), different symptom constellation (symptoms), different genetic risk factors (genetic).
- π_T(healthy baseline): A large, diffuse region where normal patients land. Low inflammatory markers, clean imaging, minimal symptoms, no high-risk genetic variants.

The realization map captures: **whenever we observe a patient with disease state u through any of the four channels and encode those observations, they land near $\pi_T(u)$ in translator space.**

What makes this a valid object in $\mathsf{SamTrans}(S)$*?* This translator system T qualifies as an object in $\mathsf{SamTrans}(S)$ because:

1. **It's a translator system** (Chapter 9): We have feature spaces, translators, and a well-defined translator metric quotient (X_T, d_T).
2. **Chartability is stipulated on $\mathcal{R}_T$** (Chapter 10): The model supplies local chart regions, coverage, regularity and bounds on the declared domains. A sample of small residuals by itself cannot establish this assumption.
3. **It has a realization map π_T**: We can coherently assign disease states u to regions of X_T. Different observations of the same disease state (blood work on Monday, imaging on Wednesday, symptoms reported Friday) all land near the same point $\pi_T(u)$, up to small observation noise.

Partial and approximate:

The realization map π_T is:

- **Partial**: Only defined for disease states in the trained regime. Novel diseases, pediatric cases, or adversarial inputs don't have well-defined $\pi_T(u)$ because they're outside $\mathcal{R}_T$.
- **Approximate**: Observations don't land *exactly* at $\pi_T(u)$; they land in a small ball around it due to measurement noise, biological variability, and encoder imperfections. The "near" statement means: with high probability (say, 95%), a genuine observation of disease state u lands within distance $\tau_{\text{noise}} \approx 0.1$ of $\pi_T(u)$ in d_T.

Why we need the realization map π_T. Without π_T, we'd just have a metric space (X_T, d_T)—numbers and distances with no semantic meaning. The realization map π_T connects the abstract structure S (disease states, benign evolution) to the concrete translator system (feature vectors, distances).

It lets us formulate three distinct checks:

- Observations of the same entity should lie near its chosen reference point $\pi_T(u)$ under the observation model. To infer identity from proximity, distinct entities must also have adequate separation at the task's resolution.
- Benign evolution should follow the admissible paths specified by S. A geometrically short path is benign only if the task model makes it so.
- An observation outside $\mathcal{R}_T$ lacks the stated guarantee. An unfamiliar observation can also fall inside a familiar region, so distance alone need not detect every regime violation.

The realization map is the bridge from latent semantics to computational geometry.

Morphisms: structure-preserving maps between translator realizations

Next we define morphisms between translator realizations. Intuitively, a morphism $f : T \to T'$ in $\mathsf{SamTrans}(S)$ is a way to move from one translator-based realization of S to another without breaking the semantic structure.

Typical examples include:

- Replacing one encoder with another while keeping the underlying semantics fixed.
- Migrating to a new translator graph that refactors feature spaces but preserves the induced translator metric on the in-regime region.
- Compressing or coarsening a translator system while tracking how points in $\mathcal{R}_T$ map into $\mathcal{R}_{T'}$.

Formally, we ask for a map between the in-regime regions that respects both the semantics and the metric structure up to controlled slack.

Definition 13.2 (Morphisms in $\mathsf{SamTrans}(S)$)**.** Given two translator-based realizations T and T' of S with in-regime regions $\mathcal{R}_T \subseteq X_T$ and $\mathcal{R}_{T'} \subseteq X_{T'}$, a *morphism* $f : T \to T'$ in $\mathsf{SamTrans}(S)$ consists of a map

$$f : \mathcal{R}_T \to \mathcal{R}_{T'}$$

such that:

1. **Semantic compatibility:** For all latent entities u in the regime of interest,

   $$d_{T'}\big(f(\pi_T(u)),\, \pi_{T'}(u)\big)$$

   is bounded by a declared finite allowance η_f (ideally zero). That is, f carries the representation of u in T close to the representation of u in T'.

2. **Regularity:** In the smooth category, $\mathcal{R}_T$ and $\mathcal{R}_{T'}$ carry the smooth structures supplied by their chosen realizations, and f is smooth in these coordinates. The designated regions are open manifold regions, or have a separately specified smooth structure. A distance inequality alone does not imply smoothness.
3. **Metric control:** There are finite constants $0 < L_{\min} \le L_{\max}$ and $\varepsilon_f \ge 0$, valid on the declared region:

$$L_{\min}\, d_T(x,y) - \varepsilon_f \;\le\; d_{T'}\big(f(x), f(y)\big) \;\le\; L_{\max}\, d_T(x,y) + \varepsilon_f.$$

Concrete examples of morphisms in $\mathsf{SamTrans}(S)$. To make this concrete, let's continue with the medical diagnosis domain and examine several typical morphisms between translator-based realizations.

Example 1: Laboratory test panel upgrade.
System T (2018): Hospital uses a standard inflammatory marker panel:

- Blood feature space $V_{\text{blood}}^{\text{old}} = \mathbb{R}^{64}$: encodes CRP, ESR, white blood cell count, basic antibody panel (12 markers total)
- Translator metric quotient X_T, in-regime region $\mathcal{R}_T$
- Crohn's disease maps to reference point $\pi_T(\text{Crohn's}) \in \mathcal{R}_T$

System T' (2024): Same hospital upgrades to comprehensive panel:

- Blood feature space $V_{\text{blood}}^{\text{new}} = \mathbb{R}^{128}$: includes all old markers *plus* calprotectin, cytokine profiles, microbiome markers, metabolite panels (35 markers total)
- Translator metric quotient $X_{T'}$, in-regime region $\mathcal{R}_{T'}$
- Crohn's disease maps to reference point $\pi_{T'}(\text{Crohn's}) \in \mathcal{R}_{T'}$

The morphism $f : T \to T'$:
The map $f : \mathcal{R}_T \to \mathcal{R}_{T'}$ tracks how disease states move from the old system to the new one:

1. **Semantic compatibility**: A patient with active Crohn's disease has representation $\pi_T(\text{Crohn's})$ in the old system. When we run the new comprehensive panel on the same patient (or a cohort of Crohn's patients), their results encode to $\pi_{T'}(\text{Crohn's})$ in the new system.
   The morphism ensures:

$$d_{T'}\big(f(\pi_T(\text{Crohn's})),\, \pi_{T'}(\text{Crohn's})\big) \approx 0.$$

   In words: the old system's representation of Crohn's maps to the new system's representation of Crohn's. Same disease, different coordinates.
2. **Metric control**: Consider two disease states in the old system: active Crohn's and Crohn's in remission. Suppose $d_T(\text{active}, \text{remission}) = 0.4$ in the old translator metric.
   For this hypothetical comparison, suppose the new system gives $d_{T'}(f(\text{active}), f(\text{remission})) = 0.5$.
   The bi-Lipschitz condition bounds this change:

$$L_{\min} \cdot 0.4 - \varepsilon_f \le 0.5 \le L_{\max} \cdot 0.4 + \varepsilon_f.$$

   With $L_{\min} = 0.9$, $L_{\max} = 1.3$, $\varepsilon_f = 0.05$:

$$0.9 \cdot 0.4 - 0.05 = 0.31 \le 0.5 \le 1.3 \cdot 0.4 + 0.05 = 0.57.$$

This pair satisfies the stipulated interval. Whether the panel improves discrimination requires labeled observations and a specified decision rule. The circular-track construction below lets us check the distance bound across the entire model.

Why this is important: Hospitals have historical patient records in the old system's coordinates. When they upgrade lab tests, they need to ensure that:

- Historical diagnoses remain valid (semantic compatibility)
- Risk thresholds calibrated on old data transfer approximately to new data (metric control)
- Longitudinal patient tracking isn't disrupted (same patient before/after upgrade maps consistently)

The morphism f formalizes this migration path and provides explicit bounds $(L_{\min}, L_{\max}, \varepsilon_f)$ on how much the representation changes.

Example 2: Multi-site deployment with different imaging equipment.

System T (Academic Medical Center):

- Imaging space $V_{\text{imaging}}^{\text{AMC}} = \mathbb{R}^{256}$: Uses state-of-the-art 3 Tesla MRI, latest CT scanners
- Encoder trained on high-quality reference imaging
- $\pi_T(\text{active IBD})$: Detailed mucosal inflammation patterns visible

System T' (Regional Hospital):

- Imaging space $V_{\text{imaging}}^{\text{regional}} = \mathbb{R}^{256}$: Uses older 1.5 Tesla MRI, standard CT
- Encoder adapted for lower-resolution, higher-noise imaging
- $\pi_{T'}(\text{active IBD})$: Coarser inflammation patterns, less detail

The morphism $f : T \to T'$:

Both hospitals participate in a multi-center clinical trial. They need to ensure that "active IBD" diagnosed at the academic center corresponds reliably to "active IBD" at the regional hospital, even though imaging quality differs.

1. **Semantic compatibility**: The morphism f maps:

$$f(\pi_T(\text{active IBD})) \approx \pi_{T'}(\text{active IBD}).$$

   Both sites agree on what constitutes active disease, even though their imaging representations differ in resolution and noise.

2. **Metric control**: At the academic center, mild IBD ($d_T = 0.2$ from healthy baseline) and severe IBD ($d_T = 0.6$ from baseline) are clearly distinguished.
   At the regional hospital, lower resolution compresses distances somewhat:

$$0.85 \cdot d_T(x, y) - 0.03 \leq d_{T'}(f(x), f(y)) \leq 1.1 \cdot d_T(x, y) + 0.03.$$

   The regional hospital's representation slightly compresses the severity spectrum ($L_{\min} = 0.85 < 1$) but doesn't collapse distinctions entirely. With $d_T = 0.4$ (moderate disease), the regional system gives $d_{T'} \approx 0.31$–0.47, preserving the rough ordering.

Why this is important: Multi-center trials must aggregate data from hospitals with varying equipment quality. The morphism f provides:

- A formal way to map diagnoses across sites (semantic compatibility)

- Explicit bounds on how much measurement uncertainty increases at lower-quality sites (metric control via $L_{\min}, L_{\max}, \varepsilon_f$)
- Guidance for which severity distinctions are reliable across sites (where metric control is tight) and which aren't (where ε_f is large)

Example 3: Compression for telemedicine / rural clinic deployment.
System T (Full system):

- Four feature spaces: blood ($\mathbb{R}^{128}$), imaging ($\mathbb{R}^{256}$), symptoms ($\mathbb{R}^{64}$), genetic ($\mathbb{R}^{512}$)
- Full translator graph with all pairwise translators
- Requires high-bandwidth connection, GPU inference, 500ms latency

System T' (Compressed system):

- Compressed feature spaces: blood ($\mathbb{R}^{32}$), imaging ($\mathbb{R}^{64}$), symptoms ($\mathbb{R}^{16}$), genetic ($\mathbb{R}^{128}$)
- Sparse translator graph (only most-reliable translators retained)
- Runs on CPU, low bandwidth, 50ms latency

The morphism $f : T \to T'$:
A rural clinic with limited connectivity deploys the compressed system. The morphism f maps patients' representations from the full system (used at the main hospital) to the compressed system (used at the rural clinic):

1. **Semantic compatibility**: A patient diagnosed with lupus flare at the main hospital (using full system T) should still be recognized as having lupus flare when their data is processed by the rural clinic's compressed system (T'):

$$d_{T'}\big(f(\pi_T(\text{lupus flare})),\, \pi_{T'}(\text{lupus flare})\big) \lesssim 0.05.$$

2. **Metric control**: Compression inevitably loses some information. Distances between disease states in the compressed system are looser:

$$0.7 \cdot d_T(x, y) - 0.1 \leq d_{T'}(f(x), f(y)) \leq 1.2 \cdot d_T(x, y) + 0.1.$$

The compressed system has $L_{\min} = 0.7$ (some distances shrink due to loss of detail) and larger additive slack $\varepsilon_f = 0.1$ (more uncertainty).

Decision rule adaptation in this constructed example. Suppose the reference rule accepts a match when $d_T \leq 0.3$. Write $d' = d_{T'}(f(x), f(y))$. The two bounds imply

$$d' \leq 0.11 \implies d_T \leq 0.3, \qquad d' > 0.46 \implies d_T > 0.3.$$

Indeed, the lower inequality gives $d_T \leq (d' + 0.1)/0.7$, and the upper gives $d_T \geq (d' - 0.1)/1.2$. The interval $0.11 < d' \leq 0.46$ is unresolved by these bounds. A distance of 0.25 can therefore require abstention or an additional calibrated model; it is not a transferred confidence threshold.

Think of comparing a ruler with a less precise replacement. Well inside the acceptance region, both measurements support the same decision. Near its edge, the uncertainty interval crosses the line. The bounds tell us where a decision is justified and where the replacement measurement leaves a question open. Clinical confidence would additionally require an appropriate labeled population and validation; the example supplies only the distance calculation.

Properties of morphisms. In each example, the morphism $f : T \to T'$:

- **Preserves identity**: Same disease state in T maps to same disease state in T' (semantic compatibility)
- **Controls distortion**: Distance changes obey the declared lower factor, upper factor and additive slack (metric control)
- **Is auditable**: A shared validation cohort gives sampled distortions and semantic residuals. Uniform constants need a domain-wide argument; population guarantees need a justified sampling model.
- **Supports threshold transfer**: Knowing $L_{\min}, L_{\max}, \varepsilon_f$ lets us adapt decision thresholds when migrating from T to T'

Composition and the category structure. If we have three systems—say, 2018 lab panel (T), 2024 lab panel (T'), and 2024 panel compressed for telemedicine (T'')—and morphisms $f : T \to T'$ (lab upgrade) and $g : T' \to T''$ (compression), we can compose them:

$$g \circ f : T \to T''$$

The composed morphism tracks how a 2018 representation maps all the way to the 2024 compressed system. Write $a_f = L_{\min}^{(f)}$, $b_f = L_{\max}^{(f)}$, and similarly for g. Substituting the bound for f into the bound for g gives

$$\begin{aligned} d_{T''}(gf(x), gf(y)) &\geq a_g a_f d_T(x, y) - (a_g \varepsilon_f + \varepsilon_g), \\ d_{T''}(gf(x), gf(y)) &\leq b_g b_f d_T(x, y) + (b_g \varepsilon_f + \varepsilon_g). \end{aligned}$$

Thus valid factors are $a_g a_f$ and $b_g b_f$, and a common additive allowance is

$$\varepsilon_{gf} = b_g \varepsilon_f + \varepsilon_g.$$

The later map acts on the earlier error. If the first conversion can miss by 0.1 and the second magnifies distances tenfold, the first error can become 1, before adding the second map's own allowance.

Semantic allowances also compose: the triangle inequality gives a valid allowance $b_g \eta_f + \varepsilon_g + \eta_g$. Smooth maps compose, and their image domains must fit the next map's domain. To state the category axioms, we allow all finite declared certificates; a fixed numerical tolerance need not be closed under composition. The arrows are the underlying maps, with bounds recorded for their audit. Associativity and identities then hold exactly for maps. If bounds are supplied only for selected pairs or scales, the composite claim is restricted to pairs whose intermediate images satisfy the next bound.

What morphisms let us do. The morphisms in $\mathsf{SamTrans}(S)$ formalize:

- **System upgrades**: How to migrate from old to new implementations while preserving semantics
- **Multi-site deployments**: How to relate systems at different hospitals with different equipment
- **Compression/adaptation**: How to deploy lighter-weight versions while tracking what's lost
- **Threshold recalibration**: How decision boundaries must change when moving between realizations

Chapter 20 transfers a false-positive upper bound α through a controlled comparison with an additional allowance for geometric failure and threshold-band mass. Every probability is conditional on the negative class. The morphism supplies the distance bound; the population supplies the mass that can cross the decision boundary.

Morphisms: structure-preserving maps between geometric realizations

A morphism between geometry-based realizations should preserve both semantics and geometry up to controlled distortion.

Definition 13.3 (Morphisms in $\mathsf{SamGeom}(S)$)**.** Given two geometry-based realizations $G = (M, g, \mathcal{R}_M, \pi_M)$ and $G' = (M', g', \mathcal{R}_{M'}, \pi_{M'})$ of S, a *morphism* $h : G \to G'$ in $\mathsf{SamGeom}(S)$ consists of a smooth map

$$h : \mathcal{R}_M \to \mathcal{R}_{M'}$$

such that:

1. **Semantic compatibility:** For all latent entities u in the regime of interest,

   $$d_{M'}\big(h(\pi_M(u)),\, \pi_{M'}(u)\big)$$

   is bounded by a declared finite allowance η_h.
2. **Metric control:** h has a bi-Lipschitz comparison with additive slack on $\mathcal{R}_M$, in the sense that there exist finite $0 < L_{\min} \le L_{\max}$ and $\varepsilon_h \ge 0$ such that

   $$L_{\min}\, d_M(x, y) - \varepsilon_h \;\le\; d_{M'}\big(h(x), h(y)\big) \;\le\; L_{\max}\, d_M(x, y) + \varepsilon_h$$

   for all $x, y \in \mathcal{R}_M$. Bounds on selected scales remain useful comparison certificates when each successive map is evaluated on its declared pairs.

Concrete examples of morphisms in $\mathsf{SamGeom}(S)$*.* To make geometric morphisms tangible, let's examine several scenarios where we have multiple Riemannian realizations of the same medical diagnosis structure S and need to relate them.

Example 1: A flat chart and a spherical model.

Use two normalized marker coordinates $x = (x_1, x_2)$ in a hypothetical comparison. They can index a model state without being claimed to determine a clinical diagnosis. Take the flat disk $B_r \subset \mathbb{R}^2$ and the radius-R sphere. The inverse stereographic map, scaled to have unit length factor at the origin, is

$$h(x) = \frac{1}{1 + \|x\|^2/(4R^2)} \left(x_1, x_2, R\left[1 - \frac{\|x\|^2}{4R^2}\right] \right).$$

Its image lies on the sphere because the squared norm of the displayed numerator equals $R^2(1 + \|x\|^2/(4R^2))^2$. The origin maps to the north pole. Both spaces are two-dimensional; the third ambient coordinate locates the sphere in space.

For an entity u with planar representation $\pi_M(u)$, define its spherical representation by $\pi_{M'}(u) = h(\pi_M(u))$. This supplies exact semantic compatibility for the chosen model. It does not show that marker coordinates are a sufficient representation of every entity.

Differentiating h gives the pullback round metric

$$h^* g_{S^2} = \lambda(x)^2 I, \qquad \lambda(x) = \frac{1}{1 + \|x\|^2/(4R^2)}.$$

Consequently, for a curve contained in B_r,

$$\lambda_r L_{\text{flat}}(\gamma) \le L_{S^2}(h\gamma) \le L_{\text{flat}}(\gamma), \qquad \lambda_r = \frac{1}{1 + r^2/(4R^2)}.$$

The disk is convex, so its intrinsic flat distance is Euclidean distance. Taking infima over the corresponding curves proves

$$\lambda_r \|x - y\| \le d_{h(B_r)}(h(x), h(y)) \le \|x - y\|.$$

Here the spherical paths stay inside $h(B_r)$. If $r < 2R$, that image is a cap smaller than a hemisphere and is geodesically convex, so the shorter spherical arc stays in the cap as well. The intrinsic and ambient spherical distances then agree.

For $R = 5$ and $r = 1$, the lower factor is $100/101 \approx 0.9901$. For $r = 2$, it is $25/26 \approx 0.9615$. These constants follow from the map; they are not estimates from a patient cohort. A pair separated by 1.5 in the second disk has spherical distance between 1.4423 and 1.5. Its classification can change only if the threshold lies within the corresponding uncertainty interval.

This is the map/globe lesson in a complete calculation. Keep the same entity assignments, specify the actual conversion, carry its metric and derive the effect on distances. A bounded spherical model is a modeling choice; the radius alone establishes no biological severity law.

Example 2: Dimension reduction - full manifold vs. 2D projection.

Geometric realization G (Full 5D manifold):

A comprehensive model uses five dimensions:

- Manifold M: A 5-dimensional Riemannian manifold embedded in $\mathbb{R}^8$
- Dimensions: inflammation intensity, duration, systemic spread, immune response profile, genetic risk score
- In-regime region $\mathcal{R}_M$: A 5D submanifold covering common inflammatory conditions
- π_M(Crohn's) lives in a 5D region with specific genetic risk profile

Geometric realization G' (Clinical 2D summary):

For clinical use, physicians want a simple 2D visualization:

- Manifold M': A 2-dimensional Riemannian manifold (the "clinical summary space")
- Dimensions: composite severity score (axis 1), treatment urgency (axis 2)
- In-regime region $\mathcal{R}_{M'}$: A 2D region where most patients land
- $\pi_{M'}$(Crohn's) is the 2D projection of the 5D Crohn's region

The morphism $h : G \to G'$:

A learned dimensionality reduction maps the full five-dimensional model to the two-dimensional summary. This is a hypothetical visualization example. A smooth map into two dimensions has differential rank at most two, so it cannot preserve a positive lower infinitesimal length bound on all five independent directions. A proposed lower distance bound must therefore specify a restricted set, resolution scale or additive slack. For illustration, assume the following inequalities have been established for the selected pairs:

1. **Semantic compatibility**: For the chosen model assignments, require semantic compatibility:

$$h(\pi_M(\text{Crohn's})) = \pi_{M'}(\text{Crohn's}),$$
$$h(\pi_M(\text{lupus})) = \pi_{M'}(\text{lupus}).$$

2. **Metric control**: Dimension reduction inevitably distorts distances. Some disease states that were far apart in 5D might appear close in 2D projection:

$$0.4 \cdot d_M(p,q) - 0.2 \leq d_{M'}(h(p), h(q)) \leq 0.9 \cdot d_M(p,q) + 0.1.$$

Constants: $L_{\min} = 0.4$ (substantial compression due to dimensionality loss), $L_{\max} = 0.9$ (projection doesn't magnify distances much), $\varepsilon_h = 0.2$ (additive slack from local distortions).

Clinical implication: The 2D projection is useful for visualization but loses information. Two patients who appear close in 2D might actually be far apart in the full 5D space (different genetic risk, different immune profiles). The morphism's metric bounds quantify this:

- If two model points have $d_{M'} = 0.5$, the displayed inequalities imply $4/9 \leq d_M \leq 1.75$
- The 2D visualization should not be used for fine-grained clinical decisions where those distinctions matter
- Any decision use needs validation of the information and uncertainty retained by the selected representation

Example 3: Reparametrization - different chart choices for the same manifold.

Assume in this constructed example that each measurement triple is a smooth, locally invertible coordinate system on the same region, and that the metric is transformed by the coordinate rule. Actual immune and symptom measurements need not satisfy that assumption. It is precisely the hypothesis that distinguishes a coordinate change from a lossy observation.

Geometric realization G (Immunological coordinates):

Disease states parametrized by immune response markers:

- Manifold M: A 3D submanifold of $\mathbb{R}^5$
- Coordinates: (c_1, c_2, c_3) = (T-cell activation, B-cell activity, cytokine levels)
- Metric g: Riemannian metric reflecting how immune states vary
- π_M(Crohn's flare) at immune coordinates $(1.5, 0.8, 2.1)$

Geometric realization G' (Clinical symptom coordinates):

Same disease states, but parametrized by observable symptoms:

- Manifold $M' = M$ (same underlying manifold!)
- Coordinates: (c'_1, c'_2, c'_3) = (pain score, inflammation biomarker, functional impairment)
- Metric g': Pullback of g under the coordinate change
- $\pi_{M'}$(Crohn's flare) at symptom coordinates $(7.2, 4.5, 3.1)$

The morphism $h : G \to G'$:

This is a **reparametrization**: the map $h : M \to M'$ is the identity map on the underlying manifold, but coordinates change.

1. **Semantic compatibility**: Trivially satisfied—same points, same disease states:

$$h(\pi_M(u)) = \pi_{M'}(u) \text{ exactly.}$$

2. **Metric control**: Since it's a reparametrization of the same Riemannian manifold, geodesic distances are preserved exactly:

$$d_{M'}(h(p), h(q)) = d_M(p,q) \quad \text{for all } p, q \in M.$$

Here $L_{\min} = L_{\max} = 1$, $\varepsilon_h = 0$. The morphism is an isometry.

Why this matters: Within the constructed locally invertible model, the transformed metric makes the two coordinate calculations agree. Applying that conclusion to actual immune and symptom observations would require establishing the assumed correspondence and the information it retains; the coordinate theorem alone does not supply that clinical evidence.

Properties and composition. In all three examples:

- **Semantic compatibility** ensures disease identities are preserved: h maps Crohn's to Crohn's, lupus to lupus
- **Metric control** quantifies geometric distortion: $(L_{\min}, L_{\max}, \varepsilon_h)$ bound how much distances change
- **Smoothness** supplies derivatives of every order. It does not prevent rank loss: $h(x) = x^3$ is smooth but has derivative zero at the origin. A nonsingular coordinate change needs a full-rank derivative and a smooth local inverse.

If we have three geometric realizations—say, full 5D model (G), spherical 3D model (G'), and 2D clinical projection (G'')—with morphisms:

$$h_1 : G \to G' \quad \text{(5D} \to \text{3D spherical)},$$
$$h_2 : G' \to G'' \quad \text{(3D} \to \text{2D projection)},$$

we can compose:

$$h_2 \circ h_1 : G \to G''.$$

The multiplicative factors multiply. Earlier additive errors are multiplied by later distortion factors, as shown in the composition calculation at the beginning of this chapter. Those bounds apply to the specified pairs and scales, not automatically to every point of a dimension-reducing map.

Why morphisms in $\mathsf{SamGeom}(S)$ *matter.* Geometric morphisms formalize:

- **Model comparisons**: When two research groups use different geometric models of the same disease space, morphisms let us compare their results and transfer findings between models
- **Simplification for deployment**: Full high-dimensional models may be too complex for clinical use. Morphisms to lower-dimensional realizations quantify what's lost in simplification
- **Coordinate changes**: Different parametrizations (immunological vs. clinical, Cartesian vs. spherical) are related by isometric morphisms, confirming they're equivalent descriptions
- **Threshold transfer**: For a reference acceptance threshold τ_M, accept certainly when $d_{M'} \leq L_{\min}\tau_M - \varepsilon_h$ and reject certainly when $d_{M'} > L_{\max}\tau_M + \varepsilon_h$. The interval between is unresolved by the comparison; if the sufficient acceptance bound is negative, it admits no nonnegative distance.

In Chapter 20, error budgets will transfer along these morphisms: if G has false positive rate α, then the compared rule on G' has a false-positive bound obtained from α, the geometry-failure probability and the boundary-band mass, all conditioned on the negative class.

Again, composition and identities are given by composition and identity of maps between in-regime regions. With these, the geometry-based realizations of S and their morphisms form a category $\mathsf{SamGeom}(S)$.

We can now name a complete implementation and the maps taking it to another implementation. The next test returns to the engineer replacing a representation: after going into the new description and back, how far can a reconstructed observation or comparison have moved? That is the practical content of the round-trip conditions.

Approximate Equivalence of Realizations

On the circular track, the translator route and the surface ruler measured the same arc. More generally, F_S sends a chosen translator realization to geometry and G_S returns geometry to a translator realization. We now ask what their round trips preserve within the stated region.

Round-tripping realizations

Given a translator-based realization T in $\mathsf{SamTrans}(S)$, we can:

- Apply F_S to get a geometry-based realization $G_T = F_S(T)$.
- Apply G_S to G_T to get a new translator-based realization $T' = G_S(G_T)$.

Similarly, given a geometry-based realization G in $\mathsf{SamGeom}(S)$, we can:

- Apply G_S to get a translator-based realization $T_G = G_S(G)$.
- Apply F_S to T_G to get a new geometry-based realization $G' = F_S(T_G)$.

For the chosen diffeomorphic realization Φ_T of Chapter 12, let Q_{G_T} be the exact isometry from the atlas quotient of T' to the manifold of G_T. The forward map $Q_{G_T}^{-1}\Phi_T$ and reverse map $\Phi_T^{-1}Q_{G_T}$ are inverses on the designated regions. Their distance bounds come from the chosen realization and its inverse. In the other direction, choose Q_G as the realization of T_G and the round trip recovers the given geometry G on that region.

For an approximate realization, two separate comparisons need evidence. A forward distance bound with additive slack can send distinct points to the same image and need not cover the whole target. If the task requires a return, supply a reconstruction map with its own domain, distance bounds and round-trip allowances in both directions. An ε-equivalence of realizations records those maps and checks. Decision consequences also require the threshold margins and population assumptions of Chapter 20.

Equivalence of realizations (informal definition)

For practical purposes, we say that a translator-based realization T and a geometry-based realization G of S are *ε-equivalent realizations* if there exist maps

$$\Phi : \mathcal{R}_T \to \mathcal{R}_G, \qquad \Psi : \mathcal{R}_G \to \mathcal{R}_T$$

such that:

- Both Φ and Ψ are semantically compatible: they carry the representation of each latent entity u in one realization to the representation of u in the other, up to distance at most ε.
- Both are approximate isometries on the scales of interest: distances are distorted by at most small multiplicative factors and additive slack bounded by ε.
- The round-trips $\Psi \circ \Phi$ and $\Phi \circ \Psi$ move points by at most ε in the respective metrics on $\mathcal{R}_T$ and $\mathcal{R}_G$.

The exact constructions give explicit maps when the chosen realization is invertible on its image. In a lossy setting, Φ and Ψ are additional inputs. A comparison on selected scales must keep that scope when a later argument uses its bounds.

In Chapter 20, we will see that ε-equivalence of realizations implies *equivalence of detection error budgets*: if a detection system built on T has certain error guarantees, then a detection system built on G will have the bounds supplied by the transfer theorem when its comparison and population hypotheses hold, and conversely for the reverse comparison.

A concrete object and its two descriptions

Use the aircraft orientation space $SO(3)$ with its metric normalized by physical angle. Quaternion representatives form its two-sheeted cover. On local neighborhoods one can choose smooth representatives and chart transitions; their exact path construction returns the same rotation-space distance. This supplies a fully specified geometric object, translator description and comparison.

For telemetry, an exact invertible change of calibrated continuous coordinates supplies another instance. A lossless encoding supplies an exact record round trip but becomes a smooth chart only after a smooth domain and metric have been defined. The AMC and GIGI worked chapters make these different operations explicit.

What the construction preserves

The maps between realizations let us follow one comparison through a system change. In the chosen smooth categories, identities and composition are exact operations on maps. Distance distortion and reconstruction error have their separately stated bounds. The two directions remain

$$F_S : \mathsf{SamTrans}(S) \to \mathsf{SamGeom}(S),$$
$$G_S : \mathsf{SamGeom}(S) \to \mathsf{SamTrans}(S),$$

with the realization and domain conditions specified in the constructions.

The composition bound carries error through each intermediate map. The predictive example shows why a changing state also requires the order of its observations. Chapter 20 will combine the distance bound with a decision rule and its population.

For a replacement, identify the claim to check: an observation's identity, a bounded comparison, or an ordered update. Next we examine an exact identification in which two representatives belong to one geometric state.

Practice with the construction

Back to the maps and the globe. When a route changes maps, carry each conversion error through the maps that follow it.

13.1. Two maps have upper bounds $(b_1, \eta_1) = (1.02, 0.01)$ and $(b_2, \eta_2) = (1.03, 0.02)$. Compute the composite upper bound and explain its units.

13.2. A reconstruction moves each endpoint by at most 0.004. How far from a single threshold must a reference distance be to guarantee its classification survives that round trip?

Part IV

Double cover and transported sameness

Chapter 14

Symmetry, quotients and two sheets

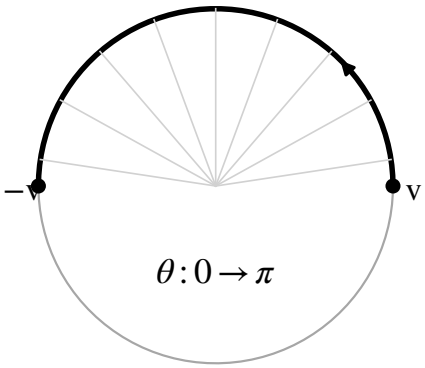
The realization maps told us how descriptions correspond. They also force a sharper question: can distinct descriptions represent exactly the same state? A bar viewed without a marked end makes that identification visible. We calculate which distances and observations survive it, and which information a record of motion must retain.

Remove the arrowhead

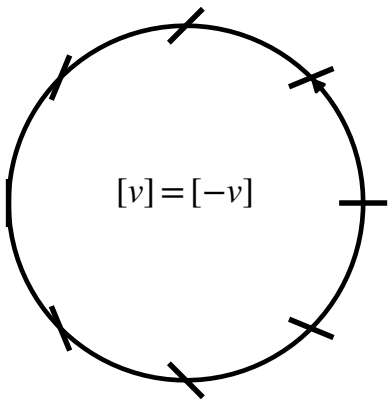
Place a straight bar on a turntable under an overhead camera. Paint an arrow along the bar so the picture shows which end is marked. Now cover the paint. The camera can still measure the line the bar lies on. A half-turn brings that line back to its starting orientation and swaps the ends.

Represent a planar unit arrow by $u(\theta) = (\cos\theta, \sin\theta)$. Its unoriented axis identifies u and $-u$, so θ and $\theta + \pi$ describe the same line. The quotient is the real projective line $\mathbb{RP}^1$. The projection $S^1 \to \mathbb{RP}^1$ has exactly two representatives over each line. Small neighborhoods separate into two disjoint arcs, each mapping homeomorphically onto the same neighborhood of lines. That is a double cover.

A *The lifted arrow changes sign*



B *The physical bar returns*



$S^1 \longrightarrow \mathbb{RP}^1, \qquad v \longmapsto \{v, -v\}$

The base loop closes; its continuous lift ends on the other sheet.

Figure 14.1: The two-sheet cover of unoriented axes. The left path travels from v to $-v$ on the representative circle. Its image on $\mathbb{RP}^1$ completes a loop, drawn using the circle coordinate 2θ. Small line segments at sampled base points show their unoriented axes.

The observable must respect the identification. Relative to a fixed unit arrow a, the squared inner product $(a \cdot u)^2$ is unchanged by $u \mapsto -u$. The signed inner product $a \cdot u$ is not. It therefore answers a question about an oriented representative, not just the underlying line.

If θ is the arrow angle relative to a, the invariant score is $\cos^2 \theta$. The expression $\cos^2(\theta/2)$ changes under $\theta \mapsto \theta + \pi$ and is not an observable of this line quotient. In the aircraft example the half-angle belongs to a different map and a different angle convention. Writing the domain and the covering map keeps both constructions intelligible.

A lifted journey remembers the sheet

Turn the bar through a half-turn while tracking the painted arrow. Its unmarked axis returns to its starting orientation. The arrow now points in the opposite direction. The endpoint on the base has returned; the endpoint on the cover has changed sheet.

This is not a failure of identity of the base object. It is retained information about the path. If an instrument reports only the axis, the two endpoints agree. If an experiment continuously tracks the arrow, it also knows which lift was followed. The measurement protocol determines which question the record can answer.

For a general covering $p : \widetilde{M} \to M$, a continuous path in M has a unique lift after choosing its initial representative. One can construct it by dividing the path into intervals lying in evenly covered neighborhoods and using the appropriate inverse on each interval. The local inverses agree at the joining endpoints, which fixes the continuation. Standard covering theory develops the associated fundamental-group description [5].

A camera watching a rotating bar

An overhead camera sees the bar on the turntable with the paint covered. What angle should its software report when the same axis could be described by either of two arrows? We will give the axis a coordinate and derive its distance rule.

This is a useful place to slow down. The double cover is not the assertion that every object has exactly two descriptions. A street address can have many spellings. A point on a globe can belong to many maps. The special feature of the bar experiment is a specified symmetry with exactly two alternatives, acting without a fixed point on the unit-arrow circle. The mathematics describes that experiment precisely.

Write the arrow as a complex number $z = e^{i\alpha}$. A convenient circular coordinate for its axis is

$$w = z^2 = e^{2i\alpha}.$$

If the camera reports w, recovering z gives two possibilities. At $w = 1$ they are 1 and -1; at $w = i$ they are $e^{i\pi/4}$ and $-e^{i\pi/4}$. The choice cannot be made from one unmarked picture. An earlier marked picture can supply the missing initial choice, and sufficiently frequent subsequent observations can continue it.

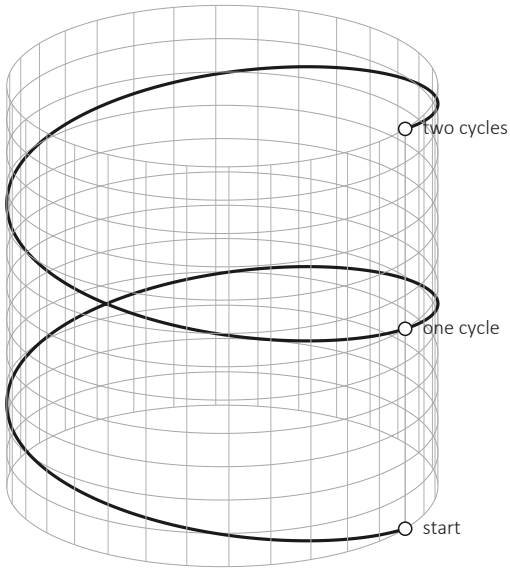
Proposition 14.1 (The metric belongs to the object)**.** *Equip the arrow circle with $d\alpha^2$. The quotient metric on the axis circle, in the coordinate $\beta = 2\alpha$, is $\frac{1}{4}d\beta^2$. Its circumference is π, and the projection is a local isometry.*

Proof. On an arc short enough to choose one inverse, $d\alpha = \frac{1}{2}d\beta$. Substitution into the arrow metric gives the stated metric. Integrating the quotient line element as β runs from 0 to 2π gives $\int_0^{2\pi} \frac{1}{2} d\beta = \pi$. The length of every sufficiently short lifted arc equals the length of its image. □

Drawing both circles with the same radius on paper does not give them the same geometric circumference. One circle in the drawing is a coordinate display. The factor 1/4 is the instruction that tells the ruler how to read it. This is the map/globe distinction in its smallest useful form: the ink on the map does not choose the metric of the object.

A return in the object can retain a history

Height counts completed object cycles; it is a history coordinate, not a physical axis.



Unoriented bar: u and −u

Display: (cos 2α, sin 2α)

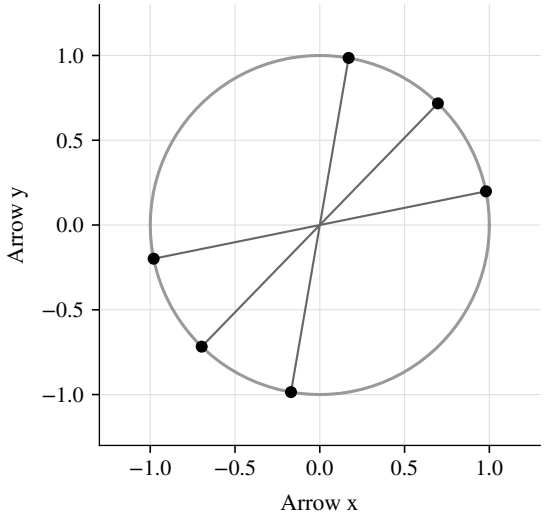
After one bar cycle: u → −u

After two: u → u



The physical bar is unchanged.

A Opposite arrows describe the same bar



B Doubling the angle removes the sign choice

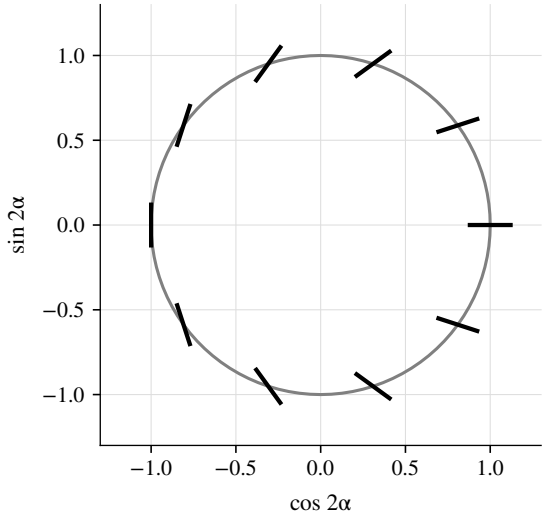
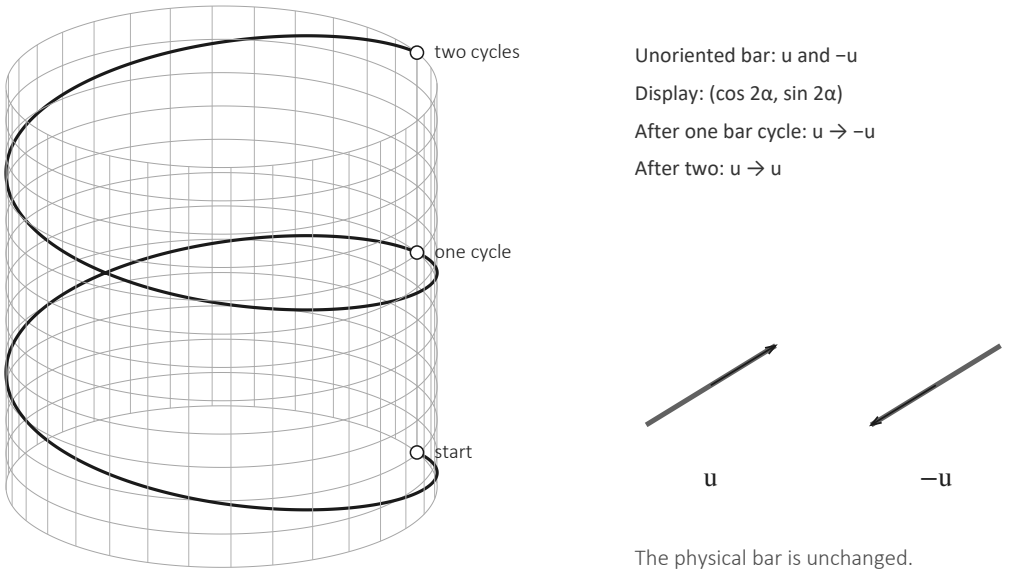


Figure 14.2: *The camera sees an axis; the record can retain a turn.* The study continues on the following figure page.

A return in the object can retain a history

Height counts completed object cycles; it is a history coordinate, not a physical axis.



C A physical half-turn reverses the arrow

D The quotient ruler preserves local length

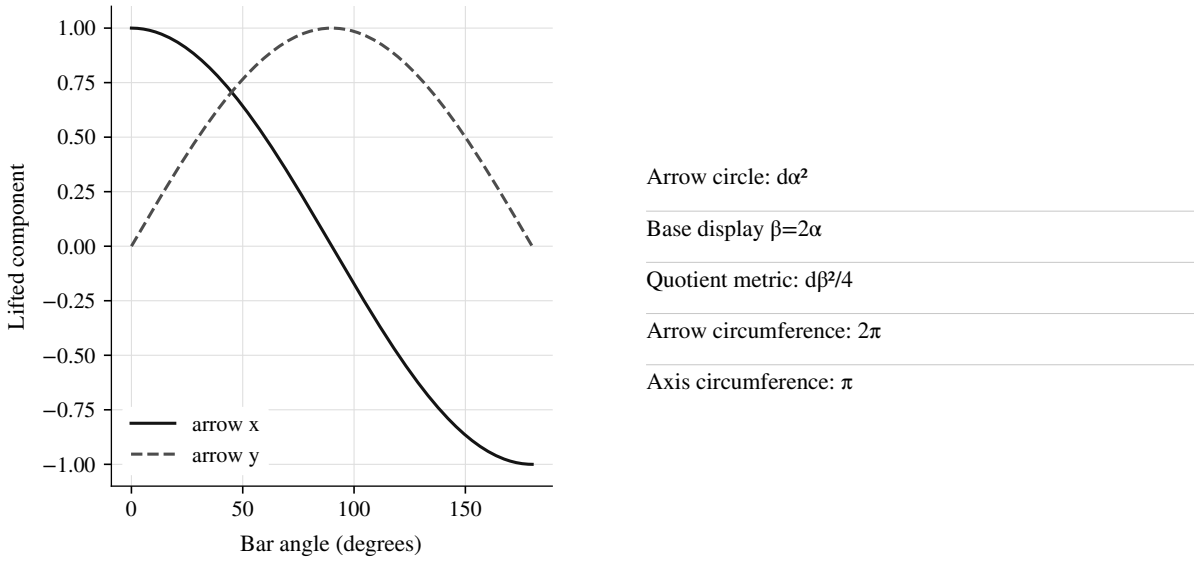


Figure 14.2: *The camera sees an axis; the record can retain a turn (continued).* The upper construction pairs antipodal arrows with the same observed axis. The lower calculation follows a continuous lift through one axis cycle, a physical half-turn of the bar. The base returns while the chosen representative changes sign. All angles are in radians unless degrees are explicitly printed.

A distance that ignores the paint

For two axes represented by unit vectors u, v, define

$$\delta([u],[v]) = \arccos |u^T v|, \qquad 0 \le \delta \le \pi/2.$$

This is the shorter angle between their lines. A bar at 10° and a bar at 170° differ by 20° as axes, although the particular arrows drawn on them differ by 160°. The absolute value performs the symmetry reduction before the inverse cosine measures the angle.

There is also an entirely matrix-based description. Replace the arrow by its rank-one orthogonal projector,

$$P_u = uu^T.$$

The matrix is the same for u and $-u$. It acts on any vector x by keeping the component along the bar: $P_u x = (u^T x)u$. For $u = (\cos\alpha, \sin\alpha)$,

$$P_u = \frac{1}{2}\begin{pmatrix} 1+\cos 2\alpha & \sin 2\alpha \\ \sin 2\alpha & 1-\cos 2\alpha \end{pmatrix}.$$

The doubled angle is now visible in actual entries of the observable. No choice of an arrowhead remains.

Expanding a Frobenius norm gives another invariant comparison:

$$\begin{aligned} \|P_u - P_v\|_F^2 &= \operatorname{tr}(P_u) + \operatorname{tr}(P_v) - 2\operatorname{tr}(P_u P_v) \\ &= 2\bigl(1 - (u^T v)^2\bigr) = 2\sin^2\delta. \end{aligned}$$

Thus the normalized projector discrepancy is $\sin\delta$. It orders pairs in the same way as the geodesic angle on $[0, \pi/2]$, but it is not the angle itself. At a right angle the discrepancy is one; the geodesic distance is $\pi/2$. Choosing a threshold in one convention and applying it in the other changes the decision.

Axis angle δ	Overlap $\cos^2\delta$	Projector discrepancy	Arc distance
0°	1	0	0
30°	$3/4$	$1/2$	$\pi/6$
45°	$1/2$	$1/\sqrt{2}$	$\pi/4$
60°	$1/4$	$\sqrt{3}/2$	$\pi/3$
90°	0	1	$\pi/2$

Which comparisons live downstairs?

Proposition 14.2 (Descent of an observable)**.** *Let $p : X \to Y$ be a surjection. A function $f : X \to \mathbb{R}$ can be written $f = \bar{f} \circ p$ for a unique function $\bar{f} : Y \to \mathbb{R}$ if and only if f is constant on every fiber $p^{-1}(y)$. If p is a quotient map and f is continuous, the resulting $\bar{f}$ is continuous.*

Proof. Necessity follows because equal images under p have equal values under $\bar{f}$. For sufficiency, set $\bar{f}(y) = f(x)$ for any x over y; constancy on the fiber makes the definition independent of the representative. Surjectivity gives uniqueness. The continuity statement is the defining property of the quotient topology. □

Try it with the bar. Turn it end for end and take the same reading again. The bar still lies along the same line, but the arrow points the other way. If the score changes, it's picking up something about the arrow that the line alone can't tell you.

Which harmonics survive the missing arrowhead?

Suppose a sensor records a signal throughout one complete turn of the marked bar. Now retain only the axis seen by the camera with the paint covered. Angles a half-turn apart correspond to the same camera observation. Which parts of the signal can still describe the unmarked bar? A peak that becomes a trough after a half-turn cannot belong to the axis alone.

I read and translate old Soviet mathematics as a hobby, working through constructions in my *Zadachnik* notebook [37]. Harmonics are a particular pleasure. Vygodsky's line-angle discussion and his treatment of Fourier orthogonality offer a useful meeting place: a line repeats after a half-turn, while a signal can remember a whole turn [36, §21, pp. 34–36; §§413–414, pp. 605–609].

Let θ be the painted end's angle from a fixed horizontal reference. For a square-integrable signal on the full circle, use the uniform angular mean

$$\|f\|_2^2 = \frac{1}{2\pi}\int_0^{2\pi} |f(\theta)|^2\, d\theta, \qquad (Uf)(\theta) = f(\theta+\pi).$$

The operator U pairs photographs taken half a turn apart. An axis observable satisfies $Uf = f$. Split any signal into the pair's average and half-difference:

$$P_+ f = \frac{f+Uf}{2}, \qquad P_- f = \frac{f-Uf}{2}, \qquad f = P_+ f + P_- f. \tag{14.1}$$

The average repeats after a half-turn. The half-difference reverses sign. Averaging is an explicit operation on the readings; identifying the two bar ends does not itself prescribe a new sensor output.

Proposition 14.3 (The closest signal that belongs to the axis)**.** *For this uniform angular norm, $P_+ f$ is the unique best approximation, up to equality almost everywhere, among square-integrable axis observables. For every such h,*

$$\|f-h\|_2^2 = \|P_- f\|_2^2 + \|P_+ f - h\|_2^2.$$

Proof. Shifting both arguments of the inner product by π preserves its integral. If a repeats and b changes sign under that shift, then $\langle a, b\rangle = \langle a, -b\rangle = 0$. Apply this to $a = P_+ f - h$ and $b = P_- f$, then expand the squared norm. Its minimum occurs precisely when $h = P_+ f$ almost everywhere. □

Uniform weighting matters: unequal weighting of the two representatives generally changes the best average.

A worked comparison. Take the two readout models already used in this chapter,

$$f(\theta) = \cos^2(\theta/2) = \tfrac{1}{2} + \tfrac{1}{2}\cos\theta, \qquad g(\theta) = \cos^2\theta = \tfrac{1}{2} + \tfrac{1}{2}\cos 2\theta.$$

These specify mathematical signals, not measured intensity laws. For f, the retained part is $1/2$ and the discarded part is $\frac{1}{2}\cos\theta$. For g, everything survives. The mean of either cosine is zero and its mean square is $1/2$, so

Signal	Full $\|\cdot\|_2^2$	Retained	Discarded
$f = \cos^2(\theta/2)$	3/8	1/4	1/8
$g = \cos^2\theta$	3/8	3/8	0

Take two pictures of the bar a half-turn apart. The camera sees the same line both times, but the sensor signal f can be high in one picture and low in the other. If I need one reading for that line, the average fits best. What we drop accounts for one third of f's squared norm.

The Fourier explanation is immediate. Since $Ue^{ik\theta} = (-1)^k e^{ik\theta}$, even-index modes survive and odd-index modes disappear. For $f = \sum_{k\in\mathbb{Z}} c_k e^{ik\theta}$ in L^2, orthogonality gives

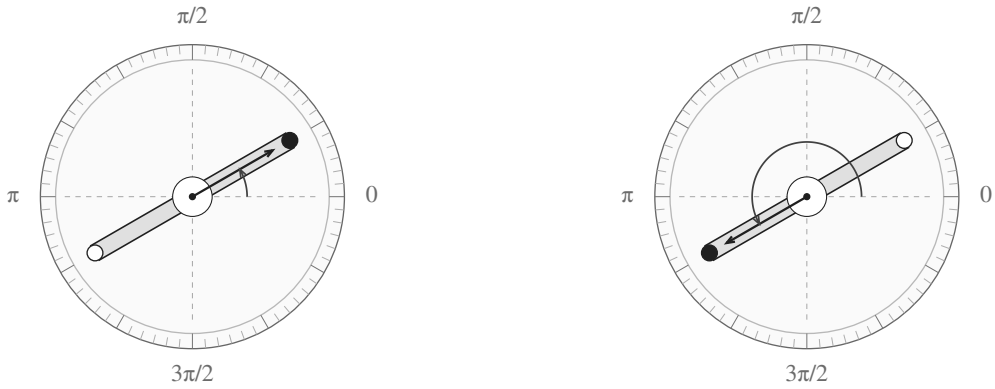
$$P_+ f = \sum_{k \text{ even}} c_k e^{ik\theta}, \qquad \|P_- f\|_2^2 = \sum_{k \text{ odd}} |c_k|^2.$$

Here "even" refers to the frequency index, not to reflection $\theta \mapsto -\theta$: $\sin 2\theta$ survives too. As a check, the signal $1 + \frac{1}{2}\cos\theta + \frac{1}{4}\sin 2\theta$ retains $1 + \frac{1}{4}\sin 2\theta$ and loses squared norm $1/8$.

DOUBLE COVER / FOURIER MODES

What the missing arrowhead costs

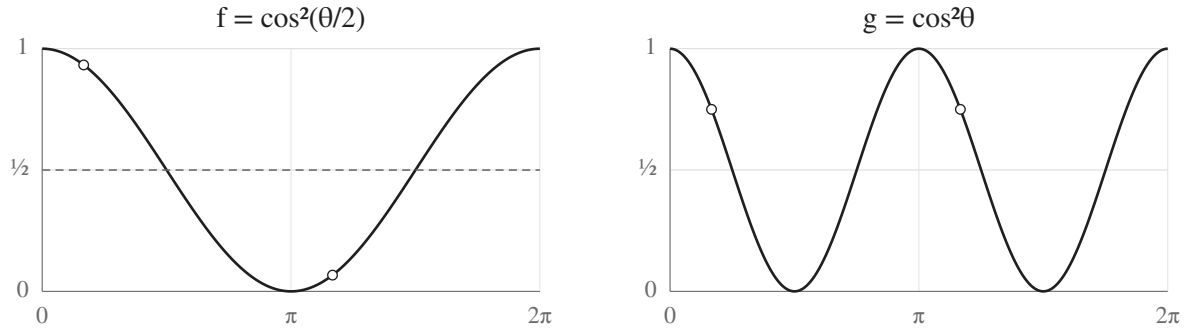
01 Pair readings half a turn apart



$f(\pi/6) \approx 0.933$ $f(7\pi/6) \approx 0.067$

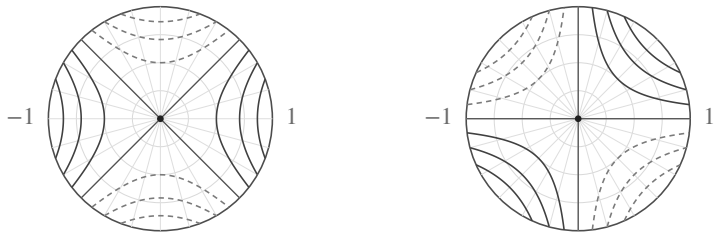
Same axis; mean reading = ½

02 The pairing selects the surviving signal



$P_+f = ½$; removed: ½ cosθ $P_+g = g$; removed: 0

03 The same modes inside the disk



$u = x^2 - y^2$ $v = 2xy$

$z^2 = u + iv$

$\Delta u = \Delta v = 0$

On r = 1:

cos 2θ, sin 2θ

Contours: ±¼, ±½, ±¾ and zero; dashed curves are negative.

Figure 14.3: *The axis retains the even-index harmonics.* The paired turntable positions represent the same unmarked bar. Solid curves show the readouts; the half-turn pairing determines their retained and discarded components. The bottom construction links the surviving angular mode to a harmonic field on the disk.

There is a direct bridge to the harmonic fields of Chapter 24. For integer $k \geq 1$, the real and imaginary parts of $z^k = r^k e^{ik\theta}$ are harmonic partners: their planar Laplacians vanish. On a fixed-radius circle they become Fourier modes. Replacing z by $-z$ preserves exactly the even powers. The two uses of "harmonic" meet in this calculation: angular oscillations form the boundary pattern of a field satisfying a differential equation. The geometric identification tells us which patterns describe the same object.

Why a continuous sign choice needs a history

Suppose the camera samples a bar rotating from 0 to 180°. A convenient display wraps its reported angle into $[-90°, 90°)$. The display therefore jumps from a value near 90° to one near −90°. The bar did not jump. The selected coordinate interval ended.

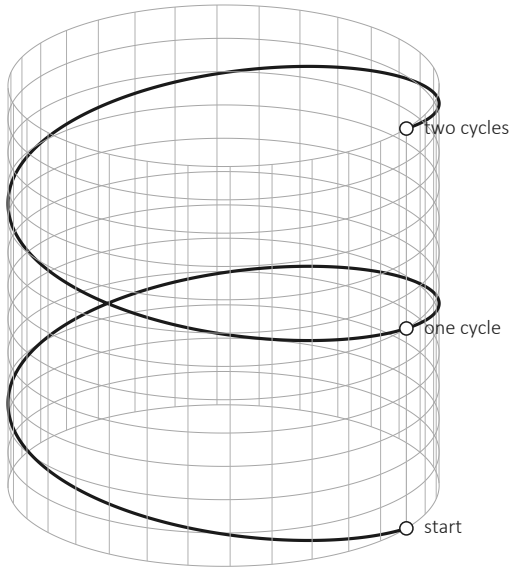
To reconstruct an arrow sequence, choose an initial unit vector u_0. For each subsequent observed axis, choose the representative u_k satisfying $u_{k-1}^T u_k > 0$. This chooses the smaller incremental arrow rotation. It works provided the actual incremental rotation is strictly less than 90° and the observation errors are small enough to leave that inequality unambiguous. At exactly 90° the two choices tie. With a faster unobserved motion, the nearest representative can follow the wrong lift.

That last condition is an information limit. A sequence of still images cannot recover arbitrary turns between exposures. The same problem occurs when a wheel seems to rotate backward in a film. The covering-space construction tells us what history can be retained; the sampling rate tells us whether that history was observed.

There can be no continuous rule selecting one arrow over every axis. If such a rule existed, following it around a complete axis loop would return to the same selected arrow. The unique lift starting at u_0 instead ends at $-u_0$. The contradiction is global. On any sufficiently small interval of axes, a smooth choice is easy.

A return in the object can retain a history

Height counts completed object cycles; it is a history coordinate, not a physical axis.



Unoriented bar: u and −u

Display: (cos 2α, sin 2α)

After one bar cycle: u → −u

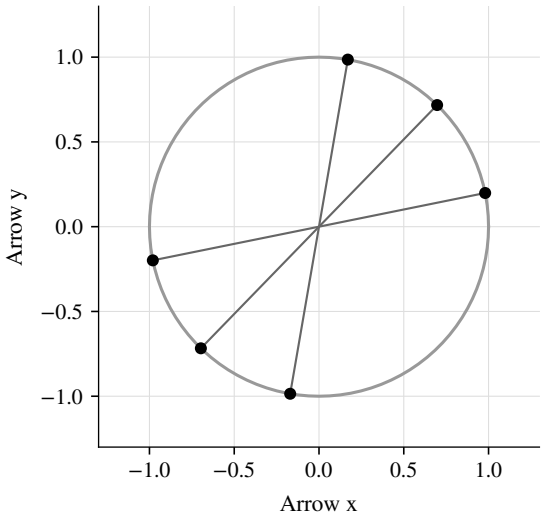
After two: u → u



u −u

The physical bar is unchanged.

A Two arrow signs represent one physical axis



B Three comparison rules give different scales

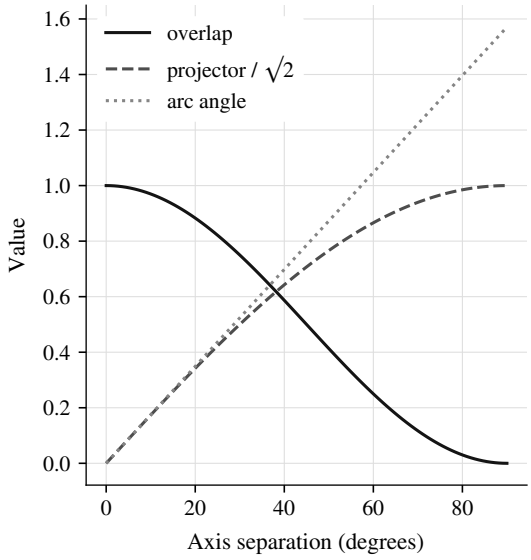
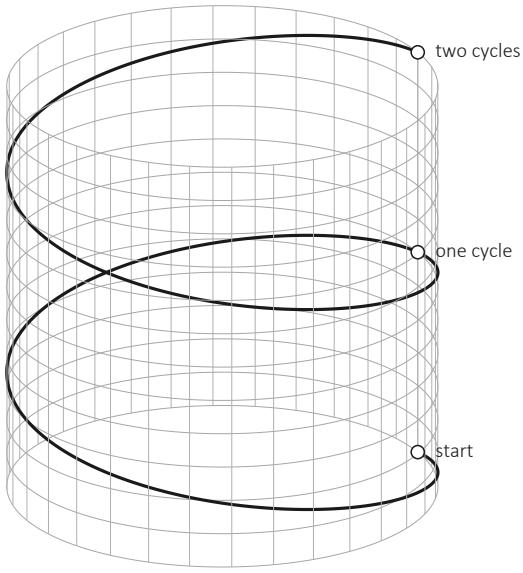


Figure 14.4: *One axis, three useful measurements.* The study continues on the following figure page.

A return in the object can retain a history

Height counts completed object cycles; it is a history coordinate, not a physical axis.



Unoriented bar: u and −u

Display: (cos 2α, sin 2α)

After one bar cycle: u → −u

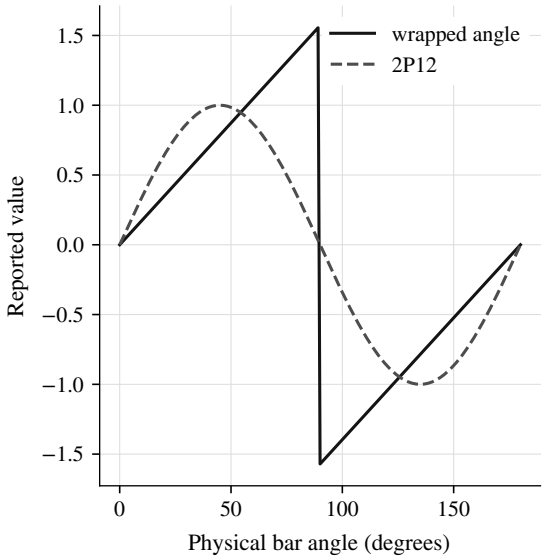
After two: u → u



The physical bar is unchanged.

C A coordinate jump need not be a physical jump

D The projector removes the sign ambiguity



$$Pu = uu^{T} = P{-}u$$

$$\mathrm{tr}(Pu\,Pv) = (u^{T}v)^{2}$$

$$\|Pu{-}Pv\|_F / \sqrt{2} = \sin\delta$$

$$\delta = \arccos|u^{T}v|$$

Figure 14.4: *One axis, three useful measurements (continued).* The projector entries, invariant overlap and shortest-axis angle are computed from the same rotating bar. The detail view identifies the branch cut of a chosen angle display. The cut belongs to the display; the axis and its projector remain continuous.

Two maps and two sheets are different kinds of information

The book's two descriptive routes remain translation and geometry. A translator says how to move a description into a form where a comparison can be made. A geometric realization says which objects those descriptions represent and what a permitted comparison measures. The bar supplies both: translate u into P_u, or measure the quotient angle directly. Their numerical agreement can be proved.

The two-sheet structure is an additional, exact fact about this example. It identifies the symmetry that a task is allowed to ignore. A translator that sends u to $-u$ preserves the axis. It does not preserve the painted end. Both statements can be true without weakening the thesis; they make the thesis operational by naming the object of sameness.

The later double-cover work develops the relation between representative spaces and invariant observables [19]. The bar makes that relation explicit on a circle. To carry it into three-dimensional attitude, we first need the local sheets and the history that tells us which representative to follow.

An overhead camera records the axis of a bar with indistinguishable ends. We will check when that stream of axes can be used to follow a particular end, once someone identifies the starting end.

Following a representative all the way around

Temporarily mark one end of the bar to fix the initial direction. Remove the mark from the camera's input before the bar moves. The recording now supplies an axis at each instant, and the initial mark tells us which of the two directions to follow.

The mathematical condition behind that continuation is stronger than saying there are two answers. A map $\pi : \widetilde{M} \to M$ is a covering map when each base point has a neighborhood U for which $\pi^{-1}(U)$ splits into disjoint open sets, each mapped homeomorphically onto U. Such a neighborhood is *evenly covered*. On a double cover there are exactly two local copies. Within either copy, a continuous local choice is unambiguous.

Construct the local copies for the bar

Write the bar angle as β modulo π, while its marked-end representative has angle θ modulo 2π. The projection sends θ to its class modulo π. Around any bar orientation choose an interval of width less than π. Its inverse image is two disjoint intervals separated by π in the representative circle. Each maps one-to-one to the original interval.

Choose a starting representative θ_0. A short observed motion stays in one evenly covered interval, so it has a unique continuation in the interval containing θ_0. When the path leaves that neighborhood, continue through an overlapping one. The representative already chosen on the overlap determines which sheet to use next.

For a continuous path on a compact parameter interval, finitely many such neighborhoods suffice after subdividing the interval. This constructs a lift. If two lifts start at the same representative, local uniqueness makes them agree on the first subinterval, then the second, and so on. They agree everywhere. This is the path-lifting argument in a form that directly describes the tracking experiment [5].

A closed observation need not close upstairs

Let the marked end turn through π. The bar configuration returns to its starting value. The lifted representative ends at the opposite point of the circle. After another half-turn, both the bar and its representative return.

The operation $D(\theta) = \theta + \pi$ exchanges the two representatives without changing the bar. It is a *deck transformation*: $\pi \circ D = \pi$. Applying it twice is the identity. The two possible starting representatives generate two lifts of the same observed path, related by D at every time.

This gives a concrete distinction between endpoint and path information. An endpoint photograph answers which bar configuration is present. A lifted history additionally answers how a chosen representative was continued. Neither record is a universal replacement for the other.

The airplane uses the same logic on a different space

For aircraft attitude, the representative is a unit quaternion and the base is $SO(3)$. The covering map sends q and $-q$ to the same rotation matrix. Around any sufficiently small rotation neighborhood, one can select one quaternion sign continuously. A continuous attitude path has a unique quaternion lift after its starting sign is chosen.

For a full physical turn about a fixed axis u,

$$q(t) = \big(\cos(\pi t),\, u\sin(\pi t)\big), \qquad 0 \le t \le 1.$$

The attitude returns to the identity, while the lift runs from 1 to -1. A second full turn closes the lift. This is an exact statement about the specified rotation path. It does not mean a rigid airplane needs two turns to recover its physical orientation; it already recovered that after one.

For more general closed attitude paths, the lifted endpoint is either the starting quaternion or its negative. Under a continuous deformation through closed attitude paths with fixed starting orientation, that endpoint sign cannot change continuously between two isolated choices. The sign therefore distinguishes at least two classes. It distinguishes exactly two here because S^3 is simply connected: two lifts with the same start and endpoint can be deformed into one another with endpoints fixed, and projecting that deformation gives one between the original attitude loops.

What the sign-continuity rule assumes

The common discrete rule selects the sign of q_{j+1} for which $q_j^T q_{j+1}$ is positive. It chooses the closer of the two representatives on S^3. If the actual physical rotation increment is known to be strictly less than π along the short local continuation, this agrees with the local lift.

At exactly π, the dot product is zero and the choice is tied. Above that sampling limit, the endpoint reports cannot distinguish a long physical motion from a shorter motion in the opposite direction. The continuous path-lifting theorem is not a theorem about arbitrary sparse samples: it assumes a continuous path has actually been supplied.

A practical record should therefore contain the sample times and the bound used to exclude hidden turns. Angular-rate measurements can help establish the continuation. They also introduce their own uncertainty. A tracking algorithm that silently chooses the nearest representative has made a motion assumption, whether or not its code comments say so.

Two descriptions and two sheets

The book's translation and geometric accounts are two methods of describing a sameness structure. They do not, merely by being two methods, create a two-sheeted covering map. In the bar and

aircraft constructions the double cover is supplied by a specified space, projection, and discrete fiber. That makes its consequences provable.

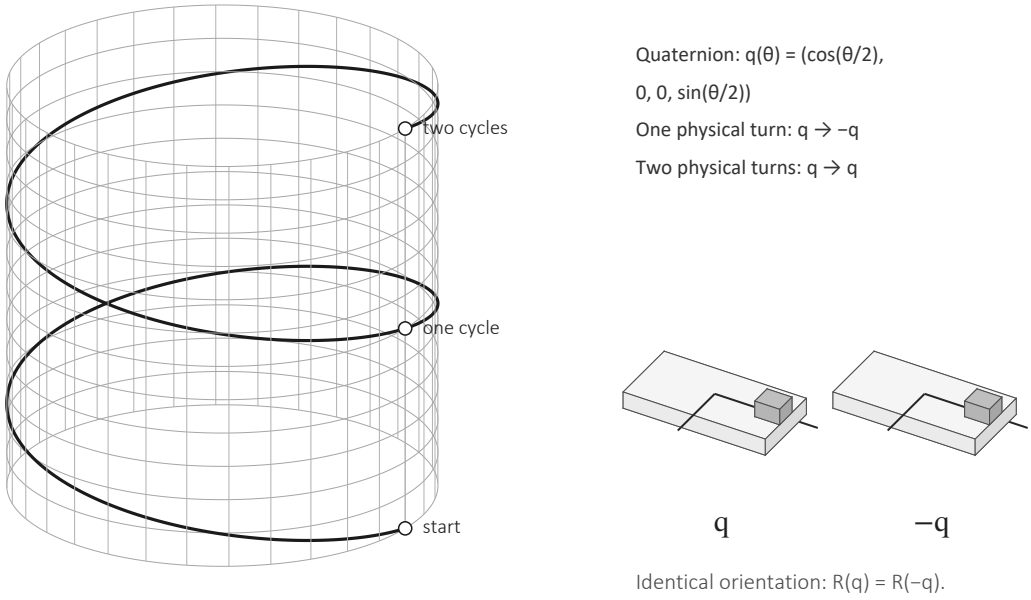
The distinction strengthens the organizing idea. A translator system can act on representatives while its comparison descends to the base object. The base distance ignores arbitrary sign, whereas a lifted path can preserve a history. The two descriptions cooperate because the projection and the preserved observables are explicit.

To compare bar, aircraft and atlas, identify the base, representative space, projection and fiber. State which observables ignore the representative and which history the lift retains. These data determine the shared construction.

A covering audit. For $f : S^1 \to S^1$, $f(e^{i\theta}) = e^{3i\theta}$, construct an evenly covered neighborhood, identify the three deck transformations, and lift one full base-circle turn from 1. Compare its endpoint with the double-cover examples. The same method works; the number of sheets is now three.

A return in the object can retain a history

Height counts completed object cycles; it is a history coordinate, not a physical axis.



A One physical turn changes the lifted sheet

B A continuous representative retains the history

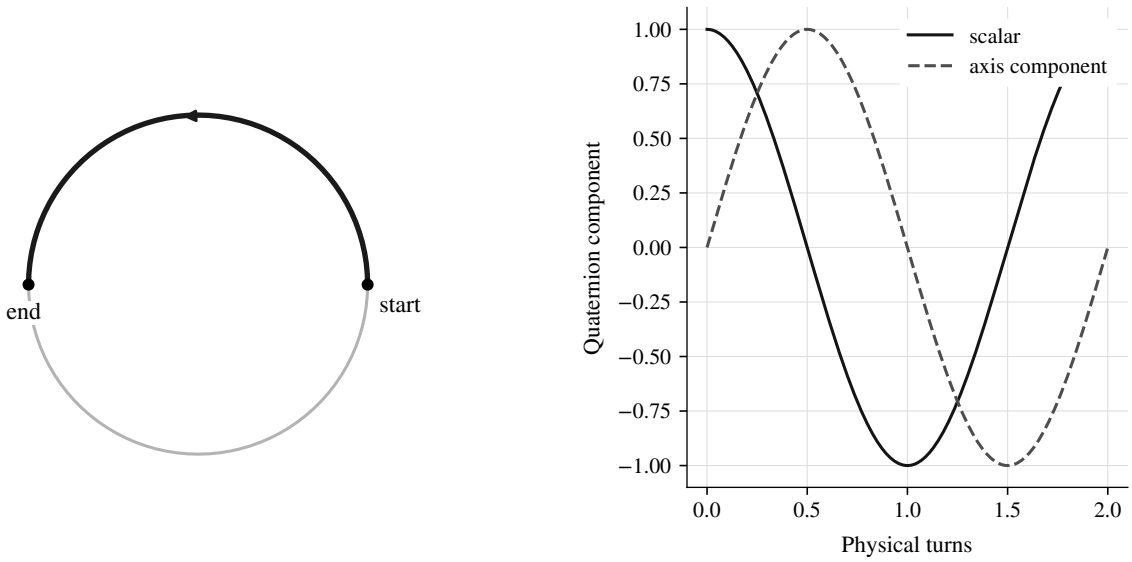
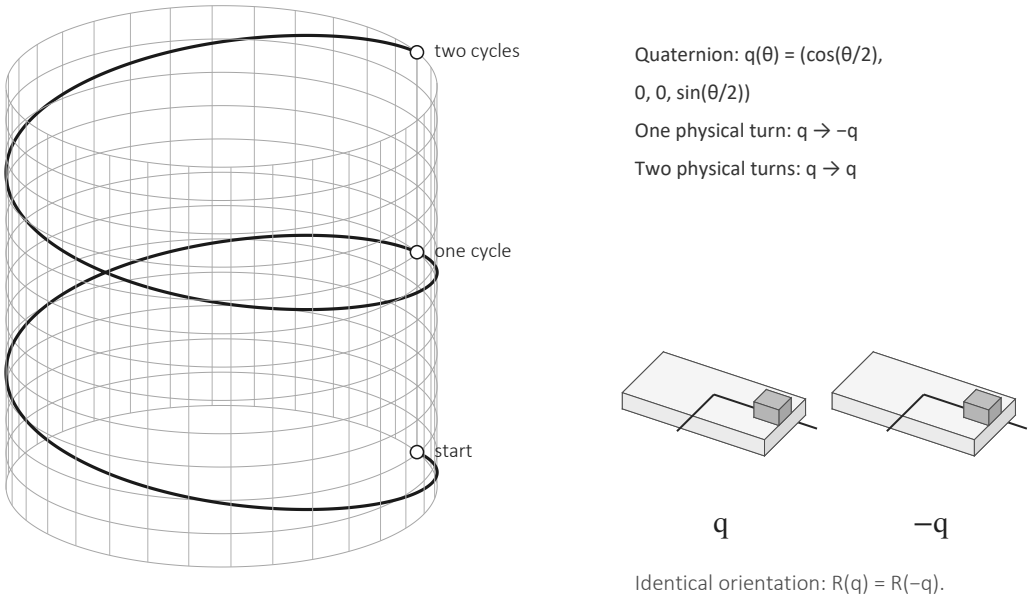


Figure 14.5: *The representative records a completed turn.* The study continues on the following figure page.

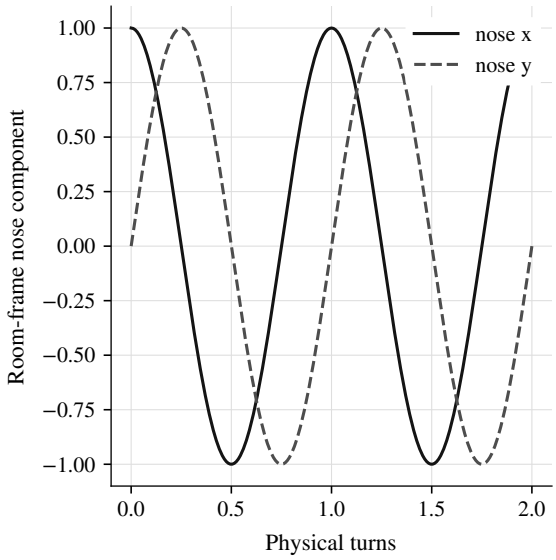
A return in the object can retain a history

Height counts completed object cycles; it is a history coordinate, not a physical axis.



C The physical attitude returns before the lift

D The represented object determines the comparison



After 1 turn: R = I; q = −1

After 2 turns: R = I; q = +1

Choose one initial sign.

Sparse endpoints alone do not reveal hidden turns.

Figure 14.5: *The representative records a completed turn (continued).* The unit-quaternion great-circle slice takes one physical turn from the initial representative to its negative. The room-frame nose direction has already returned. A second turn closes both histories.

The unmarked bar has one axis and two opposite arrow representatives. The quotient angle compares that axis consistently, while the harmonic calculation identifies precisely which signal components survive erasing the end labels. A continuous record of motion can retain the representative chosen at the start. The comparison must therefore say whether it concerns the visible axis or the recorded journey. We now ask the same question of an aircraft orientation.

Practice with the construction

Back to the maps and the globe. Two descriptions can locate one point on the globe while disagreeing about information retained on a covering space. We preserve the base object by using invariant observables, and preserve the path by keeping the lift when the task needs it.

14.1. For an arrow at angle $\pi/3$, compare $\cos^2\theta$ and $\cos^2(\theta/2)$ before and after adding π. Which descends to the line?

14.2. Describe what a sensor would have to record to distinguish the two lifted endpoints of the half-turn experiment.

Chapter 15

One orientation, two quaternion descriptions

Hold a model aircraft level above a table. Two attitude files describe this pose, but the four quaternion components in one file are the negatives of those in the other. Subtracting the numbers reports a difference even while the nose and wings stay fixed. As with the two arrows over one turntable axis, we need a comparison that respects the shared physical state.

This can happen when the numbers form a unit quaternion. ICARUS uses quaternions in its attitude calculations. Both quaternion signs describe one rotation, so the attitude comparison has to give the same answer for either sign.

The globe in this example is the space of physical orientations. On the description side, two quaternion labels point to the same attitude. We will write down the covering map that makes that statement exact. A change of label must not become a fictitious turn of the aircraft. A control implementation also needs to interpolate orientations and combine several reports. We will require all three operations to respect the same sign identification, then inspect the outcomes of the specified simulator.

The object is an orientation

A physical orientation is represented by a rotation matrix $R \in SO(3)$. A unit quaternion $q \in S^3$ maps to that orientation through

$$\rho : S^3 \longrightarrow SO(3), \qquad \rho(q) = \rho(-q).$$

Every orientation has exactly two unit-quaternion representatives. This is a double cover with a specified total space, base space and map.

The raw Euclidean distance between q and $-q$ is 2. Their physical orientation distance is 0. The first calculation measures separation in the four stored coordinates; the second measures the smallest physical rotation between the represented states. A detector that uses the wrong comparison can mistake a representation choice for motion.

Normalize the Riemannian metric on $SO(3)$ so that its geodesic distance is the principal physical rotation angle, measured in radians. Then

$$\theta = 2 \arccos |\langle q_1, q_2 \rangle|, \qquad 0 \le \theta \le \pi.$$

The absolute value makes the result independent of the sign chosen for either quaternion. The pullback of this metric to the unit three-sphere is four times its usual round metric. The factor matters: a quaternion's half-angle and a physical rotation angle describe the same motion at different scales.

Why the quaternion map has exactly two representatives

Mark the nose, left wing and upward directions on the model aircraft. Its orientation tells us where those three directions point relative to the room. The quaternion encodes that rotation. Write a quaternion as $q = a + bi + cj + dk$, with $i^2 = j^2 = k^2 = ijk = -1$. Identify a spatial vector

$v \in \mathbb{R}^3$ with a pure imaginary quaternion. For $|q| = 1$, define

$$\rho(q)v = qvq^{-1}.$$

Quaternion norm multiplicativity shows that this linear map preserves vector lengths. It preserves orientation as well: its determinant is continuously either $+1$ or -1 on the connected unit sphere S^3, and is $+1$ at $q = 1$. Thus $\rho(q) \in SO(3)$.

If q commutes with every imaginary vector, its imaginary part must vanish, so the kernel on the unit sphere is $\{1, -1\}$. Conversely, a rotation through physical angle θ about unit axis u is produced by

$$q = \cos(\theta/2) + u \sin(\theta/2).$$

The map is onto, and two quaternions produce the same rotation exactly when they differ by sign. Its derivative at the identity maps an imaginary tangent w to the infinitesimal action $v \mapsto wv - vw = 2w \times v$. This is invertible between the three-dimensional tangent spaces. Translation in the groups gives the same local property everywhere. Choose an open neighborhood of q on which ρ is a diffeomorphism onto its image, small enough to be disjoint from its antipodal copy. The other copy maps diffeomorphically onto the same image. Every rotation in that image has exactly two representatives, so these two copies exhaust its inverse image. This is the evenly covered neighborhood of Chapter 14.

Calculate the same comparison twice

Start at the identity quaternion $q_0 = (1, 0, 0, 0)$. A quarter-turn about the vertical axis has representative

$$q_1 = \left(\frac{\sqrt{2}}{2}, 0, 0, \frac{\sqrt{2}}{2} \right).$$

The distance formula gives $\theta = \pi/2$. Changing q_1 to $-q_1$ leaves that answer unchanged.

Now define a sign-invariant overlap and a chordal comparison:

$$S_{\text{rot}} = \langle q_0, q_1 \rangle^2 = \cos^2(\theta/2), \qquad d_{\text{ch}} = \sin(\theta/2).$$

For this quarter-turn both S_{rot} and d_{ch}^2 equal $1/2$. In general,

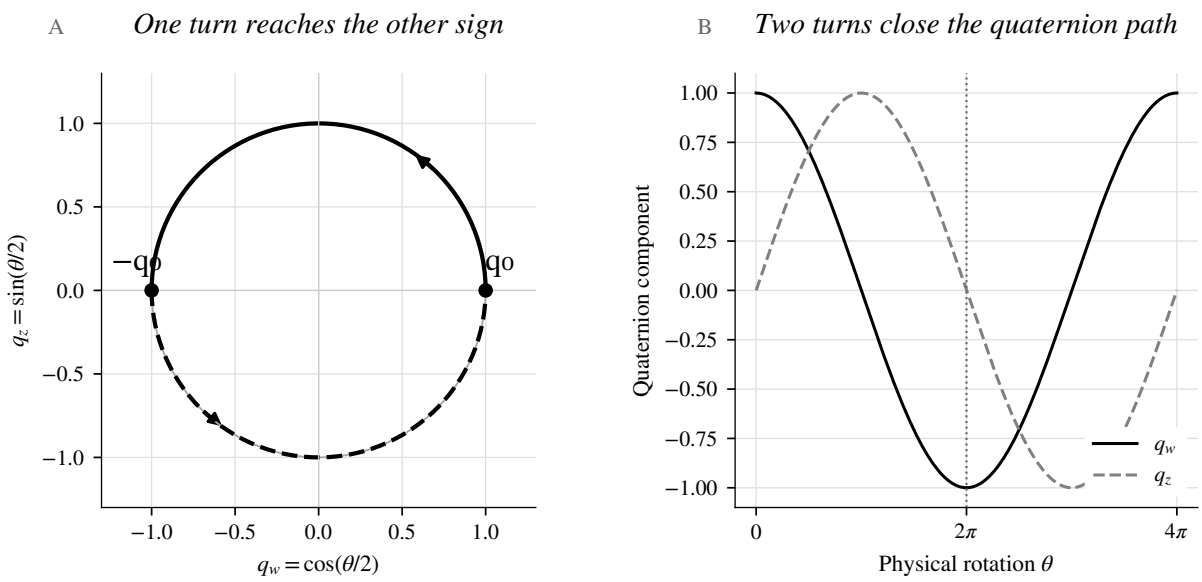
$$S_{\text{rot}} + d_{\text{ch}}^2 = 1.$$

The intrinsic distance is θ; the chordal quantity is a specified function of it. To carry an angular threshold into a chordal comparison, convert it using the same formula that relates the two quantities.

Let $R = \rho(q_0)^{\mathsf{T}} \rho(q_1)$ be the relative rotation. Its eigenvalues are $1, e^{i\theta}, e^{-i\theta}$, so

$$\|I - R\|_2 = 2 \sin(\theta/2) = 2d_{\text{ch}}.$$

Here $\| \cdot \|_2$ is the matrix operator norm. We have reached the same comparison through quaternion coordinates and through the rotation matrix. The underlying object, the maps and the measurement rules are all explicit.



Physical turn	*Quaternion (w, x, y, z)*	*Body attitude*
0	(1, 0, 0, 0)	Identity
π	(0, 0, 0, 1)	Half-turn about z
2π	(−1, 0, 0, 0)	Identity
4π	(1, 0, 0, 0)	Identity

Figure 15.1: A fixed-axis rotation lifted to unit quaternions. A physical 2π turn returns the body to its original attitude but takes the continuous lift from q_0 to $-q_0$. A 4π turn returns the lift as well. Only the nonzero scalar and axial components are plotted.

Returning to a state is not the same as erasing a path

Rotate the aircraft model through one complete turn about the vertical axis. It returns to its initial physical orientation. Along the continuous quaternion path

$$q(\alpha) = (\cos(\alpha/2), 0, 0, \sin(\alpha/2)),$$

the endpoint at $\alpha = 2\pi$ is $(-1, 0, 0, 0)$. A second complete turn, at $\alpha = 4\pi$, returns the lift to $(1, 0, 0, 0)$.

An endpoint comparison and a lifted history therefore answer different questions. The former says whether the orientation is the same now. The latter retains information about the route by which it was reached. Neither answer replaces the other. The mathematical construction tells us which answer a calculation is using.

The factor of two is a measurement convention

For the fixed-axis curve above, unit-sphere arc length is $|\theta|/2$ until the chosen branch reaches its endpoint. Physical rotation angle is $|\theta|$. Accordingly the pullback of the metric on $SO(3)$ normalized by rotation angle is four times the unit-round metric on S^3. Lengths scale by two because squared lengths scale by four.

This normalization explains the half-angle in the sign-invariant score. With physical angle $\theta \in [0, \pi]$, squared quaternion inner product gives $\cos^2(\theta/2)$. In the unoriented-line example the invariant was $\cos^2 \theta$ because the angle and covering map were different. The similarity of the formulas is useful after their domains and rulers are stated.

The covering map has explained the sign pair. To implement the comparison, we now attach a complete frame to the same aircraft and follow its rotation through matrices, quaternion interpolation and normalization.

Mark the aircraft before measuring its attitude

Place three short painted arrows on an aircraft model: one toward the nose, one toward the left wing, and one upward. Rotate the model while the room stays fixed. The three arrows give the columns of an orientation matrix R: their coordinates in the room. They remain mutually perpendicular and right-handed, so $R^T R = I$ and $\det R = 1$. The object is an element of $SO(3)$.

Now imagine two independent flight-control programs reporting the same attitude. One returns a unit quaternion q; the other returns $-q$. Comparing the four stored numbers directly would report a difference of two. Comparing the painted arrows reports zero. The geometric question is which of those comparisons the control task needs.

Write a quaternion as $q = (w, v)$, with $v \in \mathbb{R}^3$ and $w^2 + v^T v = 1$. With the convention used here, its action on a vector x is

$$R(q)x = (w^2 - v^T v)x + 2v(v^T x) + 2w(v \times x).$$

Every term is unchanged when (w, v) is replaced by $(-w, -v)$. This verifies the identification at the level of an actual vector, not merely by a declaration about signs. The inverse is $q^{-1} = (w, -v)$, and quaternion multiplication composes the corresponding rotations in the same order.

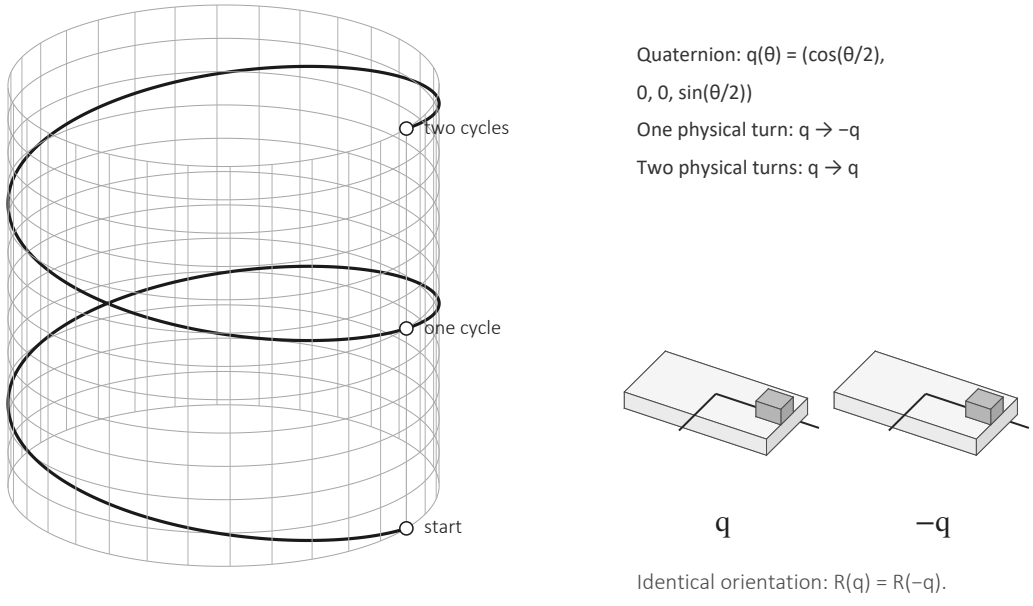
For a rotation through angle θ about a unit axis n,

$$q(\theta) = \big(\cos(\theta/2),\, n \sin(\theta/2)\big).$$

The half-angle belongs to this parametrization of a physical rotation. One physical turn takes q from $(1, 0)$ to $(-1, 0)$. Two physical turns return the quaternion to $(1, 0)$. The aircraft has returned after the first turn; its continuous lift has retained the route.

A return in the object can retain a history

Height counts completed object cycles; it is a history coordinate, not a physical axis.



A The marked body turns in the room

B The quaternion takes two turns to return

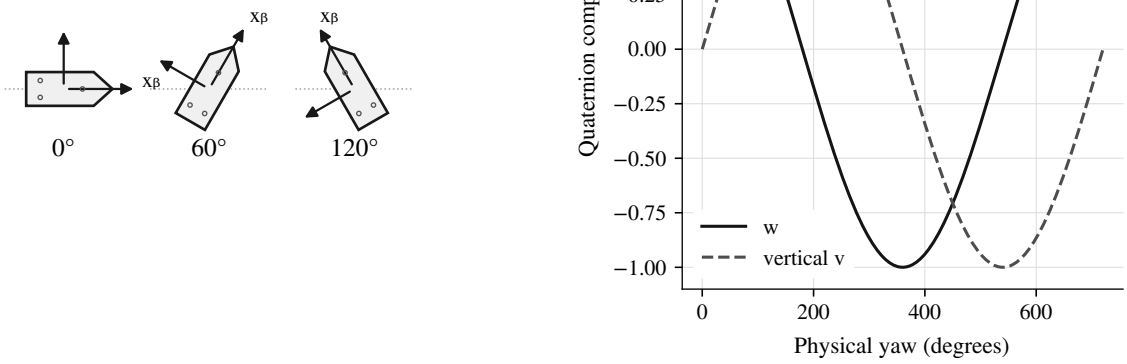
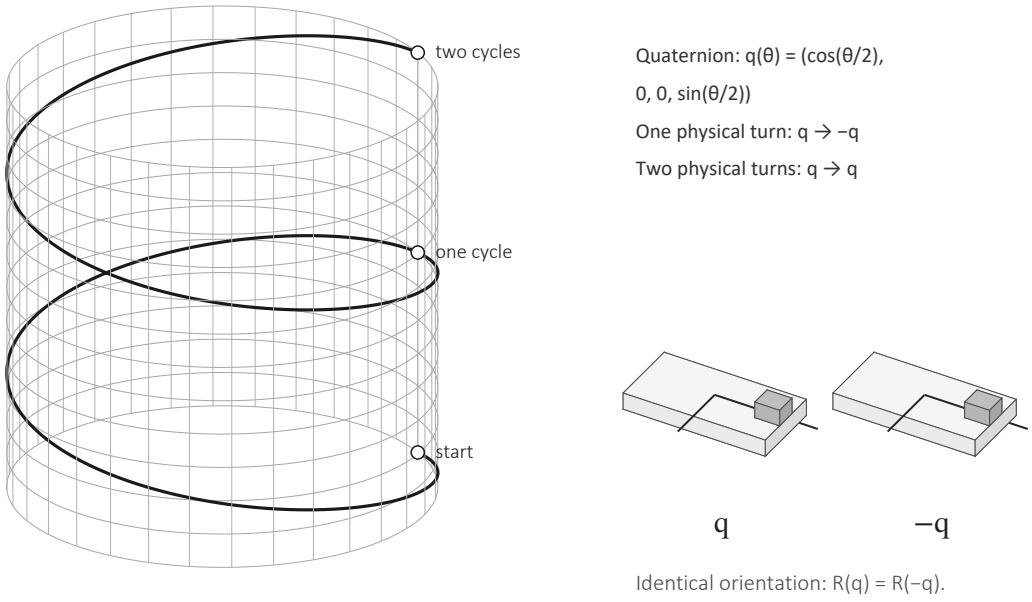


Figure 15.2: *The aircraft returns before the lift does.* The study continues on the following figure page.

A return in the object can retain a history

Height counts completed object cycles; it is a history coordinate, not a physical axis.



C Half-turn checkpoints track the sign

YAW	*q = (w, 0, 0, z)*	*BODY*
0°	(1, 0, 0, 0)	initial
180°	(0, 0, 0, 1)	half-turn
360°	(−1, 0, 0, 0)	returned
540°	(0, 0, 0, −1)	half-turn
720°	(1, 0, 0, 0)	returned

D The physical nose repeats after one turn

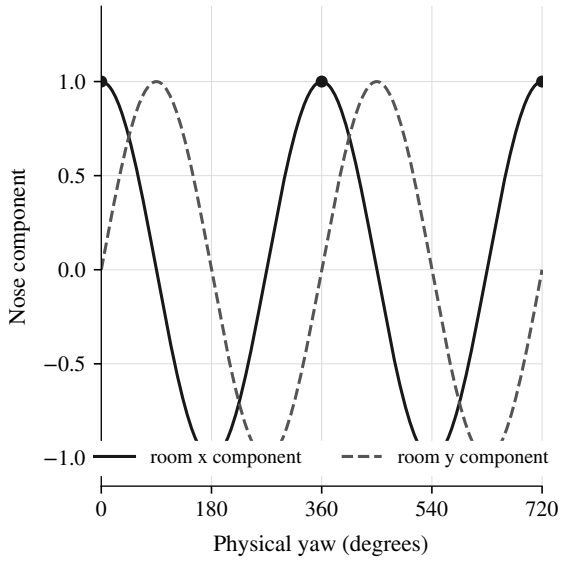


Figure 15.2: *The aircraft returns before the lift does (continued).* Body axes are shown at three yaw angles. The quaternion components then follow a continuous two-turn history, while the register identifies the representatives at each half-turn. Opposite representatives generate the same rotation matrix; the physical angle and its half-angle have different periods.

Deriving the attitude distance twice

Let q_1, q_2 be normalized representatives. Their relative quaternion $q_1^{-1}q_2$ has scalar part $q_1^T q_2$. A sign-invariant shortest rotation angle is therefore

$$d_{SO(3)}(R_1, R_2) = 2\arccos(|q_1^T q_2|).$$

The absolute value selects the shorter of the two antipodal routes in the cover. Its output lies in $[0, \pi]$. This is the geodesic distance for the bi-invariant metric whose tangent-vector norm is the physical rotation angle.

There is a second route through the rotation matrix. A rotation through θ has eigenvalues $1, e^{i\theta}, e^{-i\theta}$, so

$$\operatorname{tr}(R_1^T R_2) = 1 + 2\cos\theta, \qquad \theta = \arccos\frac{\operatorname{tr}(R_1^T R_2) - 1}{2}.$$

The quaternion and matrix calculations are two descriptions of one measured separation.

As a complete instance, let R_1 be a 30° yaw and R_2 a 70° yaw, both about the room's vertical axis. Their representatives are

$$q_1 = (\cos 15^\circ, 0, 0, \sin 15^\circ), \qquad q_2 = (\cos 35^\circ, 0, 0, \sin 35^\circ).$$

Their dot product is $\cos 20^\circ$. The quaternion distance is 40°. The relative matrix has trace $1 + 2\cos 40^\circ$ and gives the same answer. Negating either reported quaternion changes neither result.

The Euclidean distance between the nearest representatives is

$$c = \min\{\|q_1 - q_2\|, \|q_1 + q_2\|\} = 2\sin(\theta/4).$$

This is the quaternion chord on the unit three-sphere. The earlier $d_{\text{ch}} = \sin(\theta/2)$ is half the operator-norm difference between rotation matrices. The two chords use different spaces. The quaternion chord is approximately $\theta/2$ at small angles. A software tolerance copied from physical radians into raw quaternion distance is therefore off by a factor approaching two near the identity. Away from the identity, the full nonlinear conversion matters.

A continuous interpolation has an explicit contract

For interpolation choose the sign of q_2 so that $q_1^T q_2 \geq 0$, and put $\omega = \arccos(q_1^T q_2)$. When $0 < \omega < \pi$, the great-circle interpolation is

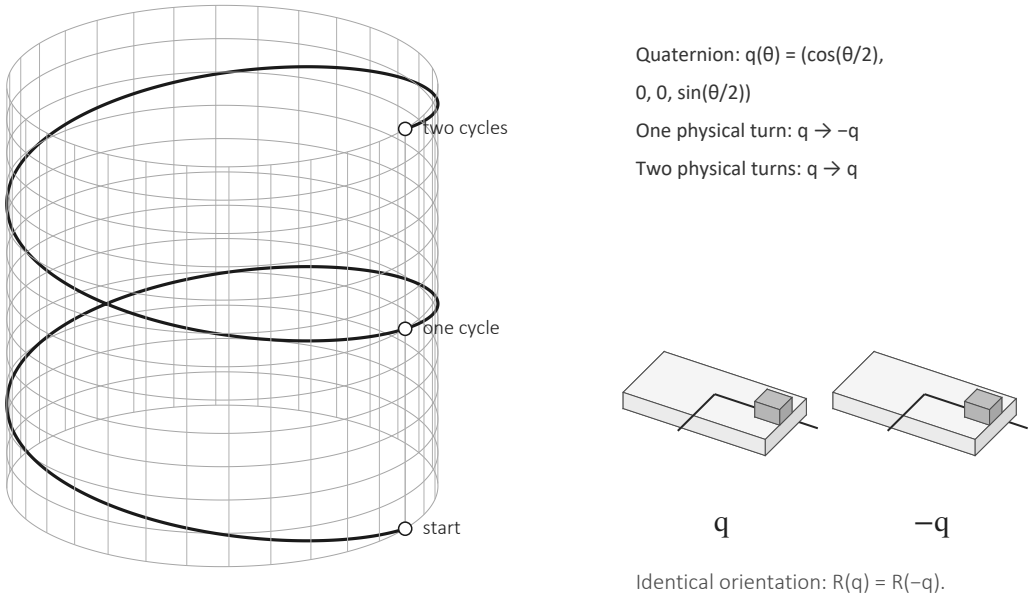
$$q(t) = \frac{\sin((1-t)\omega)}{\sin\omega} q_1 + \frac{\sin(t\omega)}{\sin\omega} q_2, \qquad 0 \leq t \leq 1.$$

The two coefficients do not sum to one in general. They combine two unit vectors in their common plane to remain on the sphere. Taking the dot product with q_1 shows $q_1^T q(t) = \cos(t\omega)$; the lifted angular speed is constant. The corresponding physical rotation progresses at speed 2ω.

At identical orientations, use the continuous limit. At a relative physical rotation of 180°, the axis line is fixed and either direction along it gives a shortest route. A previous angular velocity, a mechanical constraint or a declared tie-breaking rule must decide. The manifold does not remove that ambiguity by changing notation.

A return in the object can retain a history

Height counts completed object cycles; it is a history coordinate, not a physical axis.



A A body supplies the physical reference

B Angle, cover arc and chord differ

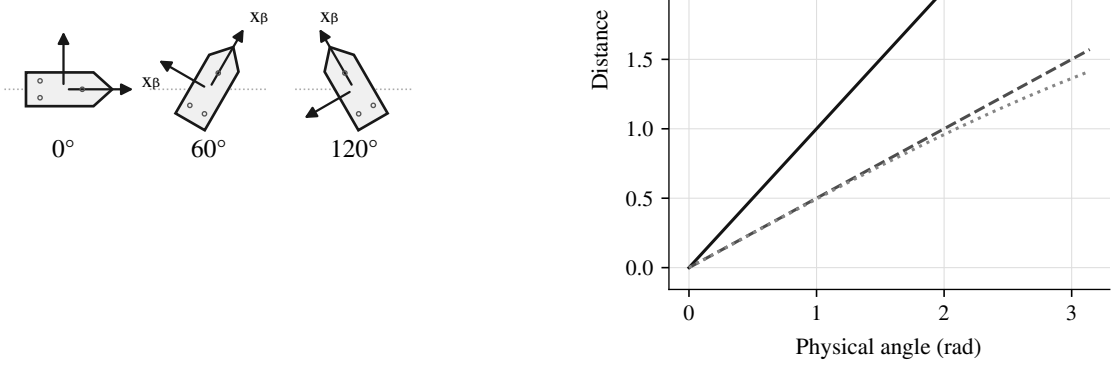
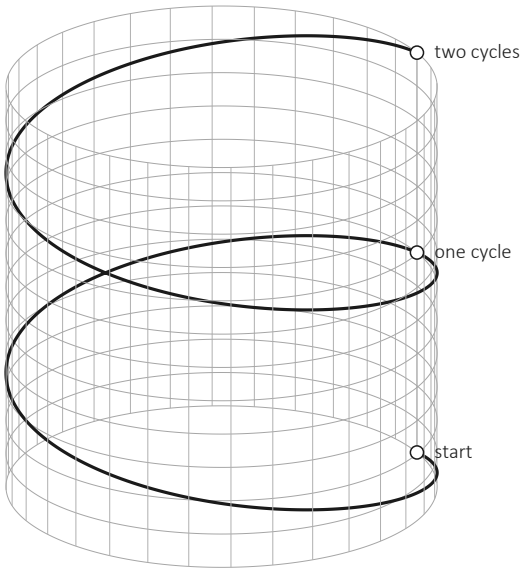


Figure 15.3: *A distance survives a change of representation.* The study continues on the following figure page.

A return in the object can retain a history

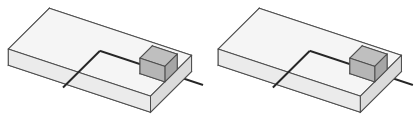
Height counts completed object cycles; it is a history coordinate, not a physical axis.



Quaternion: q(θ) = (cos(θ/2), 0, 0, sin(θ/2))

One physical turn: q → −q

Two physical turns: q → q



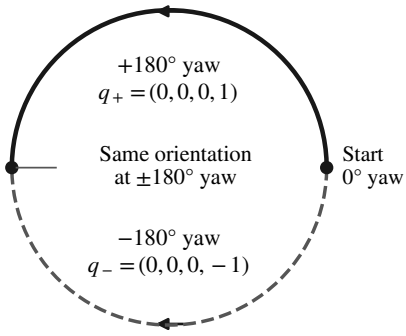
q −q

Identical orientation: R(q) = R(−q).

C One 40-degree turn, three numerical measures

D Two yaw histories, one final orientation

QUANTITY	*EXPRESSION*	*VALUE*
Quaternion dot	cos 20°	0.939693
Relative trace	1 + 2 cos 40°	2.532089
Physical angle	2 arccos $\lvert q_1 \cdot q_2 \rvert$	0.698132 rad
Nearest chord	2 sin 10°	0.347296



Compare yaw 30° with yaw 70°; unit quaternions.

Figure 15.3: *A distance survives a change of representation (continued).* The marked body gives a physical reference for yaw. The curves distinguish rotation angle, the nearest quaternion-cover arc and the nearest chord. The numerical table compares yaw 30° with 70°. At a half-turn, two opposite yaw histories reach the same physical endpoint and opposite unit-quaternion representatives.

Normalization and the quantities a controller actually receives

A computed quaternion is rarely exactly unit length. If an input q has norm $r > 0$, normalize it to q/r before using unit-quaternion identities. Clamping an inverse-cosine argument to $[-1, 1]$ addresses a small rounding excursion; it does not repair a missing normalization or an invalid sensor packet.

For a small attitude error, one can use a three-dimensional rotation vector. If the sign-selected relative quaternion is (w, v) with $w \geq 0$, then

$$\phi = 2\,\mathrm{atan2}(\|v\|, w)\frac{v}{\|v\|},$$

with the continuous value $\phi = 0$ when $v = 0$. Near zero, $\phi \simeq 2v$. The norm $\|\phi\|$ is the physical error angle. Its direction identifies the axis along which an attitude controller might act, subject to the actuator model and frame convention.

Put the model aircraft ten degrees off heading. Then put it on heading but with one wing ten degrees low. The angle display says ten both times. I can use that number to flag an error. The rotation vector tells me whether the correction is in yaw or roll.

A double cover is not every phase quotient

There is a nearby construction worth distinguishing because it uses similar notation. A normalized complex pair $(z_1, z_2) \in \mathbb{C}^2$ lies on S^3. If all pairs $e^{i\varphi}(z_1, z_2)$ represent the same complex line, the quotient is $\mathbb{CP}^1$, and the fiber contains a whole circle of phases. This is the Hopf fibration. The quaternion attitude quotient identifies only q and $-q$, and its base is $SO(3)$. Both start from a three-sphere; their equivalence relations and bases differ.

For normalized complex states, the invariant overlap $|z^* w|^2$ descends through the phase quotient. With the common Fubini–Study normalization,

$$d_{FS}([z], [w]) = \arccos |z^* w|.$$

For the attitude quotient the distance is $2\arccos |q_1^T q_2|$. The factors and domains are part of the experiment. A familiar half-angle is a reason to inspect a construction, not permission to erase its equivalence relation.

The ICARUS application gives the attitude construction a concrete purpose: a controller should act on the aircraft's orientation rather than an arbitrary sign in its stored representation. Its simulation records test particular controllers and scenarios. The exact sign invariance proved here holds independently of whether a particular controller performs well [20].

Laboratory 4: track a turn through quaternion signs

Use an aircraft model whose nose and wing axes are visibly marked. The computational experiment rotates it about the vertical axis. Record physical angles from zero through 720° in increments of 30°. For each angle construct $q = (\cos(\theta/2), 0, 0, \sin(\theta/2))$ and its rotation matrix.

Make two reporting systems. The first reports the continuous quaternion formula. The second multiplies selected samples by -1 without changing the aircraft orientation. Choose a deterministic sign pattern and save it. Compute the raw four-component distance between successive samples in both reports, then compute the sign-invariant physical attitude distance.

The raw distance can jump close to two when the reported sign changes. The physical distance remains 30° between adjacent samples. This is a test with a known answer: the reporting system changed representatives, not the prescribed motion.

Recover a continuous lift. Starting from the first sample, negate the next reported quaternion whenever its dot product with the previous chosen quaternion is negative. With the prescribed increments, this recovers a consistent lift up to the initial sign. At 360° the chosen quaternion is the negative of its starting value. At 720° it returns.

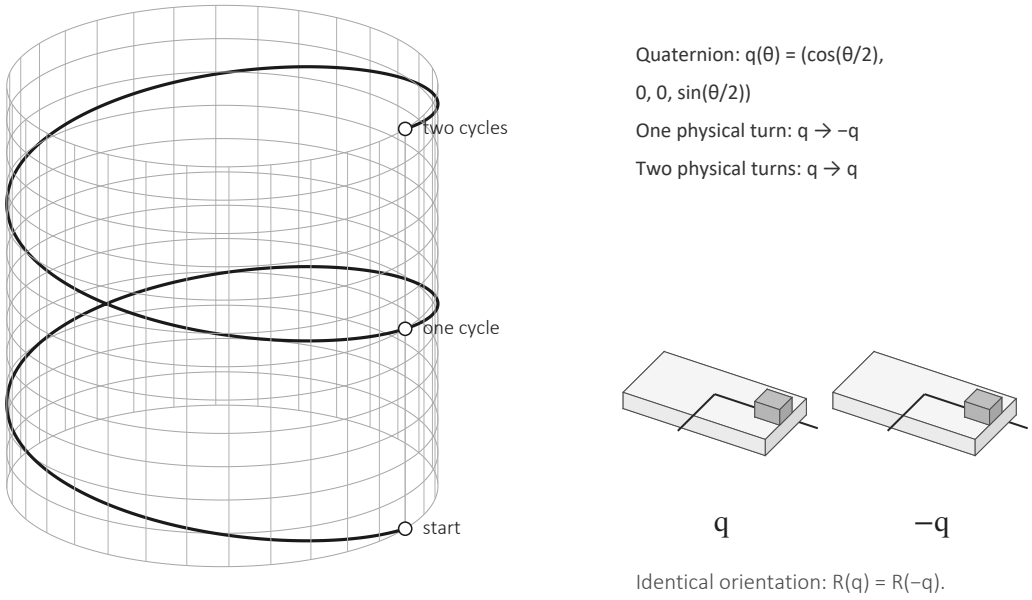
Physical turn	Scalar part	Vertical component	Aircraft attitude
0°	1	0	initial
180°	0	1	half-turn
360°	−1	0	initial
540°	0	−1	half-turn
720°	1	0	initial

Test the sampling limit. Replace the 30° increments by a physical step greater than 180°. The nearest-sign procedure now chooses the shorter representative increment and cannot reconstruct the unobserved long turn from endpoints alone. Preserve this failure. State what extra information could resolve it: a higher sample rate, an angular-velocity bound or a separately recorded turn count.

Compare two distance calculations. For each pair of samples compute the angle using the quaternion dot product and using the trace of the relative rotation matrix. Clamp only rounding excursions in inverse-cosine inputs after checking normalization. Report the largest discrepancy. Test the calculation again after negating every other reported quaternion; the physical-angle column must remain unchanged.

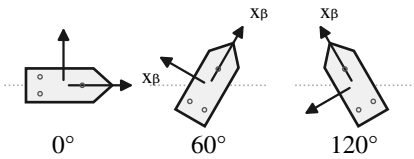
A return in the object can retain a history

Height counts completed object cycles; it is a history coordinate, not a physical axis.



A The physical body turns continuously

B Editing quaternion signs creates false jumps



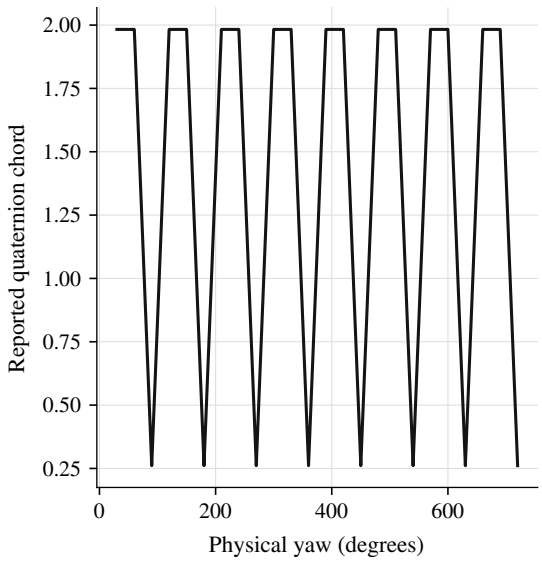
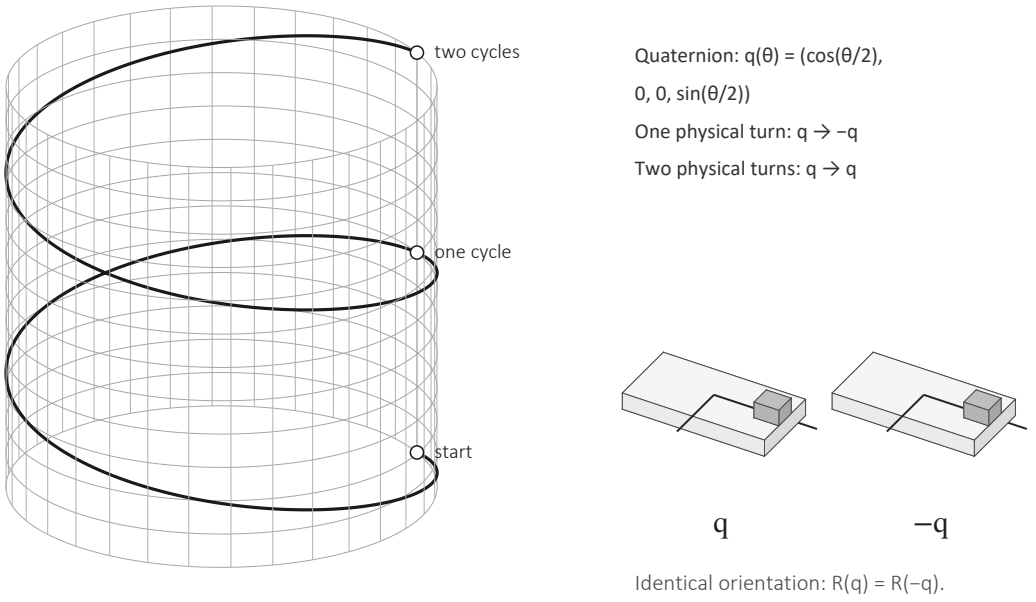


Figure 15.4: *Laboratory record: a sign jump is not an aircraft maneuver.* The study continues on the following figure page.

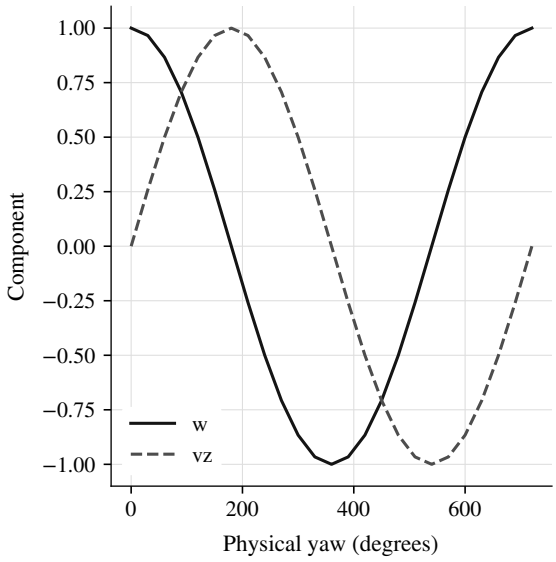
A return in the object can retain a history

Height counts completed object cycles; it is a history coordinate, not a physical axis.



C A continuous lift retains the turn history

D Independent checks distinguish sign from motion



Matrix and quaternion angles agree.

Adjacent physical angle is 30°.

At 360° the lift changes sign.

At 720° the lift returns.

Steps above 180° are ambiguous without extra history.

Figure 15.4: *Laboratory record: a sign jump is not an aircraft maneuver (continued).* The prescribed yaw sequence, deliberately changed quaternion signs and recovered lift are plotted together. Matrix and quaternion distances provide independent calculations of the same physical angle. The undersampled example records an ambiguity rather than inventing a missing turn.

Interpolation extension. Interpolate between a 30° yaw and a 150° yaw using the spherical formula in the chapter. At fractions $1/4, 1/2, 3/4$, verify physical yaws of $60°, 90°, 120°$. Repeat after negating the final quaternion and applying the sign-selection rule. Then omit that rule and inspect the longer lifted route. Explain which path the application intended to preserve.

What to submit. Include the prescribed angles, raw and invariant distances, selected signs, rotation matrices and the sampling-limit example. The experiment verifies an exact representation symmetry and a particular reconstruction procedure. It does not establish the performance of an aircraft controller or the sufficiency of any real sensor's sampling rate.

What is the average of two descriptions of the same thing?

Two attitude files report q and $-q$ for one aircraft orientation. Their componentwise average is zero, which is not a unit quaternion and represents no attitude. Averaging first and interpreting later has destroyed a perfectly consistent pair of observations. The quotient geometry must enter before the average is defined.

A general geometric mean minimizes a stated sum of squared distances:

$$F(m) = \sum_i w_i d(m, x_i)^2, \qquad w_i \geq 0, \quad \sum_i w_i = 1.$$

Such a minimizer is called a Fréchet mean. Its existence, uniqueness and computation depend on the space, the distance and the data region. A familiar arithmetic average is one special case, not a definition that can be copied into every representation.

The circular compass example

Measure heading from east with positive angles counterclockwise. Suppose two headings are 170° and −170°. Their numerical average in that display is zero. Their short circular midpoint is 180°. To calculate a circular resultant, represent the headings by unit vectors and sum them:

$$(\cos 170^\circ, \sin 170^\circ) + (\cos(-170^\circ), \sin(-170^\circ)) = (2\cos 170^\circ, 0).$$

The resultant points west, giving the appropriate midpoint for this concentrated pair. The coordinate seam no longer creates a false eastward answer.

For opposite headings 0 and π, the resultant is zero and has no direction. This is not a numerical failure to be repaired by selecting a random angle. The data are symmetric. Under squared intrinsic circle distance, there are two equal minimizing midpoints at $\pi/2$ and $3\pi/2$. Different definitions of a mean can expose that symmetry in different ways, but none should claim a unique direction supplied by these data alone.

Average line directions through their projectors

An unmarked bar has a unit-vector representative u with u and $-u$ identified. The matrix $P = uu^T$ removes that sign choice. For representatives u_i, consider the projector loss

$$F(v) = \sum_i w_i \|vv^T - u_i u_i^T\|_F^2, \qquad \|v\| = 1.$$

Expanding a term gives $2 - 2(v^T u_i)^2$. Hence minimizing the loss is equivalent to maximizing

$$v^T M v, \qquad M = \sum_i w_i u_i u_i^T.$$

The solution is any unit eigenvector corresponding to the largest eigenvalue of M, interpreted as a line. This is an exact mean for the stated projector loss. It is not automatically the same as a mean defined by squared intrinsic projective distance.

Take equal-weight bars at angles $-30°$ and $30°$. Their matrix average is $M = \mathrm{diag}(3/4, 1/4)$. The principal direction is horizontal, and the eigenvalue gap is one half. Reversing either bar representative leaves M unchanged.

For equal-weight perpendicular bars, $M = I/2$. Every unit vector has the same projector loss. The mean is not uniquely determined. The vanishing eigenvalue gap reveals the ambiguity. Reporting the gap alongside the selected direction is more informative than reporting one angle produced by a deterministic tie-breaking convention.

Use the same idea for quaternion representatives

Let normalized quaternion observations be $q_i \in S^3$, with arbitrary signs. The matrix

$$M = \sum_i w_i q_i q_i^T$$

is sign invariant. Its leading eigenvector minimizes the corresponding quaternion-projector loss. The resulting pair $\{q, -q\}$ gives one physical orientation if the top eigenspace is one-dimensional.

The distance induced by this projector loss is related to the physical attitude angle, but it is not the angle itself. If the relative physical angle is $\alpha \in [0, \pi]$, then

$$\|qq^T - rr^T\|_F^2 = 2(1 - (q^T r)^2) = 2\sin^2(\alpha/2).$$

Consequently this average minimizes a sum of squared chordal-type discrepancies. A Fréchet mean based on squared physical angles generally requires a different optimization. Naming the loss prevents a convenient eigenvector calculation from being credited with a different theorem.

For two sufficiently close attitudes of equal weight, their short-geodesic midpoint can be obtained by choosing consistent signs and normalizing $q_1 + q_2$. At a relative half-turn the choice of shortest rotational axis becomes ambiguous. The data region matters even when each individual formula is simple.

Averaging and quotienting need not commute

The bad average of q and $-q$ illustrates the issue. Quotienting first recognizes a repeated observation of the same attitude. Averaging the representatives first leaves the space entirely. A valid averaging rule should descend through the equivalence relation: changing any representative by an allowed symmetry must not change the physical answer, apart from an explicitly stated nonuniqueness.

The same requirement applies to row permutations in a distribution, chart changes on a surface, or unit changes in a measurement. An average is a construction on the chosen object. Its implementation should respect the object's symmetries and the meaning of its ruler.

A comparison system needs more than a distance function at its last stage. Every operation that combines observations must be compatible with the representation changes the system declares harmless.

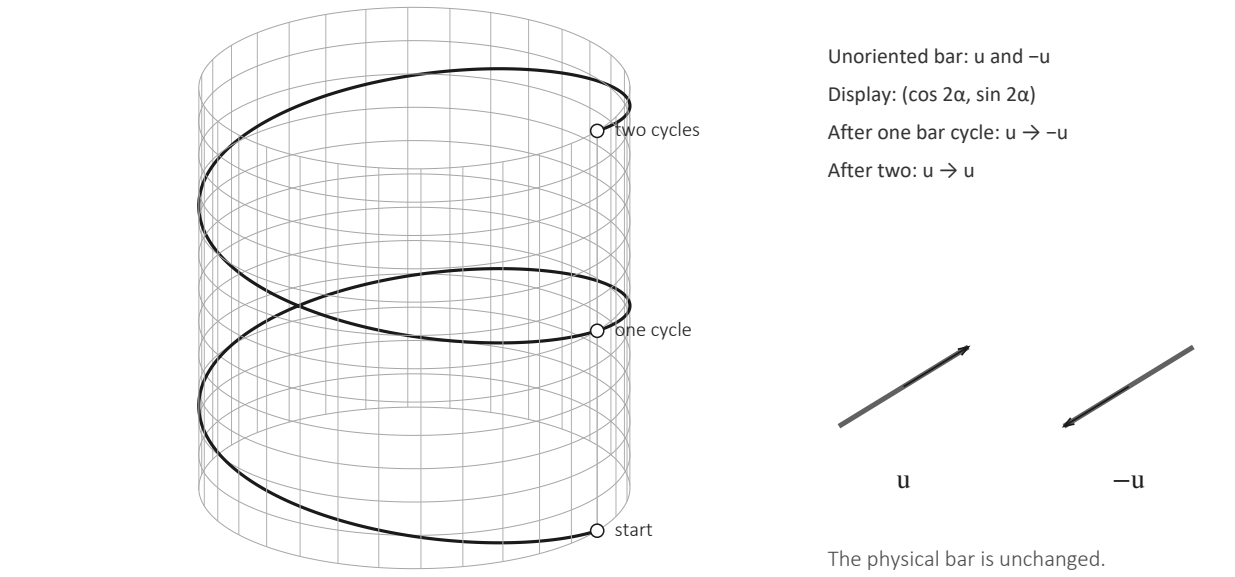
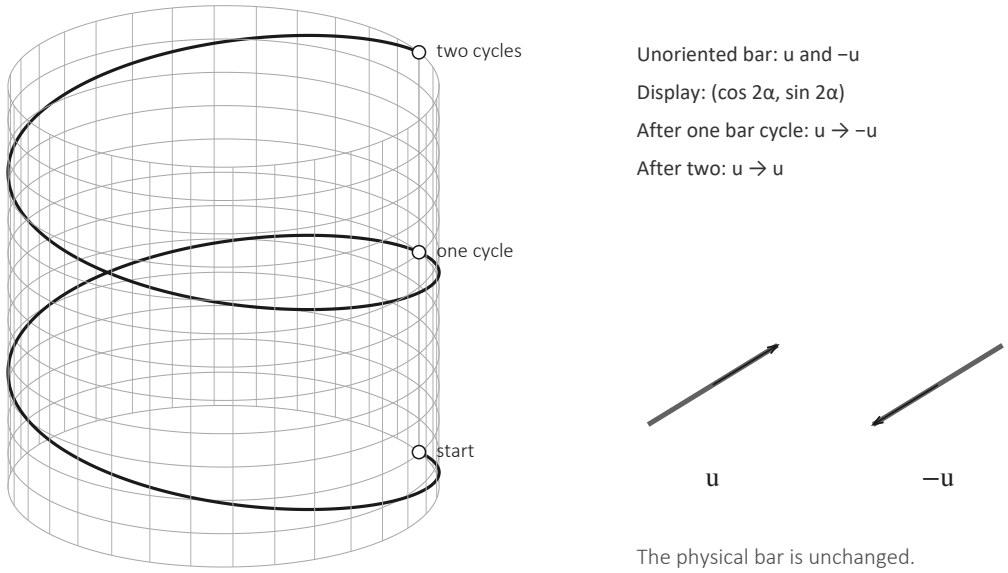


Figure 15.5: *Average the object using its declared geometry.* The study continues on the following figure page.

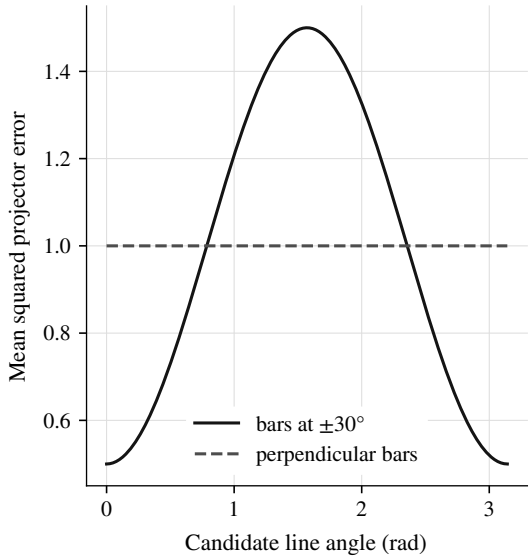
A return in the object can retain a history

Height counts completed object cycles; it is a history coordinate, not a physical axis.



C The projector loss selects the mean axis

D An eigenvalue gap makes that axis unique



$M = \Sigma\, w_i\, u_i u_i^T$

Maximize $v^T M v$ with $\|v\|=1$.

A unique leading line needs an eigenvalue gap.

A projector mean is not generally an intrinsic-distance mean.

Figure 15.5: *Average the object using its declared geometry (continued).* Circular and line-direction examples expose failures of raw representative averaging. The eigenvector construction minimizes the stated projector loss; a repeated leading eigenvalue leaves a genuine nonuniqueness.

A mean that respects the cover. Construct three quaternion reports of the same attitude with one sign reversed. Show that the projector matrix gives the same physical answer for every choice of their signs. Then construct a pair for which the leading eigenvalue is repeated and explain what the computation can legitimately report.

The implemented comparison has consequences

The ICARUS simulator uses the absolute quaternion inner product to measure the difference between the aircraft's current attitude and its target. The fixed-gain controller used as a baseline also accounts for the two quaternion signs. This matters when interpreting the experiments: the comparison is not between a system that knows the double cover and one that necessarily ignores it.

In the recorded inverted-recovery experiment, both ICARUS and the fixed-gain geometric baseline satisfy the simulator's success criterion in all 500 trials. Mean squared roll/pitch torque effort is 1.1381 for ICARUS and 13.7333 for the baseline, a ratio of about 12.07. In the flip experiment, both again succeed, but the baseline uses less effort: 0.3767 against 0.9069. These are specified simulator outcomes, not battery-energy measurements or flight-test results.

Say the aircraft holds steady and its quaternion switches from q to $-q$. The numbers change, but the controller should see zero rotation. The double cover explains why, and the invariant distance gives the controller that zero inside the model. The simulation tells us how each controller performed on the two maneuvers.

What happens when the frame we use to compare directions has to travel with us?

Practice with the construction

15.1. Write the two unit quaternions representing a half-turn about the vertical axis. Verify that negating the quaternion preserves the rotation.

15.2. For physical angular separation $\pi/3$, compute the intrinsic distance, squared-inner-product score and chordal comparison used in this chapter.

Chapter 16

What happens around a loop

Slide a small platform around a globe and bring it back to its starting mark. An arrow on the platform stays tangent to the globe as it travels. We will calculate where it points after the circuit, using a rule that keeps it from turning within the plane of the platform.

A route can change a comparison

Picture a toy sled with an arrow painted flat on its deck. Slide it along a great-circle route, keeping the deck tangent to the globe. The deck tilts with the surface. As you go, steer so that the arrow has no instantaneous turn within that local tangent plane. When the sled gets back, the arrow may point a different way.

That steering rule is parallel transport for the sphere's Levi–Civita connection. If I built the sled, I'd need a guide to enforce the rule as it moved. We can calculate exactly where the arrow would point if the guide did its job.

Use a unit globe and the triangle whose vertices are $A = (1, 0, 0)$, $B = (0, 1, 0)$ and $N = (0, 0, 1)$. Travel from A along the equator to B, then up a meridian to N, and down another meridian to A. Begin with arrow $v_A = (0, 0, 1)$.

Along the equatorial leg, the constant spatial vector $(0, 0, 1)$ stays tangent and has zero derivative, so it is parallel. Along the second leg $r(t) = (0, \cos t, \sin t)$, the tangent $r'(t) = (0, -\sin t, \cos t)$ begins as that arrow and is parallel because its derivative is normal to the sphere. It ends as $(0, -1, 0)$ at the pole. Along the last leg $r(t) = (\sin t, 0, \cos t)$, that constant vector stays tangent and parallel. At A the final arrow is therefore $(0, -1, 0)$: a right-angle change from its initial direction.

The rotation is holonomy: the transformation produced by transporting around a closed route. For this sphere triangle, its magnitude is also the enclosed Gaussian-curvature integral, $K\,\mathrm{Area} = \pi/2$. This is a particular surface case of the relation between curvature and holonomy, with orientation fixing the sign [4].

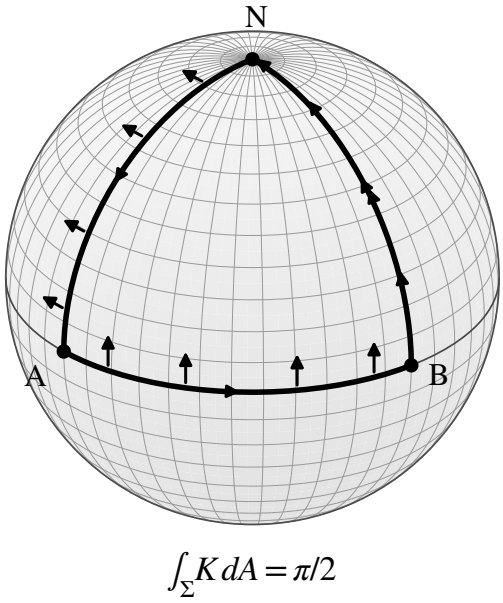
The three-leg route gave a concrete return angle. We now derive the transport equation behind that calculation, use it on the same route, and vary the loop to see which part of the result depends on the path.

A little arrow bolted to a tangent sled

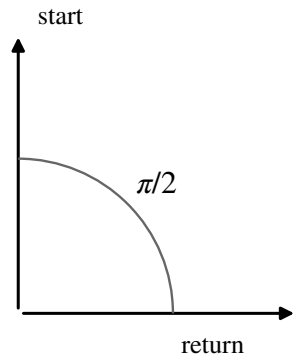
To check the triangle calculation, derive the adjustment rule for the same tangent sled. The arrow tilts in the room as its tangent plane changes. Its additional turn within that plane is the quantity we set to zero. Differentiating the tangency condition will give an equation we can verify on all three legs.

This is an idealized mechanical procedure. An ordinary compass does something else: it aligns with a magnetic field. A toy pushed by hand also introduces friction and steering errors. The sled describes a specified kinematic rule that a computation or a suitably constrained apparatus can implement. That rule is parallel transport.

A *The transport rule follows the surface*



B *The returned arrow has turned*



Tangent plane at A

A, start	(0, 0, 1)
B	(0, 0, 1)
N	(0, −1, 0)
A, return	(0, −1, 0)

Figure 16.1: Exact parallel transport around three great-circle legs of the unit sphere. The small arrows are transported tangent vectors; the arrowheads on the heavy curve indicate travel. The component register and the common tangent plane at A show the $\pi/2$ return rotation, equal to the enclosed curvature integral.

On the unit sphere let $x(t)$ be the sled's contact point and $v(t)$ its painted tangent arrow. Tangency means $x^T v = 0$. Requiring the tangential part of $\dot v$ to vanish makes $\dot v$ normal to the sphere, so $\dot v = cx$. Differentiating tangency gives

$$0 = \dot x^T v + x^T \dot v = \dot x^T v + c, \qquad \dot v = -(\dot x^T v)x.$$

This is a complete transport equation in room coordinates. It also preserves the arrow's length, because $\frac{d}{dt}(v^T v) = -2(\dot x^T v)(v^T x) = 0$.

The rule separates two motions that a picture alone can blur. The arrow tilts because its allowed plane tilts. It has no extra rotation within that allowed plane. When the sled returns, the plane is the original plane again, and the final arrow can be compared directly with its initial position.

The three-leg experiment, calculated without a shortcut

Start at $A = (1, 0, 0)$ with arrow $v = (0, 0, 1)$. Travel east along the equator to $B = (0, 1, 0)$, then along the meridian to the north pole $N = (0, 0, 1)$, then down a second meridian to A.

On the first leg $x(t) = (\cos t, \sin t, 0)$, $0 \le t \le \pi/2$. The constant arrow $(0, 0, 1)$ is tangent and satisfies the transport equation. On the second leg,

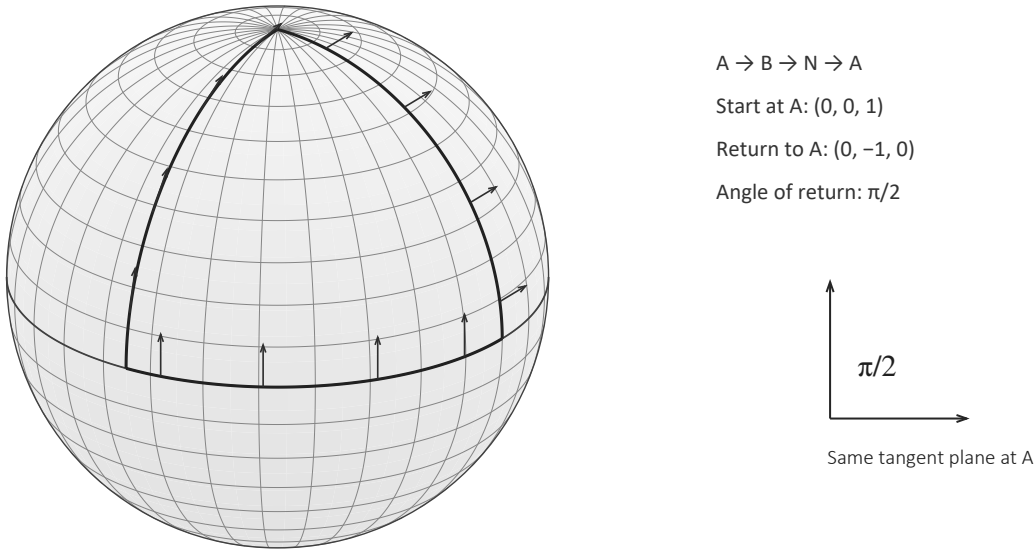
$$x(t) = (0, \cos t, \sin t), \qquad v(t) = (0, -\sin t, \cos t).$$

Here $\dot{x} = v$ and $\dot{v} = -x$, exactly as the equation requires. The arrow at the pole is $(0, -1, 0)$. On the final leg $x(t) = (\sin t, 0, \cos t)$, that arrow is constant and remains tangent. Back at A, it is a quarter-turn away from the original northward arrow.

Every leg is locally as straight as the spherical surface allows. The three-leg route encloses one eighth of the sphere, with area $\pi/2$ on the unit sphere. The Gaussian curvature is one, and the return rotation agrees, modulo a full turn, with the enclosed curvature integral $\int K\, dA = \pi/2$.

Transport follows the whole route

At each station the arrow lies in the local tangent plane.



A A transported arrow follows three surface legs

B The returned arrow differs at the same point

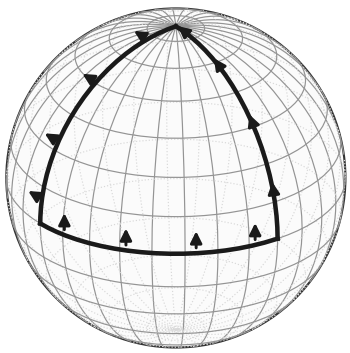
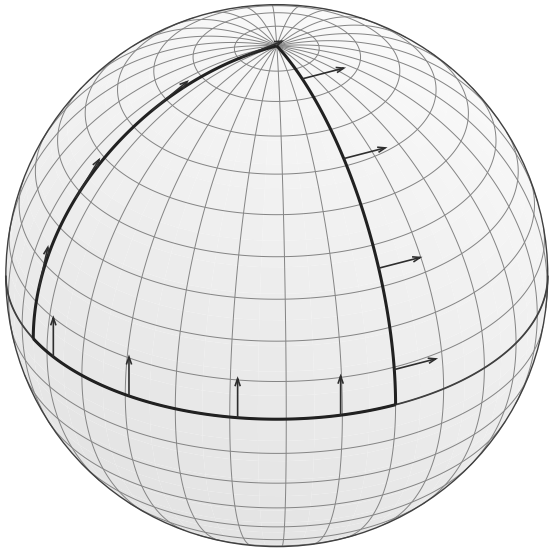


Figure 16.2: *The local rule, the three legs, and the returned arrow.* The study continues on the following figure page.

Transport follows the whole route

At each station the arrow lies in the local tangent plane.

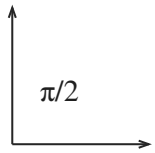


A → B → N → A

Start at A: (0, 0, 1)

Return to A: (0, −1, 0)

Angle of return: π/2



Same tangent plane at A

C The local rule preserves tangent length

x·v = 0: arrow is tangent

v′ = −(x′·v)x

Tangential part of v′ is zero.

v keeps its length.

An ordinary compass follows a different rule.

D Enclosed curvature gives the return angle

Unit-sphere triangle area = π/2

Gaussian curvature K = 1

Integral K dA = π/2

The returned arrow is measured in the original tangent plane.

Figure 16.2: *The local rule, the three legs, and the returned arrow (continued).* The tangent sled is a kinematic model of Levi-Civita parallel transport. The drawing follows the contact point and tangent frame through the three prescribed great-circle legs. The state register prints the actual room-coordinate arrows at each corner.

One family of loops, all the way to an exact return

Send the tangent sled around a circle of latitude instead of along the triangle. Changing latitude varies how much of the sphere the loop encloses, so we can test how its return changes under the same transport rule [24]. On a sphere of radius R, take the circle at colatitude ϑ, the angle down from the north pole. Its enclosed cap area is $2\pi R^2(1-\cos\vartheta)$ and its curvature is $1/R^2$. Therefore its holonomy angle, understood modulo 2π, is

$$\Theta(\vartheta) = 2\pi(1-\cos\vartheta).$$

For a unit arrow in the tangent plane, the Euclidean difference between start and return is

$$m(\vartheta) = 2\left|\sin\frac{\Theta(\vartheta)}{2}\right|.$$

At a tiny cap it is tiny. At $\vartheta = 60^\circ$, $\Theta = \pi$ and the returned arrow is reversed, giving $m = 2$. At the equator, $\vartheta = 90^\circ$, $\Theta = 2\pi$ and the arrow returns exactly. The sphere is still curved at the equator. The total rotation happens to be a full turn.

Suppose I photograph the arrow before and after the equator trip. The two pictures agree. The cap formula reaches a full turn, which looks like zero at the endpoint. A zero return on this one loop tells me the arrows line up. I need more evidence to say the surface is flat.

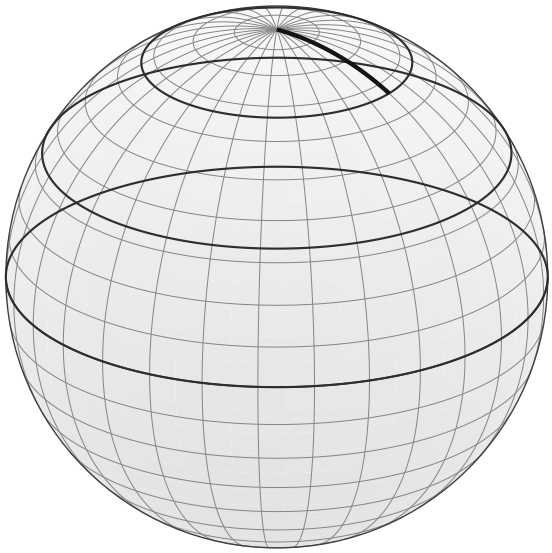
If $r = R\vartheta$ is the geodesic radius of a small cap,

$$\Theta = \pi\frac{r^2}{R^2} - \frac{\pi}{12}\frac{r^4}{R^4} + O\!\left(\frac{r^6}{R^6}\right), \qquad m = \pi\frac{r^2}{R^2} + O\!\left(\frac{r^4}{R^4}\right).$$

The leading behavior scales with area. Doubling the small loop's linear size approximately quadruples the mismatch. The approximation needs r/R small; the equatorial return cannot be inferred from the first term.

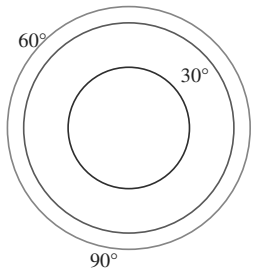
The enclosed cap controls the return

Unit sphere / the accumulated angle determines the return modulo 2π.



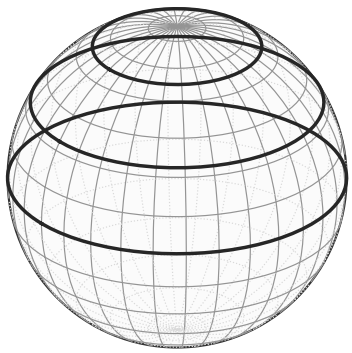
$$\Theta = 2\pi(1 - \cos\alpha)$$

The bold segment lies on the surface.
Radius is measured along that segment.
The surrounding grid supplies orientation.



Plan radii: sin α

A Each latitude encloses a different cap



B Accumulated angle and mismatch differ

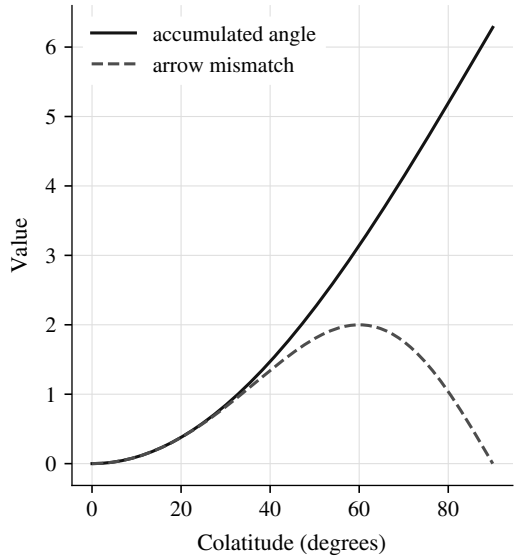
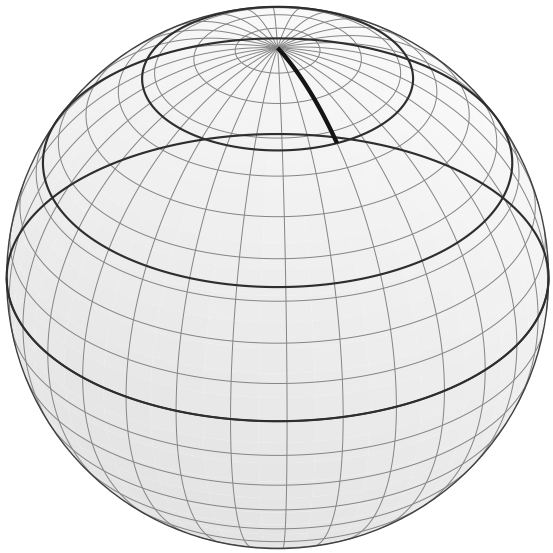


Figure 16.3: *A curved sphere can still give an exact return.* The study continues on the following figure page.

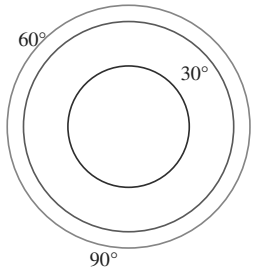
The enclosed cap controls the return

Unit sphere / the accumulated angle determines the return modulo 2π.



$$\Theta = 2\pi(1 - \cos\alpha)$$

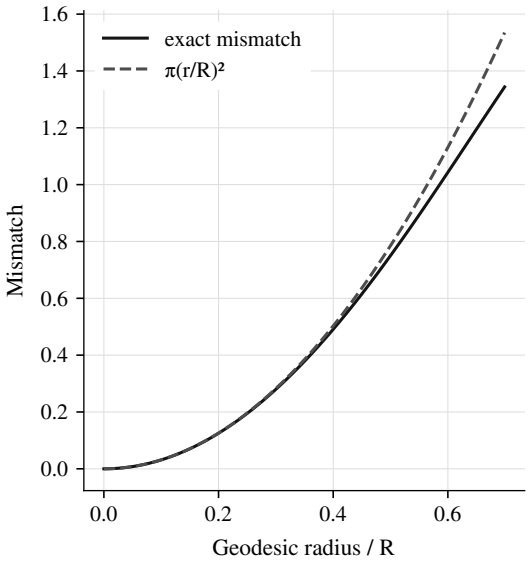
The bold segment lies on the surface.
Radius is measured along that segment.
The surrounding grid supplies orientation.



Plan radii: sin α

C The small-cap formula has a finite range

D An exact return does not imply flatness



At 0°: mismatch 0

At 60°: Θ=π, mismatch 2

At 90°: Θ=2π, mismatch 0

The equatorial return does not make the sphere flat.

Figure 16.3: *A curved sphere can still give an exact return (continued).* The cap construction connects geodesic radius, enclosed area, holonomy angle and measured arrow mismatch. The calculation includes both the small-loop approximation and the equatorial full-turn return. The curves are exact for a round sphere and the stated transport rule.

A comparison protocol can be written down

Suppose two snapshots of the sled show an arrow $a \in T_xM$ at surface point x and an arrow $b \in T_yM$ at surface point y. Each lies in the tangent plane at its own contact point. A path γ from x to y supplies a transport map $P_\gamma : T_xM \to T_yM$, after which $\|P_\gamma a - b\|$ is meaningful. A second path can give a different transported vector.

If a protocol also compares the transport around a declared closed route $\widehat{\gamma}$, an example of a combined discrepancy is

$$D_\gamma(a,b) = \|P_\gamma a - b\| + \lambda \|I - \mathrm{Hol}_{\widehat{\gamma}}\|_{\mathrm{op}}, \qquad \lambda \geq 0.$$

The parameter λ must have whatever units make the terms compatible. For unit dimensionless arrows it can be dimensionless. The definition names the norm, the route and the connection. It is a protocol-dependent discrepancy; it need not be the geodesic distance between base points or a metric on all possible records.

For the latitude loop, use that loop for both γ and $\widehat{\gamma}$ and compare the transported unit arrow with its starting direction. Each norm then equals the rotation mismatch m. Thus $D = (1+\lambda)m$. This is a complete, inspectable instance of the identity research: one object, a declared way of comparing it after travel, and a measurable return difference. It does not assert that logical equality of an object with itself has failed.

Changing the local frame does not erase the return

At each point choose an orthonormal pair of tangent arrows. Replacing that pair by a rotated pair changes the matrix entries of transport. If the starting frame changes by an orthogonal matrix Q, the matrix of a closed-loop transport changes to $Q^{-1}HQ$.

The eigenvalues and trace of H survive that conjugation. So does $\|I - H\|_{\mathrm{op}}$, because orthogonal changes of coordinates preserve the norm. The individual entries need not survive. A comparison based on a single entry must therefore explain why that entry is physically distinguished.

If two people draw their reference arrows at different angles at the same starting mark, their transport matrices can look different. They should still agree on the maximum arrow displacement. The norm gives them a result they can compare without first agreeing on which way to hold their rulers.

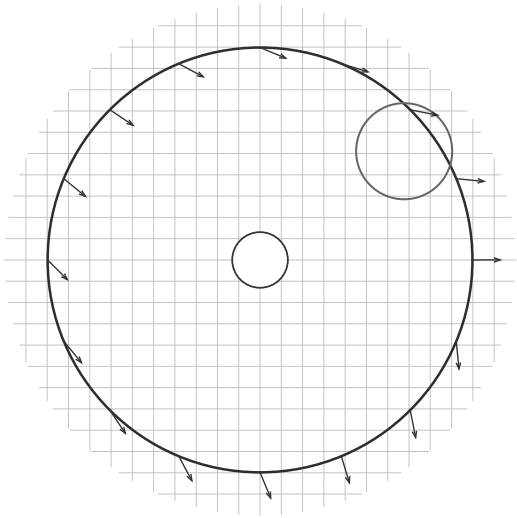
A flat local model with a nontrivial loop

Imagine a flat tabletop with its center point excluded. Compare a small circuit to one that goes all the way around the missing point. On this punctured plane use polar angle φ and a unitary line-bundle connection with local one-form $i\eta\, d\varphi$, where η is real. Its curvature is zero away from the missing origin because $d(d\varphi) = 0$ there. Parallel transport around one full turn has phase $e^{-2\pi i\eta}$ in the convention $\dot{s} + i\eta\dot{\varphi}s = 0$.

For $\eta = 1/4$, the phase is $-i$. A small loop that does not wind around the origin has phase one. The distinction lies in the allowed routes and the hole in the domain. Zero local curvature alone does not provide a single global parallel frame with trivial return along every loop.

A flat region can still have a global obstruction

A phase connection is flat away from the puncture; winding detects the missing point.



One winding: exp(−2πiη)

At η = 1/4: return phase −i

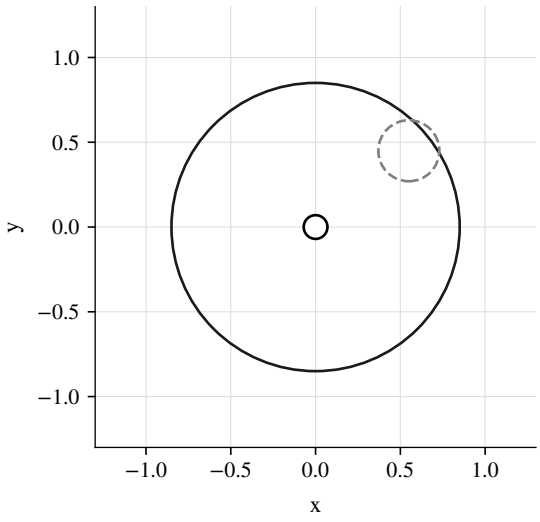
The small loop is contractible.

Its return phase is 1.

Arrow directions show the phase along the winding loop. They are fibers of a connection, not curvature of the base plane.

A Only one loop encloses the missing point

B A full winding can change the phase



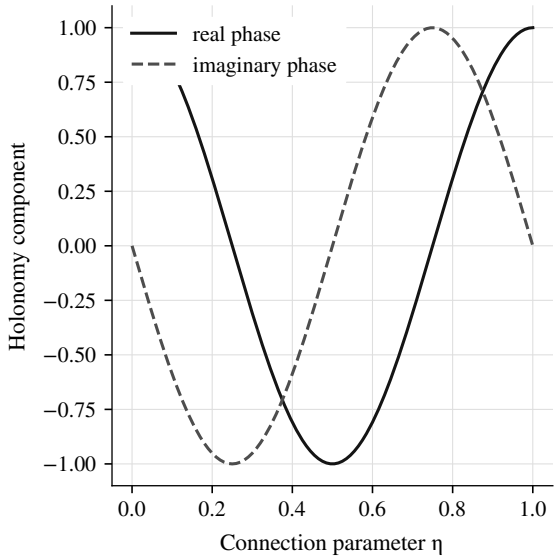
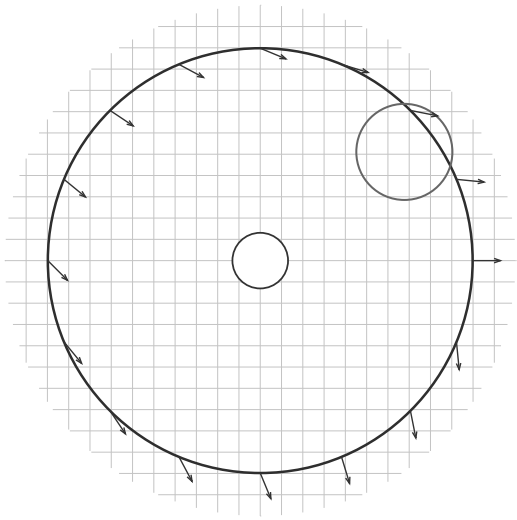


Figure 16.4: *Local flatness and global return are separate questions.* The study continues on the following figure page.

A flat region can still have a global obstruction

A phase connection is flat away from the puncture; winding detects the missing point.



One winding: exp(−2πiη)

At η = 1/4: return phase −i

The small loop is contractible.

Its return phase is 1.

Arrow directions show the phase along the winding loop.
They are fibers of a connection, not curvature of the base plane.

C The connection is flat away from the hole

D A local test cannot detect every global return

Connection form: iη dφ

Curvature is zero off the origin.

One winding gives exp(-2πiη).

η=1/4 gives -i.

A contractible small loop returns phase 1.

A loop around the hole need not.

This is bundle-connection holonomy.

It is not curvature of the flat base plane.

Figure 16.4: *Local flatness and global return are separate questions (continued).* A flat connection around a punctured plane can retain a nontrivial loop phase. The drawing separates a small contractible loop from a loop around the missing point and computes both transport integrals. This example concerns the connection on a complex line bundle, not the Gaussian curvature of the plane.

The survey construction on page 406 computes the same loop integral from local angle maps and distinguishes an integer turn count from a two-sheet sign record.

This is why a comparison system needs both its local rules and its admissible path class. "The same endpoint" is not always a complete record of how one description was transported into another.

A coordinate loop is a different experiment

Leave the sled and its arrow fixed at the starting mark. If we write the arrow's coordinates on three overlapping maps and return to the first map, the cocycle identity returns the original numbers. That operation has not moved the vector through tangent spaces at different base points. Parallel transport does. A nonzero transport holonomy is compatible with perfect coordinate bookkeeping.

Likewise, a learned translator loop can return a nonzero residual because the translators disagree. Calling both phenomena holonomy requires specifying the connection or comparison rule that makes the identification meaningful. An arbitrary conversion bug does not become curvature by receiving a geometric name.

For a connection written locally as a matrix-valued one-form A, transported components satisfy $\dot{v} + A(\dot{\gamma})v = 0$. Its curvature is $F = dA + A \wedge A$. These formulas summarize the local rule and the leading infinitesimal loop effect. Finite-loop transport still depends on the route and generally requires ordered composition. Matrices at different points need not commute.

What remains the same

The sled is back on its starting mark. The arrow pointed north when we set out. Now it points west. If I write down only where the sled ended up, I miss the quarter turn. In my work on transported identity, I ask what we need to keep when a comparison depends on the trip [24]. Here I'd keep the starting direction, the route, the rule for moving the arrow and which way it points when we get back.

A nonidentity transport matrix need not move every vector: a rotation in three dimensions fixes its axis. A diagnostic that tests only one vector can therefore miss a nontrivial loop transformation. Use the operator itself or enough independent probes to support the intended statement.

We can now name the point that returned, the route followed, and the transformation retained by transport. A scalar report of that transformation may still hide a direction. The next chapter calculates exactly how that loss can occur.

Practice with the construction

Back to the maps and the globe. Changing maps at one harbor is a coordinate operation. Sailing a closed route while transporting a local frame is a path operation. The globe supports both, and the mathematics keeps their records distinct.

16.1. Verify tangency and parallel transport of the three arrows in the worked spherical triangle by differentiating each vector.

16.2. Give a nonidentity three-dimensional rotation and a nonzero vector it fixes. What does that tell you about a one-vector holonomy test?

Chapter 17

What a single score can conceal

Two independent joints each have a sensor that measures how far the joint turned. The dashboard shows only the larger turn. Can an operator tell whether both joints moved? We use the joint readings to test several summaries, then ask how the chosen metric can certify a complete update.

One dial can hide two motions

Record each joint's angle by the two components of a unit arrow on its dial. Turning each joint rotates its own pair of components. The four coordinates describe two dials that can turn independently. For their combined rotation take

$$R = \operatorname{diag}(R_{\theta_1}, R_{\theta_2}) \in SO(4), \qquad R_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

The two coordinate planes rotate independently. With principal angles $\theta_j \in [-\pi, \pi]$, three different summaries are

$$\|I - R\|_2 = 2 \max_j |\sin(\theta_j/2)|,$$

$$\|I - R\|_F^2 = 8 \sum_{j=1}^{2} \sin^2(\theta_j/2), \qquad \frac{\operatorname{tr} R}{4} = \frac{\cos\theta_1 + \cos\theta_2}{2}.$$

The operator norm sees the largest plane deviation. The squared Frobenius norm and normalized trace encode the same pooled information: $\|I - R\|_F^2 = 8(1 - \operatorname{tr} R/4)$. Either pooled value can be recovered from the other.

Under the bi-invariant metric with Lie-algebra squared norm $\|\Omega\|^2 = -\frac{1}{2}\operatorname{tr}\Omega^2$, the distance from identity is $\sqrt{\theta_1^2 + \theta_2^2}$ using a minimizing principal-angle logarithm. This is another quantity, with a stated metric convention.

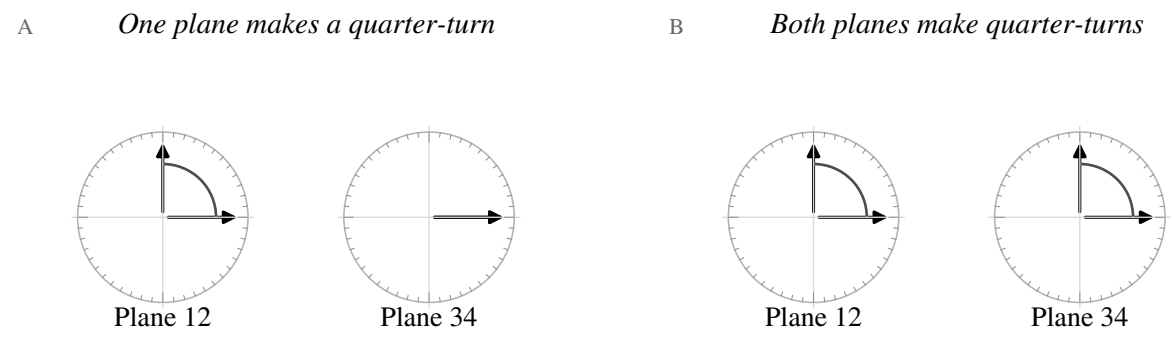
Compare $A = \operatorname{diag}(R_{\pi/2}, I_2)$ with $B = \operatorname{diag}(R_{\pi/2}, R_{\pi/2})$. Both have $\|I - R\|_2 = \sqrt{2}$. Their Frobenius distances are 2 and $2\sqrt{2}$, their normalized traces are $1/2$ and 0, and their intrinsic distances are $\pi/2$ and $\pi/\sqrt{2}$. Equal operator deviations do not make these rotations equal.

Choose the observable that serves the task

An aircraft orientation in three spatial dimensions has one principal rotation angle. A higher-dimensional representation can contain several rotation planes. Extending a one-angle formula requires specifying which aggregate it describes.

Marcella's accumulated-rotation work distinguishes a clipped diagnostic from a normalized trace [26]. The exact matrix examples above show why that distinction matters: calling both quantities a score does not make them interchangeable.

An operator deviation bounded by δ controls every unit vector's displacement by δ, by definition of the operator norm. An average-trace bound does not supply the same worst-direction control without a derived inequality. Choose the score after the question.



Summary	*A: (π/2, 0)*	*B: (π/2, π/2)*
Operator deviation	$\sqrt{2} = 1.4142$	$\sqrt{2} = 1.4142$
Frobenius deviation	2.0000	$2\sqrt{2} = 2.8284$
Normalized trace	0.5000	0.0000
Intrinsic distance	$\pi/2 = 1.5708$	$\pi/\sqrt{2} = 2.2214$

Figure 17.1: Two independent rotation planes, with one plane active in case A and both active in case B. The complete comparison shows why an equal operator deviation does not fix the Frobenius deviation, normalized trace or intrinsic distance. The metric convention is the one stated in the text.

The first pair of rotations shared the same largest error while differing in total motion. Test the converse using the same two-joint mechanism, hold the average fixed and inspect the largest error. These two tests establish which question each proposed scalar answers before we change the ruler.

Two hinges can hide inside one number

Read two independent joint dials. The first arrow turns through α in its two display coordinates, the second through β in its own pair. In the combined record these are orthogonal coordinate planes. Their action is represented by the four-dimensional orthogonal matrix

$$H = \begin{pmatrix} R(\alpha) & 0 \\ 0 & R(\beta) \end{pmatrix}, \qquad R(t) = \begin{pmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{pmatrix}.$$

Four dimensions here describe two two-component signals, not a claim that the mechanism occupies four spatial directions. The model is also a useful local example for a vector bundle with two invariant rotation planes.

The normalized trace is

$$s_{\mathrm{tr}} = \frac{\operatorname{tr} H}{4} = \frac{\cos\alpha + \cos\beta}{2}.$$

For $(\alpha, \beta) = (0, \pi)$ it is zero. For $(\alpha, \beta) = (\pi/2, \pi/2)$ it is also zero. In the first system one signal plane is unchanged and the other reversed. In the second both are quarter-turned. A zero trace has concealed that difference.

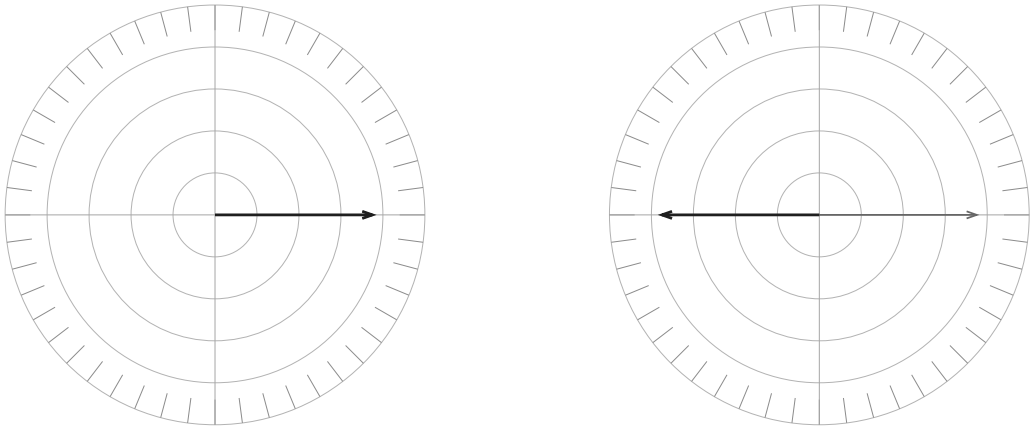
The operator discrepancy and Frobenius discrepancy answer different questions:

$$\|I - H\|_{\mathrm{op}} = 2\max\{|\sin(\alpha/2)|, |\sin(\beta/2)|\},$$
$$\|I - H\|_F^2 = 8\sin^2(\alpha/2) + 8\sin^2(\beta/2).$$

The operator norm is two for the first example and $\sqrt{2}$ for the second. The squared Frobenius norm is eight in both. If I'm checking whether either joint made a half-turn, adding up the squared changes from both joints gives me a tie. Looking at the largest change separates the two cases. The individual dials tell me which joint moved.

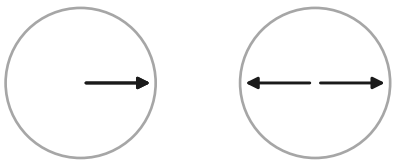
The angle pair contains what an average loses

Separate views of two invariant 2-planes in four dimensions; each disk is one plane.



Both cases: normalized trace 0; squared Frobenius error 8. The operator errors differ.

A One plane stays fixed; one makes a half-turn



B Both planes make quarter-turns

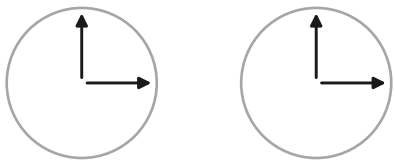
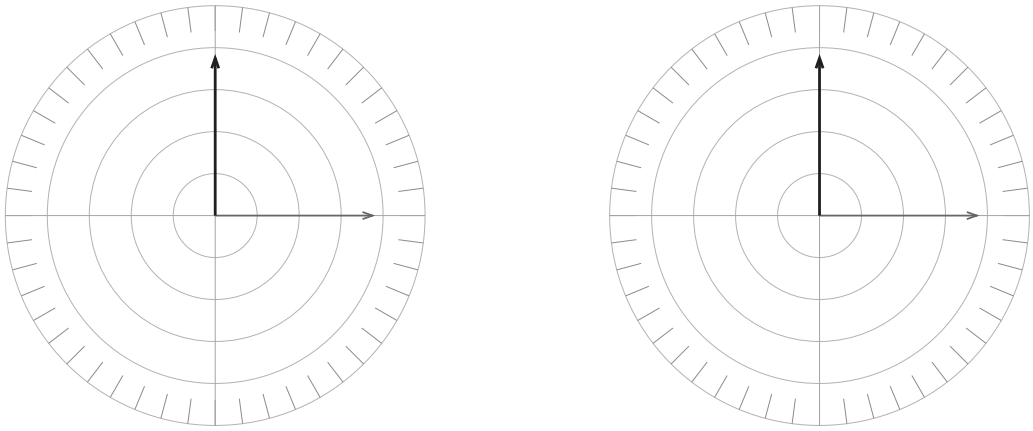


Figure 17.2: *Equal averages can conceal different directions.* The study continues on the following figure page.

The angle pair contains what an average loses

Separate views of two invariant 2-planes in four dimensions; each disk is one plane.



Plane 1: 90° Plane 2: 90°

Both cases: normalized trace 0; squared Frobenius error 8. The operator errors differ.

C Two scalar summaries fail to distinguish them

D The worst-direction errors remain different

Both normalized traces = 0	Angles (0,π): operator error 2
Both squared Frobenius errors = 8	Angles (π/2,π/2): error $\sqrt{2}$
The individual angle pairs differ.	A mean does not recover the angle spectrum.

Figure 17.2: *Equal averages can conceal different directions (continued).* Two block rotations have the same normalized trace and Frobenius discrepancy but different worst-direction errors. The figure shows the individual planes, their eigenangles and the corresponding norm calculations. No scalar is labeled a complete description of the transformation.

Which directions does the task care about?

A controller might use one joint reading to steer while it uses the other only to log motion. Equal errors in the two readings can have different consequences. A metric assigns the cost of an error, including how errors in several coordinates interact. With a positive-definite metric G, the squared error of e is $e^T G e$. The associated operator norm is

$$\|A\|_G = \sup_{x \neq 0} \frac{\sqrt{(Ax)^T G (Ax)}}{\sqrt{x^T G x}}.$$

Writing $G = L^T L$ transforms this into the Euclidean norm $\|LAL^{-1}\|_2$. The metric specifies which combinations of coordinates are expensive. Its off-diagonal terms can make correlated changes cheap and opposed changes expensive, or the reverse.

For example,

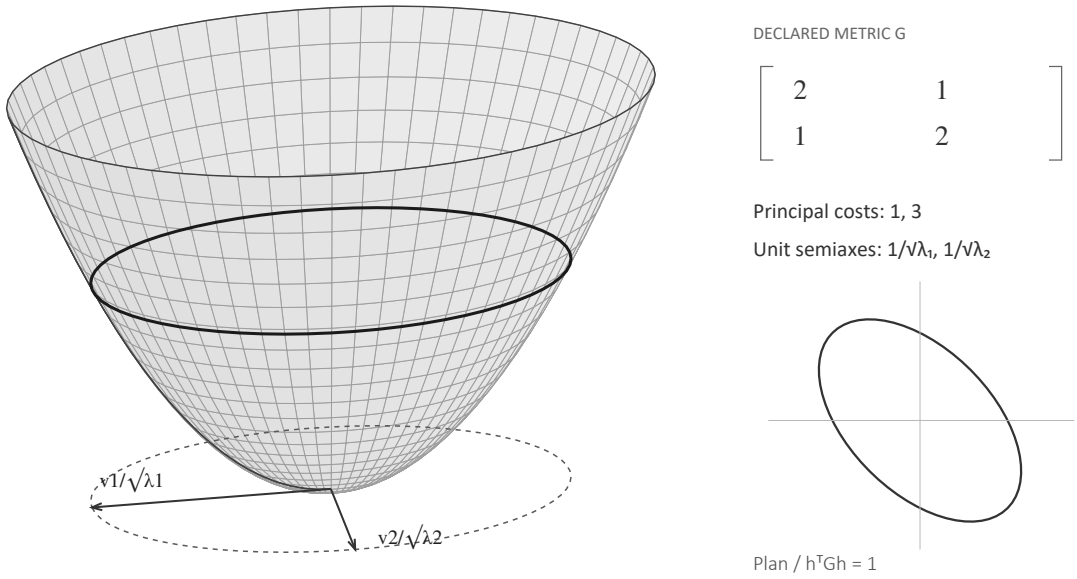
$$G = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$$

has eigenvectors along $(1, 1)$ and $(1, -1)$, with eigenvalues three and one. The vectors $(1, 1)$ and $(1, -1)$ have equal ordinary Euclidean length, but squared G-lengths six and two. Deleting the off-diagonal entries assigns both squared length four. That deletion changes the comparison rule.

Turn the ellipse without changing its shape. The eigenvalues and trace stay the same, but the same error can now cost more or less. If my decision depends on which way the error points, I need to keep track of which way the ellipse tilts.

The ruler is visible in its level sets

Constructed cost surface: height = $h^T G h$. The bold section has unit cost.



A Cross terms tilt the equal-cost ellipses

B Directional cost reveals the coupling

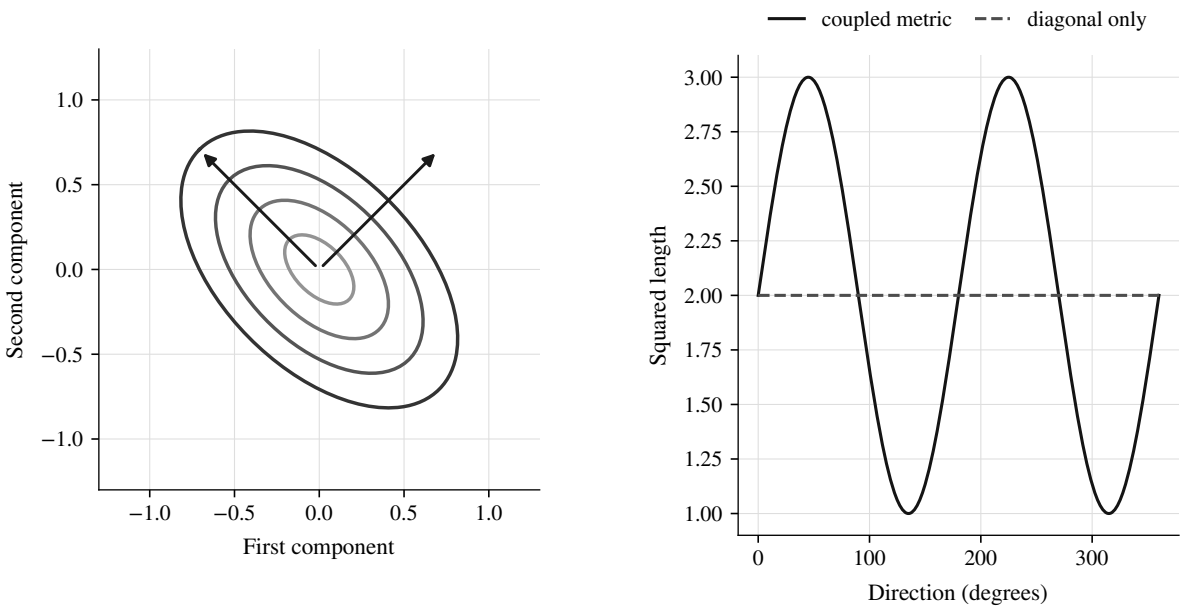
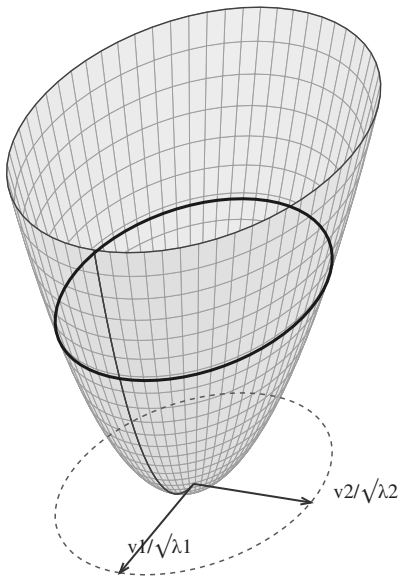


Figure 17.3: *The tilted ellipse is part of the measurement.* The study continues on the following figure page.

The ruler is visible in its level sets

Constructed cost surface: height = $h^T G h$. The bold section has unit cost.

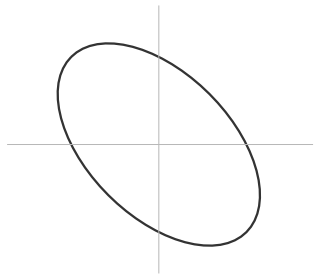


DECLARED METRIC G

$$\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

Principal costs: 1, 3

Unit semiaxes: $1/\sqrt{\lambda_1}$, $1/\sqrt{\lambda_2}$



Plan / $h^T G h = 1$

C The off-diagonal entries carry the coupling

D Equal Euclidean lengths need not have equal cost

$$\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

v=(1,1): $v^T G v=6$

w=(1,-1): $w^T G w=2$

Delete cross terms: both become 4.

The changed metric changes the comparison.

Figure 17.3: *The tilted ellipse is part of the measurement (continued).* The same two error vectors are measured by a coupled metric and its diagonal restriction. Principal directions, equal-cost ellipses and explicit quadratic-form values show the consequence of dropping cross terms. The example is exact and two-dimensional.

The two-joint examples showed what a scalar report hides. The next test asks how much the ruler itself hides or reveals about an update. We keep the complete matrix and compare its action in two metrics, so the conclusion includes both the measured effect and the price of changing the measurement rule.

A small nonnormal system gives a complete certificate

Imagine an update for two recorded error coordinates. At each step both errors are halved, but the first coordinate also receives the previous value of the second. This models the coupling studied in the constructive-metric work [31]. Let

$$A = \begin{pmatrix} 1/2 & 1 \\ 0 & 1/2 \end{pmatrix}.$$

Both eigenvalues are $1/2$, yet one application can enlarge the Euclidean norm. The eigenvalues alone do not measure that transient behavior. Seek a metric G in which the step is contractive by solving

$$G - A^T G A = I.$$

Write $G = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$. Equating entries gives

$$\frac{3}{4}a = 1, \qquad \frac{3}{4}b - \frac{1}{2}a = 0, \qquad \frac{3}{4}c - a - b = 1.$$

Consequently

$$G = \begin{pmatrix} 4/3 & 8/9 \\ 8/9 & 116/27 \end{pmatrix}.$$

Its determinant is positive and its first diagonal entry is positive, so it is positive definite. Direct substitution verifies the matrix equation. For every nonzero x,

$$\|Ax\|_G^2 = \|x\|_G^2 - \|x\|_2^2 < \|x\|_G^2.$$

If I'm watching the error on an ordinary coordinate plot, its length can still grow on the first step. In the quadratic ruler, each step takes it inside a smaller equal-cost ellipse.

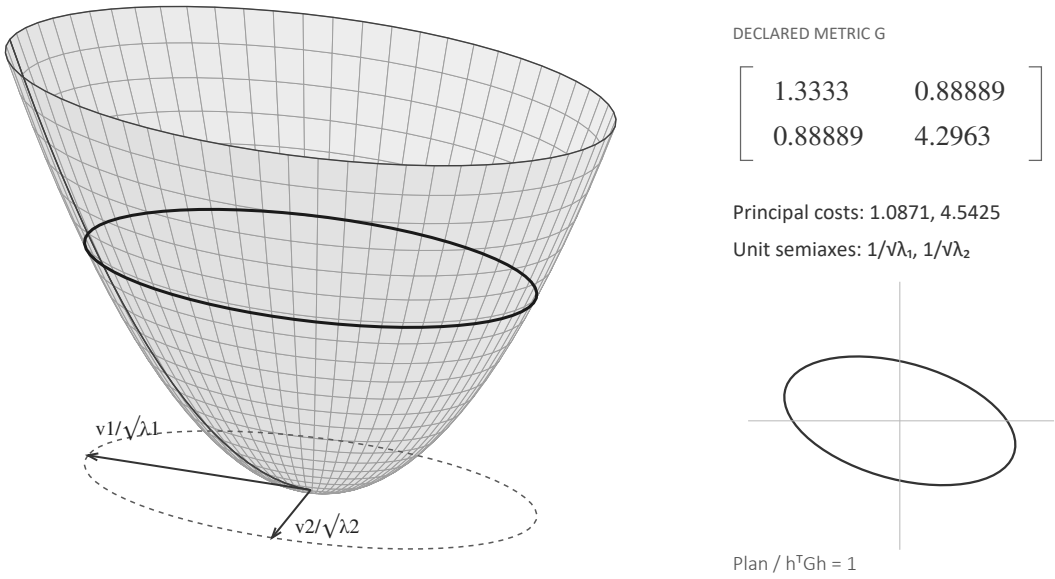
The comparison bounds

$$\lambda_{\min}(G)\|x\|_2^2 \le \|x\|_G^2 \le \lambda_{\max}(G)\|x\|_2^2$$

translate the result back to the original coordinates. If an instrument has a hard limit in those units, I still have to check it. An elongated ellipse can contain a point well beyond that limit.

The ruler is visible in its level sets

Constructed cost surface: height = $h^T G h$. The bold section has unit cost.



A The state crosses shrinking metric ellipses

B Euclidean growth can coexist with contraction

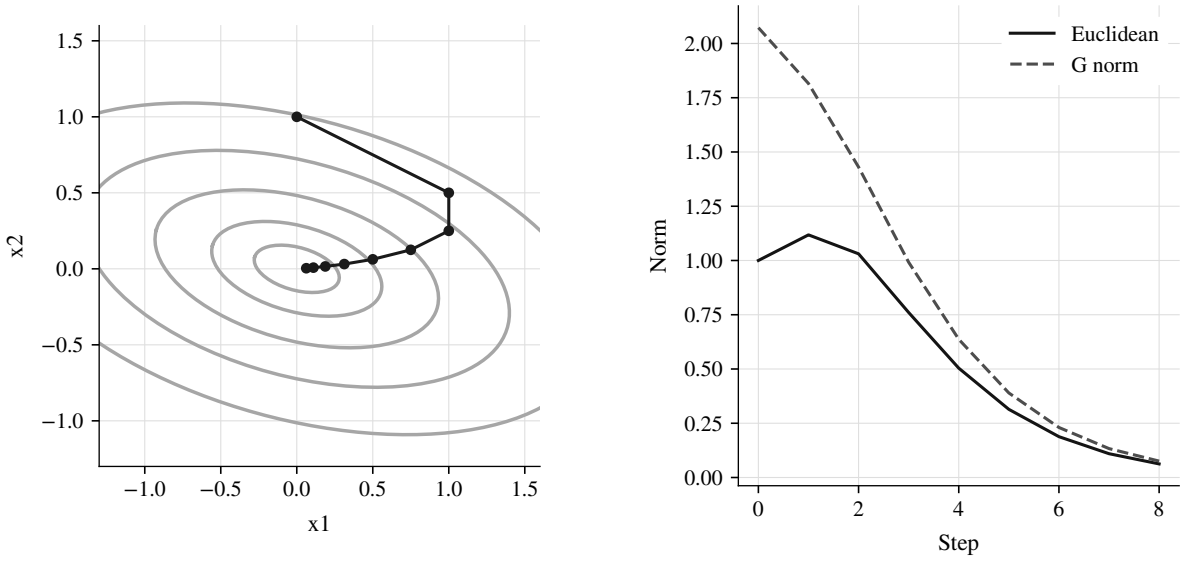
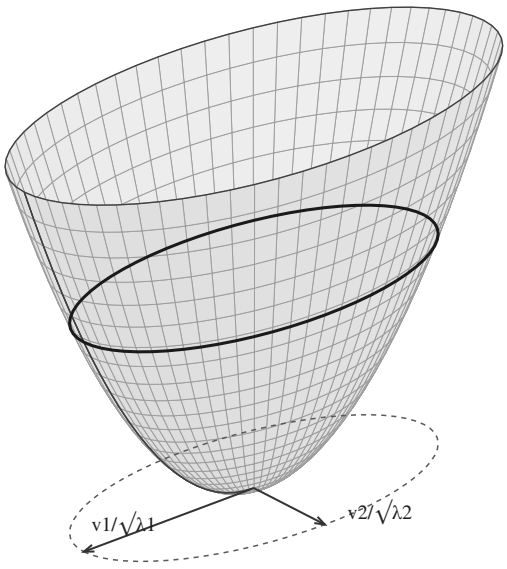


Figure 17.4: *A decreasing metric and a transient excursion.* The study continues on the following figure page.

The ruler is visible in its level sets

Constructed cost surface: height = $h^T G h$. The bold section has unit cost.

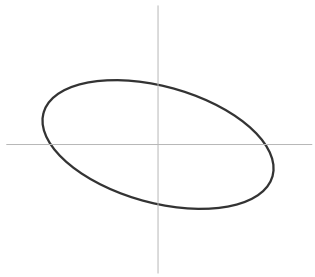


DECLARED METRIC G

$$\begin{bmatrix} 1.3333 & 0.88889 \\ 0.88889 & 4.2963 \end{bmatrix}$$

Principal costs: 1.0871, 4.5425

Unit semiaxes: $1/\sqrt{\lambda_1}$, $1/\sqrt{\lambda_2}$



Plan / $h^T G h = 1$

C The exact residual certifies the decrease

D Changing rulers has a quantified cost

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

G11=4/3; G12=8/9; G22=116/27

λmin(G)=1.087125

λmax(G)=4.542504

Norm cost $\sqrt{\kappa}$(G)=2.044127

The Euclidean transient remains real.

Figure 17.4: *A decreasing metric and a transient excursion (continued).* The nonnormal matrix is followed in ordinary coordinates and in the exact quadratic metric. Equal-cost ellipses, successive states and the identity $G - A^T G A = I$ are displayed together. The source of the cross terms and the cost of changing rulers are explicit.

What a research bridge should preserve

The later Marcella and operator work asks which geometric observables support a particular computation [26, 27]. The examples here give a test for any proposed bridge. State the object, the metric, the map and the observable. Compute them for a case with more than one active direction. Then compare the result with what the scalar summary retained.

In a single rotation plane, a chord, an angle and an overlap may determine one another on a specified interval. With several planes, the same average can arise from different angle spectra. A formula that was exact in the first model needs additional hypotheses in the second. Those hypotheses are not ornamental restrictions. They identify what has to be measured if a system is to make the same decision after a change of representation.

The largest-joint dial could not reveal whether the second joint moved. The trace could agree for mechanisms with different motions. The operator example then showed how retaining the coupled directions changes a verifiable bound. Each calculation asks what the chosen report lets us conclude about the whole state. We now assemble the observation, ruler, statistic and decision so that the required distinctions remain visible throughout the system.

Practice with the construction

Back to the maps and the globe. A map may show the steepest slope or the average elevation. Both can be useful descriptions of the same terrain. They cannot replace each other merely because both are printed as one number.

17.1. Reproduce the four summaries for the first pair of joint settings, A with one quarter-turn and B with two.

17.2. Choose $(\theta_1, \theta_2) = (\pi/3, \pi/6)$ and compute the operator deviation and intrinsic distance. Explain which motion information each retains.

Part V

The Davis manifold and defensible decisions

Chapter 18

Constructing a Davis manifold

Imagine an engineer monitoring a pump. She wants to tell which side of a reference pressure it is operating on, even as the temperature varies. The Davis construction starts with that distinction and builds the geometry that connects the measurements to a decision.

A geometry built for a question

During pump startup, temperature may drift while pressure remains on the same side of its reference. A pressure change that crosses the reference matters to this comparison. A detector must distinguish these directions and leave readings too close to the reference undecided.

A complete small construction

Model the pump state as (x, y), with pressure and temperature departures scaled to dimensionless coordinates. Let $M = \mathbb{R}^2$ with

$$g = \begin{pmatrix} 4 & 0 \\ 0 & 1 \end{pmatrix}, \qquad d_g(p, q) = \sqrt{4(x_p - x_q)^2 + (y_p - y_q)^2}.$$

The metric assigns twice as much length to a unit change in the first coordinate as to a unit change in the second. This weighting is a declared task model. A physical or statistical justification must accompany such a choice in a real application.

Let the configuration distinction be the sign of x, but restrict confident labels to $x \le -m$ and $x \ge m$ for some $m > 0$. The central strip is undecided. Any pair from the two confident regions has metric distance at least $4m$: the smallest x separation is $2m$ and its metric length is twice that. This is a proved geometric margin. If each encoded endpoint has metric error at most e, their observed separation can decrease by at most $2e$.

Choose $m = 0.1$ and $e = 0.03$. The true cross-configuration separation is at least 0.4, while the encoded separation is at least 0.34. This is a deterministic calculation in the specified model. It becomes a statistical statement only after we specify how often the endpoint-error bound and the region assumptions hold.

Record the choices in the Davis framework

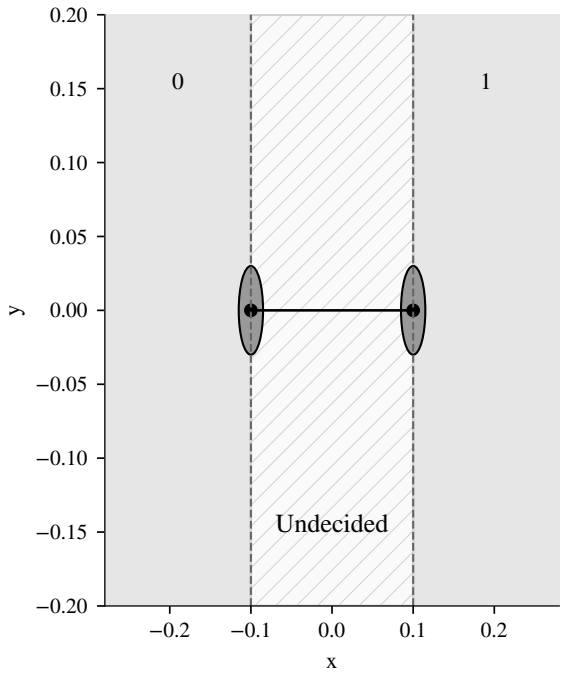
The Davis manifold framework organizes these choices into a geometric model and then adds the operational decision system [17]. The separation between construction and traversal is useful. Construction chooses the space, metric, encoding and permissible paths. Traversal describes the paths an operating system actually follows and the statistics it computes along them.

We retain the framework's geometric ingredients in the tuple

$$\mathcal{D}_{\text{geom}} = (\mathcal{X}, M, g, \phi, \psi, \mathcal{P}, \varepsilon, \mathcal{C}, c, h, \kappa_{\text{hard}}, \kappa_{\text{soft}}).$$

Here $\mathcal{X}$ is the observation space; ϕ maps observations to the manifold (M, g); ψ supplies coordinates on the operating region; $\mathcal{P}(L)$ is a declared family of allowed paths up to horizon L; and $\varepsilon(L)$

A *A margin separates the regions*



$g = \mathrm{diag}(4, 1), \quad m = 0.1, \quad e = 0.03$

Each small ellipse is an endpoint error ball: $4\,dx^2 + dy^2 \le e^2$.

B *Endpoint errors consume the margin*

Bound	*Metric length*
True separation ≥ 4m	0.40
First endpoint error	≤ 0.03
Second endpoint error	≤ 0.03
Encoded separation	≥ 0.34

Figure 18.1: The stated configuration rule leaves an undecided strip. Endpoint uncertainty balls are ellipses in the coordinate drawing because the metric is $\mathrm{diag}(4, 1)$. Their two error budgets reduce the guaranteed separation from 0.40 to 0.34.

bounds the selected distance approximation there. The configuration label set $\mathcal{C}$, configuration map c and coarse labeling h specify what distinction matters. The two margins state the separation required for the associated hard and soft decisions. Coordinates are local unless a global chart has actually been supplied.

A Davis system adds a statistic, decision rule, uncertainty accounting, operational horizon and an abstention rule. This is why a manifold alone is not a detector. A geometry can describe the data faithfully while the decision threshold or its calibration is unsuitable.

Construction does not follow from a training objective alone

A contrastive loss can encourage different configurations to separate and similar observations to approach. To prove the resulting margin, however, one needs the actual inequalities that the trained representation satisfies on the relevant domain. A small average training loss does not by itself guarantee a uniform gap for every unseen observation. The bridge from training to geometry may use a generalization result, coverage plus regularity, or a bounded operational region with a separate verification.

A smoothness penalty can encourage small derivatives. If the proof needs a derivative bound, state the bound, the norm it uses, where it holds and how you know.

The pump's geometric margin says how far apart its confident pressure regions are. Now ask whether the score reported by the instrument can detect that separation.

Build the entire instrument model on one page

Keep the pump's two dimensionless readings x and y. The first is the signed departure from reference pressure and the second is the temperature departure. Use the constant metric $G = \operatorname{diag}(4, 1)$ and the identity encoder. Straight lines are geodesics, and the distance is $\sqrt{4\Delta x^2 + \Delta y^2}$.

Now specify the operational choices. Confident configuration C_- contains $x \le -0.1$; confident configuration C_+ contains $x \ge 0.1$. The middle strip is undecided. Allow piecewise smooth paths within the observed operating rectangle. Define the statistic $s(x, y) = 2x$. It is one-Lipschitz in the chosen metric, because

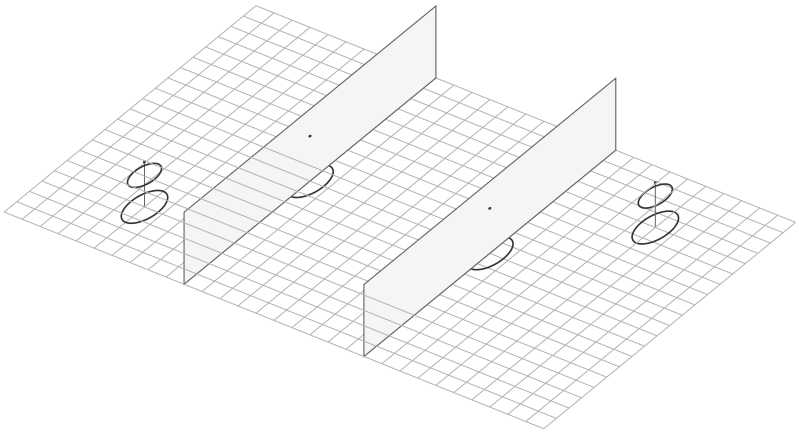
$$|s(p) - s(q)| = 2|\Delta x| \le \sqrt{4\Delta x^2 + \Delta y^2}.$$

The two confident regions satisfy $s \le -0.2$ and $s \ge 0.2$. The geometry has supplied a scalar separation because we explicitly chose a statistic that detects the configuration distinction.

If encoded observations have metric error at most $e = 0.03$, their statistic error is at most 0.03. A threshold at zero still separates both confident regions. If an additional calibration error contributes at most 0.02 to the statistic, the combined error is at most 0.05 by the triangle inequality. The remaining statistic margin is 0.15 on either side.

Error has a direction as well as a size

Constructed decision geometry / G = diag(4,1); the separating statistic is s = 2x.



Confident regions: x ≤ −0.1

x ≥ 0.1

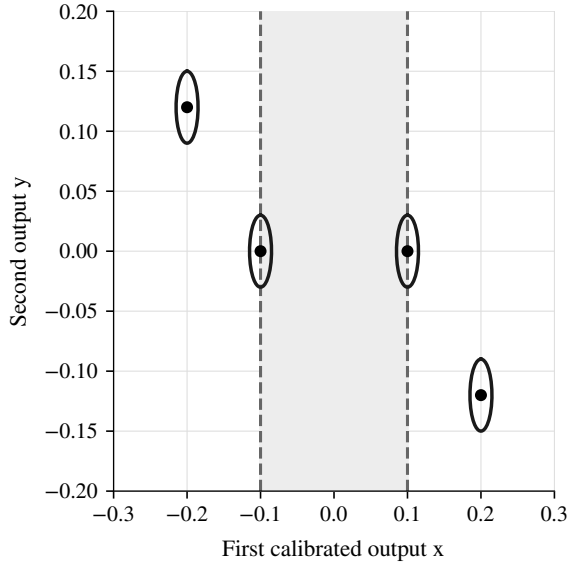
Metric error allowance: 0.03

Score calibration allowance: 0.02

The remaining margin is explicit.

Combined score allowance ≤ 0.05
Remaining confident margin ≥ 0.15
Vertical outlines are explanatory;
their height is not another variable.

A Error ellipses meet a separating region



B The statistic reads the separating direction

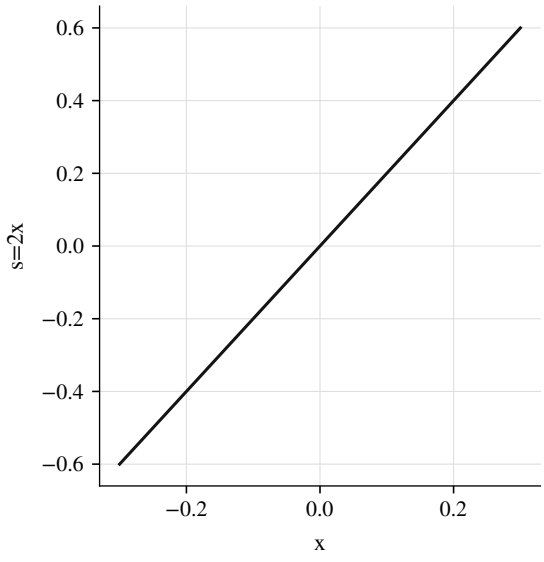
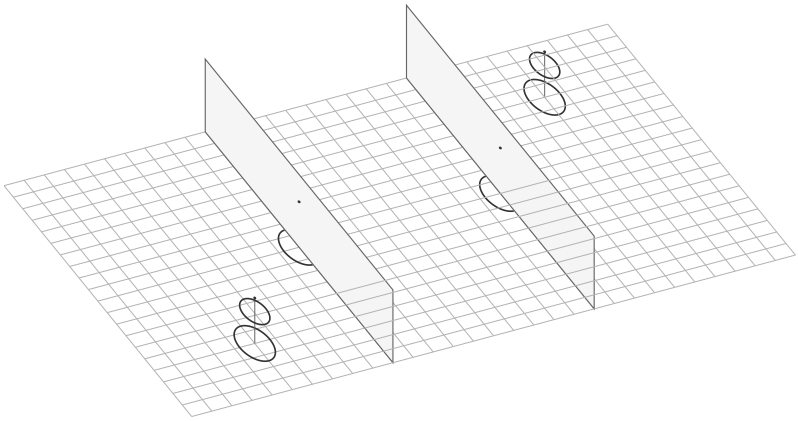


Figure 18.2: *A Davis construction from measurement to decision.* The study continues on the following figure page.

Error has a direction as well as a size

Constructed decision geometry / G = diag(4,1); the separating statistic is s = 2x.



Confident regions: x ≤ −0.1

x ≥ 0.1

Metric error allowance: 0.03

Score calibration allowance: 0.02

The remaining margin is explicit.

Combined score allowance ≤ 0.05
Remaining confident margin ≥ 0.15
Vertical outlines are explanatory;
their height is not another variable.

C The model fixes the ruler and the margin

G=diag(4,1)
Confident regions: x≤-0.1, x≥0.1
Geometric margin = 0.4
Statistic margin = 0.2 each side
s=2x is 1-Lipschitz.

D Each approximation spends part of the margin

Metric encoding error ≤0.03
Score calibration error ≤0.02
Combined score error ≤0.05
Remaining confident margin ≥0.15
The alternative statistic y misses x.

Figure 18.2: *A Davis construction from measurement to decision (continued).* The operating region, metric ellipses, configuration strip, statistic and bounded perturbations appear in one coordinated drawing. The calculation connects each error allowance to a measured or stipulated quantity. The numbers define a constructed example, not an estimated field performance.

A distant pair can still have the same score

Replace the statistic by $r(x, y) = y$. It is also one-Lipschitz, but the points $(-0.1, 0)$ and $(0.1, 0)$ have equal score zero despite belonging to different confident configurations. Their metric separation is 0.4.

This shows exactly why a geometric margin alone cannot prove a margin for an arbitrary detector. A Lipschitz condition supplies an upper bound on score change, not a lower bound. To transfer geometric separation into statistic separation, the system needs a suitable linkage condition. It can be proved for a specified statistic, imposed as a hypothesis or estimated on a declared population with a corresponding uncertainty statement.

Picture the pump display. One reading puts the pressure below reference and the other puts it above, but both show the same temperature. A temperature-only score still says zero, however precisely you measure it. To tell those states apart, the display needs the pressure channel.

Choose a path horizon with the actual dynamics in view

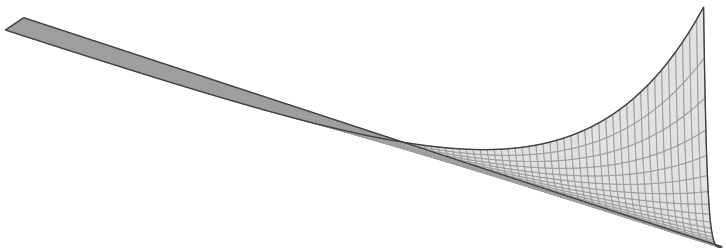
Suppose the pump's calibrated state $z = (x, y)$ evolves under a vector field $\dot{z} = f(z)$ and two nearby trajectories start with separation d_0. If f is L-Lipschitz in the chosen norm on their operating region, Gronwall's inequality gives $d(t) \le e^{Lt} d_0$. A finite initial uncertainty therefore grows into a horizon-dependent allowance.

For $d_0 = 0.01$ and $L = 0.2\,\mathrm{s}^{-1}$, the bound at five seconds is about 0.02718 and at ten seconds about 0.07389. A usable horizon depends on the available margin. It is not a universal property of the model's name.

In a detector, longer histories may reveal a developing configuration more clearly while accumulating more model mismatch. The Davis framework makes this tradeoff explicit through path families and horizon-dependent distortion [17]. A good implementation compares the gain in separation with the growing uncertainty, then records the selected horizon beside the decision.

The allowance accumulates by a stated rule

Each construction below separates an exact bound from its simplifying approximation.



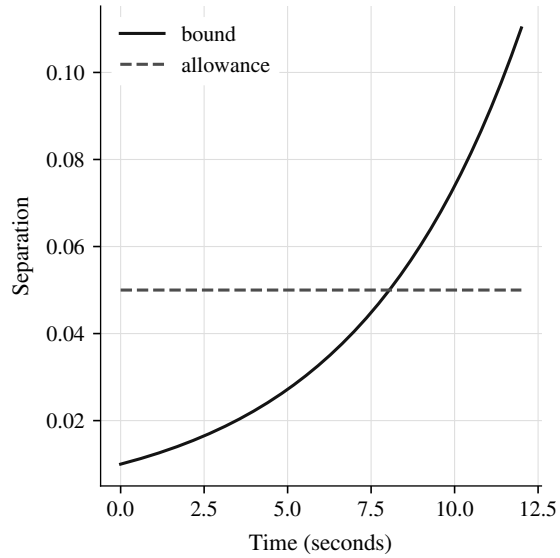
PARAMETER STUDY

$$d(t) \le 0.01 \exp(Lt)$$

Horizontal: horizon t (seconds)
Depth: Lipschitz bound L (1/s)
Height: separation bound
A constructed parameter study.

Population, model, operating region and valid parameter range belong with every reported allowance.

A A longer horizon increases uncertainty



B The declared growth law fixes the crossing

Initial separation d0=0.01

Lipschitz rate L=0.2 per second

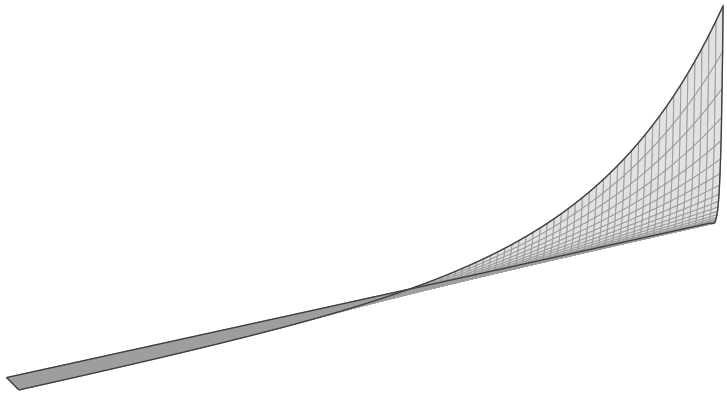
Bound d(t)≤0.01 exp(0.2t)

Allowance 0.05 lasts to 8.047190 s

Figure 18.3: *A longer observation window spends a margin.* The study continues on the following figure page.

The allowance accumulates by a stated rule

Each construction below separates an exact bound from its simplifying approximation.



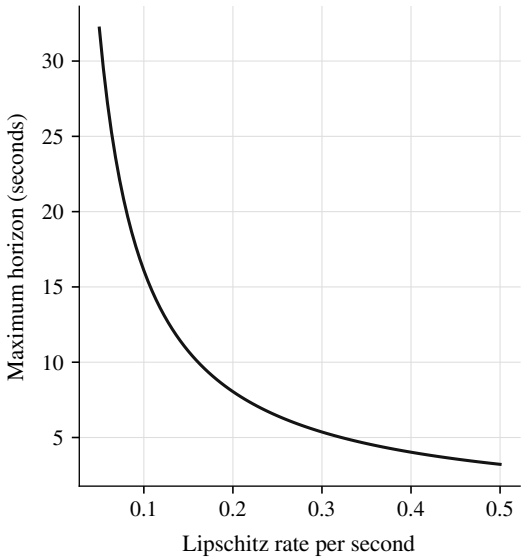
PARAMETER STUDY

$d(t) \leq 0.01 \exp(Lt)$

Horizontal: horizon t (seconds)
Depth: Lipschitz bound L (1/s)
Height: separation bound
A constructed parameter study.

Population, model, operating region and valid parameter range belong with every reported allowance.

C Faster growth shortens the useful horizon



D The bound needs a common operating region

Both trajectories stay in the region.

The same norm is used throughout.

L is a valid bound for the dynamics.

A longer history can add evidence and accumulate mismatch.

Figure 18.3: *A longer observation window spends a margin (continued).* A declared trajectory-separation bound is compared with a fixed decision allowance. The figure identifies the last horizon satisfying the inequality and shows which assumptions determine it. The curve is a model bound, not a fitted forecast of instrument errors.

In the pump model, the geometry gives pressure departures more weight than temperature departures. A statistic that reads pressure can distinguish the two confident regions; one that reads temperature alone cannot. The uncertainty allowance grows over the stated time horizon. Next we ask what survives when recalibration changes the ruler used to compare readings.

Practice with the construction

Back to the maps and the globe. The Davis construction chooses the globe that makes the task's distinctions measurable, then checks whether the observation maps preserve the needed paths and margins. Its operational value comes from keeping all those choices visible.

18.1. For the displayed metric, calculate the distance between $(0.1, 0)$ and $(-0.1, 0.2)$.

18.2. How large can e be before the lower separation bound $4m - 2e$ becomes uninformative? State the corresponding condition in plain English.

Chapter 19

Comparing with a changing metric

The pump model gave pressure and temperature departures different weights. A new calibration or a new task can change those weights while the readings stay fixed. We will bound the effect on distances, then use the same comparison to see what an operator bound costs in a different metric.

Two rulers on one region

Suppose a survey company updates its calibration. The underlying positions have not changed, but the new ruler weights small displacements differently. Before reusing a route-length tolerance, we need a bound on how much every allowable route can change.

Theorem 19.1 (Metric comparison)**.** *Let g and $\tilde{g}$ be Riemannian metrics on the same region and let all compared curves lie there. Suppose, as quadratic forms,*

$$a^2 g_p \preceq \tilde{g}_p \preceq b^2 g_p, \qquad 0 < a \le b,$$

at every point used by those curves. Then their lengths satisfy $aL_g \le L_{\tilde{g}} \le bL_g$. Taking infima over the same admissible curve family gives $ad_g \le d_{\tilde{g}} \le bd_g$.

Proof. Apply the quadratic-form bounds to the velocity of each curve, take square roots and integrate. The same curve set appears on both sides, so taking its infimum preserves the two inequalities. □

Suppose the survey crew logged a route at 100 metres. Under $g \preceq \tilde{g} \preceq 1.21g$, the same route reads between 100 and 110 metres after recalibration. A 105-metre limit would need a fresh calculation. The square root appears because the tensor controls squared speed. If instead $(1-\epsilon)g \preceq \tilde{g} \preceq (1+\epsilon)g$, the length factors are $\sqrt{1-\epsilon}$ and $\sqrt{1+\epsilon}$.

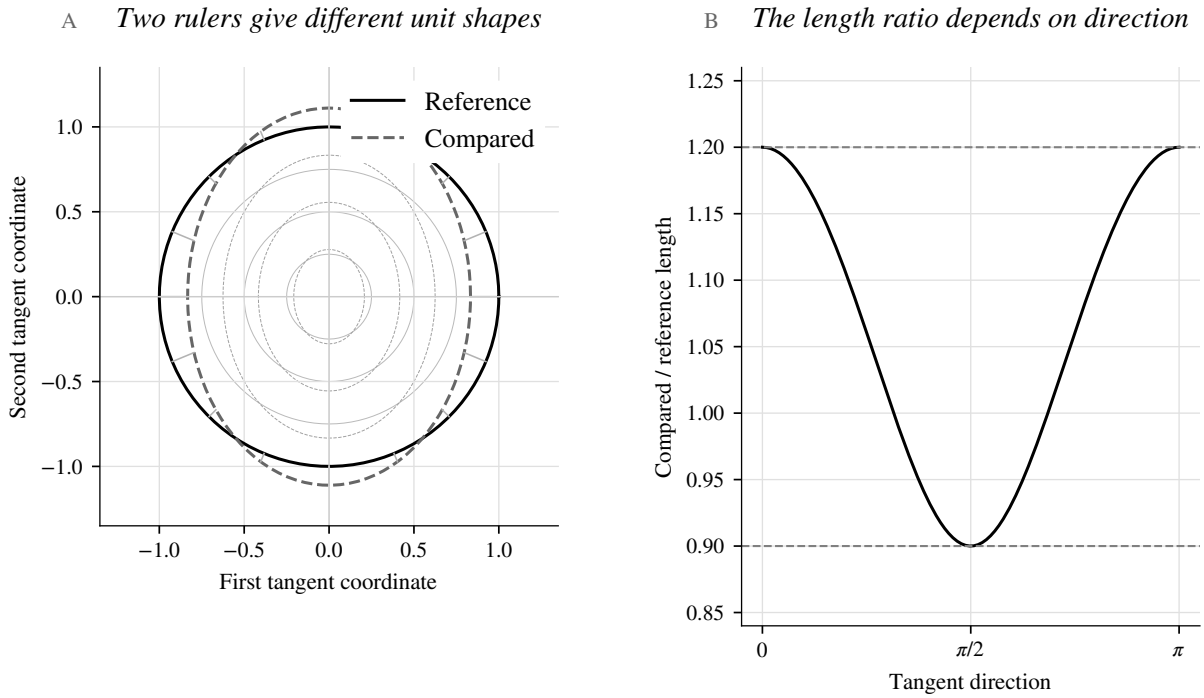
The survey calculation bounds the effect of replacing one metric with another. Calibration may also drift continuously during a survey. To keep using the original route tolerance, we need to bound how far the ruler can drift during the run.

When the ruler evolves

Suppose a time-dependent metric satisfies $|\partial_t g_t(v,v)| \le \kappa g_t(v,v)$ for every tangent vector and time in an interval. For fixed p, v, integrate the scalar differential inequality to obtain

$$e^{-\kappa t} g_0 \preceq g_t \preceq e^{\kappa t} g_0.$$

The length and distance factors are $e^{-\kappa t/2}$ and $e^{\kappa t/2}$ over a common region and common path family. If the admissible region or the path rules also change, their contribution must be handled separately. The time derivative is a statement about the actual metric evolution, not about the mere passage of time.



$$G = I, \quad H = \operatorname{diag}(1.44, 0.81), \qquad 0.9\|v\|_G \le \|v\|_H \le 1.2\|v\|_G$$

Figure 19.1: Unit-length sets for two constant positive-definite metrics, followed by their directional length ratio. The extremal ratios 0.9 and 1.2 are square roots of the generalized eigenvalue bounds; they apply to every tangent direction.

Compare the two rulers in their own principal directions

Let G and $\widetilde{G}$ be positive-definite matrices. To find the sharp local length-ratio bounds, form

$$C = G^{-1/2}\widetilde{G}G^{-1/2}.$$

For $w = G^{1/2}v$, the squared ratio is

$$\frac{v^T\widetilde{G}v}{v^TGv} = \frac{w^TCw}{w^Tw}.$$

Its extrema are the smallest and largest eigenvalues of C. The square roots give the length factors. Comparing the entries of the two matrices separately cannot replace this calculation when the rulers have different principal directions.

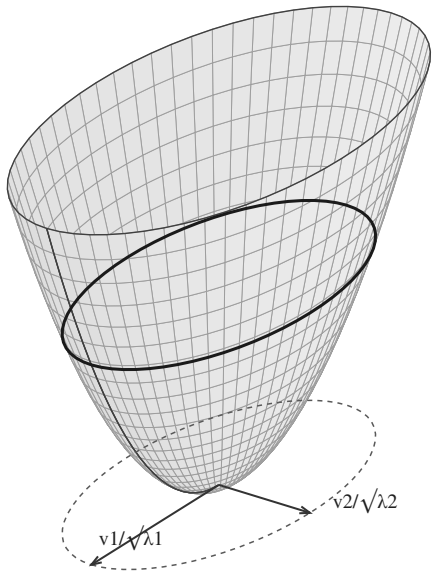
For a complete example, take $G = \operatorname{diag}(4, 1)$, Q a rotation through $30°$, and

$$\widetilde{G} = G^{1/2}Q\operatorname{diag}(0.81, 1.44)Q^TG^{1/2}.$$

Then every local length ratio lies between 0.9 and 1.2. The extremizing directions in the original coordinates are $G^{-1/2}Qe_1$ and $G^{-1/2}Qe_2$. They need not be orthogonal in the original Euclidean drawing, although they are orthogonal in the G ruler.

The ruler is visible in its level sets

Constructed cost surface: height = $h^T Gh$. The bold section has unit cost.

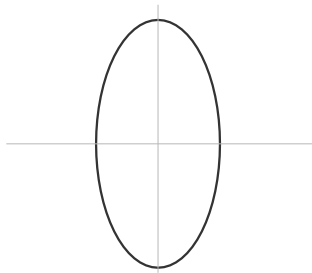


DECLARED METRIC G

$$\begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix}$$

Principal costs: 1, 4

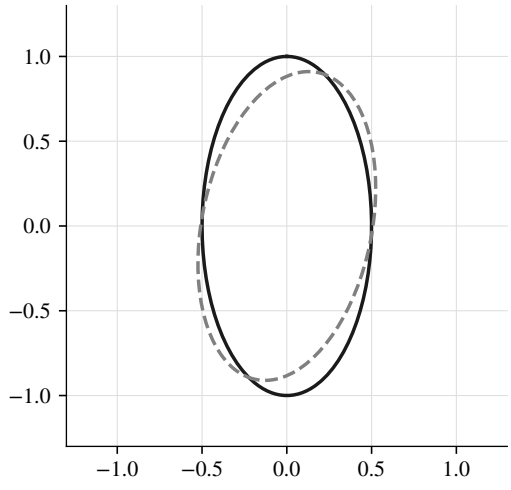
Unit semiaxes: $1/\sqrt{\lambda_1}$, $1/\sqrt{\lambda_2}$



Plan / $h^T Gh = 1$

A Two rulers assign different unit lengths

B Every direction has a bounded length ratio



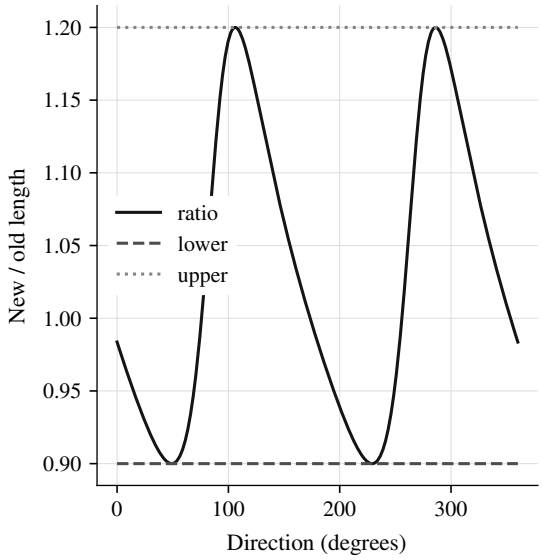
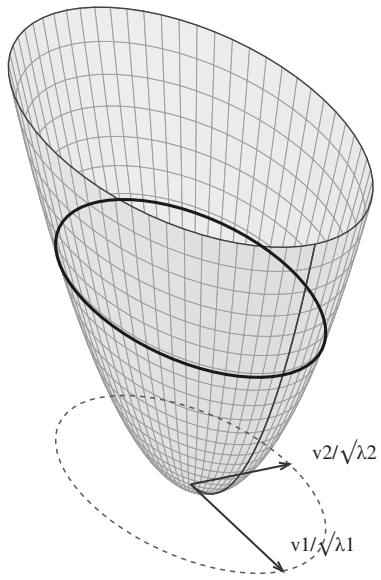


Figure 19.2: *The sharp comparison uses generalized directions.* The study continues on the following figure page.

The ruler is visible in its level sets

Constructed cost surface: height = $h^T G h$. The bold section has unit cost.

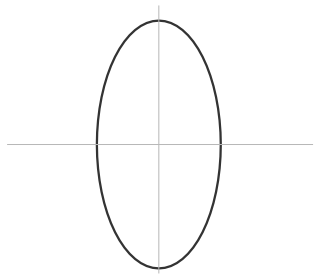


DECLARED METRIC G

$$\begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix}$$

Principal costs: 1, 4

Unit semiaxes: $1/\sqrt{\lambda_1}$, $1/\sqrt{\lambda_2}$



Plan / $h^T G h = 1$

C Whitening makes the comparison explicit

$$\begin{bmatrix} 0.967 & -0.273 \\ -0.273 & 1.28 \end{bmatrix}$$

D Generalized eigenvalues give the sharp limits

Generalized eigenvalues: 0.81, 1.44

Length factors: 0.9 and 1.2

Directions transform by G^(-1/2).

Bounds pass to distances only over the same allowed paths.

Figure 19.2: *The sharp comparison uses generalized directions (continued).* Two metric ellipses, the directional length ratio and the whitened comparison matrix describe the same pair of rulers. Generalized eigenvalues 0.81 and 1.44 give the sharp length factors 0.9 and 1.2. The registers state how directions transform and when tangent bounds extend to path distances.

Composition spends distortion multiplicatively

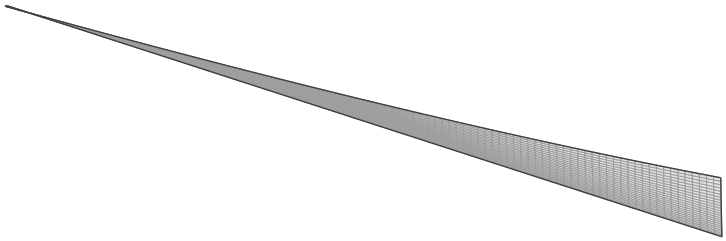
If one translator preserves distances within factors $[a_1, b_1]$ and another within $[a_2, b_2]$, their composition has factors $[a_1a_2, b_1b_2]$, provided the bounds apply on the intermediate domain. Ten steps with upper factor 1.01 have upper factor $1.01^{10} \simeq 1.104622$, not exactly 1.10.

Taking logarithms makes the multiplicative budget additive: $\log b_{\text{total}} = \sum \log b_i$. This can be a convenient way to maintain a long translation record. It does not remove the need to check every intermediate domain. A small distortion bound outside its validated region is not a bound at all.

Additive errors require a different recurrence. If $d_i \leq b_i d_{i-1} + c_i$, then two stages give $d_2 \leq b_2 b_1 d_0 + b_2 c_1 + c_2$. An early error can be amplified by later transformations. Printing only the sum $c_1 + c_2$ would miss that amplification.

The allowance accumulates by a stated rule

Each construction below separates an exact bound from its simplifying approximation.



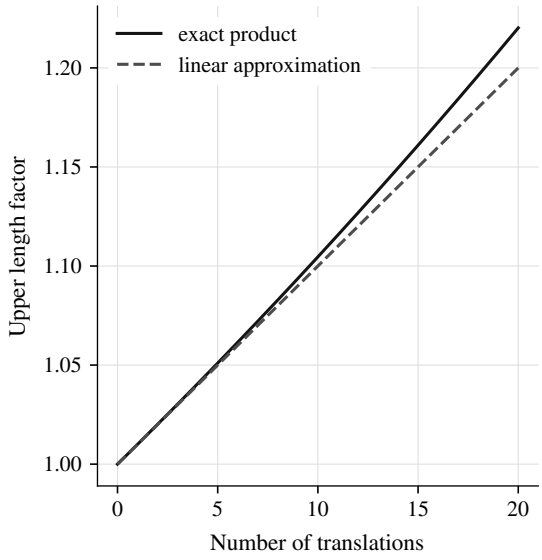
PARAMETER STUDY

$(1+\varepsilon)^n$

Horizontal: translation count n
Depth: expansion increment ε
Height: scale multiplier
A constructed parameter study.

Population, model, operating region and valid parameter range belong with every reported allowance.

A Small scale changes multiply along a chain



B Later expansions amplify earlier errors

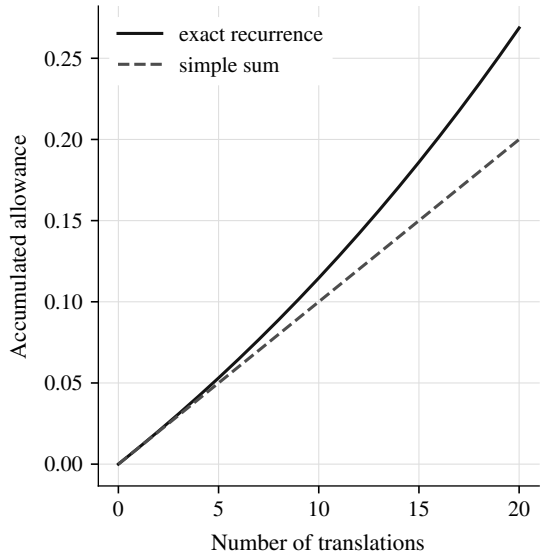
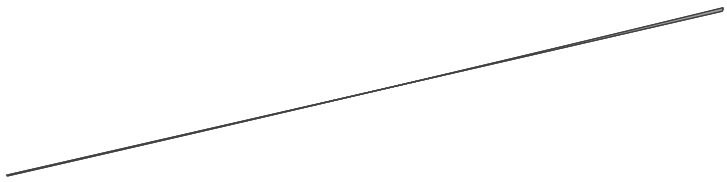


Figure 19.3: *Every translation spends part of the comparison budget.* The study continues on the following figure page.

The allowance accumulates by a stated rule

Each construction below separates an exact bound from its simplifying approximation.



PARAMETER STUDY

$0.01\ [(1+\varepsilon)^{n}-1]/\varepsilon$

Horizontal: translation count n
Depth: expansion increment ε
Height: additive-error bound
A constructed parameter study.

Population, model, operating region and valid parameter range belong with every reported allowance.

C The recurrence carries both kinds of error

$d1 \le b1\ d0 + c1$

$d2 \le b2\ b1\ d0 + b2\ c1 + c2$

A later expansion acts on earlier error.

D Each intermediate point must stay in scope

Every stage needs its valid region.

Log scale factors add.

Additive and multiplicative errors have different algebra.

An invalid intermediate input breaks the chain.

Figure 19.3: *Every translation spends part of the comparison budget (continued).* Repeated one-percent scale changes give a product rather than a linear increment. A second curve propagates a 0.01 additive allowance through successive factors of 1.03. The two-stage recurrence explains why later expansions amplify earlier errors; each intermediate input must remain in the declared domain.

The metric can change because the task changes

A recalibration may merely change coordinates, in which case the physical metric is transformed and invariant lengths should agree. A new task can instead change which errors matter, in which case the metric itself changes. The distinction is practical. A system optimized to preserve average similarity may not preserve the worst-direction margin needed for a safety-critical decision.

The same comparison of rulers will be needed for the operator certificate below. The constructive metric example in Chapter 17 shows this explicitly. A metric can certify a decreasing quantity for a nonnormal operator. Its condition number then controls the cost of translating that statement back to Euclidean coordinates. The later operator research investigates richer families, but the basic obligation remains visible in this small certificate: supply the metric, verify its definiteness, verify the operator inequality and report the comparison cost [31].

The eigenvalue comparison bounded the effect of replacing one ruler by another. We can also compare the rulers themselves. An instrument's changing uncertainty ellipse makes that second level of geometry concrete.

When the ruler itself is the object being compared

An instrument's uncertainty ellipse can rotate and change shape between two calibrations. Each ellipse is described by a positive-definite covariance matrix. We can compare the original measurements using those matrices, but we can also ask how different the two matrices themselves are. The second question requires a geometry on the space of rulers.

Let $\mathcal{P}_n$ be the set of real symmetric positive-definite $n \times n$ matrices. A tangent displacement U at a matrix G is symmetric. One useful metric is

$$\langle U, V \rangle_G = \operatorname{tr}(G^{-1} U G^{-1} V).$$

It is positive definite because

$$\langle U, U \rangle_G = \|G^{-1/2} U G^{-1/2}\|_F^2.$$

The matrix $G^{-1/2}$ converts the local ruler to an identity ruler before measuring its change. The construction compares relative deformation rather than raw entries.

Prove the coordinate invariance

An invertible change of measurement coordinates sends G to AGA^T and U to AUA^T. The inverse transforms as $A^{-T} G^{-1} A^{-1}$. Substituting into the metric and cyclically moving factors inside the trace gives

$$\langle AUA^T, AVA^T \rangle_{AGA^T} = \operatorname{tr}(G^{-1} U G^{-1} V).$$

This is why the metric is called affine invariant. The particular congruence convention depends on whether G represents covariance or a coordinate expression for a tangent metric; the identity covers any invertible congruence. The coordinates change, while the comparison of the two rulers does not.

For this metric, the distance and connecting geodesic are [34]

$$d_{\mathcal{P}}(G, H) = \|\log(G^{-1/2} H G^{-1/2})\|_F,$$
$$G(t) = G^{1/2} (G^{-1/2} H G^{-1/2})^t G^{1/2}.$$

Positive matrix powers and the logarithm are defined by the spectral decomposition. The geodesic stays positive definite throughout. Its midpoint is the matrix geometric mean.

Calculate a full two-direction example

Take

$$G = \begin{pmatrix} 1 & 0 \\ 0 & 9 \end{pmatrix}, \qquad H = \begin{pmatrix} 4 & 0 \\ 0 & 1 \end{pmatrix}.$$

The whitened comparison is $G^{-1/2} H G^{-1/2} = \operatorname{diag}(4, 1/9)$. Hence

$$d_{\mathcal{P}}(G, H) = \sqrt{(\log 4)^2 + (\log 9)^2} \simeq 2.598000.$$

The midpoint is

$$G(1/2) = \operatorname{diag}(2, 3).$$

The arithmetic midpoint is instead $\operatorname{diag}(2.5, 5)$. Both are positive definite, but they solve different averaging problems. The geometric midpoint splits each logarithmic scale ratio in half: $1 \to 2 \to 4$ in the first direction and $9 \to 3 \to 1$ in the second.

The path has entries 4^t and 9^{1-t}. Differentiation gives

$$\dot{G}(t) = G(t) \operatorname{diag}(\log 4, -\log 9).$$

Because these matrices commute, its squared speed in the displayed metric is $(\log 4)^2 + (\log 9)^2$ at every time. The path therefore has constant speed, and each half has half the total length.

Make the same example look coupled

Use the shear $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$. The transformed endpoints are

$$G' = \begin{pmatrix} 10 & 9 \\ 9 & 9 \end{pmatrix}, \qquad H' = \begin{pmatrix} 5 & 1 \\ 1 & 1 \end{pmatrix}.$$

The transformed midpoint is

$$A \operatorname{diag}(2, 3) A^T = \begin{pmatrix} 5 & 3 \\ 3 & 3 \end{pmatrix}.$$

The off-diagonal entries are now essential to the description. Setting them to zero would not be a simpler coordinate representation of the same pair. It would replace the matrices and change their geometry.

The generalized eigenvalues of (H', G') remain 4 and $1/9$, so the same distance is obtained. This is an alternative practical way to calculate it: find the positive generalized eigenvalues and take the Euclidean norm of their logarithms. A solver should exploit symmetry and positive definiteness rather than forming an unnecessary unstable inverse.

Why degeneracy is far away in this geometry

For the scalar family $g > 0$, the metric is dg^2/g^2 and the distance is $|\log(h/g)|$. As h approaches zero, that distance diverges. In a diagonal matrix family, sending one eigenvalue to zero has the same effect.

This geometry therefore treats a disappearing uncertainty or metric direction as a limiting degeneration, not a small ordinary edit of a matrix entry. That behavior can be desirable when an eigenvalue represents an essential measurement direction. It does not mean every application must use this metric. It means the choice has a concrete consequence that can be explained before the algorithm is run.

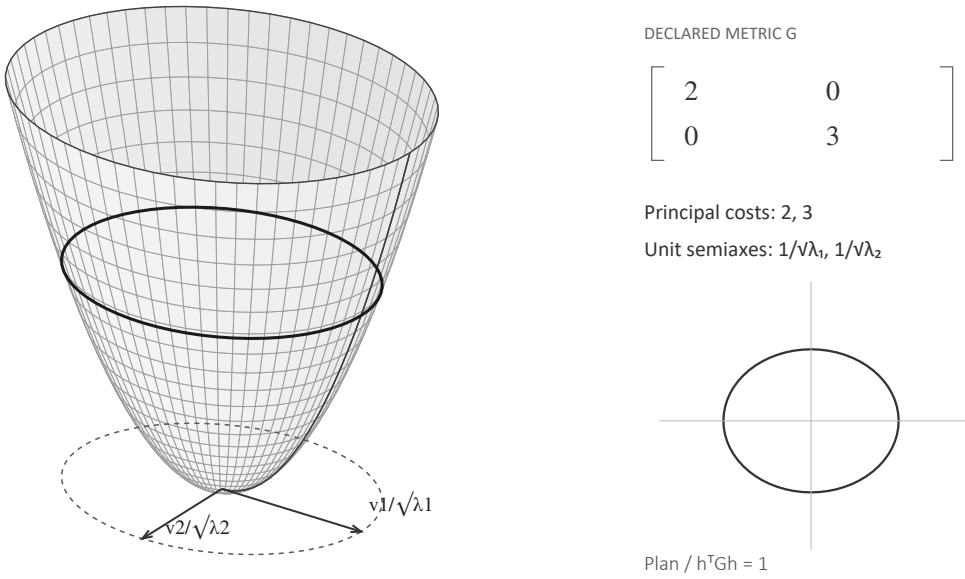
Do not confuse the two levels of distance

For two instrument states x and y, a fixed matrix G can define the displacement norm $\sqrt{(x-y)^T G(x-y)}$. For two matrices G and H, the distance $d_{\mathcal{P}}(G,H)$ compares the rulers themselves. Those are different spaces, with different units and dimensions.

The later operator certificate uses a positive-definite matrix to prove contraction and tracks its condition-number cost. The present construction shows another mathematically legitimate use of the same kind of object: measuring a change in the ruler. The common matrix notation does not make the two results interchangeable. Their connection is that both expose directional information that a single raw scalar can hide.

The ruler is visible in its level sets

Constructed cost surface: height = $h^T G h$. The bold section has unit cost.



A The unit-cost ellipse changes along the path

B Metric eigenvalues interpolate multiplicatively

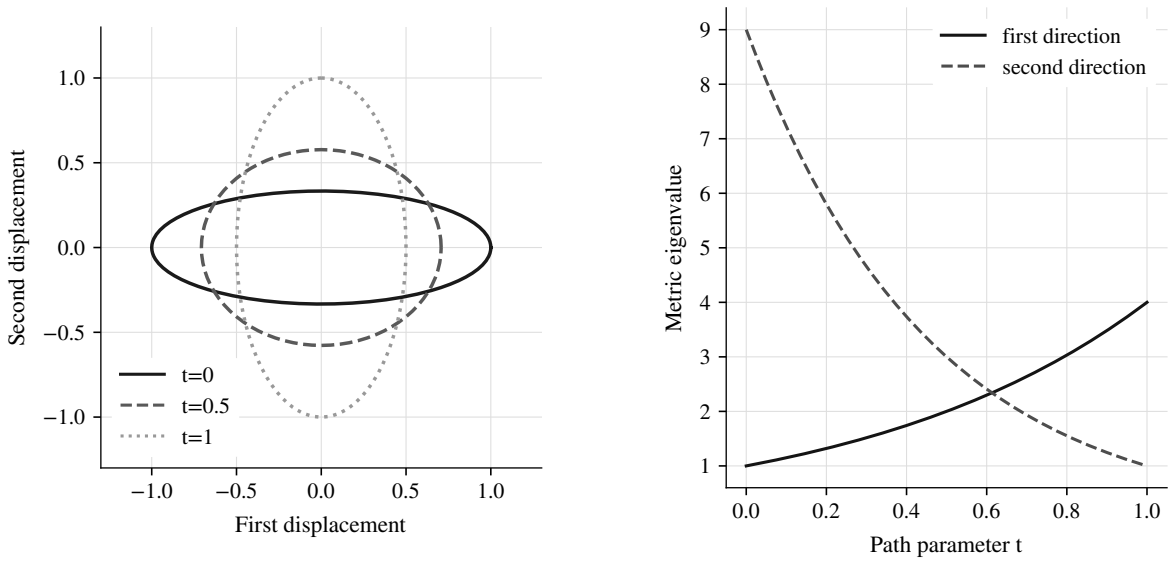
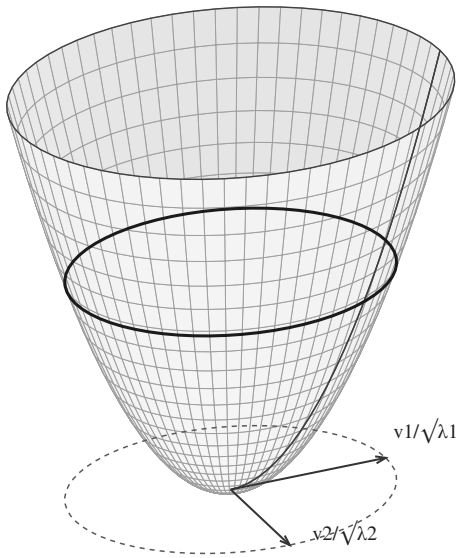


Figure 19.4: *A ruler can itself move through a geometric space.* The study continues on the following figure page.

The ruler is visible in its level sets

Constructed cost surface: height = $h^T G h$. The bold section has unit cost.

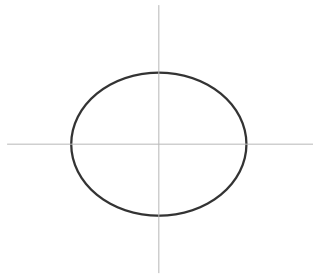


DECLARED METRIC G

$$\begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$$

Principal costs: 2, 3

Unit semiaxes: $1/\sqrt{\lambda_1}$, $1/\sqrt{\lambda_2}$



Plan / $h^T G h = 1$

C A coordinate shear preserves the midpoint

D State length and ruler distance are distinct

$$\begin{bmatrix} 5 & 3 \\ 3 & 3 \end{bmatrix}$$

State length: $\sqrt{(v^T G v)}$

Ruler distance: $\|\log(G^{-½} H G^{-½})\|_F$

Geometric midpoint: diag(2,3)

Arithmetic midpoint: diag(2.5,5)

The shear preserves ruler distance.

Figure 19.4: *A ruler can itself move through a geometric space (continued).* The diagonal metric path interpolates eigenvalues multiplicatively. A shear changes the midpoint matrix and introduces cross terms while preserving the affine-invariant distance between rulers.

A preserved midpoint. Verify the determinants and positive definiteness of G', H', and their displayed midpoint. Compute the generalized eigenvalues directly from $\det(H' - \lambda G')$. Then replace each matrix by its diagonal part and show that the comparison changes.

We have compared two rulers at rest. During an update, both the ruler and the represented vector may move. Differentiating their measured length gives the compatibility condition that joins those two changes.

A changing ruler must agree with its transport rule

An adaptive comparison system changes its metric as new information arrives. That can be sensible: some directions become more reliable or more important. But a bounded condition number alone does not guarantee that a stored vector keeps the same length while the ruler changes. The relation between state transport and metric evolution must be written down.

Let a vector evolve by $\dot{v} = A(t)v$ while a positive-definite matrix $G(t)$ supplies its squared length $s(t) = v^T G v$. Differentiate every factor:

$$\dot{s} = v^T(\dot{G} + A^T G + GA)v.$$

Length is preserved for every transported vector exactly when

$$\dot{G} + A^T G + GA = 0.$$

The condition couples the ruler and the transport. Holding one fixed while changing the other generally changes the comparison.

Construct a compatible metric rather than guessing it

Let $P(t)$ solve $\dot{P} = A(t)P$ with $P(0) = I$, and assume the finite-time fundamental matrix is invertible. Since $v(t) = P(t)v_0$, define

$$G(t) = P(t)^{-T} G_0 P(t)^{-1}.$$

Then

$$v(t)^T G(t) v(t) = v_0^T G_0 v_0$$

by direct cancellation. The matrix stays positive definite because it is an invertible congruence of G_0. Differentiating its expression also verifies the compatibility equation.

This construction does not say a physical instrument must adapt its metric in this way. It says what an exact length-preserving transport would require. If the desired metric update serves another purpose, the resulting change of length should be measured rather than silently called invariant.

A shear shows why cross terms appear

Take

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \qquad P(t) = \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}, \qquad G_0 = I.$$

The compatible metric is

$$G(t) = \begin{pmatrix} 1 & -t \\ -t & 1 + t^2 \end{pmatrix}.$$

A starting vector $v_0 = (0, 1)$ becomes $v(t) = (t, 1)$. Its ordinary Euclidean squared length grows to $1 + t^2$, while its squared length in $G(t)$ remains one. The cross term cancels the coordinate shear's apparent growth.

The determinant of $G(t)$ is one, but its eigenvalues separate as $|t|$ grows. Compatible transport can therefore preserve the relevant vector length while increasing the condition number of the coordinate ruler. Conversely, merely keeping the eigenvalues bounded does not establish compatibility.

A rotating ellipse keeps its eigenvalues and changes a measurement

Let $Q(t)$ be a planar rotation and

$$G(t) = Q(t)\,\mathrm{diag}(1, 9)Q(t)^T.$$

Every matrix has eigenvalues one and nine. If the vector is held fixed at $v = (1, 0)$, its squared length is

$$s(t) = \cos^2 t + 9\sin^2 t.$$

It varies between one and nine. The condition number stays nine throughout, yet the comparison of this physical direction changes substantially.

If instead the vector co-rotates as $v(t) = Q(t)v_0$, the corresponding quadratic length is constant. The same matrix path therefore supports different conclusions depending on what is being transported. A statement about the eigenvalues is not a statement about every direction's relation to the evolving eigenvectors.

Bound a controlled failure of compatibility

Define the symmetric residual

$$E(t) = \dot{G} + A^T G + GA.$$

Suppose it satisfies the two-sided quadratic-form bound

$$-2\alpha(t)G(t) \preceq E(t) \preceq 2\alpha(t)G(t), \qquad \alpha(t) \geq 0.$$

Then $-2\alpha s \leq \dot{s} \leq 2\alpha s$. Integrating gives

$$e^{-2\int_0^t \alpha}\, s(0) \leq s(t) \leq e^{2\int_0^t \alpha}\, s(0).$$

For lengths, take square roots: the exponents become $\pm\int_0^t \alpha$. This supplies an explicit distortion budget when exact invariance is not claimed.

To evaluate the residual in a coordinate-independent local scale, examine $G^{-1/2}EG^{-1/2}$. The stated bound is equivalent to its eigenvalues lying in $[-2\alpha, 2\alpha]$. Looking at the raw entries of E without the metric normalization would make the tolerance depend on coordinate units.

The distinction between changing coordinates and changing the task

The shear example changes the description while preserving the quadratic quantity. The rotating ellipse with a fixed vector changes which physical directions receive high cost. Both are valid mathematical operations. They should be distinguished when interpreting an adaptive detector or a frame-aware inference.

The later geometric research asks whether state, metric, and comparison can evolve coherently. This finite construction gives that question a testable form: specify $A(t)$ and $G(t)$, calculate $E(t)$, and either verify compatibility or report its bound.

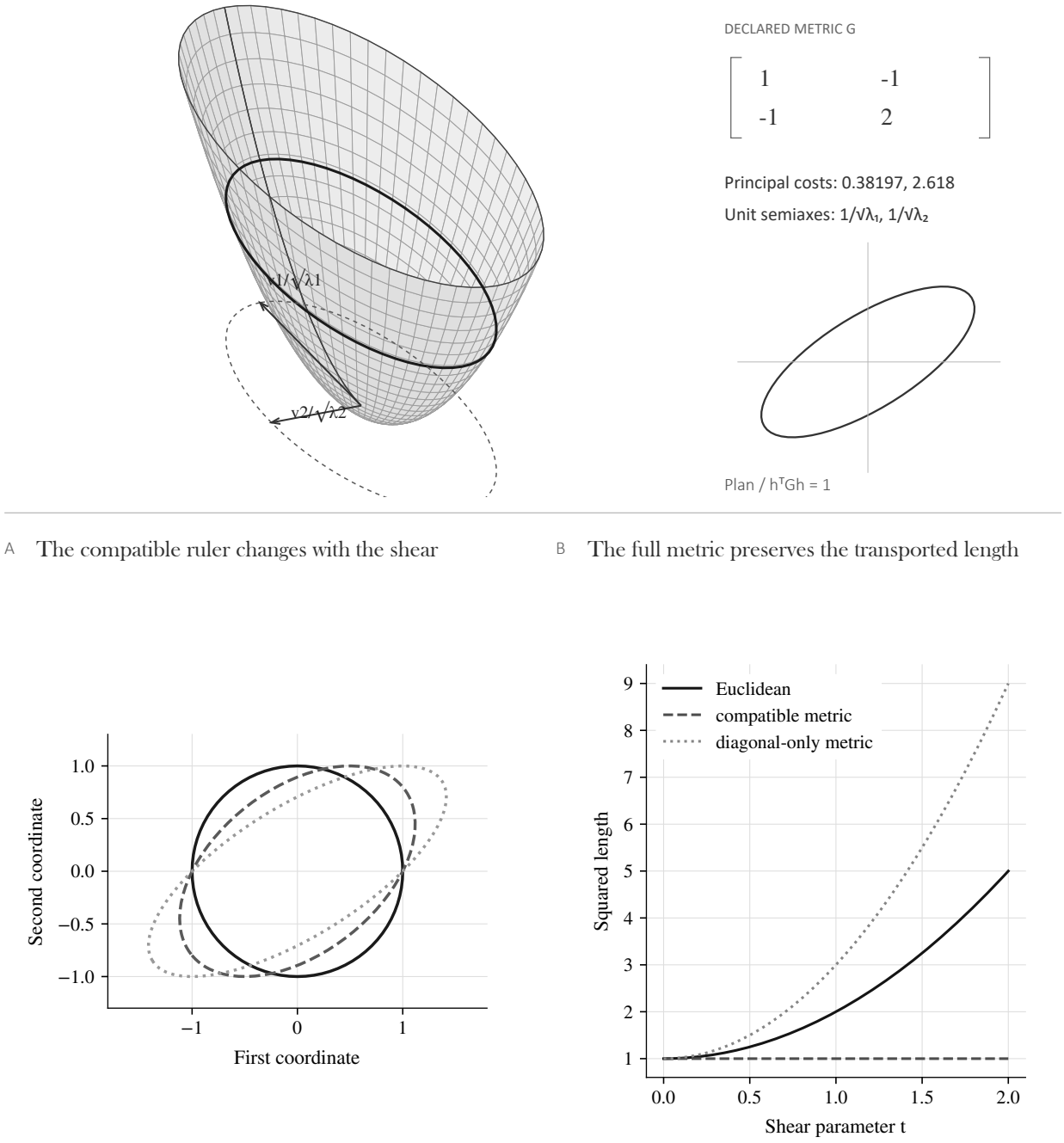
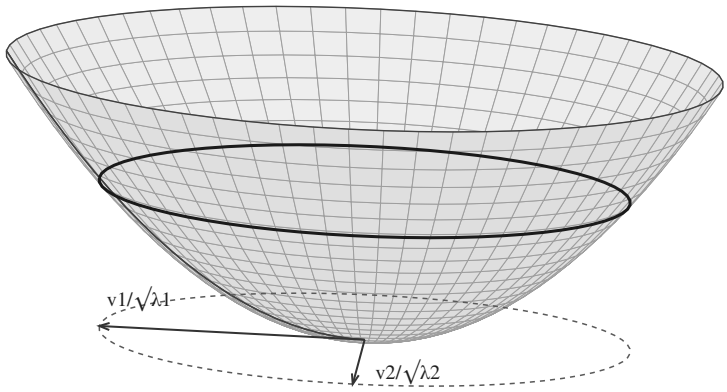


Figure 19.5: *The changing state and the changing ruler have a joint contract.* The study continues on the following figure page.

The ruler is visible in its level sets

Constructed cost surface: height = $h^T G h$. The bold section has unit cost.

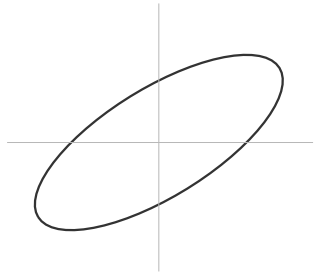


DECLARED METRIC G

$$\begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix}$$

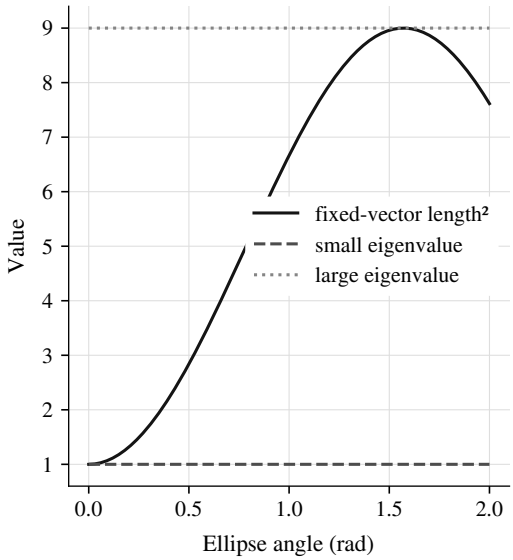
Principal costs: 0.38197, 2.618

Unit semiaxes: $1/\sqrt{\lambda_1}$, $1/\sqrt{\lambda_2}$



Plan / $h^T G h = 1$

C Fixed eigenvalues do not fix a vector length



D Compatibility links state evolution and ruler

$$dG/dt + A^T G + GA = 0$$

$$G(t) = P(t)^{-T}\, G(0)\, P(t)^{-1}$$

A bounded residual gives an exponential length allowance.

The state and ruler must be specified together.

Figure 19.5: *The changing state and the changing ruler have a joint contract (continued).* The shear preserves length in its compatible full metric while Euclidean and diagonal-only measurements grow. A rotating fixed-spectrum ruler changes the cost of a fixed vector, showing why eigenvalues alone do not establish transport compatibility.

An evolving ruler. For the shear example, compute $\dot{G}$, $A^T G$ and GA separately and verify their sum is zero. Then keep only the diagonal of $G(t)$ and calculate the squared length of $(t, 1)$. Explain precisely which invariance the diagonal replacement loses.

The price of returning to another norm

For a linear update, it may be easier to bound the output using a metric chosen for the calculation. If the original tolerance is in the Euclidean norm, we need a bound in that norm before we can use the result. Let G be a positive-definite matrix with $mI \preceq G \preceq MI$. Define $\|v\|_G = (v^{\mathsf{T}} G v)^{1/2}$. For a linear operator A,

$$\|A\|_2 \leq \sqrt{M/m}\, \|A\|_G.$$

For every vector v,

$$\begin{aligned} \|Av\|_2 &\leq \|Av\|_G / \sqrt{m} \\ &\leq \|A\|_G \|v\|_G / \sqrt{m} \\ &\leq \sqrt{M/m}\, \|A\|_G \|v\|_2. \end{aligned}$$

Taking the supremum proves the claim.

I use this norm-transport step in my operator experiments and Crouzeix investigations [27]. A matrix can be easier to control in a carefully chosen metric, but returning to the original norm spends the conditioning factor. A bound in one norm is not automatically the same numerical bound in another.

For $G = \mathrm{diag}(1, 100)$, that factor is 10. If a calculation establishes $\|A\|_G \leq 1$, the general consequence is $\|A\|_2 \leq 10$. Particular operators may satisfy a much stronger bound; the point is what follows from the stated hypotheses alone. This section uses the elementary inequality, not a claim of a universal Crouzeix proof.

In Chapter 17, an update shrank in one ruler even though it could grow in the ordinary Euclidean norm. To make that certificate useful, the new ruler must make the update shrink and stay close enough to the old ruler that we can translate the bound back. Can a ruler with no cross terms meet both requirements?

A certificate from the later operator research

The later constructive-metric research supplies a stronger result: an explicit eight-dimensional operator for which a coupled metric succeeds within a smaller conditioning budget than every metric restricted to the self-commutator's fixed directions [31]. This is a useful result for the geometry of sameness because it identifies information that a restricted ruler cannot retain.

The object is a linear operator on an eight-dimensional vector space. The two descriptions are its action in the ordinary Euclidean norm and its action in a specified positive-definite metric. The comparison cost is the metric's condition number. These are finite matrix statements; the word curvature in the source's self-commutator notation does not identify that matrix with the Riemann curvature tensor of a manifold.

Read the larger calculation as a test of a proposed simplification. Can the metric ignore couplings between certain directions and still establish the same bound within the allowed distortion? The cube labels identify the eight coordinates, the matrices supply a complete candidate, and the certificate settles that finite comparison. Its role here is to show exactly what a restricted ruler can lose.

Construct the operator

Set

$$B = \begin{pmatrix} 0 & 1 \\ 4/9 & 0 \end{pmatrix}, \qquad (w_1, w_2, w_3) = (1, 19/20, 9/10),$$

and define

$$\begin{aligned} A = {} & w_1 B \otimes I_2 \otimes I_2 \\ & + w_2 I_2 \otimes B \otimes I_2 \\ & + w_3 I_2 \otimes I_2 \otimes B. \end{aligned}$$

The tensor-state labels are the eight binary triples. Each term changes one factor, so the nonzero pattern follows the edges of a cube. The unequal weights distinguish the three factor directions.

The numerical range is $W(A) = \{x^* A x : x^* x = 1\}$. To calculate its support in direction θ, take the largest eigenvalue of $\mathrm{Re}(e^{-i\theta} A)$. Each Hermitian term acts on a separate tensor factor. The tensor product of their top eigenvectors attains the sum of their three individual maxima. With $b = 4/9$ and $s = w_1 + w_2 + w_3 = 57/20$,

$$h_A(\theta) = s\sqrt{\left(\frac{1+b}{2}\cos\theta\right)^2 + \left(\frac{1-b}{2}\sin\theta\right)^2}.$$

This is the support function of an ellipse with semiaxes $247/120$ and $19/24$. It describes the actual numerical range, rather than an enclosing circle substituted for convenience.

Specify the metric and the mapped operator

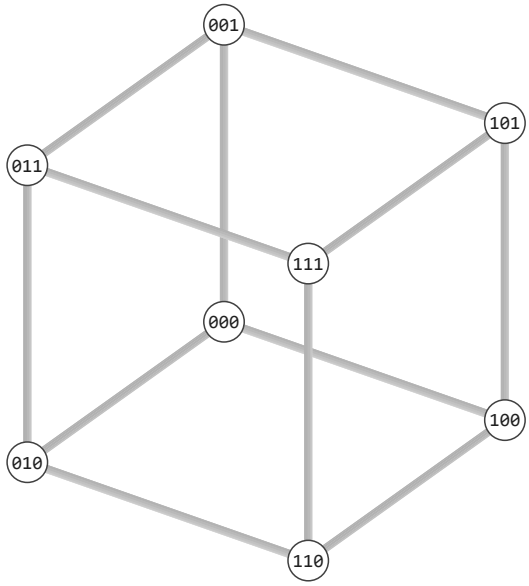
Let ϕ map the interior of that ellipse conformally to the unit disk, normalized by $\phi(0) = 0$ and $\phi'(0) > 0$. Put $T = \phi(A)$, defined by the analytic functional calculus. The retained certificate gives a rational metric G satisfying

$$I \prec G \prec \frac{13}{4} I, \qquad G - T^* G T \succ 0.$$

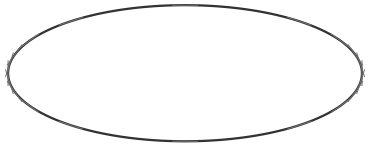
The first two inequalities bound the ruler's comparison cost. The last states that T is a strict contraction in that ruler. The entries of G and the separating matrix used below are reproduced in Appendix D; the certificate is not merely an optimizer's success flag.

Eight tensor states share one coupled ruler

Three binary factors form the state graph; the numerical range has a computed boundary.



NUMERICAL RANGE / SUPPORT LINES



Semiaxes: 247/120 and 19/24

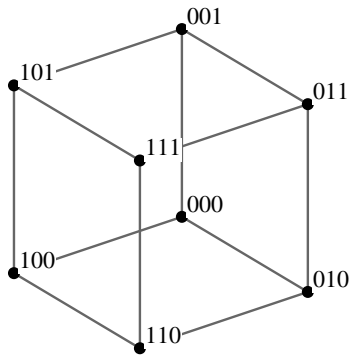
The metric has cross-direction terms.

The certificate tests the full operator.

A diagonal restriction changes the test.

A Three factors produce eight tensor states

B The numerical range is an exact ellipse



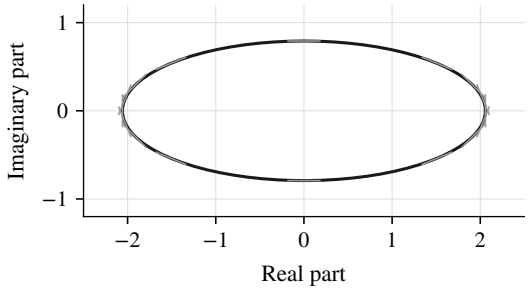
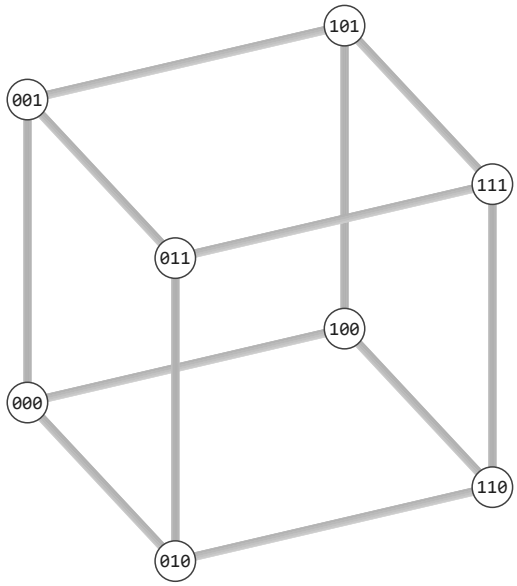


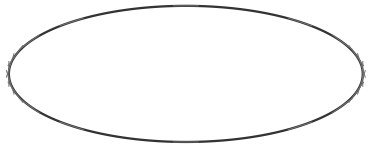
Figure 19.6: *A coupled metric succeeds where every permitted diagonal metric fails.* The study continues on the following figure page.

Eight tensor states share one coupled ruler

Three binary factors form the state graph; the numerical range has a computed boundary.



NUMERICAL RANGE / SUPPORT LINES



Semiaxes: 247/120 and 19/24

The metric has cross-direction terms.

The certificate tests the full operator.

A diagonal restriction changes the test.

C Cross terms couple the metric directions

1.02	0	0	-0.07	0	-0.10	-0.13	0
0	1.57	-0.19	0	-0.17	0	0	0.15
0	-0.19	1.67	0	-0.16	0	0	0.13
-0.07	0	0	2.29	0	-0.02	0.03	0
0	-0.17	-0.16	0	1.75	0	0	0.10
-0.10	0	0	-0.02	0	2.43	0.05	0
-0.13	0	0	0.03	0	0.05	2.56	0
0	0.15	0.13	0	0.10	0	0	3.16

D The certificate separates full and diagonal rulers

$I < G < (13/4)\, I$

$G - T^{*}\, G\, T > 0$

Norm cost $\leq \sqrt{13/2} \approx 1.802776$

Every commuting metric with $I \leq H \leq 4I$ fails.

Exact entries and interval bounds are given in the appendix.

Figure 19.6: *A coupled metric succeeds where every permitted diagonal metric fails (continued).* The eight-state tensor operator, its exact elliptical numerical range and the retained metric are shown with the rational certificate margins. The replay checks the saved matrices and analytic error allowance with exact arithmetic. The result concerns this explicit operator and the stated metric budgets.

Set $S = G^{1/2}$. Then $\|STS^{-1}\|_2 < 1$ and

$$\|S\|_2\|S^{-1}\|_2 < \frac{\sqrt{13}}{2}.$$

Applying the disk functional-calculus inequality for a contraction to $p \circ \phi^{-1}$, then returning to the original norm, gives

$$\|p(A)\|_2 \le \frac{\sqrt{13}}{2} \max_{z \in W(A)} |p(z)|$$

for every polynomial p. The constant is approximately 1.802776, below two. The argument uses both the actual conformal domain and the conditioning allowance. The elementary norm-transport inequality earlier in this chapter is the final step of that chain.

Why the cross terms are essential here

Define the self-commutator $K_A = AA^* - A^*A$. For this tensor construction it is diagonal with eight distinct entries. Any Hermitian matrix commuting with it must therefore be diagonal in this basis. The certificate also proves that no such matrix H can satisfy all three conditions

$$I \preceq H \preceq 4I, \qquad T^*HT \preceq H, \qquad HK_A = K_AH.$$

This excludes an entire restricted class of rulers. It is stronger than showing that the diagonal part of one proposed G fails. It says every allowed diagonal choice within the larger budget four fails, while the displayed coupled choice works within budget 13/4.

The proof uses a positive-definite separating matrix Z. The diagonal entries of $TZT^* - Z$ give a positive lower bound for $\operatorname{tr}(H(TZT^* - Z))$ over all diagonal H with entries in $[1, 4]$. A contractive H would instead require this trace to be nonpositive, because it equals $-\operatorname{tr}(Z(H - T^*HT))$. The contradiction is finite and inspectable.

What the computation established

The saved replay reconstructs the tensor operator, rational spectral projectors and truncated conformal series. It verifies the analytic-plus-rounding allowance, independently recomputes the positive rational decomposition pivots and checks the separation inequality. The replay passed 198 exact checks. No floating-point eigenvalue was used as the acceptance criterion.

Say you want to simplify this operator check by dropping every cross term from the ruler. You could retune all eight diagonal weights and raise the conditioning budget to four. The certificate still rules that out for this operator. It is a result about which ruler can certify contraction here. The broader Crouzeix conjecture and the book's topological covers are separate questions.

A new ruler changes old length allowances. The compatibility equation tells us when transport preserves length, and the operator certificate shows why cross-direction terms matter to that bound. Next we ask when a remaining discrepancy can change a decision.

Practice with the construction

Back to the maps and the globe. A change of ruler has a measurable cost.

19.1. If $0.81g \preceq \tilde{g} \preceq 1.44g$, bound the new length of a path whose old length is 5.

19.2. For $G = \operatorname{diag}(2, 18)$, compute the norm-transport factor and explain what information could make that general bound sharper for a specific operator.

Chapter 20

Margins and uncertainty at the boundary

We can bound how a comparison changes when its representation or ruler changes. The remaining question is whether that change alters a decision. A measurement near a threshold is exposed to an error that a measurement far from it can tolerate. We carry the geometric bound into that boundary region and then into a probability statement.

A small error at a crowded boundary

A factory checks how far a mounting hole lies from its specified location. It accepts the part if the measured displacement is at most 5 millimetres. Changing the format of the inspection record alters the computed displacement by at most 0.1 millimetres. At 2 or 8 millimetres the decision stays fixed. Near 5 it can change. The number of parts near that limit determines how often the error matters.

Let D be the reference distance, D' the distance after translation, and τ a fixed acceptance threshold. On an event E, assume $|D' - D| \le \eta$. The two binary rules are $h = \mathbf{1}\{D \le \tau\}$ and $h' = \mathbf{1}\{D' \le \tau\}$.

Theorem 20.1 (Decision transfer with a boundary term)**.** *Under the stated bound,*

$$\Pr(h' \neq h) \le \Pr(E^c) + \Pr(|D - \tau| \le \eta).$$

For any binary target Y on the same probability space,

$$|\Pr(h' \neq Y) - \Pr(h \neq Y)| \le \Pr(h' \neq h).$$

The corresponding conditional statements hold within any class or subgroup of positive probability, with every probability conditioned on that same event.

Proof. Work first on E. If $D < \tau - \eta$, both rules accept. If $D > \tau + \eta$, both reject. Disagreement is therefore contained in $E^c \cup \{|D - \tau| \le \eta\}$. Taking probabilities and the union bound proves the first inequality. Pointwise, the absolute difference between the two zero-one error indicators is at most $\mathbf{1}\{h' \neq h\}$. Taking expectations proves the second. Apply the same argument under a conditional probability measure for the last statement. □

If D has density at most L throughout the boundary band, its mass is at most $2L\eta$. More generally a margin condition $\Pr(|D - \tau| \le r) \le Cr^\alpha$ gives a term $C\eta^\alpha$. These are assumptions about the population being evaluated. They connect to the statistical study of classification margins [7].

Suppose every part is right on the acceptance line. Set $D = \tau$ for every part and let the new calculation give $D' = \tau + \eta/2$ for any positive allowance η. The shift fits inside the distance allowance, yet the old rule accepts every part and the new one rejects every part. That's why we count the parts close to the limit.

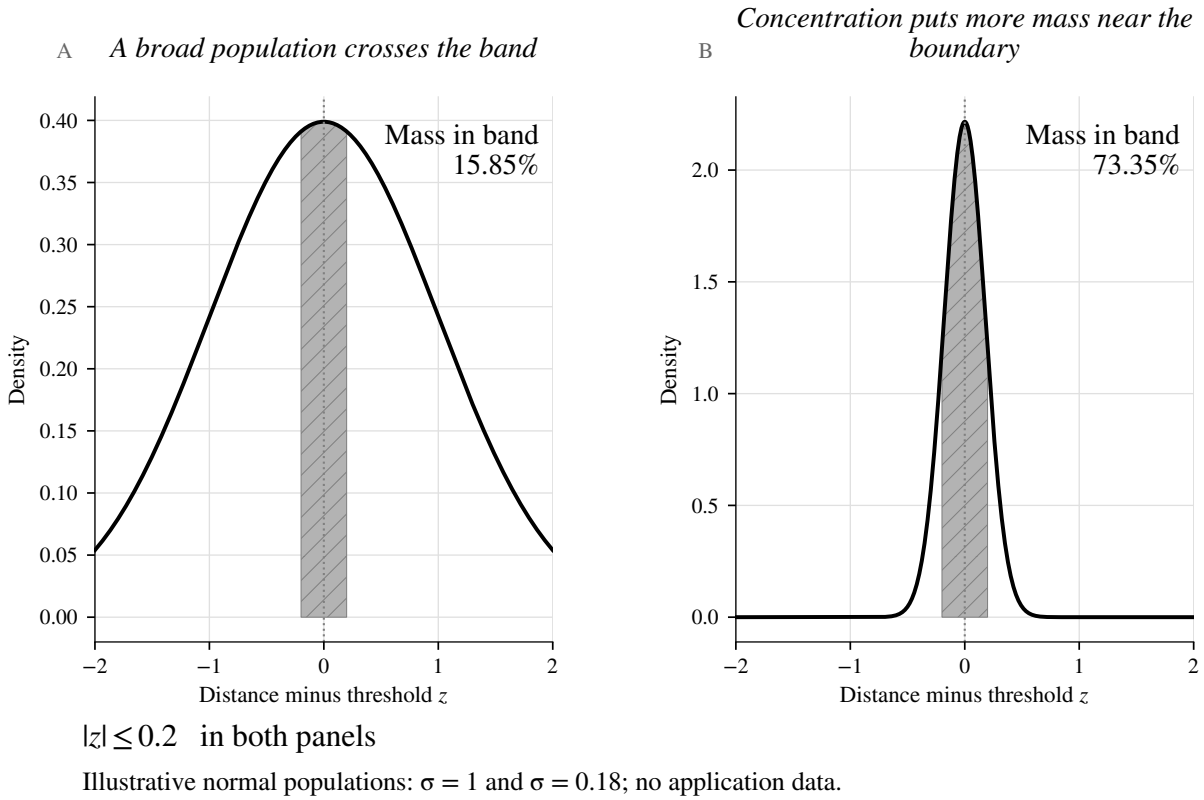


Figure 20.1: Two illustrative normal populations with the same decision-boundary band. Shading integrates the probability mass within $|z| \leq 0.2$. A narrow distance-error budget can threaten many decisions when the population is concentrated near the threshold.

From a distance allowance to a decision allowance

Chapters 11–13 supplied maps between the translator and geometric descriptions, with their realization and reconstruction conditions. Use those maps on the same observation pairs, for the same ground-truth target and a fixed binary threshold. Establish a distance-error bound η on an event E. If the reference rule has error at most B, the translated rule has error at most

$$B + \Pr(E^c) + \Pr(|D - \tau| \leq \eta).$$

The first added term covers failure of the comparison. The second counts observations close enough to the threshold for their label to change. For a false-positive bound, every probability is conditioned on the negative class; for false negatives, on the positive class.

The calculation gives the two descriptions complementary work. A geometric model can supply a distance inequality under its stated hypotheses. Matched records can test the implemented conversion and estimate the relevant population quantities. Volume comparisons require dimension, curvature and completeness conditions; converting a volume into probability also requires a population measure or density bound. A geometric proof does not supply those empirical ingredients automatically.

An engineer can follow the calculation in four steps: specify the reference rule and its population; establish the distance comparison; bound the probability of the threatened band; and combine those terms with their uncertainty statements. If the reference error is estimated from held-out data, use its confidence bound rather than treating the observed fraction as a known population

probability. Comparison and band-mass estimates need their own confidence statements; a union bound can combine their confidence failures without assuming independence.

The roles can be reversed, using the translator rule as reference and comparing the geometric rule on the same records. In either direction, the limiting term identifies what needs attention: the representation, the concentration of observations near the boundary, or the validation evidence. The proof already identified the threatened records. The following definitions specify how to report their labels, class-conditional errors and abstentions on the population used by the comparison.

Detection Systems and Error Budgets

Use the factory rule as a reference. To report an error rate, also specify which parts arrive, their true labels and the observations the detector receives. The same inventory applies to a sameness detector.

Detection tasks on a semantic sameness structure

Fix a semantic sameness structure S, with latent entities u and a notion of benign evolution as in Chapter 1. A detection task for S has the following ingredients:

- A probabilistic model for drawing *pairs of entities* (u, v) from the regime of interest, together with a *ground-truth label*
$$Y \in \{\text{same}, \text{different}\},$$
where $Y = \text{same}$ if u and v represent the same semantic entity (up to benign evolution), and $Y = \text{different}$ otherwise.
- A probabilistic model for generating *observations* of u and v through the channels of S, including noise, missing data, and other real-world effects.
- A detector that observes some representation of the data (e.g., embeddings and distances) and outputs a decision
$$\widehat{Y} \in \{\text{same}, \text{different}, \text{abstain}\}.$$

We measure performance in terms of errors with respect to the ground truth Y:

- *False positive:* $\widehat{Y} = \text{same}$ when $Y = \text{different}$.
- *False negative:* $\widehat{Y} = \text{different}$ when $Y = \text{same}$.
- *Abstention error:* either (a) failing to abstain when we are supposed to, or (b) abstaining too often, depending on the application.

For the probability budget below, define an error event that specifies which wrong labels and abstention-policy violations count. Let $\mathbb{P}$ be the probability measure over entities, observations and decisions. If the task assigns different costs to the outcomes, define those costs separately and bound the expected loss for that convention.

Distance-based detectors and thresholds

In both translator-based and geometry-based realizations, detection systems frequently operate by comparing distances in some representation space:

- On a Riemannian manifold (M, g), we compute the geodesic distance $d_M(x, y)$ between the identity representations of two entities and compare it to thresholds.

- On a translator metric space (X_T, d_T), we compute the translator distance $d_T(x, y)$ between equivalence classes and compare it to thresholds.

We will focus on *threshold-based detectors* of the form

$$\widehat{Y} = \begin{cases} \text{same} & \text{if } d \le \tau_{\text{same}}, \\ \text{different} & \text{if } d \ge \tau_{\text{diff}}, \\ \text{abstain} & \text{if } \tau_{\text{same}} < d < \tau_{\text{diff}}, \end{cases}$$

where d is either d_M or d_T depending on the representation, and $\tau_{\text{same}} < \tau_{\text{diff}}$. For a two-way rule with a single threshold, accept at equality and reject strictly above it.

Threshold-based detectors are not the only possibility, but they are common in high-stakes applications (for example, biometric verification) and make the geometry-transfer argument transparent.

Decomposed error budgets

An audit can organize failures under four questions. Use the following quantities as probability allowances for explicitly defined events:

- E_{geom} — *geometry error*: how much of the detection error arises because our representation space and metric do not perfectly realize the semantic notion of sameness (for example, curvature assumptions fail, manifolds are misspecified, translator geometry drifts).
- E_{link} — *link error*: how much arises from the observation process and encoders (sensor noise, compression artifacts, embedding noise, missed modalities).
- ξ — *calibration error*: how much arises from miscalibrated thresholds and confidence estimates (for example, the detector claims 95% confidence when the true error rate is higher).
- ζ — *abstention error*: how much arises from failures in an abstention policy (failing to abstain when required, or abstaining too often).

Where this decomposition comes from. To reconstruct a detection error, follow the observation through four checks:

- *Was the representation wrong?* (E_{geom}) Did the model use a ruler, manifold or allowable-path family that does not describe the task? Ignoring units during a telemetry conversion is one inspectable example.
- *Was the observation noisy?* (E_{link}) Did sensor noise, compression, or encoding failures introduce errors independent of the geometry? A blurry video frame, a corrupted log record, or a dropped modality contributes here.
- *Were the thresholds miscalibrated?* (ξ) Did we claim 95% confidence when the true error rate was higher? Threshold miscalibration shows up in deployment when observed error rates do not match what the model promised.
- *Did abstention fail?* (ζ) Did we fail to abstain when we should have (data too ambiguous) or abstain too often (the system is overly conservative)?

These four questions organize an audit. To attach a bound to each, define its failure event or loss and provide evidence for that bound. The transfer theorem controls disagreement between the specified decisions; it does not divide that disagreement into causal shares. Separate component transfer requires separate event or loss comparisons.

We speak of an error budget for a detector as a tuple of numbers

$$(E_{\text{geom}}, E_{\text{link}}, \xi, \zeta)$$

together with a statement of the form:

> Write E_{total} for the probability of the declared detection-error event under the distributional assumptions about entities and observations in S. Then
>
> $$E_{\text{total}} \leq E_{\text{geom}} + E_{\text{link}} + \xi + \zeta,$$
>
> provided the four named events cover the detection-error event and the four numbers bound their probabilities on the same population. The events may overlap; the inequality is then the union bound, not an additive causal attribution.

Telemetry example. A geometry error can use an unsuitable ruler; a linkage error can attach the wrong units to a field; a calibration error can misstate the frequency of threshold mistakes; an abstention failure can issue a confident answer outside the validated regime. Define these as events before adding bounds. Their overlap is permitted by a union bound.

Translate thresholds without inventing certainty

Suppose the factory now compares two distance calculations, D_T from translated records and D_M from its geometric model. The two rulers can differ in scale as well as by an additive allowance:

$$aD_T - \eta \leq D_M \leq bD_T + \eta, \qquad a, b > 0.$$

To guarantee $D_M \leq \tau$, it is sufficient that $D_T \leq (\tau - \eta)/b$. To guarantee $D_M > \tau$, it is sufficient that $D_T > (\tau + \eta)/a$. The interval between these sufficient tests is not settled by this bound. If the acceptance threshold becomes negative, nonnegative distances may provide no guaranteed acceptances at all. That is an informative limit of the guarantee.

For two thresholds $\tau_s < \tau_d$, assign same below τ_s, different above τ_d and abstain in between, with endpoint conventions fixed in the implementation. Under a uniform additive distance bound, changes in this three-way decision are confined to bands around the two thresholds or the failure event. The union bound gives the sum of the two band probabilities and the failure probability. Coverage, Pr(a definite answer), should be reported alongside error among the answered cases.

A correspondence has to carry a particular claim

In 1799, French soldiers in Egypt discovered the Rosetta Stone, bearing versions of one decree in Egyptian hieroglyphs, Demotic script and ancient Greek [2]. The known Greek text helped scholars interpret the Egyptian versions. The correspondence was useful because its shared content could be investigated; it did not make word counts or every visible feature interchangeable. That distinction is our concern when a geometric comparison reaches a decision.

Consider two altimeters on the same mountain, one using satellite positioning and one converting air pressure into altitude. Both readings must first use the same height reference. Suppose a calibration valid over the operating range establishes lower and upper factors $L_{\min} = 0.98$, $L_{\max} = 1.02$ and additive allowance $\varepsilon = 5$ metres. For a reference reading of $4,200$ metres, the corresponding interval is

$$0.98 \times 4200 - 5 \text{ to } 1.02 \times 4200 + 5 \quad \Longrightarrow \quad 4111\text{m to } 4289\text{m}.$$

This interval compares the readings under the assumed calibration. A separate uncertainty of ± 10 metres in the reference reading must also be propagated if the question concerns true altitude. A single agreement at base camp does not establish a bound across the mountain, especially as pressure conditions change.

Now put an altitude threshold on the display. A climber far below that threshold can tolerate this reading discrepancy without changing the decision. A climber close to it may receive different warnings. The calibration interval alone does not tell us how many climbers occupy that threatened band. That is a question about the population being evaluated.

The same distinction appears when a navigation program changes from latitude and longitude to UTM coordinates. Suppose its original specification reports position error within 5 metres with high probability, route-length error within 2%, and arrival-time error within 30 seconds for typical trips. An explicit conversion bound can carry a location comparison through the coordinate change. A route claim also needs corresponding paths, while a time claim needs the travel model. Each conclusion must retain the assumptions that made it true.

Applying the transfer theorem

The theorem just proved states the exact distance, event and population assumptions. Apply it on the same observation pairs and use the actual threshold rule. For class-conditional errors, condition both the geometry-failure probability and the boundary mass on the class. For quality-gated decisions, use the population remaining after the gate.

The reference distance and the translated distance must be in compatible units. The constants in a bi-Lipschitz-with-slack relation yield sufficient acceptance and rejection thresholds as derived above. The interval left between them is a legitimate abstention region.

The four-part vocabulary of geometry, linkage, calibration and abstention remains useful for diagnosis. A guarantee for each part requires its own event or loss definition; the theorem bounds the disagreement of the specified decision rules. It does not assign causal fractions of that disagreement without further evidence.

A worked transfer calculation

Consider a separate inspection of a larger assembly, with a threshold of 0.05 metre for displacement from its specified position. Suppose the old and new distance calculations differ by at most 0.01 metre on an event of probability at least 0.998. Assume the reference distances assign at most 0.03 probability to the band $[0.04, 0.06]$. The disagreement bound is $0.002 + 0.03 = 0.032$.

If the reference rule has overall error at most 0.015 on the same population, the translated rule has overall error at most 0.047. This is a bound from the specified assumptions, not an experiment. A false-positive bound needs the corresponding assumptions conditional on the negative class; the overall calculation cannot be relabeled as a false-positive rate.

The pointwise proof also identifies which records need attention. A pair with reference distance 0.02 stays below the threshold; a pair at 0.048 is unresolved by the distance bound alone. Improving the translator shrinks that band. Changing the population can change its probability mass even when the translator stays fixed.

Keep the denominators visible

A false-positive rate is conditional on the negative class. A posterior correctness is conditional on the alert event. Overall error is averaged over the population. These are different probabilities and can change differently when the population changes.

For example, suppose an alert has true-positive rate 0.9 and false-positive rate 0.01. If positives occur with probability 0.01, its precision is

$$\frac{0.9(0.01)}{0.9(0.01) + 0.01(0.99)} \approx 0.4762.$$

The same rates at prevalence 0.1 give precision 0.9091. The geometry and threshold can stay fixed while this posterior changes because the base rate changed. These are illustrative arithmetic calculations, not results from a deployed application.

For named failure mechanisms, define events before assigning budgets. Geometry failure, linkage failure and calibration failure may overlap. A union bound is valid without independence; multiplying their success probabilities requires an independence statement or a separately proved dependence correction. An observed total disagreement cannot be divided arbitrarily into causal categories and still retain a mathematical interpretation.

The factory's uncertain cases are parts whose measured displacement lies near the acceptance limit. We can draw that uncertainty for one part, then ask how often parts land in the band and how often the system declines to decide.

Draw the uncertainty before choosing the label

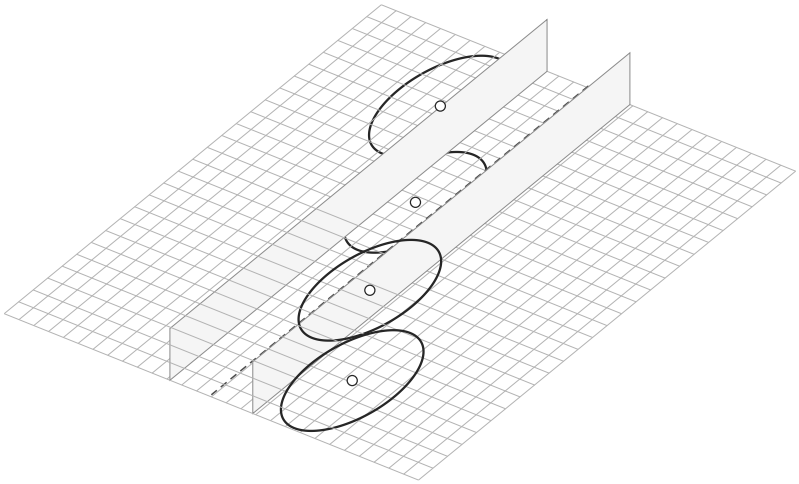
Let s be a part's reference displacement divided by a fixed scale and τ its acceptance threshold on that scale. The new software reports $\widehat{s}$ with deterministic error bound $|s - \widehat{s}| \leq \delta$. The reference score lies in $[\widehat{s} - \delta, \widehat{s} + \delta]$.

Accept when the upper endpoint is at most τ. Reject when the lower endpoint is strictly above τ. Otherwise the interval allows both outcomes, and the system abstains.

Set $\tau = 1$ and $\delta = 0.04$. If the inspection software reports a displacement score of 0.92, I'd sign off. At 1.08, I'd reject. At 0.98 the allowed score reaches 1.02, and at 1.02 it reaches down to 0.98. Both parts go back for a check. The screen puts them on opposite sides, but either label could change.

An error ellipse becomes a score interval

G = diag(4,1); score s = 2x; threshold τ = 1; deterministic allowance δ = 0.04.



Score centers:

0.92, 0.98, 1.02, 1.08

G-radius: 0.04

Semiaxes: 0.02 and 0.04

Each score range: s ± 0.04

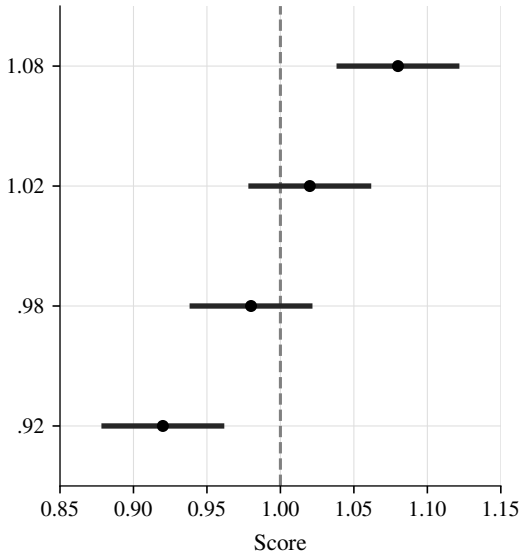
0.92 and 1.08: stable.

0.98 and 1.02: unresolved.

Strip edges: x = 0.48, 0.52.

Sheet height has no statistical meaning.

A An interval may cross the decision threshold



B The full uncertainty region matters

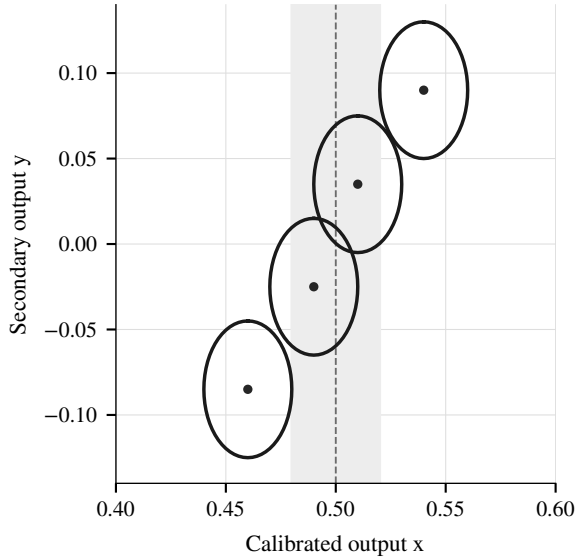
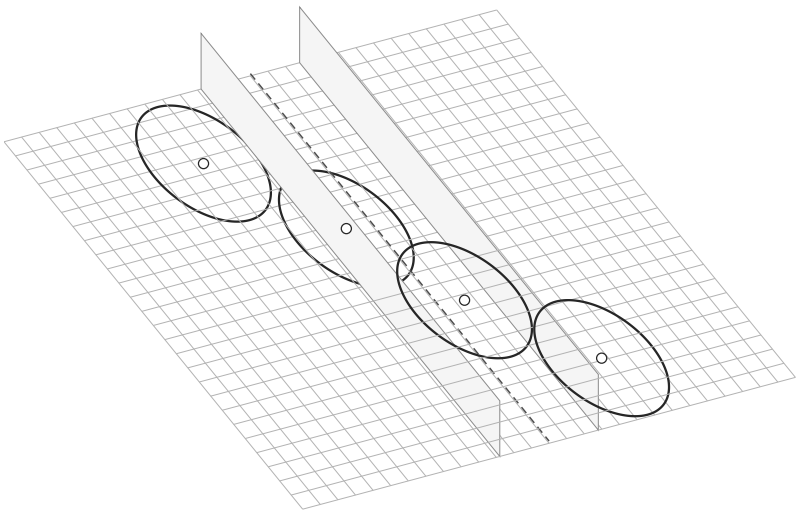


Figure 20.2: *The decision belongs to the whole error interval.* The study continues on the following figure page.

An error ellipse becomes a score interval

G = diag(4,1); score s = 2x; threshold τ = 1; deterministic allowance δ = 0.04.



Score centers:

0.92, 0.98, 1.02, 1.08

G-radius: 0.04

Semiaxes: 0.02 and 0.04

Each score range: s ± 0.04

0.92 and 1.08: stable.

0.98 and 1.02: unresolved.

Strip edges: x = 0.48, 0.52.

Sheet height has no statistical meaning.

C Four scores produce three decision states

.92 → wholly below	
.98 → crosses threshold	
1.02 → crosses threshold	
1.08 → wholly above	

D A deterministic interval is not confidence

The interval is deterministic.
A central value alone is insufficient near the boundary.
An abstention preserves uncertainty.
A confidence interval needs a separate sampling model.

Figure 20.2: *The decision belongs to the whole error interval (continued).* Four score intervals are compared with one threshold. Their geometric uncertainty regions and scalar images show why some cases are stable and others remain undecided. The shaded region is a deterministic ambiguity band, not a probability statement.

How much of the population lives near the boundary?

Suppose the approximation error bound fails with probability at most η. On the event where it holds, disagreement with the reference threshold can occur only when $|s - \tau| \leq \delta$. Hence

$$\Pr(\text{disagreement}) \leq \eta + \Pr(|s - \tau| \leq \delta).$$

If the score density is bounded by B throughout this band, its probability is at most $2B\delta$. For $\eta = 0.003$, $B = 0.4$ and $\delta = 0.02$, the bound is 0.019.

The density hypothesis is local and population-specific. A distribution can concentrate a point mass at the threshold, in which case an arbitrarily narrow band retains substantial probability. A histogram drawn from finite data can suggest where to investigate but does not alone certify a uniform density bound.

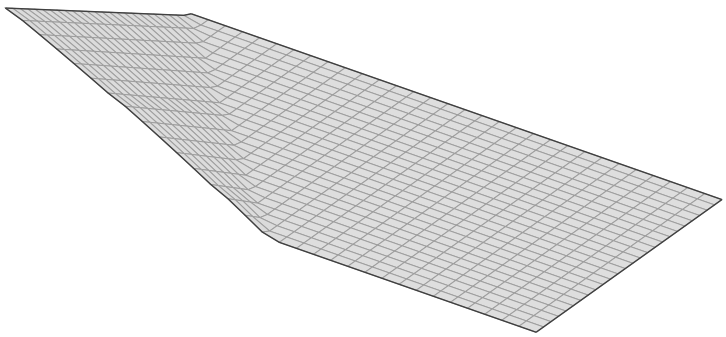
The cost of abstention should be visible

Let coverage be the probability that the system returns a label. Let selective risk be the error rate among returned labels. A system can lower selective risk by abstaining more often. Comparing selective risks without their coverage can reward a system that answers almost nothing.

For a concrete audit, record three numbers together: total cases, returned labels and wrong returned labels. If 900 of 1,000 cases receive labels and nine are wrong, coverage is 0.9 and empirical selective risk is 0.01. The unconditionally wrong-label fraction is 0.009. These are three different quantities. The remaining hundred cases require another action; they are not automatically correct outcomes.

Abstention changes what receives an answer

Height = selective risk on the same declared grid of 1,001 scores; no operational population is inferred.



Horizontal: abstention half-width

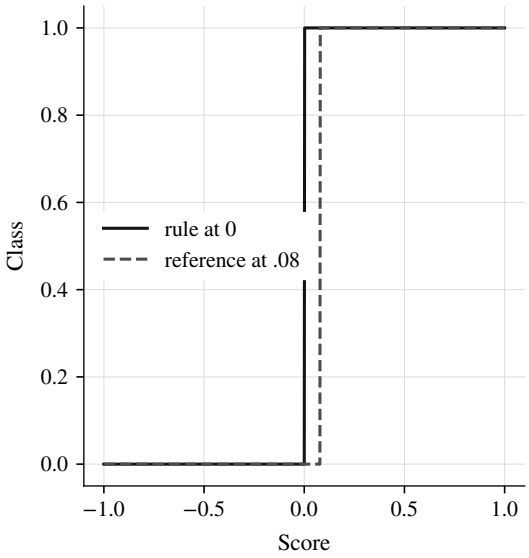
Depth: reference threshold

Height: wrong / returned

Prediction threshold is fixed at zero.

The book compares the slice at 0.08.
Coverage counts returned answers.
Selective risk conditions on return.
Neither substitutes for wrong / all.

A A constructed population separates two rules



B More abstention trades coverage for fewer errors

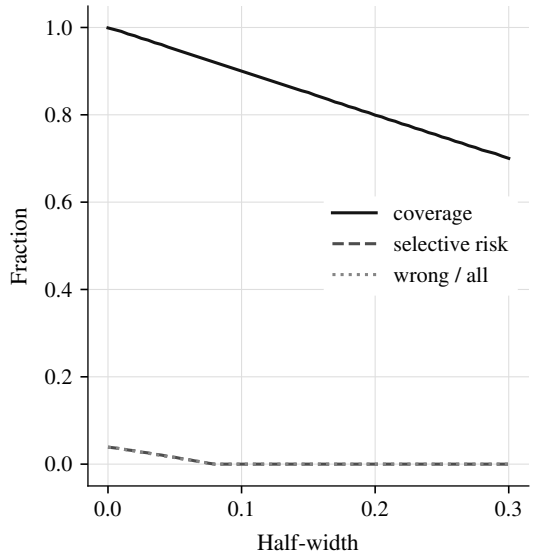
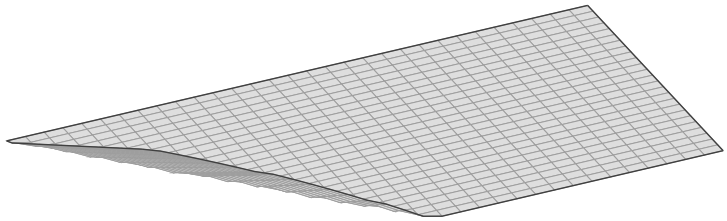


Figure 20.3: *A smaller error rate can come with fewer answers.* The study continues on the following figure page.

Abstention changes what receives an answer

Height = selective risk on the same declared grid of 1,001 scores; no operational population is inferred.



Horizontal: abstention half-width

Depth: reference threshold

Height: wrong / returned

Prediction threshold is fixed at zero.

The book compares the slice at 0.08.
Coverage counts returned answers.
Selective risk conditions on return.
Neither substitutes for wrong / all.

C The stated fixture fixes the comparison

1,001 equally spaced scores [-1,1]
Reference class: score > .08
Rule: score > 0
Return a label only if |score| > width.

D Three rates answer three different questions

Coverage counts returned answers.
Selective risk conditions on return.
Wrong/all uses every input.
This grid is not a sampled operational population.

Figure 20.3: *A smaller error rate can come with fewer answers (continued).* A fully specified illustrative score population is evaluated at several abstention widths. Coverage, selective risk and the fraction of all cases receiving a wrong label are displayed together. The comparison is labeled a constructed audit, not a new empirical GIGI result.

Laboratory 5: preserve a decision through a metric change

Replay a threshold decision on a grid of synthetic two-channel readings (x, y). Use $G = \mathrm{diag}(4, 1)$, score $s = 2x$ and threshold zero. Build the grid over $x \in [-0.3, 0.3]$ and $y \in [-0.2, 0.2]$. Treat $x \le -0.1$ and $x \ge 0.1$ as the confident regions and leave the middle strip undecided.

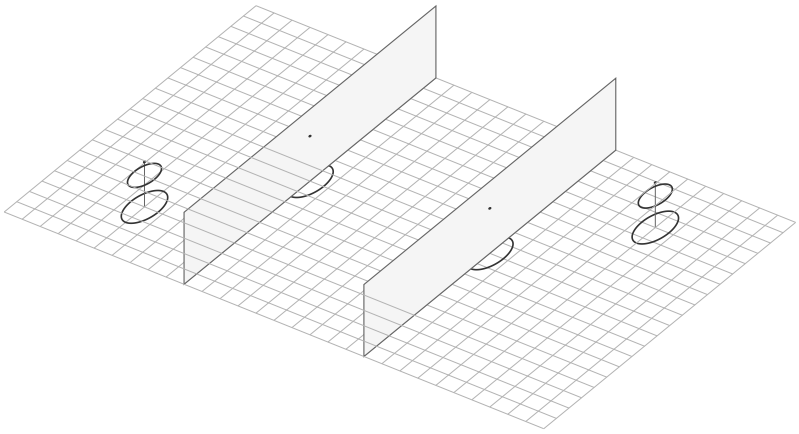
Translate the coordinates. Apply $z = A(x, y)^T$ with the shear matrix from Laboratory 2. Transform the metric to $A^{-T} G A^{-1}$ and the statistic to $s(z) = 2e_1^T A^{-1} z$. Verify pointwise agreement of labels and score values. Reusing the formula $2z_1$ instead is a deliberately incorrect control.

Add bounded uncertainty. At selected grid points draw the G-ball of radius 0.03. Its coordinate semiaxes are 0.015 and 0.03. The score varies by at most 0.03 over the ball. Translate the entire ellipse and verify the same score range in display coordinates.

Account for a second error. Allow a calibration error of at most 0.02 in score units. The total worst-case score error is 0.05. Leave the declared middle strip $|x| < 0.1$ undecided. At a point in a confident region, report its label only if its complete score interval lies strictly on the corresponding side of zero. Count decisions, abstentions and disagreement with the unperturbed reference on your grid. As a separate check, find the narrower band where uncertainty alone prevents a sign decision.

Error has a direction as well as a size

Constructed decision geometry / G = diag(4,1); the separating statistic is s = 2x.



Confident regions: x ≤ −0.1

x ≥ 0.1

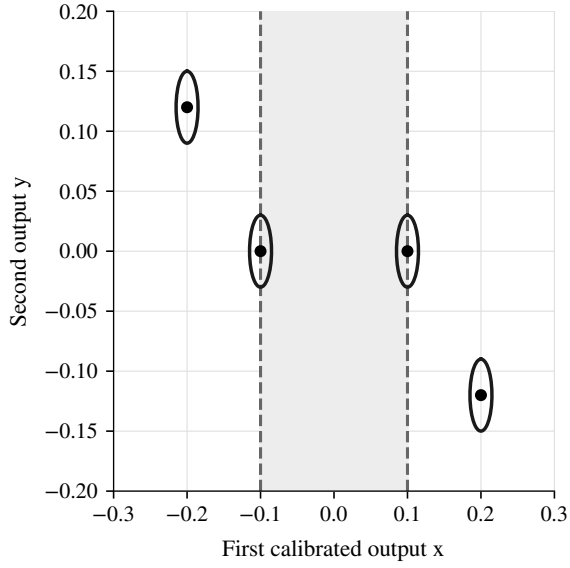
Metric error allowance: 0.03

Score calibration allowance: 0.02

The remaining margin is explicit.

Combined score allowance ≤ 0.05
Remaining confident margin ≥ 0.15
Vertical outlines are explanatory;
their height is not another variable.

A Error ellipses approach the decision boundary



B The shear moves the displayed boundary

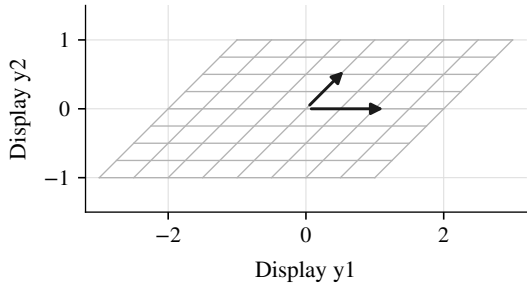
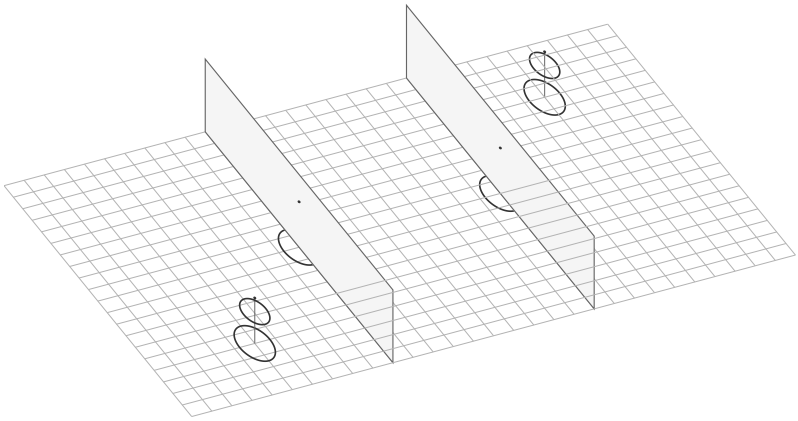


Figure 20.4: *Laboratory record: the boundary moves on the map.* The study continues on the following figure page.

Error has a direction as well as a size

Constructed decision geometry / G = diag(4,1); the separating statistic is s = 2x.



Confident regions: x ≤ −0.1

x ≥ 0.1

Metric error allowance: 0.03

Score calibration allowance: 0.02

The remaining margin is explicit.

Combined score allowance ≤ 0.05
Remaining confident margin ≥ 0.15
Vertical outlines are explanatory;
their height is not another variable.

C The transformed statistic gives the same score

D An unchanged formula is the failed control

G=diag(4,1)	Metric error ≤.03
z=A x	Calibration score error ≤.02
$Gz=A^{-T}\,G\,A^{-1}$	Combined score interval ±.05
$s(z)=2\,e1^{T}\,A^{-1}\,z$	Reusing 2z1 is deliberately wrong.
Scores agree point by point.	Using y ignores the separating direction.

Figure 20.4: *Laboratory record: the boundary moves on the map (continued).* A decision boundary, its uncertainty ellipses and its abstention band are transformed by a shear. Correctly transformed scores agree; a deliberately untransformed score exposes the error. The grid is a deterministic test fixture, not a population sample.

Separate a grid audit from a risk claim. The fraction of grid points that abstain is a property of this chosen grid. It becomes a population coverage estimate only if the grid or another sample represents the intended population under a stated sampling scheme. Repeat the audit with a distribution concentrated near the boundary. The geometry remains the same while coverage changes.

Change the statistic. Use $r = y$ and repeat the cross-configuration comparison. Find two points from opposite confident regions with identical score. Explain why the metric margin survives while this statistic fails to detect it. The exercise tests linkage, not merely numerical stability.

What to submit. Provide the transformed metric and statistic, four representative uncertainty ellipses, a confusion-and-abstention table for the fixture and the failed-statistic example. State which conclusions are exact consequences of the model and which are counts from the finite grid.

Laboratory 5 tested score changes on a fixed grid of two-channel readings. A guarantee for future comparisons also needs to account for how often readings land near the decision threshold.

A complete transfer calculation for a threshold decision

An engineer compares two observations and accepts a match when a distance is at most τ. A second implementation uses translated observations and a different numerical distance. The question is not whether the two distances are identical in every decimal place. It is whether their difference can change the decision, and how often the relevant uncertainty occurs.

Let D be the reference distance and $\widehat{D}$ the translated distance for the same pair. On a good event G, suppose

$$|\widehat{D} - D| \le \varepsilon.$$

Define $a = \mathbf{1}\{D \le \tau\}$ and $\widehat{a} = \mathbf{1}\{\widehat{D} \le \tau\}$. If $D < \tau - \varepsilon$, both accept. If $D > \tau + \varepsilon$, both reject. Consequently,

$$\{a \ne \widehat{a}\} \subseteq G^c \cup \{|D - \tau| \le \varepsilon\}.$$

The boundary convention is included in the closed band. Taking probabilities gives

$$P(a \ne \widehat{a}) \le P(G^c) + P(|D - \tau| \le \varepsilon).$$

No independence between these two events is required. The union bound may be conservative, but its meaning is explicit.

The set inclusion just proved answers the factory inspector's question. A changed label must come from a failed comparison bound or from a part close enough to the acceptance boundary. Counting those possibilities bounds disagreement between implementations. To say how often either implementation is wrong about the part itself, we must now compare with its target label.

From disagreement to error relative to a label

Let $Y \in \{0, 1\}$ be the desired match label under the task's ground-truth convention. Whenever $\widehat{a} \ne Y$, either $a \ne Y$ or $a \ne \widehat{a}$. Therefore

$$P(\widehat{a} \ne Y) \le P(a \ne Y) + P(G^c) + P(|D - \tau| \le \varepsilon).$$

This is a transfer bound. It does not prove that the reference decision is accurate; its error remains in the first term. Nor does it prove that the geometric good event holds often; that is the second term. The third measures how much of the actual population lies near the decision boundary.

These distinctions make the bound useful. A detector can have precise distance preservation and still make poor decisions if its chosen representation or labels do not match the task. It can also have a good reference decision but transfer poorly when many observations sit near its threshold.

Turn multiplicative distortion into a usable band

Suppose the available comparison is

$$(1-\delta)D - b \le \widehat{D} \le (1+\delta)D + b, \qquad 0 \le \delta < 1.$$

On a region with $D \le R$, this yields the simpler allowance $\varepsilon = \delta R + b$. But using the threshold directly can be sharper. A disagreement requires

$$\frac{\tau - b}{1+\delta} \le D \le \frac{\tau + b}{1-\delta},$$

with the lower endpoint replaced by zero if necessary. The interval is generally asymmetric about τ.

For $\tau = 1$, $\delta = 0.02$ and $b = 0.01$, the possible disagreement band runs from approximately 0.970588 to 1.030612. A pair at reference distance 0.5 has ample room; a pair at 0.999 does not. Reporting one mean distortion over a dataset would hide this difference.

A fully specified population example

To inspect the arithmetic, suppose D is uniform on $[0, 2]$, $\tau = 1$, and the good-event additive allowance is 0.03. The threshold band has width 0.06, so its probability is 0.03. If $P(G^c) \le 0.01$ and the reference error is at most 0.04, the transferred error is at most 0.08.

These numbers define a constructed example. They are not performance measurements from AMC, ICARUS or GIGI. Their purpose is to show how three independently meaningful statements combine. If the population instead concentrates half its mass in the threshold band, the same geometric allowance can yield a much less informative bound.

For the asymmetric multiplicative band above and the same uniform population, the band probability is half its width, approximately 0.030012. The bound is computed from the actual interval, not from a symmetric shortcut that silently changes its endpoints.

Abstention makes the guarantee visible at one observation

Suppose an observation yields a certified reference-distance interval $[L, U]$. Accept if $U \le \tau$, reject if $L > \tau$, and otherwise abstain. Every non-abstaining decision agrees with the reference rule whenever the interval contains the true distance. The guarantee is pointwise on that event.

Coverage is the proportion of observations on which the system decides. Selective error is the error rate conditional on deciding. They should be reported together. A rule that abstains almost everywhere can have a small selective error while providing little operational service. The aim is to understand the tradeoff, not to hide the unanswered cases.

If interval containment fails with probability at most α and the decision coverage is $c > 0$, the probability of disagreement conditional on deciding is at most α/c. This follows because the joint event of deciding and disagreeing is contained in the interval-failure event. No claim of a smaller conditional rate follows from the marginal failure rate alone.

What a finite audit can and cannot certify

On a fixed collection of audited pairs, one can calculate exact counts: the number of interval failures, disagreements, abstentions and correct labels under the recorded convention. Those counts describe that collection. A statement about future pairs requires a sampling assumption or another argument connecting the audit to the future population.

For a deterministic domain audit, a finite net plus a known regularity bound can extend coverage. Let $e(x)$ be a scalar discrepancy on a metric domain and suppose it is L-Lipschitz. If every domain point lies within distance h of an audited point and every audited value is at most e_0, then

$$e(x) \leq e_0 + Lh$$

everywhere on the domain. The proof is one application of the Lipschitz inequality to a nearest audited point. The net radius and Lipschitz constant are essential evidence; a dense-looking scatterplot does not supply them automatically.

For a statistical audit, independent representative draws can support confidence bounds on a probability. Dependence, duplicated records, selected easy cases and distribution change all affect that inference. These are not reasons to abandon the experiment. They are reasons to identify what population and what statement the experiment was designed to test.

The central claim in operational form

The translator and geometric implementations share a comparison task when their representations, path geometry and decision statistic are linked by stated bounds. The transfer calculation shows exactly what follows. It preserves the possibility of using either description while keeping the price of approximation visible.

The practical conclusion is not that every nearby pair is the same individual. It is that a specified decision about an individual can survive a change of description when the margin exceeds the transported uncertainty. To check whether the decision survives, compare its margin with the uncertainty bound.

A shifted operating population. Keep the distance allowance and threshold fixed, but let D have density $f(d) = 1 - |d - 1|$ on $[0, 2]$. Verify that the density integrates to one and calculate the mass of the symmetric band $|D - 1| \leq 0.03$. Compare the resulting transfer bound with the uniform-population example.

What reaches the decision

The inspection example carried a distance error of at most 0.01 metre, a comparison-failure probability of at most 0.002, and a boundary-band probability of at most 0.03 into a disagreement bound of 0.032. The laboratory then made acceptance, rejection and abstention inspectable on individual records.

The proof identifies the only records whose labels can change under its assumptions: those for which the comparison fails or whose reference score lies close enough to a threshold. Geometry, linkage, calibration and abstention remain useful diagnostic categories, but their names do not assign causal shares of the total error. Each claimed bound needs its event, target population and uncertainty statement. A false-positive statement conditions all its terms on the negative class.

A distance bound has become a decision bound through the failure event and the mass near the threshold. Its practical use now depends on establishing those assumptions on the population being evaluated. The next chapter follows that evidence through the audit.

Practice with the construction

Back to the maps and the globe. Two maps can agree to within a small positional error. Whether that error changes a navigation decision depends on how close the route runs to a boundary and how often the boat enters that band.

20.1. If the geometric bound fails with probability at most 0.003, $\eta = 0.02$ and the boundary density is bounded by 0.4, give a bound on disagreement.

20.2. For $a = 0.95$, $b = 1.05$, $\eta = 0.02$ and manifold threshold 1, find sufficient translator thresholds for certain acceptance and rejection.

Chapter 21

What the guarantee actually assumes

The factory in Chapter 20 decides whether to accept a part by measuring its displacement from a specified position. A software update changes how it computes that distance. Before relying on the old decision guarantee, the engineer needs evidence about the new computation and the parts it will inspect.

Can the inspection guarantee be used after the update?

Take one test part. Its old inspection report places the hole 3 millimetres along one axis and 4 along the other, exactly at the 5-millimetre acceptance limit. The new system writes those positions in different coordinates. The engineer checks whether both records describe this hole, whether the conversion preserves the displacement, and how often parts like it lie near the limit. Chapter 1's calibration rail gives us a known conversion before we audit the less predictable cases.

From theory to diagnostics

The theoretical conditions in Chapters 9–10 can be summarized (informally) as:

- Translators should behave like smooth, well-conditioned coordinate changes in the regime of interest.
- The metric quotient (X_T, d_T) should be locally Euclidean in that regime.
- There should exist a map Φ such that $d_M(\Phi(x), \Phi(y))$ and $d_T(x, y)$ agree up to controlled multiplicative and additive slack.
- A specified binary decision guarantee transfers when the distance bound and population terms in Chapter 20 are established.

Each statement has measurable diagnostics. A diagnostic is not automatically a certificate of the whole hypothesis: a histogram of local errors does not prove topology, and a sampled maximum does not bound every unseen input.

- **Local Euclidean structure:** distributions of local distortion $\hat{\varepsilon}_{\mathrm{loc}}$.
- **Smooth, well-conditioned translators:** Jacobian singular values and condition numbers.
- **Round-trip and cocycle consistency:** $\hat{\varepsilon}_{\mathrm{rt}}$ and $\hat{\varepsilon}_{\mathrm{cyc}}$.
- **Riemannian realizability:** empirical bi-Lipschitz constants $(\hat{\alpha}_{\min}, \hat{\alpha}_{\max})$ and slack $\hat{\varepsilon}_{\mathrm{geom}}$.
- **Error budget slack:** empirical disagreement, with separate mechanism frequencies when their events are defined.

Table 21.1 summarizes the mapping.

For the test part, a unit conversion should leave the 5-millimetre result intact. The diagnostics below test the routes that convert those coordinates, the directions they stretch, and the larger population on which the inspection rule operates.

Table 21.1: From theory to diagnostics. Each hypothesis from Chapters 9–12 has a corresponding empirical diagnostic in this chapter.

Theoretical condition	**Measurable quantity**	**Section**
Local Euclidean structure on X_T	Local distortion $\hat{\varepsilon}_{\mathrm{loc}}(r)$	Step 2: Local distortion and the paths an audit covers
Smooth translators with bounded distortion	Jacobian singular values, condition numbers	Computing Jacobian spectra
Round-trip consistency of translators	Round-trip error $\hat{\varepsilon}_{\mathrm{rt}}$	Step 3: Estimating round-trip error $\varepsilon_{\mathrm{rt}}$
Cocycle consistency on triples	Mean residual $\hat{\varepsilon}_{\mathrm{cyc}}$	Step 4: Measuring cocycle residuals
Riemannian realizability of $\mathcal{R}_T$	$\hat{\alpha}_{\min}$, $\hat{\alpha}_{\max}$, $\hat{\varepsilon}_{\mathrm{geom}}$	Fitting and validating realization maps Φ
Error budget transfer slack	Comparison failures, threshold-band mass and direct disagreement	Measure decision disagreement and coverage

Estimating chartability constants

We start with the chartability constants:

- $\varepsilon_{\mathrm{loc}}$: local mismatch between within-space metric and quotient metric.
- $\varepsilon_{\mathrm{rt}}$: round-trip error for translator pairs.
- $\varepsilon_{\mathrm{cyc}}$: uniform bound on cocycle residuals.

These checks examine whether the implemented conversions support the distance bound used by the inspection rule.

Step 1: Define your in-regime data set

All of these estimates are conditional on a choice of *regime* $\mathcal{R}_T$.

1. **Identify channels and feature spaces.** List the observation channels and associated feature spaces V_i you care about.
2. **Define in-regime sampling.** Specify the operating population for the application:
   - ICARUS: normalized quaternions, specified frames and a bounded operating scenario.
   - HERALD (High-resolution Early Recognition of Antigenic Landscape Divergence): strains within major variants, assays on well-characterized platforms.
   - AMC (Adaptive Multi-Vector Correlation): benign telemetry windows and common incident types.
   - PRISM: high-volume transaction corridors with stable schemas.
3. **Collect overlapping tuples.** Collect a large sample of pairs and triples of embeddings:

$$\{v_k^{(i)} \in V_i\}, \quad \{(v_k^{(i)}, v_k^{(j)})\}, \quad \{(v_k^{(i)}, v_k^{(j)}, v_k^{(\ell)})\}$$

   where k indexes a record and i, j, ℓ index channels. The factory might log the same hole inspection in the old format, the new format and an archive. Each tuple should refer to the same inspection.

4. **Choose sample size for the claim.** State the population, independence or dependence assumptions, desired precision and confidence level. Five thousand independent representative Bernoulli observations, for example, give a two-sided Hoeffding allowance of $\sqrt{\log(40)/(2 \cdot 5000)} \approx 0.0192$ at 95% confidence [8]. That is an allowance of about 1.92 percentage points, not a certificate of zero rare failures. Reusing the same records in many pairs does not manufacture that many independent observations.

Step 2: Local distortion and the paths an audit covers

For the test part, the comparison is between the measured and specified hole locations. The quotient allows routes through other formats. Could one of those routes make the displacement look smaller than it is? Chapter 10 uses the multiplicative comparison

$$(1 - \varepsilon_{\mathrm{loc}})d_i(v, w) \le d_T(\iota_i(v), \iota_i(w)) \le (1 + \varepsilon_{\mathrm{loc}})d_i(v, w).$$

Its allowance is dimensionless. For $d_i(v, w) > 0$, the corresponding discrepancy is

$$e_{\mathrm{rel}}(v, w) = \left| \frac{d_T(\iota_i(v), \iota_i(w))}{d_i(v, w)} - 1 \right|.$$

An absolute discrepancy $|d_T - d_i|$ instead has distance units and belongs in an additive allowance. Keep both if the instrument has both a relative scale error and an absolute resolution limit. Do not divide by distances below a declared resolution floor; report those pairs separately, including their absolute errors and decisions.

What a short-path search measures. Choose in-regime anchors $v_a \in U_i$, neighbors v_b with $0 < d_i(v_a, v_b) \le r$, and explicit finite families of admissible paths in both directions. Define

$$U_{ab} = \min_{\gamma \in \Gamma_{ab}} L(\gamma), \qquad U_{ba} = \min_{\gamma \in \Gamma_{ba}} L(\gamma).$$

Each selected family is a subset of all allowed paths. Consequently

$$D(v_a, v_b) \le U_{ab}, \qquad d_T(\iota_i(v_a), \iota_i(v_b)) \le U = \max\{U_{ab}, U_{ba}\}.$$

The search supplies an upper bound, not a lower bound. It can expose a shortcut: if $U < (1 - \varepsilon)d_i$, the proposed lower chart bound fails for that pair. Failure to find a shortcut does not prove that none exists.

Think of a map application that tests three driving routes. The shortest of those three is a route you can actually drive, but a fourth road may be shorter. A search through a few translator cycles has the same limitation. Adding more tested paths can lower the computed distance; it cannot turn the search into an exclusion of every untested path.

A small residual can accumulate into a collapse. Take three feature spaces $V_A, V_B, V_C = \mathbb{R}$. All translators are the identity except $T_{AC}(x) = x + 0.01$ and $T_{CA}(x) = x - 0.01$. Every pairwise round trip is exact and every Jacobian is 1. Triangle residuals are at most 0.01.

Yet the zero-cost loop $A \to C \to B \to A$ moves a feature value by 0.01. Repeat it one hundred times to move from 0 to 1 at no path cost. The reverse loop moves back at no cost. Thus $d_T(\iota_A(0), \iota_A(1)) = 0$ although $d_A(0, 1) = 1$. The quotient remains a mathematical metric on its classes, but this chart has lost a distinction. Checking one loop, or only its derivative, would miss the accumulated effect.

How to obtain an actual bound. A uniform local allowance needs control of the true quotient distance, including its admissible path family. On a fully specified finite weighted graph, an exact shortest-path algorithm can inspect all routes implicitly; this certifies that graph's distances. Relating such a graph to a continuous translator system also requires a discretization bound. Alternatively, a proof covering every admissible continuous path can supply a lower bound.

For an audited pair, suppose these arguments establish $L \le d_T \le U$. Then

$$e_{\mathrm{rel}}(v, w) \le \max\left\{\left|\frac{L}{d_i(v, w)} - 1\right|, \left|\frac{U}{d_i(v, w)} - 1\right|\right\}.$$

A bound on sampled pairs extends to a whole compact audit domain if, for example, the samples form a verified finite net and the discrepancy has a known Lipschitz constant. The finite-net argument in Chapter 10 gives the extra allowance. It covers untested inputs only after the path-distance calculation itself has been justified.

What to report from finite data. Report the number of pairs, radii, resolution floor, path families, sample maximum and selected quantiles. A 95th percentile is a description of the sampled error distribution. It does not replace the supremum required by a deterministic chart bound. If only U is available, label the statistic as a short-path diagnostic rather than an estimate certified to bound the true relative error.

Worked calibration example. On the calibration rail, x measures position in metres and $T(x) = 100x + 5$ reports it in shifted centimetres. Transport the metric coefficient to 10^{-4} in the output coordinate. Lengths agree and the relative distortion is zero. Raw coordinate differences alone would report a hundredfold magnification. The known separation tests whether the implementation converts the ruler as well as the numbers.

Step 3: Estimating round-trip error $\varepsilon_{\mathrm{rt}}$

The pointwise round-trip residual is $d_i(T_{ji}(T_{ij}(v)), v)$. Chapter 10's uniform allowance bounds it at every admissible point. A sampled mean and a sampled percentile answer different questions. The Euclidean norms below apply to the chosen feature coordinates; use the declared metric for other charts.

In practice:

1. For each pair of spaces (V_i, V_j) with translators T_{ij}, T_{ji}:
   (a) Sample a batch of in-regime points $\{v_k^{(i)}\} \subset V_i$.
   (b) Compute
   $$\delta_{\mathrm{rt}}(v_k^{(i)}) = \|T_{ji}(T_{ij}(v_k^{(i)})) - v_k^{(i)}\|.$$
2. Define
   $$\hat{\varepsilon}_{\mathrm{rt}}^{i\to j\to i} = \frac{1}{N}\sum_{k=1}^{N} \delta_{\mathrm{rt}}(v_k^{(i)}), \quad \hat{\varepsilon}_{\mathrm{rt},0.99}^{i\to j\to i} = \mathrm{quantile}_{0.99}(\delta_{\mathrm{rt}}).$$
3. Repeat for all translator pairs of interest.

Example: AMC round-trip errors across telemetry stacks. Adaptive Multi-Vector Correlation (AMC) compares telemetry held in different representations. Suppose one operating window appears in a raw log, an alternate metrics pipeline and an aggregation store. Convert the records

there and back, then compare them with the originals. The table uses example numbers to show how to read the check.

Translator pair	Mean $\hat{\varepsilon}_{\mathrm{rt}}$	99th percentile
Logs ↔ Alts	0.023	0.067
Logs ↔ Agg	0.041	0.112
Alts ↔ Agg	0.035	0.089

The logs–aggregation pair has larger reported residuals than the logs–alternate pair. Whether 0.112 is acceptable depends on the declared scale and decision margin. It is a residual value, not an eleven-percent failure rate. Schema drift is one possible cause to investigate against records; the table alone does not identify a cause or prove that a point is outside the regime.

Practical guidance.

- Normalizing embeddings makes $\varepsilon_{\mathrm{rt}}$ dimensionless and easier to interpret.
- Track $\hat{\varepsilon}_{\mathrm{rt}}$ over time; sharp increases are a red flag for schema changes or training problems.

Step 4: Measuring cocycle residuals

For each triple (i, j, k), let $\mathrm{Cyc}_{i\to j\to k}$ be the residual from Chapter 9. A uniform $\varepsilon_{\mathrm{cyc}}$ bounds that discrepancy throughout the stated overlap domain. The empirical mean and percentile below summarize the sampled residuals.

In practice:

1. For each triple (i, j, k) where you have translators T_{ij}, T_{jk}, T_{ik}:
   (a) Sample in-regime points $\{v_\ell^{(i)}\}_{\ell=1}^{N} \subset V_i$ whose images lie in the declared domains in V_j and V_k.
   (b) Compute
   $$r_\ell = \mathrm{Cyc}_{i\to j\to k}(v_\ell^{(i)}) = d_k(T_{jk}(T_{ij}(v_\ell^{(i)})), T_{ik}(v_\ell^{(i)})).$$
2. Aggregate:
   $$\hat{\varepsilon}_{\mathrm{cyc}}^{i,j,k} = \frac{1}{N}\sum_{\ell=1}^{N} r_\ell, \quad \hat{\varepsilon}_{\mathrm{cyc},0.99}^{i,j,k} = \mathrm{quantile}_{0.99}((r_\ell)_{\ell=1}^{N}).$$

Practical guidance.

- High cocycle slack often indicates misaligned training objectives or missing context in one of the spaces.
- When several platforms report the same underlying observations, record $\hat{\varepsilon}_{\mathrm{rt}}$ for each platform pair and $\hat{\varepsilon}_{\mathrm{cyc}}$ for each triple. The register shows where the routes disagree.

Interpreting the constants

Report local length distortion, round-trip residual, cocycle residual and Jacobian conditioning with their units, norms, domains and sampling rule. A practical threshold is chosen from the task's tolerance and downstream margin; there is no domain-independent traffic-light cutoff. Sampling estimates and uniform mathematical bounds must remain distinguishable in the report.

Computing Jacobian spectra

A conversion might return the test part's coordinates after a round trip and still stretch nearby positions differently in each direction. The Jacobian measures what happens to a small displacement of the hole before the reverse conversion has a chance to undo it.

For $T_{ij} : V_i \to V_j$, the derivative at $v \in V_i$ maps source tangent vectors to target tangent vectors. Use orthonormal coordinates for the declared tangent metrics. If their positive-definite matrices are G_i and G_j, inspect the whitened matrix $G_j^{1/2} JT_{ij}(v) G_i^{-1/2}$. Ordinary Euclidean singular values are appropriate when both matrices are the identity. Changing units without transporting these metrics can otherwise look like a change in conditioning.

Automatic differentiation vs. finite differences

There are two main approaches:

- **Automatic differentiation:** If T_{ij} is implemented in a differentiable framework (for example, as a neural network), you can compute Jacobian–vector products efficiently. This lets you estimate leading singular values without forming the full Jacobian matrix.
- **Finite differences:** If you treat T_{ij} as a black box, you can approximate directional derivatives by
$$JT_{ij}(v)\, h \approx \frac{T_{ij}(v + \epsilon h) - T_{ij}(v)}{\epsilon}$$
for small ϵ and random directions h.

Common pitfalls.

- **Finite difference instability:** if ϵ is too large, derivatives are inaccurate; if too small, numerical noise dominates. For normalized embeddings, $\epsilon \sim 10^{-4}$ to 10^{-6} often works.
- **Sampling bias:** estimating Jacobians only on easy, high-quality data gives overly optimistic condition numbers. Sample across your full in-regime distribution, including boundary cases.
- **Ignoring direction dependence:** a single average condition number can hide bad directions. Always look at the distribution.

Estimating singular values via power iteration

To estimate the largest singular value $\sigma_{\max}$ at v:

1. Choose a random unit tangent vector $h_0 \in T_v V_i$, expressed in the source orthonormal coordinates.
2. Iteratively apply:
$$h_{k+1} = \frac{JT_{ij}(v)^\top \big(JT_{ij}(v) h_k\big)}{\big\| JT_{ij}(v)^\top \big(JT_{ij}(v) h_k\big) \big\|}.$$
3. After convergence checks and comparison with an independent calculation where feasible, estimate
$$\sigma_{\max}(v) \approx \big\| JT_{ij}(v) h_k \big\|.$$

Finite differences approximate forward Jacobian–vector products; a transpose action needs an adjoint, automatic differentiation or an explicit matrix. The displayed iteration uses orthonormal coordinates. If its denominator vanishes, restart with another direction or inspect the kernel.

Power iteration estimates the leading direction; numerical convergence alone is not a rigorous upper bound on all singular values.

Estimating $\sigma_{\min}$ is more involved. Options include:

- For an exact differentiable local inverse between equal-dimensional spaces,

$$JT_{ji}(T_{ij}(v)) = JT_{ij}(v)^{-1}.$$

  Hence $\sigma_{\min}(JT_{ij}(v)) = 1/\sigma_{\max}(JT_{ji}(T_{ij}(v)))$ in the corresponding tangent metrics. A learned reverse map needs a derivative-level bound, not only a small value-level round trip.
- On moderate-dimensional spaces, form the forward Jacobian and compute its full spectrum with numerical error checks. Define the lower stretch as $\inf_{\|h\|=1} \|Jh\|$ over all source directions. If the target has fewer dimensions than the source, a kernel exists and this value is zero, even if an economical rectangular SVD prints only positive singular values.

Why accurate positions do not guarantee accurate derivatives. Let $f(x) = x + 0.9\epsilon \sin(x/\epsilon)$ and let the learned reverse be $g(x) = x$. Both round-trip value errors are at most 0.9ϵ. But $f'(\pi\epsilon) = 0.1$ while $g' = 1$, so using the reverse derivative would predict a lower stretch of 1 where the forward stretch is 0.1. A conversion can return a position very accurately while misrepresenting how nearby positions separate.

For a square forward Jacobian A and learned reverse derivative B evaluated at the corresponding point, a certified residual $\|BA - I\|_2 \le \eta < 1$ gives

$$\|Ah\|_2 \ge \frac{1-\eta}{\|B\|_2}\|h\|_2.$$

Indeed, $\|BAh\| \ge (1-\eta)\|h\|$ and $\|BAh\| \le \|B\|\|Ah\|$. This is a usable lower-stretch bound once the derivative residual and operator norm are established.

Aggregating Jacobian spectra

The distribution of sampled singular values reveals where a translator behaves poorly. A uniform chartability claim additionally needs bounds covering all points and source directions in its domain.

1. Sample $v_k^{(i)}$ from $U_i \subseteq V_i$ with $\iota_i(v_k^{(i)}) \in \mathcal{R}_T$.
2. For each source point $v_k^{(i)}$ and translator T_{ij}:
   - Estimate $\sigma_{\max}(v_k)$ and (when possible) $\sigma_{\min}(v_k)$.
3. Build histograms of:

$$\{\sigma_{\min}(v_k)\}, \quad \{\sigma_{\max}(v_k)\}, \quad \left\{\frac{\sigma_{\max}(v_k)}{\sigma_{\min}(v_k)}\right\}.$$

A reproducible condition-number example. Construct 5,000 two-dimensional local linear maps with Jacobians $J = \mathrm{diag}(1, k)$. Use $k = 2.4$ for 3,000 maps, $k = 8.7$ for 1,999, and $k = 23.1$ for one. These are stipulated matrices, not observations from an application. Their singular values are 1 and k, so their condition numbers are exactly k.

The sample median is 2.4, the nearest-rank 95th percentile is 8.7, and the maximum is 23.1. The table shows a central group, a broad region of greater directional distortion, and one more extreme case. None of those values is universally acceptable or unacceptable: compare the resulting uncertainty with the decision margin.

Even a condition number of one can hide a scale problem. The matrices $0.001I$ and $1000I$ both have condition number one. One compresses all lengths a thousandfold; the other expands them a thousandfold. Report $\sigma_{\min}$ and $\sigma_{\max}$ as well as their ratio.

The HERALD connection. HERALD's source manuscript, *High-resolution Early Recognition of Antigenic Landscape Divergence*, develops a pullback-metric construction and a conditional comparison between latent and geodesic distances [16]. The relevant question here is what ruler a representation induces and over which region it supports the desired comparison. An application audit must identify the actual maps, tangent metrics, sampled records and singular-value calculations. The constructed matrix example above explains how to read such a report without attributing its numbers to HERALD measurements.

Fitting and validating realization maps Φ

For the test part, we also need to relate the distance computed from stored records, d_T, to the distance measured by the inspection model, d_M. A realization map Φ locates each record in that model. A fitted embedding is a candidate for Φ. Its dimensionality, topology and metric agreement need their own evidence.

This section is about fitting such a map and validating the empirical bi-Lipschitz-with-slack inequality

$$\alpha_{\min}\, d_T(x,y) - \varepsilon_{\text{geom}} \;\le\; d_M(\Phi(x),\Phi(y)) \;\le\; \alpha_{\max}\, d_T(x,y) + \varepsilon_{\text{geom}}$$

for most in-regime pairs.

Choosing a model family for Φ

In many applications, a simple model is sufficient:

- **Scalar rescaling:** if both d_T and d_M are distances on the same latent space, Φ may be close to identity; you mainly calibrate distances.
- **Global linear map:** $\Phi(x) = Ax + b$ with matrix A and vector b to align two linear embeddings.
- **Piecewise linear/charts:** fit separate linear maps on different chart regions of X_T.
- **Distance calibration:** given already specified point maps, fit a scalar monotone function f so that

  $$d_M(\Phi(x),\Phi(y)) \approx f\big(d_T(x,y)\big),$$

  by regressing d_M on d_T over sampled pairs. This compares a statistic; it does not itself define a point map Φ or establish a manifold realization.

Common pitfalls.

- **Overfitting Φ:** complex models (deep networks) can perfectly fit noise. Start simple.
- **Train/test leakage:** fitting and validating Φ on the same pairs yields overly optimistic $\hat{\varepsilon}_{\text{geom}}$. Always use a held-out test set.
- **Extrapolation:** a good fit on in-regime pairs does not guarantee good behavior out-of-regime. Do not transfer budgets outside the region where Φ was validated.

Fitting Φ on paired samples

Suppose you have:

- Points $\{x_k \in \mathcal{R}_T\}$ sampled from the translator metric space.
- Their images under a manifold encoding $\psi(x_k) \in \mathcal{R}_M$ (for example, via chart encoders).

You can fit Φ by minimizing

$$\mathcal{L}(\Phi) = \sum_k d_M\big(\Phi(x_k), \psi(x_k)\big)^2$$

for a chosen model family.

For pure distance calibration, you can instead work with pairwise distances:

$$\mathcal{L}(f) = \sum_{k,\ell} \Big(f(d_T(x_k, x_\ell)) - d_M(\psi(x_k), \psi(x_\ell))\Big)^2,$$

with a monotone function f.

Split your data into training, validation, and test sets.

Validating the bi-Lipschitz-with-slack inequality

Separate fitting, calibration and final testing. Fit Φ on training data. On a calibration set, choose fixed positive factors $a \le b$ and an additive allowance $e \ge 0$ for

$$ad_T - e \le d_M(\Phi(x), \Phi(y)) \le bd_T + e.$$

A regression can suggest a and b. If its intercept is β, that offset must be accounted for: a residual bound r around a slope $\alpha > 0$ gives the candidate $a = b = \alpha$ and allowance $|\beta| + r$ on the pairs satisfying that residual bound. A fitted percentile of residuals still describes a sample, not a uniform envelope.

For each calibration pair, compute the required additive allowance

$$s(x, y) = \max\{0,\ ad_T - d_M,\ d_M - bd_T\}.$$

The sample maximum covers those calibration pairs for the chosen factors. A percentile proposes a smaller allowance with some sample failures. Freeze the map, factors, allowance, regime and decision thresholds before testing them on fresh data.

A probability statement from a held-out audit. On N independent representative test pairs, let $\widehat{p}_{\text{fail}}$ be the observed fraction with $s > e$. For fixed choices made before inspecting this test set, Hoeffding's inequality gives, with confidence at least $1 - \beta_0$,

$$p_{\text{fail}} \le \min\left\{1, \widehat{p}_{\text{fail}} + \sqrt{\frac{\log(1/\beta_0)}{2N}}\right\}.$$

This bounds a population failure probability under the sampling assumptions. It does not certify every input. If the claim is conditional on a class or quality gate, sample that population and use its sample count. Reused observations, time dependence and post-test tuning require an appropriate different analysis.

Fifty failures in five thousand comparisons. With $N = 5{,}000$, $\widehat{p}_{\text{fail}} = 0.01$ and $\beta_0 = 0.05$, the one-sided allowance is about 0.01731, giving an upper bound of about 0.02731. In the inspector's tally, 50 of 5,000 distance comparisons exceeded the agreed error allowance. That's one percent in this test. The 95 percent upper bound for the population is about 2.73 percent.

The transfer theorem then needs the probability mass near the decision threshold as well as this good-event failure bound. If both probabilities are estimated, allocate confidence allowances to both audits and combine them, for example by a union bound. A deterministic version instead requires a uniform comparison proved over its full domain and path family.

Practical checks. Inspect residuals across distance, channel and operating region; a fit near zero may fail at larger scales. Track changes over time, but validate any revised constants on new held-out records. Keep a scalar distance calibration separate from evidence that a pointwise realization map exists. This makes the final statement readable: which map, which region, which distance allowance, which probability, and under which sampling assumptions.

Measure decision disagreement and coverage

Return to the part at the 5-millimetre limit. The transfer theorem bounds disagreements using comparison failures and the probability of lying near a threshold. Measure both quantities on the intended population. A direct disagreement audit checks how often the two complete rules return different answers.

Run both specified detectors on the same labeled pairs and measure disagreement. To investigate its causes, define separate observable events and inspect their overlaps; total disagreement alone does not allocate causal contributions.

Step 1: Fix the decision rules

Choose:

- For a binary comparison in common units, declare the acceptance threshold and the distance allowance that fixes its boundary band.
- For a three-way comparison, declare the manifold-side thresholds $(\tau^M_{\text{same}}, \tau^M_{\text{diff}})$. Specify which endpoint conventions apply and when the rule abstains.
- Fix the translator-side rules using the comparison bounds in Chapter 20. Record the population on which those bounds apply.

Step 2: Sample labeled pairs

Draw a labeled sample of pairs (x_i, y_i) from your regime with ground-truth labels $Y_i \in \{\text{same}, \text{different}\}$.

For each pair:

- Compute $d_T(x_i, y_i)$ and $d_M(\Phi(x_i), \Phi(y_i))$.
- Compute manifold decision $\hat{Y}_i^M$ and translator decision $\hat{Y}_i^T$ (with abstentions if using a three-zone policy).

Step 3: Measure disagreement probability

Estimate

$$\widehat{p}_{\text{disagree}} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}\{\hat{Y}_i^T \neq \hat{Y}_i^M\}.$$

Report disagreement together with coverage for each rule if either can abstain. The observed fractions describe the sample. Confidence bounds on future fractions require a justified sampling

model. The numerical distance allowance determines a band, and the observed population determines how much mass lies in it.

Attribute mechanisms only with evidence

Define geometry, linkage, calibration and abstention failures as separate observable events or adjudicated mechanisms. They can overlap. Report their individual frequencies and overlaps when the records support that classification. Use a union bound for a total upper bound. Do not divide the total disagreement among named mechanisms by arbitrary proportions and present the resulting numbers as measured causal contributions.

Practical guidance.

- Compare proposed abstention widths on calibration data and report disagreement and coverage together. Freeze the selected width before testing the chosen policy on fresh records.
- Choose a confidence method consistent with the sampling unit and dependence. Resample independent units or suitable blocks where justified; an ordinary pair bootstrap does not remove dependence caused by shared records.

When is error budget transfer justified?

We have now seen how to measure chartability, local conditioning, realizability, and empirical slack. The remaining question is: *when are these measurements good enough to trust error budget transfer?*

Checklist: preconditions for transfer

Table 21.2 summarizes a pre-flight checklist.

Table 21.2: Checklist for deciding whether to apply error budget transfer.

Condition	Notes
In-regime region $\mathcal{R}_T$ defined and sampled	Clear definition of what data you care about (channels, demographics, regimes).
Chartability constants $\hat{\varepsilon}_{\mathrm{loc}}, \hat{\varepsilon}_{\mathrm{rt}}, \hat{\varepsilon}_{\mathrm{cyc}}$ estimated	Empirical summaries identified as such; uniform claims need a domain argument.
Both extreme singular values controlled on the claimed region	Tangent metrics, rank, units and relevant decision margins specified.
Point map Φ and any separate scalar calibration validated	Fitted and calibrated first; fixed choices checked on fresh test records.
Empirical bi-Lipschitz constants $(\hat{\alpha}_{\min}, \hat{\alpha}_{\max}, \hat{\varepsilon}_{\mathrm{geom}})$ bounded	Distortion factors close to 1; slack small relative to decision margins.
Labeled pair sample collected for disagreement estimation	Sampling unit, target population, precision and confidence specified.
Empirical disagreement $\widehat{p}_{\mathrm{disagree}}$ small	Observed fraction and justified upper confidence bound reported separately.
Comparison failures and boundary-band mass bounded	Population and confidence statements recorded.

A decision from the audit

The audit passes the intended claim only when its required bounds, population, path family and decision margin are supported. A restricted pass can apply on a smaller declared region, with coverage reported. If the needed assumptions are unverified or the bound is uninformative, retain the observations and withhold that particular guarantee. The thresholds for this judgment come from the application and the theorem, not universal numerical cutoffs.

When audits fail: recovery strategies

If the evidence falls short of the requested guarantee, identify which bound failed before choosing a remedy.

Strategy 1: Restrict the regime. If chartability fails globally but holds in subregions:

- Define a smaller $\mathcal{R}_T$ (for example, exclude very low-quality data, rare flows, or extreme operating conditions).
- Re-run diagnostics on this restricted regime.
- Accept reduced coverage in exchange for tighter guarantees.

Example: A PRISM-style audit could restrict its comparison guarantee to well-supported payment routes and send cases on unverified routes for review.

Strategy 2: Improve translators. If $\hat{\varepsilon}_{\mathrm{rt}}$ or $\hat{\varepsilon}_{\mathrm{cyc}}$ are large:

- Collect more paired/overlapping data across platforms or channels.
- Add explicit cycle-consistency or round-trip losses to translator training.
- Introduce domain-specific features that resolve ambiguity between spaces.

Example: In a HERALD comparison, new overlap records can be used to test whether revised translators reduce the sampled cocycle residual on held-out observations. The improvement must be measured, not assumed from having more data.

Strategy 3: Improve encoders. If Jacobian condition numbers are very large:

- Increase embedding dimensionality to reduce collapse.
- Add auxiliary losses to preserve local geometry (for example, contrastive or triplet losses).
- Experiment with architectures that produce smoother embeddings.

Example: In an AMC-style telemetry system, two encoder versions can be compared on the same overlapping records. Test whether moving a feature into the encoder improves local distortion and the smallest and largest stretches.

Strategy 4: Work in one realization only. If chartability fundamentally fails (for example, channels are too heterogeneous to stitch):

- Do not force error budget transfer.
- Work entirely in translator space *or* entirely in manifold space.
- Treat geometry and translators as separate tools rather than equivalent realizations.

This is honest engineering: the framework is falsifiable, and sometimes the verdict is "not applicable."

Choosing geometry-first vs. translator-first

Even when diagnostics look good, you still face a design choice: should you start from geometry or from translators?

Key questions

Ask yourself:

1. **Is there a natural geometric notion of distance?**
   - If yes (for example, a specified antigenic model in HERALD or physical orientations in ICARUS), geometry-first is attractive.
   - If distance is emergent or opaque, translator-first may be more realistic.
2. **How heterogeneous are your observation channels?**
   - Highly heterogeneous legacy systems (AMC, PRISM) push toward translator-first: it is easier to train translators than to impose a unified geometry at the start.
   - Homogeneous or tightly controlled channels (many vision and scientific settings) can support geometry-first more easily.
3. **Where is your expertise?**
   - Strong domain theory about geometry (immunology, physics) suggests geometry-first.
   - Strong engineering expertise in existing pipelines suggests translator-first.
4. **Where will the detector be deployed?**
   - If deployment must sit inside existing translator-heavy systems, translator-first is often the only practical choice.
   - If you control the representation end-to-end, geometry-first is more plausible.

A simple decision tree

A rough decision pattern:

- If your domain has:
  - An established notion of distance,
  - A small number of channels,
  - And freedom to design encoders,

  then start **geometry-first**: design a manifold model G, then construct a translator realization T, and use F_S to analyze compatibility.
- If your domain has:
  - Many heterogeneous legacy channels,
  - Evolving schemas and platforms,
  - Tight deployment constraints,

  then start **translator-first**: build T, audit chartability, and induce G with F_S where possible.

In practice, many systems use both: start translator-first to integrate heterogeneous pipelines quickly, then add geometry-first analysis once you understand the empirical structure. The framework supports both directions.

Example audit workflows

To make these tools less abstract, here are compressed audit workflows for two representative scenarios.

An orientation audit

For the exact quaternion model, test that q and $-q$ give the same rotation matrix and orientation distance. Compare the quaternion formula with the rotation-matrix trace formula on sampled orientations, including near zero and near a half-turn. Separately test continuous lifted paths, since endpoint sign invariance does not retain a winding history. Then evaluate the specified controller in its simulator with its recorded baseline and trial protocol. These are distinct geometry, implementation and dynamics checks.

AMC-style audit (translator-first)

A translator-first audit in an AMC-like system (telemetry across stacks):

1. **Setup.**
   - Define telemetry regimes: $\mathcal{R}_{\text{benign}}$, $\mathcal{R}_{\text{incident}}$, $\mathcal{R}_{\text{unknown}}$.
   - For each regime, collect overlapping telemetry from multiple stacks (logs, alternate metrics, aggregations).
2. **Chartability.**
   - Train translators between stacks on overlapping windows.
   - Estimate $\hat{\varepsilon}_{\text{loc}}, \hat{\varepsilon}_{\text{rt}}, \hat{\varepsilon}_{\text{cyc}}$ in each regime.
   - Compute Jacobian spectra and flag unstable regions or stacks.
3. **Induced geometry.**
   - On $\mathcal{R}_{\text{benign}}$, fit a manifold model M_{AMC} (for example, via dimensionality reduction on X_T).
   - Fit a realization map Φ and validate bi-Lipschitz constants.
4. **Slack estimation.**
   - Use historical labeled incidents to build error budgets on X_T.
   - Use disagreement experiments with justified confidence bounds to compare the specified decisions on X_T and M_{AMC}. Audit each named failure mechanism separately.
5. **Decision.**
   - Check the requested guarantee separately on ordinary windows $\mathcal{R}_{\text{benign}}$ and incident windows $\mathcal{R}_{\text{incident}}$.
   - Use automatic actions where the required bounds are supported. Record any restricted scope and send unsupported cases to human review.

An audit has to reproduce the decision

An engineer should be able to choose one rejected record, reopen the run and reconstruct why it was rejected. That requires more than a plot of average distances. Keep the raw observation identifiers, units, source versions, calibration, translator route, metric convention, thresholds and the error bounds actually used. If the system abstained, retain the reason. If a record was excluded, retain that reason too.

To reproduce a worked example, start with its inputs. For a simulation, record the model and comparator. For a measured-data study, record the dataset and protocol. Those details determine which decisions we can reconstruct and what the results support.

The transfer calculation also needs a reference error bound. One way to obtain such a bound is to specify a score model with known class means, variances and prevalence. We now derive that route to an alert-reliability claim, then ask what finite testing can establish when those population quantities must be estimated.

A finite-variance guarantee with all its conditions

The Davis system includes a useful finite-variance calculation [17]. Let S be a real statistic and $Y \in \{0, 1\}$ its target. On a specified population or quality-gated region, let $\mu_y = \mathbb{E}[S \mid Y = y]$, $\sigma_y^2 = \mathrm{Var}(S \mid Y = y)$ and $\pi_y = \Pr(Y = y)$. Assume finite variances and $\Delta = \mu_1 - \mu_0 > 0$. At the midpoint $s_* = (\mu_0 + \mu_1)/2$, set

$$c_y = \frac{\sigma_y^2}{\sigma_y^2 + (\Delta/2)^2}.$$

Then Cantelli's inequality gives $\Pr(S > s_* \mid Y = 0) \le c_0$ and $\Pr(S \le s_* \mid Y = 1) \le c_1$. For the actual alert event $A = \{S > s_*\}$, Bayes' rule yields

$$\Pr(Y = 1 \mid A) \ge \frac{\pi_1(1 - c_1)}{\pi_0 c_0 + \pi_1(1 - c_1)},$$

provided the denominator is positive. The expression increases with true-positive rate and decreases with false-positive rate, so substituting their respective bounds has the stated direction.

Here is a short proof of the inequality itself. For centered X with variance σ^2 and $a, b > 0$, the event $X \ge a$ implies $(X + b)^2 \ge (a + b)^2$. Markov's inequality therefore bounds its probability by $(\sigma^2 + b^2)/(a + b)^2$. Choosing $b = \sigma^2/a$ gives $\sigma^2/(\sigma^2 + a^2)$; the zero-variance case follows directly. Apply this to $S - \mu_0$ and to $\mu_1 - S$ in their respective classes.

If we additionally filter alerts, the selected event changes. A lower bound for $\Pr(Y = 1 \mid A)$ is not automatically a lower bound for $\Pr(Y = 1 \mid A \cap Q)$. The filter could retain a disproportionate share of false alerts. Either redo the conditional calculation inside Q or establish an explicit selection assumption. If all moments were estimated inside an operating region, every prevalence and probability in the displayed guarantee must use that same region.

What finite testing can certify

For a fixed detector evaluated on n independent held-out pairs from the target distribution, let $\hat{p}$ be the observed disagreement frequency. Hoeffding's inequality gives the one-sided upper bound

$$p \le \min\left\{1, \hat{p} + \sqrt{\frac{\log(1/\alpha)}{2n}}\right\}$$

with confidence at least $1-\alpha$ [8]. This is a distribution-free sampling bound for that fixed procedure under independent sampling. It is not valid merely because a dataset contains n rows. Repeated frames from one sequence, repeated readings from one instrument and threshold selection on the same test data change the sampling problem.

With $n = 1000$, zero observed disagreements and $\alpha = 0.05$, the displayed bound is about 0.0387. A zero in a table is not a proof of zero population risk. More tailored binomial methods can be sharper under the same Bernoulli setup; the simple bound here makes the assumptions and arithmetic visible.

A *Retain the inputs to the decision*

Retained field	*Declared example*
Observation	p = (−0.1, 0), q = (0.1, 0)
Frame / units	Same two-axis frame
Metric version	G = diag(4, 1)
Endpoint bound	e = 0.03 each
Decision threshold	$\tau = 0.30$
Measured separation	$\hat{d} = 0.40$

B *Every allowed distance clears the threshold*

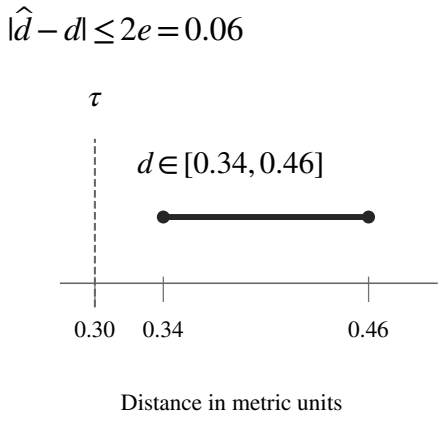


Figure 21.1: A constructed audit record with enough information to reconstruct a decision. In this example each endpoint has metric error at most 0.03, so a measured distance of 0.40 admits true distances from 0.34 to 0.46. Every admitted value exceeds the declared threshold 0.30.

An engineer assessing the new factory inspection software can combine bounds for several failure events. A finite experiment must establish those bounds on the parts the software will inspect.

A guarantee is a chain of events

Trace an incorrect rejection back through the inspection. The part may have been outside the calibrated range. The encoding may have exceeded its distortion allowance. The statistic may have ignored displacement in a relevant direction. The threshold may have been miscalibrated. These events can lead to the same decision and can occur together.

If successful operation requires events $A_1, \ldots, A_k$, the probability that at least one fails satisfies

$$\Pr\left(\bigcup_i A_i^c\right) \le \sum_i \Pr(A_i^c).$$

No independence assumption is needed for this union bound. Adding estimated component failure rates does not, however, make the estimates exact or ensure they refer to the same operational population. The record must state how each was obtained.

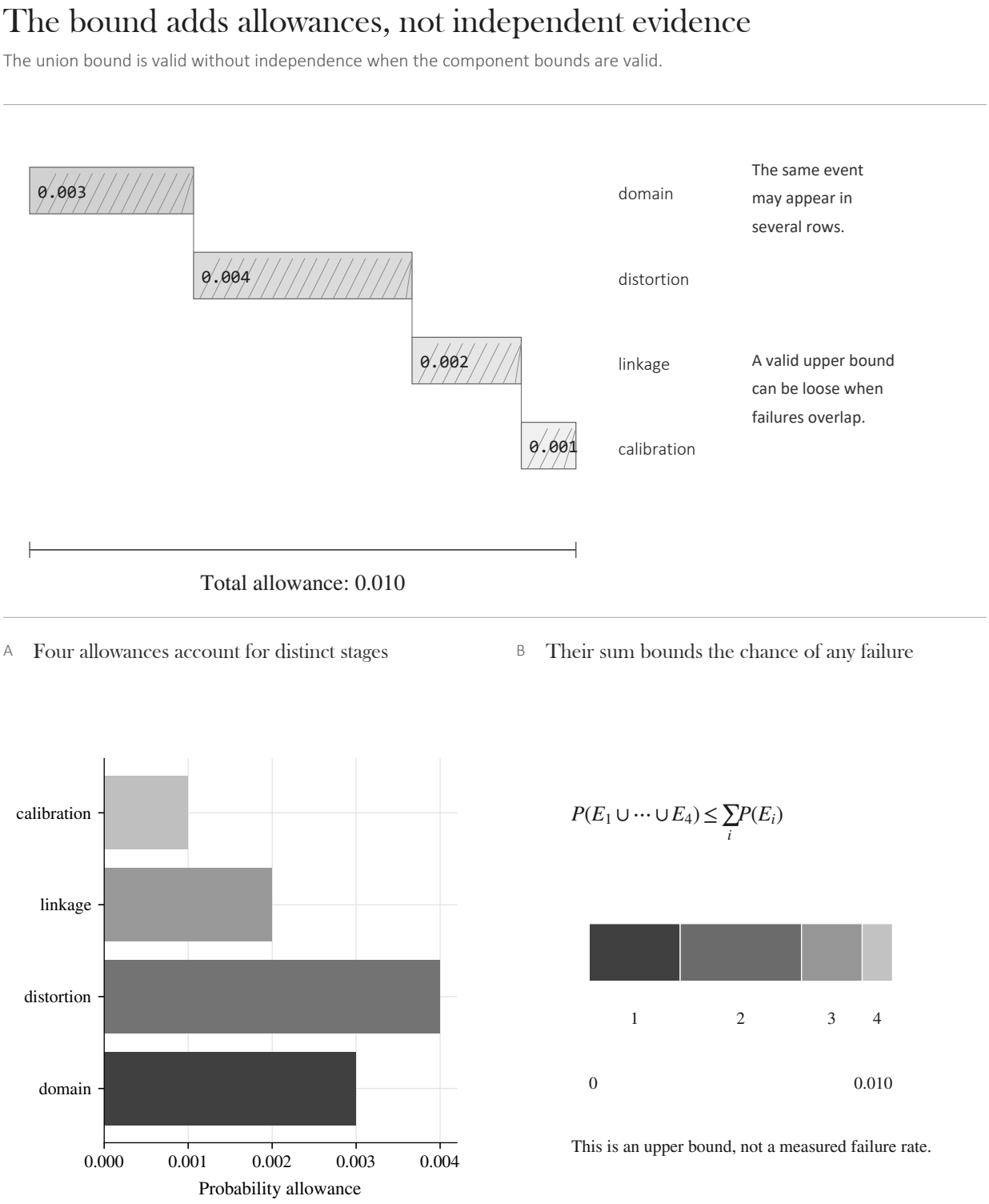
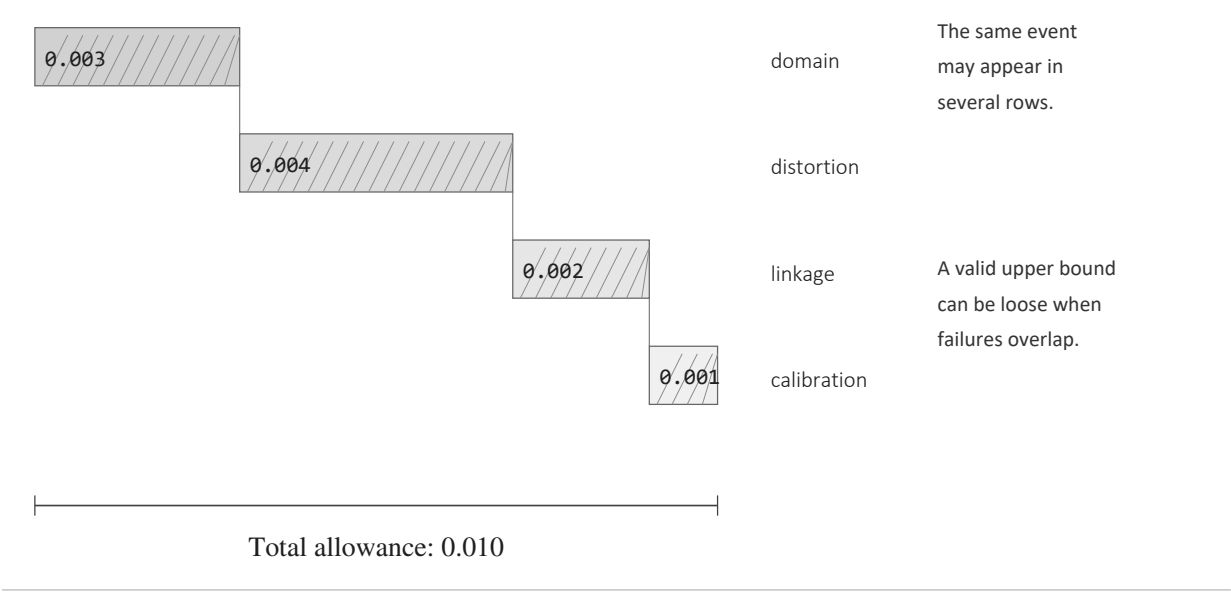


Figure 21.2: *A decision error has several possible entry points.* The study continues on the following figure page.



C The same event can occupy all four rows

D Every allowance needs supporting evidence

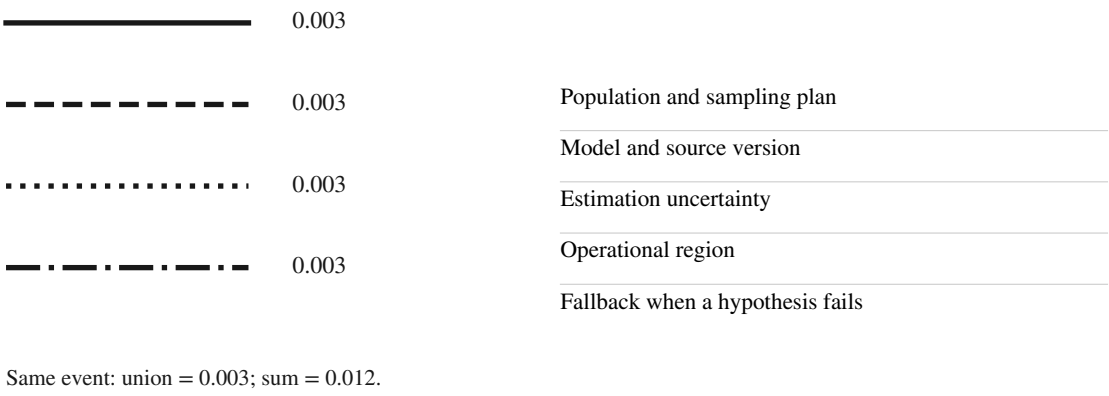


Figure 21.2: *A decision error has several possible entry points (continued).* Four declared component allowances sum to 0.010. The union-bound diagram states what this total bounds. In the correlated example, all four rows describe the same event: the union has probability 0.003 although the sum is 0.012. Independence is unnecessary, but every component allowance needs valid supporting evidence.

Inspect how conservative the variance bound can be

For a scalar random variable Z with mean μ and variance σ^2, Cantelli's inequality states

$$\Pr(Z - \mu \geq a) \leq \frac{\sigma^2}{\sigma^2 + a^2}, \qquad a > 0.$$

One proof applies Markov's inequality to $(Z - \mu + t)^2$, giving $(\sigma^2 + t^2)/(a + t)^2$, then minimizes over $t \geq 0$ at $t = \sigma^2/a$.

The bound does not require Gaussian tails. That generality has a price: it can be loose. With $\sigma = 1$ and $a = 3$, it gives 0.1. A normal distribution has a much smaller upper tail at three standard deviations, but using that smaller number requires the normal model or other tail information.

The mean and variance in the theorem are population quantities. Plugging sample estimates into the same formula without accounting for their uncertainty does not produce the same guarantee. A complete audit either supplies bounds on those quantities or states clearly that the calculation is a descriptive estimate.

Zero observed failures does not mean zero failure probability

For n independent Bernoulli trials with a common failure probability p, the probability of observing no failures is $(1 - p)^n$. A one-sided exact upper confidence endpoint after zero failures is therefore

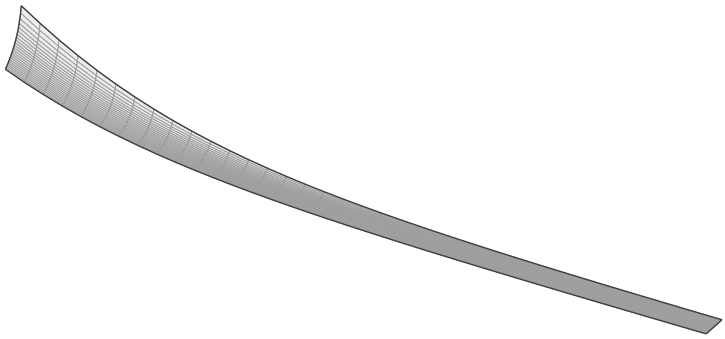
$$p_U = 1 - \alpha^{1/n}$$

at confidence level $1 - \alpha$. With $n = 100$ and $\alpha = 0.05$, it is approximately 0.02951. With $n = 1{,}000$, it is approximately 0.002991.

Those numbers concern the declared independent sampling experiment. Repeating the same deterministic input a thousand times does not provide a thousand independent operational trials. A seeded simulation can be fully reproducible and still cover a narrow model family. Reproducibility and population coverage answer different questions, and both belong in an evidence table.

Zero failures leaves a nonzero upper bound

Exact binomial endpoint: height = 1 − α^(1/n); the sampling assumptions still matter.



Horizontal: $\log_{10}$ trial count

Depth: tail probability α

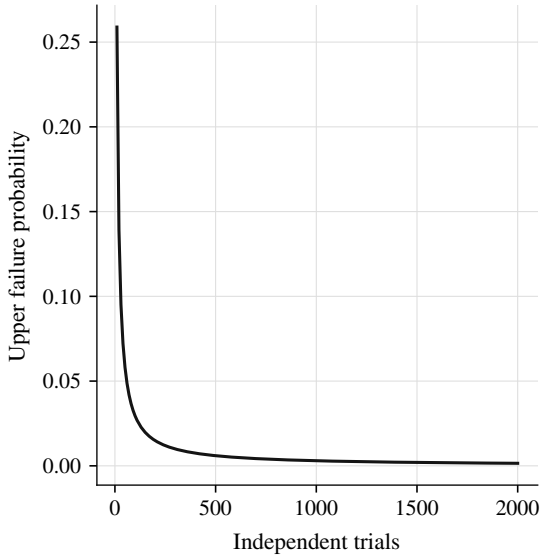
Height: upper failure probability

At α = 0.05 and n = 100:

0.0295130

Independent trials and a common failure probability are hypotheses. Replaying one input is a different test.

A No observed failures still leaves uncertainty



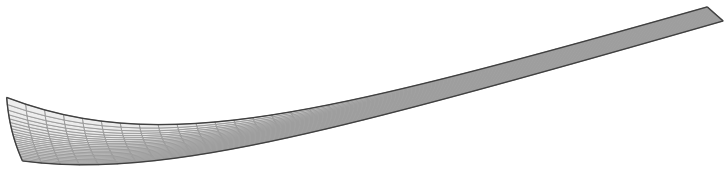
B More independent trials tighten the limit

n=10: upper 0.2588656
n=30: upper 0.0950339
n=100: upper 0.0295130
n=300: upper 0.0099361
n=1000: upper 0.0029912
n=2000: upper 0.0014967

Figure 21.3: *A clean test run still has a sampling limit.* The study continues on the following figure page.

Zero failures leaves a nonzero upper bound

Exact binomial endpoint: height = 1 − α^(1/n); the sampling assumptions still matter.



Horizontal: $\log_{10}$ trial count

Depth: tail probability α

Height: upper failure probability

At α = 0.05 and n = 100:

0.0295130

Independent trials and a common failure probability are hypotheses. Replaying one input is a different test.

C The binomial limit needs a sampling model

Independent trials

One common failure probability

A declared failure criterion

A sampled population relevant to use

D Replaying a result does not add new trials

Repeating identical deterministic input verifies repeatability.

It does not create independent population trials.

Both checks have useful, different roles.

Figure 21.3: *A clean test run still has a sampling limit (continued).* Exact zero-failure upper endpoints are plotted against the number of independent trials. The calculation is paired with the assumptions that make the binomial model applicable. Deterministic replay checks are kept separate from population error estimates.

From an assumption to an audit record

The audit keeps a measured result attached to its observations, units and operating region. Round-trip and cycle residuals test conversions; Jacobian bounds inspect local distortion; the chosen realization map connects the comparison spaces. A decision audit also needs threshold-band mass, class prevalence, coverage and confidence bounds.

The worked examples expose the consequences of omitting one of those ingredients. A small variance does not determine alert precision without prevalence. A quality filter can change the population on which a guarantee was established. A finite sample with no observed failures still leaves uncertainty about the population failure probability. Keeping those quantities in the record makes the guarantee replayable.

For the test part exactly on the limit, a small distance error can change the label. A known conversion proves agreement for that part, while the inspection population and its threshold-band mass govern the transfer claim. The same chain lets us read the application records and see what each body of evidence supports.

Practice with the construction

21.1. With equal class prevalence, equal conditional variance 1 and mean gap 4, calculate c_0, c_1 and the posterior lower bound.

21.2. Why can a quality filter worsen alert precision even if it retains fewer records? Construct a ten-alert example.

Chapter 22

Worked realizations and reproducible experiments

Chapter 1 followed a temperature reading through different records of one acquisition. We return to that question at the scale of a telemetry system. Each application specifies what it compares, which changes it allows and what its evidence establishes.

A query returns two copies of a telemetry record. One came through the ordinary ingestion path. The other was rewritten during a storage migration. Both have the same timestamp and source, but their vectors have different shapes. A comparison routine reports that the arrays differ. The engineer needs to establish whether the stored descriptions refer to one measurement and still support the same calculation.

This is the engineering problem behind Adaptive Multi-Vector Correlation, AMC. Developed from telemetry work at NASA, AMC chooses among embedding formats and keeps track of the translations between them. A record can be inexpensive to store in one format and useful to compare in another. The representation must be allowed to change while the relationships needed by the task remain available.

Keep the maps and the globe beside this example. A navigator can find the same harbor on two overlapping coastal charts, although it sits at different positions on the sheets. Changing charts has not moved the ship. The charts need translation rules and the right scale to describe the same route. Here the representations play the role of the maps; the underlying measurement model plays the role of the globe. The globe stands for the shared geometric object, whose shape we must specify in each example.

AMC's published evaluation uses public MODIS land-surface-temperature and vegetation-index data [14]. Its telemetry origins explain the engineering problem, and the MODIS study supplies a specified test of format switching. The monitoring and manifold model below asks how that architecture could be applied to comparisons of operating states.

First decide what is supposed to remain the same

There are two questions here, and a working system needs both. *Record identity* asks whether two stored objects descend from the same observation. *Task equivalence* asks whether those objects retain what a particular calculation needs. A shared identifier helps answer the first question. It cannot, by itself, answer the second.

A source identifier, timestamp, calibration version and transformation history can establish that a sparse vector and a dense vector came from the same measurement. If one transformation has discarded a small but important component, that history will still be true. The downstream calculation may nevertheless give a different answer. Conversely, two distinct measurements can be close enough for a declared tolerance without becoming the same event.

AMC makes these distinctions operational through its format graph, translation checks and UUID-linked lineage. Dual writes keep both versions available during a transition. For the mathematics, let m denote a measurement record and let

$$e_i = E_i(m) \in V_i, \qquad e_j = E_j(m) \in V_j$$

be two representations. A translator T_{ij} tries to make the two routes agree:

$$T_{ij}(E_i(m)) \approx E_j(m).$$

The right-hand side is the independently encoded target representation. That is what makes the comparison useful: the translator is being checked against another route to the same measurement.

A change of coordinates changes the metric's coefficients

Begin with a transformation simple enough to inspect completely. A temperature record is stored in Celsius in one channel and Fahrenheit in another. If x is the Celsius value and y the Fahrenheit value, then

$$y = T(x) = \frac{9}{5}x + 32.$$

The measurements 20 and 21 degrees Celsius become 68 and 69.8 degrees Fahrenheit. Subtracting the displayed numbers gives 1 in one channel and 1.8 in the other. The physical temperature difference is unchanged.

Choose Celsius degrees as the unit of length on this one-dimensional state space. In the x coordinate the metric is $ds^2 = dx^2$. Since $dx = (5/9)\,dy$, the same metric written in the other coordinate is

$$ds^2 = \left(\frac{5}{9}\right)^2 dy^2.$$

The two calculations now agree:

$$|21 - 20| = \frac{5}{9}\,|69.8 - 68| = 1.$$

The geometry is flat. The translator changes the coordinate, and the metric coefficients change to preserve the same length.

The two temperature scales are two maps of the same line. Multiplying Fahrenheit differences by 5/9 is the map's ruler doing its job: it makes the description agree with the measurement on the globe.

For a smooth, invertible coordinate change $y = T(x)$ in more dimensions, the rule is

$$G_x(x) = DT(x)^{\mathsf{T}}\, G_y(T(x))\, DT(x).$$

Take a small displacement v in the first coordinate system. Its translated displacement is $DT(x)v$. The displayed identity says that both descriptions assign the same squared length to that displacement. Integrating along a curve gives the same curve length; taking the infimum over the same family of curves gives the same intrinsic distance.

That exact construction is the reference case. A learned translator must be assessed against stated approximation bounds. A representation that discards information cannot simply be declared an invertible coordinate chart.

An exact representation change

Consider a small vector whose values have already been put in agreed units:

$$v = (0, 3.42, 0, -1.18).$$

A dense format stores all four entries. A sparse format can store the dimension and the two nonzero index–value pairs,

$$(2, 3.42), \qquad (4, -1.18).$$

Let E_{s} encode that sparse record and D_{s} decode it. If indices, dimension and numerical values are retained exactly, then

$$D_{\mathrm{s}}(E_{\mathrm{s}}(v)) = v.$$

The storage objects have different shapes. Their decoded vectors are equal. Any specified calculation on the decoded vector therefore has the same input.

This equality is an encoding statement. The geometric statement comes after a metric has been chosen. On a fixed-support patch, the active values can serve as local coordinates, and their metric follows the same change-of-coordinates rule. When support changes, that patch description must be reconsidered. A variable-length storage record does not become a single smooth chart merely because it is called a vector.

A lossy representation needs a budget

Now round each coordinate to the nearest multiple of a step Δ, assuming no clipping or overflow. Denote the result by $Q_\Delta(v)$. Every coordinate error is at most $\Delta/2$, so in n dimensions

$$\|Q_\Delta(v) - v\|_2 \le \frac{\sqrt{n}\,\Delta}{2}.$$

For two records v, w, the reverse triangle inequality gives

$$\begin{aligned} \big|\|Q_\Delta(v) - Q_\Delta(w)\|_2 - \|v - w\|_2\big| &\le \|Q_\Delta(v) - v\|_2 \\ &\quad + \|Q_\Delta(w) - w\|_2 \\ &\le \sqrt{n}\,\Delta. \end{aligned}$$

The bound explains what the format change is permitted to do to a distance. It is not a performance estimate taken from AMC's evaluation; it is a complete worked calculation for this particular quantizer.

Take $w = (0, 3.57, 0, -1.09)$ and $\Delta = 0.1$. Before rounding,

$$\|v - w\|_2 = \sqrt{0.15^2 + 0.09^2} \approx 0.17493.$$

After rounding, the nonzero entries are $(3.4, -1.2)$ and $(3.6, -1.1)$, giving

$$\|Q_\Delta(v) - Q_\Delta(w)\|_2 = \sqrt{0.2^2 + 0.1^2} \approx 0.22361.$$

Set the tolerance at 0.20. The engineer can reopen the same two readings and get a match from the original file and a rejection from the rounded copy. If that match matters, keep the original precision for comparisons this close to the cutoff.

More generally, suppose the distance error is at most δ. A decision of the form "distance at most t" is preserved whenever the original distance lies strictly below $t - \delta$ or strictly above $t + \delta$. The band between them is the part that needs closer examination. To turn this deterministic statement into a probability of disagreement, we must also know how much of the data distribution falls in that band.

Following more than one translation

AMC's graph makes the next question unavoidable. A record may pass from format i to j to k, although a direct route from i to k also exists. On a region where both routes are defined, compare

$$T_{jk} \circ T_{ij} \quad \text{and} \quad T_{ik}.$$

Exact coordinate transitions agree on their common domain. Approximate translators have a measurable composition defect. Moreover, the first error may be amplified by the next translator. If T_{jk} is L-Lipschitz and the two one-step discrepancies are bounded by ε_{ij} and ε_{jk}, then

$$\|T_{jk}(T_{ij}(E_i(m))) - E_k(m)\| \leq L\varepsilon_{ij} + \varepsilon_{jk}.$$

The bound requires the relevant points and intermediate images to stay inside the region on which those assumptions hold. The route and its conditioning belong in the calculation.

This is where the two descriptions meet. The translator account follows the record through changes of representation. The Riemannian account asks which lengths, neighborhoods and paths those changes preserve. AMC supplies the engineering question; the manifold construction states precisely when the two accounts agree and how much disagreement an approximation can introduce.

From one preserved record to a comparison of states

The migration calculation held one measurement fixed. AMC also asks whether different telemetry windows describe an equivalent operating state. We now specify that task, then change the domain to test two additional requirements. HERALD asks how observation platforms affect a biological comparison. PRISM asks which payment records may represent one transfer. Each case applies the earlier mathematics to a different proposed invariant, with its own observation maps and evidence.

AMC: compare the operating state across telemetry windows

Problem and semantic structure

Imagine that two monitoring stacks recorded the same interval of a distributed service. One has a raw log and the other summarizes latency and queue depth. AMC (Adaptive Multi-Vector Correlation) supplies an architecture for moving between formats. This operating-state model poses a further question:

> Do two telemetry windows—possibly from different monitoring stacks or vendors—represent the same underlying system state, up to benign fluctuations?

The semantic sameness structure S_{AMC} includes:

- Latent entities u representing the state of a distributed system (latencies, error rates, queue depths, resource utilization).
- Benign evolution: small deviations and fluctuations that do not correspond to meaningful state changes (for example, routine load variation).
- Observation channels:
  - C_{logs}: structured logs and metrics from one monitoring stack.
  - C_{alts}: metrics and traces from an alternative stack.
  - C_{agg}: aggregated dashboards and alerts.

Realization: translator-first with induced geometry

Use AMC's translator-first architecture to model this operating-state comparison:

- Each telemetry source defines a feature space V_i based on its schema and embedding methods.
- Translators T_{ij} are trained between these spaces using overlapping periods where multiple stacks observe the same system.

- The translator metric space (X_T, d_T) is the primary representation for cross-telemetry comparison.

To build a geometric realization of T_{AMC}, the model needs these additional steps:

- Smooth chartability diagnostics identify in-regime regions where translators behave like smooth, well-conditioned coordinate changes.
- On these regions, fit a candidate manifold $(M_{\mathrm{AMC}}, g_{\mathrm{AMC}})$ and establish its realization map and bounds. The categorical construction organizes a chosen realization; it is not itself a fitting algorithm.
- A verified manifold can describe benign state trajectories and test directions associated with incidents.

Smooth chartability diagnostics and regimes

Three conditions can affect chartability in the monitoring model:

- **Schema changes.** When a vendor modifies metric names or aggregation logic, test whether translators trained on the older schema show increases in $\varepsilon_{\mathrm{rt}}$ and $\varepsilon_{\mathrm{cyc}}$.
- **Incident regimes.** Relationships learned in benign regimes can fail under high stress. Test whether the Jacobian magnifies incident-sensitive directions.
- **Sampling gaps.** If overlapping telemetry between stacks is limited, estimates of $\varepsilon_{\mathrm{loc}}$ and Jacobians become unreliable.

An audit can partition X_T into candidate operating regimes:

- $\mathcal{R}_{\mathrm{benign}}$: typical operating conditions, where the audit tests for small chartability constants.
- $\mathcal{R}_{\mathrm{incident}}$: incident conditions, where a separate geometry and audit may be needed.
- $\mathcal{R}_{\mathrm{unknown}}$: out-of-regime conditions (unseen configurations, schema breaks).

A proposed deployment policy can automate supported benign windows, alert on a recognized incident and route unknown windows to operators. Regime membership and confidence in that assignment are separate quantities.

Error budgets and interpretation

On the translator side, AMC's detector might have an error budget such as

$$E^T_{\mathrm{geom}} = 0.04, \quad E^T_{\mathrm{link}} = 0.06, \quad \xi_T = 0.03, \quad \zeta_T = 0.02$$

for benign regimes.

What these numbers mean. For this hypothetical telemetry audit, interpret 0.04, 0.06, 0.03 and 0.02 as upper bounds on four named failure events. If they cover the reference decision's errors, its total error is at most 0.15 by a union bound. Their overlap is allowed; they are not counts of distinct mistakes.

The total decision guarantee transfers through Chapter 20. To transfer a component separately, establish a corresponding event inclusion. For example, let B^T_c and B^M_c denote the translator and manifold failure events for component c. If

$$B^M_c \subseteq B^T_c \cup D_c, \qquad P(D_c) \leq \Delta_c,$$

then $P(B_c^M) \leq P(B_c^T) + \Delta_c$. Under these additional component-specific hypotheses, the displayed budgets give

$$\begin{aligned} E_{\text{geom}}^M &\leq 0.04 + \Delta_{\text{geom}}, \\ E_{\text{link}}^M &\leq 0.06 + \Delta_{\text{link}}, \\ \xi_M &\leq 0.03 + \Delta_\xi, \\ \zeta_M &\leq 0.02 + \Delta_\zeta. \end{aligned}$$

Fitting a manifold alone supplies none of those event inclusions. The audit has to identify the changed failure mechanism and justify its allowance. This retains the useful diagnostic vocabulary while making every claimed component bound inspectable.

What this framework enables

Two tasks in the proposed monitoring audit are:

- Build translators and evaluate their agreement empirically.
- Fit a candidate manifold to the relevant telemetry features.

The comparison bounds connect those tasks and govern what a decision can reuse:

- **Regime-aware guarantees.** Chartability diagnostics define regimes where manifold-based reasoning is justified and where it is not.
- **Transfer from a measured reference.** Use the measured translator-side error on the declared population together with a bound on comparison failure and the probability mass near the threshold.
- **Upgrade paths.** When telemetry stacks change, morphisms in $\mathsf{SamTrans}(S_{\text{AMC}})$ plus chartability checks indicate whether existing guarantees carry over.

Domain-specific and general lessons

Domain-specific aspects:

- Telemetry schemas and instrumentation evolve quickly.
- Ground truth labels for "same state" are approximate (e.g., based on incident tickets).
- Incident regimes may have dramatically different geometry than benign regimes.

General lessons:

- Translator-first systems can still support meaningful manifold models when chartability holds locally.
- Reverse error budget transfer (translator $\to$ manifold) is valuable when manifolds are used primarily for analysis and visualization.
- Regime partitioning (benign/incident/unknown) is a natural way to make chartability assumptions explicit.

What to notice

AMC illustrates:

- The *translator-first* route to geometry: start with heterogeneous data, then ask whether a manifold emerges.

- How the same error budget decomposition works even when ground truth is noisy.
- How different regimes can have different chartability properties and error budgets.

In AMC, changes in telemetry can come from the system or from the representation. The next case poses that distinction for biological measurements reported by several platforms. A difference in two reports could reflect the objects being compared or the methods that produced the reports.

HERALD: High-resolution Early Recognition of Antigenic Landscape Divergence

Problem and semantic structure

HERALD (High-resolution Early Recognition of Antigenic Landscape Divergence) operates in viral surveillance. The latent entities are viral strains, and the sameness question is antigenic:

> Do two viral strains provoke essentially the same immune response, up to benign antigenic drift, or has a meaningful antigenic shift occurred?

The semantic sameness structure S_{HERALD} includes:

- Latent entities u representing viral strains.
- Benign evolution: genetic and antigenic changes that do not meaningfully change immune recognition (for example, within-variant drift).
- Observation channels:
  - C_{seq}: nucleotide or amino acid sequences.
  - C_{assay}: neutralization assay results (for example, titers across panels of sera).
  - C_{meta}: epidemiological metadata (location, date, lineage annotations).

Antigenic sameness is not sequence equality: two sequences can differ at multiple positions yet be antigenically similar.

Realization: geometry-first antigenic landscapes

HERALD's source manuscript develops the geometric comparison under explicit assumptions [16]. The model below applies that perspective to the book's representation question:

- It posits an *antigenic manifold* $(M_{\text{Ag}}, g_{\text{Ag}})$ whose points represent antigenic states.
- Geodesic distance d_M corresponds to antigenic distance (for example, log-fold changes in neutralization titers).
- Charts are induced by mapping assay profiles and sequence-based features into $\mathbb{R}^d$ via encoders and embedding methods.

Translator systems integrate heterogeneous platforms:

- Different labs and consortia produce feature spaces V_i (different assay platforms, sera panels, embeddings).
- Translators T_{ij} align these spaces where overlaps exist.
- A translator-based realization T_{HERALD} and metric quotient (X_T, d_T) integrate this heterogeneous data.

Smooth chartability and realizability

An audit of this multi-platform HERALD model would need the following evidence:

- **Platform overlaps.** Overlapping sets of strains tested on platforms A and B provide pairs for training and auditing T_{AB} and T_{BA}.
- **Round-trip and cocycle.** For such overlaps, measure round-trip errors and cocycle residuals, reporting sampled estimates of $\varepsilon_{\mathrm{rt}}$ and $\varepsilon_{\mathrm{cyc}}$.
- **Local Euclidean patches.** Within clusters of closely related strains, local linear models are fit to assay and sequence embeddings to justify approximate Euclidean structure (small $\varepsilon_{\mathrm{loc}}$).

Record the available overlaps and residual estimates in a platform compatibility matrix. Its entries should identify the data, maps and audit population, so that a proposed cross-platform comparison can be checked against its actual support.

Well-sampled regions can support a more informative audit, but the values of $\varepsilon_{\mathrm{loc}}, \varepsilon_{\mathrm{rt}}, \varepsilon_{\mathrm{cyc}}$ and the quality of M_{Ag} must still be established. Sparse records can leave a condition unverified rather than proving it false. The declared operating region should reflect that distinction.

Error budgets and interpretation

On the manifold side, HERALD may declare two strains antigenically "the same" if d_M is below a threshold corresponding to, say, no more than a two-fold change in neutralization titers across a reference panel. An indicative manifold-side error budget might be

$$E^M_{\mathrm{geom}} = 0.05, \quad E^M_{\mathrm{link}} = 0.07, \quad \xi_M = 0.03, \quad \zeta_M = 0.05.$$

What these numbers mean. Treat the four displayed numbers as stipulated upper bounds on named failure-event probabilities on one declared population. Geometry mismatch has allowance 0.05, linkage failure 0.07, calibration failure 0.03, and an explicitly defined abstention-policy violation 0.05.

If these events cover the decision-error event, the union bound gives total error at most 0.20. It does not predict exactly twenty errors in a hundred observations or assign disjoint causal counts. The same observation can belong to several failure events. Whether that upper bound is useful depends on the task and its costs.

For a new translator-based decision, the transfer theorem adds its separately established comparison-failure and threshold-band probabilities on that same population. Neither allowance can be inferred merely from the size of a chart residual. This is a conditional calculation, not a new measured HERALD result.

What this framework enables

The framework connects two kinds of work:

- Antigenic cartography and manifold-based visualizations, rooted in immunology and assay data.
- Practical pipelines for integrating data from new platforms, often using ad hoc similarity heuristics.

The framework enables:

- **Platform comparison.** Specify the maps and data supporting a new platform, then establish a distance comparison and the decision consequences on the intended population.
- **Explicit out-of-regime marking.** Chartability diagnostics provide a falsifiable criterion for declaring certain regions of sequence space outside the safe region for manifold-based reasoning.

- **Separation of concerns.** The domain model specifies M_{Ag} and d_M. The engineering audit specifies translator distortion and the population and margin conditions needed for transfer.

Domain-specific and general lessons

Domain-specific aspects:

- Assay platforms have complex noise patterns and idiosyncrasies.
- Antigenic similarity must respect immunological constraints (for example, cross-reactivity patterns).
- Data is sparse and expensive to obtain; out-of-regime regions are significant.

General lessons:

- Geometry-first frameworks are natural when the scientific quantity of interest *is* a distance.
- Smooth chartability can fail sharply; it is important to know where, not just whether, it holds.
- Translator systems can relate measurement platforms when semantic correspondence, geometric comparison and their domains are established.

What to notice

HERALD shows the framework under harsher conditions:

- In the stipulated budget, the linkage allowance is the largest; an empirical claim would need the corresponding records.
- Out-of-regime marking is not optional; it is central.
- Transfer depends on the actual distance and probability bounds; broader uncertainty can reduce the region in which a decision is supported.

Payment reconciliation starts with records on different payment rails. Two transactions can have the same amount and still be separate events. The system must specify which records may refer to the same transfer before it can use geometry to rank candidate matches.

PRISM: Payment Rail Identity Semantic Mapper

A ledger example that cannot hide behind similarity

Consider a constructed ledger for an authorization of 100 USD. The reconciliation program looks for its settlement among records with nearby timestamps and amounts. Another authorization for the same amount could produce a settlement in the same window, so a nearby amount and time alone cannot identify the event.

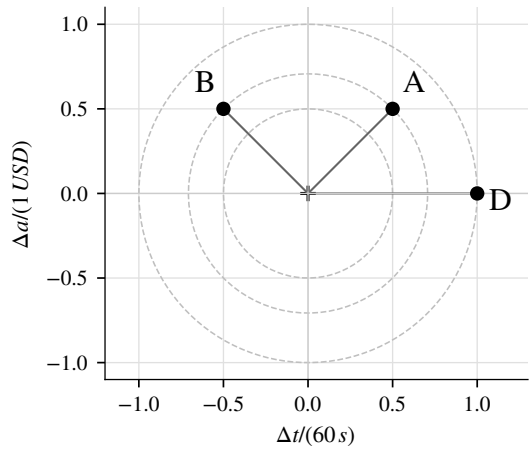
Define a permitted candidate set using currency, account and event type. Within that set, compare time and amount with stated scales, for example

$$d^2 = \left(\frac{\Delta t}{60\,\mathrm{s}}\right)^2 + \left(\frac{\Delta a}{1\,\mathrm{USD}}\right)^2.$$

The formula is dimensionless. It is an illustrative local metric on the continuous fields, not a claim that every ledger is one smooth manifold. Discrete constraints determine which records can enter the comparison. One-to-one assignment across several candidates is an additional combinatorial problem.

In the drawing, A and B are both admissible and equally distant from the reference. C is excluded by its currency even though its raw numbers match. D is admissible but farther away. The tie calls for additional evidence or an unresolved result. A test fixture should also include duplicate amounts, missing identifiers and unmatched records.

A *Two admissible records can tie*



B *The currency constraint comes first*

Candidate	*Currency*	*Score*
A: +30 s, +$0.50	USD	0.7071
B: −30 s, +$0.50	USD	0.7071
C: 0 s, +€0.00	EUR	Rejected
D: +60 s, +$0.00	USD	1.0000

A and B tie. The score cannot recover a missing transaction identifier.

Figure 22.1: A constructed reconciliation fixture with dimensionless time and amount coordinates. Currency compatibility determines admission before the metric is applied. Two admitted records tie at $\sqrt{1/2}$; the score preserves this ambiguity rather than inventing an event identifier.

Problem and semantic structure

PRISM's reconciliation implementation retrieves candidates, scores pairs and checks currency before accepting a match [15]. Its geometric model below starts with economic transfers (for example, "Alice pays Bob $100") and asks:

> Do two transaction records—potentially from different payment rails, currencies, and messaging formats—represent the same underlying economic transfer, up to benign processing differences?

The semantic sameness structure S_{PRISM} includes:

- Latent entities u representing economic transfers between parties.
- Benign evolution: processing artifacts that do not change the underlying transfer (for example, intermediate correspondent banks, agreed currency conversions, message splitting for regulatory reasons).
- Observation channels:

- C_{SWIFT}: SWIFT-style messages.
- C_{RTGS}: real-time gross settlement entries.
- C_{cards}: card network transaction records.
- C_{internal}: internal ledger entries in different systems.

Realization: translator-first with sparse geometry

A translator-first model for this setting would specify:

- Each rail defines a feature space V_i via schemas and embeddings (counterparties, amounts, times, free-text fields).
- Translators T_{ij} fitted on matched transaction tuples believed to represent the same transfer.
- A translator metric space (X_T, d_T) for comparison of admitted candidates.

Suppose a Riemannian model $(M_{\text{PRISM}}, g_{\text{PRISM}})$ can be justified on a subset $\mathcal{R}_T^{\text{simple}} \subset X_T$. An audit would distinguish:

- Relatively simple corridors, such as standard SWIFT-to-RTGS flows, where records can test whether a local manifold approximation is supported.
- Complex or rare patterns, such as multi-step correspondent chains, where a translator metric may be the only supported comparison.

Smooth chartability and partial realizability

The modeled chartability claim is local:

- **Simple corridors.** Matched examples and stable schemas give an opportunity to test $\varepsilon_{\text{loc}}, \varepsilon_{\text{rt}}, \varepsilon_{\text{cyc}}$.
- **Rare patterns.** Scarce overlaps can leave translator behavior and the manifold comparison unverified.

Any resulting manifold guarantee applies on the region $\mathcal{R}_T^{\text{simple}}$ for which the realization and its bounds have been established.

Error budgets and interpretation

In simple corridors, PRISM's translator-side matching engine might have a budget like

$$E_{\text{geom}}^T = 0.01, \quad E_{\text{link}}^T = 0.02, \quad \xi_T = 0.01, \quad \zeta_T = 0.03.$$

What these numbers mean. For this hypothetical payment comparison, the displayed numbers are upper bounds on geometry, linkage, calibration and abstention-policy failure events. If those events cover the reference rule's errors, the total error is at most $0.01 + 0.02 + 0.01 + 0.03 = 0.07$. It is not an observed seven-percent rate, and the four categories need not be disjoint.

A local manifold and distance comparison support the total decision-transfer argument only with its population and margin assumptions. If the separate component event inclusions just

derived for AMC are also established here, then

$$\begin{aligned} E^M_{\text{geom}} &\le 0.01 + \Delta_{\text{geom}}, \\ E^M_{\text{link}} &\le 0.02 + \Delta_{\text{link}}, \\ \xi_M &\le 0.01 + \Delta_{\xi}, \\ \zeta_M &\le 0.03 + \Delta_{\zeta}. \end{aligned}$$

Each Δ_c now has a specific obligation: bound the event that creates a new component failure after changing representation. An overall disagreement count cannot determine these four allowances by itself.

What this framework enables

Payment matching can begin with two kinds of inputs:

- Heuristic field-level similarity scores (amount within tolerance, fuzzy name matching, time window overlap).
- Rules about which record pairs are eligible for a match.

The framework brings:

- **A semantic core.** Latent economic transfers and benign evolutions provide a language for what matches *mean*, beyond fields.
- **Local geometric windows.** In simple corridors, manifold models support principled analysis of how far benign processing differences can move a transaction.
- **Hybrid automation.** Error budgets and chartability diagnostics support fully automated matching in well-behaved regions and human-in-the-loop workflows in complex regions.

Domain-specific and general lessons

Domain-specific aspects:

- Strong regulatory constraints on error rates, with asymmetric costs for false positives and negatives.
- Highly heterogeneous schemas with many optional and free-text fields.
- Skewed distributions, with a small number of very common patterns and a long tail of rare ones.

General lessons:

- Global smooth chartability is not required; local, patchy chartability is still valuable.
- The framework naturally supports hybrid automation: automatic decisions where chartability holds, human review where it does not.
- Abstention budgets play a central role in high-stakes domains and must be tracked explicitly.

What to notice

PRISM shows:

- How the framework behaves when chartability is fundamentally local.
- That even partial manifold models can be useful when backed by translator-side guarantees.
- How error budgets guide where automation is safe and where human oversight is needed.

Follow one observation through the whole argument

The application chapters now bring the mathematics back to records someone can inspect. AMC begins with telemetry descriptions and their conversion. ICARUS supplies an actual $SO(3)$ orientation model and its quaternion double cover. GIGI supplies exact encoding fixtures and measured-data fingerprints. Their evidence answers different questions, and the comparisons are stated at that level.

The shared connection is operational: specify the object, record the description, choose the transformation and test what must survive. A thermostat record, an aircraft attitude and an analytic fingerprint do not become the same kind of object. They instantiate the same discipline of comparison.

Antigenic comparison and payment reconciliation expose two further difficulties. A biological label can refer to a functional relation rather than sequence identity. A payment match can require discrete constraints, such as currency and account, before any continuous tolerance is meaningful. Numerical budgets introduced with chosen or hypothetical inputs are model calculations under those stated assumptions.

The cases have made their objects and allowed changes explicit. We can now compare their evidence on equal terms: which statements follow from a construction, which were measured in a simulation, and which require observations of a deployed system?

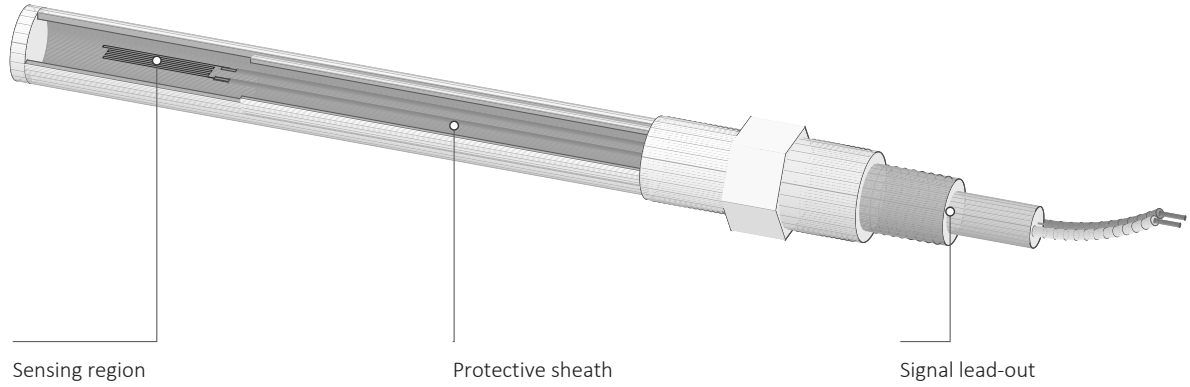
One application, several kinds of evidence

The same project can supply a theorem, a simulation and a field observation. They do not become interchangeable because they share a name. The sign invariance of quaternion attitude is an exact mathematical statement. A recorded ICARUS run compares particular controllers under specified simulated conditions. A flight test would add physical sensors, actuators and environmental effects. Each can be valuable while supporting a different conclusion.

The AMC record illustrates another separation. A lossless format conversion can be tested by exact round trip. A quantization step needs an error bound. A detection claim needs a population, labels and a decision protocol. A throughput measurement needs hardware, workload and timing conventions. A single word such as “validated” is too coarse to describe all four.

The record must stay attached to its origin

Illustrative probe anatomy / retained GIGI values are identified separately below.



PROBE	ACQUISITION	VALUE	UNIT
T-001	1710000000	22.4 °C	celsius

A **AMC:** conversion preserves the telemetry record

EVIDENCE	*WHAT IS CHECKED*
Object	One observed record
Exact check	Decode(encode(x)) = x
Separate check	Quantization error
Timing claim	Needs workload + hardware

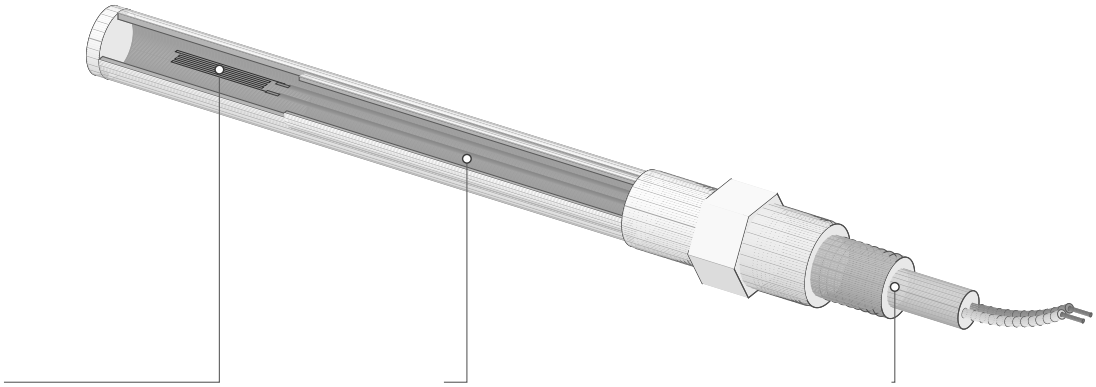
B **ICARUS:** two signs, one physical attitude

EVIDENCE	*WHAT IS CHECKED*
Object	Rigid-body orientation
Exact check	R(q) = R(−q)
Recorded evidence	Controller simulation
Still needed	Physical flight tests

Figure 22.2: *Match each application claim to the evidence that can support it.* The study continues on the following figure page.

The record must stay attached to its origin

Illustrative probe anatomy / retained GIGI values are identified separately below.



PROBE	ACQUISITION	VALUE	UNIT
T-001	1710000000	22.4 °C	celsius

C **GIGI:** fewer bytes, the same typed fields

EVIDENCE	*WHAT IS CHECKED*
Object	Three typed records
Stored sizes	292 bytes → 113 bytes
Exact check	Decoded fields equal originals
Scope	This documented fixture

D **GIGI:** withheld queries test the completion

EVIDENCE	*WHAT IS CHECKED*
Input	17 binned distributions
Audit	408 withheld queries
Provenance	Source rows + bin rules
Meaning	Feasible sets, not confidence

Figure 22.2: *Match each application claim to the evidence that can support it (continued).* AMC, ICARUS and GIGI examples are arranged by object, operation, preserved quantity and verification type. Exact identities, saved computations and empirical samples remain visibly distinct. The table is a guide to reading the application records, not a ranking of projects.

A paired experiment removes an avoidable ambiguity

Suppose two telemetry converters are timed on the same recorded input windows. For each window i, compute the difference D_i in processing time. Both converters face the same mix of short records and complex records. Timing them on different windows would mix the implementation difference with differences in input difficulty.

The paired design still needs a prespecified outcome. "Better" might mean lower decision error, lower latency, fewer abstentions or smaller distortion. A method can improve one and worsen another. The result table should preserve those tradeoffs rather than collapsing them into an unexplained overall score.

A useful reproducibility record contains the source version, input identifiers, parameters, random seeds where applicable, output values and the comparison rule. It should also identify which quantities were computed from the full data before any values were withheld. That last detail is essential to understanding the GIGI inference experiment in Chapter 24.

What a failure teaches the geometry

When an example fails, inspect which declared invariance was broken. Was a quaternion sign treated as a physical change? Did two histograms use different bins? Did a converter alter units without transforming the metric? Did a score ignore the direction that separates configurations? These are geometric questions with concrete diagnostics.

Run the checks on failures as well as successes. A deliberately inconsistent input can show whether a system recognizes that its own model has no compatible answer. A same-size edit can show whether the output depends on the evidence rather than just the number of records. A restart can test whether the preserved state really carries the computation.

The common question across the applications

The applications share a question while placing different demands on the comparison. ICARUS preserves physical orientation through a quaternion sign change and evaluates controllers in declared simulations. AMC follows system state through telemetry conversions. HERALD's antigenic model needs matched observations and platform overlap. PRISM's transaction comparison must respect currency and admissible payment routes before a distance can rank candidate matches.

For each case, specify the object, channels, permitted changes and operating region. A finite audit supports the claim justified by its sampling and coverage. The realization map connects the two descriptions. Transferring a decision bound also requires evidence about comparison failures and the population near the threshold.

GIGI takes the same question down to individual stored fields, then follows a coarser description into geometric inference.

Practice with the construction

22.1. Two admissible candidate matches differ by $(\Delta t, \Delta a) = (30\,\mathrm{s}, 0.5\,\mathrm{USD})$ and $(60\,\mathrm{s}, 0)$. Compute their illustrative scores.

22.2. Give a pair that the formula ranks as close but the currency rule must reject. Explain why this is a domain restriction rather than a small metric correction.

Chapter 23

Preserving observations through computation

Take a small GIGI software test. Three temperature records arrive from sensors T-001, T-002 and T-003. The first two say normal and T-003 reports 45.8 °C with an alert. GIGI's DHOOM format packs the three records into a shorter file. Can the program decode the alert and attach it to the right sensor and time?

Imagine the monitoring program has cached a normal result for an earlier batch of three observations. The new batch also contains three, but the third now carries the alert. Reusing the answer because the count matches would discard the change the system exists to detect. What must the software retain before an old answer or an old comparison can be reused? We will follow that question through decoding, addresses, calibration, frames and the cached result.

We can answer that question without guessing what the shorter file looks like. Decode it. Check the sensor identifiers, the timestamps, the temperatures and the alert. If the third sensor's warning survives, the new description has preserved it. If the warning has become "normal," a plausible temperature alone will not rescue the translation.

GIGI brings the question of what survives a change of description into a working data system. A description is stored, transformed, compared and queried. At each step we can ask what the operation preserves. My *GIGI Builds* develops the storage and execution side of that program; *GIGI Thinks* develops the geometric structures and the conditions under which an answer can be supported. Here we test which records survive encoding, which comparisons survive a frame change, and which conclusions the retained constraints support.

The maps and the globe remain our guide. The map is the description the program can read; the globe is the common object or geometric model against which that description is checked. *GIGI Builds* shows how descriptions survive storage and transmission. *GIGI Thinks* asks what their geometric comparisons require. Table 23.1 puts those connections beside AMC, ICARUS and the measured-data study.

One record, two encodings

Table 23.2 gives the three records used in the DHOOM round-trip example in *GIGI Builds*. This is a deliberately small software fixture, chosen so that every field can be checked by eye. It is not a field experiment.

DHOOM stores repeated structure once and records the departures. In this example, the identifier advances from T-001 to T-002 to T-003, and the timestamp advances by sixty seconds. The unit is always Celsius. The usual status is normal. The temperatures still have to be supplied, and the third record must override the usual status.

The live encoder produces the following header, wrapped here only to fit the page:

```
readings{sensor_id@T-001, timestamp@1710000000+60,
value, unit|celsius, status|normal}:
```

The three record lines are

```
22.4
```

Table 23.1: The maps and the globe: shared questions across the working examples.

Example	The map: how it is described	The globe: what must agree
AMC Telemetry	One measurement in different units or storage formats.	The measurement's identity and the distances needed by the task; approximate translations carry an error budget.
ICARUS Orientation	Quaternion descriptions q and $-q$ label one physical attitude.	The same rotation in $SO(3)$ and the same orientation distance. The double cover states exactly how the labels meet.
GIGI Builds Records	JSON and DHOOM encode the three sensor readings differently.	The decoded identifiers, times, values and alert agree. The shorter description preserves every field in this fixture.
GIGI Thinks Frames	Local coordinates and frame labels specify how a quantity is written.	Transition maps and the transformed metric preserve the modeled vector and its length. A shared name alone cannot do that.
GIGI Fingerprints	Bin counts become coefficients of an analytic polynomial.	With common bin edges, equal polynomials mean equal bin proportions. Individual readings within a bin are not retained.
GIGI Completion	Seven known function values supply a partial description.	One bounded analytic function must fit them jointly. The query disk shows values still possible; a unique function need not be determined.

Each row states its own test of agreement. Exact encoding, a smooth coordinate change, a double cover, a lossy summary and constrained inference preserve different things. The calculations that follow make those distinctions explicit.

Table 23.2: Three sensor records preserved by the live encoding round trip.

Sensor	Timestamp	Temperature (°C)	Status
T-001	1710000000	22.4	normal
T-002	1710000060	23.1	normal
T-003	1710000120	45.8	alert

All three records also retain the unit field, celsius. Timestamps are the fixture's integer time values.

```
23.1
45.8, , :alert
```

The empty slot in the last line retains the default unit; the explicit alert overrides the default status. These marks have meaning because the header supplies a decoding rule. Without that rule, omitted text would simply be missing information.

Let $\mathcal{E}$ be the encoder and $\mathcal{D}$ the decoder. For this fixture X, the test establishes

$$\mathcal{D}(\mathcal{E}(X)) = X.$$

The compact JSON object containing the three records occupies 292 bytes; the DHOOM document occupies 113 bytes, including its header and newline framing. The bytes differ. The decoded records agree, field for field. The measured size ratio is about $2.58 : 1$ for this example; another dataset has to pay for its own header and its own exceptions.

The equality is the first useful kind of sameness. It tells us that any deterministic comparison made from the decoded fields receives the same inputs. If d compares two decoded records, we can

define the corresponding wire comparison by

$$d_{\text{wire}}(\mathcal{E}(x), \mathcal{E}(y)) = d(\mathcal{D}(\mathcal{E}(x)), \mathcal{D}(\mathcal{E}(y))).$$

Where the round-trip identity holds, this equals $d(x, y)$. A byte-by-byte distance between the two documents would answer a different question. The translator route preserves the geometric calculation by recovering its inputs; shorter storage alone does not establish that preservation.

A format change that preserves every tested field

The useful engineering question in the three-record GIGI example is exact: does decoding restore the original records, including their warning status? The compression ratio is secondary to that identity. A shorter representation that loses the third sensor's alert would fail the comparison even if it saved more bytes.

Let E encode the three sensor records and D decode them. We checked $D(E(X)) = X$ for those records X. A deterministic comparison that uses only the restored fields gets the same inputs either way. To say that for every valid file, we'd need a format specification and checks that cover its full grammar, or a proof that the encoder and decoder preserve every valid record.

A *All typed fields survive decoding*

Sensor	*Timestamp*	*Temperature*	*Status*
T-001	1710000000	22.4 °C	normal
T-002	1710000060	23.1 °C	normal
T-003	1710000120	45.8 °C	alert

B *The record contains regularity and exceptions*

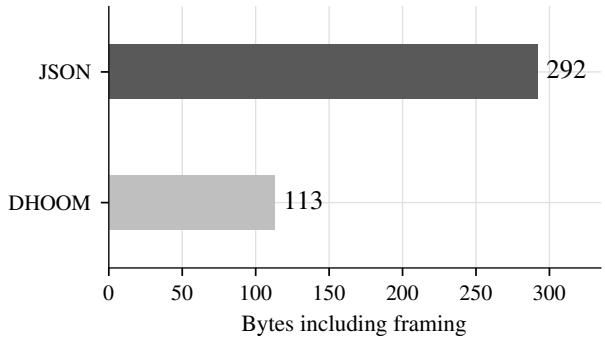
Identifier advances by one.

Timestamp advances by 60 s.

Unit defaults to Celsius.

Third status overrides normal.

C *This fixture uses fewer bytes*



$\mathcal{D}(\mathcal{E}(X)) = X$: identifiers, times, values, units and alert all survive.

Figure 23.1: The three-record DHOOM fixture: its decoded fields, reusable structure, explicit exception and measured storage sizes. The compact representation retains the third sensor's alert. These are the verified fixture sizes, not a universal compression ratio.

What the bundle language adds

The decoded alert is useful only if it remains attached to the right sensor and acquisition event. A sensor reading has an address and a value. For these records, the address can be the pair consisting of sensor identifier and timestamp. Over each address sits the set of values the schema permits: a temperature, a status and a unit. The bundle description makes that separation explicit.

Write B for the address set and F_p for the allowed values at address p. The total space and projection are

$$E = \{(p, v) : p \in B,\ v \in F_p\}, \qquad \pi(p, v) = p.$$

The stored observations define a partial section σ on the set $A \subseteq B$ of addresses actually observed:

$$\sigma : A \longrightarrow E, \qquad \pi \circ \sigma = \mathrm{id}_A \,.$$

An individual row is the value of that section at one address. The equation says that a reading remains attached to the address under which it was filed. It is the difference between a temperature from T-003 at its recorded time and an unattached number that happens to be 45.8.

The DHOOM header supplies the default status “normal” for a record that exists. A sensor that sent no record has no entry in A. A present address with a default status belongs to A and has a value under σ. The file format must let the monitoring program tell those cases apart.

The storage address set can be discrete. A smooth Riemannian manifold enters when we specify a geometric model for the quantities being compared. Merely naming keys, fibers and sections does not supply a metric tensor or make a categorical status into a tangent vector. Temperature and wind components admit numerical models; a status flag may instead select a separate comparison rule. The model must say how each participates.

The comparison has to survive a change of frame

Exact fields are only the first dependency of a reusable comparison. A vector-valued sensor record also needs its frame. Imagine two instruments measuring the same horizontal wind at the same location. The first instrument’s axes point east and north. The second has been turned through a quarter-turn. The first reports $(10, 0)$ metres per second; the second reports $(0, -10)$ in its own axes. Subtracting the stored pairs produces an apparent disagreement of $\sqrt{200}$ metres per second. Rotating the second pair into the first frame produces $(10, 0)$. The physical disagreement is zero.

This is a calculated coordinate example, not a wind measurement from the GIGI experiments. Its purpose is to expose the information an implemented comparison needs. Let x denote the first coordinates and $y = \phi(x)$ the second. The metric tensors must satisfy

$$g_x = (D\phi_x)^{\mathsf{T}}\, g_y(\phi(x))\, D\phi_x.$$

For a rotation, both coordinate metrics are the Euclidean identity matrix, because the rotation is orthogonal. For a general change of coordinates, their matrix entries need not agree. The equality above is what preserves the length of every tangent displacement.

This is the harbor-chart problem with wind instruments: the arrows are written in different frames, but the wind on the globe is the same. The coordinate transition and the metric together let us compare it.

Integrating those lengths along corresponding curves preserves their lengths. If the coordinate change also takes the admissible paths in one description onto the admissible paths in the other, their distance calculations agree. This is a complete route from translator to geometry: an explicit map, a transformed metric, and the same path problem. If the instruments occupy different

locations on a curved model, comparing their tangent vectors additionally requires a specified transport rule and path. A common field name cannot supply those choices.

GIGI Thinks develops this relationship through charts, transitions and transport. For the present book the central requirement is concrete: retain enough information about the observation and its frame that a change of description cannot masquerade as a change of the thing observed.

A score with a stated ruler

Suppose the program now compares an instrument's reported orientation with a reference orientation. The principal angle $r \in [0, \pi]$ is the angle of the relative rotation between them. Each orientation has two unit-quaternion representatives, q and $-q$. The cached score must agree for either representative. Use

$$S(r) = \cos^2(r/2), \qquad d_{\mathrm{ch}} = \sin(r/2), \qquad S + d_{\mathrm{ch}}^2 = 1.$$

The half-angle score also appears in *GIGI Thinks*. Here r is an angle in radians, and d_{ch} is half the operator norm of the difference of the two rotation matrices, as in Chapter 15. The chord between nearest unit-quaternion representatives has a different length, $2\sin(r/4)$. Changing either representative's sign leaves the physical orientation and this score unchanged.

Suppose the geometric angle is known only through an estimate $\widehat{r}$ with a verified error bound $|\widehat{r} - r| \le \varepsilon$. Since

$$S'(r) = -\tfrac{1}{2}\sin r,$$

the mean-value theorem gives

$$|S(\widehat{r}) - S(r)| \le \tfrac{1}{2}\varepsilon.$$

That bound turns a tolerance on the measurement into a tolerance on the decision. Table 23.3 uses $\varepsilon = 0.02$ radians and a declared acceptance threshold $S \ge 0.75$.

Table 23.3: A bounded angle error becomes a bounded score error.

$\widehat{r}$ (rad)	$S(\widehat{r})$	Allowed score	Decision
1.00	0.770151	[0.760151, 0.780152]	Accept
1.05	0.748786	[0.738785, 0.758786]	Undetermined
1.10	0.726798	[0.716798, 0.736799]	Reject

Calculated example: $|\widehat{r} - r| \le 0.02$ radians; accept when $S \ge 0.75$. The 0.01 score-error bounds are rounded outward. These are not measured confidence levels.

Take the two orientations that look 1.05 radians apart. If the dashboard rounds the score to 0.75, it looks like a match. But the allowed scores run from about 0.739 to 0.759. I'd leave that comparison undecided or get a more precise angle reading. That range uses the angle allowance we chose for this example.

This is the evolution relevant to *The Geometry of Sameness*. The question has moved from whether two descriptions can agree to what a working operation needs in order to preserve their agreement: an address, a decoding rule, a frame, a metric, and a bound when exact agreement is unavailable. The implementation study below follows those requirements into the saved state and cached answers of a program.

Which ruler belongs to a field?

A timestamp, temperature and status flag do not share one obvious numerical ruler. A flag is categorical. Temperature has units. Time identifies both an observation address and, in another model, a continuous coordinate in a trajectory. A good schema can store all three, but the schema does not by itself determine their Riemannian geometry.

The map/globe analogy therefore asks a concrete question before each construction: what common object are these rows describing? For the lossless encoding, it is the full record. For the wind-frame calculation, it is a vector at a stated location. For the angle score, it is an orientation comparison with a specified normalization. The shared-connection table keeps those objects visible instead of letting the same notation blur them together.

To reuse a saved comparison, the program needs the same decoded values, addresses, calibration and score definition. A vector comparison also depends on the sensor axes. If a wind instrument turns during a recording, the frame at each time matters.

A record, a section and a vector are not interchangeable words

A sensor record begins as a finite collection of fields: identifier, time, value and status. A section of a bundle is a mathematical assignment that selects a fiber element over each base point in a domain. A vector is an element of a specified vector space. A software representation can implement relationships among all three, but those relationships have to be constructed.

For example, let time be the base and let a temperature value lie in a one-dimensional fiber over each acquisition time. A time series becomes a section over the sampled times. If a calibration assigns a new numerical coordinate in each fiber, the section values transform by that calibration. If the calibration changes with time, comparing raw values at different times without accounting for it can manufacture a trend.

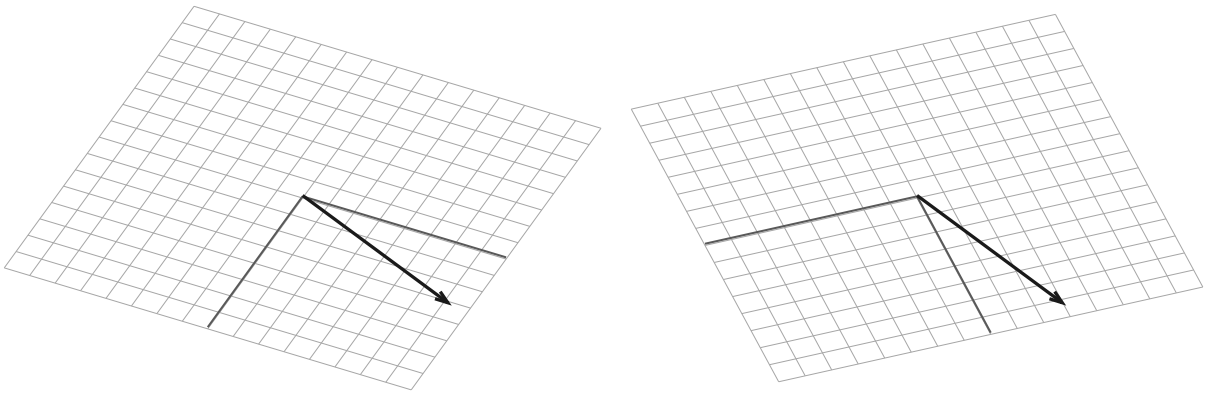
Now let the instrument rotate on its mount. Write $x(t)$ for the wind vector in fixed room axes and let the columns of $E(t)$ be the instrument's orthonormal axes in that room. The instrument reports $c(t) = E(t)^T x(t)$. Differentiating gives

$$\dot{c} = E^T \dot{x} - \Omega c, \qquad \Omega = E^T \dot{E},$$

where Ω is skew-symmetric. A stationary physical vector can have changing components solely because the frame rotates. The connection term records that contribution.

A fixed vector meets a moving coordinate frame

The two drawings represent the same room vector x = (0.9, 0.4).



c = (0.9000, 0.4000)

Frame rotation: 0.0 rad

c = (0.9140, -0.3669)

Frame rotation: 0.8 rad

A One fixed vector has two coordinate reports

B The reports change as the sensor frame turns

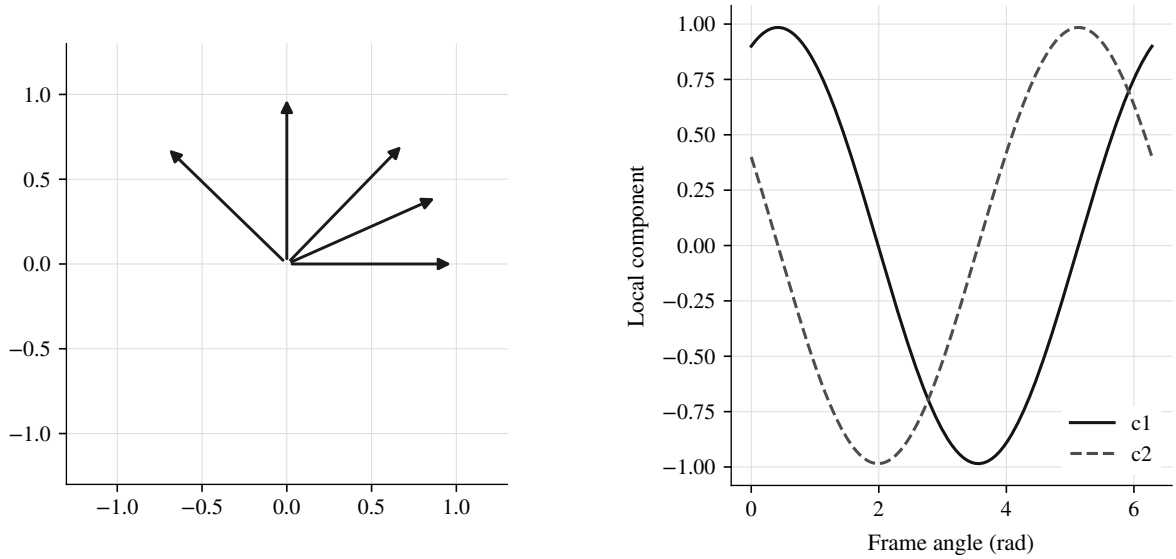
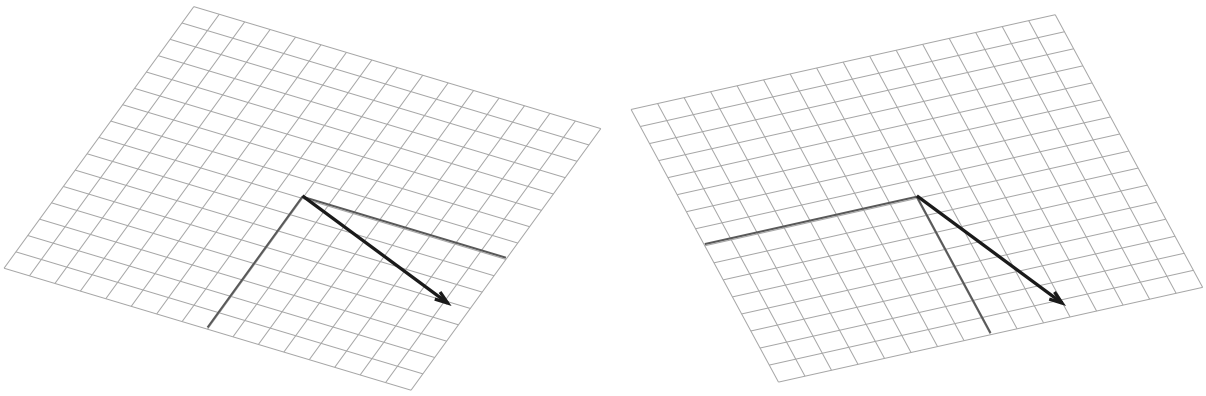


Figure 23.2: *A changing record can describe a stationary vector.* The study continues on the following figure page.

A fixed vector meets a moving coordinate frame

The two drawings represent the same room vector x = (0.9, 0.4).



c = (0.9000, 0.4000)

Frame rotation: 0.0 rad

c = (0.9140, -0.3669)

Frame rotation: 0.8 rad

C The frame term explains the component change

$c = E^T x$

$c' = E^T x' - \Omega c$

$\Omega = E^T E'$ is skew-symmetric.

Even $x' = 0$ can give $c' \neq 0$.

D A measurement needs its frame and timestamp

Physical quantity and units

Frame identifier

Acquisition time

Calibration version

The transformation used for comparison

Figure 23.2: *A changing record can describe a stationary vector (continued).* A fixed room vector is expressed in a rotating sensor frame. Component traces and the frame-derivative term are calculated together. The diagram separates physical motion from a change in the local descriptive frame.

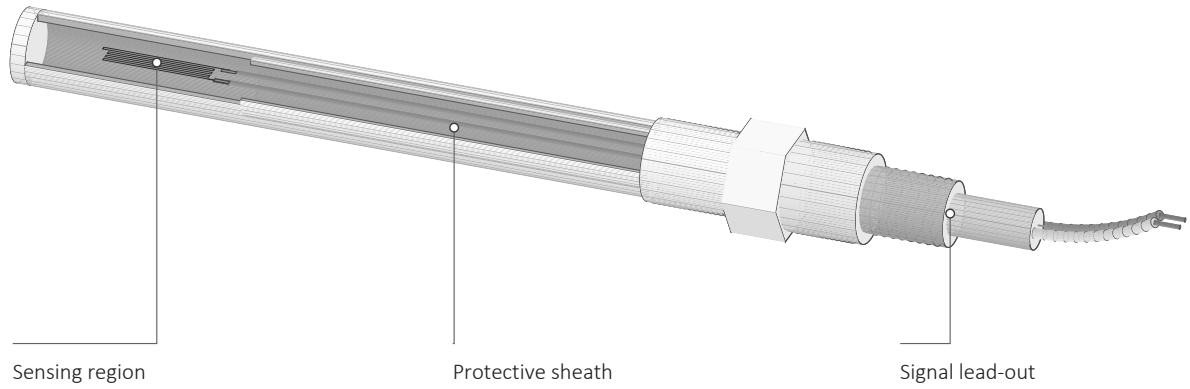
The lossless round trip has a narrow, valuable conclusion

The three-record GIGI example has an exact decoding check. That establishes equality of the decoded record fields under the specified encoder and decoder. It does not establish that every compressed format preserves arbitrary downstream geometry. A geometric score can still be changed by quantization, omitted metadata or a different calibration convention.

The relevant sequence is therefore: identify the record, encode it, decode or query it, apply the comparison in the correct units and frame, then inspect the resulting invariant or error bound. The GIGI books supply both implementation narratives and geometric constructions [21, 22]. Following one actual record through the sequence keeps those contributions connected.

The record must stay attached to its origin

Illustrative probe anatomy / retained GIGI values are identified separately below.



PROBE	ACQUISITION	VALUE	UNIT
T-001	1710000000	22.4 °C	celsius

A Three typed records supply the example

PROBE	*READING*	*STATUS*
T001	22.4 °C	normal
T002	23.1 °C	normal
T003	45.8 °C	alert

Each record also retains its acquisition time.

B The compact record decodes to the same fields

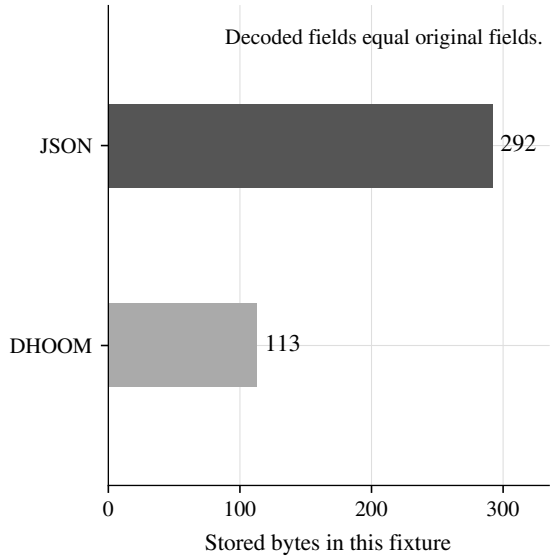
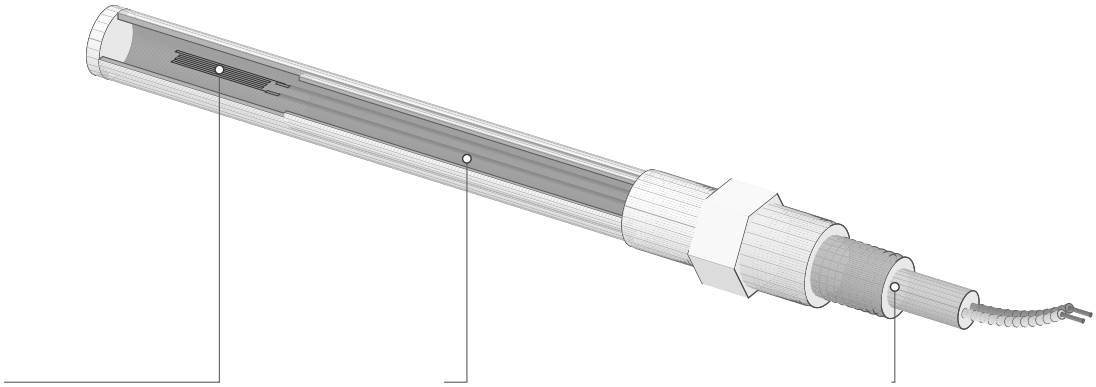


Figure 23.3: *Exact records and geometric comparisons meet at a contract.* The study continues on the following figure page.

The record must stay attached to its origin

Illustrative probe anatomy / retained GIGI values are identified separately below.



PROBE	ACQUISITION	VALUE	UNIT
T-001	1710000000	22.4 °C	celsius

C Geometric comparison needs a separate ruler

D Controls test whether the result uses the data

Choose the metric and calibration.	Same-size edit: outputs must use values.
Transform vector components with the frame.	Restart: persisted state must reproduce results.
Carry approximation error into the score.	Record count alone is an insufficient cache key.

Figure 23.3: *Exact records and geometric comparisons meet at a contract (continued).* The documented three-record example is followed through its lossless representation and into a declared comparison. The drawing labels which equalities follow from round-trip decoding and which require separate metric or calibration checks.

The restored record must also remain the input to the result we return. Reusing a result calculated from different evidence would break that connection even if encoding and decoding were exact. The cache tests below check the last link between the observation and its reported computation.

A cache is valid only for the state it represents

At the chapter opening, the monitoring program had saved a normal result for three readings. The next batch still has three readings, but T-003 now carries an alert. Using only the record count to look up a result would retrieve the stale answer. A same-size edit tests for that defect. The cache key must account for the values or a faithful content identifier, model version, calibration, query and any other state the result uses.

A restart test asks a different question. Does the persisted representation contain enough information to reproduce the result after the process ends? A restart can pass while a same-size edit fails, or vice versa. The controls in the GIGI study are informative because they test these distinct ways a computation can lose its connection to the object it claims to describe.

The third sensor's alert survives when decoding restores its actual fields. A geometric comparison survives when the restored values carry their units, frame and ruler. A computed answer remains valid when it uses the same evidence and model state. These are successive requirements on one record's journey through software. The next chapter deliberately preserves a coarser object, a distribution, and asks what partial observations of that object determine.

Practice with the construction

23.1. Which field in the three-sensor fixture would reveal a converter that incorrectly replaced every status by its default? Why would checking temperature alone miss the failure?

23.2. In the frame example, rotate $(0, -10)$ by a positive quarter-turn. Compare the physical discrepancy with the discrepancy obtained by subtracting the unconverted arrays.

Chapter 24

The geometry of a measured distribution

Chapter 23 checked that a shorter file still decoded to the same sensor readings. Here the task changes. In a table of 6,497 wine records, we want to compare how often alcohol concentrations fall in fixed intervals. Reordering the rows leaves those proportions alone. Replacing concentrations by bin midpoints keeps the proportions but changes the individual readings. We will track what each representation retains, then ask what partial information about it determines.

The GIGI geometric inference study makes that choice explicit. It constructs bounded analytic fingerprints from public wine and power-plant measurements. We will follow the construction from ordinary numbers to a geometric inference, with enough actual data visible to repeat the calculation.

Here we must first decide what our globe represents. A map of the wine's distribution can preserve the proportion in each alcohol interval while losing which bottle supplied each reading. We will construct that distribution object explicitly, then examine how much of its analytic fingerprint a partial description determines.

Begin with the measurements

Table 24.1 reproduces the first six source rows for four selected fields. The wine entries begin the red-wine file; the full experiment concatenates 1,599 red-wine and 4,898 white-wine records. The plant entries come from the first worksheet of the supplied workbook. Its other worksheets contain shuffled copies, which the study does not count as additional observations.

Table 24.1: Selected measurements, in their original file order.

Row	Wine alcohol (% vol.)	Wine acidity (g/dm^3)	Plant air (°C)	Plant output (MW)
1	9.4	7.4	14.96	463.26
2	9.8	7.8	25.18	444.37
3	9.8	7.8	5.11	488.56
4	9.8	11.2	20.86	446.48
5	9.4	7.4	10.82	473.90
6	9.4	7.4	26.27	443.67

Wine and plant columns come from separate datasets; sharing a displayed row does not pair their observations. Acidity is fixed acidity. Sources: UCI Wine Quality and Combined Cycle Power Plant.

The row number is a position in a source file, not a timestamp or a match between a wine sample and a plant observation. Each dataset is processed separately. All twelve wine fields and all five plant fields enter the study, including wine quality and electrical output. Those two fields are observed distributions here, not targets being predicted from the others.

That leaves 17 fields from 16,065 records. For each field the same procedure constructs 32 histogram bins, supplies seven complex fingerprint values, and asks for 24 withheld values. The resulting $17 \times 24 = 408$ queries give us both individual examples and a complete field-by-field summary.

Construct the object before using its geometry

Divide a field's observed range into 32 equal-width bins. Let n_k count the observations in bin k, let N be the total count, and set $p_k = n_k/N$. The fingerprint is

$$H(z) = \sum_{k=0}^{31} p_k z^k.$$

For wine alcohol, the range is 8.0 to 14.9 percent by volume. Its bin edges are

$$b_k = 8.0 + 0.215625k, \qquad k = 0, \ldots, 32.$$

Each bin includes its left edge and excludes its right edge, except that the last includes the maximum. Table 24.2 gives every count.

Table 24.2: The complete wine-alcohol histogram.

Bin k	Count n_k	Bin k	Count n_k
0	2	16	188
1	5	17	174
2	33	18	202
3	189	19	239
4	311	20	186
5	440	21	134
6	892	22	124
7	355	23	64
8	375	24	25
9	399	25	51
10	294	26	9
11	429	27	15
12	444	28	2
13	360	29	0
14	262	30	0
15	293	31	1

Total: 6,497. Coefficient $p_k = n_k/6497$; bin edges are $8 + 0.215625k$. Read each half from top to bottom.

Divide each count by 6,497 to obtain the coefficients. The polynomial begins

$$H(z) = \frac{2 + 5z + 33z^2 + 189z^3 + \cdots + z^{31}}{6497}.$$

At $z = 0$, every positive power disappears. Thus $H(0) = 2/6497$ reads off the share of wines in the first interval. A nonzero query mixes several bin proportions. Later we use the value at zero and six other evaluations to constrain a withheld value. The nonnegative coefficients sum to one, so

$$|H(z)| \le \sum_{k=0}^{31} p_k |z|^k \le 1 \qquad (|z| \le 1).$$

We have constructed an analytic function with a known bound. The complex variable z probes the distribution; it is not a complex alcohol concentration. The measurements determine the coefficients, while the construction supplies the analytic structure.

With common bin edges, two datasets produce the same polynomial exactly when their ordered bin proportions agree. Permuting rows preserves those proportions. Changing a measurement within its bin preserves them too. Multiplying every count by the same positive integer also leaves H unchanged. The fingerprint therefore preserves this binned distribution; the record count and row identities must be retained separately when an application needs them.

Common edges are essential. If two datasets separately choose their minima and maxima, coefficient p_4 may refer to different physical intervals. A comparison over time can freeze reference edges, report observations falling outside them, and retain that convention with the fingerprint. The experiment's per-field range choice is specified; comparing changing distributions additionally requires a shared range convention.

The histogram fingerprint has a declared resolution

The retained wine-alcohol polynomial has degree at most thirty-one and coefficients given by thirty-two bin proportions. Its information content is therefore tied to those bins. Moving a measurement within one bin leaves the polynomial unchanged. Moving it across a boundary changes two coefficients by $1/N$ in opposite directions.

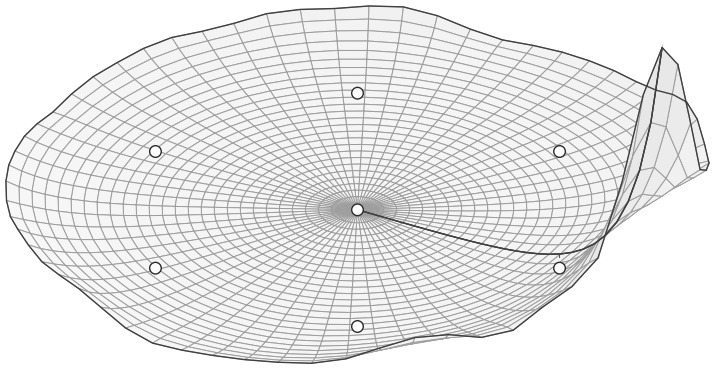
If one sample moves from bin j to bin k, the fingerprint change is $(z^k - z^j)/N$. On $|z| \le r < 1$,

$$|\Delta H(z)| \le \frac{r^j + r^k}{N} \le \frac{2}{N}.$$

That is a deterministic sensitivity bound for this one-record edit. It is not a statement about the probability of an edit or about predicting the original continuous measurement. The retained count N and fixed bin convention make the bound interpretable.

The prepared data defines a surface of values

Height = |H(z)|; H(z) = Σ $p_k z^k$. This is the constructed analytic fingerprint.



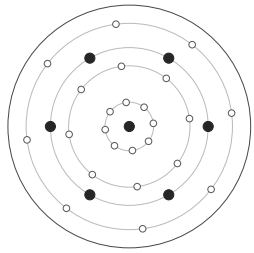
6,497 source records

32 nonnegative coefficients

Σ p_k = 1

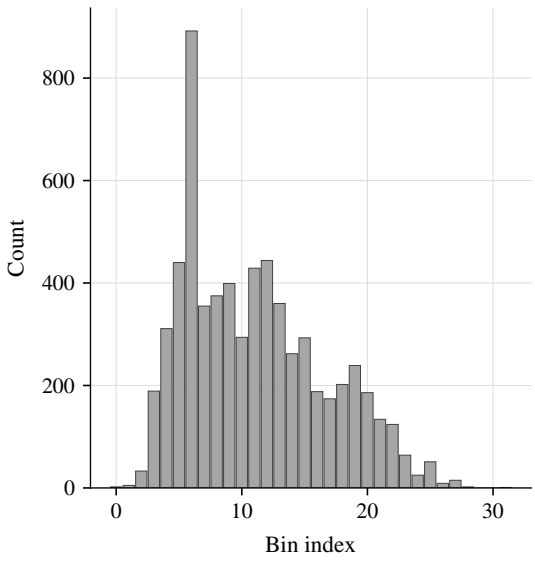
7 supplied function values

24 withheld queries in this field



Filled: supplied / open: withheld locations

A The retained histogram supplies the baseline



B Moving one record changes two coefficients

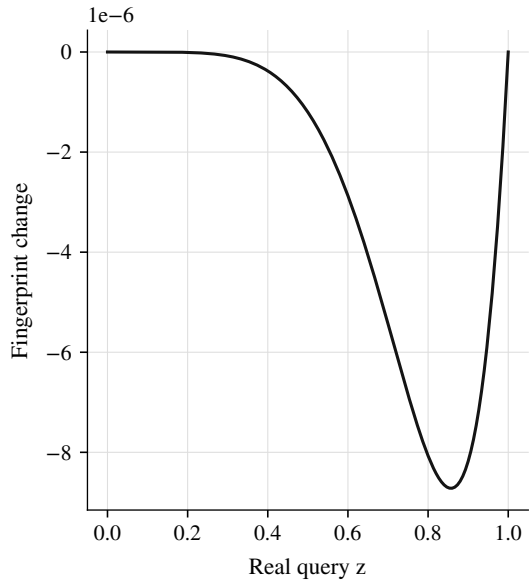
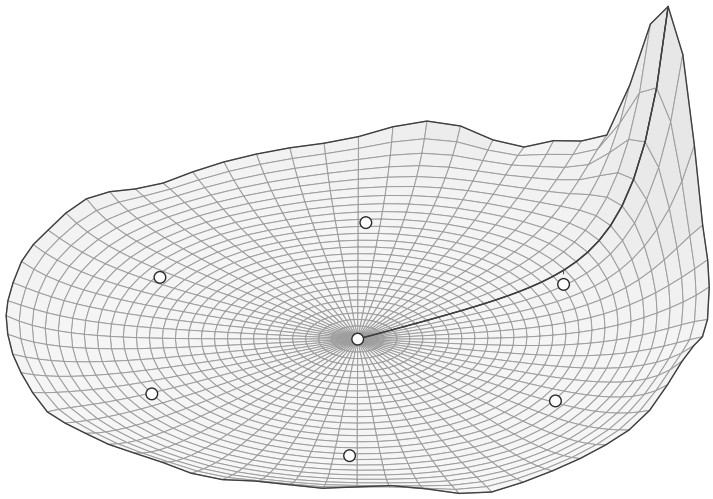


Figure 24.1: *A one-record change has a computable fingerprint effect.* The study continues on the following figure page.

The prepared data defines a surface of values

Height = |H(z)|; H(z) = Σ $p_k z^k$. This is the constructed analytic fingerprint.



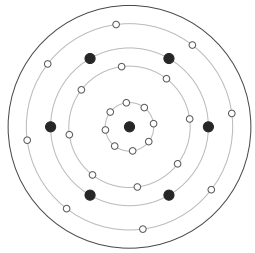
6,497 source records

32 nonnegative coefficients

Σ p_k = 1

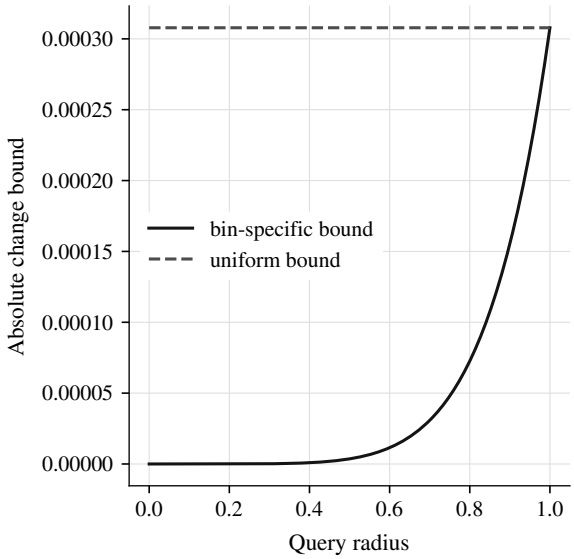
7 supplied function values

24 withheld queries in this field



Filled: supplied / open: withheld locations

C The edit has a computable output bound



D The sample size and bin grid stay fixed

Decrease count 6 by one.

Increase count 7 by one.

N=6497 remains fixed.

$\Delta H(z)=(z^7-z^6)/6497$

The original empirical counts are preserved.

Figure 24.1: *A one-record change has a computable fingerprint effect (continued).* The actual wine-alcohol bin grid is used to show an illustrative move between two bins. Coefficient differences and query-value changes are computed from the declared edit. The original empirical table remains unchanged.

Two geometries occupy the same construction

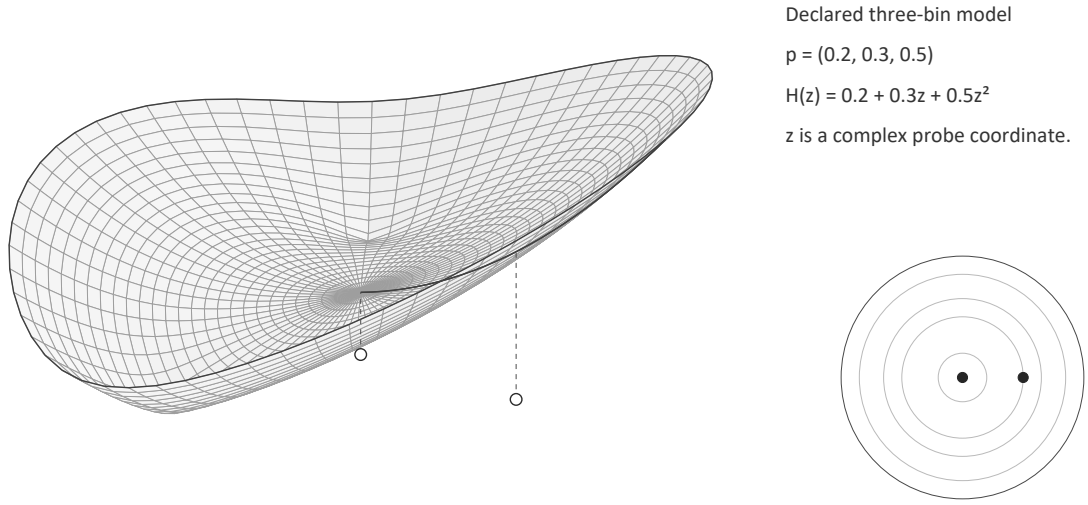
The coefficient vector $p = (p_0, \ldots, p_{31})$ belongs to a probability simplex. The complex probe z belongs to a disk. These are different spaces. A map is holomorphic on the disk if it has a complex derivative at every point there. A Fisher metric can compare nearby coefficient distributions on an interior simplex; the Poincaré metric controls holomorphic maps of the probe disk. Neither is a substitute for the other.

For strictly positive probabilities, the simplex Fisher form is $\sum_k dp_k^2/p_k$ restricted to $\sum_k dp_k = 0$. The square-root map $p \mapsto (\sqrt{p_k})$ turns it into four times the sphere metric on the positive orthant. Zero bins lie at the boundary of that regular chart. Adding pseudocounts would change the distribution and must be declared rather than silently performed.

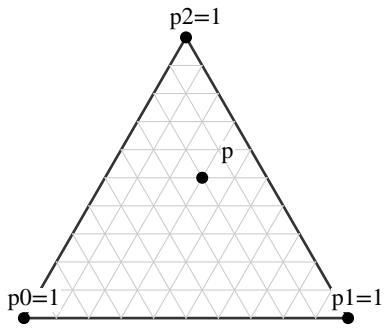
The analytic fingerprint then maps a chosen coefficient vector to a function H. A partial observation of that function creates a completion problem over a larger function class. At each step GIGI works with a different quantity. The histogram keeps bin proportions from the records, the polynomial encodes those proportions, and the supplied values constrain what the function can do elsewhere.

The prepared data defines a surface of values

Height = $|H(z)|$; $H(z) = \Sigma\, p_k z^k$. This is the constructed analytic fingerprint.



A Three proportions define one distribution



B Square roots expose the Fisher arc

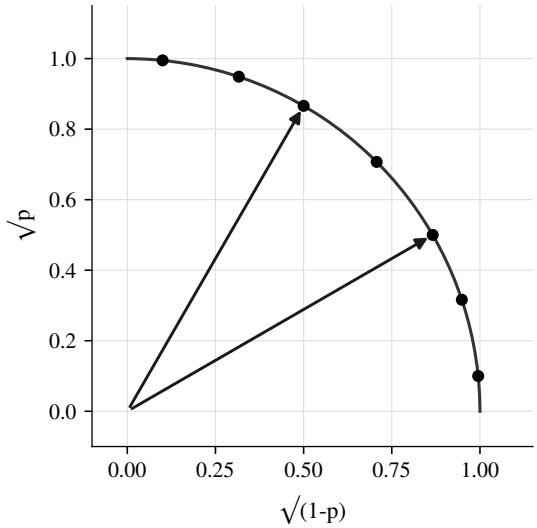
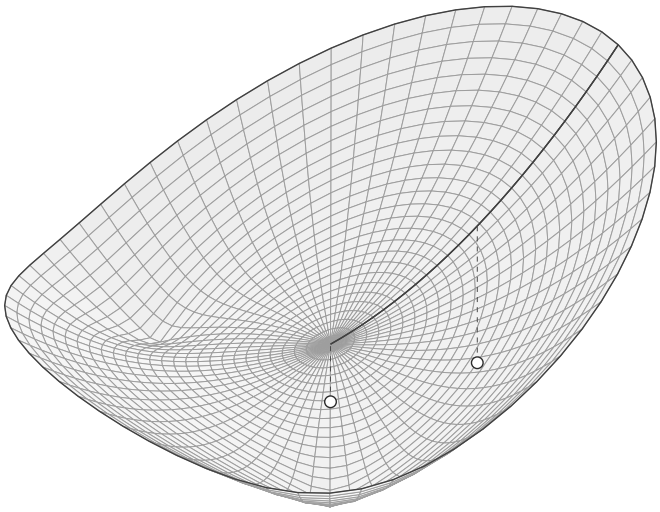


Figure 24.2: *The distribution space and the probe space have different rulers.* The study continues on the following figure page.

The prepared data defines a surface of values

Height = |H(z)|; H(z) = Σ $p_k z^k$. This is the constructed analytic fingerprint.

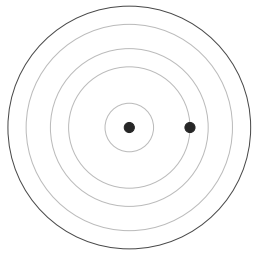


Declared three-bin model

p = (0.2, 0.3, 0.5)

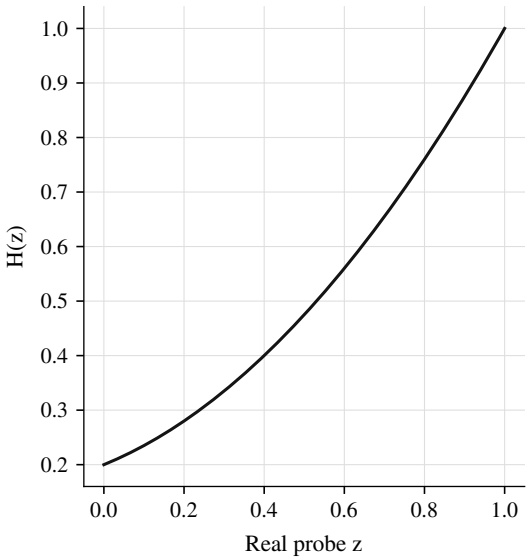
H(z) = 0.2 + 0.3z + 0.5z²

z is a complex probe coordinate.



Probe disk / marked values at 0 and 1/2

C The coefficients define an analytic fingerprint



D Distribution space and probe space differ

p=(.2,.3,.5) is a distribution.

Fisher form: Σ dp²/p on Σdp=0

z is a complex probe in the disk.

H(z)=.2+.3z+.5z²

The Poincaré ruler acts on the probe disk.

Figure 24.2: *The distribution space and the probe space have different rulers (continued).* The three-bin simplex is paired with the simpler two-outcome square-root arc that demonstrates the Fisher length construction. The three-bin polynomial is evaluated along a real radius of the complex probe disk. Distribution coordinates and probe coordinates play different roles; the full empirical study retains all thirty-two bins.

The binning step has already discarded distinctions within each interval. Before asking what analytic completion recovers, we can calculate one precise measure of the information lost by a coarser distribution description.

A divergence, a ruler, and a decision are different objects

Two histogram summaries can be compared by a divergence as well as by an intrinsic distance. The quantities are related, but they obey different rules. A divergence can distinguish distributions even when it fails the conditions for a metric.

For strictly positive probability vectors p, q, define

$$D_{\mathrm{KL}}(p\|q) = \sum_i p_i \log \frac{p_i}{q_i}.$$

It is nonnegative. Jensen's inequality for the concave logarithm gives

$$\sum_i p_i \log \frac{q_i}{p_i} \le \log\left(\sum_i q_i\right) = 0.$$

Equality requires the ratios q_i/p_i to agree, hence $p = q$. This establishes a useful separation property, but not all metric axioms. In general the divergence is asymmetric and does not satisfy the triangle inequality.

A small table makes the asymmetry visible

For Bernoulli probabilities $p = 0.2$ and $q = 0.6$,

$$D_{\mathrm{KL}}(p\|q) = 0.2\log(1/3) + 0.8\log 2,$$
$$D_{\mathrm{KL}}(q\|p) = 0.6\log 3 + 0.4\log(1/2).$$

The values are approximately 0.334795 and 0.381909. Reversing the arguments changes the answer. Symmetrizing removes that asymmetry, but it does not automatically supply the triangle inequality.

The Fisher distance between the same probabilities is

$$2\left|\arcsin\sqrt{0.6} - \arcsin\sqrt{0.2}\right|.$$

It is symmetric because it is the length of the interval under a Riemannian ruler. The divergence and the distance can both be useful, but the operation being performed should be named.

Derive the local metric from the divergence

Write $q_i = p_i + \delta_i$ with $\sum_i \delta_i = 0$ and each $|\delta_i|$ small compared with p_i. Expanding the logarithm,

$$\begin{aligned} D_{\mathrm{KL}}(p\|p+\delta) &= -\sum_i p_i \log(1 + \delta_i/p_i) \\ &= \frac{1}{2}\sum_i \frac{\delta_i^2}{p_i} - \frac{1}{3}\sum_i \frac{\delta_i^3}{p_i^2} + O(\|\delta\|^4). \end{aligned}$$

The linear term vanishes because the probability increments sum to zero. The leading quadratic term is one half of the Fisher squared tangent length.

The remainder is local. If some p_i is extremely small, the requirement that δ_i/p_i be small can fail even when the ordinary Euclidean norm of δ looks modest. A local approximation must state the operating region, especially near the boundary of the simplex.

For Bernoulli $p = 1/2$ and increment h, the exact expression is

$$D_{\mathrm{KL}}(1/2\|1/2+h) = -\tfrac{1}{2}\log(1-4h^2) = 2h^2 + 4h^4 + O(h^6).$$

The Fisher tangent form at $1/2$ is $4h^2$. Its half is $2h^2$, matching the expansion. The higher terms explain why substituting a finite divergence directly for a squared geodesic distance is only an approximation.

Merge bins and calculate what is lost

Partition the original bins into groups B_j. Let $P_j = \sum_{i\in B_j} p_i$ and $Q_j = \sum_{i\in B_j} q_i$. Within a group with positive totals, define conditional proportions $p_{i|j} = p_i/P_j$ and $q_{i|j} = q_i/Q_j$. Then

$$\frac{p_i}{q_i} = \frac{P_j}{Q_j}\frac{p_{i|j}}{q_{i|j}}.$$

Substituting and summing gives the exact decomposition

$$D_{\mathrm{KL}}(p\|q) = D_{\mathrm{KL}}(P\|Q) + \sum_j P_j D_{\mathrm{KL}}(p_{\cdot|j}\|q_{\cdot|j}).$$

Each conditional divergence is nonnegative, so merging bins cannot increase this discrepancy. The lost amount is precisely the weighted difference in the within-group conditional distributions.

Take $p = (0.4, 0.1, 0.2, 0.3)$ and $q = (0.1, 0.4, 0.3, 0.2)$. Merging the first two bins and the last two gives $P = Q = (0.5, 0.5)$. The coarse divergence is zero even though the fine distributions differ. Their full divergence is

$$0.3\log 4 + 0.1\log(3/2) \simeq 0.456435.$$

The coarse representation has deliberately identified the two distributions. No later inference based only on those two totals can recover which internal arrangement was present.

Connect the calculation to GIGI's object

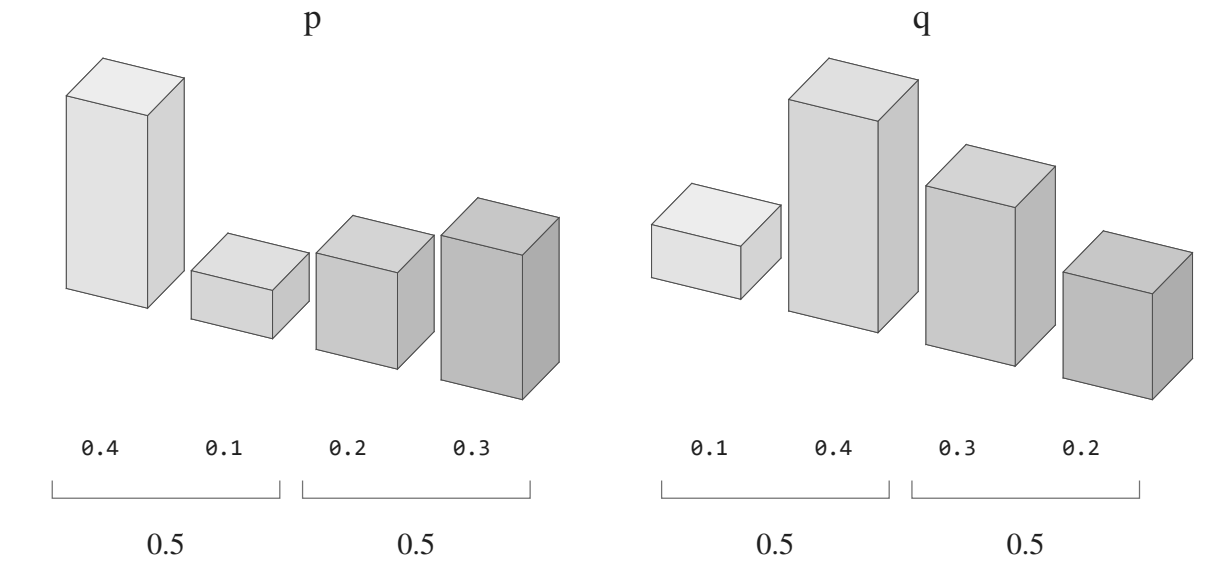
The positivity assumptions also matter for actual histograms. If $p_i > 0$ and $q_i = 0$, the corresponding divergence is infinite. A bin with $p_i = 0$ contributes zero under the usual limiting convention. The interior Fisher formula cannot be evaluated at a zero coefficient by ordinary division. One can work on a specified common-support face or choose a smoothing model, but adding pseudocounts changes the represented distribution and must be recorded.

GIGI's polynomial fingerprint preserves the declared ordered bin proportions. It does not preserve distinctions already lost when raw measurements were placed into bins. The completion stage then asks what follows from a limited set of observations of that polynomial. There are therefore two different reductions: measurement-to-histogram and polynomial-to-supplied-evaluations.

The first is controlled by the bin convention and the retained metadata. The second is controlled by the analytic interpolation constraints. A narrow completion disk does not reverse the original binning. Conversely, a histogram that retains a task distinction may still require enough well-placed analytic observations before that distinction is determined by completion.

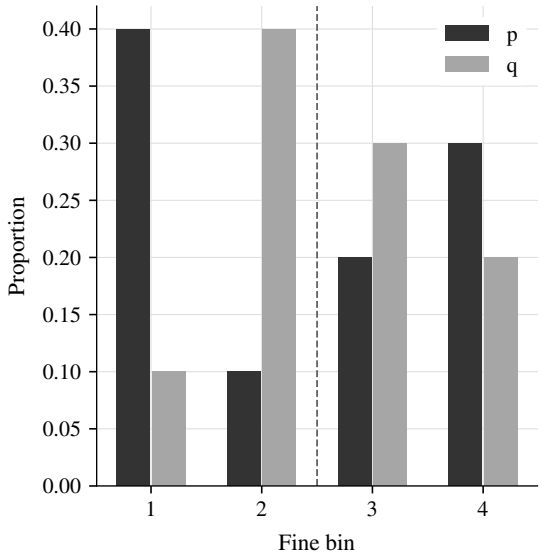
Coarse totals can conceal different fine records

Two four-bin distributions become the same two-bin distribution after merging.



Fine KL = 0.456435…; coarse KL = 0. The missing distinction lies inside the merged groups.

A Different fine bins give identical coarse totals



B The Fisher form captures the local quadratic term

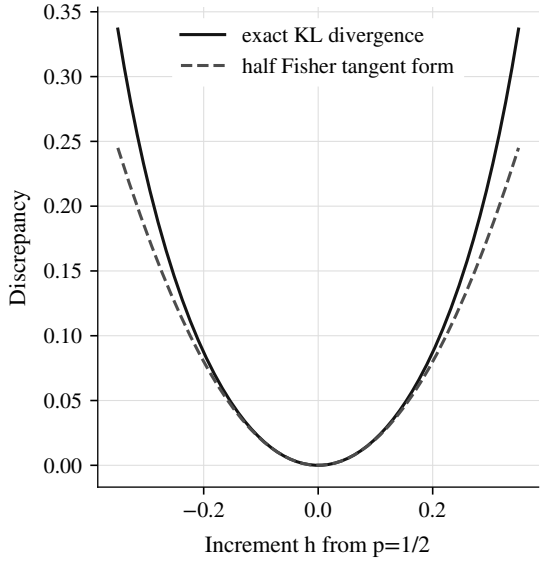
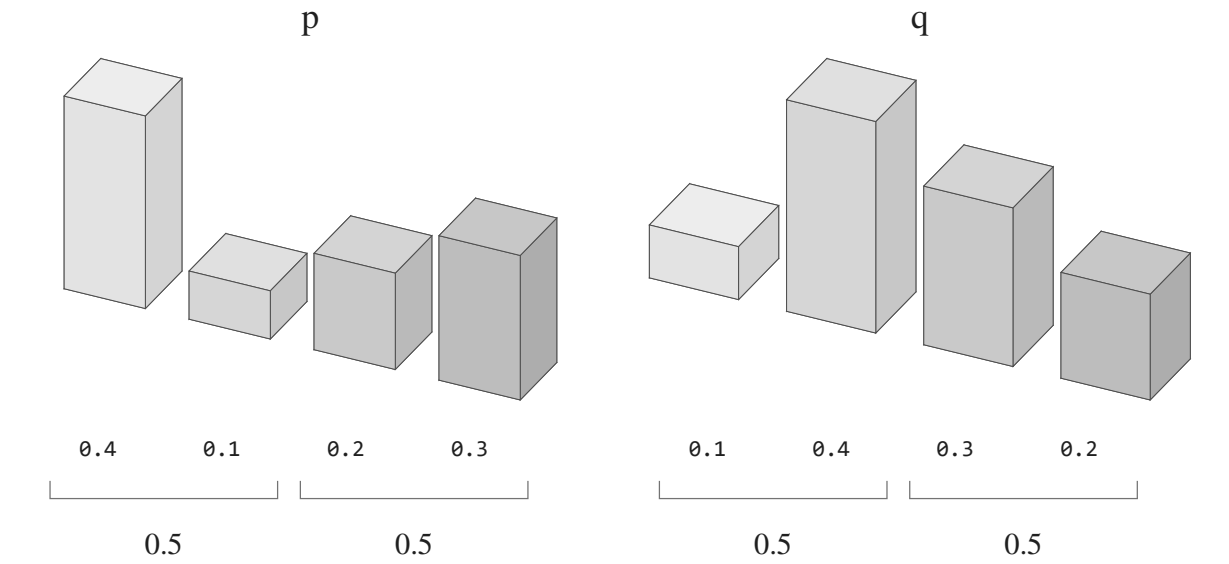


Figure 24.3: *A local ruler retains only the leading term of a finite divergence.* The study continues on the following figure page.

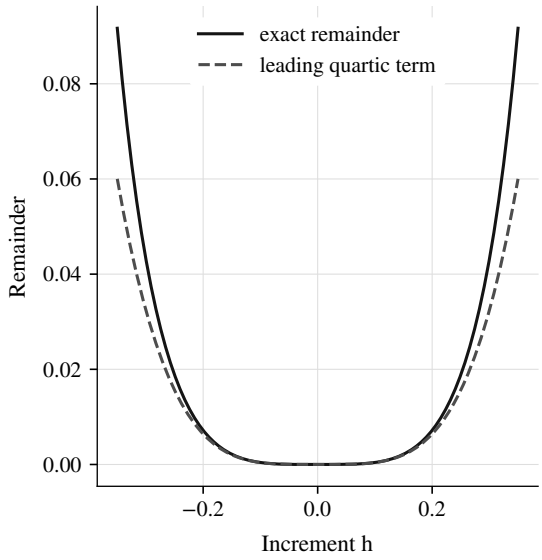
Coarse totals can conceal different fine records

Two four-bin distributions become the same two-bin distribution after merging.



Fine KL = 0.456435…; coarse KL = 0. The missing distinction lies inside the merged groups.

C The higher-order remainder grows away from zero



D Merging bins discards a measurable divergence

Fine KL: 0.3 log4 + 0.1 log(3/2)

Fine KL ≈ 0.456435

Coarse totals: (0.5, 0.5) for both

Coarse KL = 0

The lost terms are conditional divergences.

Figure 24.3: *A local ruler retains only the leading term of a finite divergence (continued).* The four-bin distributions become identical after the stated bin merger. The Bernoulli expansion shows the Fisher quadratic term and the higher-order remainder of the finite divergence.

The divergence calculation makes this hierarchy explicit. It supplies a way to quantify one form of information loss, while the Riemannian metric supplies local lengths and the threshold rule supplies a decision. The same underlying observations can participate in all three constructions without turning them into one interchangeable number.

A coarse equality with a fine difference. Verify the four-bin divergence numerically and derive the two conditional-divergence terms. Give one additional retained statistic that distinguishes the two fine distributions after the coarse totals alone have become identical.

We now know what binning preserved and what merging bins discarded. Hold that chosen histogram fixed. The next question concerns partial access to its fingerprint: if we know the function at several probe locations, which values remain possible elsewhere? The disk geometry supplies the constraints for that inference. It measures a different stage of the same observation-to-answer process.

Why the disk has its own distance

Equip the open unit disk $\mathbb{D}$ with the Poincaré metric:

$$ds^2 = \frac{4\,|dz|^2}{(1-|z|^2)^2},$$

$$d_{\mathbb{D}}(z,w) = 2\operatorname{artanh}\left|\frac{z-w}{1-\overline{w}z}\right|.$$

A small coordinate step has a different intrinsic length at different positions. The radial distance from zero to radius 0.5 is $\log 3 \simeq 1.099$; to radius 0.85 it is $\log(37/3) \simeq 2.512$. The second point is less than twice as far away in the drawing and more than twice as far away in this metric.

The disk drawn on the page is our map. The Poincaré metric supplies its ruler. In the map/globe analogy, the underlying geometry here is hyperbolic, so the word globe does not mean that the modeled space must be a sphere.

A nonconstant holomorphic disk map obeys Schwarz–Pick:

$$d_{\mathbb{D}}(H(z),H(w)) \le d_{\mathbb{D}}(z,w).$$

An allowed fingerprint cannot expand this distance. One known value constrains another; several values must also be jointly compatible.

Table 24.3 shows the seven supplied values for wine alcohol: one probe at zero and six on the circle of radius 0.65. Each is computed from the full histogram. Seven observations here means seven evaluations of H, after the records have already contributed to its construction. It does not mean seven wine samples reconstruct thousands of unseen bottles.

For $H(z_i) = w_i$, form the Pick matrix

$$P_{ij} = \frac{1 - w_i\overline{w_j}}{1 - z_i\overline{z_j}}.$$

For distinct nodes in the disk, positive semidefiniteness of P is the classical criterion for a holomorphic interpolant bounded by one. It means $a^*Pa \ge 0$ for every complex vector a, and tests whether all the observations can belong to one allowed function.

For example, $H(0) = 0$ and $H(1/2) = 0.9$ individually lie within the allowed range. Jointly, however, they give

$$P = \begin{pmatrix} 1 & 1 \\ 1 & 19/75 \end{pmatrix}, \qquad \det P = -56/75 < 0.$$

Table 24.3: The seven supplied wine-alcohol fingerprint values.

No.	$\operatorname{Re} z$	$\operatorname{Im} z$	$\operatorname{Re} H(z)$	$\operatorname{Im} H(z)$
1	0.0000000	0.0000000	0.0003078	0.0000000
2	0.6500000	0.0000000	0.0455185	0.0000000
3	0.3250000	0.5629165	0.0010187	-0.0087965
4	-0.3250000	0.5629165	0.0078383	0.0000327
5	-0.6500000	0.0000000	0.0031549	0.0000000
6	-0.3250000	-0.5629165	0.0078383	-0.0000327
7	0.3250000	-0.5629165	0.0010187	0.0087965

Displayed decimals are rounded. The calculation uses the full-precision values, available from the histogram and probe formula.

No function in this class fits both readings. Schwarz–Pick explains the same failure: a disk map fixing zero cannot send 1/2 as far out as 0.9. Checking individual ranges would miss the contradiction.

A single observation already draws a precise disk

Take a holomorphic function $f : \mathbb{D} \to \mathbb{D}$ and suppose $f(z_0) = w_0$. Schwarz–Pick gives

$$\left| \frac{f(q) - w_0}{1 - \overline{w_0} f(q)} \right| \le r, \qquad r = \left| \frac{q - z_0}{1 - \overline{z_0} q} \right|.$$

This is a pseudohyperbolic disk in the output variable. Expanding the squared inequality and completing the square converts it into an ordinary Euclidean disk with center and radius

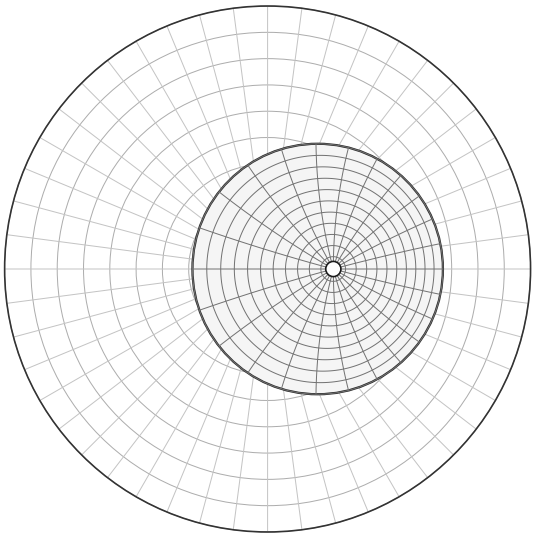
$$c = \frac{(1 - r^2) w_0}{1 - r^2 |w_0|^2}, \qquad \rho = \frac{r(1 - |w_0|^2)}{1 - r^2 |w_0|^2}.$$

The denominator is positive for $r < 1$ and $|w_0| < 1$. When $w_0 = 0$, the disk is centered at zero with radius r. Away from zero, the Euclidean center is not simply w_0.

The two-node Pick criterion gives the converse: its determinant condition is precisely the Schwarz–Pick inequality above, so every point of the disk is attained. For $z_0 = 0$, $w_0 = 1/4$ and $q = 1/2$, the exact values are $c = 4/21$ and $\rho = 10/21$. The constant function $f \equiv 1/4$ gives one compatible query value inside the disk.

A feasible set is a geometric constraint

The disk describes compatible values of a holomorphic disk-to-disk function.



Given f(0) = 1/4

Query at z = 1/2

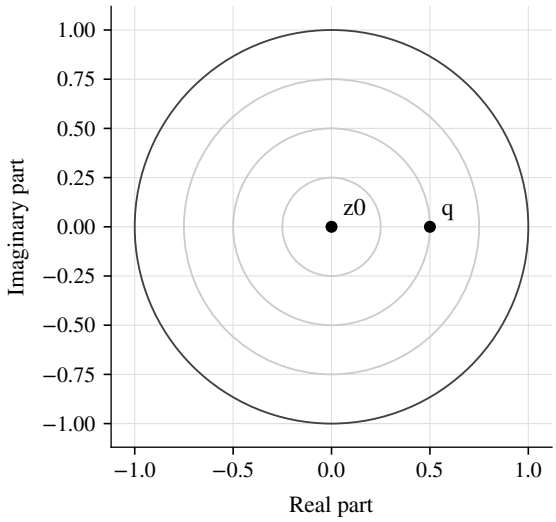
Euclidean center: 4/21

Euclidean radius: 10/21

Pseudohyperbolic radius: 1/2

The set carries no probability law. Separate feasible disks do not guarantee a jointly feasible selection.

A One observed value constrains a new query



B Every compatible answer lies in this disk

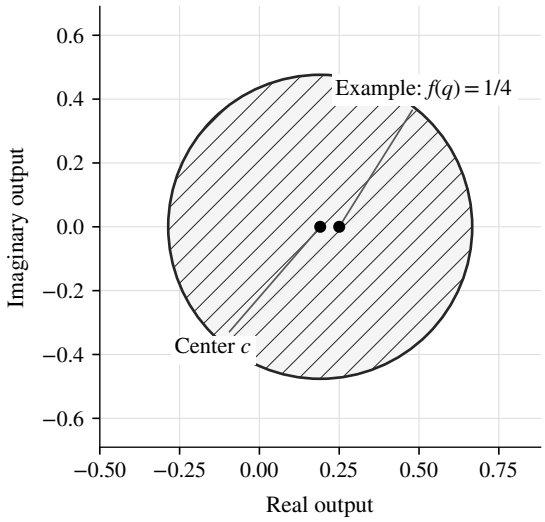
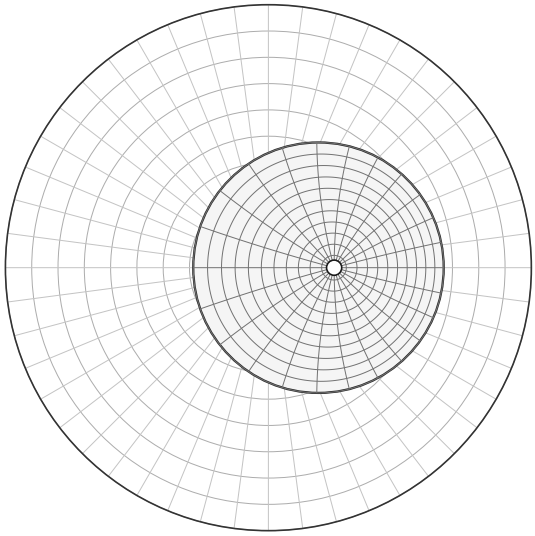


Figure 24.4: *One observed value constrains an entire family of answers.* The study continues on the following figure page.

A feasible set is a geometric constraint

The disk describes compatible values of a holomorphic disk-to-disk function.



Given f(0) = 1/4

Query at z = 1/2

Euclidean center: 4/21

Euclidean radius: 10/21

Pseudohyperbolic radius: 1/2

The set carries no probability law. Separate feasible disks do not guarantee a jointly feasible selection.

C Center and radius follow exactly

D The disk belongs to a specified function class

z0=0; w0=1/4; q=1/2	Holomorphic maps from disk to disk
Pseudohyperbolic radius r=1/2	One exact supplied value
Euclidean center c=4/21	Every compatible query value is included.
Euclidean radius ρ=10/21	No probability distribution on functions is assumed.

Figure 24.4: *One observed value constrains an entire family of answers (continued).* One exact observation at $z_0 = 0$ and a query at $q = 1/2$ determine a Schwarz–Pick output disk. For $w_0 = 1/4$, its Euclidean center is $4/21$ and radius $10/21$. The shaded region is the set of compatible values in the declared function class, not a probability region.

Derive the many-observation disk from the matrix

For observed pairs (z_i, w_i) with positive-definite Pick matrix P, append query (q, w). The new column has entries

$$b_i(w) = \frac{1 - w_i\overline{w}}{1 - z_i\overline{q}} = a_i - d_i\overline{w},$$

where $a_i = (1 - z_i\overline{q})^{-1}$ and $d_i = w_i a_i$. The last diagonal entry is $(1 - |w|^2)/(1 - |q|^2)$. Positive semidefiniteness of the augmented matrix is equivalent to the Schur-complement inequality

$$\frac{1 - |w|^2}{1 - |q|^2} - b(w)^* P^{-1} b(w) \geq 0.$$

Write $s = (1 - |q|^2)^{-1}$, $A = a^* P^{-1} a$, $D = d^* P^{-1} d$ and $E = a^* P^{-1} d$. Expansion gives

$$(s + D)|w|^2 - 2\,\mathrm{Re}(E\overline{w}) \leq s - A.$$

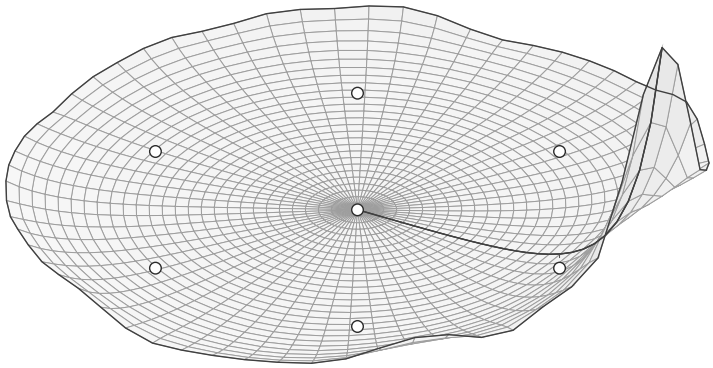
Thus

$$c = \frac{E}{s + D}, \qquad \rho^2 = |c|^2 + \frac{s - A}{s + D}.$$

The formula is evaluated by linear solves, not by explicitly forming an inverse in a numerical implementation. A slightly negative computed ρ^2 can be roundoff near a degenerate case; a substantially negative value requires investigating feasibility or the calculation. Singular Pick matrices require their own range condition or a specialized interpolation treatment, not division by an unavailable inverse.

The prepared data defines a surface of values

Height = $|H(z)|$; $H(z) = \Sigma\, p_k z^k$. This is the constructed analytic fingerprint.



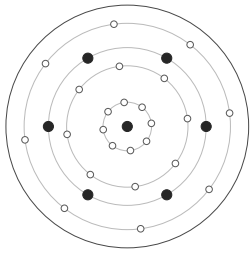
6,497 source records

32 nonnegative coefficients

$\Sigma\, p_k = 1$

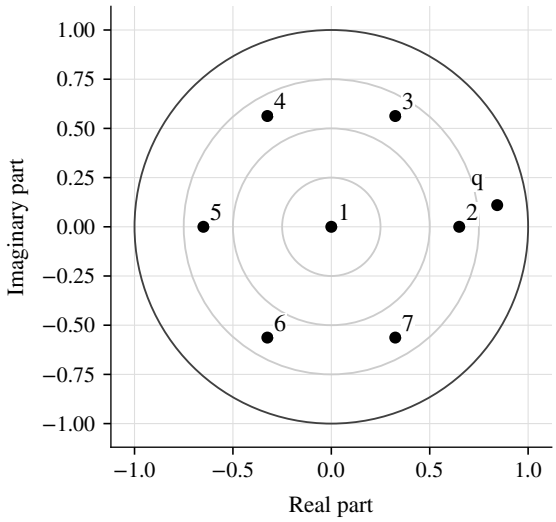
7 supplied function values

24 withheld queries in this field



Filled: supplied / open: withheld locations

A Seven values constrain the prepared fingerprint



B The withheld truth tests the feasible disk

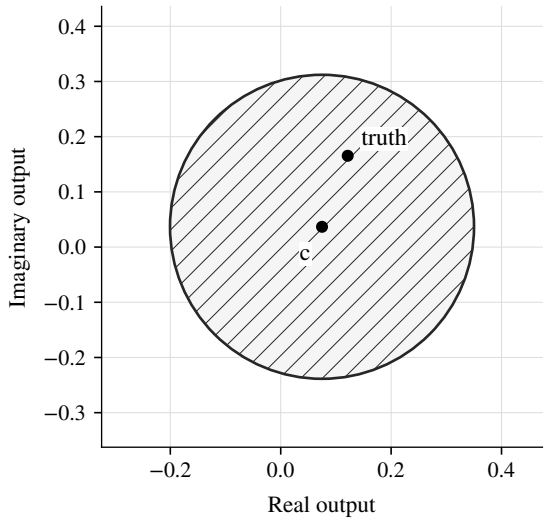
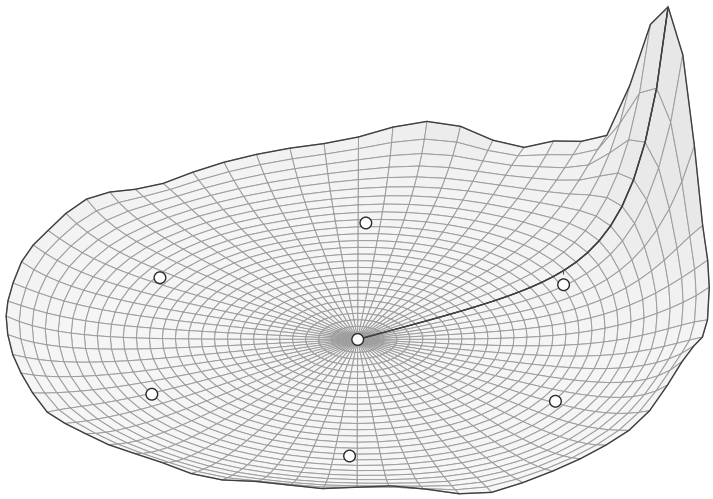


Figure 24.5: *The feasible disk is a matrix inequality made visible.* The study continues on the following figure page.

The prepared data defines a surface of values

Height = $|H(z)|$; $H(z) = \Sigma\, p_k z^k$. This is the constructed analytic fingerprint.



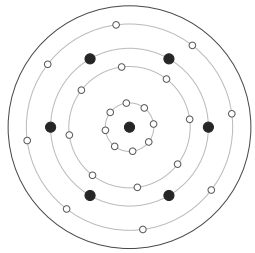
6,497 source records

32 nonnegative coefficients

$\Sigma\, p_k = 1$

7 supplied function values

24 withheld queries in this field



Filled: supplied / open: withheld locations

C The Pick matrix couples the observations

1.00	1.00	1.00	1.00	1.00	1.00	1.00
1.00	1.73	1.15	0.79	0.70	0.79	1.15
1.00	1.15	1.73	1.15	0.79	0.70	0.79
1.00	0.79	1.15	1.73	1.15	0.79	0.70
1.00	0.70	0.79	1.15	1.73	1.15	0.79
1.00	0.79	0.70	0.79	1.15	1.73	1.15
1.00	1.15	0.79	0.70	0.79	1.15	1.73

D The outer query gives a numerical check

Center = 0.0748215 + 0.0367339i

Radius = 0.2753669

Truth gap = 0.1367974

Gap / radius = 0.4967824

The truth is withheld from the solve.

Figure 24.5: *The feasible disk is a matrix inequality made visible (continued).* The seven retained input probes, feasible output disk and absolute entries of the observed-node Pick matrix describe the wine-alcohol query. The register gives its computed center, radius and held-out containment ratio. The independently known polynomial value validates the region; it is not an input to the completion solve.

The answer is a region, and we can inspect it

At an unobserved query point q, append an unknown value w and require the enlarged Pick matrix to remain positive semidefinite. For the strictly feasible examples here, the possible values form a Euclidean disk in the complex output plane:

$$\mathcal{V}(q) = \{w : |w - c(q)| \le \rho(q)\}.$$

The Poincaré metric governs the analytic restriction; ρ describes the resulting set in ordinary output coordinates. The plotted radius is not a geodesic distance.

The engine considers all bounded holomorphic functions fitting the supplied values. It does not additionally impose nonnegative coefficients or degree-31 polynomial form. The original fingerprint belongs to this larger family, but need not be its only member. That explains why a nonzero uncertainty disk can remain.

Use two query locations on the same ray:

$$q_{\text{inner}} = 0.5e^{0.13i}, \qquad q_{\text{outer}} = 0.85e^{0.13i}.$$

Table 24.4 follows the radii as the observations are supplied in the displayed order. Before any are supplied, the allowed set is the unit disk. The intermediate disks are independently reconstructed from the recorded data; the final radii also reproduce on the preserved GIGI executable.

Table 24.4: Adding observations narrows the admissible value set.

Values supplied	Radius at q_{inner}	Radius at q_{outer}
0	1.0000000	1.0000000
1	0.5000000	0.8500000
2	0.1226144	0.4108183
3	0.0761991	0.3537682
4	0.0626840	0.3362660
5	0.0542032	0.3241005
6	0.0454243	0.3093338
7	0.0312438	0.2753669

Radii describe complex output values. These are constraints under the analytic model, not uncertainty in percent alcohol.

The outer query remains much less constrained. In this interpolation problem, the same number of observations buys different precision at different locations. The complete polynomial, retained for validation but withheld from the completion query, lets us inspect the final answers.

Table 24.5 shows the disk centers, radii and independently evaluated truth. In each case the center-to-truth separation is smaller than the radius. The engine's representative is one compatible choice; it need not equal the center or recover the withheld polynomial's value.

These disks are admissible-value sets, not statistical confidence intervals. Their widths come from a function class and supplied equations. Choosing a value independently from each query's disk need not produce one globally compatible function: the disks describe marginal possibilities, and a collection of choices must still satisfy the joint interpolation conditions.

What additional observations have bought

Imagine handing someone seven values of the wine fingerprint and keeping the histogram in your drawer. Ask what the fingerprint could be at the farther of the two points we tested. The disk

Table 24.5: Two final query results, with the held-out truth.

Quantity	Inner query	Outer query
$\operatorname{Re} q$	0.4957809	0.8428276
$\operatorname{Im} q$	0.0648171	0.1101890
$\operatorname{Re} H(q)$ (truth)	0.0115230	0.1213582
$\operatorname{Im} H(q)$ (truth)	0.0067790	0.1653725
$\operatorname{Re} c$	0.0184189	0.0748215
$\operatorname{Im} c$	0.0064237	0.0367339
Radius ρ	0.0312438	0.2753669
$\lvert H(q) - c \rvert$	0.0069050	0.1367974
Re representative	0.0199780	0.0992059
Im representative	0.0069164	0.0481757

Containment is checked as $|H(q) - c| \leq \rho$ using full precision. Representatives need not equal the true polynomial value.

rules out some answers and leaves others possible. To see where the actual value lands, open the drawer and evaluate the polynomial from the saved histogram.

Because the observation sets are nested, the exact feasible function classes are nested. Evaluating those classes at a fixed query produces nested possible-value sets. This statement follows from set inclusion, before any implementation is chosen. It gives an algorithm an invariant to check: a later exact feasible set should not acquire a value excluded by an earlier exact set.

The farther query retains more freedom because analytic interpolation couples locations through the disk's geometry. This supplies a precise example of why the position of evidence matters as well as its count. It does not establish that every empirical dataset or every inference problem follows the same numerical contraction curve.

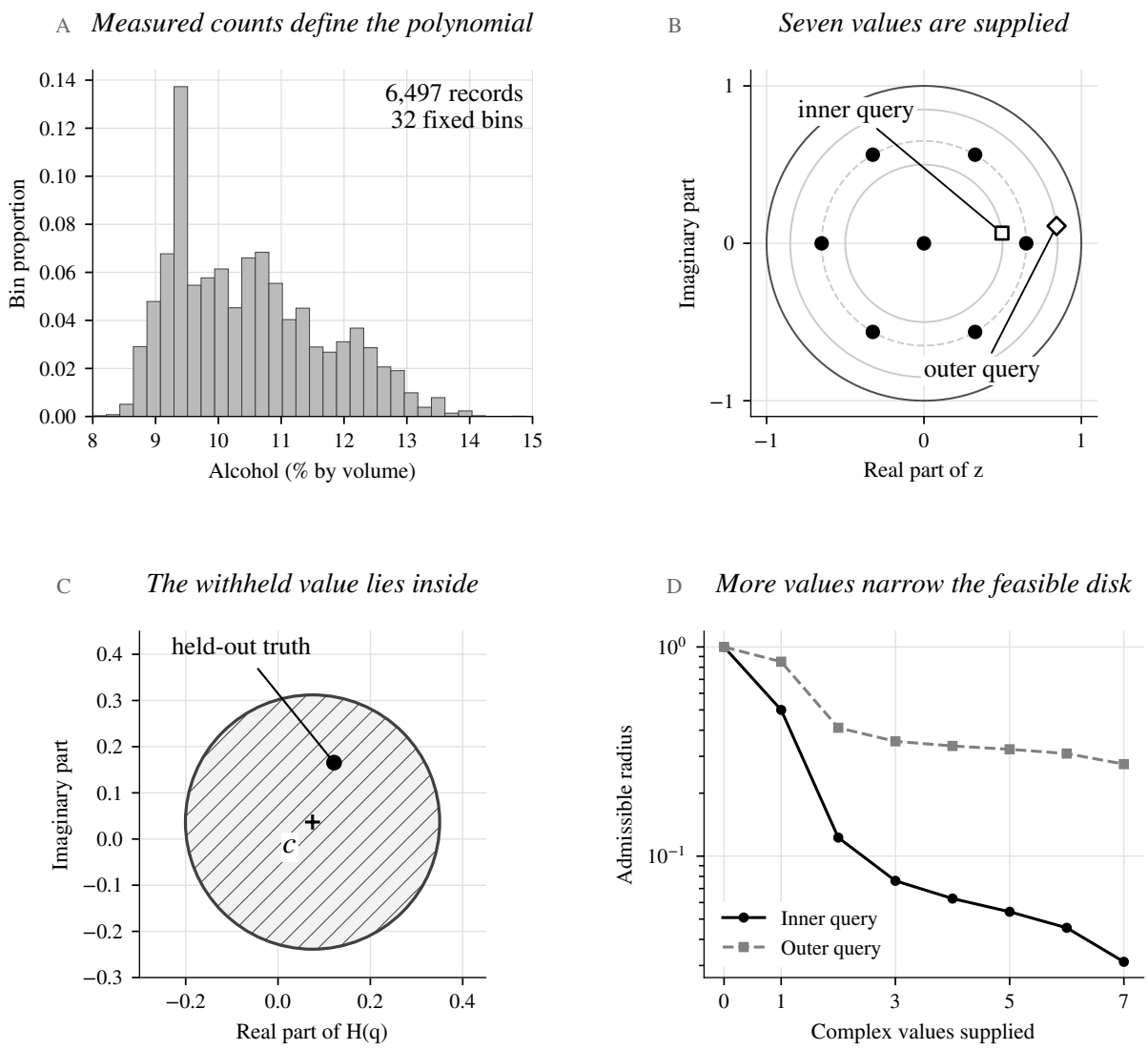


Figure 24.6: The wine-alcohol field from the measured-data study, represented by 32 fixed bins from 6,497 records. Seven complex values constrain the analytic fingerprint. The outer-query disk contains the held-out value, and the prefix curves show how the admissible radius changes at the inner and outer query locations. The disk is an admissible set under the analytic model, not a statistical confidence region.

Read the complete field table

Table 24.6 compares each measured field to a baseline with the same seven observation nodes and all seven supplied values set to zero. Every row uses the same 24 queries at eight angles on each circle of radius 0.2, 0.5 and 0.85.

The measured median radius is 0.0326711 pooled over all 408 queries. The zero baseline median is 0.0381204. Compared with radius one, those are reductions of 96.73% and 96.19%. Ten field medians are within 3% of baseline. Chlorides and residual sugar are at 67% and 69%, the smallest ratios.

Appendix H derives the baseline and records all 408 withheld values. Each lies in its computed disk because the generating polynomial is admissible. The largest field radius is 0.3615180.

The study also changes the evidence deliberately. An inconsistent edit must produce an infeasible result. Restoring the original values must restore the answers. Scaling a coherent observation set

Table 24.6: Final disk radii over 24 queries per field.

Dataset	Field	Median	Largest	Median / baseline
	Zero values	0.038120	0.361535	1.000
Wine	Fixed acidity	0.037136	0.359627	0.974
Wine	Volatile acidity	0.032881	0.356064	0.863
Wine	Citric acid	0.033666	0.354450	0.883
Wine	Residual sugar	0.026238	0.301759	0.688
Wine	Chlorides	0.025470	0.327083	0.668
Wine	Free sulfur dioxide	0.029563	0.351521	0.776
Wine	Total sulfur dioxide	0.037100	0.359987	0.973
Wine	Density	0.030989	0.353088	0.813
Wine	pH	0.037984	0.361279	0.996
Wine	Sulphates	0.032739	0.355401	0.859
Wine	Alcohol	0.037090	0.359531	0.973
Wine	Quality score	0.038047	0.361212	0.998
Plant	Air temperature	0.037990	0.361326	0.997
Plant	Exhaust vacuum	0.038095	0.361478	0.999
Plant	Ambient pressure	0.038107	0.361507	1.000
Plant	Relative humidity	0.038111	0.361518	1.000
Plant	Electrical output	0.037770	0.360942	0.991

The baseline fixes all seven values at zero. Wine has 6,497 records per field and plant has 9,568. The disks contain all 408 held-out fingerprint values.

while keeping its record count unchanged must change the answers. Restarting must preserve the recovered answers for all 17 fields. The controls passed in the original runs and when we repeated them. They matter because the same number of rows can contain different evidence.

A separate planted check supplies $H(0) = 0$ and $H(1/2) = 1/2$. Equality in Schwarz's lemma forces $H(z) = z$. The engine returns a numerically unique solution, with zero-radius disks to the checked tolerance. At $0.2 + 0.3i$ and $-0.4 + 0.1i$, the outputs agree with those query coordinates. The answer is known in advance; this checks the operation against a mathematical consequence.

The field table shows which individual queries were constrained. We now ask a stronger question: can the selected answers all belong to one function?

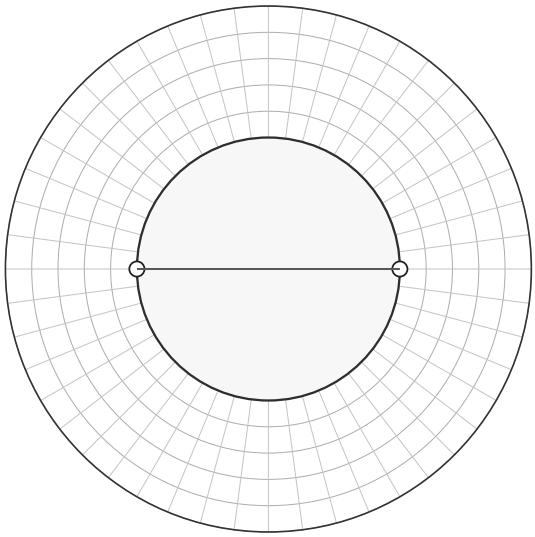
Two individually possible values can be jointly impossible

Consider a new completion test with only one observation of a holomorphic disk map f. Knowing $f(0) = 0$ allows $|f(q)| \le |q|$. At $q_1 = 1/2$, the value $1/2$ is allowed. At $q_2 = -1/2$, the value $1/2$ is also individually allowed. But selecting both is impossible: equality at the first nonzero point forces $f(z) = z$, which would give $f(-1/2) = -1/2$.

Each output disk describes the possibilities at one query. A joint choice must satisfy the enlarged Pick matrix for all selected values at once. The three clocks in Chapter 10 posed the same kind of problem: pairwise time offsets could be plausible in isolation while failing to fit one common time scale.

A feasible set is a geometric constraint

The disk describes compatible values of a holomorphic disk-to-disk function.



Given f(0) = 0

Select f(1/2) = 1/2.

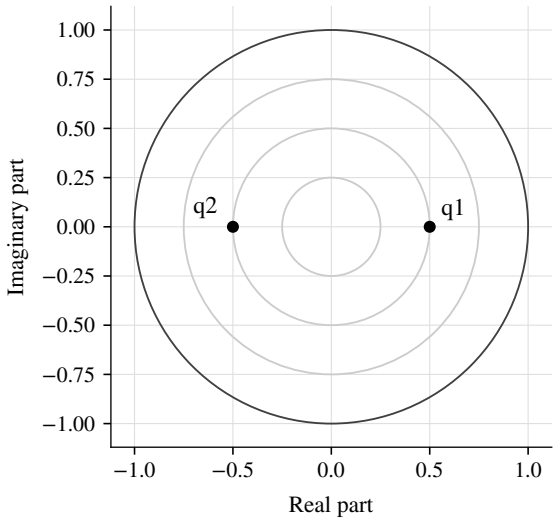
Equality forces f(z) = z.

Then f(−1/2) = −1/2.

A second choice of +1/2 fails jointly.

The set carries no probability law. Separate feasible disks do not guarantee a jointly feasible selection.

A Each query is feasible on its own



B Each separate query permits a radius-1/2 disk

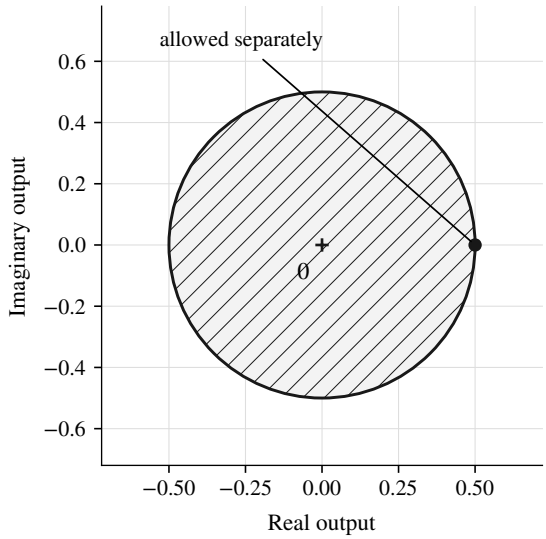
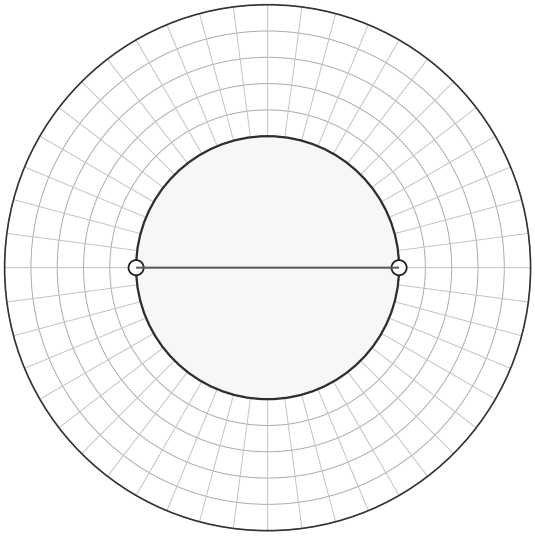


Figure 24.7: *Marginally possible answers need not belong to one function.* The study continues on the following figure page.

A feasible set is a geometric constraint

The disk describes compatible values of a holomorphic disk-to-disk function.



Given f(0) = 0

Select f(1/2) = 1/2.

Equality forces f(z) = z.

Then f(−1/2) = −1/2.

A second choice of +1/2 fails jointly.

The set carries no probability law. Separate feasible disks do not guarantee a jointly feasible selection.

C The joint matrix rejects the chosen pair

D The first extreme value forces the second

$$z=(0,\tfrac{1}{2},-\tfrac{1}{2}),\quad w=(0,\tfrac{1}{2},\tfrac{1}{2})$$

$$P=\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 3/5 \\ 1 & 3/5 & 1 \end{bmatrix}$$

$$\det P=-\tfrac{4}{25}<0 \implies \lambda_{\min}(P)<0$$

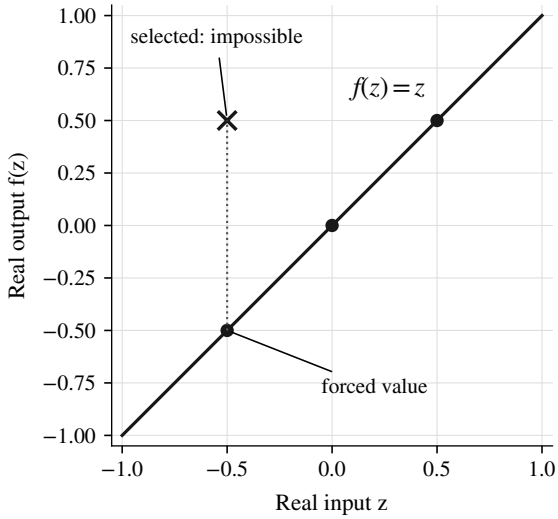


Figure 24.7: *Marginally possible answers need not belong to one function (continued).* With $f(0)=0$, the choices $f(1/2)=1/2$ and $f(-1/2)=1/2$ are separately feasible but jointly impossible. Their three-node Pick matrix has determinant $-4/25$. The equality case forces $f(z)=z$, so the second value must instead be $-1/2$. The graph shows that contradiction directly.

The fingerprint has a real part and an imaginary part defined on the same disk. Their slopes give us a way to check whether they could have come from one analytic function.

Check that the two fields describe one analytic function

My obsession with harmonics started in a bar scene. In *Good Will Hunting* (1997), Sean and Lambeau are arguing about Will's future [45]. Somewhere in the middle of all that, Robin Williams says, "Specifically, bounded harmonic functions." It was like an echo of an echo in my brain. Weird way to put it, I know, but those words stuck like a fly on flypaper.

Later I found a thin book on holomorphic functions in a used bookstore. Two hundred and ten pages at most. I read it over and over and over.

A few weeks before I wrote this, I was working on the Crouzeix conjecture and had one of those moments where you stop and go, *wait, these two things are connected.* The Crouzeix work got me asking how much a few values can tell you about a bounded holomorphic function. In GIGI I could try that with real data, and I could ask another question. If I know the whole real part of a function, how much of the imaginary part can I recover? Both questions made it into *The Geometry of Flight* [20]. Here we can see the second one with the wine fingerprint we already have.

Picture a thin metal plate with insulated faces. Hold its rim at fixed temperatures and wait for the readings to settle. In the ideal model with uniform thermal conductivity and no internal heat sources, the steady temperature $u(x, y)$ is harmonic. Draw a circle with its whole interior inside the plate. The temperature at the center equals the average around the circle. The local rule is Laplace's equation, $u_{xx} + u_{yy} = 0$, for a twice continuously differentiable field. Each second derivative measures how fast the slope changes along one axis. Those changes can cancel even when the surface bends. In the film, *bounded* puts one more condition on the function. Its absolute value has a finite ceiling throughout the region.

A holomorphic function has a complex derivative at every point of an open region. Write its value at $z = x + iy$ as $H(z) = u(x, y) + iv(x, y)$. Take a small complex step h and divide the change in H by that step. The quotient $[H(z + h) - H(z)]/h$ must tend to the same value however h approaches zero. A step east gives the limiting quotient $u_x + iv_x$. A step north gives $(u_y + iv_y)/i = v_y - iu_y$. Dividing by i rotates the northward measurement back a quarter-turn. Matching the real and imaginary parts of the two quotients gives the Cauchy–Riemann equations

$$u_x = v_y, \qquad u_y = -v_x.$$

Differentiate the first equation in x and the second in y, then add. The mixed derivatives of v cancel, giving $u_{xx} + u_{yy} = 0$. The same argument gives $v_{xx} + v_{yy} = 0$. Each part of a holomorphic function is therefore harmonic. The pairing imposes a further condition. Try $u = x$ and $v = 0$. Both fields are harmonic. Here $u_x = 1$ and $v_y = 0$, so the pair fails the first Cauchy–Riemann equation.

For $H(z) = z^2$, the two fields are $u = x^2 - y^2$ and $v = 2xy$. The real field bends up along x and down along y by equal amounts, so its second derivatives cancel. The Cauchy–Riemann equations give both slopes of its harmonic conjugate v from the slopes of u. On the disk, harmonicity of u makes the integral of those slopes independent of the path. One value of v fixes the additive constant.

For the wine-alcohol fingerprint, the harmonic-conjugate study starts with the entire real field $u = \operatorname{Re} H$ and fixes the imaginary part at one point. The earlier completion engine had only seven evaluations of H. Table 24.7 checks the reconstructed field against the original polynomial at $z = 0.2 + 0.3i$.

Table 24.7: Wine-alcohol fingerprint and derivatives at $0.2 + 0.3i$.

Quantity	Real part	Imaginary part
H	-0.001308978	0.000052343
H'	-0.012377215	0.006311034
H''	-0.083621857	0.070958944

These are values of the analytic fingerprint and its derivatives, not physical rates of change of alcohol concentration.

The first derivative gives $u_x = v_y = \operatorname{Re} H'$ and $v_x = -u_y = \operatorname{Im} H'$. The second gives $u_{xx} = \operatorname{Re} H''$ and $u_{yy} = -\operatorname{Re} H''$, whose sum is zero. Corresponding identities hold for v. These checks tie the reconstructed field to the independently known polynomial. The companion records 816 matching value-and-derivative query checks across all 17 fields.

The pair (u, v) is not the quaternion example's two-sheet cover: its relationship is expressed by differential equations and a normalization. The shared question remains precise. What makes two descriptions belong to one object is a map or compatibility rule we can write down and check.

GIGI's disk let a harmonic conjugate be fixed by a normalization. The next experiment changes only the domain: a circular site has an inaccessible middle. Following the angle around that site reveals the extra global condition.

One circuit of the site, two kinds of memory

Imagine a surveyor walking around a circular site whose middle is inaccessible. Each local map records distance from the center and angle from a reference line. Adjacent maps translate correctly. Yet a continuously recorded angle gains 2π on one circuit. The surveyor returns to the same stake with a different accumulated angle. A single position and a record of the journey answer different questions.

This is the global condition hidden inside the phrase "recover the harmonic conjugate." Vygodsky calculates the differentials of angle and log-distance, then treats the conditions for path-independent integration [36, §433, p. 638; §§475–476, pp. 696–700]. Those calculations let us test precisely what changes when the disk in the GIGI construction is replaced by a ring.

Measure coordinates in units of a fixed reference length, so $r = \sqrt{x^2 + y^2}$ is dimensionless. On the annulus $A = \{1/2 < r < 1\}$, set

$$u = \log r, \qquad u_x = \frac{x}{r^2}, \quad u_y = \frac{y}{r^2},$$
$$u_{xx} + u_{yy} = \frac{y^2 - x^2 + x^2 - y^2}{r^4} = 0.$$

Thus u is harmonic everywhere in the site. Its conjugate would have to satisfy

$$dv = -u_y\,dx + u_x\,dy = \frac{-y\,dx + x\,dy}{x^2 + y^2} =: \alpha. \tag{24.1}$$

Locally this is $dv = d\theta$. But on the counterclockwise route $\gamma(t) = (r_0 \cos t, r_0 \sin t)$, $0 \le t \le 2\pi$, direct substitution gives

$$\alpha(\dot\gamma(t)) = 1, \qquad \oint_\gamma \alpha = 2\pi.$$

A single-valued v would return to its starting value, giving integral zero. No choice of additive constant removes this contradiction.

Proposition 24.1 (The missing global condition)**.** *A C^2 harmonic function u on a connected planar domain has a single-valued harmonic conjugate exactly when every closed-path integral of $-u_y\,dx + u_x\,dy$ is zero. When it exists, the conjugate is unique up to a constant.*

Proof. Necessity follows by integrating dv around a closed path. For sufficiency, fix a base point and integrate the displayed form to each destination. Zero closed-path integrals make the answer independent of the route. Differentiation gives the Cauchy–Riemann equations; two such answers have zero difference gradient and therefore differ by a constant. Harmonicity makes the form closed. On a simply connected domain that closedness already guarantees the required zero periods. □

This is why the disk calculation above needs only a normalization. For the log-radius field on this annulus, recovering the angle also needs a record of winding.

A worked lift: one turn and then another. Write $z = e^{s+i\theta}$ with $\log(1/2) < s < 0$ and now allow θ to run over all of $\mathbb{R}$. On this unwrapped strip the logarithm is the ordinary function $L = s + i\theta$. A circuit changes L by $2\pi i$; two change it by $4\pi i$. Remembering its value requires the full integer turn count.

Instead retain $w = e^{L/2}$. One circuit changes its sign; two restore it:

$$w(0) = \sqrt{r_0}, \qquad w(2\pi) = -\sqrt{r_0}, \qquad w(4\pi) = \sqrt{r_0}, \qquad w^2 = z.$$

The resulting map is the genuine double cover

$$q : \{1/\sqrt{2} < |w| < 1\} \longrightarrow A, \qquad q(w) = w^2.$$

Every base point has exactly two roots. The derivative $2w$ never vanishes here, so there is no branch point in either annulus. The bounded holomorphic function $F(w) = w$ lives on the covering annulus; it cannot be selected as a continuous single-valued square root throughout the base annulus [5, Chapter 1].

HARMONIC CONJUGATES / WINDING

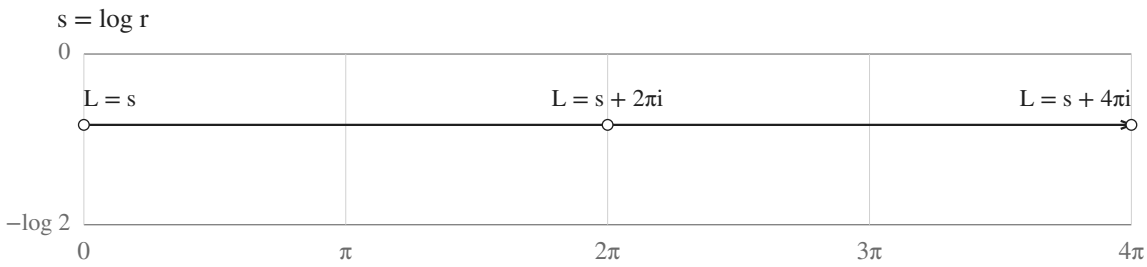
What a circuit leaves in the record

01 Two maps of the same ring



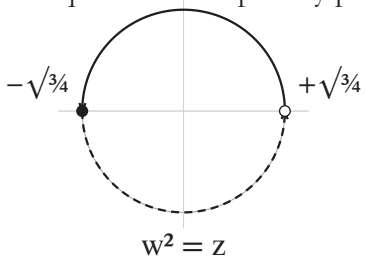
½ < r < 1; route $r_0 = ¾$; positive sense is counterclockwise.

02 The logarithm keeps every turn



Accumulated angle θ; successive returns to the same stake

03 The square root keeps only parity



Circuits	Base z	Root w
0	¾	$+\sqrt{¾}$
1	¾	$-\sqrt{¾}$
2	¾	$+\sqrt{¾}$

One base circuit traverses half the root circle; two complete it.

Figure 24.8: *The same stake, a different record of the walk.* A dashed ray marks each chart's cut; the route at $r_0 = 3/4$ continues using both maps. Their angles agree on the upper overlap and differ by 2π on the lower overlap. The logarithm retains every circuit; the root retains its parity. The root circle has radius $\sqrt{3/4}$, not $3/4$.

The connection with Chapter 16 can also be calculated. Choose the transport law $\dot{a} + (i/2)\alpha(\dot{\gamma})a = 0$. Its curvature is zero, since $d\alpha = 0$, but one circuit multiplies a by $\exp(-i\pi) = -1$. The coefficient $1/2$ specifies that transport rule; the hole alone does not determine a phase. Position, integer winding and sign are three different records. As with maps of a globe, local agreement becomes a global description only after we specify what must agree when the maps close around a journey.

The annulus experiment shows what the disk assumption contributed to GIGI's harmonic construction: local differential agreement must also meet the global conditions of its domain. Laboratory 6 returns to the recorded fingerprint with that distinction in view and makes the completion and harmonic checks reproducible.

How the blackboard graph folds

The movie that put harmonic functions in my head also left a graph problem on a hallway blackboard. In Oliver Knill's numbering, its four tasks ask for the adjacency matrix, the three-step walk counts, and the generating functions for all pairs and for $1 \to 3$ [38]. Picture four stations connected by roads. Stations 1, 2 and 4 form a triangle, and two separate roads run from 2 to 3. How many walks of a given length take you from 1 to 3? You can revisit a station or reuse a road. Each crossing counts as a step, and either road into 3 counts as a separate choice.

Use station order $1, 2, 3, 4$. Entry A_{ij} counts roads from i to j. Matrix multiplication chains those choices, so A^3 counts three-step walks.

$$A = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 2 & 1 \\ 0 & 2 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{pmatrix}, \qquad A^3 = \begin{pmatrix} 2 & 7 & 2 & 3 \\ 7 & 2 & 12 & 7 \\ 2 & 12 & 0 & 2 \\ 3 & 7 & 2 & 2 \end{pmatrix}.$$

Think of the generating function as a tally sheet with one row for each walk length. The count for three steps goes in front of z^3, and the count for length n goes in front of z^n. The matrix geometric series gives $F_{ij}(z) = \sum_{n \ge 0} (A^n)_{ij} z^n = [(I - zA)^{-1}]_{ij}$ as a power series about zero.

Try folding the drawing along the dashed line in Figure 24.9. Station 1 lands on station 4, and 2 and 3 stay put. On the smaller map I give that pair one label, S.

From either station in S, one road takes me to 2 and one takes me to the other station in S. The choices into each group match, so I only need three labels in my tally. Station 3 keeps its own label, so I can still count exactly the walks that end there.

From 2 there are two ways into S and two into 3. In order $S, 2, 3$, the folded matrix is

$$B = \begin{pmatrix} 1 & 1 & 0 \\ 2 & 0 & 2 \\ 0 & 2 & 0 \end{pmatrix}.$$

Put a 1 in row i of P at station i's group and zeros elsewhere. For a group C, $(AP)_{i,C} = \sum_{j \in C} A_{ij}$ counts roads from i into C, including parallel roads. These are exactly the entries of PB [40]. Iterating gives $A^n P = P B^n$. The column for $\{3\}$ gives the count to station 3 itself.

Let a, b, c be the series for walks ending at 3 from $S, 2, 3$. Counting by the first step gives $a = z(a+b)$, $b = 2z(a+c)$ and $c = 1+2zb$. The 1 counts the zero-step walk already at 3. Eliminating b, c recovers the full-matrix answer in Knill and in Horváth, Kórándi and Szabó [38, 42].

$$F_{13}(z) = \frac{2z^2(1+z)}{1 - 7z^2 - 2z^3 + 4z^4} = \frac{2z^2}{1 - z - 6z^2 + 4z^3}.$$

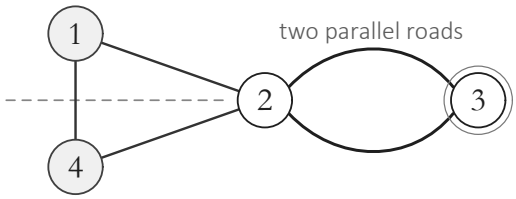
GOOD WILL HUNTING / GRAPH SYMMETRY

The fold keeps every count to 3

Each road costs one step. The folded arrows count choices from one starting station.

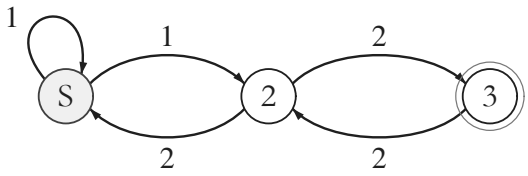
01 Two maps of the same walks

Original stations



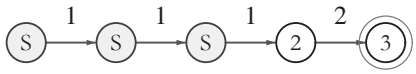
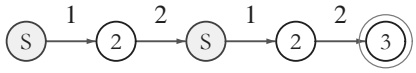
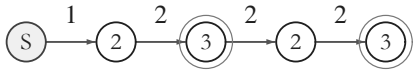
--- Mirror line fixes 2 and 3

Folded states



S groups 1 and 4. Its loop costs one step.

02 Exactly four steps from 1 to 3

Original stations	*Folded route and choices by step*	*Product*	*Walks*
$1 \to 4 \to 1 \to 2 \to 3$		$1 \times 1 \times 1 \times 2$	2
$1 \to 2 \to 1 \to 2 \to 3$ $1 \to 2 \to 4 \to 2 \to 3$		$1 \times 2 \times 1 \times 2$	4
$1 \to 2 \to 3 \to 2 \to 3$		$1 \times 2 \times 2 \times 2$	8
		$2 + 4 + 8$	14

03 Why the count survives

The group map P

	S	2	3
1	1	0	0
2	0	1	0
3	0	0	1
4	1	0	0

Rows 1 and 4 map to S.

$AP = PB \Rightarrow A^nP = PB^n$

The mode that reverses sign

	1	2	3	4
u	+1	0	0	−1
Au	−1	0	0	+1

The pattern changes sign and stays zero at 3.

Its factor $1 + z$ cancels in the walk series.

Figure 24.9: The reflection groups stations 1 and 4. The register pairs each original station sequence with its folded sequence and counts all 14 walks, distinguishing the parallel roads.

Start at station 1 and give yourself exactly two steps to get to 3. You have to go through 2, then pick either of the two roads to 3. With three steps, you have time to go through 4 first. On the folded drawing those routes are $S \to 2 \to 3$ and $S \to S \to 2 \to 3$. The loop is the real crossing from 1 to 4. You have used up a step even though your label on the small map still says S. Either route still gives you two choices for the last road.

Take the four-step coefficient as a check. You have to be at 2 with one step left. You can cross inside S twice, head to 2, and pick either road to 3. That's two walks. Or go straight to 2, take a side trip to either station in S and back, then pick either road to 3. That's four. Or go straight to 2, then go to 3, back to 2, and out to 3 again. Each of those last three crossings offers two roads. That's eight. A walk can pass through 3 early as long as it ends there. Add up the three counts and you get fourteen.

I ran the GIGI walk example for lengths 0 through 128. Both graphs and the rational series give the same integers, starting with $0, 0, 2, 2, 14, 18, 94$.

The half-turn in Chapter 14 suggests checking which patterns change sign. Here 2 and 3 stay fixed, and the swap reverses the odd mode $u = e_1 - e_4$. Direct multiplication gives $Au = -u$. The columns of P span $u^{\perp}$, where A acts as B in that basis. Thus $\det(I - zA) = (1 + z)\det(I - zB)$. The vector e_3 is a column of P, so $(I - zA)^{-1}e_3$ stays in $u^{\perp}$. This is why the count to 3 has no contribution from the -1 mode.

Now suppose someone is waiting for you at station 1 and you have exactly one step left. If you're at 4, you can cross over and meet them. If you're already at 1, every road takes you away. The folded label S tells you which pair of stations you're in. To answer this question you need to know which one.

Meet halfway and count the whole walk

Before I cross a road, I'm already standing somewhere. That's the single-vertex starting case in my paper *Zero Does Not Exist* [39]. Its Peano embedding uses the path P_{n+1} for n crossings. Take $1 \to 2 \to 1 \to 2 \to 3$. There are five places in the record, even though I visit some stations twice. To distinguish walks that take different roads between 2 and 3, record the road at each crossing too.

Split that walk after the second crossing. Each half has three positions. Glue the halves at their shared stop and count it once. Writing just the path part of the paper's construction, $P_3 \oplus P_3 = P_5$. The addition gives the shape of the joined record. The station and road labels tell us which walks we can join and which walks are different.

I can write the two halves on slips of paper. The first says $1 \to 2 \to 1$. The second says $1 \to 2 \to 3$, with one of the two roads into 3 marked. The first slip ends where the second begins. Try them in the other order and I'd have to get from 3 back to 1 before I could use the next slip. The bare paths have the same lengths either way. Once I've written down the stations, the order matters.

A zero-crossing slip has one station on it and no road. If I start at 1, there's one way to be at 1 without crossing a road and zero ways to be at 3. The diagonal of $A^0 = I$ records those starting positions.

Try that with every four-step walk from 1 to 3. At each halfway station, pair every two-step walk into it from 1 with every two-step walk out to 3. Each complete walk has exactly one halfway station.

Halfway station	From 1	To 3	Joined walks
1	2	2	4
2	1	0	0
3	2	4	8
4	1	2	2
Total			14

Look at the first row. I can get back to 1 in two crossings by going through 2 or through 4. Either of those slips joins to a slip that goes $1 \to 2 \to 3$, and I have two choices for its last road. That gives four complete walks with 1 at the halfway mark.

I could cut after r crossings instead. If there are s crossings left, the number of pairs that meet at k is $(A^r)_{ik}(A^s)_{kj}$. Each complete walk gives back its own pair when I cut it at that same place. That's what the matrix product is doing. The two factors count the halves, and the sum runs over the places they can meet.

$$(A^4)_{13} = \sum_{k=1}^{4} (A^2)_{1k}(A^2)_{k3} = 14.$$

Which stations can share a label?

My path paper gave me a record I could join at a stop. My partition paper gave me a way to put several stops under one label. This board lets me ask what happens if I do both. If I write S on the slips where I used to write 1 or 4, can I still join and count the walks to 3?

A connected partition groups stations that can reach one another by roads inside their group [41, 30]. The triangle on 1, 2 and 4 has five partitions. For each one, station 3 can stand alone or join the group containing 2. That gives ten connected partitions out of the fifteen possible groupings.

Grouping 1 and 2 is allowed by that rule. But stand at 1 and ask how many roads take you straight to 3. None. Stand at 2 and you have two. For the walk calculation we need the same number of choices into every group from each station that shares a label. This is the weighted equitable condition [43]. It requires equal total road counts within a group. Station 2 has four and the others have two, so 2 stands alone. From 3 there are two roads into 2, and from either 1 or 4 there's one. That leaves 1 and 4 as the only pair that can share a label, and their counts match.

The standard contraction counts connected groupings that keep an edge's two ends together [41]. I use it in my cycle proof [30]. On this board, give each group a letter and bring stations 1 and 4 together. If they carry the same letter, S carries it once. If they carry different letters, S would belong to two groups at once. Joining those groups would change the grouping you started with. Figure 24.10 shows both cases.

For the connectivity count, draw one link between neighboring stations on the smaller map. The result is the chain S–2–3. Cut or keep each of its two links and you get four groupings. Paths in an original block give walks in the chain after repeated copies of S are dropped. Expand S back into 1 and 4 and their road keeps the block connected. These are exactly the four original groupings that keep 1 and 4 together. The other six split the pair.

For the walk count, B keeps the loop at S and counts outgoing choices per starting station. Stand at either 1 or 4 and there's one road to 2. Stand at 2 and there are two roads into S.

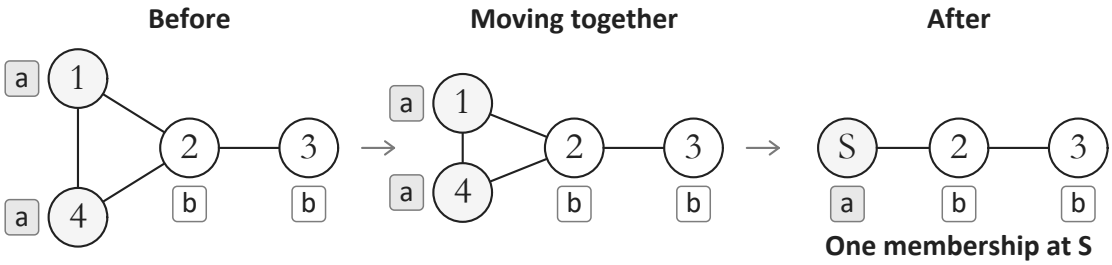
GOOD WILL HUNTING / CONNECTED GROUPINGS

Which groups survive the merge?

Circles name the stations. Letter tags name their groups.

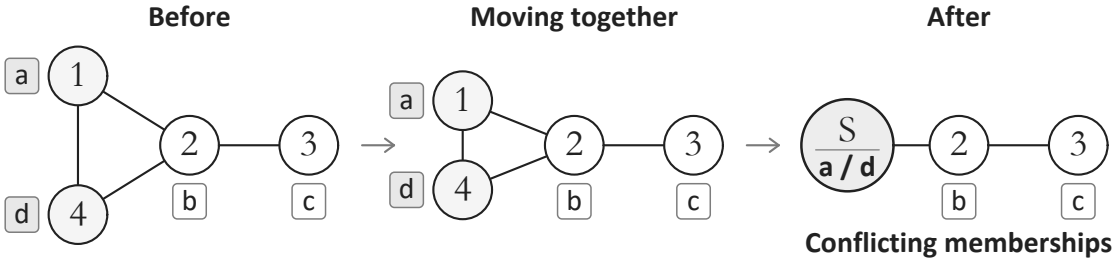
The same letter reaches S

Stations 1 and 4 both carry the group letter a.



Two letters reach S

Stations 1 and 4 start in different groups.



The four groupings that survive

Cut or keep each link of the simple chain S–2–3.

Links kept	On the chain	On the original board
Neither link	S | 2 | 3	{1,4} | {2} | {3}
S–2 only	{S,2} | 3	{1,2,4} | {3}
2–3 only	S | {2,3}	{1,4} | {2,3}
Both links	{S,2,3}	{1,2,3,4}

Four of the ten connected groupings keep 1 and 4 together.

Figure 24.10: The same merge applied to two connected groupings. Links show adjacency in the simple connectivity graph. The block letters show whether the grouping can be read after the merge.

How many original walks fit one folded route?

Look at the top row of Figure 24.10. It keeps 2 and 3 in one group, and we can still read that grouping after 1 and 4 merge. Now ask someone in the group $\{2,3\}$ how many roads lead straight to $\{1,4\}$. Standing at 2, I have two. Standing at 3, I have zero. That group survives the drawing, but one number can't describe both starting stations.

For the walk calculation, go back to the three groups $S, \{2\}, \{3\}$. Fix the actual start at station 1 and write down a folded route. For every crossing, I can take whichever road reaches the next group on my list. The number of choices for the crossing after that is the same whichever station in the new group I reached. That's what it means for this partition to be equitable. For a sequence of groups $C_0, \dots, C_n$ with $C_0 = S$, the number of original walks from 1 that follow it is

$$\prod_{t=0}^{n-1} B_{C_t, C_{t+1}}.$$

At every step, each partial walk has exactly that step's number of continuations. Multiply the choices along one route, then add over the folded routes that end in $\{3\}$. Because 3 has a group to itself, every walk counted this way ends at 3 itself.

If my first slip ends at 1 and the second starts at 4, I still have a crossing to make between them. In the product I always continue from the station the walk actually reached. I can use a shared count for 1 and 4 because they offer the same number of choices into each next group.

Now erase one road between 2 and 3. The ten connected groupings stay the same, but the four-step count drops from 14 to 4. Instead, keep both roads and add a second road from 1 to 2. The count rises to 40, and now 1 and 4 offer different numbers of roads into 2. The little map for counting walks needs the road choices from each starting station, including the crossing that became a loop at S.

Could two states do the job?

Let $a_n = (A^n)_{13}$. In the Hankel matrix, the entry depends on the sum of the row and column indices [44].

$$H = (a_{i+j})_{i,j=0}^{2} = \begin{pmatrix} 0 & 0 & 2 \\ 0 & 2 & 2 \\ 2 & 2 & 14 \end{pmatrix}, \qquad \det H = -8.$$

A d-dimensional linear realization $a_n = c^{\mathsf{T}} M^n b$ with fixed M, b, c factors H into a matrix with rows $c^{\mathsf{T}} M^i$ and one with columns $M^j b$. Hence $\operatorname{rank} H \le d$. Here H has rank three, and $a_n = (B^n)_{S,3}$ attains that dimension. Three is the minimum for a fixed linear realization of this sequence.

What the fingerprint keeps

The fold kept the walk counts to station 3. Now go back to the wine table and make a copy of every row. Each bin has twice as many readings and the table is twice as long. The proportions stay the same, so with the same bin edges, both tables give the same polynomial. Shuffling the rows or moving a reading within its bin also leaves the polynomial alone. If you need the original order, the sample count or the exact readings, keep that information in the record.

The uncertainty disk answers a second question. With just a few values of the polynomial in hand, what can we say at a new probe? The disk contains all values an analytic function bounded by one can take there while matching the supplied values.

Laboratory 6: reproduce a GIGI completion calculation

Start with the thirty-two wine-alcohol bin counts in Table 24.2. Divide them by $N = 6497$ to obtain the coefficients p_k. Check that every coefficient is nonnegative and the sum is one. Evaluate $H(z) = \sum p_k z^k$ using Horner's rule and retain the original counts separately.

Prepare the seven observations. Use $z_0 = 0$ and six nodes $0.65e^{2\pi ij/6}$ for $j = 0, \dots, 5$. Compute values from the full coefficient vector. Compare the results with the printed observation table at its displayed precision. The full histogram is used in this preparation step; the later completion query receives only the nodes and values.

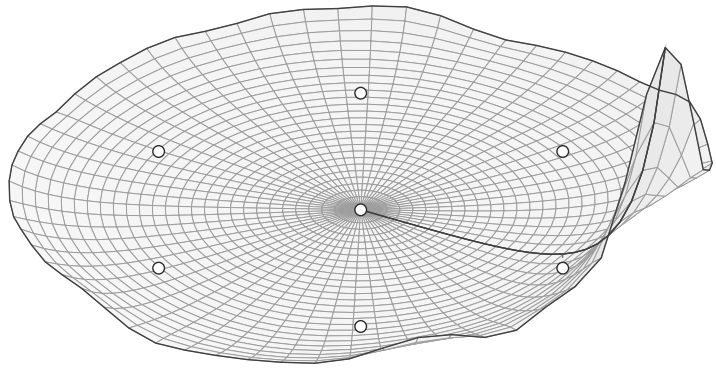
Form the compatibility matrix. Compute $P_{ij} = (1 - w_i\overline{w_j})/(1 - z_i\overline{z_j})$. Check its Hermitian residual and smallest eigenvalue. Small residuals should be interpreted against the numerical precision and matrix scale. Do not make an inconsistent matrix positive by silently deleting a negative eigenvalue.

Calculate two query disks. Use $q_1 = 0.5e^{0.13i}$ and $q_2 = 0.85e^{0.13i}$. Apply the Schur-complement construction in this chapter. Solve linear systems with P to obtain the needed quadratic forms. Compare centers and radii with the retained ledger. The expected final radii, rounded to seven decimals, are 0.0312438 and 0.2753669.

Open the withheld answer only for validation. Evaluate the full polynomial at each query. Both containment ratios $|H(q) - c|/\rho$ should be at most one within numerical tolerance. The withheld value may lie anywhere in the disk.

The prepared data defines a surface of values

Height = |H(z)|; H(z) = Σ $p_k z^k$. This is the constructed analytic fingerprint.



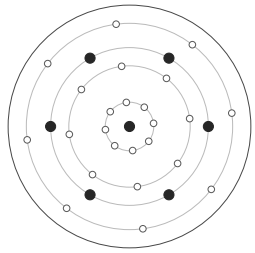
6,497 source records

32 nonnegative coefficients

Σ p_k = 1

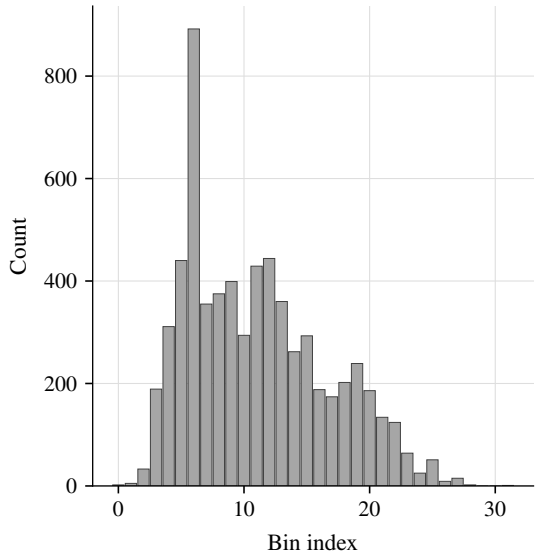
7 supplied function values

24 withheld queries in this field



Filled: supplied / open: withheld locations

A The measured counts define the fingerprint



B Seven supplied values constrain completion

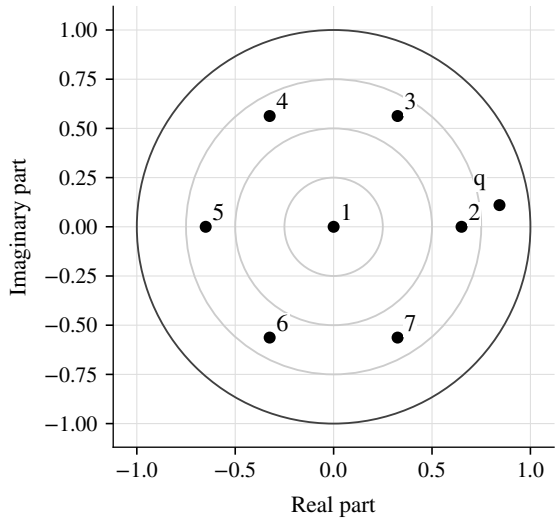
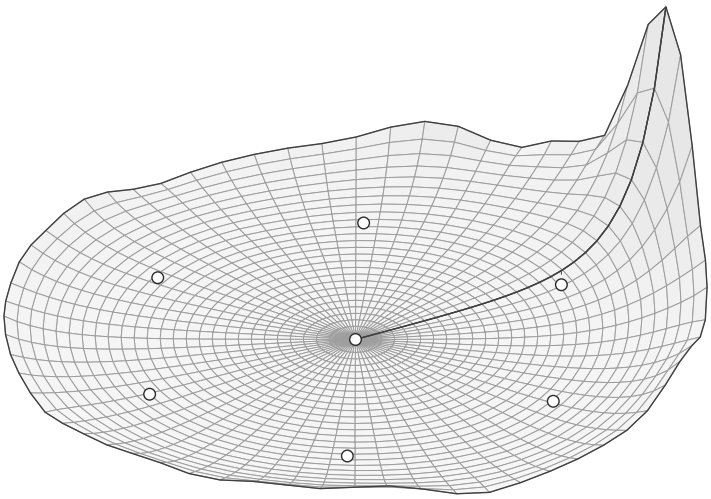


Figure 24.11: *Laboratory record: from counts to a withheld value.* The study continues on the following figure page.

The prepared data defines a surface of values

Height = $|H(z)|$; $H(z) = \Sigma\, p_k z^k$. This is the constructed analytic fingerprint.



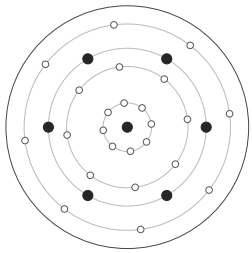
6,497 source records

32 nonnegative coefficients

$\Sigma\, p_k = 1$

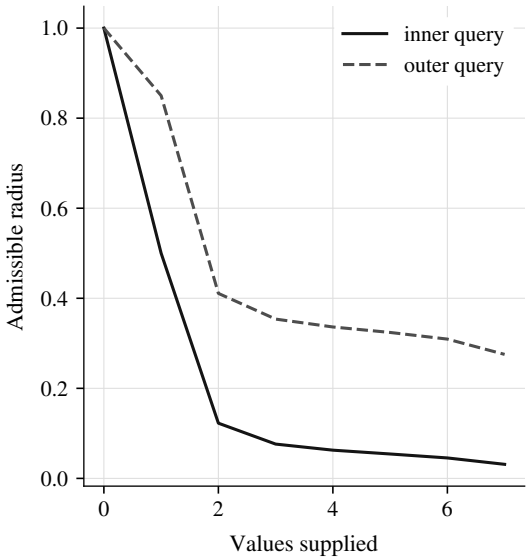
7 supplied function values

24 withheld queries in this field



Filled: supplied / open: withheld locations

C More supplied values shrink the feasible radii



D Exact checks distinguish feasible from impossible

Inner final radius: 0.0312438

Outer final radius: 0.2753669

H(0)=0, H(1/2)=1/2 → f(z)=z

H(0)=0, H(1/2)=.9 → infeasible

Same-size edits must change dependent results.

Figure 24.11: *Laboratory record: from counts to a withheld value (continued).* The retained wine-alcohol histogram and seven supplied probe locations define the input. The two query radii are plotted against the number of supplied values; the register states their final values and exact feasibility checks. The full polynomial provides the withheld validation values, not inputs to the completion solve.

Supply the observations one at a time. For each prefix, reconstruct the two disks. Added exact constraints can only shrink the feasible sets. For Euclidean disks, nesting can be tested by $|c_{k+1} - c_k| + \rho_{k+1} \leq \rho_k$, allowing a declared numerical tolerance. A smaller radius alone does not prove nesting if the center has moved.

Run a known-answer control. Use $f(0) = 0$ and $f(1/2) = 1/2$. Schwarz's lemma forces $f(z) = z$. This is a singular interpolation case: the feasible query value is unique. Test it with a method that handles that case, or verify it analytically rather than applying the positive-definite inverse formula outside its assumptions.

Run a contradiction control. Change the second value to 0.9. The two-node Pick determinant is $-56/75$, so the data are infeasible. A routine that returns a plausible-looking disk has failed to enforce the model. Restore the original observations and verify that the original results return.

Run an evidence-dependence control. Multiply every supplied value by 0.8 while keeping the nodes and count of observations unchanged. This produces another coherent bounded function, $0.8H$, but its completion problem is different. Recalculate rather than reusing a result keyed only by record count. Keep the altered problem separate from the retained empirical study.

What to submit. Save the input counts, bin convention, node order, observation values, matrix diagnostics, disk parameters, containment checks and controls. The experiment reproduces inference about this bounded analytic fingerprint. It neither reconstructs all original wine measurements from seven probes nor predicts wine quality from alcohol alone.

The wine records now answer three distinct preservation questions. Their bin proportions define the chosen distribution; the polynomial represents those proportions; partial function values constrain the possible answer at another probe. The withheld-value checks test that last calculation against the known fingerprint. Harmonic compatibility and the domain determine when two field descriptions belong to one global function. At each stage, the evidence determines particular features of one object and leaves other features unresolved. Those remaining distinctions motivate the final chapter.

Practice with the construction

24.1. Create two four-reading datasets that have identical proportions in bins $[0, 1)$ and $[1, 2]$ but different means. What does their shared fingerprint preserve?

24.2. At the inner query, use the final table to verify $|H(q) - c| \leq \rho$. Explain why choosing the disk center as a representative does not prove that it equals $H(q)$.

Chapter 25

The questions that follow

Chapter 24 used samples of the wine-alcohol fingerprint to bound its values at new points. A separate disk-map example showed how values that are possible one at a time can fail to fit a single function. When do local reports determine a common object? Which measurements, histories and decisions must other descriptions of it preserve?

What we have built

The map and globe have accompanied a sequence of explicit constructions. We can now use their common question to organize what remains to be done.

The constructions share five ingredients.

- A question: when are two things really the same?
- A hidden world of entities u, observed through noisy channels.
- Two ways to model sameness: as translators between views, and as geometry on a manifold.
- A bridge between those ways: smooth chartability, specified Riemannian realizations, and quantitative comparison maps.
- An error budget transfer theorem that says: guarantees do not live in a single coordinate system.

In one image, the story looks like this:

- There is a *territory*: the semantic sameness structure S.
- There are many *maps*: translator-based realizations and manifold-based realizations.
- There is an *atlas*: charts and translators that overlap, stitch, and sometimes wrinkle.
- There is an *inspector*: the audits that tell us when the stitching is good enough to trust.

Underneath all of this is one claim: if we are careful about how we stitch—if the quilt lies smoothly enough over the globe—then error, risk, and responsibility can be compared across representations. We can say, with a straight face, that a bound proved in one coordinate system still means something when the system is actually deployed in another.

The final chapter follows the remaining requirements into further constructions, then asks what they imply for three broader questions:

1. What does this perspective offer for *machine learning systems* in practice?
2. What does it say—and not say—about *fairness and accountability*?
3. What does it illuminate about *sameness itself*, as a mathematical and human idea?

The finite examples below test the outstanding requirements, compatibility of local reports, additional geometric structure, retained history and computational resources. Their consequences for deployment and responsibility then bring us back to why the invariant had to be named at the beginning.

ML systems as moving coordinate frames

Most modern machine learning systems are really systems of systems. A decision like "approve this loan," "flag this transaction," or "unlock this phone" is rarely made in a single space by a single model. It is the product of many steps:

- Data is collected in one format, cleaned in another, normalized in a third.
- Features are built, sometimes in a warehouse, sometimes on the fly.
- Models are trained, updated, distilled, compressed, deployed in variants.
- Outputs are fed into downstream aggregators, heuristics, and policies.

As the pipeline changes, its descriptions of the world change too. A schema update may recode the same observation. A new embedding may keep some of its features and discard others.

An encoder replacement or a post-processing step can alter the decision even if it changes only one part of the pipeline. The geometry of sameness asks whether it preserves the comparison on the same underlying structure.

Representation changes as chart transitions

When a replacement preserves the same modeled state through an invertible smooth coordinate conversion, it supplies a chart transition. A lossy encoder or a changed task requires its own comparison map and uncertainty allowance. The earlier constructions let us ask which kind of replacement a software release actually makes.

The smooth chartability conditions from Chapters 9 and 10 can be read as engineering desiderata:

- **Smoothness:** small changes in the old representation should correspond to small changes in the new one.
- **Bounded distortion:** distances should not be wildly inflated or collapsed.
- **Round-trip consistency:** mapping forward and back should return us close to where we started, on in-regime data.
- **Cocycle consistency:** different migration paths should agree.

These are not just mathematical niceties; they are checks we can run whenever we change a representation in a live system. They are ways of asking: "Are we still talking about the same thing, or did we silently change the meaning of sameness when we refactored?"

Example: a telemetry representation upgrade. An old representation stores a dense vector; a new one stores its nonzero coordinates and indices. If decoding returns the original vector exactly, any deterministic comparison of that decoded vector is unchanged. If the new format also quantizes values, use the explicit distance-error bound and threshold margin before transporting a guarantee. The AMC worked example calculates both cases. A byte-format upgrade and a geometric model change have different proof obligations.

Error budgets as invariants across releases

In software engineering, invariants are the properties we promise never to break: this counter never goes negative, this lock is always acquired before that one, this field is unique.

Error budget transfer lets us treat certain detection guarantees as *approximate invariants* across representation changes. Given bounds on the relevant operating population, an upgrade could support claims such as:

- "This upgrade will not increase the false positive rate by more than 1.5%."
- "The abstention rate in this corridor will remain within 3% of its previous level, modulo known slack."
- "Our fairness constraint, expressed as an error budget difference between groups, will not degrade beyond a specified tolerance."

These illustrative promises require numerical evidence for the comparison-failure probability and threshold-band mass on each stated population. Theorem 20.1 supplies the resulting bound when those quantities and the reference error bound have been established. The audit records which part was proved, stipulated or estimated.

Debugging by changing coordinates

When a system fails, we often respond by staring harder at the representation we already have. Sometimes the right move is to change coordinates: embed the same data in a new space where the failure is easier to see.

In this book, that took the form of:

- Moving between a translator description and a specified manifold to inspect the geometry (AMC or ICARUS).
- Moving from a manifold back to a translator graph to see where paths stop composing (AMC, *PRISM*).

If a defect changes a quantity that an exact coordinate map preserves, it persists in both descriptions. A different view may make that quantity easier to inspect. A lossy representation can erase the evidence, so check the relevant preservation claim before interpreting what you see.

When do local reports determine one object?

My Field Equations of Semantic Coherence develops a wider research program around completion, compatibility, retained history and resource limits [18]. The geometry of sameness raises questions about all four. Each needs its own objects, assumptions and evidence. The broader program supplies conjectures and research directions.

The map/globe question now becomes operational: how much of the common object is determined by the descriptions available, how stable is that determination, and what must a computation retain to preserve it?

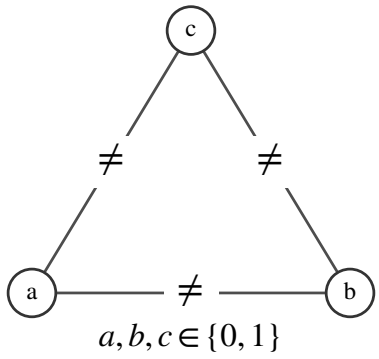
The three clocks in Chapter 10 had to fit their reported offsets to one time scale. We start with an even smaller compatibility test. Then we add further requirements: the geometry may have to support a prescribed operation, and a prediction may have to preserve the order of observations.

Local agreement and global completion

Put three two-position switches on a test bench and write their settings as $a, b, c \in \{0, 1\}$. Three pairwise checks require $a \neq b$, $b \neq c$ and $c \neq a$. Each check can pass when run by itself. All three cannot pass on one setting: the first two imply $a = c$, contradicting the third. The drawing checks all eight settings.

Now replace each pair constraint by a compatible restriction of one common assignment. In a sheaf-style construction, matching local sections can be glued when the stipulated overlap and gluing conditions hold. One must define the local data, restriction maps and the compatibility

A *Every pair can differ*



B *No binary triple satisfies all three*

abc	*Failed constraint*
000	ab, bc, ca
001	ab
010	ca
011	bc
100	bc
101	ca
110	ab
111	ab, bc, ca

Each pair can differ; three binary values cannot all differ.

Figure 25.1: A binary compatibility problem checked exhaustively. Every pair can be assigned different values, but all eight global assignments violate at least one edge condition. Pairwise feasibility therefore does not establish global compatibility.

being tested before invoking that language. A nonzero numerical loop residual is a useful diagnostic in a specified translator system, but is not automatically every possible global obstruction.

The GIGI interpolation chapter provides a different, analytic completion problem. There the Pick matrix states the joint compatibility criterion for the chosen function class. A positive-semidefinite condition and a small explicit witness carry more information than a vague statement that the observations fit together.

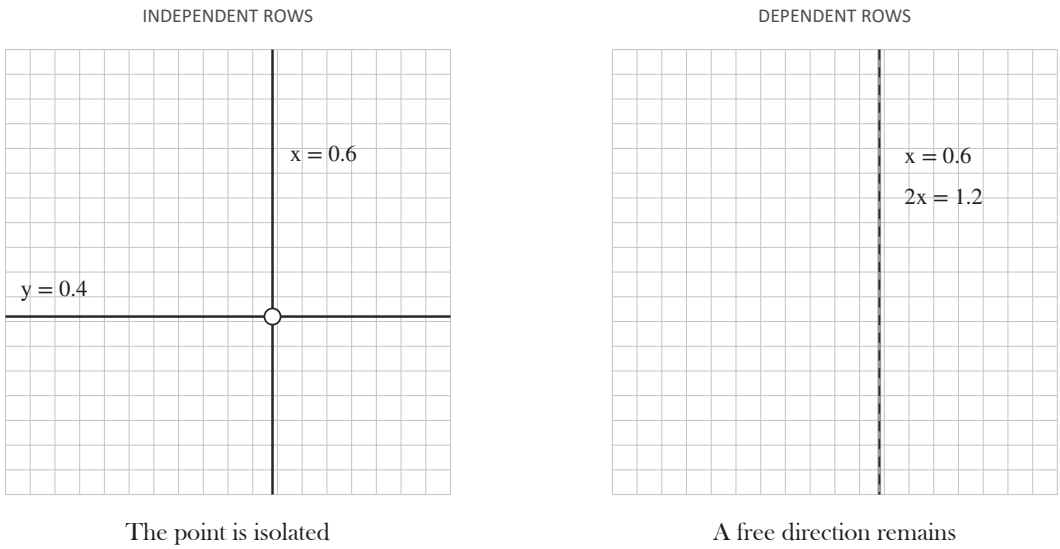
A finite completion problem can be solved all the way down

Return to the three clocks. Their unknown offsets are x_1, x_2, x_3. Comparing two clocks gives a difference, and calibrating one against a reference fixes the common origin. Write these constraints as $Ax = b$. Existence requires b to lie in the column space of A. If a solution exists, every solution has the form $x_0 + v$ with $v \in \ker A$. Uniqueness requires a trivial kernel.

Run one more check and ask what it tells you about the clocks. A check that repeats an exact offset leaves the same possible settings. An independent exact check can pin down a setting that was free. If the new check contradicts the old ones, no setting fits them all. Repeating a noisy reading can sharpen an estimate under the right noise model. It doesn't add a new exact equation.

Constraint directions determine what remains free

A repeated equation adds a record, but it does not add an independent direction.



The repeated constraints coincide exactly. Their shared line leaves y undetermined.

A Independent readings isolate one point

B A repeated equation leaves a free direction

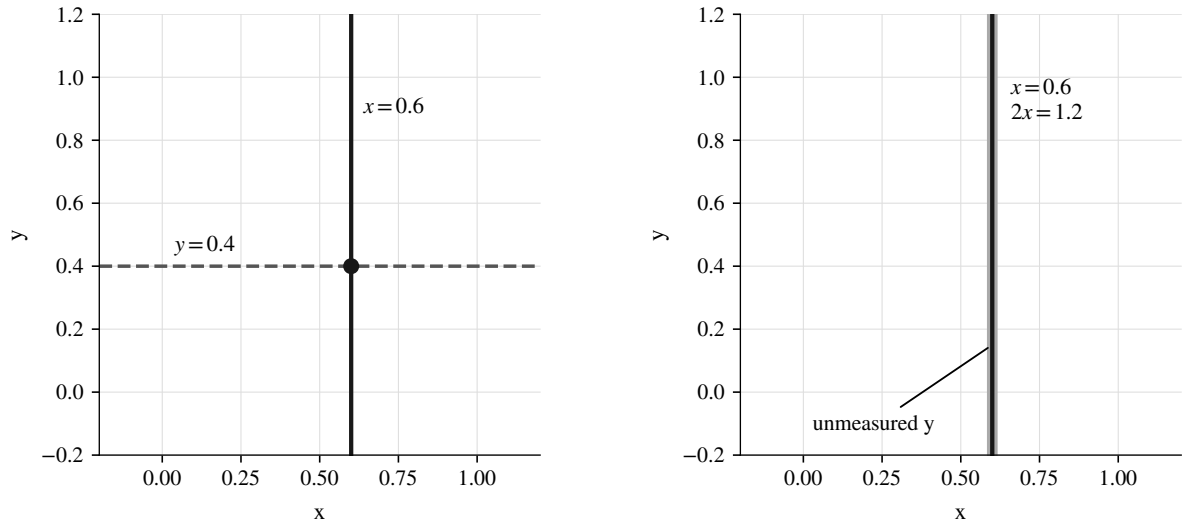
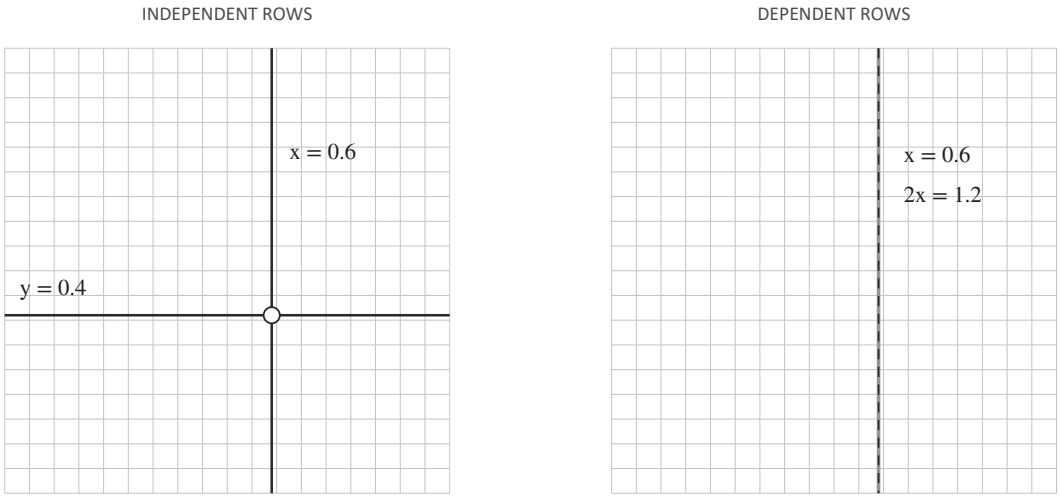


Figure 25.2: *More rows do not always mean more information.* The study continues on the following figure page.

Constraint directions determine what remains free

A repeated equation adds a record, but it does not add an independent direction.



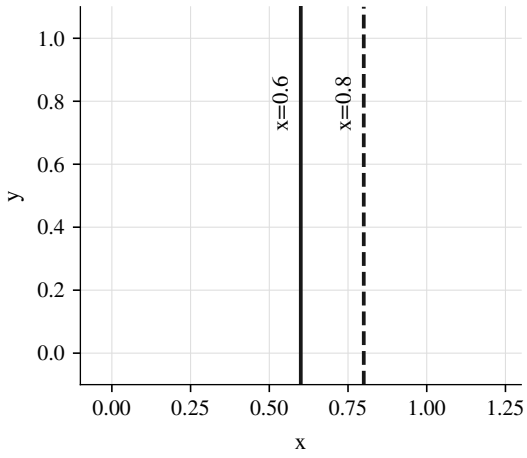
The point is isolated

A free direction remains

The repeated constraints coincide exactly. Their shared line leaves y undetermined.

C Incompatible readings leave no solution

D Independent noisy repeats reduce uncertainty



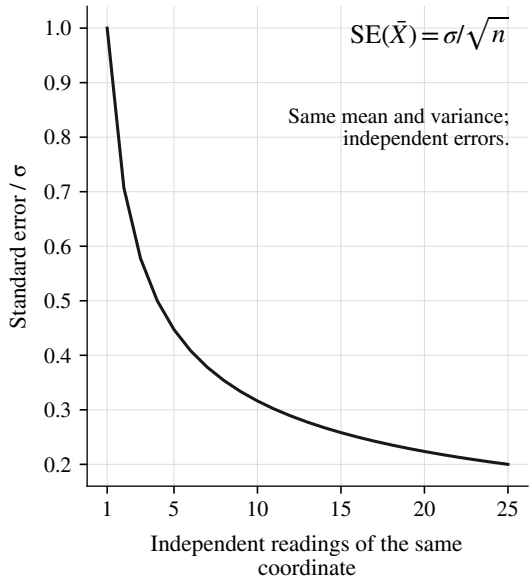


Figure 25.2: *More rows do not always mean more information (continued).* The readings $x = 0.6$ and $y = 0.4$ intersect at one point. Replacing the second reading by $2x = 1.2$ leaves an entire line; adding the incompatible reading $x = 0.8$ leaves no solution. Independent noisy repeats answer a different question: under the stated common-variance model, the standard error decreases as $\sigma/\sqrt{n}$.

What must another description preserve?

Once a common object has been identified, its descriptions may still differ in what they preserve. Start with a decision based on a distance. Chapter 20's transfer theorem is exact for the stated allowance and population. A linear rate requires a bounded-density or comparable margin condition. Further work can sharpen the inputs:

- Find sharper constants for a particular comparison.
- Use higher-order curvature terms to improve the distance allowance.
- Establish the relevant distance bounds in function spaces.

Natural questions:

- Can we derive second-order distance allowances that track curvature more precisely before applying the transfer theorem?
- Can we extend the framework to infinite-dimensional manifolds (for example, spaces of functions, distributions, or policies)?
- Can we characterize exactly when the approximate categorical equivalence of Chapter 13 lifts to stronger notions of equivalence?

These are technical questions, but they matter. In some domains, the slack we can tolerate is tiny, and first-order approximations may not be enough.

A completed state can carry more than one kind of structure. The next example asks whether the ruler and a prescribed quarter-turn operation agree. This is the same compatibility obligation that joined coordinate changes to measurement earlier in the book, now tested on an additional geometric operation.

A metric and a quarter-turn can be compatible

The later Kähler work asks how a metric, a complex structure and transport can coexist [32]. A concrete two-dimensional model makes the compatibility conditions visible. On $\mathbb{R}^2$ with Euclidean metric, define

$$J = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

Then $J^2 = -I$ and $J^T J = I$. It rotates each tangent vector by a quarter-turn while preserving length. The two-form $\omega(u, v) = g(Ju, v)$ is the oriented area form $dx \wedge dy$ and is closed.

On a conformal surface metric $g = \lambda(x, y)^2 I$ with $\lambda > 0$, the same local J preserves g, and $\omega = \lambda^2 dx \wedge dy$. In two real dimensions every smooth two-form is closed because there are no nonzero three-forms. With the compatible surface complex structure, this gives the elementary Kähler picture. In higher dimensions, closedness and integrability are substantive requirements; a collection of quarter-turn matrices alone does not establish them.

The disk ruler expands toward its boundary

Height = conformal factor $\lambda = 2/(1 - |z|^2)$, shown only for $|z| \le 0.9$.

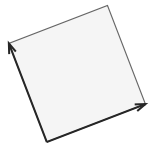


Metric: $g = \lambda^2 I$

J is a quarter-turn.

$J^2 = -I$

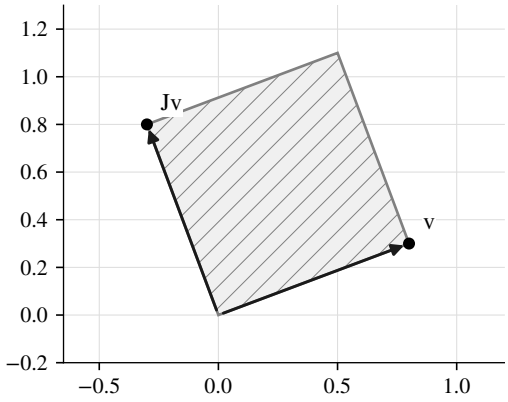
$J^T g J = g$



The same structure relates length and

A A quarter-turn links length and oriented area

B The disk ruler expands toward the boundary



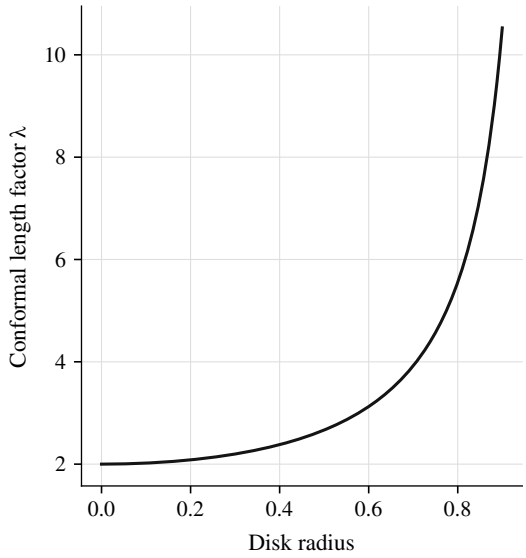
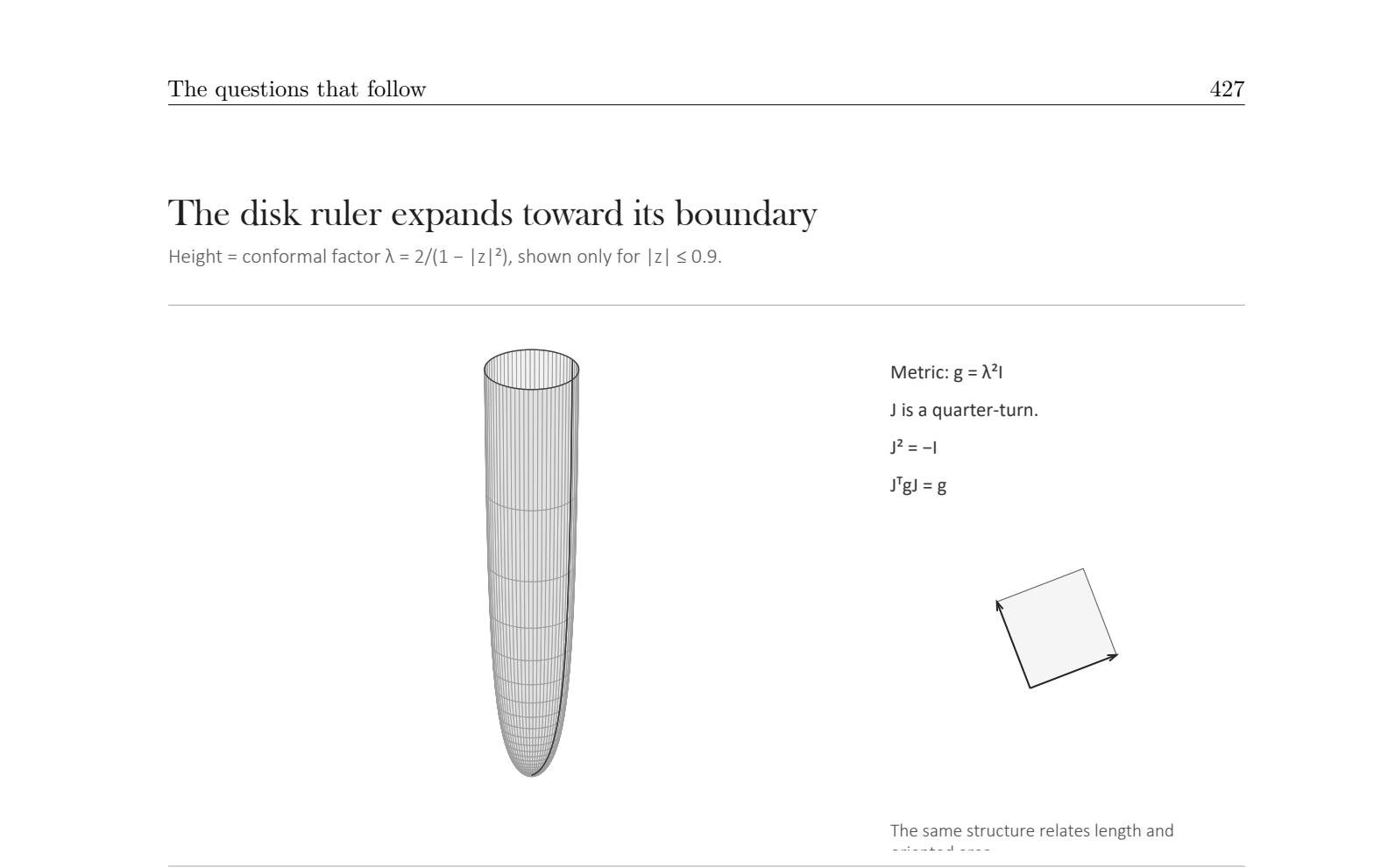


Figure 25.3: *Length, quarter-turn and area can share one geometry.* The study continues on the following figure page.



C The compatibility identities agree

D Higher dimensions require additional structure

$J^2=-I$

$J^TgJ=g$ for $g=\lambda^2 I$

$\omega(u,v)=g(Ju,v)$

$\omega=\lambda^2\, dx\wedge dy$

Every smooth surface 2-form is closed.

An oriented surface has a compatible complex structure.

In higher dimensions, integrability and closedness need additional checks.

A matrix quarter-turn alone is not the full Kähler structure.

Figure 25.3: *Length, quarter-turn and area can share one geometry (continued).* A vector and its quarter-turn span an oriented area. The disk conformal length factor is plotted separately, and the registers give the compatibility identities and their two-dimensional scope. Higher-dimensional complex and Kähler structures require the additional checks stated in the text.

Adversarial and strategic settings

Most of this book assumes that errors arise from noise, drift, and modeling limitations—not from an adversary actively trying to fool the system. In adversarial settings:

- Upstream channels may be manipulated to break chartability selectively in some regions.
- Realization maps Φ might be well-behaved on benign data but fail under targeted attacks.
- The very notion of benign evolution can be exploited (for example, by constructing paths that move along the edge of what the system considers "still the same").

Adversarial geometry of sameness is largely open. What does it mean to have a representation that is robust to someone seeking to break the equivalence between translator and manifold views? Can error budgets be made robust to worst-case, rather than typical-case, chartability violations?

Fairness-specific generalizations

While I have gestured toward fairness and accountability, there is much more to do:

- Error budgets could be refined to track not just global error, but disparities between groups or regions of the manifold.
- Chartability diagnostics could be stratified by protected attributes, asking explicitly where the quilt is wrinkled for specific populations.
- Guarantees could be framed as obligations not just on average error, but on invariance of error budgets across representation changes.

In short: we can ask not just "are these two detections the same in different realizations?" but "does this realization change affect different people differently?"

Dynamics, causality, and intervention

This book treats benign evolution as an abstract path in a latent space. In many real systems, the path is governed by dynamics or policies:

- A patient's state evolves under treatment.
- A system's telemetry evolves under control.
- A financial account evolves under user and system actions.

Integrating the geometry of sameness with causal and dynamical models is an open frontier:

- How do interventions affect the geometry of benign evolution?
- Can error budgets be conditioned not just on passive observation, but on actions taken in the system?
- Can sameness structures be used to design control policies that are robust across representations?

I do not have clean answers here—only a sense that this is where many interesting problems live.

Order is part of the experiment

Two sensors report whether a trolley is parked at the left or right end of a rail. Suppose it stays parked, with prior odds 1, and the reports are independent given its location. One report has

likelihood ratio 3 and the other 1/2. Bayesian updating gives final odds 3/2 in either order because both multipliers describe the same fixed state.

If the trolley may move to the other end between reports, the update must include that transition. Take probability vector $(1/2, 1/2)$, first likelihood vector $(1, 3)$, a transition that flips the state with probability 0.2, and second likelihood vector $(2, 1)$. The first update gives $(1/4, 3/4)$; transition gives $(0.35, 0.65)$; the second update gives $(14/27, 13/27)$. Reversing the reports around the transition gives $(1/3, 2/3)$. These are different observation histories, not two conflicting answers to one timeless evidence set.

Suppose someone hands you both trolley reports without their timestamps. While it stays parked, you can multiply the two likelihood ratios either way. If it may move between checks, ask which report came last before predicting where it is. That is the question in ordered-update and causal-state research [25]. For two linear updates on the same space, I can compose them in both orders and subtract one result from the other. That difference, called the commutator, tells me whether order matters.

What can we verify with the evidence and resources available?

The framework assumes we can:

- Sample pairs and triples from in-regime data across channels.
- Compute Jacobian spectra or finite-difference approximations for translators.
- Fit and validate realization maps Φ between translator spaces and manifolds.
- Run disagreement experiments to estimate slack terms $\Delta_\bullet(\delta)$.

In production systems processing millions of decisions per second:

- Computing full Jacobian spectra for every translator may be infeasible.
- Sampling paired observations across all channels may require complex and expensive joins.
- Fitting Φ may need to happen online as data drifts, under latency constraints.
- Monitoring chartability constants in real time competes with other monitoring budgets.

Open engineering questions include:

- Can chartability diagnostics be approximated with sketches or lightweight sampling?
- Can we develop cheap proxies for Jacobian condition numbers that still flag dangerous regions?
- How do we integrate chartability monitoring into existing MLOps pipelines (feature stores, model registries, A/B testing frameworks)?
- What is the acceptable computational overhead for these audits, and how do we prioritize which regimes to monitor most closely?

These are not obstacles to the framework; they are part of making it real. The theory is only useful if the audits are cheap enough to run and easy enough to interpret that teams will actually use them.

Capacity needs a unit and a task

The field-equation program proposes capacity relations such as $C = \tau/K$, relating a declared tolerance to a curvature-related cost. To use such a relation, first define what C counts or measures, which curvature quantity K means, and what units make the expression meaningful. A scaling

conjecture is most useful when it specifies an experiment that could distinguish it from other scalings.

The sphere calculation in Chapter 7 gives a proved nearby example with a particular task. If the tolerated relative error in circumference is τ, the sufficient radius condition is $s^2 \leq 6\tau/K$ in the stated range. The radius scales as the square root of τ/K, while an area scale can be proportional to τ/K. The difference comes from what is being measured. This does not prove a universal completion-capacity law; it shows how to make one proposed relationship concrete enough to derive and test.

Flat geometry also needs care. As $K \to 0$, a local curvature penalty can disappear while memory, observation count, numerical precision and identifiability remain limiting. A capacity claim must account for the relevant limits instead of reading a formal division by zero as an unlimited machine.

Where to start?

If I had to point future work in one direction, it would be this: *adversarial and strategic robustness* combined with *fairness-specific generalizations*.

The framework as presented assumes benign noise and drift. But real deployments face:

- Adversaries who exploit representation boundaries.
- Populations who game the system by mimicking benign evolution.
- Regulators who demand group-stratified guarantees.

Can we derive error budget transfer theorems that remain valid under worst-case chartability violations? Can we bound how much an adversary can degrade error budgets by selectively breaking smooth chartability in targeted regions? Can we make fairness constraints themselves invariant across representation changes?

The mathematical machinery is here. The next step is making it robust to the fact that the world fights back.

Fairness, power, and accountability

Every time a system decides that two records, two faces, two transactions, two histories are "the same," it is taking a stand in a world full of asymmetry and power. It is deciding who is linked to what, who inherits which risk, who is allowed to move cleanly through a bureaucracy and who is stopped because some database somewhere decided they were "close enough" to a different person.

A geometry of sameness comes with responsibilities.

Sameness as a normative choice

A claim of semantic sameness depends on three choices.

- with respect to a choice of latent entities u,
- with respect to a domain-specific notion of benign evolution,
- and with respect to thresholds we choose for "same enough."

For faces, we may decide that aging is benign but swapping people is not. For viruses, we may decide that certain mutations are benign and others constitute a new variant. For transactions, we may decide that some rearrangements within a household are benign while others trigger fraud.

These are not choices the math can make for us. The math does not tell us *what* should count as benign evolution. But it *does* tell us when we have changed our minds without admitting it—when a representation change has silently shifted the boundary between "same" and "different."

What the framework can do is insist that once we have made those choices, we are not allowed to quietly change them by swapping encoders or translators without noticing.

Audits as proof obligations

Fairness work in machine learning often focuses on aggregate statistics: group-wise error rates, calibration plots, disparity metrics. Those are important. But there is another layer: *representation audits*.

If we claim that a system respects a cast of individuals across featurizations—that a person's identity is not suddenly reinterpreted when their data is ported from one system to another—then chartability becomes a kind of fairness proof obligation:

- Do round-trip errors spike for certain subgroups?
- Do Jacobian condition numbers blow up at specific intersectional slices of the population?
- Does $\hat{\varepsilon}_{\text{loc}}$ stay small in some neighborhoods of X_T but balloon in others?

We can ask: who lives in the wrinkled parts of the quilt? Whose identities get stretched and torn when we drape our representation over the manifold we claimed to be modeling?

Example: age-stratified chartability in face recognition. A representation can perform differently across operational groups. Define the group and the relevant conditional error before pooling results. For example, an illustrative audit might show a narrow threshold band in one camera regime and a much larger band in another. The same global mean distance error would then conceal different decision risks. Report subgroup sample sizes, coverage and uncertainty, together with the overall result. This is a requirement on evidence, not a claim about a specific identity deployment.

Responsibility under representation drift

Real-world systems drift. Data distributions shift. New sensors are added. Old schemas are retired. In practice, deployment looks less like a static map and more like a time-lapse of maps being redrawn while people are still driving on the roads.

The geometry of sameness offers a language for responsibility under drift:

- If you change representations but keep your error budget guarantees on paper, you have an obligation to demonstrate that chartability still holds.
- If you know that chartability fails in certain regions, you have an obligation to mark those regions as out-of-regime and adjust your claims accordingly.
- If you rely on invariance across representations to argue that a system is fair or safe, you have an obligation to show that the invariance is real, not just aspirational.

These are not mathematical theorems; they are ethical and regulatory norms that the theorems can support. They turn a vague promise—"we did not change the semantics"—into something testable.

What is sameness, really?

Behind the charts, the manifolds, and the error budgets is a more basic question that has been tugging at me since the beginning: what does it mean for something to be the same?

Mathematically, we made a choice: we modeled sameness as living in an equivalence relation on pairs of observations, and we lifted that to an underlying hidden variable u. But humans do not experience identity as a crisp equivalence relation. We experience it as a path.

The path view of identity

I think of sameness along a path this way.

> Two things are the same when there is a gentle path between them that never leaves the space of that thing being itself.

For faces, that path might be aging, changes in lighting, changes in hairstyle. For viruses, it might be mutations that preserve antigenic identity. For systems, it might be configuration changes that do not alter semantic behavior.

We used geodesics and benign evolution to turn that intuition into math. But I think it is worth naming that the picture we are really chasing is a story about continuity: about which changes we agree to treat as "still me" and which ones break the thread.

The framework here formalizes that intuition with geodesics and benign evolution. But the *choice* of what counts as a "gentle path"—which changes preserve identity and which break it—remains a human, social, political choice. The math just makes that choice explicit and testable.

In a world where many of us have had to fight to have our own identities recognized as continuous—across time, across categories other people thought were fixed—I am suspicious of any system that claims to detect sameness without ever articulating its view of benign evolution. "Same according to what path?" is not just a geometric question. It is a human one.

Sameness across contexts

Semantic sameness is always contextual. A pair of transactions may be the same from the point of view of a risk model, but not from the point of view of a customer. A pair of scientific measurements may be the same in one theory and distinct in another.

What the framework here offers is not an answer to "what is really the same?" but a way of pinning down: "given this notion of sameness, and these observation channels, and this decision rule, here is how consistency behaves across representations."

In other words: the geometry of sameness is not a metaphysics. It is a bookkeeping system for a promise we make in context.

Shadows, not essences

It is tempting, when drawing identity manifolds and equivalence classes, to imagine that we are getting closer to some pure essence of things—that the manifold is the "true" space and the embeddings are just shadows.

I do not believe that. The manifold is also a shadow; it is just a shadow that is particularly convenient for the kinds of arguments we want to make. Translator graphs, manifolds, and error budgets are all ways of tracing the outline of something we cannot write down fully: the structure

of a world where different things hold together or come apart as we look at them from different angles.

The safest way to use geometry here is not to confuse the picture with the territory, but to use the picture to keep ourselves honest about how much of the territory we are actually seeing.

A useful next question has a discriminating experiment

Choose an experiment in which competing explanations predict distinguishable records. If a proposed capacity depends only on curvature, compare two models with the same curvature but different observation noise or available precision. If a scalar score is claimed to retain all relevant information, construct two transformations with the same score but different task behavior. If a completion engine is claimed to depend on observations, edit values without changing their count.

The worked constructions keep the newer research tied to objects, maps, rulers and inspectable records.

What survives translation

The central task remains the geometry of singular objects and their descriptions. Name the object. Identify the allowable changes. State the maps. Put a ruler on the relevant space. Calculate what survives translation. Track the path when the path matters. Carry the uncertainty into the decision rather than dropping it at the final line.

The translator account follows changing representations; the geometric account follows the object, its distances and admissible paths. Exact constructions prove their agreement. Approximate constructions quantify its cost. The double-cover examples distinguish a state from its representatives and the history of its motion. The experiments make those comparisons executable. The next question can begin with an object, a calculation, and a record someone else can inspect.

Practice with the construction

25.1. Enumerate the eight binary triples and verify that none satisfies all three inequality constraints.

25.2. Reproduce both ordered Bayesian calculations. Which assumption in the timeless multiplication example has changed?

Appendix A

Working foundations and further derivations

The calculations below develop the proofs using linear algebra, calculus and probability.

Quadratic forms and the shape of a ruler

A real symmetric matrix G defines the quadratic form $q(v) = v^{\mathsf{T}} G v$. The spectral theorem supplies an orthogonal matrix Q and real eigenvalues λ_j with $G = Q\,\mathrm{diag}(\lambda_j)Q^{\mathsf{T}}$. Writing $w = Q^{\mathsf{T}} v$ gives

$$q(v) = \sum_j \lambda_j w_j^2.$$

Thus G is positive definite exactly when every eigenvalue is positive. If $m \le \lambda_j \le M$, then $m\|v\|_2^2 \le q(v) \le M\|v\|_2^2$. This is the linear algebra behind metric and norm comparison.

The set $v^{\mathsf{T}} G v = 1$ is the unit ellipse or ellipsoid of that ruler. Its principal semiaxes have lengths $1/\sqrt{\lambda_j}$. A large eigenvalue means a small coordinate displacement already counts as unit length. If an eigenvalue tends to zero, the ellipsoid stretches without bound in that direction, expressing the loss of sensitivity seen in the plate-sensor example.

For two positive-definite matrices G and H, the optimal constants in $aG \preceq H \preceq bG$ are the extreme eigenvalues of $G^{-1/2} H G^{-1/2}$. Set $w = G^{1/2} v$ to write the ratio $v^{\mathsf{T}} H v / (v^{\mathsf{T}} G v)$ as the Rayleigh quotient of that matrix. Turn the same little arrow under both rulers to see where their squared readings disagree most. The ratios survive a change of grid when both rulers go with it. The condition number of one coordinate matrix can change.

Pulling a curve through a nonlinear map

Let $T : U \to V$ be a smooth map between Riemannian coordinate domains and let $\gamma : [a, b] \to U$ be piecewise smooth. The chain rule gives $\dot\eta = DT_\gamma \dot\gamma$ on each smooth piece for $\eta = T \circ \gamma$. If $G_U = DT^{\mathsf{T}} G_V DT$, the equality of speeds is exact. If only the inequalities

$$a^2 G_U \preceq DT^{\mathsf{T}} G_V DT \preceq b^2 G_U$$

hold, then $a L_U(\gamma) \le L_V(T \circ \gamma) \le b L_U(\gamma)$.

The upper distance bound $d_V(Tx, Ty) \le b\, d_U(x, y)$ follows by transporting each source curve and taking the infimum. The lower bound needs more: a short target curve might not be the image of a source curve. A diffeomorphism onto the designated target region, with matching allowable curve families, supplies the needed reverse path comparison. This is why a local derivative lower bound on a noninjective map does not prove a global lower distance bound. The map $t \mapsto (\cos t, \sin t)$ has unit derivative norm everywhere, yet points 2π apart have the same image.

Infima, minima and exact quotient distance

An infimum is a greatest lower bound. It can exist without being attained. The numbers $1 + 1/n$ have infimum 1 although none equals 1. A path-distance definition therefore need not assume a shortest curve exists. Its triangle inequality follows from approximate minimizers: choose a

first path within $\epsilon/2$ of one infimum and a second within $\epsilon/2$ of the other, concatenate, then let ϵ approach zero.

For a pseudometric d, the zero relation is transitive because $d(x,z) \le d(x,y) + d(y,z) = 0$ whenever $d(x,y) = d(y,z) = 0$. The reverse triangle inequality gives

$$|d(x,y) - d(x',y')| \le d(x,x') + d(y,y').$$

Thus choosing other representatives of the same zero classes leaves distance unchanged. The quotient separates classes because distance zero is exactly the defining identification. This quotient uses an exact zero relation, even when its input pseudometric was constructed from approximate operational maps.

Quantization and a comparison budget

Round each of n coordinates to the nearest multiple of $\Delta > 0$, with a fixed tie rule and no clipping. Each coordinate error is at most $\Delta/2$, hence

$$\|Q_\Delta(v) - v\|_2 \le \sqrt{n}\,\Delta/2.$$

For two vectors, apply the reverse triangle inequality to their difference:

$$\big|\|Q_\Delta(v) - Q_\Delta(w)\|_2 - \|v - w\|_2\big| \le \sqrt{n}\,\Delta.$$

Think of the two ends of the inspection rail measurement. If rounding nudges the left reading left and the right reading right, both nudges count in the measured gap. The bound leaves room for both, even when rounding usually works out better.

For a constant metric $G \preceq MI$, the same endpoint error is at most $\sqrt{Mn}\,\Delta/2$ in G-norm, and the pair-distance bound is $\sqrt{Mn}\,\Delta$. With a changing metric, use a uniform bound along the relevant small adjustment paths. Clipping invalidates the coordinate error bound unless its additional error is separately included.

A statistical model supplies a different ruler

Suppose the observation is a binary sensor outcome whose success probability is $p \in (0,1)$. The probability model is $P_p(X=1) = p$, $P_p(X=0) = 1-p$. The score derivative is $\partial_p \log P_p(X) = X/p - (1-X)/(1-p)$. Its expected square is

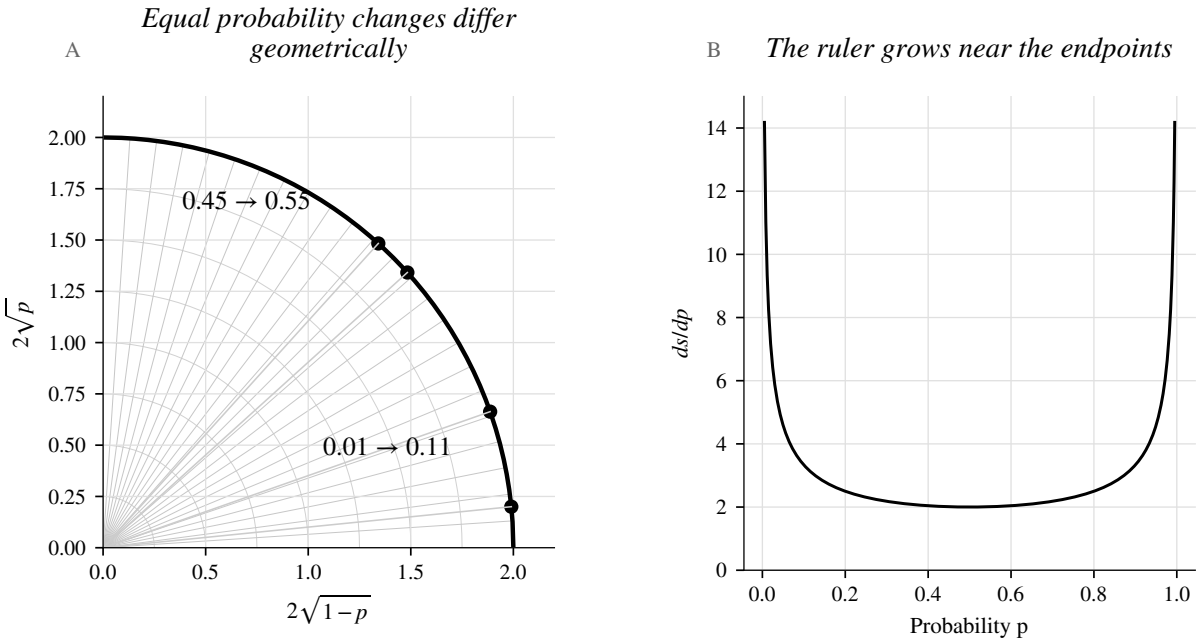
$$I(p) = p\frac{1}{p^2} + (1-p)\frac{1}{(1-p)^2} = \frac{1}{p(1-p)}.$$

The Fisher metric is therefore $ds^2 = dp^2/[p(1-p)]$. It measures distinguishability of nearby probability models under this statistical construction. It does not assign physical metres to the sensor or identify which individual observation occurred.

Set $u = 2\arcsin\sqrt{p}$. Then $du/dp = 1/\sqrt{p(1-p)}$, so the metric becomes du^2. The Fisher distance between two Bernoulli models is consequently

$$d_F(p,q) = 2|\arcsin\sqrt{p} - \arcsin\sqrt{q}|.$$

The endpoints $p = 0, 1$ lie outside the regular open model, though they are at finite distances in its metric completion. The coordinate coefficient diverges there while the integrated distance remains finite. This gives another concrete distinction between a coordinate coefficient and a global distance.



Probability pair	*Raw difference*	*Fisher distance*
0.01 → 0.11	0.10	0.475796
0.45 → 0.55	0.10	0.200335

Figure A.1: Bernoulli probabilities embedded as $(2\sqrt{1-p}, 2\sqrt{p})$ on a radius-two quarter-circle. Its arc-length element is the Fisher ruler. Equal raw probability differences produce different intrinsic lengths near the edge and near the middle of the interval.

Under a smooth change of parameter $p = p(\theta)$, the score transforms by the chain rule and $I_\theta = (dp/d\theta)^2 I_p$. The metric transforms just as the surveyor's ruler did. More generally, information geometry studies the behavior of statistical models and their metrics under parameter changes and statistical transformations [9].

To see why losing observations can reduce information, let $T(X)$ be a statistic and suppose regularity permits differentiating its distribution under the integral. Its score is the conditional expectation of the full-data score given T. Jensen's inequality then gives

$$\mathbb{E}\big[\mathbb{E}[S(X) \mid T]^2\big] \le \mathbb{E}[S(X)^2].$$

In several parameters the same argument is a positive-semidefinite matrix inequality, applied to every linear combination of score coordinates. Equality requires the relevant score information to be retained by the statistic. The assumptions on differentiability and integrability belong to this result; an arbitrary compressed file is not assigned a Fisher metric by analogy.

Composition of endpoint errors

Suppose a correct transformation sequence is $F_k \cdots F_1$ and implemented steps introduce errors e_j at their outputs. If the subsequent steps have Lipschitz constants $L_{j+1}, \ldots, L_k$ on every intermediate

point used, then the final discrepancy is bounded by

$$\sum_{j=1}^{k} e_j \prod_{r=j+1}^{k} L_r.$$

Induction proves the formula. After one more step, multiply the old bound by its Lipschitz constant and add its new error. Empty products equal one. This bookkeeping exposes why early errors can dominate a long route, and why all intermediate states must remain inside the regions supporting the constants.

For three steps with $e_1 = e_2 = e_3 = 0.01$ and $L_2 = L_3 = 2$, the bound is $0.04 + 0.02 + 0.01 = 0.07$. The first converter's mistake goes through two steps that can each double it. The last converter's mistake is already in the output, so we count it as it is. If every later constant is at most one, simple addition is valid.

Risk, conditioning and selective answers

Let A be the event that a system answers. Report coverage $\Pr(A)$ and, for positive coverage, selective error $\Pr(\widehat{Y} \neq Y \mid A)$. Their product is the probability of an answered mistake. Neither quantity describes performance alone.

A regime-specific guarantee concerns $\Pr(\cdot \mid G)$; an overall bound also allows for $\Pr(G^c)$. An alert-conditioned posterior needs the alert prevalence and its relation to G. This continues the map/globe discipline: every number retains its event, description and measured population.

Appendix B

Solutions and routes through the exercises

These solutions cover the numbered exercises. The detector workshop retains its own checks and answer discussion. An open-ended exercise has a proposed construction rather than a unique verbal answer.

The object and its descriptions

1.1. $y = 1000(0.025) - 12 = 13$. Since $dx = dy/1000$, the metric coefficient is 10^{-6}. The offset changes addresses, not differential lengths.

1.2. Two copies of one instrument record, with the same source identifier and acquisition time, can describe one measurement. Two instruments can also record the same temperature at different places. Value equality alone cannot decide which occurred; retain source, time and location or the corresponding observation identifiers.

2.1. The chain $0, 0.6, 1.2$ already has endpoints farther than 1. All three lie in one connected component of the adjacent-threshold graph. The clustering rule is therefore broader than the original pairwise tolerance.

2.2. For a rigid rotation of the same mug, its serial number remains unchanged while pixel locations change. An inspection of the mark on the object, linked to both photographs, can test that identity claim. Other justified invariants are possible.

3.1. The translated readings are 205 and 208. Their raw difference is 3 centimetres, or 0.03 metres. Integrating $ds = |dy|/100$ gives the same length as integrating $|dx|$ between 2 and 2.03.

3.2. Specify the representative space, base space and projection, and show that every base point has two representatives with evenly covered neighborhoods. Merely storing a document in two formats does not establish those properties.

Riemannian foundations

4.1. Both chart pairs are $(0, 1)$. Substitution into the overlap rule divides by $0^2 + 1^2 = 1$ and leaves the pair unchanged.

4.2. A different chart covers the pole smoothly. The failure is in assigning a longitude, not in the local Euclidean structure of the sphere.

5.1. The squared length is $0.04^2 + 9(0.02)^2 = 0.0052$, so the length is about 0.072111. Multiplying $(0.04, 0.02)$ by the displayed J gives the same squared Euclidean norm.

5.2. If α is in degrees, $\theta = (\pi/180)\alpha$, so $ds^2 = dr^2 + r^2(\pi/180)^2 d\alpha^2$.

6.1. The arc is $2\pi/3 \approx 2.0944$, while the chord is $4\sin(\pi/6) = 2$.

6.2. Enlarge the permitted region or compare routes through overlapping charts. A shorter valid path that leaves the original chart disproves the global optimality of the confined result. Failure to find such a route is not by itself a proof that none exists.

7.1. $s/R \le \sqrt{0.06} \approx 0.244949$ is sufficient and lies inside the stated range $s/R \le 1$.

7.2. $f(1+h) - f(1) - f'(1)h = h^2$. The second derivative is 2, so the bound $Bh^2/2$ equals h^2 and is attained.

8.1. The pullback matrix is $\mathrm{diag}(4, 9)$; singular values are $2, 3$; the inverse derivative has operator norm $1/2$.

8.2. The local pulled-back metric is $4dt^2$. Every circle point has two representatives modulo 2π. The derivative is nonzero, so no local direction is lost; the ambiguity is global.

Translators and realization

9.1. $D(0, 2\varepsilon) = 0$, $D(2\varepsilon, 0) = 2\varepsilon$, and $d_{\mathrm{sym}} = 2\varepsilon$. Each free loop moves forward by ε; backward travel is paid.

9.2. Use directed edges $A \to B$ and $B \to C$, each of cost 1, with no direct $A \to C$. Under the one-hop rule the last cost is infinite, so the usual triangle inequality would demand $\infty \le 2$, which fails.

10.1. The two-step conversion is $10(100x) + 7 = 1000x + 7$, exactly the direct conversion. Adding 0.2 millimetre to the direct route produces an absolute residual of 0.2 millimetre.

10.2. The bound is $0.012 + 0.4(0.01) = 0.016$. Without an h-net, some points may have no nearby tested point, so the extension from tested residuals to the whole region is unsupported.

11.1. Choose positions at angles $-\epsilon$ and $+\epsilon$. One local chart near the seam describes both; alternatively use overlapping angular intervals with a 2π transition. The short arc has length $2R\epsilon$ in either description.

11.2. The proof only needs finitely many chart segments along this particular path. Compactness of its parameter interval supplies a finite subcover and suitable subdivision, even on a noncompact manifold.

12.1. The shortest angular lift has displacement $\pi/2$. The distance is $\sqrt{4(\pi/2)^2 + 9} = \sqrt{\pi^2 + 9}$.

12.2. The lower bound gives $d_T(x, y) \le \eta/a$. Unless $\eta = 0$ or injectivity is separately supplied, different nearby points can share an image.

13.1. The upper multiplier is $1.03(1.02) = 1.0506$. The additive term is $1.03(0.01) + 0.02 = 0.0303$, in the output distance units after the first error has been converted by the second multiplier.

13.2. A margin strictly greater than 0.008 is sufficient. At the boundary of that interval, the threshold's equality convention can matter.

Covers and transported comparison

14.1. Initially $\cos^2\theta = 1/4$ and $\cos^2(\theta/2) = 3/4$. After adding π, they are $1/4$ and $1/4$. Only the first is invariant under the line identification.

14.2. A visible end mark distinguishes the endpoints directly. An unmarked camera can also follow the axis over time if the starting arrow is known and the record supports a unique lift. For discrete samples, a bound of less than $\pi/2$ on the total physical rotation between frames suffices for the nearest-arrow rule. A pair of endpoint snapshots without that motion information cannot distinguish a half-turn from staying still.

15.1. The quaternions are $(0, 0, 0, 1)$ and $(0, 0, 0, -1)$. Conjugation by either sends (x, y, z) to $(-x, -y, z)$.

15.2. The intrinsic angle is $\pi/3$; the score is $\cos^2(\pi/6) = 3/4$; the chordal comparison is $\sin(\pi/6) = 1/2$. Its square and the score sum to one.

16.1. Each vector has zero dot product with the corresponding sphere position. The constant vectors have zero derivative. On the middle leg, $v' = -(0, \cos t, \sin t) = -r(t)$ is normal, so its tangential derivative vanishes. The endpoints match the next segment's initial arrow.

16.2. A quarter-turn about z fixes $(0, 0, 1)$. Testing only that vector reports no change although the operator is not the identity.

17.1. For A, B, the operator norms are both $\sqrt{2}$. The Frobenius norms are $2, 2\sqrt{2}$; normalized traces $1/2, 0$; and intrinsic distances $\pi/2, \pi/\sqrt{2}$.

17.2. The operator deviation is $2\sin(\pi/6) = 1$, determined by the larger angle. The intrinsic distance is $\sqrt{\pi^2/9 + \pi^2/36} = \pi\sqrt{5}/6$, retaining both angles' squared contribution.

Decisions and computation

18.1. The squared distance is $4(0.2)^2 + (0.2)^2 = 0.2$, so the distance is $\sqrt{0.2} \approx 0.447214$.

18.2. The lower bound is positive when $e < 2m$. The combined uncertainty of the two endpoints must be smaller than the geometric separation available between the confident regions.

19.1. The length factors are 0.9 and 1.2, giving the interval $[4.5, 6]$.

19.2. The factor is $\sqrt{18/2} = 3$. Information about the operator's action relative to the metric's eigendirections can improve the general worst-case estimate.

20.1. The bound is $0.003 + 2(0.4)(0.02) = 0.019$.

20.2. Guaranteed acceptance uses $D_T \le 0.98/1.05 \approx 0.933333$. Guaranteed rejection uses $D_T > 1.02/0.95 \approx 1.073684$. The intervening range remains unresolved by this bound.

21.1. Each $c_y = 1/(1+4) = 0.2$. Equal prevalence makes the posterior lower bound $0.8/(0.2+0.8) = 0.8$.

21.2. Start with nine true alerts and one false alert. A filter retaining only the false alert reduces precision from 0.9 to zero. A claim for the filtered subset needs evidence about the filter's selection.

22.1. The distances are $\sqrt{0.5} \approx 0.707107$ and 1. This ranking uses the chosen scales; it is not proof of transaction identity.

22.2. Equal numerical amounts and timestamps in USD and EUR can have zero numerical difference in the displayed formula, but fail the fixed-currency candidate rule. They are outside that formula's intended common domain.

23.1. The third record's alert status would be overwritten. The temperatures can remain correct while the decision-bearing categorical field is wrong.

23.2. The quarter-turn matrix maps $(0, -10)$ to $(10, 0)$. The physical discrepancy is zero; direct subtraction of the unconverted arrays has length $\sqrt{200}$.

24.1. For example, $(0.1, 0.2, 1.1, 1.2)$ and $(0.8, 0.9, 1.8, 1.9)$ have half their readings in each bin. Their means differ. The fingerprint preserves the two bin proportions, not the within-bin positions or the mean.

24.2. The displayed center-to-truth separation is about 0.0069050, below radius 0.0312438. Containment permits many values; the center is one representative of the admissible disk, not a proof of the hidden value.

25.1. All-equal triples fail every inequality. Each other triple has two equal entries and therefore fails the inequality on that pair. This accounts for all eight possibilities.

25.2. The forward updates are $(1/4, 3/4)$, $(7/20, 13/20)$ and $(14/27, 13/27)$. In reverse they are $(2/3, 1/3)$, $(3/5, 2/5)$ and $(1/3, 2/3)$. The state can evolve between observations, so report order determines which evidence applies before and after the transition.

Appendix C

How to train your detector

Today I want to talk about something that quietly sits underneath a lot of modern systems: how they decide when two things are "the same."

When your phone unlocks using your face, when a hospital compares two scans, when a bank decides whether a charge matches something you really did—in all of these cases, there is some detector sitting behind the scenes, trying to answer a very simple-looking question:

Are these two observations coming from the same underlying thing, or not?

Tonight's lecture is really a story about training that detector, and about holding it to its promises.

I am going to tell this as a narrative. You don't need any advanced math for this, just a feel for distance and for change over time. I will occasionally wave at deeper results in the book, but you can ignore those details on a first read.

Part 0: A Hidden World and Its Many Views

Let me start with the picture I have in my head.

Behind the data, there is some hidden world: people, machines, viruses, ships, accounts, and all the other things we care about. Each of those things has a state that changes over time.

A person ages. A machine heats up, then cools down, then wears out. A virus mutates. A payment starts as an intent and gets turned into different messages on different rails. A ship leaves port, changes speed and heading, and eventually arrives somewhere else.

We never see these states directly. Instead, we see them through *channels*.

A camera gives us images or video. A logging system gives us text lines. A telemetry system gives us time series. A lab test gives us numbers. A payment network gives us structured messages. A neural network encoder gives us vectors.

The important thing is that the *same* underlying state can show up in many ways:

- The same person can appear as a passport photo, a selfie, and a 512-dimensional embedding vector.
- The same payment can appear as a SWIFT message, a card authorization, and a ledger entry.
- The same machine state can appear as a particular pattern of metrics and logs.

So when I say "hidden world," think of that: the true things and their states, which we only glimpse through instruments.

What does the detector do with a pair of observations? Given maybe two images, maybe an image and a vector, maybe two financial messages—it has to decide:

"Same underlying thing under acceptable change? Or genuinely different?"

That is the whole game.

Take a second and check yourself: in your own mind, can you say what the "hidden world" is for faces, or for payments? If you can, you are already on the right track.

Part 1: Paths, Benign Evolution, and Curved Space

To talk about sameness, I want to introduce two ideas: *paths* and *curved spaces*.

Let me start with distance, because that's familiar. In a city, there are different ways to talk about distance.

There is the straight-line distance—as the crow flies. And there is the street distance—how far you actually walk along sidewalks and roads.

The street distance is almost always longer, because you have to follow the roads. And there is a basic principle that says: if you go from A to C by way of B, that cannot be shorter than going from A to C along the best possible route. Detours do not beat the optimal path. That is the triangle inequality in plain clothes.

Now, let me switch to time.

Think about one person aging. From day to day, the change is tiny. From year to year, a bit more. Over decades, a lot. But we still regard it as the same person throughout. You can imagine their state moving along some path in an abstract "face space"—a little drift each day.

You can do the same for other things.

A healthy machine wiggles around inside a normal region of behavior. A virus drifts as it mutates. A payment moves through multiple systems, picking up different annotations. In each case you can imagine a path of states.

Now I want to name a kind of change. There are changes we are comfortable calling "the same"—small lighting changes on a face, small fluctuations in metrics, a payment picking up an extra log line. And then there are changes we are *not* comfortable calling the same—a different person, a machine jumping into a failure mode, a completely unrelated transaction.

I will call the first kind *benign evolution*. A *benign path* is a path along which the changes are all benign.

So one way to say "these two states are the same" is:

> I can get from one to the other by a benign path that is not too long.

And one way to say "no, they are different" is:

> Any path between them is either not benign, or unreasonably long.

That is informal, but it is the intuition behind the geometry in the main text.

Now, what about curved spaces?

The metaphor I like here is the Earth. The Earth's surface is curved, but if you stand on a small patch of it, it looks flat. When pilots talk about distance between two cities, they mean distance along the surface—those arcs you see on a globe.

And you cannot draw a single flat map of the whole Earth without some distortion. That is why we have many maps: one for Europe, one for North America, and so on. Each one is accurate locally. Each one is a *chart* on the sphere.

I want you to take that feeling and move it to the world of states.

Imagine a "surface of all possible faces," or "all possible system states," or "all possible virus strains." Each point is a possible state. Distance along the surface is how much that state has changed. Nearby points correspond to benign changes; far points correspond to large changes or different identities.

For example: a 25-year-old's face and a 26-year-old's face of the same person would be nearby points—a small benign change. Two different people's faces, even if they look somewhat similar in

a photo, would be far apart on this surface, because there is no benign path connecting one to the other.

In the main theory, that surface is a Riemannian manifold. Here, it is enough if you imagine something that behaves like the Earth: curved overall, flat in small patches, with distances measured along the surface.

If you can hold that picture, we are ready to talk about two kinds of engineers.

Check your understanding. Why can't one flat map cover Earth perfectly without distortion? In the space of faces, what might it mean for two faces to be close together on this surface? What might it mean for them to be far apart?

Part 2: Two Ways to Build a System

The first kind of engineer I will call the *translator builder*. The second I will call the *geometer*. They are both trying to solve the same problem—they are just using different language.

Let me start with the translator builder.

In her head, there are many vector spaces:

- an image embedding space,
- a text embedding space,
- a log-feature space,
- a financial message space,
- ...

Each channel has an encoder that maps raw observations into its space. On top of that, she trains translators: image to embedding, logs to embedding, rail A to rail B, and so on. If you drew it, it would look like a graph: spaces as nodes, translators as edges.

She worries a lot about *drift*: distributions change over time, new kinds of inputs arrive, models get retrained. And she worries about *round-trips*. If she takes a point in space A, translates it to space B and then back, she wants to land near where she started. If she does not, somewhere in that chain, identity has been damaged.

So the translator builder's goal is clear: make this graph of spaces and translators behave so that observations from the same underlying thing can move around without losing who they are, and observations from different underlying things eventually drift apart or fail consistency checks.

Now, the geometer.

The geometer thinks in terms of that curved sameness space I described earlier. In their head, there is one underlying surface whose points are states. Distances along that surface measure how much a thing has changed.

Each channel gives them a way of putting coordinates on a piece of the surface. One chart might look like a space of image features. Another might look like a space of telemetry signatures. Another might look like a space of economic variables.

It is just like the Earth again. In one part of town, people give directions with street addresses. In another, they give GPS coordinates. In another, they say "the third building past the fountain." **All of these are different languages for describing points on the same curved surface.**

The geometer chooses the shape of the surface and a way of measuring distance on it. They draw regions: inside this small radius, we are comfortable saying "same"; beyond that radius, we say "different"; in between, we maybe refuse to answer.

So the translator builder and the geometer are working on the same underlying concept—sameness under bounded benign evolution—but one is talking about spaces and translators, and the other is talking about manifolds and geodesics.

Keep both pictures loosely in mind; we will need them later.

Check your understanding. In your own words, what is the translator builder trying to achieve? In your own words, what is the geometer trying to achieve? Which mental picture feels more natural to you right now, and why?

Part 3: The Oracle, the Mapper, the Judge, and the Keeper

Now I want to introduce the three characters who will matter most for the heart of this lecture: the *Mapper*, the *Judge*, and the *Keeper*. They form a relay team between an imaginary oracle and the decisions we see.

Let me start with the oracle.

The oracle is a thought experiment. She lives directly in the hidden world. She sees the true states and all the benign paths. For any pair of states, she knows, without hesitation, whether they belong to the same identity or regime under benign evolution, or whether they are truly different.

We do not have access to her. But she gives us a reference: the ground truth we would like to approximate.

Between the oracle and us, imagine three roles.

First, the **Mapper**. The Mapper is responsible for building how the system *represents* the world:

- encoders from raw data into spaces,
- translators between spaces,
- or, in the geometric view, the curved sameness space and its charts.

The Mapper's job is to make sure that states that are truly the same under benign change land close together in these maps, and states that are truly different land far apart.

Second, the **Judge**. The Judge looks at whatever maps the Mapper has built and says, "Given these maps, how should we decide?" The Judge:

- chooses thresholds or templates,
- combines evidence from multiple channels,
- and decides: same, different, or "I abstain."

The Judge's job is to make cautious decisions on top of the representations.

Third, the **Keeper**. The Keeper lives at the boundary between the system and the world. The Keeper:

- keeps track of which sensors and subsystems are trusted,
- logs what the Judge decided,
- decides how and where to surface those decisions.

The Keeper's job is to make sure good decisions are not ignored, lost, or quietly overwritten.

Now, I want you to think about how these three roles can fail.

If the Mapper fails, two different people might be mapped to almost the same point in embedding space. They are truly different in the hidden world, but they look too similar to the system. That is a *mapping error*.

If the Judge fails, the maps might be fine—two people are clearly far apart—but the threshold between "same" and "different" has been set so wide that we still call them "same." That is a *decision error*.

If the Keeper fails, the system might correctly decide "different," log it, and then some downstream process ignores the alert, or a logging bug drops the record. That is a *usage error*.

Chapter 20 audits four kinds of failure inside the detector. The Mapper's work covers geometry and the link from observation to record. The Judge sets calibrated thresholds and decides when to abstain. The Keeper delivers the result to the people who need it and can enforce the abstention policy. A dropped alert can spoil a correct decision. For this lecture, count that as a failure of the whole relay, and assign each miss to the first failed stage.

Check your understanding. Can you name one failure that belongs to the Mapper, one to the Judge, and one to the Keeper?

Part 4: Error Budgets — Turning Local Limits into a Global Promise

Now we arrive at what I think of as the crown of the whole story: *error budgets*.

We are going to move from talking about errors informally to talking about how often they are allowed to happen, and how that translates into a global promise about the system as a whole.

Let us focus on one particular kind of failure: the system says "same" when, according to the oracle in the hidden world, the two states are truly different. This is the kind of miss that can be especially serious in safety and security applications.

For each of our three roles, we are going to attach a number:

- p_{map} — how often the Mapper is allowed to betray the oracle on truly-different pairs;
- p_{judge} — how often the Judge is allowed to betray the oracle on truly-different pairs, given correct maps;
- p_{keep} — how often the Keeper is allowed to betray the oracle on truly-different pairs, given a correct internal decision.

These are miss *budgets*. You can think of them as promises: "I will mess up at most this often, on this kind of case."

For example, you might decide that:

- the Mapper's budget should be at most 0.5%,
- the Judge's budget at most 0.7%,
- the Keeper's budget at most 0.3%.

Now, here is the theorem in plain language:

> Among truly different pairs, the chance that the final system treats the pair as the same is at most
>
> $$p_{\text{map}} + p_{\text{judge}} + p_{\text{keep}}.$$

If those three budgets are 0.5%, 0.7%, and 0.3%, then the global miss probability is at most 1.5%.

Use the same population of truly different pairs throughout. For a stage whose budget is conditional on the preceding stages succeeding, multiply that conditional probability by the probability of reaching the stage. The product is at most the budget. The sum bounds the miss rate when the three first-failure events cover every miss.

Now, I want to make at least one of these budgets feel very concrete. Let me talk about p_{map} in a face recognition setting.

Suppose we have a system where we know, in the ideal semantic sense, that two people are different. Maybe they have different IDs in a database; we have done the human work to be sure.

In the hidden identity manifold, let us imagine that their distance should really be something like 0.8 on whatever scale we are using. But when we run them through our encoder, we get embedding vectors that are only 0.15 apart. And let us say our system calls any pair with distance less than 0.2 "same person."

In that situation, the Mapper has failed the oracle. Two truly-different states have landed too close together. The Judge might then honestly apply its threshold and say "same." The fault, in this example, lies with the representation.

How do we measure how often that kind of failure happens?

Here is a simple empirical procedure:

1. Take 1,000 pairs of faces where you are certain the two people are different people.
2. Encode each face, compute the embedding distance for each pair.
3. Count how many of those 1,000 pairs end up with distance below your "same person" threshold, say 0.2.

If that happens 3 times, then the Mapper has caused a miss on 3 out of 1,000 truly-different pairs in this test. That is 0.3%. You might then set $p_{\text{map}} \approx 0.003$.

In practice, getting thousands of confidently-labeled pairs can be hard. Smaller samples give noisier estimates, which is one reason we keep budgets conservative and measure repeatedly as the system runs.

You can do analogous things for the Judge and the Keeper: measure how often their particular failure modes occur on cases where the oracle would say "different." In the main text, those budgets are tied to more refined quantities—round-trip errors, cocycle slack, local distortion, and so on—but the idea is the same: these numbers come from measurement, not imagination.

So, to summarize this part: once you have per-role budgets, you can turn them into a global promise. If the Mapper, the Judge, and the Keeper all keep their own promises, then the system as a whole cannot miss truly-different pairs more often than the sum of those budgets.

So we have error budgets, and we know how to measure them. But here is a practical problem: what happens to those budgets when we change the system's representation?

Check your understanding. Given budgets of 0.4%, 0.4%, and 0.2%, what would the pessimistic total bound be? And if you had to cut the total in half, which role would you try to improve first?

Part 5: Two Ways to Map a City

We are almost done. There is one more idea I want to explain: *why equivalence between different representations matters.*

Up to now, I have spoken as if the system's representation is fixed. In reality, we change encoders, retrain translators, introduce manifolds, merge or split spaces—all the time. So the natural question is:

What happens to my error budgets when I change how I represent the world?

To talk about that, I want to go back to cities.

There are at least two ways you can describe navigation in a city.

One way is with neighborhood maps and conversion rules. Downtown, you might use street names and numbers. At the harbor, people talk in pier and slip numbers. On a campus, everyone gives directions with building names and room numbers.

You also have conversion rules. "Pier 7 is at 3rd and Market." "Room 2.103 is in Building E, second floor." If you stack all of these together, you get something that feels a lot like the translator builder's graph: many local coordinate systems, plus translations between them.

Another way is with a single GPS-style map. Every point on the city has a latitude and longitude. Distances are measured in meters along roads. There is one global coordinate system.

Now: when are these two views effectively the same?

Informally, when the neighborhood maps overlap correctly, the conversion rules are reliable, going from one system to another and back again brings you close to where you started, and taking different routes to the same place gives you consistent answers.

When that is true, both systems—the neighborhood-map view and the GPS view—agree on the things we care about: how far apart two places are, which routes are short, which are long.

Back to our detector.

The translator-builder's system is like the neighborhood-map view: many spaces, many translators. The geometer's system is like the GPS view: one curved sameness space, with charts on top.

In the main text of this book, I make that correspondence precise. At a high level, the conditions are exactly the ones you would expect from the city analogy:

- the translators are *smooth*: small changes in input cause small changes in output, no wild jumps;
- they are *roughly invertible*: going $A \to B \to A$ puts you back near where you started;
- they are *consistent around loops*: going $A \to B \to C$ and $A \to C$ give you the same result up to small error.

When those conditions hold, you can show that the translator system and the geometric system are just two descriptions of one underlying semantic structure. They agree on which pairs are close and which are far, which paths are short and which are long, and where the boundaries between "same" and "different" lie.

Why does that matter for error budgets? Because it means that if you prove an error budget for your detector in one realization—say, in the manifold picture—you can transfer that budget to the translator picture, or vice versa, with only a controlled slack. Changing representation, when done within these equivalence conditions, does not magically erase your guarantees; it only perturbs them within a known bound. In other words: representation changes, when done under these conditions, become *governed operations*: your promises survive the upgrade.

The main text of this book makes this precise using the language of *categories* (collections of related structures) and *functors* (maps that preserve structure between those collections). You do not need that formalism here. The important idea is simpler: we have two different ways of

drawing the same city, and as long as we keep the conversions honest, the promises we make in one view continue to hold in the other.

Check your understanding. In the city analogy, what are the "translators"? What is the "manifold"? Why do we care that both descriptions agree on which paths are short?

Part 6: Why This Matters

Let me bring this back to earth.

We started with very ordinary questions: is this the same face? The same tumor state? The same payment? The same machine health? These are questions systems are answering, quietly, all the time.

Every time we change how those systems represent the world—a new model, a new embedding, a new translator, a new manifold—we are changing, implicitly:

- who is considered close to whom,
- where the line between "same" and "different" lies,
- which cases are easy to see, and which become harder.

Representation changes are not neutral. They move the geometry of sameness. They shift who is grouped together and who is pulled apart.

What I have tried to do in this book, and in this lecture, is give you a way to talk about that geometry explicitly:

- to think in terms of hidden worlds, paths, and curved spaces;
- to see translator-based systems and geometry-based systems as two realizations of the same semantic structure;
- to decompose detector failure into the contributions of the Mapper, the Judge, and the Keeper;
- to turn those contributions into an error budget that gives a global bound on how often we miss truly-different cases;
- and to understand how those budgets move when we change representations.

If you are a builder of systems, the invitation is: treat your detector as a relay trying to approximate an oracle. Make per-role budgets explicit. Treat representation changes as governed acts, not casual version bumps. Insist on replay and audit so that, when something goes wrong, you can walk back through the Mapper, the Judge, and the Keeper and see who spent their budget.

If you are a student of mathematics or of systems, the invitation is: notice that geometry, probability, and system design are not separate worlds. They meet here, in the quiet act of deciding what is "the same."

Deciding what is "the same" is a quiet kind of power. It shapes identities, boundaries, access, and risk. The geometry of sameness, and the error budgets built on top of it, are one way of making that power explicit and accountable to the promises we make.

That, in the end, is what I mean by *training your detector*.

Suggested Answers to Check Your Understanding Questions

Part 1 (Paths, benign evolution, curved space).

- One flat map cannot cover Earth perfectly because Earth's surface is curved. Whenever you flatten it, you must distort something: distances, angles, or areas.

- In the "face surface," two faces are close if they differ only by small, benign changes (lighting, pose, a little aging). They are far apart if they belong to different people or require a large, non-benign change to get from one to the other.

Part 2 (Translator vs. geometer).

- The translator builder is trying to make a network of encoders and translators where the same underlying thing can move between spaces without losing its identity, and different things separate over time.
- The geometer is trying to find a curved space and a distance rule so that benign changes are short paths and truly different states are separated by gaps.
- Which picture feels more natural is a matter of taste; many engineers find the translator view intuitive, while many mathematicians like the manifold view.

Part 3 (Mapper, Judge, Keeper).

- A mapping error example: two different people are encoded to nearly the same embedding.
- A decision error example: the embeddings are well separated, but the threshold between "same" and "different" is set so loosely that they are still called "same."
- A usage error example: the system correctly decides "different," but the alert is ignored, overridden, or dropped from the logs.

Part 4 (Error budgets).

- With budgets of 0.4%, 0.4%, and 0.2%, the pessimistic total bound is $0.4\%+0.4\%+0.2\% = 1.0\%$.
- To cut the bound in half, you would focus on the role(s) with the largest contribution to the sum (for example, whichever of the three budgets is biggest), or on the role you can most effectively improve in practice.

Part 5 (City maps and equivalence).

- In the city analogy, the "translators" are the conversion rules between neighborhood coordinate systems (street addresses, pier numbers, room numbers). The "manifold" is the underlying city surface described in a global coordinate system like GPS.
- We care that both descriptions agree on which paths are short because that is what lets us transfer guarantees: if "short" and "far" mean the same thing in both pictures, then decisions based on distance (like same vs. different) are consistent across representation changes.

Appendix D

The retained operator certificate

Chapter 19 introduces the eight-dimensional example from my constructive-metric research [31]. This appendix supplies the finite data and the error-control argument behind its certificate. The basis order is $000, 001, 010, 011, 100, 101, 110, 111$ for the three tensor factors. Matrix indices in the tables run from one to eight in that order.

Two exact rational matrices

The entries printed below are the nonzero upper-triangular entries of $10^9 G$ and $10^9 Z$. Reflect them across the diagonal; every unlisted entry is zero. Those integers specify both matrices exactly.

i	j	$10^9 G_{ij}$	$10^9 Z_{ij}$
1	1	1020992621	186229780
1	4	-70400501	136577786
1	6	-100421142	133725122
1	7	-128972260	130399910
2	2	1568375432	161146824
2	3	-186498274	156620258
2	5	-170872240	152769018
2	8	150608379	104742808
3	3	1665059990	152221036
3	5	-164622094	148477874
3	8	127640045	101800677
4	4	2288588162	100164003
4	6	-20521473	98071806
4	7	29181359	95633150
5	5	1745145677	144826953
5	8	102746274	99297433
6	6	2429082996	96023506
6	7	50901557	93635686
7	7	2559682835	91307439
8	8	3157807346	68081260

The metric satisfies $I \prec G \prec (13/4)I$. The matrix Z is positive definite and supplies the separation witness. A witness here means an explicit mathematical object whose substitution verifies the claimed property. Here you can multiply the matrices and check the resulting inequalities directly.

Enclose the conformal operator

Write $b = 4/9$, $\sqrt{b} = 2/3$ and $s = 57/20$. For $-1 \le r \le 1$, the ellipse map used in the source has the series representation

$$\phi(s\sqrt{b}\,r) = \frac{2\sqrt{b}\sum_{j=0}^{\infty} b^{2j(j+1)}\mathcal{T}_{2j+1}(r)}{1 + 2\sum_{j=1}^{\infty} b^{2j^2}\mathcal{T}_{2j}(r)},$$

where $\mathcal{T}_k$ is the Chebyshev polynomial of the first kind. The source gives the conformal-map identities used to obtain this expression. The certificate replay then works with rational arithmetic at the eight spectral arguments

$$r_\sigma = \frac{\sum_j w_j\sigma_j}{s}, \qquad \sigma \in \{-1, 1\}^3.$$

Truncate both sums at $N = 8$ and call the quotient $f_N(r)$. The rational projectors

$$P_\sigma = \bigotimes_{j=1}^{3} \frac{I_2 + \sigma_j B/\sqrt{b}}{2}$$

satisfy $P_\sigma^2 = P_\sigma$ and sum to the identity. They need not be orthogonal projectors in the Euclidean norm because B is nonnormal. Define $T_N = \sum_\sigma f_N(r_\sigma)P_\sigma$.

Using $|\mathcal{T}_k(r)| \le 1$ on this interval, put

$$\begin{aligned} L &= 1 - \frac{2b^2}{1 - b^6} > 0, \\ a_N &= \frac{2\sqrt{b}\,b^{2(N+1)(N+2)}}{1 - b^{4(N+2)}}, \\ d_N &= \frac{2b^{2(N+1)^2}}{1 - b^{4N+6}}. \end{aligned}$$

The denominators in the exact and truncated expressions exceed L. A scalar error allowance is

$$\epsilon_N = \frac{a_N}{L} + \frac{2\sqrt{b}}{1 - b^4}\frac{d_N}{L^2}.$$

The tensor eigenbasis has condition number $27/8$. If T_0 rounds every entry of T_N to the nearest multiple of 10^{-12}, then

$$\|T - T_0\|_2 \le \frac{27}{8}\epsilon_N + 8\max_{ij}|(T_N - T_0)_{ij}| < 10^{-10} =: \eta.$$

We stopped the series early and then rounded the matrix we got. That's two chances to move away from the true operator. The first term pays for the omitted tail after the change of basis, and the second pays for rounding the entries. The next test has to leave room for both.

Positive pivots give the certificate

Set $M = \max\{\|T_0\|_1, \|T_0\|_\infty\}$ and $\rho = 2M\eta + \eta^2$. Exact rational LDL^* decompositions establish the following lower bounds on every diagonal pivot:

Matrix	Each pivot exceeds
$G - I$	$1/100$
$(13/4)I - G$	$1/20$
$G - T_0^* G T_0 - (13/4)\rho I$	$1/200000$
Z	$1/10000000$

These are decomposition pivots, not eigenvalue bounds. Their positivity proves positive definiteness. The operator error estimate gives

$$\|T^* G T - T_0^* G T_0\|_2 \le \|G\|_2 \rho \le (13/4)\rho,$$

so the certified margin survives the uncertainty in the true conformal operator. Here's what that margin buys us. We can do the check with the rounded matrix on the page, allow the true matrix to differ by the full amount we bounded, and still conclude that the update shrinks vectors in the G ruler.

Exclude all diagonal metrics in the stated budget

Let r_i be the diagonal entries of $T_0 Z T_0^* - Z$. The retained rational calculation proves

$$\sum_i (r_i)_+ - 4 \sum_i (-r_i)_+ - 32 \operatorname{tr}(Z)\rho > \frac{249}{25000} > 0,$$

where $x_+ = \max(x, 0)$. The function $x \mapsto x_+ - 4(-x)_+$ is four-Lipschitz. The error allowance therefore preserves positivity for the true operator T.

For a diagonal H with each entry between one and four, the smallest possible contribution from a positive diagonal entry of $TZT^* - Z$ uses weight one, while the smallest contribution from a negative entry uses weight four. The displayed inequality therefore makes $\operatorname{tr}(H(TZT^* - Z))$ positive for every such H. A Stein inequality $H - T^* H T \succeq 0$ would make the same trace nonpositive when paired with $Z \succeq 0$. No allowed diagonal H can satisfy it.

Think of a diagonal ruler as eight knobs, one per coordinate, each set somewhere between one and four. For the update to pass, the trace we just computed has to be zero or negative. We set the knobs to make it as small as possible and it's still positive, with a certified lower bound of approximately 0.0099600061. We checked the sign in rational arithmetic before printing the decimal. Turning the knobs another way can't get us past that test.

Appendix E

Laboratory solutions and diagnostic readings

The laboratories are designed to produce a record of both successful invariances and deliberate failures. The reference calculations below explain the expected outcomes. Numerical residuals depend on arithmetic precision and the integration method; a claimed tolerance should be stated alongside the result.

Two maps of one globe

For the north inverse map F_N, substitution of $(1,1)$ gives $(2/3, 2/3, 1/3)$. The south coordinates of this point are $(1/2, 1/2)$. At $(1,1)$ the transition derivative exchanges and negates the two coordinate directions, each with scale one half:

$$J = \begin{pmatrix} 0 & -1/2 \\ -1/2 & 0 \end{pmatrix}.$$

For $h = (0.02, -0.01)$, the north quadratic form is

$$h^T g_N h = \frac{4}{9}(0.02^2 + 0.01^2) = \frac{1}{4500}.$$

The transformed displacement is $(0.005, -0.01)$ and its squared south length is $(16/9)(0.005^2 + 0.01^2)$, the same number. The two small coordinate arrows are different lengths on the printed maps because the local rulers are different.

The exact transition at $x + h$ is computed with its own denominator, $1.02^2 + 0.99^2 = 2.0205$. Subtracting $T(x)$ gives a finite displacement that differs from Jh by terms of second order in h. Halving h should reduce that leading remainder by approximately a factor of four. The derivative identity is exact; using it for a finite displacement is the approximation.

The route length is $2\arctan(2) - \pi/2 \simeq 0.6435011088$. At sphere radius three, the same angular route has length approximately 1.9305033264. Its dimensionless chart metric becomes $36I/(1 + u^2 + v^2)^2$. If the chart plane itself uses length coordinates $U = Ru$, $V = Rv$, the metric becomes $4I/[1 + (U^2 + V^2)/R^2]^2$. The coefficient changes because the coordinate units changed.

The deliberately wrong identity-metric calculation gives coordinate lengths one and one half. It should fail. If a program reports agreement there, check whether it accidentally reused one route's result or normalized both answers after calculation. A failed control is part of the explanation of what the successful check actually established.

A sheared survey map

The inverse shear is $A^{-1} = \begin{pmatrix} 1/2 & -1/2 \\ 0 & 1 \end{pmatrix}$. Multiplication gives $G_y = \begin{pmatrix} 1/4 & -1/4 \\ -1/4 & 5/4 \end{pmatrix}$. The off-diagonal terms are required even though the physical ruler is Euclidean.

The displayed quarter-circle velocity is

$$\dot{y}(t) = (-2\sin t + \cos t, \cos t).$$

Substitution into $\dot{y}^T G_y \dot{y}$ simplifies to $\sin^2 t + \cos^2 t = 1$. The integrated physical length is therefore $\pi/2$. Measuring the displayed curve with the identity matrix instead integrates

$$\sqrt{4\sin^2 t - 4\sin t\cos t + 2\cos^2 t},$$

which is a different experiment. It measures ink distance on the distorted display.

For the uncertainty example, $\Sigma_y = \begin{pmatrix} 0.08 & 0.04 \\ 0.04 & 0.04 \end{pmatrix}$. Multiplying by G_y and taking the trace gives 0.05. The off-diagonal covariance is generated by the shear and does not by itself imply a new physical dependence in the original instrument errors.

When the physical task metric changes to $G_x = \operatorname{diag}(4, 1)$, its display matrix becomes

$$G_y = \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}.$$

The three squared lengths become $4, 1, 5$. These changed answers are correct because the task deliberately changed its ruler. In contrast, a coordinate change alone should preserve the original $1, 1, 2$.

For the statistical continuation, the substitution $\theta = 2\arcsin\sqrt{p}$ makes the Fisher line element $d\theta^2$. Ten equal intrinsic increments are therefore obtained by equally spacing θ and transforming back through $p = \sin^2(\theta/2)$. Equal increments of probability or log odds give different sample placements on the same family.

A translator network with one altered edge

With the four-clock basis A, B, C, D, the incidence matrix is

$$B = \begin{pmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \\ -1 & 0 & 0 & 1 \\ -1 & 0 & 1 & 0 \end{pmatrix}.$$

The exact offsets $(0, 2, 5, 4)$ produce $(2, 3, -1, 4, 5)$. Changing the last edge to six seconds gives the equal-weight fit

$$\widehat{x} = (0, 2.25, 5.5, 4.25), \qquad B\widehat{x} - b = (0.25, 0.25, -0.25, 0.25, -0.5).$$

The residual is orthogonal to every admissible column variation: $B^T(B\widehat{x} - b) = 0$. Its squared norm is 0.5. The zero in the first offset fixes the otherwise arbitrary common origin.

For the cycle $A \to B \to C \to A$, the observed sum is $2 + 3 - 6 = -1$. For $A \to C \to D \to A$, it is $6 - 1 - 4 = 1$. These two residuals are unchanged by adding a gradient Bu to all edge observations. That invariance makes them a useful diagnostic of compatibility rather than a choice of clock origin.

In the rigid-frame extension, the convention $T_{ji} = F_j^{-1} F_i$ gives

$$T_{AC} T_{CB} T_{BA} = F_A^{-1} F_C F_C^{-1} F_B F_B^{-1} F_A = I.$$

The cancellation supplies a direct proof of the round-trip identity. Introducing an error in one factor breaks that cancellation. The remaining product records the error after it has been transported through the other frame transformations. Its translation and rotation parts should be inspected together.

A continuous attitude lift

For adjacent prescribed yaw angles separated by 30°, the quaternion dot product is $\cos 15°$ before arbitrary sign edits. The nearest representative chord is $2 \sin 7.5° \simeq 0.261052$. An artificial sign change makes the raw chord $2 \cos 7.5° \simeq 1.982890$. Both report the same physical angle after the absolute-dot-product comparison.

At one full turn the continuous representative has scalar part minus one. Its rotation matrix is the identity. The base orientation has returned while the lift remembers the turn. The prescribed motion travels only 30° between samples, so each true lifted endpoint is the nearer of the two sign choices. A small angle computed from the endpoints alone would not bound the motion between them.

For a physical 210° step, the endpoint orientation is also reachable by a 150° step around the opposite axis. The nearest-sign rule cannot decide which route actually occurred. An angular-speed bound whose product with the sampling interval is below 180° would exclude the long route.

In the interpolation extension, the relative physical yaw is 120° and the cover angle is 60°. Spherical interpolation follows that short great-circle segment. Negating the final quaternion without first selecting its nearest sign instead presents a longer cover arc. The sign-invariant endpoints still agree, but the path differs. Endpoint sameness and path sameness remain separate questions.

A boundary after a coordinate change

For $z = Ax$ with the laboratory shear, $x_1 = (z_1 - z_2)/2$. Thus the transformed statistic is $s(z) = z_1 - z_2$, and the decision boundary is the diagonal line $z_1 = z_2$. The original vertical boundary has changed appearance on the display but not its physical meaning.

The transformed task metric is $\begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}$. A G-ball of radius 0.03 permits $|\Delta x_1| \leq 0.015$, so $|\Delta s| \leq 0.03$. The calibration allowance makes the total score interval's half-width 0.05. At $|x_1| = 0.1$, the score magnitude is 0.2, leaving 0.15 after both allowances. The declared middle strip still abstains. The separate sign-only diagnostic abstains when $|2x_1| \leq 0.05$, or $|x_1| \leq 0.025$.

The deliberately untransformed statistic $2z_1$ equals $4x_1 + 2x_2$. Its sign can disagree with the intended sign of x_1 because it has introduced an unwanted dependence on the secondary coordinate. The alternative statistic y fails even more directly: opposite configurations can share the same y. Neither defect is repaired by a more accurate matrix inversion.

The finite grid audit gives deterministic counts for that grid. Reweighting points near the boundary changes the coverage fraction without changing the coordinate invariance. This is the distinction between a geometric guarantee on a region and a population statement about how often the region is visited.

Completion from the retained GIGI observations

The wine-alcohol counts sum to 6497. Horner evaluation at zero returns 2/6497. The six nonzero preparation nodes occur in conjugate pairs, so their polynomial values also occur in conjugate pairs because the coefficients are real. This provides a useful structural check before any completion solve.

For a strictly positive Pick matrix, expanding its Schur complement yields the center $c = E/(s + D)$ and squared radius $|c|^2 + (s - A)/(s + D)$ defined in Chapter 24. The inner query has

center approximately $0.0184189 + 0.0064237i$ and radius 0.0312438. The outer query has center approximately $0.0748215 + 0.0367339i$ and radius 0.2753669.

The center-to-truth gaps are approximately 0.0069050 and 0.1367974. Both are smaller than their corresponding radii. The outer ratio is approximately 0.4967824. That containment is the expected result; equality with the center is not required.

The two-point identity control is singular because it already determines the whole function. Schwarz's lemma gives $|f(z)| \le |z|$ after $f(0) = 0$, and equality at the nonzero point $1/2$ forces a rotation $f(z) = e^{i\theta}z$. The observed positive value fixes $e^{i\theta} = 1$. The contradiction control changes that value to 0.9, violating the same inequality and producing the negative determinant $-56/75$.

For the same-size edit, multiplying the observations by 0.8 preserves feasibility because $0.8H$ is still a bounded analytic function. It changes the constraints and therefore generally changes the feasible disks. The observation count is unchanged. A cache that returns the old answer has failed to retain the dependency on the evidence.

The successful calculation links four different records: the original counts, the prepared observations, the completion result and the separately opened validation truth. Keeping those records distinct is what makes the empirical study readable and reproducible.

Appendix F

Worked constructions: solutions and checks

The extended investigations end with calculations that connect the model to an observable result. These solutions retain the intermediate steps so that a discrepancy can be traced to a convention, a derivative, a route choice, or a numerical calculation.

Rescaling the surveyed globe

The dimensionless north-chart metric on a sphere of radius R is

$$g_R(u,v) = \frac{4R^2}{(1+u^2+v^2)^2} I.$$

The route from $(1,0)$ to $(2,0)$ has length

$$L_R = R\left(2\arctan 2 - \frac{\pi}{2}\right).$$

At $R = 10$, this is approximately 6.435011088. The same angular arc in the south chart gives the same result. The chord becomes $10\sqrt{2/5} \simeq 6.324555320$. The long great-circle route becomes $10(2\pi - \alpha)$, where $\alpha = \arccos(4/5)$.

Now use map coordinates with length units, $U = Ru$ and $V = Rv$. Their differentials are $du = dU/R$ and $dv = dV/R$. Substituting into the line element gives

$$ds^2 = \frac{4(dU^2 + dV^2)}{(1 + (U^2+V^2)/R^2)^2}.$$

The explicit factor R^2 has disappeared from the numerator because the coordinate displacements already carry length units. The endpoints are now $(R,0)$ and $(2R,0)$. Integrating this expression still gives L_R. A formula cannot be compared across two implementations without checking which of these coordinate conventions each uses.

For the geodesic's room-coordinate acceleration, $x''(s) = -x(s)/R^2$ when s is physical arc length. If the parameter is instead the dimensionless angle $t = s/R$, then $d^2x/dt^2 = -x$. Both equations describe the same route. The derivatives differ because their parameters differ.

Adding information to the pointer observation

An end-identifying mark makes the pair $(\cos\theta, \sin\theta)$ available. Equality of this pair for two angles implies equality modulo 2π, so the physical pointer state is recovered. Its derivative has norm one at every angle.

By contrast, adding coordinates such as

$$h(\theta) = (\cos 2\theta, \sin 2\theta, \cos 4\theta, \sin 4\theta)$$

does not distinguish the two ends. Every component has period π. The squared derivative norm is $4 + 16 = 20$, so the map has full rank as a map from a one-dimensional manifold, while retaining

the two-sheet ambiguity. More sensitivity and more recorded coordinates did not imply global injectivity.

If a measurement vector is multiplied by an invertible matrix, its fiber structure is unchanged: equality before and after that transformation is equivalent. A noninvertible projection can enlarge fibers. These observations provide a simple design test for a proposed feature transformation before any training experiment is discussed.

At the horizontal sensor's maximum, the exact inverse relation for a report $1 - e$ is $\theta = \pm \arccos(1 - e)$. As $e \downarrow 0$, its magnitude is $\sqrt{2e} + O(e^{3/2})$. The square-root response and sign ambiguity explain why the usual linear inverse fails there. It is a property of the observation geometry, not an arithmetic accident.

The altered camera mark

In Chapter 10, cameras A and B report three marked points on one workbench. Camera A's triangle has side lengths 1, 1 and $\sqrt{2}$. Moving the second point in camera B's report to $(3, 3.1)$ gives lengths 1.1, 1 and $\sqrt{2.21}$. A rigid transform cannot change those lengths.

The centered-coordinate calculation gives

$$A = -\frac{1}{30}, \qquad B = \frac{7}{5}.$$

The fitted angle is

$$\widehat{\alpha} = \operatorname{atan2}(42, -1),$$

slightly larger than 90°. In particular,

$$\cos\widehat{\alpha} = -\frac{1}{\sqrt{1765}}, \qquad \sin\widehat{\alpha} = \frac{42}{\sqrt{1765}}.$$

The centroids are $\bar{p} = (1/3, 1/3)$ and $\bar{q} = (8/3, 71/30)$. The fitted translation is

$$\widehat{t} = \left(\frac{8}{3} + \frac{43}{3\sqrt{1765}}, \frac{71}{30} - \frac{41}{3\sqrt{1765}}\right).$$

These exact expressions can be used to verify the displayed drawing without rounding the angle first.

The rotation-dependent part of the objective is $-2(A\cos\alpha + B\sin\alpha)$. At the displayed angle, the bracket is $\sqrt{A^2 + B^2} = \sqrt{1765}/30$, its maximum. This proves global optimality for the planar rotation in this least-squares problem. It does not prove that the point correspondence or the rigid-camera model is correct; those were assumptions of the problem.

The three fitted residuals sum to zero because the optimal translation matches the centroids. They need not vanish individually. This is the expected signature of a shared transform fitting inconsistent data, and a useful implementation check.

Three sheets instead of two

For $f(e^{i\theta}) = e^{3i\theta}$, choose a base arc with angular width less than 2π. Its inverse image is three disjoint arcs of one-third that width. Restricting f to any of those arcs gives a homeomorphism onto the base arc.

The deck transformations multiply representatives by 1, $e^{2\pi i/3}$ and $e^{4\pi i/3}$. A full positive base turn $e^{2\pi i t}$ lifts from 1 to

$$\widetilde{\gamma}(t) = e^{2\pi i t/3},$$

ending at $e^{2\pi i/3}$. Three base turns are required to close that chosen lift. The physical base point returns after each turn.

The example has the same local-lifting mechanism as a double cover, with a different fiber. That is why the number of descriptions must be established from the projection rather than inferred from a resemblance between formulas. The aircraft cover has a two-point fiber; the phase quotient of a normalized complex vector has a continuous circle fiber; neither can be replaced by this three-sheet map without changing the object.

Calibration in different units

Two instrument calibrations give models $N(5, 4)$ and $N(7, 4)$. The second argument is variance, so both standard deviations are $\sigma = 2$. Substitution into the normal-family formula gives

$$d_F = \sqrt{2}\ \operatorname{arcosh}\left(1 + \frac{4}{16}\right) = \sqrt{2}\log 2.$$

After multiplying readings by 100, the distributions are $N(500, 40000)$ and $N(700, 40000)$. The mean difference is 200, and both standard deviations are 200. The ratio inside the distance expression is still $40000/160000 = 1/4$.

Multiplying the variance by 100 rather than 100^2 would represent a different noise model. This common unit error changes the geometry even though the transformed mean looks plausible. Keeping mean and standard deviation together in the derivation prevents it.

The constant-spread path has Fisher length $|7 - 5|/2 = 1$, slightly larger than $\sqrt{2}\log 2$. The full-family geodesic has a maximum standard deviation larger than two. That extra spread is a mathematical feature of the shortest route in the complete family, not a recommendation for operating the instrument.

A midpoint after a shear

The determinants of G', H' and the transformed midpoint M' are respectively 9, 4 and 6. Their leading diagonal entries are positive, so each symmetric 2×2 matrix is positive definite.

The generalized characteristic polynomial is

$$\begin{aligned}\det(H' - \lambda G') &= \det\begin{pmatrix} 5 - 10\lambda & 1 - 9\lambda \\ 1 - 9\lambda & 1 - 9\lambda \end{pmatrix} \\ &= (1 - 9\lambda)(4 - \lambda).\end{aligned}$$

Its roots are $1/9$ and 4, exactly those of the diagonal pair. Their logarithms give the preserved ruler distance.

If the off-diagonal terms are discarded, the matrices become $\operatorname{diag}(10, 9)$ and $\operatorname{diag}(5, 1)$. Their generalized eigenvalues are $1/2$ and $1/9$. One ratio has changed from four to one half. The diagonal-only calculation therefore compares different rulers; it is not an approximation justified solely by the existence of a simpler display.

For scalar $g, h > 0$, the geodesic midpoint is $\sqrt{gh}$. The matrix example inherits that formula in a commuting eigenbasis. In a general noncommuting pair, entrywise square roots or entrywise geometric means do not give the matrix midpoint. The conjugation by $G^{1/2}$ in the full formula is essential.

A circular graph with uncertain endpoints

With $n = 36$ and radius $R = 2$, each arc edge has length $\pi/9$. A ten-edge shortest route has arc length

$$d = \frac{10\pi}{9} \simeq 3.490658504.$$

Its chord-graph length is

$$d_n = 40\sin(\pi/36) \simeq 3.486229710.$$

The exact convention error is about 0.004428794 meters. If the true endpoints lie within one centimeter of the reported station endpoints, their true circle distance lies within 0.02 meters of d. Thus the direct interval is $[d - 0.02, d + 0.02]$.

If only d_n and the general relative-error bound are retained, let $\beta = \pi^2/(6n^2)$. Since $(1 - \beta)d \leq d_n \leq d$,

$$d_n \leq d \leq \frac{d_n}{1 - \beta}.$$

The uncertainty interval becomes

$$\left[\max(0, d_n - 0.02), \frac{d_n}{1 - \beta} + 0.02\right].$$

This interval combines the two mechanisms without treating the known one-sided chord error as a symmetric random noise.

Charging a fixed cost $c > 0$ for a chart switch changes the atlas path cost by c whenever the route must change reports. Refining the atlas can introduce additional required switches without changing the circle. Such an overhead can model computation, but it is not an intrinsic surface length unless that distinction is explicitly incorporated into the object being compared.

Three stages of transported error

The upper factor is

$$1.02 \times 1.05 \times 1.01 = 1.08171.$$

The additive allowance is

$$0.01(1.05)(1.01) + 0.02(1.01) + 0.005 = 0.035805.$$

For a hypothetical reversed route whose maps and domains support the stated reversed bounds, the factor is unchanged but the allowance is

$$0.005(1.05)(1.02) + 0.02(1.02) + 0.01 = 0.035755.$$

Early errors experience more subsequent amplification. Reversing an actual pipeline may be impossible because its maps have different domains or are not invertible; the numerical reversal is meaningful only after a reversed route has been supplied.

In the exact reconstruction example $a(x) = 100x$, $b(y) = y/100$, the round-trip allowance is zero. With ordinary Euclidean rulers the forward Lipschitz factor is 100 and the inverse factor is $1/100$. Their product is one, and the resulting two-sided distance bound is exact. With the transformed Y ruler, both factors are one. The two certificates report different coordinate conventions for the same underlying correspondence.

A population concentrated at the threshold

The triangular density $f(d) = 1 - |d - 1|$ on $[0, 2]$ integrates to one: each side is a right triangle of area one half. For $0 \le h \le 1$,

$$P(|D - 1| \le h) = 2 \int_0^h (1 - u)\, du = 2h - h^2.$$

At $h = 0.03$, the mass is 0.0591. The same reference error 0.04 and geometric-failure allowance 0.01 now give the transferred upper bound 0.1091, instead of 0.08 under the uniform example.

The geometric approximation did not change. The operational population placed more observations close to the boundary, where the same approximation can alter a decision. This calculation is the practical content of the boundary-mass term: it connects an error in the representation to the actual comparison task.

If a selective rule has interval-failure probability at most 0.01 and coverage 0.8, its disagreement rate conditional on deciding is bounded by 0.0125. At coverage 0.1, the corresponding bound is 0.1. A marginal coverage and a marginal failure bound do not justify claiming the same conditional reliability at every coverage level.

Three statistical midpoints

For $p = (0.8, 0.1, 0.1)$ and $q = (0.1, 0.8, 0.1)$, the mixture midpoint is $(0.45, 0.45, 0.1)$. The natural-coordinate midpoint is obtained by normalizing $(\sqrt{0.08}, \sqrt{0.08}, 0.1)$, giving approximately

$$(0.424889, 0.424889, 0.150221).$$

The Fisher midpoint is the componentwise square of the normalized vector $\sqrt{p} + \sqrt{q}$, giving approximately

$$(0.439965, 0.439965, 0.120071).$$

The unrounded vectors sum exactly to one. Independently rounded displayed components can sum to 1.000001; that is a display effect, not a failed normalization.

For coincident endpoints, the Fisher great-circle formula has a removable zero-angle expression. Define the interpolation to be the constant vector p, or use the continuous small-angle limit. The mixture and natural-coordinate formulas already return p directly. The special case should be handled deliberately rather than allowing division by zero to decide the output.

The three-outcome Fisher geometry is a piece of a radius-two sphere. The different paths in probability coordinates therefore do not contradict the existence of two exact coordinate descriptions: a path straight in one affine connection need not be shortest for the Riemannian connection.

The information inside merged bins

Both groups in the four-bin example have total mass $1/2$ under both distributions. Their conditional probability vectors are $(0.8, 0.2)$ versus $(0.2, 0.8)$ in the first group, and $(0.4, 0.6)$ versus $(0.6, 0.4)$ in the second. The weighted conditional contributions are

$$\tfrac{1}{2}(0.6 \log 4) = 0.3 \log 4, \qquad \tfrac{1}{2}(0.2 \log(3/2)) = 0.1 \log(3/2).$$

Their sum is approximately 0.4564352191. The coarse divergence is zero, so these two terms account for the entire difference.

Retaining the first fine-bin proportion would already distinguish the two distributions: it is 0.4 in one and 0.1 in the other. Retaining the conditional proportion within the first merged group also works. A statistic of the two coarse totals alone cannot distinguish them because those totals are identical.

A representative-invariant attitude mean

Three reports $q, q, -q$ produce

$$M = \tfrac{1}{3}(qq^T + qq^T + (-q)(-q)^T) = qq^T.$$

Any choice of their signs gives the same matrix. Its eigenvalues are one and three zeros, and its leading line is $\{q, -q\}$. The physical attitude is uniquely determined even though the eigenvector sign is not.

For equal-weight orthogonal quaternion representatives q_1, q_2, the matrix has eigenvalues $1/2, 1/2, 0, 0$. Every unit vector in their two-dimensional span maximizes the stated quadratic objective. The implementation can report the leading subspace or identify the ambiguity, but a selected single eigenvector is then a convention rather than a unique consequence of the projector loss.

An intrinsic attitude-distance mean solves a different objective. The projector example is valuable precisely because its optimization and its symmetry can be checked completely; changing the objective requires redoing that argument.

The exact evolving shear ruler

In Chapter 19, a shear carries $(0, 1)$ to $(t, 1)$. For the metric $G(t)$ constructed to preserve length under that shear,

$$\dot{G} = \begin{pmatrix} 0 & -1 \\ -1 & 2t \end{pmatrix}, \quad A^T G = \begin{pmatrix} 0 & 0 \\ 1 & -t \end{pmatrix}, \quad GA = \begin{pmatrix} 0 & 1 \\ 0 & -t \end{pmatrix}.$$

Their sum is zero. The transported vector $(t, 1)$ consequently has constant squared length one in $G(t)$.

The diagonal replacement is $\mathrm{diag}(1, 1 + t^2)$. It assigns squared length $1 + 2t^2$ to the same vector. The cross terms had contributed $-2t^2$, exactly the amount needed to preserve the original quadratic quantity. Dropping them changes the comparison even though the replacement remains positive definite.

Appendix G

From a geometric specification to a working comparison

A theorem tells us what a construction preserves. An implementation must decide how the objects, maps, rulers, and observations are stored and evaluated. This workshop follows that passage through small examples with known answers. It concentrates on the geometric contract: what the program is allowed to identify, what its numbers measure, and how a failed assumption becomes visible.

First build a record that can survive a change of format

Start with one sensor event. The event has an identifier, acquisition time, calibrated quantity, unit, coordinate frame, and calibration version. Those fields do different jobs. A value and unit state a measurement; the event identifier and acquisition time connect that measurement to an occurrence; the frame identifies its coordinates, and the calibration specifies how the raw reading becomes a measured value.

Use the constructed record

$$(\text{event } E17,\ t = 12.5\,\text{s},\ x = 2.03\,\text{m}, \text{frame } A,\ \text{calibration } C4).$$

A second encoding might store 203 centimeters, with time 12500 milliseconds. The numeric arrays differ. The translations $x' = 100x$ and $t' = 1000t$ preserve the stated quantities when the units change with them. The event and calibration identifiers should remain attached.

Separate the four equality tests

Byte equality asks whether the two serialized files are identical. Record equality asks whether decoding yields the same typed fields under a declared canonical convention. Measurement equality asks whether the values agree after unit and frame conversion. Event identity asks whether both records describe the same occurrence. None of these tests is a universal replacement for the others.

For example, sorting fields can change bytes without changing a record. Changing units can change the decoded number without changing the measurement. Copying an old event's value into a new event can preserve the measurement while changing the occurrence. A detector that uses only the number cannot recover the missing distinction.

The GIGI DHOOM round trip belongs to record preservation. The Riemannian coordinate examples belong to geometric preservation on stated smooth domains. The connection becomes useful when a pipeline records both contracts rather than assuming one proves the other.

Quantization is a map with fibers

Suppose the output format rounds meters to a grid of width $\Delta = 0.01$. Define

$$Q_\Delta(x) = \Delta \left\lfloor \frac{x}{\Delta} + \frac{1}{2} \right\rfloor,$$

with an explicit tie convention. Every grid value has an interval of inputs that round to it. The map is therefore many-to-one. For all ordinary non-overflow inputs,

$$|Q_\Delta(x) - x| \le \Delta/2.$$

For a pair of values,

$$||Q_\Delta(x) - Q_\Delta(y)| - |x - y|| \le \Delta.$$

This follows from the reverse triangle inequality and the two endpoint allowances. It gives the comparison stage a valid bound even though the encoding is not reversible.

If the metric on the measurement line is $g(x)dx^2$ with $g(x) \le M$ on the relevant intervals, each quantization displacement has intrinsic length at most $\sqrt{M}\,\Delta/2$. Two endpoint errors therefore contribute at most $\sqrt{M}\,\Delta$ to a distance comparison. The unit grid alone does not determine this intrinsic bound; the ruler is part of the calculation.

Clipping needs its own flag

If the format clips values above a maximum b, then the output b can represent arbitrarily many larger inputs. There is no uniform finite reconstruction allowance without an independent bound on the input range. A saturation flag distinguishes a measured value near the limit from an unknown value beyond it.

An exact decoder can restore the saturation flag and the clipped number. It cannot restore the lost physical quantity. This is the simplest example of why an excellent encoding round trip does not imply that the acquisition process retained every relevant measurement.

Choose a chart before a numerical solver chooses it for you

The sphere offers a concrete chart-management problem. North stereographic coordinates become large near the omitted north pole. South coordinates are well behaved there. A program can use both, but its switching rule should be based on numerical conditioning and declared domains.

For north coordinates $x = (u, v)$ with $r^2 = u^2 + v^2$, the transition to the south chart is

$$T(x) = \frac{x}{r^2}.$$

Its derivative is

$$DT(x) = \frac{1}{r^2} I - \frac{2}{r^4} x x^T.$$

Along the radial direction it has eigenvalue $-1/r^2$; along the perpendicular direction it has eigenvalue $1/r^2$. The transition is conformal up to an orientation reversal in these conventions. Its Euclidean condition number is one, while its absolute scale can be very large or very small. A condition-number report alone would miss that absolute amplification.

A simple switching rule

Suppose the active north chart is used while $r \le 2$. When r exceeds two, switch to the south chart, where the new radius is $1/r < 1/2$. To avoid repeated switches due to noise near a boundary, retain the current chart until its own radius crosses the outer threshold. The two-chart symmetry gives an overlapping operating band.

The thresholds are implementation choices, not geometric singularities. They should be selected with arithmetic precision and expected measurement errors in view. The state record should retain which chart is active, since the same pair of numbers can otherwise denote different points.

When switching a covariance or tangent quantity, transform it as well. If a small coordinate uncertainty has covariance Σ_N, first-order propagation gives

$$\Sigma_S \simeq DT\,\Sigma_N\,DT^T.$$

The metric transforms by the inverse rule, so the scalar $\operatorname{tr}(g\Sigma)$ is invariant under the linearized coordinate transformation. Copying the covariance entries unchanged while changing the coordinates breaks that invariance.

Test the point, tangent and uncertainty separately

At $(u, v) = (2, 0)$, the south coordinate is $(1/2, 0)$. The transition derivative is $DT = \operatorname{diag}(-1/4, 1/4)$. For the tangent arrow, the transformation gives

$$(0.02, 0.01) \longmapsto (-0.005, 0.0025).$$

The north and south metric matrices are $(4/25)I$ and $(64/25)I$, respectively. Both assign squared length 0.00008 to the corresponding arrow.

For $\Sigma_N = 10^{-4}I$, the propagated covariance is $\Sigma_S = 6.25 \times 10^{-6}I$. Both traces $\operatorname{tr}(g_N\Sigma_N)$ and $\operatorname{tr}(g_S\Sigma_S)$ equal 3.2×10^{-5}. These are three different checks: point reconstruction, tangent-length preservation, and uncertainty interpretation. A program can pass one while failing another.

Measure a route with a stated numerical tolerance

The north-chart route $u \in [1, 2]$, $v = 0$ has integrand $f(u) = 2/(1 + u^2)$. Its exact length is available, so it makes a useful independent reference for a numerical integrator. The purpose is not to demonstrate that a library can reproduce a known decimal. It is to connect its error tolerance to the eventual comparison.

For the composite midpoint rule with n equal intervals, the standard bound is

$$|L - L_n| \le \frac{(b-a)^3}{24n^2} \max_{[a,b]} |f''(u)|.$$

Here

$$f''(u) = \frac{4(3u^2 - 1)}{(1 + u^2)^3}.$$

On $[1, 2]$ it is nonnegative and at most one: differentiating again gives $48u(1 - u^2)/(1 + u^2)^4 \le 0$, so the maximum occurs at one. Therefore

$$|L - L_n| \le \frac{1}{24n^2}.$$

At $n = 100$, the allowance is $1/240000$, about 4.17×10^{-6}. A smaller observed difference from the exact value is welcome, but the proved allowance is what supports a uniform claim for the stated computation.

Keep distinct numerical approximations distinct

There may be several error sources: endpoint localization, approximation of the metric field, path discretization, quadrature, and floating-point arithmetic. The midpoint bound controls the quadrature of the specified integrand. It does not control uncertainty in the endpoints or prove that the prescribed route is shortest.

If the metric used by the program satisfies

$$(1 - \eta)g \preceq \widehat{g} \preceq (1 + \eta)g, \qquad 0 \le \eta < 1,$$

then the exact length of the same route under $\widehat{g}$ lies between $\sqrt{1 - \eta}L_g$ and $\sqrt{1 + \eta}L_g$. A numerical integral of that approximate metric adds another allowance. The square roots matter: the inequality compares squared tangent lengths, while the path integral measures lengths.

For a decision threshold, combine the intervals before deciding. If every admissible numerical and model value lies below the threshold, accept. If the interval crosses it, further precision might help, but only if numerical error is a material part of its width. Recomputing the same uncertain endpoint to more decimal places cannot remove acquisition uncertainty.

Use two genuinely different calculations

A useful validation pair measures the route in two coordinate charts with independently constructed metric integrands. Another pair compares a room-coordinate speed integral against a chart-coordinate integral. Repeating the same formula in two functions can reproduce the same mistake.

The map/globe example makes independent checking possible: the room curve $(\cos s, 0, \sin s)$ has unit speed, while the two chart integrals use different coordinate intervals and scale factors. Their agreement follows from the geometry, but their implementations need not share the same numerical path.

Carry an uncertainty ellipse along a route

Suppose a tangent displacement on the sphere is uncertain. In an orthonormal tangent frame at the starting point, its covariance is

$$\Sigma_0 = \begin{pmatrix} 4 & 0 \\ 0 & 1 \end{pmatrix}.$$

The entries are in the square of the chosen displacement unit. The long axis describes greater variance in one physical tangent direction. Moving this ellipse to another point requires a transport rule; copying its entries into an arbitrary new frame is not a geometric operation by itself.

Let $P : T_pM \to T_qM$ be Levi–Civita parallel transport along a specified route. It is a linear isometry for the Riemannian metric. In orthonormal start and end frames its matrix is orthogonal. The transported covariance is

$$\Sigma_q = P\Sigma_0 P^T.$$

Its trace, determinant and eigenvalues are preserved. Its displayed axes can rotate because the covariance directions have been transported relative to the chosen end frame.

A loop can rotate the ellipse without changing its spread

For a spherical cap with colatitude θ, the accumulated transport angle is $2\pi(1 - \cos\theta)$ modulo 2π, with sign fixed by traversal convention. Choose $\cos\theta = 7/8$, giving a return angle $\pi/4$. In the original frame,

$$P = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}.$$

Then

$$\Sigma_{\text{return}} = P\Sigma_0 P^T = \begin{pmatrix} 2.5 & 1.5 \\ 1.5 & 2.5 \end{pmatrix}.$$

The trace remains five and the determinant remains four. The eigenvalues remain four and one. The distribution of uncertainty across the original coordinate directions has changed.

The invariant scalar trace therefore cannot tell the whole story of the transported ellipse. It records total variance in this orthonormal frame, while the cross terms retain directional information. This is a direct geometric example of the book's warning against asking one score to represent every feature of a comparison.

Some objects are insensitive to the same holonomy

If $\Sigma_0 = cI$, then $P\Sigma_0 P^T = cI$ for every orthogonal P. An isotropic covariance returns unchanged even when a marked tangent arrow rotates. The connection has not become flat; the particular observable is invariant under that rotation.

Similarly, a line direction identifies opposite arrows. A return angle of π changes an oriented arrow to its negative but preserves its associated rank-one projector. An observation's symmetry determines which part of the transport it can reveal. This connects the bar's double cover to the choice of observable, without claiming that covariance matrices themselves form a two-sheeted cover.

Transport is not an uncertainty-growth model

Pure parallel transport preserves the metric length of every tangent vector. A physical estimator may instead propagate covariance through a dynamics Jacobian and add process noise:

$$\Sigma_{k+1} = F_k \Sigma_k F_k^T + Q_k.$$

That equation contains dynamics and new uncertainty. Its matrix F_k need not be an isometry, and Q_k need not be zero. When geometric transport is one component of an estimator, identify it separately from those mechanisms. A growing ellipse is not automatically evidence of curvature; a rotated ellipse is not automatically evidence of increased uncertainty.

Compare orientation histories without inventing a turn

Let normalized samples q_k represent aircraft attitudes at times t_k. Before subtracting samples, verify their norms and the component convention. Scalar-first and scalar-last arrays can have the same length and still describe different rotations if interpreted incorrectly.

Choose representative signs continuously where the sampling permits it. The relative quaternion is

$$r_k = q_k^{-1} q_{k+1}.$$

After choosing its short representative with nonnegative scalar part, write $r_k = (w_k, v_k)$. Its physical increment is

$$\alpha_k = 2\,\mathrm{atan2}(\|v_k\|, w_k), \qquad 0 \le \alpha_k \le \pi.$$

If $v_k \neq 0$, its axis is $u_k = v_k/\|v_k\|$. The vector $\alpha_k u_k/(t_{k+1} - t_k)$ is the rate for a constant-speed shortest rotation in the relative frame. The samples alone do not establish the actual angular velocity between them.

Small increments need a stable expression

For small α, the vector part has magnitude $\sin(\alpha/2) \simeq \alpha/2$. Computing $2 \arccos w$ can lose relative precision when w rounds close to one. The two-argument arctangent uses both components and is often numerically preferable. Any normalization step should be recorded: normalizing a badly corrupted quaternion hides a defect unless the original norm residual is retained.

The quantity to compare between two reported histories depends on the task. A sequence of sign-invariant endpoint errors tests attitude agreement at the sample times. A sequence of relative increments tests local path agreement under the sampling assumptions. Integrated angular-rate information can test for missed motion between samples. These are related records, but their conclusions have different domains.

An exact example with a sign edit

Let $q_0 = (1, 0, 0, 0)$ and let the next physical attitude be a 30° yaw, represented by

$$q_1 = (\cos 15^\circ, 0, 0, \sin 15^\circ).$$

If the file stores $-q_1$ instead, the physical increment remains 30°. A componentwise subtraction changes substantially. The sign-invariant quaternion dot product and the rotation matrices expose the agreement.

Now let the physical system make a 390° yaw between the two samples. Its final attitude is the same as after 30°, but the continuously lifted history differs. The two endpoint attitudes alone cannot settle which motion occurred. An angular-rate bound excluding an extra full turn is information about the experiment; it is not supplied by the endpoint distance formula.

Separate a representative jump from a physical jump

When an apparent discontinuity is found, check the matrix $R(q)$ as well as the stored quaternion. If the quaternion jumps in sign while R changes smoothly, the discontinuity is representational. If the rotation matrix jumps too, possible causes include actual motion, timestamp disorder, convention mismatch or bad data. The geometry narrows the diagnosis but does not determine the physical cause from one symptom.

A useful test fixture contains a known continuous rotation, deliberate sign changes, a shuffled timestamp, and a nonunit sample. The expected outcomes are different. A program that treats all four as one generic “orientation error” loses the distinctions the double-cover construction was meant to preserve.

Prepare a completion problem without looking at the answer

The measured-distribution experiment begins by reading all source records and constructing the histogram. That preparation can legitimately use the entire source because the task is then to answer queries about the prepared analytic object. The withheld quantities are function evaluations, not raw physical measurements that the preparation was forbidden to see.

For a field with counts $n_0, \dots, n_{31}$ and total N, store the ordered coefficients $p_k = n_k/N$, the physical bin convention, and the source-selection record. Evaluate the seven supplied nodes by Horner’s rule:

$$H(z) = p_0 + z\big(p_1 + z(p_2 + \cdots + zp_{31})\big).$$

The query stage can then be restricted to those seven observations and the bound $|H| \le 1$. A separate validator retains the complete polynomial and calculates the withheld truth after the query result has been recorded.

What should be counted as query work

Preparation reads the source records and computes the supplied values. A later query using a fixed-size prepared representation can have work independent of the original row count, while the overall pipeline still depends on that count. Both statements can be true. Report preparation and query costs separately so that one is not mistaken for the other.

The independent Pick-matrix calculation solves systems with the seven-by-seven matrix. It should solve linear systems rather than explicitly invert the matrix merely to form the same products. The dimension is set by the number of supplied complex observations. If that number grows, the work and conditioning change even when the raw source size stays fixed.

A singular matrix is a mathematical case, not an invitation to divide

A positive-definite Pick matrix gives the Schur-complement disk formula used in the worked examples. For distinct nodes in scalar Pick interpolation, a feasible singular Pick matrix gives a unique interpolant. To test a proposed extra value by a generalized Schur complement, its new column must also lie in the range of the original matrix. A pseudoinverse alone omits this condition.

The planted identity example makes this visible. The observations $f(0) = 0$ and $f(1/2) = 1/2$ force $f(z) = z$ by equality in Schwarz's lemma. The value set at every query is a singleton. The inference should recognize that mathematical constraint rather than report a broad region merely because an ordinary positive-definite solve fails.

The contradiction example changes the second value to 0.9. Its Pick determinant is negative, and no bounded disk map fits both observations. An implementation should preserve that infeasibility. Silently enlarging the bound or dropping an observation may be useful in a separately specified noisy model, but it is not a solution of the original exact problem.

Round the displayed answer outward when it is a bound

The book's tables round centers and radii for readability. Those printed decimals should not be treated as exact inputs to a fresh certification. If a displayed disk must itself be guaranteed to contain the computed disk, account for center-rounding error: a center displaced by at most e_c requires the displayed radius to increase by at least e_c in addition to its own rounding allowance.

If the stored calculation has a certified numerical error e_n, that allowance also belongs in the containment statement. An observed small floating-point residual is evidence about that run, but it is not automatically a rigorous rounding bound. The exact rational operator certificate elsewhere in the book uses a different verification method precisely because its acceptance criterion requires a certified inequality.

Control changes that leave the file size unchanged

A cache keyed only by observation count will miss a same-size edit. In the retained control, scaling every supplied value by 0.8 preserves feasibility but changes the constraints. The query result should change accordingly. Restoring the original values should restore the original problem.

The dependency record should therefore identify the observations themselves, their ordering or canonical set convention, the analytic bound, the query point, and the relevant implementation version. A restart test checks whether those dependencies survive process lifetime. These are practical consequences of asking which mathematical object the program is actually answering questions about.

A comparison record that another person can inspect

The final record should make the route from observation to decision recoverable. For the constructions in this book, that means identifying the object, its representation, the metric, the allowed path family, the chosen observable and its uncertainty. The necessary detail varies with the task, but the logical roles remain stable.

For a camera registration, keep correspondences, frames, fitted transforms and individual residuals. For a lifted attitude history, keep timestamps, conventions, sign choices and the motion assumption. For a GIGI completion, keep coefficients or their source record, supplied nodes and

values, query result, and separately computed validation truth. For a metric certificate, keep the exact matrices and the bounds used to certify their inequalities.

A successful comparison is then more than a low final score. It is a chain in which each map has a defined domain, each ruler has a meaning, and each conclusion follows from a checkable statement. If a comparison fails, trace the chain to find the first assumption whose failure changes the answer.

A change in software or coordinates can preserve the object and the comparison. The geometry specifies what has to survive, and the audit checks whether the implementation preserved it.

Appendix H

The complete measured-field register

Chapter 24 follows one GIGI calculation in detail. We ran the same construction on all seventeen measured fields. It retains every bin count and every final withheld-query radius from the recorded study, with the separately computed polynomial values and containment ratios.

The wine source contains 6,497 records and twelve fields. The power-plant source contributes 9,568 records from its first worksheet and five fields. The source archives, selection rules and licenses are identified in the references [12, 13, 23]. The other worksheets are shuffled copies, not additional independent samples.

The preparation is common to every field: 32 equal-width bins over that field's observed range; a normalized polynomial of degree at most 31; and seven supplied values at the declared disk nodes. The 24 withheld probe locations are the same for every field.

The replay reproduced the saved engine radii. This register reconstructs each histogram from the pinned raw archive and independently recomputes its disks by the Pick-matrix Schur complement. The 408 rows test prepared fingerprints; they do not predict physical measurements.

For the baseline in Table 24.6, require zeros at 0 and the six nodes $ae^{2\pi ij/6}$ with $a = 0.65$. Applying Schwarz's lemma after moving each zero to the origin gives $H = BG$ with $\|G\|_\infty \le 1$, where B is the product of the seven disk automorphism factors. Constant choices of G fill the disk centered at zero, of radius

$$\rho_0(q) = |B(q)| = \left| \frac{q(q^6 - a^6)}{1 - a^6 q^6} \right|.$$

The median over the 24 common probes is 0.0381204. The table divides each field median by this number before rounding.

The displayed containment ratio is a deterministic check: a value of 0.5 means that the known polynomial value lies halfway from the disk center to its boundary. It does not mean 50 percent confidence. The completion family consists of all bounded holomorphic functions satisfying the supplied observations, without additionally restricting them to nonnegative polynomial coefficients.

Each field follows the same sequence: measurement range, bin counts, supplied values, and withheld queries. Units belong to the measurements and bin edges; complex outputs and radii are dimensionless.

Wine: Fixed acidity

This distribution is concentrated toward the lower half of its observed range. The polynomial records the ordered bin proportions; it does not retain a pairing with the alcohol or quality measurements from the same bottle.

The source supplies 6,497 records, with observed minimum 3.8 and maximum 15.9 in g/dm^3. There are 31 occupied bins. The largest count is 1188, in bin 7; its lower and upper edges are 6.446875 and 6.825. The bin index measures position in this declared interval system.

The complete coefficient record

Use $p_k = n_k/6497$ and $H(z) = \sum_{k=0}^{31} p_k z^k$. The edges follow $b_k = 3.8 + (0.378125)k$. Decimal edge parameters are displayed to nine significant digits; the saved source endpoints define the unrounded construction.

k	n_k	k	n_k
0	2	16	84
1	6	17	60
2	25	18	30
3	122	19	24
4	150	20	33
5	524	21	24
6	892	22	13
7	1188	23	20
8	1093	24	8
9	635	25	5
10	603	26	4
11	397	27	2
12	203	28	0
13	144	29	2
14	125	30	2
15	74	31	3

The exact count sum is 6497. Empty bins are retained as zero coefficients; shifting the later coefficients left would create a different fingerprint.

The values supplied to completion

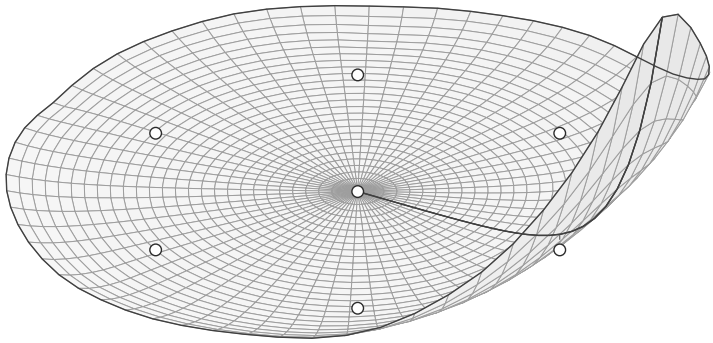
The observation nodes are $z_0 = 0$ and $z_{j+1} = 0.65e^{2\pi ij/6}$ for $j = 0, \ldots, 5$. The index in the following table identifies those nodes. These seven values are calculated after the entire histogram has been prepared.

Node	$\operatorname{Re} H(z_j)$	$\operatorname{Im} H(z_j)$
0	0.0003078	0.0000000
1	0.0500054	0.0000000
2	0.0072070	0.0012104
3	0.0020597	-0.0016466
4	-0.0034859	0.0000000
5	0.0020597	0.0016466
6	0.0072070	-0.0012104

The displayed decimals are for reading, not a replacement for the exact counts. Horner evaluation of the coefficient record reconstructs the values at any desired arithmetic precision.

The prepared data defines a surface of values

Height = |H(z)|; H(z) = Σ $p_k z^k$. This is the constructed analytic fingerprint.



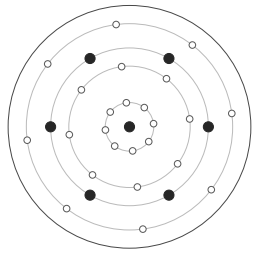
6,497 source records

32 nonnegative coefficients

Σ p_k = 1

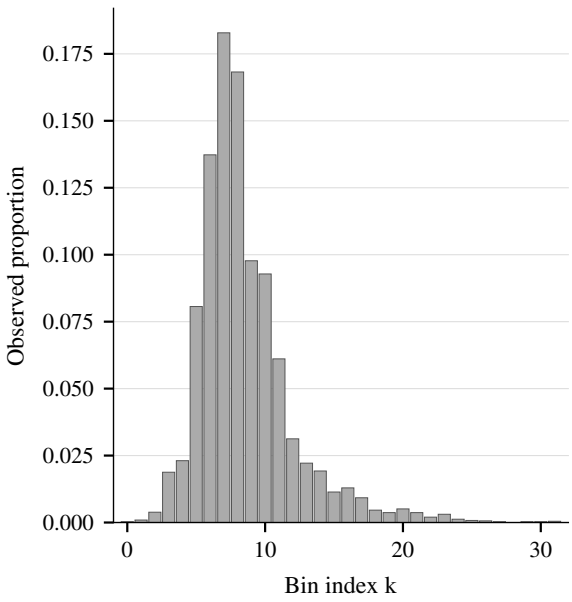
7 supplied function values

24 withheld queries in this field



Filled: supplied / open: withheld locations

Measured distribution



Completion by probe location

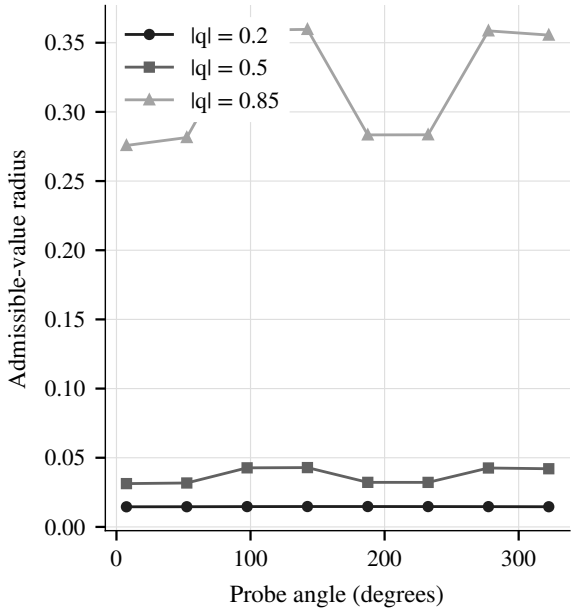


Figure H.1: Wine: Fixed acidity. The histogram and all retained query radii are computed from the same coefficient record. Connecting segments organize the eight sampled angles; they do not assert a bound between those angles.

The complete withheld-query register

For each radius $r = 0.2, 0.5, 0.85$, the eight directions are $q_{r,k} = r\exp(i(0.13 + 2\pi k/8))$, $k = 0, \dots, 7$. The last column is $|H(q) - c(q)|/\rho(q)$, computed from the full-precision values. A ratio at most one records containment in the feasible disk. It is not a probability or a percentage prediction error.

r, k	$\operatorname{Re} H(q)$	$\operatorname{Im} H(q)$	$\rho(q)$	Ratio
0.2,0	0.0008392	0.0001631	0.0145466	0.26461
0.2,1	0.0002154	0.0003036	0.0145952	0.22848
0.2,2	0.0002045	0.0000363	0.0146850	0.17704
0.2,3	0.0002466	0.0000849	0.0147362	0.14345
0.2,4	0.0001510	-0.0000341	0.0147324	0.13490
0.2,5	0.0002698	-0.0000545	0.0147054	0.15189
0.2,6	0.0001658	-0.0000904	0.0146620	0.19296
0.2,7	0.0003721	-0.0004059	0.0145905	0.24476
0.5,0	0.0099616	0.0069498	0.0313078	0.41761
0.5,1	-0.0003375	-0.0016205	0.0317559	0.25787
0.5,2	-0.0002995	-0.0010663	0.0427102	0.13468
0.5,3	0.0012329	0.0010572	0.0428860	0.08018
0.5,4	-0.0012188	-0.0006778	0.0322170	0.06770
0.5,5	0.0014640	0.0001633	0.0321951	0.09293
0.5,6	-0.0008724	-0.0000396	0.0425970	0.16519
0.5,7	-0.0048071	-0.0002308	0.0420552	0.31878
0.85,0	0.1618263	0.2153513	0.2757391	0.72009
0.85,1	0.0610397	-0.0080361	0.2814669	0.21892
0.85,2	0.0057453	-0.0161308	0.3590777	0.07735
0.85,3	0.0146065	0.0077857	0.3596272	0.03618
0.85,4	-0.0077851	0.0001483	0.2833704	0.01463
0.85,5	0.0040242	0.0099640	0.2834430	0.04165
0.85,6	-0.0263508	0.0042574	0.3585770	0.10405
0.85,7	-0.0289395	0.1096412	0.3555086	0.35290

The median radius is 0.0371361 and the largest is 0.3596272 at $r = 0.85$, $k = 3$. At that query the center is $0.0017278 + 0.0096277i$ and the truth-to-center gap is 0.0130098. All 24 retained values satisfy the containment check.

What this field contributes

At probe radius 0.2, the eight admissible radii range from 0.0145466 to 0.0147362. At radius 0.5 they range from 0.0313078 to 0.0428860; at radius 0.85, from 0.2757391 to 0.3596272. These intervals separate the radial change from the variation across the eight directions.

The largest truth-to-center ratio among this field's 24 queries is 0.720090. The shared model and probe arrangement make the radii and containment ratios comparable across fields. Physical units belong to the histogram convention.

Wine: Volatile acidity

The long upper tail occupies relatively few bins with small counts. A bounded-function completion respects the supplied evaluations, but it does not enforce a chemical model or infer an unmeasured acidity.

The source supplies 6,497 records, with observed minimum 0.08 and maximum 1.58 in g/dm^3. There are 27 occupied bins. The largest count is 1104, in bin 4; its lower and upper edges are 0.2675 and 0.314375. The bin index measures position in this declared interval system.

The complete coefficient record

Use $p_k = n_k/6497$ and $H(z) = \sum_{k=0}^{31} p_k z^k$. The edges follow $b_k = 0.08 + (0.046875)k$. Decimal edge parameters are displayed to nine significant digits; the saved source endpoints define the unrounded construction.

k	n_k	k	n_k
0	74	16	32
1	483	17	25
2	1019	18	12
3	1008	19	10
4	1104	20	9
5	799	21	3
6	392	22	2
7	363	23	2
8	259	24	1
9	161	25	0
10	211	26	2
11	207	27	0
12	131	28	0
13	98	29	0
14	60	30	0
15	29	31	1

The exact count sum is 6497. Empty bins are retained as zero coefficients; shifting the later coefficients left would create a different fingerprint.

The values supplied to completion

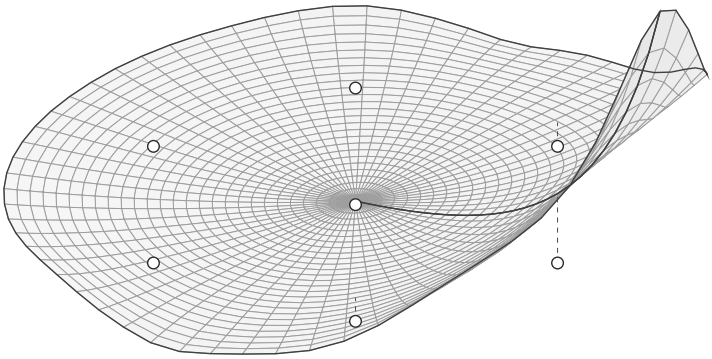
The observation nodes are $z_0 = 0$ and $z_{j+1} = 0.65e^{2\pi ij/6}$ for $j = 0, \ldots, 5$. The index in the following table identifies those nodes. These seven values are calculated after the entire histogram has been prepared.

Node	$\mathrm{Re}\, H(z_j)$	$\mathrm{Im}\, H(z_j)$
0	0.0113899	0.0000000
1	0.2231850	0.0000000
2	-0.0434047	0.0635232
3	-0.0228169	-0.0001876
4	0.0055929	-0.0000000
5	-0.0228169	0.0001876
6	-0.0434047	-0.0635232

The displayed decimals are for reading, not a replacement for the exact counts. Horner evaluation of the coefficient record reconstructs the values at any desired arithmetic precision.

The prepared data defines a surface of values

Height = |H(z)|; H(z) = Σ $p_k z^k$. This is the constructed analytic fingerprint.



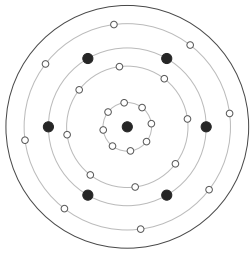
6,497 source records

32 nonnegative coefficients

Σ p_k = 1

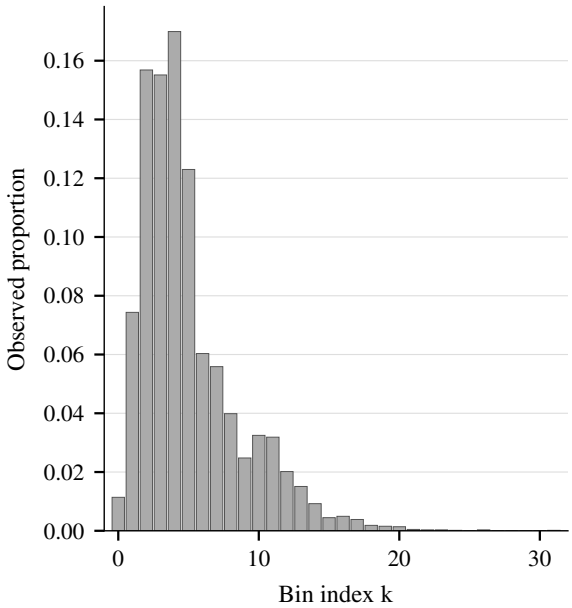
7 supplied function values

24 withheld queries in this field



Filled: supplied / open: withheld locations

Measured distribution



Completion by probe location

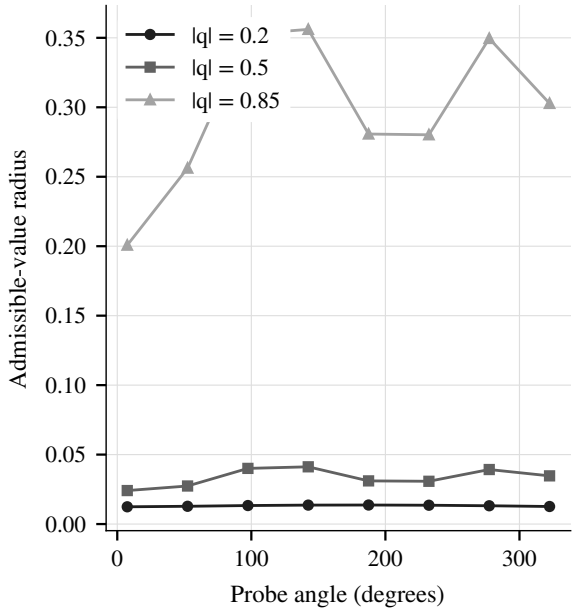


Figure H.2: Wine: Volatile acidity. The histogram and all retained query radii are computed from the same coefficient record. Connecting segments organize the eight sampled angles; they do not assert a bound between those angles.

The complete withheld-query register

For each radius $r = 0.2, 0.5, 0.85$, the eight directions are $q_{r,k} = r\exp(i(0.13 + 2\pi k/8))$, $k = 0, \ldots, 7$. The last column is $|H(q) - c(q)|/\rho(q)$, computed from the full-precision values. A ratio at most one records containment in the feasible disk. It is not a probability or a percentage prediction error.

r, k	$\mathrm{Re}\, H(q)$	$\mathrm{Im}\, H(q)$	$\rho(q)$	Ratio
0.2,0	0.0336141	0.0041745	0.0124207	0.15652
0.2,1	0.0174557	0.0181520	0.0127900	0.12660
0.2,2	0.0040815	0.0121455	0.0133250	0.09412
0.2,3	0.0014935	0.0040062	0.0136417	0.07862
0.2,4	0.0017687	-0.0006730	0.0137012	0.07512
0.2,5	0.0016320	-0.0063019	0.0135487	0.08219
0.2,6	0.0070391	-0.0151062	0.0131669	0.10293
0.2,7	0.0240347	-0.0163964	0.0126381	0.13936
0.5,0	0.1173991	0.0311149	0.0241076	0.30036
0.5,1	-0.0025523	0.0653035	0.0273909	0.11810
0.5,2	-0.0173434	0.0163701	0.0400685	0.05814
0.5,3	-0.0064908	-0.0027827	0.0412061	0.04917
0.5,4	0.0011135	0.0012664	0.0310710	0.04313
0.5,5	-0.0118379	-0.0011367	0.0307916	0.05186
0.5,6	-0.0184284	-0.0270862	0.0392474	0.06680
0.5,7	0.0298896	-0.0819743	0.0346911	0.17603
0.85,0	0.4283715	0.2135076	0.2007395	0.68525
0.85,1	-0.1227518	0.0942105	0.2561859	0.08271
0.85,2	-0.0258725	0.0126696	0.3526927	0.01764
0.85,3	-0.0060659	-0.0280304	0.3560645	0.03942
0.85,4	0.0131300	0.0043835	0.2807468	0.01023
0.85,5	-0.0346316	0.0216484	0.2801952	0.05065
0.85,6	-0.0409514	-0.0039323	0.3495311	0.03146
0.85,7	-0.0775690	-0.2369573	0.3027708	0.20442

The median radius is 0.0328811 and the largest is 0.3560645 at $r = 0.85$, $k = 3$. At that query the center is $-0.0105456 - 0.0147295i$ and the truth-to-center gap is 0.0140350. All 24 retained values satisfy the containment check.

What this field contributes

At probe radius 0.2, the eight admissible radii range from 0.0124207 to 0.0137012. At radius 0.5 they range from 0.0241076 to 0.0412061; at radius 0.85, from 0.2007395 to 0.3560645. These intervals separate the radial change from the variation across the eight directions.

The largest truth-to-center ratio among this field's 24 queries is 0.685250. The shared model and probe arrangement make the radii and containment ratios comparable across fields. Physical units belong to the histogram convention.

Wine: Citric acid

Zero is an observed endpoint in this field. The first bin therefore has a direct physical interpretation. The value at probe zero is the first-bin proportion, not the mean concentration.

The source supplies 6,497 records, with observed minimum 0 and maximum 1.66 in g/dm^3. There are 21 occupied bins. The largest count is 1605, in bin 5; its lower and upper edges are 0.259375 and 0.31125. The bin index measures position in this declared interval system.

The complete coefficient record

Use $p_k = n_k/6497$ and $H(z) = \sum_{k=0}^{31} p_k z^k$. The edges follow $b_k = 0 + (0.051875)k$. Decimal edge parameters are displayed to nine significant digits; the saved source endpoints define the unrounded construction.

k	n_k	k	n_k
0	345	16	2
1	192	17	2
2	187	18	0
3	320	19	7
4	733	20	0
5	1605	21	0
6	1093	22	0
7	662	23	1
8	400	24	0
9	494	25	0
10	177	26	0
11	85	27	0
12	74	28	0
13	48	29	0
14	57	30	0
15	12	31	1

The exact count sum is 6497. Empty bins are retained as zero coefficients; shifting the later coefficients left would create a different fingerprint.

The values supplied to completion

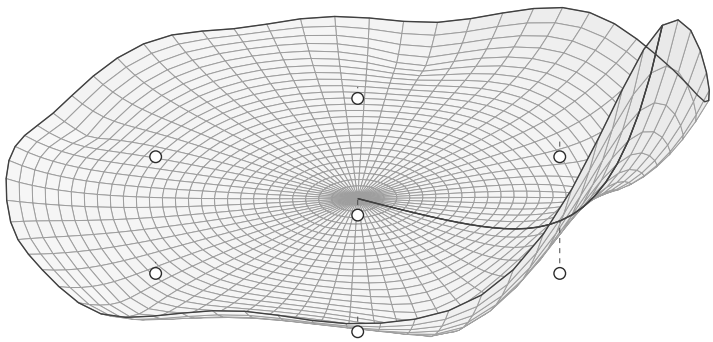
The observation nodes are $z_0 = 0$ and $z_{j+1} = 0.65e^{2\pi ij/6}$ for $j = 0, \ldots, 5$. The index in the following table identifies those nodes. These seven values are calculated after the entire histogram has been prepared.

Node	$\mathrm{Re}\, H(z_j)$	$\mathrm{Im}\, H(z_j)$
0	0.0531014	0.0000000
1	0.1686176	0.0000000
2	0.0599300	-0.0094480
3	0.0371282	0.0015729
4	0.0323903	0.0000000
5	0.0371282	-0.0015729
6	0.0599300	0.0094480

The displayed decimals are for reading, not a replacement for the exact counts. Horner evaluation of the coefficient record reconstructs the values at any desired arithmetic precision.

The prepared data defines a surface of values

Height = $|H(z)|$; $H(z) = \Sigma\, p_k z^k$. This is the constructed analytic fingerprint.



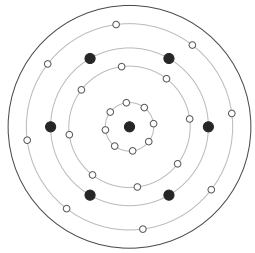
6,497 source records

32 nonnegative coefficients

$\Sigma\, p_k = 1$

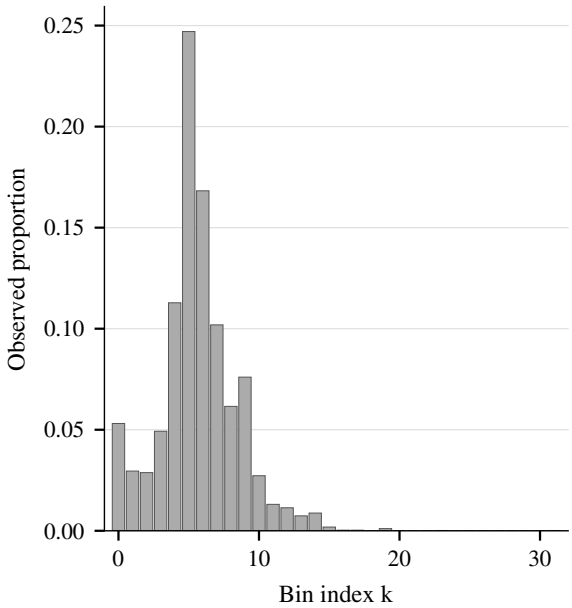
7 supplied function values

24 withheld queries in this field



Filled: supplied / open: withheld locations

Measured distribution



Completion by probe location

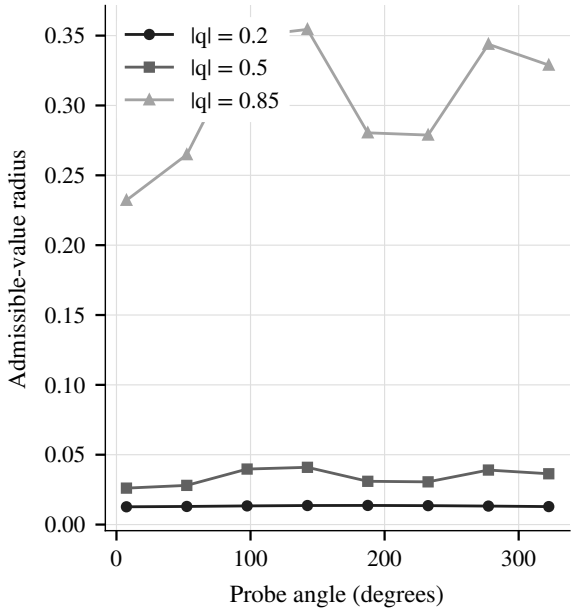


Figure H.3: Wine: Citric acid. The histogram and all retained query radii are computed from the same coefficient record. Connecting segments organize the eight sampled angles; they do not assert a bound between those angles.

The complete withheld-query register

For each radius $r = 0.2, 0.5, 0.85$, the eight directions are $q_{r,k} = r\exp(i(0.13 + 2\pi k/8))$, $k = 0, \dots, 7$. The last column is $|H(q) - c(q)|/\rho(q)$, computed from the full-precision values. A ratio at most one records containment in the feasible disk. It is not a probability or a percentage prediction error.

r, k	$\mathrm{Re}\, H(q)$	$\mathrm{Im}\, H(q)$	$\rho(q)$	Ratio
0.2,0	0.0606672	0.0013583	0.0127178	0.23055
0.2,1	0.0558857	0.0057748	0.0129630	0.18640
0.2,2	0.0514747	0.0053445	0.0133592	0.13126
0.2,3	0.0487777	0.0027594	0.0136329	0.10710
0.2,4	0.0480897	-0.0005715	0.0136892	0.10356
0.2,5	0.0494272	-0.0037441	0.0135463	0.11177
0.2,6	0.0528011	-0.0057720	0.0132365	0.14674
0.2,7	0.0576889	-0.0051484	0.0128662	0.20599
0.5,0	0.0952107	0.0171392	0.0260826	0.41681
0.5,1	0.0500892	0.0084622	0.0280482	0.20884
0.5,2	0.0467635	0.0146603	0.0397286	0.05948
0.5,3	0.0453944	0.0044141	0.0409800	0.05677
0.5,4	0.0411333	-0.0022621	0.0309688	0.07093
0.5,5	0.0440659	-0.0048757	0.0305798	0.04770
0.5,6	0.0542625	-0.0146819	0.0390236	0.08935
0.5,7	0.0488660	-0.0211968	0.0363637	0.28854
0.85,0	0.3462739	0.2277272	0.2321179	0.76823
0.85,1	0.0428409	-0.0981693	0.2647807	0.21062
0.85,2	-0.0007790	0.0600526	0.3493676	0.08746
0.85,3	0.0628904	-0.0012665	0.3544497	0.04624
0.85,4	0.0119747	-0.0249654	0.2804218	0.07120
0.85,5	0.0480354	0.0186737	0.2788036	0.03060
0.85,6	0.0933671	-0.0813304	0.3438257	0.12507
0.85,7	-0.1119406	0.0151821	0.3288737	0.33983

The median radius is 0.0336663 and the largest is 0.3544497 at $r = 0.85$, $k = 3$. At that query the center is $0.0481966 + 0.0059905i$ and the truth-to-center gap is 0.0163881. All 24 retained values satisfy the containment check.

What this field contributes

At probe radius 0.2, the eight admissible radii range from 0.0127178 to 0.0136892. At radius 0.5 they range from 0.0260826 to 0.0409800; at radius 0.85, from 0.2321179 to 0.3544497. These intervals separate the radial change from the variation across the eight directions.

The largest truth-to-center ratio among this field's 24 queries is 0.768228. The shared model and probe arrangement make the radii and containment ratios comparable across fields. Physical units belong to the histogram convention.

Wine: Residual sugar

A concentrated lower-range distribution and a long tail are visible in the counts. The fixed 32-bin construction may summarize fine structure near the concentration poorly; the counts make that resolution choice inspectable.

The source supplies 6,497 records, with observed minimum 0.6 and maximum 65.8 in g/dm^3. There are 15 occupied bins. The largest count is 3011, in bin 0; its lower and upper edges are 0.6 and 2.6375. The bin index measures position in this declared interval system.

The complete coefficient record

Use $p_k = n_k/6497$ and $H(z) = \sum_{k=0}^{31} p_k z^k$. The edges follow $b_k = 0.6 + (2.0375)k$. Decimal edge parameters are displayed to nine significant digits; the saved source endpoints define the unrounded construction.

k	n_k	k	n_k
0	3011	16	0
1	758	17	0
2	656	18	0
3	624	19	0
4	392	20	0
5	371	21	0
6	332	22	0
7	189	23	0
8	118	24	0
9	37	25	0
10	3	26	0
11	1	27	0
12	2	28	0
13	0	29	0
14	0	30	0
15	2	31	1

The exact count sum is 6497. Empty bins are retained as zero coefficients; shifting the later coefficients left would create a different fingerprint.

The values supplied to completion

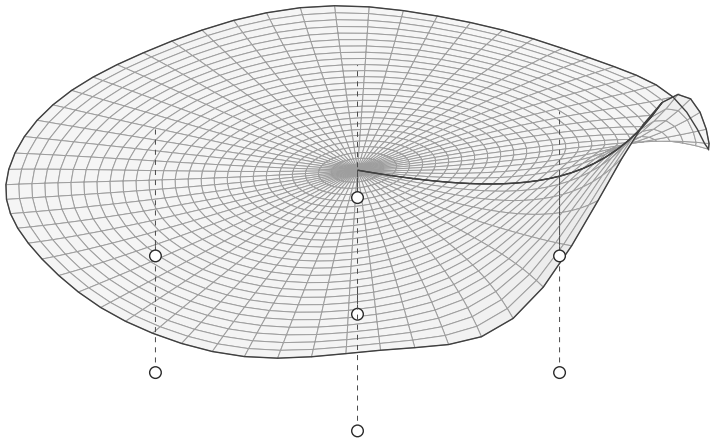
The observation nodes are $z_0 = 0$ and $z_{j+1} = 0.65e^{2\pi ij/6}$ for $j = 0, \ldots, 5$. The index in the following table identifies those nodes. These seven values are calculated after the entire histogram has been prepared.

Node	$\mathrm{Re}\, H(z_j)$	$\mathrm{Im}\, H(z_j)$
0	0.4634447	0.0000000
1	0.6316979	0.0000000
2	0.4557423	0.0892838
3	0.4248435	0.0330581
4	0.4109326	0.0000000
5	0.4248435	-0.0330581
6	0.4557423	-0.0892838

The displayed decimals are for reading, not a replacement for the exact counts. Horner evaluation of the coefficient record reconstructs the values at any desired arithmetic precision.

The prepared data defines a surface of values

Height = |H(z)|; H(z) = Σ $p_k z^k$. This is the constructed analytic fingerprint.



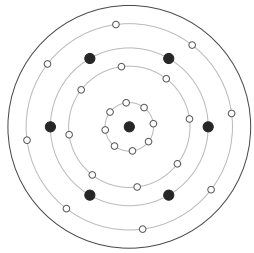
6,497 source records

32 nonnegative coefficients

Σ p_k = 1

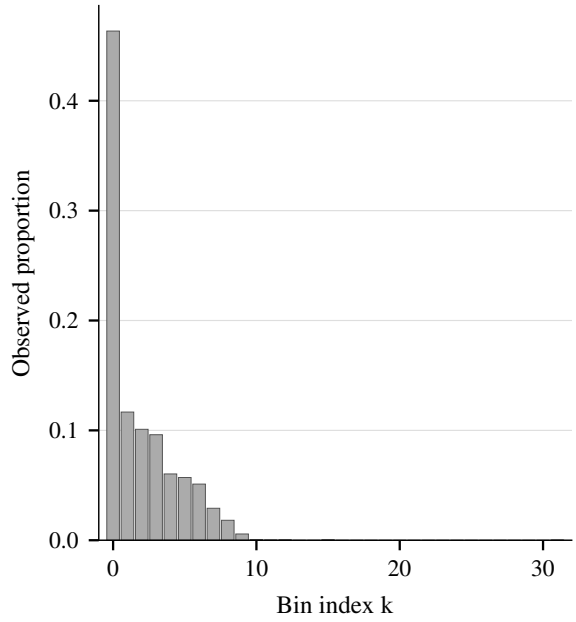
7 supplied function values

24 withheld queries in this field



Filled: supplied / open: withheld locations

Measured distribution



Completion by probe location

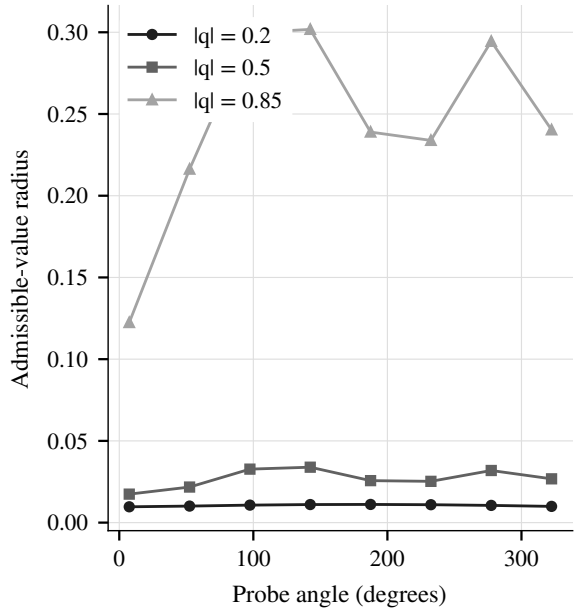


Figure H.4: Wine: Residual sugar. The histogram and all retained query radii are computed from the same coefficient record. Connecting segments organize the eight sampled angles; they do not assert a bound between those angles.

The complete withheld-query register

For each radius $r = 0.2, 0.5, 0.85$, the eight directions are $q_{r,k} = r\exp(i(0.13 + 2\pi k/8))$, $k = 0, \ldots, 7$. The last column is $|H(q) - c(q)|/\rho(q)$, computed from the full-precision values. A ratio at most one records containment in the feasible disk. It is not a probability or a percentage prediction error.

r, k	$\mathrm{Re}\, H(q)$	$\mathrm{Im}\, H(q)$	$\rho(q)$	Ratio
0.2,0	0.4912962	0.0044169	0.0096789	0.15856
0.2,1	0.4758351	0.0226299	0.0101144	0.12934
0.2,2	0.4568796	0.0214480	0.0107065	0.09576
0.2,3	0.4462118	0.0109790	0.0110522	0.07891
0.2,4	0.4435714	-0.0022397	0.0111236	0.07541
0.2,5	0.4488147	-0.0149243	0.0109546	0.08270
0.2,6	0.4623666	-0.0234332	0.0105331	0.10519
0.2,7	0.4825820	-0.0188763	0.0099322	0.14189
0.5,0	0.5622257	0.0223825	0.0174289	0.29788
0.5,1	0.4787345	0.0711565	0.0217596	0.12836
0.5,2	0.4378829	0.0429075	0.0327237	0.06092
0.5,3	0.4262986	0.0205231	0.0339052	0.04517
0.5,4	0.4212040	-0.0044099	0.0256852	0.04408
0.5,5	0.4298252	-0.0271287	0.0252570	0.04331
0.5,6	0.4456998	-0.0531401	0.0318883	0.07018
0.5,7	0.5059740	-0.0718013	0.0267916	0.17595
0.85,0	0.7567948	0.1115980	0.1224056	0.70913
0.85,1	0.4446654	0.1240229	0.2162449	0.15150
0.85,2	0.4094920	0.0463119	0.2997330	0.13819
0.85,3	0.4130528	0.0334686	0.3017594	0.14578
0.85,4	0.3982192	-0.0068838	0.2389389	0.10186
0.85,5	0.4219932	-0.0332294	0.2338438	0.11825
0.85,6	0.4086254	-0.0711807	0.2944442	0.10337
0.85,7	0.4747570	-0.1699629	0.2402968	0.02387

The median radius is 0.0262384 and the largest is 0.3017594 at $r = 0.85$, $k = 3$. At that query the center is $0.3693179 + 0.0287464i$ and the truth-to-center gap is 0.0439890. All 24 retained values satisfy the containment check.

What this field contributes

At probe radius 0.2, the eight admissible radii range from 0.0096789 to 0.0111236. At radius 0.5 they range from 0.0174289 to 0.0339052; at radius 0.85, from 0.1224056 to 0.3017594. These intervals separate the radial change from the variation across the eight directions.

The largest truth-to-center ratio among this field's 24 queries is 0.709133. The shared model and probe arrangement make the radii and containment ratios comparable across fields. Physical units belong to the histogram convention.

Wine: Chlorides

The comparatively narrow main concentration and sparse tail illustrate why a small fingerprint value at some probes need not mean that the measurements themselves are small. Probe position and bin order determine the value.

The source supplies 6,497 records, with observed minimum 0.009 and maximum 0.611 in g/dm^3. There are 23 occupied bins. The largest count is 2825, in bin 1; its lower and upper edges are 0.0278125 and 0.046625. The bin index measures position in this declared interval system.

The complete coefficient record

Use $p_k = n_k/6497$ and $H(z) = \sum_{k=0}^{31} p_k z^k$. The edges follow $b_k = 0.009 + (0.0188125)k$. Decimal edge parameters are displayed to nine significant digits; the saved source endpoints define the unrounded construction.

k	n_k	k	n_k
0	292	16	0
1	2825	17	5
2	1773	18	2
3	902	19	2
4	395	20	3
5	123	21	7
6	45	22	0
7	32	23	0
8	37	24	2
9	13	25	0
10	16	26	0
11	6	27	0
12	8	28	0
13	5	29	0
14	1	30	0
15	1	31	2

The exact count sum is 6497. Empty bins are retained as zero coefficients; shifting the later coefficients left would create a different fingerprint.

The values supplied to completion

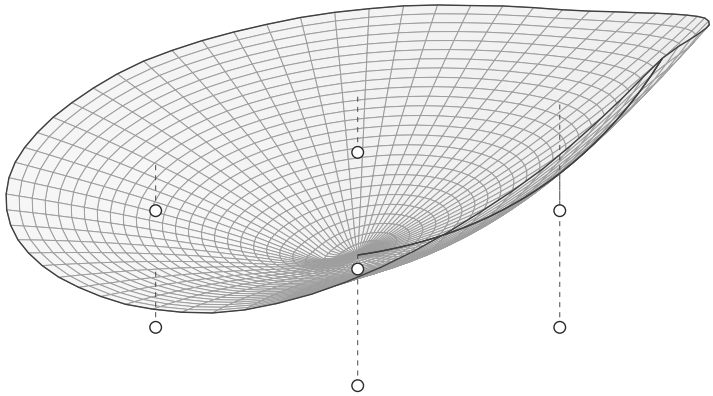
The observation nodes are $z_0 = 0$ and $z_{j+1} = 0.65e^{2\pi ij/6}$ for $j = 0, \ldots, 5$. The index in the following table identifies those nodes. These seven values are calculated after the entire histogram has been prepared.

Node	$\operatorname{Re} H(z_j)$	$\operatorname{Im} H(z_j)$
0	0.0449438	0.0000000
1	0.4950883	0.0000000
2	0.0866615	0.3336483
3	-0.1220808	0.1524857
4	-0.1514096	0.0000000
5	-0.1220808	-0.1524857
6	0.0866615	-0.3336483

The displayed decimals are for reading, not a replacement for the exact counts. Horner evaluation of the coefficient record reconstructs the values at any desired arithmetic precision.

The prepared data defines a surface of values

Height = |H(z)|; H(z) = Σ $p_k z^k$. This is the constructed analytic fingerprint.



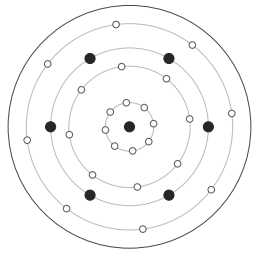
6,497 source records

32 nonnegative coefficients

Σ p_k = 1

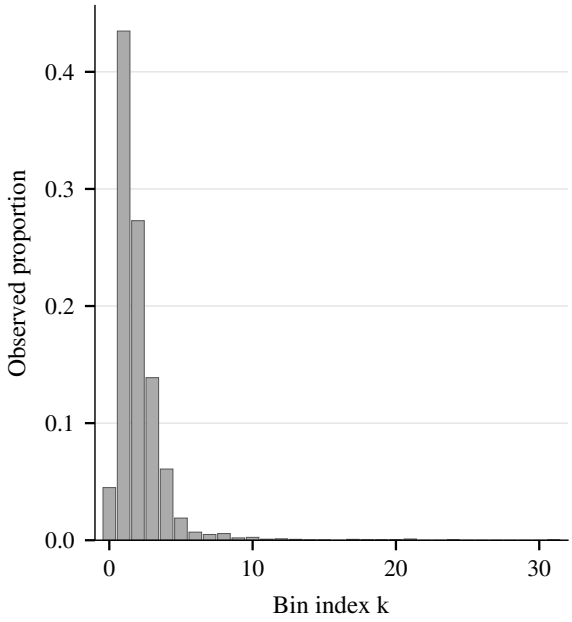
7 supplied function values

24 withheld queries in this field



Filled: supplied / open: withheld locations

Measured distribution



Completion by probe location

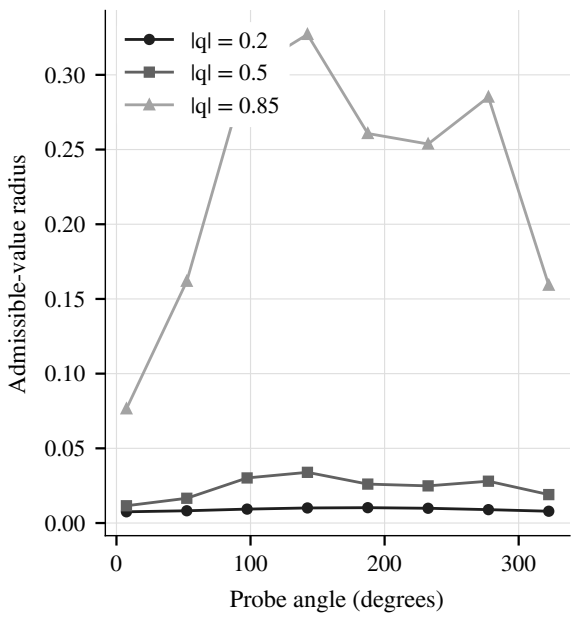


Figure H.5: Wine: Chlorides. The histogram and all retained query radii are computed from the same coefficient record. Connecting segments organize the eight sampled angles; they do not assert a bound between those angles.

The complete withheld-query register

For each radius $r = 0.2, 0.5, 0.85$, the eight directions are $q_{r,k} = r\exp(i(0.13 + 2\pi k/8))$, $k = 0, \dots, 7$. The last column is $|H(q) - c(q)|/\rho(q)$, computed from the full-precision values. A ratio at most one records containment in the feasible disk. It is not a probability or a percentage prediction error.

r, k	$\operatorname{Re} H(q)$	$\operatorname{Im} H(q)$	$\rho(q)$	Ratio
0.2,0	0.1428390	0.0145543	0.0075171	0.16140
0.2,1	0.0940297	0.0798670	0.0082068	0.12760
0.2,2	0.0236242	0.0824487	0.0093176	0.08887
0.2,3	-0.0208457	0.0434291	0.0100887	0.06789
0.2,4	-0.0316840	-0.0088445	0.0102756	0.06296
0.2,5	-0.0099227	-0.0588664	0.0098811	0.07291
0.2,6	0.0453338	-0.0879651	0.0089639	0.09989
0.2,7	0.1161764	-0.0646229	0.0078962	0.14205
0.5,0	0.3463817	0.0546989	0.0115590	0.32848
0.5,1	0.1406384	0.2424538	0.0165744	0.12159
0.5,2	-0.0396723	0.1842614	0.0301808	0.05079
0.5,3	-0.1059680	0.0806772	0.0339387	0.02826
0.5,4	-0.1178597	-0.0156489	0.0260653	0.02350
0.5,5	-0.0922538	-0.1144832	0.0248751	0.03583
0.5,6	0.0041608	-0.2156812	0.0279946	0.07068
0.5,7	0.2242134	-0.2161245	0.0190372	0.18471
0.85,0	0.7188653	0.1577137	0.0764600	0.76578
0.85,1	0.1156547	0.4922165	0.1618393	0.15572
0.85,2	-0.1389934	0.2577111	0.3022243	0.10899
0.85,3	-0.1846333	0.0978227	0.3270827	0.08577
0.85,4	-0.1873400	-0.0170911	0.2607578	0.06570
0.85,5	-0.1778933	-0.1438369	0.2537040	0.08691
0.85,6	-0.0995267	-0.3291654	0.2851262	0.15011
0.85,7	0.3196520	-0.5046631	0.1592020	0.30326

The median radius is 0.0254702 and the largest is 0.3270827 at $r = 0.85$, $k = 3$. At that query the center is $-0.1578368 + 0.0895227i$ and the truth-to-center gap is 0.0280525. All 24 retained values satisfy the containment check.

What this field contributes

At probe radius 0.2, the eight admissible radii range from 0.0075171 to 0.0102756. At radius 0.5 they range from 0.0115590 to 0.0339387; at radius 0.85, from 0.0764600 to 0.3270827. These intervals separate the radial change from the variation across the eight directions.

The largest truth-to-center ratio among this field's 24 queries is 0.765776. The shared model and probe arrangement make the radii and containment ratios comparable across fields. Physical units belong to the histogram convention.

Wine: Free sulfur dioxide

This field and total sulfur dioxide share a physical unit but measure different quantities. Their polynomials are not interchangeable observations of one chemical variable.

The source supplies 6,497 records, with observed minimum 1 and maximum 289 in mg/dm^3. There are 18 occupied bins. The largest count is 1271, in bin 3; its lower and upper edges are 28 and 37. The bin index measures position in this declared interval system.

The complete coefficient record

Use $p_k = n_k/6497$ and $H(z) = \sum_{k=0}^{31} p_k z^k$. The edges follow $b_k = 1 + (9)k$. Decimal edge parameters are displayed to nine significant digits; the saved source endpoints define the unrounded construction.

k	n_k	k	n_k
0	694	16	1
1	1147	17	0
2	1232	18	0
3	1271	19	0
4	887	20	0
5	653	21	0
6	372	22	0
7	145	23	0
8	53	24	0
9	17	25	0
10	9	26	0
11	7	27	0
12	2	28	0
13	3	29	0
14	2	30	0
15	1	31	1

The exact count sum is 6497. Empty bins are retained as zero coefficients; shifting the later coefficients left would create a different fingerprint.

The values supplied to completion

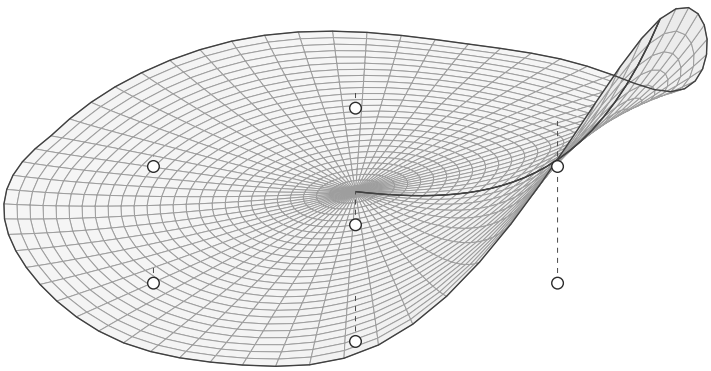
The observation nodes are $z_0 = 0$ and $z_{j+1} = 0.65e^{2\pi ij/6}$ for $j = 0, \ldots, 5$. The index in the following table identifies those nodes. These seven values are calculated after the entire histogram has been prepared.

Node	$\operatorname{Re} H(z_j)$	$\operatorname{Im} H(z_j)$
0	0.1068185	0.0000000
1	0.3972045	0.0000000
2	0.0687361	0.1387077
3	0.0487741	0.0417330
4	0.0346065	0.0000000
5	0.0487741	-0.0417330
6	0.0687361	-0.1387077

The displayed decimals are for reading, not a replacement for the exact counts. Horner evaluation of the coefficient record reconstructs the values at any desired arithmetic precision.

The prepared data defines a surface of values

Height = |H(z)|; H(z) = Σ $p_k z^k$. This is the constructed analytic fingerprint.



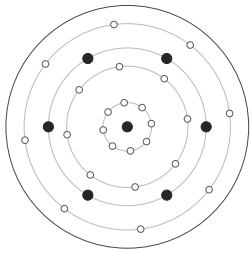
6,497 source records

32 nonnegative coefficients

Σ p_k = 1

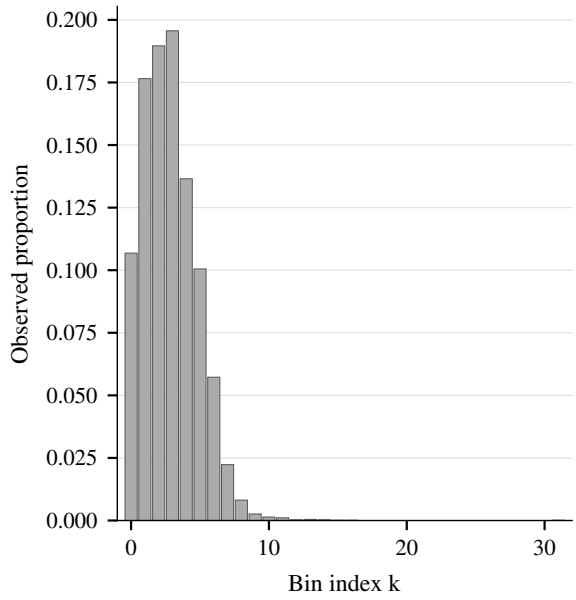
7 supplied function values

24 withheld queries in this field



Filled: supplied / open: withheld locations

Measured distribution



Completion by probe location

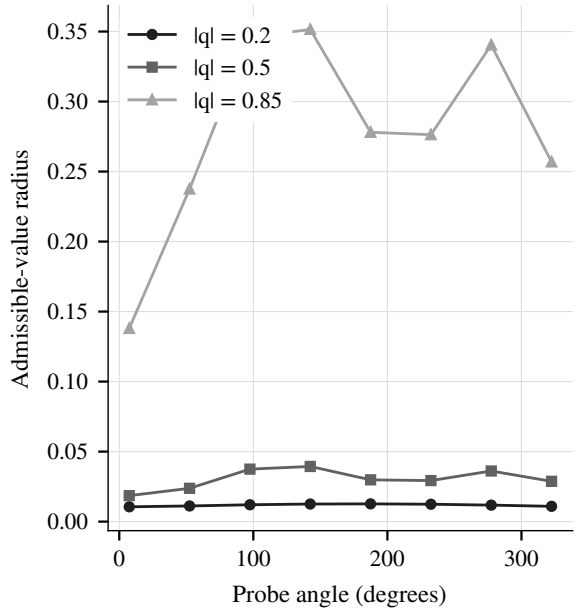


Figure H.6: Wine: Free sulfur dioxide. The histogram and all retained query radii are computed from the same coefficient record. Connecting segments organize the eight sampled angles; they do not assert a bound between those angles.

The complete withheld-query register

For each radius $r = 0.2, 0.5, 0.85$, the eight directions are $q_{r,k} = r\exp(i(0.13 + 2\pi k/8))$, $k = 0, \ldots, 7$. The last column is $|H(q) - c(q)|/\rho(q)$, computed from the full-precision values. A ratio at most one records containment in the feasible disk. It is not a probability or a percentage prediction error.

r, k	$\mathrm{Re}\, H(q)$	$\mathrm{Im}\, H(q)$	$\rho(q)$	Ratio
0.2,0	0.1508248	0.0072530	0.0105732	0.20217
0.2,1	0.1247530	0.0357828	0.0111700	0.16382
0.2,2	0.0956740	0.0317446	0.0120288	0.11825
0.2,3	0.0812182	0.0155233	0.0125494	0.09414
0.2,4	0.0778569	-0.0031308	0.0126603	0.08882
0.2,5	0.0846102	-0.0213448	0.0124075	0.09974
0.2,6	0.1036768	-0.0354326	0.0117709	0.13125
0.2,7	0.1359345	-0.0303954	0.0109128	0.18046
0.5,0	0.2734426	0.0398689	0.0185419	0.37920
0.5,1	0.1188495	0.1172801	0.0237780	0.16540
0.5,2	0.0638766	0.0587542	0.0374755	0.07173
0.5,3	0.0535547	0.0263778	0.0393793	0.04544
0.5,4	0.0479356	-0.0056986	0.0298319	0.04450
0.5,5	0.0568912	-0.0353498	0.0292946	0.05019
0.5,6	0.0717035	-0.0758551	0.0361448	0.09054
0.5,7	0.1684235	-0.1251576	0.0287702	0.23626
0.85,0	0.6176594	0.1866537	0.1379529	0.78565
0.85,1	0.0071688	0.2059786	0.2375217	0.02892
0.85,2	0.0264277	0.0556774	0.3458776	0.00723
0.85,3	0.0381167	0.0385103	0.3515211	0.01776
0.85,4	0.0175452	-0.0104868	0.2780326	0.01855
0.85,5	0.0444808	-0.0387763	0.2763075	0.01502
0.85,6	0.0175088	-0.0823173	0.3404121	0.03498
0.85,7	0.0945984	-0.3398237	0.2568600	0.24278

The median radius is 0.0295632 and the largest is 0.3515211 at $r = 0.85$, $k = 3$. At that query the center is $0.0331965 + 0.0346648i$ and the truth-to-center gap is 0.0062447. All 24 retained values satisfy the containment check.

What this field contributes

At probe radius 0.2, the eight admissible radii range from 0.0105732 to 0.0126603. At radius 0.5 they range from 0.0185419 to 0.0393793; at radius 0.85, from 0.1379529 to 0.3515211. These intervals separate the radial change from the variation across the eight directions.

The largest truth-to-center ratio among this field's 24 queries is 0.785651. The shared model and probe arrangement make the radii and containment ratios comparable across fields. Physical units belong to the histogram convention.

Wine: Total sulfur dioxide

The distribution combines red and white wine records under the stated source selection. The histogram preserves this pooled distribution and does not retain wine type as an additional label.

The source supplies 6,497 records, with observed minimum 6 and maximum 440 in mg/dm^3. There are 26 occupied bins. The largest count is 691, in bin 8; its lower and upper edges are 114.5 and 128.0625. The bin index measures position in this declared interval system.

The complete coefficient record

Use $p_k = n_k/6497$ and $H(z) = \sum_{k=0}^{31} p_k z^k$. The edges follow $b_k = 6 + (13.5625)k$. Decimal edge parameters are displayed to nine significant digits; the saved source endpoints define the unrounded construction.

k	n_k	k	n_k
0	335	16	77
1	396	17	43
2	283	18	14
3	257	19	2
4	264	20	3
5	369	21	2
6	491	22	2
7	664	23	0
8	691	24	1
9	552	25	0
10	563	26	1
11	454	27	0
12	396	28	0
13	299	29	0
14	196	30	0
15	141	31	1

The exact count sum is 6497. Empty bins are retained as zero coefficients; shifting the later coefficients left would create a different fingerprint.

The values supplied to completion

The observation nodes are $z_0 = 0$ and $z_{j+1} = 0.65e^{2\pi ij/6}$ for $j = 0, \ldots, 5$. The index in the following table identifies those nodes. These seven values are calculated after the entire histogram has been prepared.

Node	$\operatorname{Re} H(z_j)$	$\operatorname{Im} H(z_j)$
0	0.0515623	0.0000000
1	0.1525681	0.0000000
2	0.0558062	0.0441897
3	0.0291157	0.0209231
4	0.0232454	0.0000000
5	0.0291157	-0.0209231
6	0.0558062	-0.0441897

The displayed decimals are for reading, not a replacement for the exact counts. Horner evaluation of the coefficient record reconstructs the values at any desired arithmetic precision.

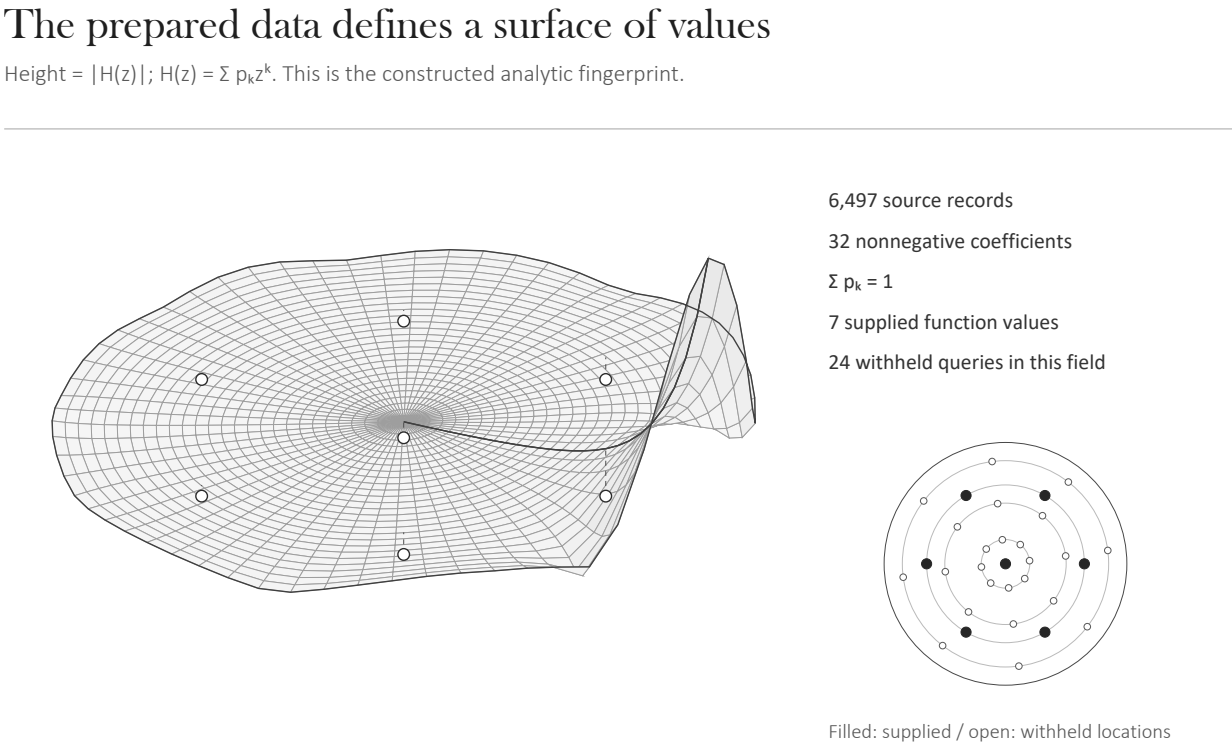


Figure H.7: Wine: Total sulfur dioxide. The histogram and all retained query radii are computed from the same coefficient record. Connecting segments organize the eight sampled angles; they do not assert a bound between those angles.

The complete withheld-query register

For each radius $r = 0.2, 0.5, 0.85$, the eight directions are $q_{r,k} = r\exp(i(0.13 + 2\pi k/8))$, $k = 0, \dots, 7$. The last column is $|H(q) - c(q)|/\rho(q)$, computed from the full-precision values. A ratio at most one records containment in the feasible disk. It is not a probability or a percentage prediction error.

r, k	$\operatorname{Re} H(q)$	$\operatorname{Im} H(q)$	$\rho(q)$	Ratio
0.2,0	0.0657014	0.0021965	0.0145485	0.15448
0.2,1	0.0581979	0.0114168	0.0146247	0.12930
0.2,2	0.0484616	0.0113896	0.0147405	0.09766
0.2,3	0.0424260	0.0060055	0.0147993	0.07910
0.2,4	0.0409107	-0.0012288	0.0147959	0.07455
0.2,5	0.0439250	-0.0081203	0.0147674	0.08364
0.2,6	0.0514016	-0.0122272	0.0147112	0.10699
0.2,7	0.0614750	-0.0094303	0.0146085	0.14030
0.5,0	0.1021252	0.0130764	0.0308927	0.27192
0.5,1	0.0619984	0.0333258	0.0317485	0.13688
0.5,2	0.0399527	0.0245495	0.0428678	0.06839
0.5,3	0.0310481	0.0114489	0.0430395	0.04412
0.5,4	0.0285986	-0.0023793	0.0323207	0.03761
0.5,5	0.0330408	-0.0157479	0.0323176	0.05006
0.5,6	0.0452299	-0.0287056	0.0427281	0.08342
0.5,7	0.0721863	-0.0327001	0.0418790	0.17855
0.85,0	0.2603661	0.1870437	0.2652603	0.60287
0.85,1	0.0720108	0.0610495	0.2810092	0.10545
0.85,2	0.0324020	0.0320021	0.3594090	0.03411
0.85,3	0.0260332	0.0144079	0.3599866	0.02344
0.85,4	0.0169760	-0.0017562	0.2835893	0.00381
0.85,5	0.0195007	-0.0171526	0.2837610	0.01727
0.85,6	0.0274034	-0.0373698	0.3584614	0.02523
0.85,7	0.0717162	-0.0321289	0.3525807	0.12461

The median radius is 0.0370999 and the largest is 0.3599866 at $r = 0.85$, $k = 3$. At that query the center is $0.0177525 + 0.0160274i$ and the truth-to-center gap is 0.0084376. All 24 retained values satisfy the containment check.

What this field contributes

At probe radius 0.2, the eight admissible radii range from 0.0145485 to 0.0147993. At radius 0.5 they range from 0.0308927 to 0.0430395; at radius 0.85, from 0.2652603 to 0.3599866. These intervals separate the radial change from the variation across the eight directions.

The largest truth-to-center ratio among this field's 24 queries is 0.602871. The shared model and probe arrangement make the radii and containment ratios comparable across fields. Physical units belong to the histogram convention.

Wine: Density

The numerical range of density is narrow. Rescaling the bin index spreads that range over 32 polynomial coefficients. The complex-output radius is consequently not a density error in grams per cubic centimeter.

The source supplies 6,497 records, with observed minimum 0.98711 and maximum 1.03898 in g/cm^3. There are 13 occupied bins. The largest count is 1285, in bin 5; its lower and upper edges are 0.995214688 and 0.996835625. The bin index measures position in this declared interval system.

The complete coefficient record

Use $p_k = n_k/6497$ and $H(z) = \sum_{k=0}^{31} p_k z^k$. The edges follow $b_k = 0.98711 + (0.0016209375)k$. Decimal edge parameters are displayed to nine significant digits; the saved source endpoints define the unrounded construction.

k	n_k	k	n_k
0	47	16	0
1	457	17	0
2	888	18	0
3	1029	19	0
4	1076	20	0
5	1285	21	0
6	1092	22	0
7	478	23	0
8	118	24	0
9	22	25	0
10	2	26	0
11	0	27	0
12	0	28	0
13	0	29	0
14	2	30	0
15	0	31	1

The exact count sum is 6497. Empty bins are retained as zero coefficients; shifting the later coefficients left would create a different fingerprint.

The values supplied to completion

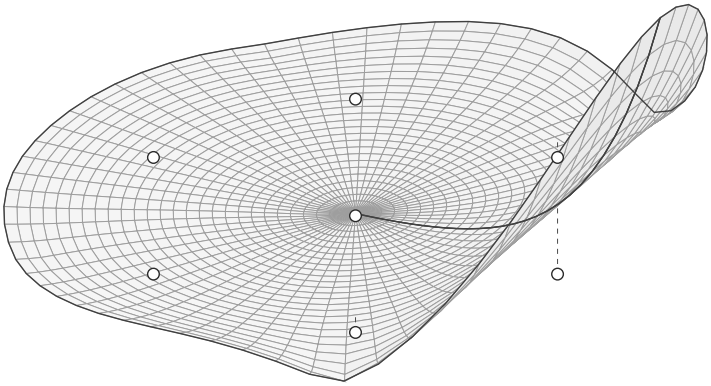
The observation nodes are $z_0 = 0$ and $z_{j+1} = 0.65e^{2\pi ij/6}$ for $j = 0, \ldots, 5$. The index in the following table identifies those nodes. These seven values are calculated after the entire histogram has been prepared.

Node	$\operatorname{Re} H(z_j)$	$\operatorname{Im} H(z_j)$
0	0.0072341	0.0000000
1	0.2236458	0.0000000
2	-0.0314637	0.0477508
3	-0.0166092	-0.0020607
4	-0.0080380	0.0000000
5	-0.0166092	0.0020607
6	-0.0314637	-0.0477508

The displayed decimals are for reading, not a replacement for the exact counts. Horner evaluation of the coefficient record reconstructs the values at any desired arithmetic precision.

The prepared data defines a surface of values

Height = |H(z)|; H(z) = Σ $p_k z^k$. This is the constructed analytic fingerprint.



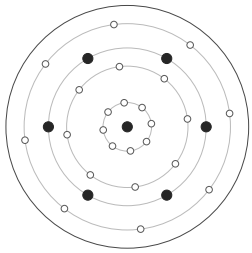
6,497 source records

32 nonnegative coefficients

Σ p_k = 1

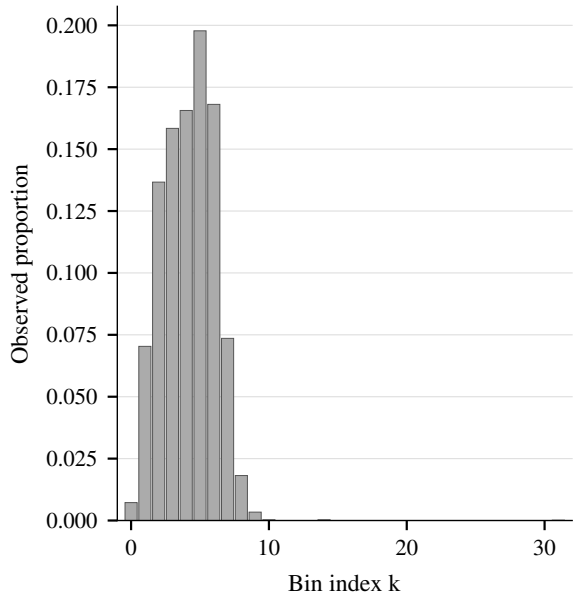
7 supplied function values

24 withheld queries in this field



Filled: supplied / open: withheld locations

Measured distribution



Completion by probe location

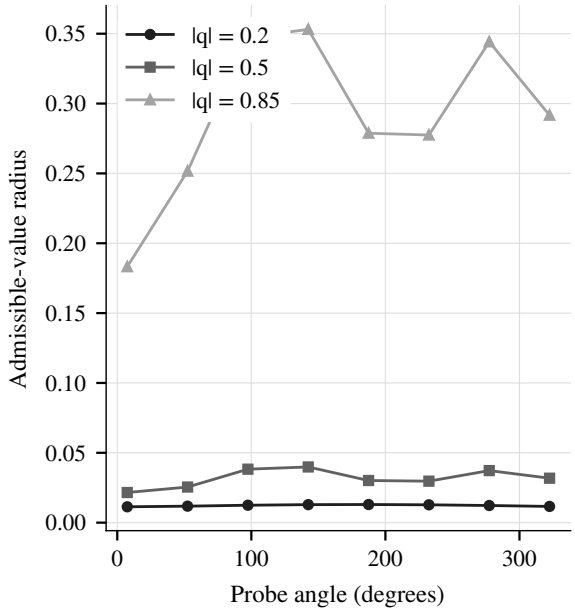


Figure H.8: Wine: Density. The histogram and all retained query radii are computed from the same coefficient record. Connecting segments organize the eight sampled angles; they do not assert a bound between those angles.

The complete withheld-query register

For each radius $r = 0.2, 0.5, 0.85$, the eight directions are $q_{r,k} = r\exp(i(0.13 + 2\pi k/8))$, $k = 0, \ldots, 7$. The last column is $|H(q) - c(q)|/\rho(q)$, computed from the full-precision values. A ratio at most one records containment in the feasible disk. It is not a probability or a percentage prediction error.

r, k	$\mathrm{Re}\, H(q)$	$\mathrm{Im}\, H(q)$	$\rho(q)$	Ratio
0.2,0	0.0279273	0.0038892	0.0113489	0.28025
0.2,1	0.0130035	0.0167228	0.0118054	0.23339
0.2,2	0.0007935	0.0115459	0.0124746	0.17099
0.2,3	-0.0022002	0.0043266	0.0128911	0.13471
0.2,4	-0.0024171	-0.0007997	0.0129790	0.12637
0.2,5	-0.0017910	-0.0064345	0.0127727	0.14335
0.2,6	0.0035526	-0.0141086	0.0122729	0.18962
0.2,7	0.0190045	-0.0151413	0.0116123	0.25426
0.5,0	0.1096010	0.0321224	0.0215709	0.48021
0.5,1	-0.0057516	0.0555342	0.0254976	0.26124
0.5,2	-0.0189607	0.0157053	0.0382777	0.11335
0.5,3	-0.0088431	0.0020620	0.0398692	0.06533
0.5,4	-0.0073190	-0.0004505	0.0301626	0.05968
0.5,5	-0.0115491	-0.0033897	0.0296960	0.07354
0.5,6	-0.0183103	-0.0265598	0.0372659	0.14782
0.5,7	0.0192929	-0.0745345	0.0318162	0.34348
0.85,0	0.4591586	0.2344633	0.1831946	0.83496
0.85,1	-0.0881780	0.0132711	0.2515896	0.18951
0.85,2	-0.0617237	0.0021837	0.3473817	0.07251
0.85,3	-0.0052302	0.0015805	0.3530884	0.00780
0.85,4	-0.0086210	0.0000993	0.2787837	0.01382
0.85,5	-0.0040593	0.0101328	0.2774632	0.00651
0.85,6	-0.0497574	-0.0478489	0.3441835	0.10137
0.85,7	-0.1636737	-0.1797372	0.2916472	0.39472

The median radius is 0.0309894 and the largest is 0.3530884 at $r = 0.85$, $k = 3$. At that query the center is $-0.0045584 + 0.0042526i$ and the truth-to-center gap is 0.0027552. All 24 retained values satisfy the containment check.

What this field contributes

At probe radius 0.2, the eight admissible radii range from 0.0113489 to 0.0129790. At radius 0.5 they range from 0.0215709 to 0.0398692; at radius 0.85, from 0.1831946 to 0.3530884. These intervals separate the radial change from the variation across the eight directions.

The largest truth-to-center ratio among this field's 24 queries is 0.834956. The shared model and probe arrangement make the radii and containment ratios comparable across fields. Physical units belong to the histogram convention.

Wine: pH

The source quantity is already a logarithmic chemical scale. The experiment bins the reported pH numbers directly; it does not transform them into hydrogen-ion concentrations before constructing the fingerprint.

The source supplies 6,497 records, with observed minimum 2.72 and maximum 4.01 in pH units. There are 31 occupied bins. The largest count is 693, in bin 10; its lower and upper edges are 3.123125 and 3.1634375. The bin index measures position in this declared interval system.

The complete coefficient record

Use $p_k = n_k/6497$ and $H(z) = \sum_{k=0}^{31} p_k z^k$. The edges follow $b_k = 2.72 + (0.0403125)k$. Decimal edge parameters are displayed to nine significant digits; the saved source endpoints define the unrounded construction.

k	n_k	k	n_k
0	3	16	356
1	7	17	244
2	6	18	169
3	42	19	126
4	90	20	96
5	170	21	58
6	228	22	27
7	321	23	24
8	459	24	18
9	546	25	8
10	693	26	7
11	665	27	2
12	625	28	1
13	541	29	2
14	527	30	0
15	434	31	2

The exact count sum is 6497. Empty bins are retained as zero coefficients; shifting the later coefficients left would create a different fingerprint.

The values supplied to completion

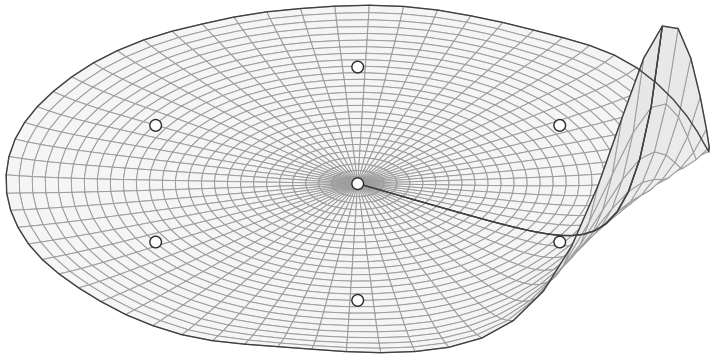
The observation nodes are $z_0 = 0$ and $z_{j+1} = 0.65e^{2\pi ij/6}$ for $j = 0, \ldots, 5$. The index in the following table identifies those nodes. These seven values are calculated after the entire histogram has been prepared.

Node	$\operatorname{Re} H(z_j)$	$\operatorname{Im} H(z_j)$
0	0.0004618	0.0000000
1	0.0214842	0.0000000
2	0.0003399	-0.0014261
3	0.0001902	0.0005237
4	-0.0005431	0.0000000
5	0.0001902	-0.0005237
6	0.0003399	0.0014261

The displayed decimals are for reading, not a replacement for the exact counts. Horner evaluation of the coefficient record reconstructs the values at any desired arithmetic precision.

The prepared data defines a surface of values

Height = |H(z)|; H(z) = Σ $p_k z^k$. This is the constructed analytic fingerprint.



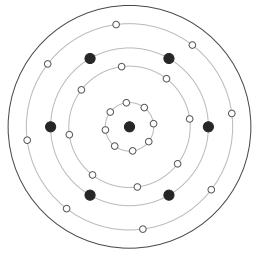
6,497 source records

32 nonnegative coefficients

Σ p_k = 1

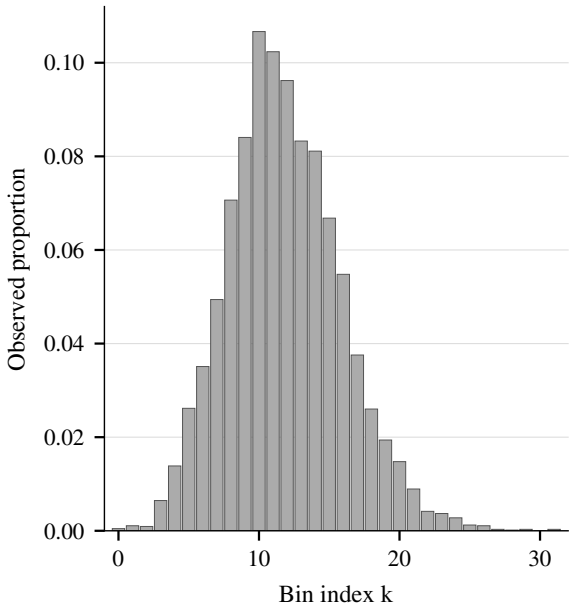
7 supplied function values

24 withheld queries in this field



Filled: supplied / open: withheld locations

Measured distribution



Completion by probe location

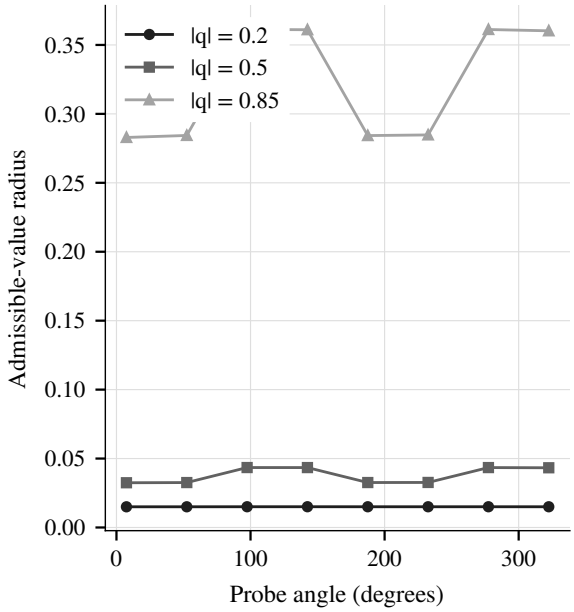


Figure H.9: Wine: pH. The histogram and all retained query radii are computed from the same coefficient record. Connecting segments organize the eight sampled angles; they do not assert a bound between those angles.

The complete withheld-query register

For each radius $r = 0.2, 0.5, 0.85$, the eight directions are $q_{r,k} = r\exp(i(0.13 + 2\pi k/8))$, $k = 0, \ldots, 7$. The last column is $|H(q) - c(q)|/\rho(q)$, computed from the full-precision values. A ratio at most one records containment in the feasible disk. It is not a probability or a percentage prediction error.

r, k	$\mathrm{Re}\, H(q)$	$\mathrm{Im}\, H(q)$	$\rho(q)$	Ratio
0.2,0	0.0007869	0.0000755	0.0150026	0.08543
0.2,1	0.0005178	0.0002058	0.0150115	0.06840
0.2,2	0.0004309	0.0001722	0.0150419	0.04884
0.2,3	0.0003080	0.0001323	0.0150497	0.03850
0.2,4	0.0002498	-0.0000309	0.0150338	0.03604
0.2,5	0.0003516	-0.0001593	0.0150299	0.04097
0.2,6	0.0004567	-0.0001720	0.0150380	0.05434
0.2,7	0.0005931	-0.0002222	0.0150253	0.07573
0.5,0	0.0042304	0.0026321	0.0324430	0.17110
0.5,1	-0.0001577	-0.0002588	0.0325830	0.06863
0.5,2	0.0006503	0.0003994	0.0434727	0.02956
0.5,3	0.0001870	0.0004887	0.0434523	0.01934
0.5,4	-0.0002462	-0.0001469	0.0326002	0.01552
0.5,5	0.0003179	-0.0004465	0.0326591	0.02203
0.5,6	0.0007759	-0.0001151	0.0434533	0.03604
0.5,7	-0.0009494	-0.0006432	0.0433081	0.10052
0.85,0	0.0543093	0.1505834	0.2828916	0.48493
0.85,1	-0.0035134	-0.0001409	0.2843542	0.02401
0.85,2	0.0014557	0.0008902	0.3612786	0.00955
0.85,3	0.0024964	0.0013534	0.3610352	0.00848
0.85,4	-0.0010771	0.0025155	0.2842579	0.00692
0.85,5	0.0000157	0.0013816	0.2847375	0.00904
0.85,6	0.0017057	-0.0030862	0.3611950	0.00557
0.85,7	0.0099482	0.0077483	0.3602464	0.07495

The median radius is 0.0379836 and the largest is 0.3612786 at $r = 0.85$, $k = 2$. At that query the center is $-0.0015140 + 0.0026436i$ and the truth-to-center gap is 0.0034488. All 24 retained values satisfy the containment check.

What this field contributes

At probe radius 0.2, the eight admissible radii range from 0.0150026 to 0.0150497. At radius 0.5 they range from 0.0324430 to 0.0434727; at radius 0.85, from 0.2828916 to 0.3612786. These intervals separate the radial change from the variation across the eight directions.

The largest truth-to-center ratio among this field's 24 queries is 0.484928. The shared model and probe arrangement make the radii and containment ratios comparable across fields. Physical units belong to the histogram convention.

Wine: Sulphates

The same preparation and completion rule is used here as for every other wine field. Differences in the resulting disks come from the supplied values, while their common broad pattern also reflects the shared probe geometry.

The source supplies 6,497 records, with observed minimum 0.22 and maximum 2 in g/dm^3. There are 24 occupied bins. The largest count is 1218, in bin 5; its lower and upper edges are 0.498125 and 0.55375. The bin index measures position in this declared interval system.

The complete coefficient record

Use $p_k = n_k/6497$ and $H(z) = \sum_{k=0}^{31} p_k z^k$. The edges follow $b_k = 0.22 + (0.055625)k$. Decimal edge parameters are displayed to nine significant digits; the saved source endpoints define the unrounded construction.

k	n_k	k	n_k
0	23	16	8
1	209	17	10
2	634	18	1
3	1055	19	4
4	1029	20	4
5	1218	21	0
6	778	22	0
7	575	23	0
8	368	24	3
9	211	25	1
10	172	26	0
11	78	27	0
12	52	28	0
13	32	29	0
14	14	30	0
15	14	31	4

The exact count sum is 6497. Empty bins are retained as zero coefficients; shifting the later coefficients left would create a different fingerprint.

The values supplied to completion

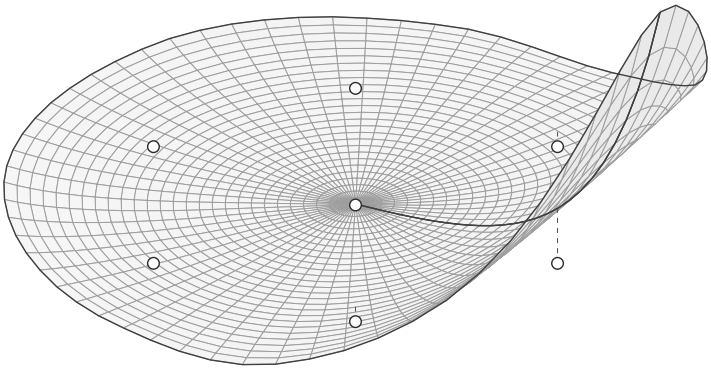
The observation nodes are $z_0 = 0$ and $z_{j+1} = 0.65e^{2\pi ij/6}$ for $j = 0, \ldots, 5$. The index in the following table identifies those nodes. These seven values are calculated after the entire histogram has been prepared.

Node	$\mathrm{Re}\, H(z_j)$	$\mathrm{Im}\, H(z_j)$
0	0.0035401	0.0000000
1	0.1766809	0.0000000
2	-0.0449254	0.0154308
3	-0.0015095	-0.0095272
4	-0.0081094	0.0000000
5	-0.0015095	0.0095272
6	-0.0449254	-0.0154308

The displayed decimals are for reading, not a replacement for the exact counts. Horner evaluation of the coefficient record reconstructs the values at any desired arithmetic precision.

The prepared data defines a surface of values

Height = |H(z)|; H(z) = Σ $p_k z^k$. This is the constructed analytic fingerprint.



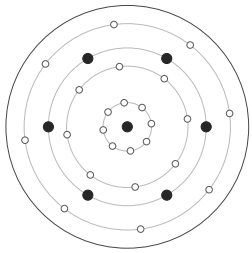
6,497 source records

32 nonnegative coefficients

Σ p_k = 1

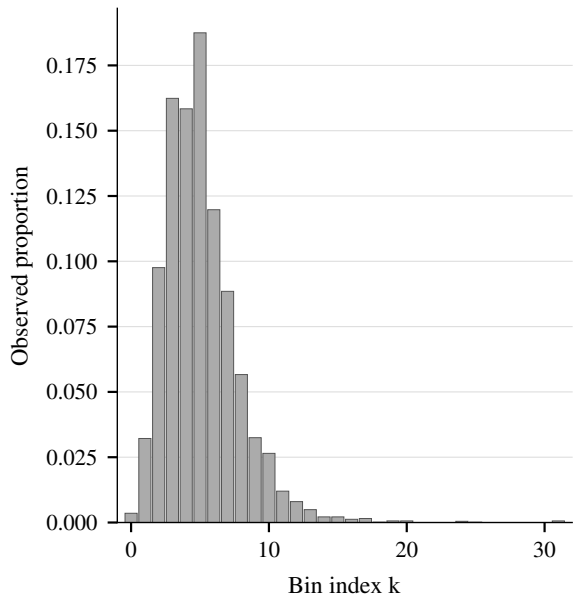
7 supplied function values

24 withheld queries in this field



Filled: supplied / open: withheld locations

Measured distribution



Completion by probe location

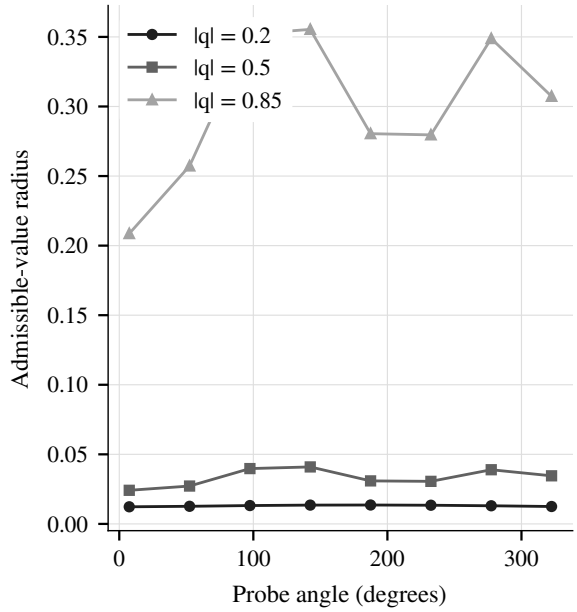


Figure H.10: Wine: Sulphates. The histogram and all retained query radii are computed from the same coefficient record. Connecting segments organize the eight sampled angles; they do not assert a bound between those angles.

The complete withheld-query register

For each radius $r = 0.2, 0.5, 0.85$, the eight directions are $q_{r,k} = r\exp(i(0.13 + 2\pi k/8))$, $k = 0, \ldots, 7$. The last column is $|H(q) - c(q)|/\rho(q)$, computed from the full-precision values. A ratio at most one records containment in the feasible disk. It is not a probability or a percentage prediction error.

r, k	$\mathrm{Re}\, H(q)$	$\mathrm{Im}\, H(q)$	$\rho(q)$	Ratio
0.2,0	0.0151671	0.0025000	0.0122970	0.23126
0.2,1	0.0050376	0.0091827	0.0126574	0.19024
0.2,2	-0.0003931	0.0043422	0.0131912	0.14004
0.2,3	-0.0002224	0.0012183	0.0135172	0.11438
0.2,4	-0.0000917	-0.0002302	0.0135812	0.10907
0.2,5	-0.0003932	-0.0019009	0.0134210	0.12015
0.2,6	0.0003581	-0.0061078	0.0130317	0.15436
0.2,7	0.0088591	-0.0090033	0.0125087	0.20823
0.5,0	0.0769914	0.0266822	0.0241539	0.41101
0.5,1	-0.0189804	0.0324291	0.0272128	0.19858
0.5,2	-0.0100913	-0.0010630	0.0397626	0.08435
0.5,3	0.0009377	0.0000568	0.0409564	0.06291
0.5,4	-0.0026805	-0.0012344	0.0309163	0.06383
0.5,5	-0.0008575	0.0026171	0.0305851	0.06438
0.5,6	-0.0152497	-0.0039396	0.0389337	0.10623
0.5,7	-0.0008531	-0.0540215	0.0345626	0.27544
0.85,0	0.3905855	0.2347230	0.2086363	0.77697
0.85,1	-0.1140411	-0.0046209	0.2572524	0.11448
0.85,2	-0.0175248	-0.0132086	0.3520795	0.02054
0.85,3	0.0257369	0.0020836	0.3554014	0.04376
0.85,4	-0.0213577	-0.0132207	0.2804509	0.04498
0.85,5	0.0108076	0.0218822	0.2795986	0.02742
0.85,6	-0.0195478	0.0141913	0.3488120	0.02681
0.85,7	-0.1642639	-0.1347051	0.3072903	0.30070

The median radius is 0.0327394 and the largest is 0.3554014 at $r = 0.85$, $k = 3$. At that query the center is $0.0106649 + 0.0059226i$ and the truth-to-center gap is 0.0155532. All 24 retained values satisfy the containment check.

What this field contributes

At probe radius 0.2, the eight admissible radii range from 0.0122970 to 0.0135812. At radius 0.5 they range from 0.0241539 to 0.0409564; at radius 0.85, from 0.2086363 to 0.3554014. These intervals separate the radial change from the variation across the eight directions.

The largest truth-to-center ratio among this field's 24 queries is 0.776974. The shared model and probe arrangement make the radii and containment ratios comparable across fields. Physical units belong to the histogram convention.

Wine: Alcohol

This is the field developed in the main worked example. The complete register below extends the two displayed query directions to all 24 withheld probes in the retained experiment.

The source supplies 6,497 records, with observed minimum 8 and maximum 14.9 in percent by volume. There are 30 occupied bins. The largest count is 892, in bin 6; its lower and upper edges are 9.29375 and 9.509375. The bin index measures position in this declared interval system.

The complete coefficient record

Use $p_k = n_k/6497$ and $H(z) = \sum_{k=0}^{31} p_k z^k$. The edges follow $b_k = 8 + (0.215625)k$. Decimal edge parameters are displayed to nine significant digits; the saved source endpoints define the unrounded construction.

k	n_k	k	n_k
0	2	16	188
1	5	17	174
2	33	18	202
3	189	19	239
4	311	20	186
5	440	21	134
6	892	22	124
7	355	23	64
8	375	24	25
9	399	25	51
10	294	26	9
11	429	27	15
12	444	28	2
13	360	29	0
14	262	30	0
15	293	31	1

The exact count sum is 6497. Empty bins are retained as zero coefficients; shifting the later coefficients left would create a different fingerprint.

The values supplied to completion

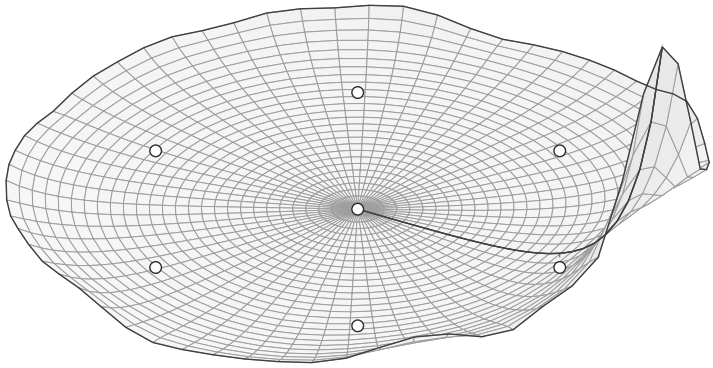
The observation nodes are $z_0 = 0$ and $z_{j+1} = 0.65e^{2\pi ij/6}$ for $j = 0, \ldots, 5$. The index in the following table identifies those nodes. These seven values are calculated after the entire histogram has been prepared.

Node	$\operatorname{Re} H(z_j)$	$\operatorname{Im} H(z_j)$
0	0.0003078	0.0000000
1	0.0455185	0.0000000
2	0.0010187	-0.0087965
3	0.0078383	0.0000327
4	0.0031549	-0.0000000
5	0.0078383	-0.0000327
6	0.0010187	0.0087965

The displayed decimals are for reading, not a replacement for the exact counts. Horner evaluation of the coefficient record reconstructs the values at any desired arithmetic precision.

The prepared data defines a surface of values

Height = $|H(z)|$; $H(z) = \Sigma\, p_k z^k$. This is the constructed analytic fingerprint.



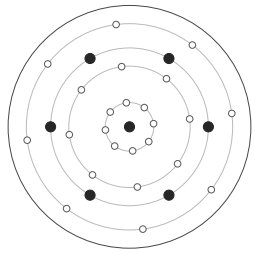
6,497 source records

32 nonnegative coefficients

$\Sigma\, p_k = 1$

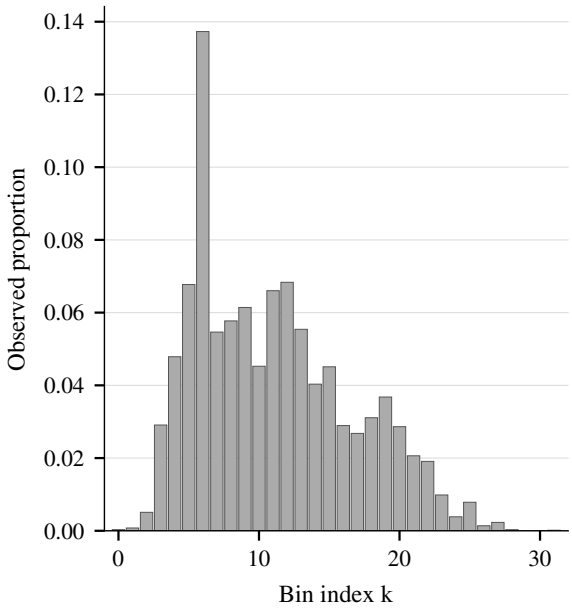
7 supplied function values

24 withheld queries in this field



Filled: supplied / open: withheld locations

Measured distribution



Completion by probe location

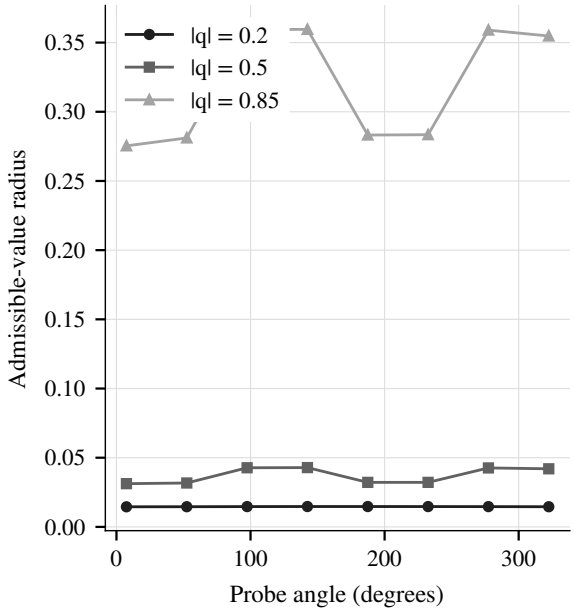


Figure H.11: Wine: Alcohol. The histogram and all retained query radii are computed from the same coefficient record. Connecting segments organize the eight sampled angles; they do not assert a bound between those angles.

The complete withheld-query register

For each radius $r = 0.2, 0.5, 0.85$, the eight directions are $q_{r,k} = r\exp(i(0.13 + 2\pi k/8))$, $k = 0, \dots, 7$. The last column is $|H(q) - c(q)|/\rho(q)$, computed from the full-precision values. A ratio at most one records containment in the feasible disk. It is not a probability or a percentage prediction error.

r, k	$\mathrm{Re}\, H(q)$	$\mathrm{Im}\, H(q)$	$\rho(q)$	Ratio
0.2,0	0.0009625	0.0002187	0.0145311	0.11208
0.2,1	0.0000722	0.0003425	0.0145844	0.08825
0.2,2	0.0002277	-0.0000660	0.0146779	0.05577
0.2,3	0.0002766	0.0000769	0.0147258	0.03372
0.2,4	0.0001914	-0.0000255	0.0147201	0.02730
0.2,5	0.0003186	-0.0000381	0.0146965	0.03954
0.2,6	0.0001158	0.0000256	0.0146545	0.06578
0.2,7	0.0002984	-0.0005330	0.0145776	0.09882
0.5,0	0.0115230	0.0067790	0.0312438	0.22101
0.5,1	-0.0040812	-0.0018434	0.0317306	0.10840
0.5,2	0.0005596	-0.0016991	0.0427370	0.04408
0.5,3	-0.0001963	0.0018435	0.0428600	0.02563
0.5,4	0.0000401	0.0003775	0.0321812	0.01467
0.5,5	0.0020157	-0.0013446	0.0321893	0.03033
0.5,6	-0.0000499	0.0013349	0.0426216	0.05886
0.5,7	-0.0064371	-0.0038892	0.0419911	0.14477
0.85,0	0.1213582	0.1653725	0.2753669	0.49678
0.85,1	0.0050115	-0.0530645	0.2810867	0.11971
0.85,2	-0.0153886	-0.0113906	0.3594611	0.04708
0.85,3	-0.0191815	0.0247146	0.3595308	0.06575
0.85,4	0.0146536	0.0135275	0.2831873	0.04351
0.85,5	0.0273249	-0.0225176	0.2834254	0.07339
0.85,6	-0.0087019	-0.0149181	0.3589960	0.05544
0.85,7	-0.0679210	0.0218078	0.3547572	0.14179

The median radius is 0.0370902 and the largest is 0.3595308 at $r = 0.85$, $k = 3$. At that query the center is $-0.0005223 + 0.0102038i$ and the truth-to-center gap is 0.0236374. All 24 retained values satisfy the containment check.

What this field contributes

At probe radius 0.2, the eight admissible radii range from 0.0145311 to 0.0147258. At radius 0.5 they range from 0.0312438 to 0.0428600; at radius 0.85, from 0.2753669 to 0.3595308. These intervals separate the radial change from the variation across the eight directions.

The largest truth-to-center ratio among this field's 24 queries is 0.496782. The shared model and probe arrangement make the radii and containment ratios comparable across fields. Physical units belong to the histogram convention.

Wine: Quality score

Quality is an observed ordered score in this study. Its histogram contains empty intervals because a discrete score is placed into 32 equal-width bins. The calculation does not predict quality from the other wine fields.

The source supplies 6,497 records, with observed minimum 3 and maximum 9 in ordered score. There are 7 occupied bins. The largest count is 2836, in bin 16; its lower and upper edges are 6 and 6.1875. The bin index measures position in this declared interval system.

The complete coefficient record

Use $p_k = n_k/6497$ and $H(z) = \sum_{k=0}^{31} p_k z^k$. The edges follow $b_k = 3 + (0.1875)k$. Decimal edge parameters are displayed to nine significant digits; the saved source endpoints define the unrounded construction.

k	n_k	k	n_k
0	30	16	2836
1	0	17	0
2	0	18	0
3	0	19	0
4	0	20	0
5	216	21	1079
6	0	22	0
7	0	23	0
8	0	24	0
9	0	25	0
10	2138	26	193
11	0	27	0
12	0	28	0
13	0	29	0
14	0	30	0
15	0	31	5

The exact count sum is 6497. Empty bins are retained as zero coefficients; shifting the later coefficients left would create a different fingerprint.

The values supplied to completion

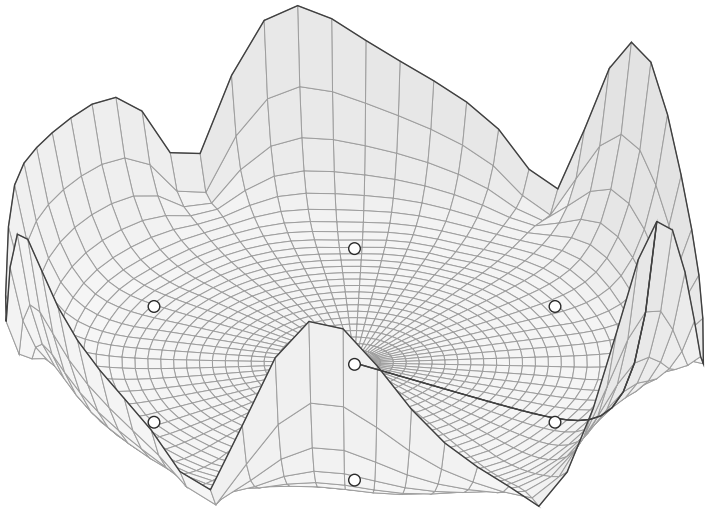
The observation nodes are $z_0 = 0$ and $z_{j+1} = 0.65e^{2\pi ij/6}$ for $j = 0, \ldots, 5$. The index in the following table identifies those nodes. These seven values are calculated after the entire histogram has been prepared.

Node	$\mathrm{Re}\, H(z_j)$	$\mathrm{Im}\, H(z_j)$
0	0.0046175	0.0000000
1	0.0133685	0.0000000
2	0.0040898	-0.0075609
3	0.0002714	0.0008795
4	0.0056143	-0.0000000
5	0.0002714	-0.0008795
6	0.0040898	0.0075609

The displayed decimals are for reading, not a replacement for the exact counts. Horner evaluation of the coefficient record reconstructs the values at any desired arithmetic precision.

The prepared data defines a surface of values

Height = |H(z)|; H(z) = Σ $p_k z^k$. This is the constructed analytic fingerprint.



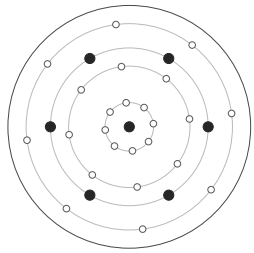
6,497 source records

32 nonnegative coefficients

Σ p_k = 1

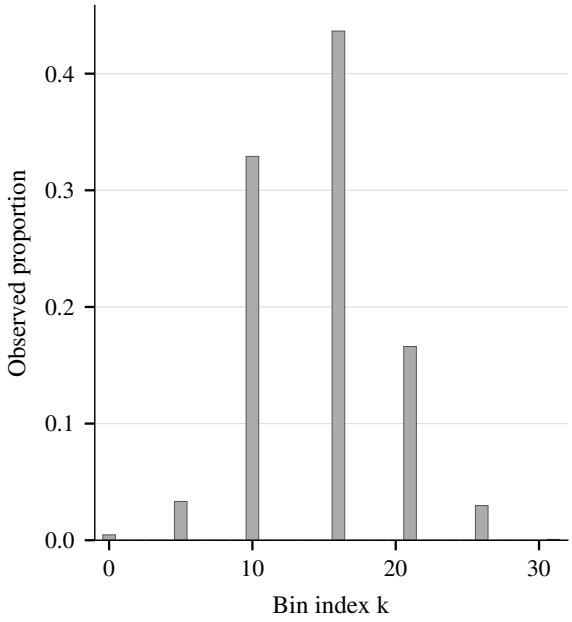
7 supplied function values

24 withheld queries in this field



Filled: supplied / open: withheld locations

Measured distribution



Completion by probe location

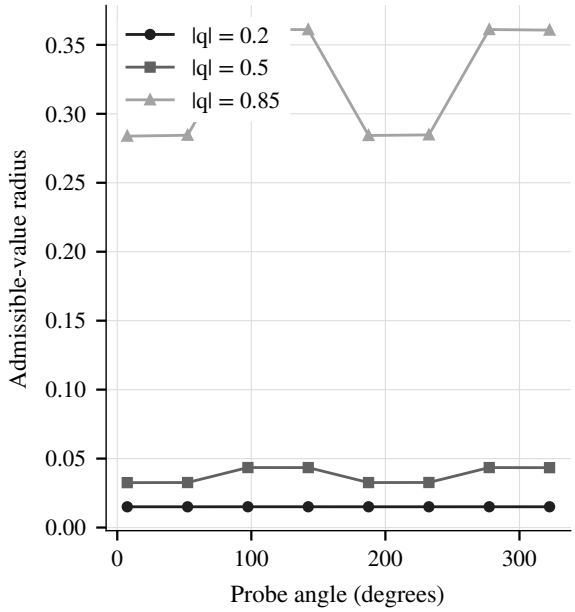


Figure H.12: Wine: Quality score. The histogram and all retained query radii are computed from the same coefficient record. Connecting segments organize the eight sampled angles; they do not assert a bound between those angles.

The complete withheld-query register

For each radius $r = 0.2, 0.5, 0.85$, the eight directions are $q_{r,k} = r\exp(i(0.13 + 2\pi k/8))$, $k = 0, \ldots, 7$. The last column is $|H(q) - c(q)|/\rho(q)$, computed from the full-precision values. A ratio at most one records containment in the feasible disk. It is not a probability or a percentage prediction error.

r, k	$\mathrm{Re}\, H(q)$	$\mathrm{Im}\, H(q)$	$\rho(q)$	Ratio
0.2,0	0.0046260	0.0000065	0.0150413	0.00418
0.2,1	0.0046160	-0.0000105	0.0150435	0.00325
0.2,2	0.0046111	0.0000084	0.0150655	0.00208
0.2,3	0.0046281	-0.0000014	0.0150690	0.00407
0.2,4	0.0046091	-0.0000064	0.0150521	0.00202
0.2,5	0.0046189	0.0000106	0.0150501	0.00345
0.2,6	0.0046239	-0.0000085	0.0150640	0.00371
0.2,7	0.0046070	0.0000014	0.0150603	0.00154
0.5,0	0.0055272	0.0009443	0.0325621	0.05510
0.5,1	0.0041645	-0.0009376	0.0326425	0.04803
0.5,2	0.0038995	0.0005232	0.0435015	0.03725
0.5,3	0.0059533	-0.0002203	0.0434752	0.05045
0.5,4	0.0038732	-0.0003133	0.0326179	0.04131
0.5,5	0.0044448	0.0011212	0.0326764	0.04956
0.5,6	0.0051571	-0.0011308	0.0434913	0.04779
0.5,7	0.0038945	0.0000600	0.0434176	0.03428
0.85,0	0.0124517	0.1017348	0.2838049	0.32894
0.85,1	-0.0704079	0.0325887	0.2844660	0.28327
0.85,2	-0.0392110	-0.0272916	0.3612124	0.16111
0.85,3	0.0637585	0.0144984	0.3610441	0.17775
0.85,4	-0.0009998	0.0795098	0.2842835	0.28036
0.85,5	-0.0766028	0.0578267	0.2847302	0.32084
0.85,6	-0.0169705	-0.0407556	0.3611108	0.14361
0.85,7	0.0385226	0.0082836	0.3606878	0.13494

The median radius is 0.0380470 and the largest is 0.3612124 at $r = 0.85$, $k = 2$. At that query the center is $0.0065553 + 0.0086566i$ and the truth-to-center gap is 0.0581964. All 24 retained values satisfy the containment check.

What this field contributes

At probe radius 0.2, the eight admissible radii range from 0.0150413 to 0.0150690. At radius 0.5 they range from 0.0325621 to 0.0435015; at radius 0.85, from 0.2838049 to 0.3612124. These intervals separate the radial change from the variation across the eight directions.

The largest truth-to-center ratio among this field's 24 queries is 0.328940. The shared model and probe arrangement make the radii and containment ratios comparable across fields. Physical units belong to the histogram convention.

Power plant: Air temperature

The field contains ambient air temperatures recorded with the power-plant observations. Rows are treated as a distribution sample; their file order is not interpreted as a dynamical temperature trajectory.

The source supplies 9,568 records, with observed minimum 1.81 and maximum 37.11 in °C. There are 32 occupied bins. The largest count is 548, in bin 21; its lower and upper edges are 24.975625 and 26.07875. The bin index measures position in this declared interval system.

The complete coefficient record

Use $p_k = n_k/9568$ and $H(z) = \sum_{k=0}^{31} p_k z^k$. The edges follow $b_k = 1.81 + (1.103125)k$. Decimal edge parameters are displayed to nine significant digits; the saved source endpoints define the unrounded construction.

k	n_k	k	n_k
0	6	16	388
1	29	17	400
2	61	18	475
3	161	19	536
4	199	20	526
5	253	21	548
6	318	22	499
7	400	23	414
8	327	24	409
9	398	25	318
10	455	26	277
11	479	27	181
12	363	28	85
13	314	29	38
14	346	30	8
15	356	31	1

The exact count sum is 9568. Empty bins are retained as zero coefficients; shifting the later coefficients left would create a different fingerprint.

The values supplied to completion

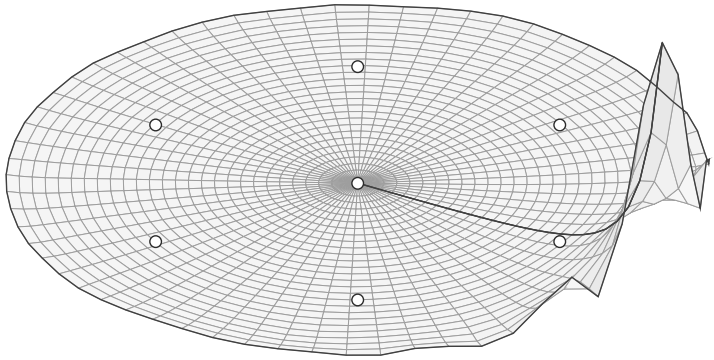
The observation nodes are $z_0 = 0$ and $z_{j+1} = 0.65e^{2\pi ij/6}$ for $j = 0, \ldots, 5$. The index in the following table identifies those nodes. These seven values are calculated after the entire histogram has been prepared.

Node	$\mathrm{Re}\, H(z_j)$	$\mathrm{Im}\, H(z_j)$
0	0.0006271	0.0000000
1	0.0248910	0.0000000
2	-0.0024679	0.0000893
3	0.0009343	0.0009868
4	-0.0015906	0.0000000
5	0.0009343	-0.0009868
6	-0.0024679	-0.0000893

The displayed decimals are for reading, not a replacement for the exact counts. Horner evaluation of the coefficient record reconstructs the values at any desired arithmetic precision.

The prepared data defines a surface of values

Height = |H(z)|; H(z) = Σ $p_k z^k$. This is the constructed analytic fingerprint.



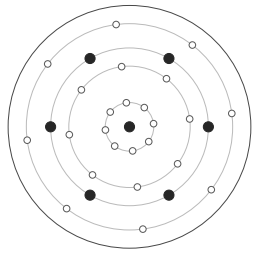
9,568 source records

32 nonnegative coefficients

Σ p_k = 1

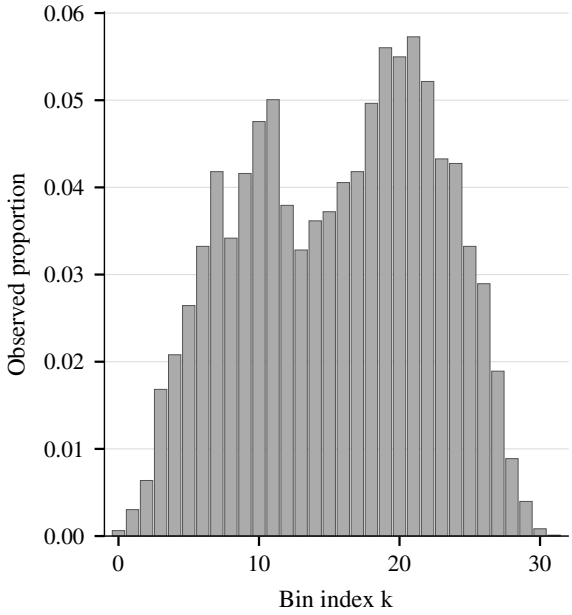
7 supplied function values

24 withheld queries in this field



Filled: supplied / open: withheld locations

Measured distribution



Completion by probe location

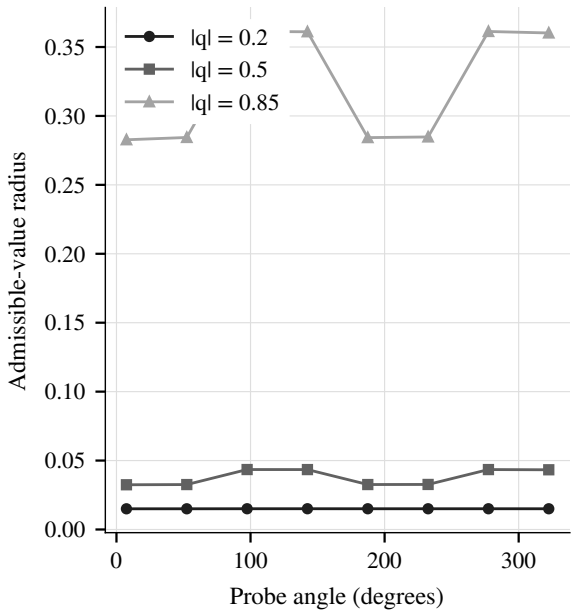


Figure H.13: Power plant: Air temperature. The histogram and all retained query radii are computed from the same coefficient record. Connecting segments organize the eight sampled angles; they do not assert a bound between those angles.

The complete withheld-query register

For each radius $r = 0.2, 0.5, 0.85$, the eight directions are $q_{r,k} = r\exp(i(0.13 + 2\pi k/8))$, $k = 0, \ldots, 7$. The last column is $|H(q) - c(q)|/\rho(q)$, computed from the full-precision values. A ratio at most one records containment in the feasible disk. It is not a probability or a percentage prediction error.

r, k	$\mathrm{Re}\, H(q)$	$\mathrm{Im}\, H(q)$	$\rho(q)$	Ratio
0.2,0	0.0016366	0.0002190	0.0150067	0.06256
0.2,1	0.0007788	0.0007526	0.0150160	0.05202
0.2,2	0.0003759	0.0004325	0.0150466	0.04003
0.2,3	0.0002420	0.0002306	0.0150540	0.03376
0.2,4	0.0001713	-0.0000517	0.0150379	0.03235
0.2,5	0.0002896	-0.0002957	0.0150343	0.03524
0.2,6	0.0004402	-0.0005334	0.0150428	0.04340
0.2,7	0.0010826	-0.0007534	0.0150297	0.05653
0.5,0	0.0080840	0.0034513	0.0324407	0.11325
0.5,1	-0.0013726	0.0019370	0.0325909	0.05023
0.5,2	0.0002086	-0.0000084	0.0434851	0.02820
0.5,3	0.0000349	0.0006971	0.0434602	0.02101
0.5,4	-0.0005966	-0.0002656	0.0326057	0.02045
0.5,5	0.0003306	-0.0006482	0.0326658	0.02249
0.5,6	-0.0004060	-0.0000938	0.0434670	0.03209
0.5,7	-0.0007279	-0.0041440	0.0433137	0.06932
0.85,0	0.0476594	0.0978792	0.2827053	0.29113
0.85,1	-0.0072346	-0.0037068	0.2843717	0.01120
0.85,2	0.0037505	-0.0055508	0.3613258	0.02076
0.85,3	-0.0011273	0.0007708	0.3610481	0.00756
0.85,4	-0.0039396	-0.0031578	0.2842700	0.01372
0.85,5	-0.0002917	-0.0024137	0.2847512	0.00961
0.85,6	-0.0051473	0.0009230	0.3612599	0.01961
0.85,7	-0.0095982	0.0006884	0.3601866	0.04250

The median radius is 0.0379898 and the largest is 0.3613258 at $r = 0.85$, $k = 2$. At that query the center is $-0.0003549 + 0.0007283i$ and the truth-to-center gap is 0.0075021. All 24 retained values satisfy the containment check.

What this field contributes

At probe radius 0.2, the eight admissible radii range from 0.0150067 to 0.0150540. At radius 0.5 they range from 0.0324407 to 0.0434851; at radius 0.85, from 0.2827053 to 0.3613258. These intervals separate the radial change from the variation across the eight directions.

The largest truth-to-center ratio among this field's 24 queries is 0.291125. The shared model and probe arrangement make the radii and containment ratios comparable across fields. Physical units belong to the histogram convention.

Power plant: Exhaust vacuum

Exhaust vacuum is retained as a measured coordinate. The analytic fingerprint is constructed from its marginal histogram, so its relation to electrical output is not part of this univariate completion problem.

The source supplies 9,568 records, with observed minimum 25.36 and maximum 81.56 in cm Hg. There are 29 occupied bins. The largest count is 1184, in bin 8; its lower and upper edges are 39.41 and 41.16625. The bin index measures position in this declared interval system.

The complete coefficient record

Use $p_k = n_k/9568$ and $H(z) = \sum_{k=0}^{31} p_k z^k$. The edges follow $b_k = 25.36 + (1.75625)k$. Decimal edge parameters are displayed to nine significant digits; the saved source endpoints define the unrounded construction.

k	n_k	k	n_k
0	13	16	95
1	0	17	134
2	0	18	311
3	0	19	548
4	13	20	245
5	83	21	387
6	157	22	374
7	478	23	364
8	1184	24	516
9	898	25	565
10	690	26	412
11	466	27	340
12	184	28	208
13	312	29	97
14	256	30	38
15	193	31	7

The exact count sum is 9568. Empty bins are retained as zero coefficients; shifting the later coefficients left would create a different fingerprint.

The values supplied to completion

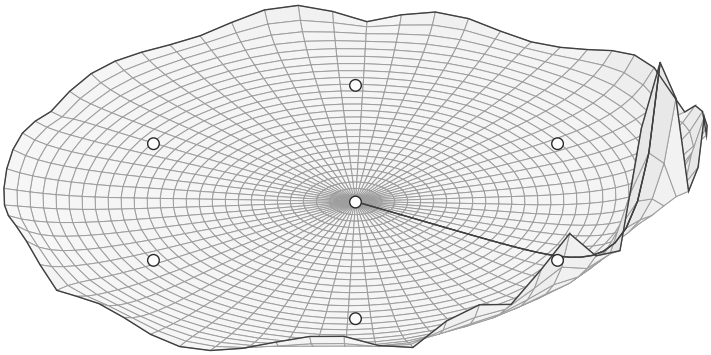
The observation nodes are $z_0 = 0$ and $z_{j+1} = 0.65e^{2\pi ij/6}$ for $j = 0, \ldots, 5$. The index in the following table identifies those nodes. These seven values are calculated after the entire histogram has been prepared.

Node	$\operatorname{Re} H(z_j)$	$\operatorname{Im} H(z_j)$
0	0.0013587	0.0000000
1	0.0139718	0.0000000
2	0.0001366	0.0034026
3	0.0000668	-0.0014229
4	0.0019493	-0.0000000
5	0.0000668	0.0014229
6	0.0001366	-0.0034026

The displayed decimals are for reading, not a replacement for the exact counts. Horner evaluation of the coefficient record reconstructs the values at any desired arithmetic precision.

The prepared data defines a surface of values

Height = |H(z)|; H(z) = Σ $p_k z^k$. This is the constructed analytic fingerprint.



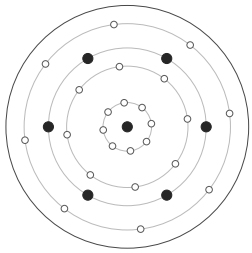
9,568 source records

32 nonnegative coefficients

Σ p_k = 1

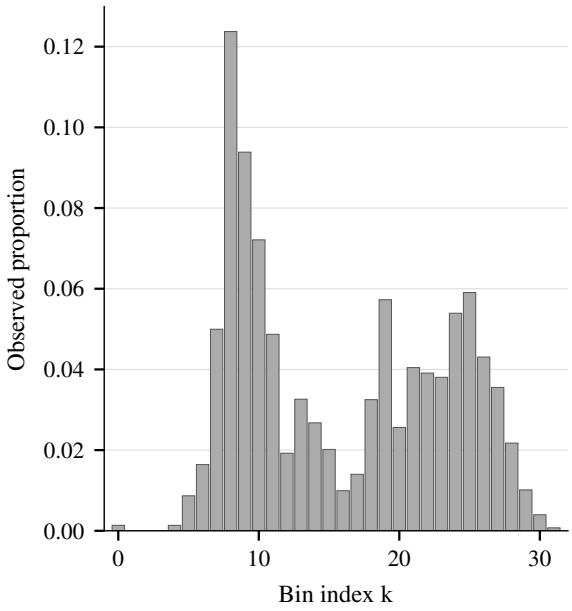
7 supplied function values

24 withheld queries in this field



Filled: supplied / open: withheld locations

Measured distribution



Completion by probe location

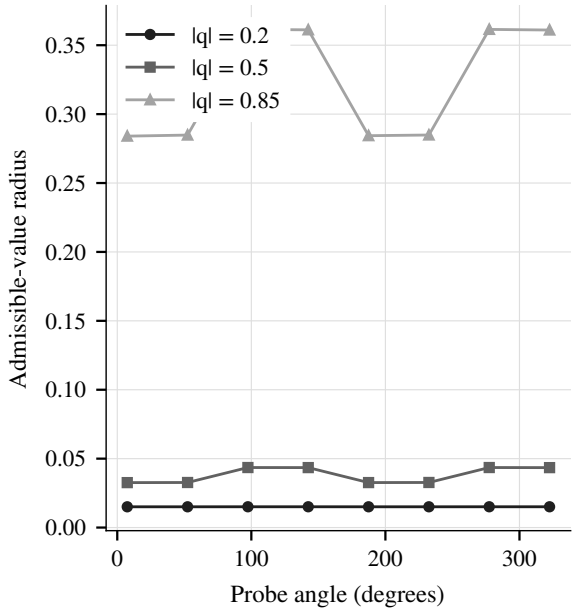


Figure H.14: Power plant: Exhaust vacuum. The histogram and all retained query radii are computed from the same coefficient record. Connecting segments organize the eight sampled angles; they do not assert a bound between those angles.

The complete withheld-query register

For each radius $r = 0.2, 0.5, 0.85$, the eight directions are $q_{r,k} = r\exp(i(0.13 + 2\pi k/8))$, $k = 0, \ldots, 7$. The last column is $|H(q) - c(q)|/\rho(q)$, computed from the full-precision values. A ratio at most one records containment in the feasible disk. It is not a probability or a percentage prediction error.

r, k	$\mathrm{Re}\, H(q)$	$\mathrm{Im}\, H(q)$	$\rho(q)$	Ratio
0.2,0	0.0013641	0.0000043	0.0150614	0.08585
0.2,1	0.0013579	-0.0000042	0.0150632	0.07120
0.2,2	0.0013588	0.0000024	0.0150834	0.04921
0.2,3	0.0013590	-0.0000011	0.0150847	0.03277
0.2,4	0.0013589	-0.0000001	0.0150671	0.02784
0.2,5	0.0013575	0.0000011	0.0150665	0.03723
0.2,6	0.0013612	-0.0000012	0.0150827	0.05625
0.2,7	0.0013535	0.0000010	0.0150803	0.07785
0.5,0	0.0024070	0.0013830	0.0326084	0.16406
0.5,1	0.0019097	0.0001405	0.0326937	0.10040
0.5,2	0.0014719	0.0002583	0.0435553	0.04763
0.5,3	0.0015514	0.0000624	0.0435096	0.01963
0.5,4	0.0013483	0.0000445	0.0326394	0.00355
0.5,5	0.0013774	0.0002734	0.0327061	0.02831
0.5,6	0.0014810	0.0001699	0.0435525	0.06122
0.5,7	0.0012798	0.0010041	0.0434839	0.12413
0.85,0	0.0364031	0.1066101	0.2840168	0.35502
0.85,1	0.0143010	0.0360640	0.2847972	0.12754
0.85,2	0.0085609	0.0124303	0.3614777	0.04761
0.85,3	0.0077228	0.0108683	0.3611613	0.03687
0.85,4	0.0059810	0.0040330	0.2843431	0.01711
0.85,5	0.0055009	0.0151021	0.2848449	0.04737
0.85,6	0.0153384	0.0098734	0.3614715	0.05992
0.85,7	0.0420154	0.0429960	0.3610175	0.18602

The median radius is 0.0380950 and the largest is 0.3614777 at $r = 0.85$, $k = 2$. At that query the center is $-0.0034014 + 0.0000567i$ and the truth-to-center gap is 0.0172106. All 24 retained values satisfy the containment check.

What this field contributes

At probe radius 0.2, the eight admissible radii range from 0.0150614 to 0.0150847. At radius 0.5 they range from 0.0326084 to 0.0435553; at radius 0.85, from 0.2840168 to 0.3614777. These intervals separate the radial change from the variation across the eight directions.

The largest truth-to-center ratio among this field's 24 queries is 0.355023. The shared model and probe arrangement make the radii and containment ratios comparable across fields. Physical units belong to the histogram convention.

Power plant: Ambient pressure

Ambient pressure is represented in the source unit and then binned over its observed range. The bins establish the physical meaning of each coefficient; an independent change of the range would change that meaning.

The source supplies 9,568 records, with observed minimum 992.89 and maximum 1033.3 in millibars. There are 32 occupied bins. The largest count is 836, in bin 16; its lower and upper edges are 1013.095 and 1014.35781. The bin index measures position in this declared interval system.

The complete coefficient record

Use $p_k = n_k/9568$ and $H(z) = \sum_{k=0}^{31} p_k z^k$. The edges follow $b_k = 992.89 + (1.2628125)k$. Decimal edge parameters are displayed to nine significant digits; the saved source endpoints define the unrounded construction.

k	n_k	k	n_k
0	5	16	836
1	4	17	653
2	6	18	615
3	8	19	639
4	16	20	428
5	46	21	488
6	89	22	291
7	144	23	192
8	202	24	159
9	276	25	130
10	431	26	86
11	613	27	54
12	667	28	24
13	754	29	25
14	834	30	28
15	805	31	20

The exact count sum is 9568. Empty bins are retained as zero coefficients; shifting the later coefficients left would create a different fingerprint.

The values supplied to completion

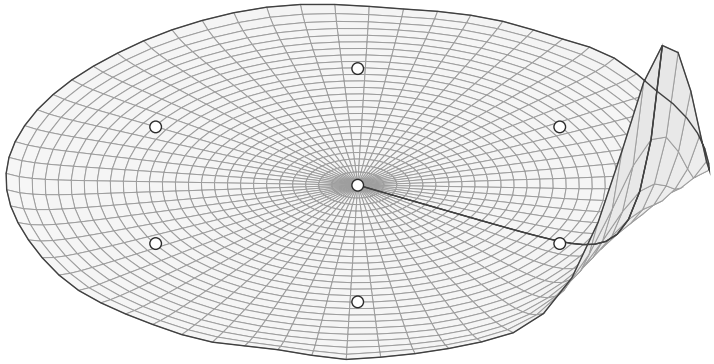
The observation nodes are $z_0 = 0$ and $z_{j+1} = 0.65e^{2\pi ij/6}$ for $j = 0, \ldots, 5$. The index in the following table identifies those nodes. These seven values are calculated after the entire histogram has been prepared.

Node	$\operatorname{Re} H(z_j)$	$\operatorname{Im} H(z_j)$
0	0.0005226	0.0000000
1	0.0072490	0.0000000
2	0.0008507	0.0002725
3	0.0002955	-0.0000035
4	0.0003512	0.0000000
5	0.0002955	0.0000035
6	0.0008507	-0.0002725

The displayed decimals are for reading, not a replacement for the exact counts. Horner evaluation of the coefficient record reconstructs the values at any desired arithmetic precision.

The prepared data defines a surface of values

Height = $|H(z)|$; $H(z) = \Sigma\, p_k z^k$. This is the constructed analytic fingerprint.



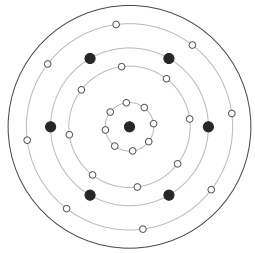
9,568 source records

32 nonnegative coefficients

$\Sigma\, p_k = 1$

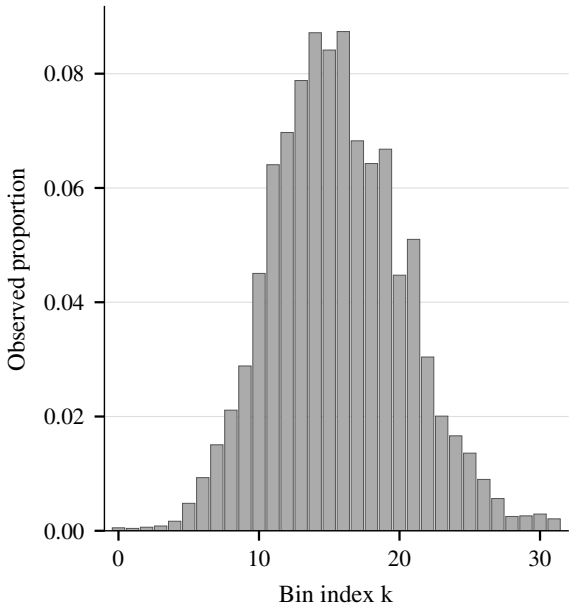
7 supplied function values

24 withheld queries in this field



Filled: supplied / open: withheld locations

Measured distribution



Completion by probe location

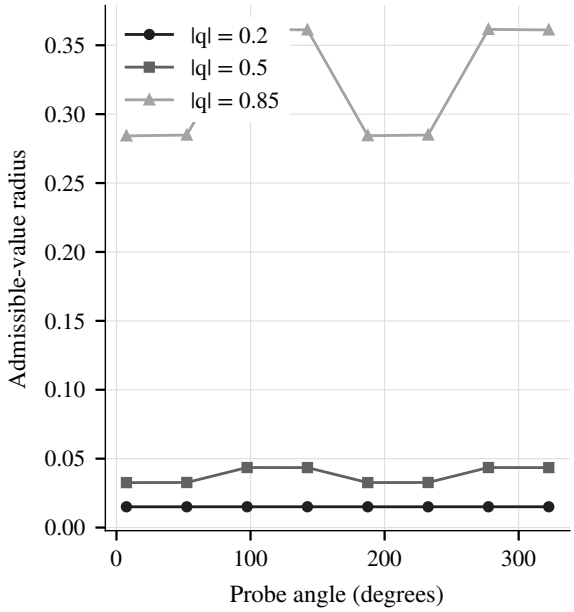


Figure H.15: Power plant: Ambient pressure. The histogram and all retained query radii are computed from the same coefficient record. Connecting segments organize the eight sampled angles; they do not assert a bound between those angles.

The complete withheld-query register

For each radius $r = 0.2, 0.5, 0.85$, the eight directions are $q_{r,k} = r\exp(i(0.13 + 2\pi k/8))$, $k = 0, \ldots, 7$. The last column is $|H(q) - c(q)|/\rho(q)$, computed from the full-precision values. A ratio at most one records containment in the feasible disk. It is not a probability or a percentage prediction error.

r, k	$\mathrm{Re}\, H(q)$	$\mathrm{Im}\, H(q)$	$\rho(q)$	Ratio
0.2,0	0.0006400	0.0000227	0.0150676	0.03199
0.2,1	0.0005590	0.0000899	0.0150685	0.02563
0.2,2	0.0004912	0.0000723	0.0150876	0.01875
0.2,3	0.0004641	0.0000316	0.0150883	0.01508
0.2,4	0.0004591	-0.0000062	0.0150705	0.01421
0.2,5	0.0004695	-0.0000448	0.0150702	0.01596
0.2,6	0.0005093	-0.0000833	0.0150873	0.02068
0.2,7	0.0005885	-0.0000818	0.0150860	0.02830
0.5,0	0.0014392	0.0006774	0.0326291	0.06803
0.5,1	0.0005866	0.0001659	0.0327045	0.02385
0.5,2	0.0003755	0.0001262	0.0435639	0.01258
0.5,3	0.0004234	0.0000390	0.0435157	0.00778
0.5,4	0.0004001	-0.0000144	0.0326436	0.00711
0.5,5	0.0004080	-0.0000371	0.0327112	0.00897
0.5,6	0.0004033	-0.0002242	0.0435622	0.01477
0.5,7	0.0004732	-0.0001546	0.0435027	0.03488
0.85,0	-0.0025209	0.0887587	0.2842258	0.30006
0.85,1	0.0016758	-0.0017722	0.2848304	0.00790
0.85,2	0.0018963	-0.0007411	0.3615068	0.00889
0.85,3	-0.0008004	0.0018809	0.3611757	0.00445
0.85,4	0.0005728	-0.0000763	0.2843544	0.00157
0.85,5	0.0000084	0.0022454	0.2848606	0.00840
0.85,6	-0.0027817	-0.0004881	0.3614994	0.01070
0.85,7	0.0014377	-0.0047424	0.3611097	0.01161

The median radius is 0.0381070 and the largest is 0.3615068 at $r = 0.85$, $k = 2$. At that query the center is $-0.0009177 + 0.0008087i$ and the truth-to-center gap is 0.0032126. All 24 retained values satisfy the containment check.

What this field contributes

At probe radius 0.2, the eight admissible radii range from 0.0150676 to 0.0150883. At radius 0.5 they range from 0.0326291 to 0.0435639; at radius 0.85, from 0.2842258 to 0.3615068. These intervals separate the radial change from the variation across the eight directions.

The largest truth-to-center ratio among this field's 24 queries is 0.300063. The shared model and probe arrangement make the radii and containment ratios comparable across fields. Physical units belong to the histogram convention.

Power plant: Relative humidity

Relative humidity is a percentage-valued measurement. The nonnegative normalized histogram gives a bounded polynomial by construction, without requiring a probabilistic distribution over the unknown analytic functions.

The source supplies 9,568 records, with observed minimum 25.56 and maximum 100.16 in percent. There are 32 occupied bins. The largest count is 602, in bin 25; its lower and upper edges are 83.84125 and 86.1725. The bin index measures position in this declared interval system.

The complete coefficient record

Use $p_k = n_k/9568$ and $H(z) = \sum_{k=0}^{31} p_k z^k$. The edges follow $b_k = 25.56 + (2.33125)k$. Decimal edge parameters are displayed to nine significant digits; the saved source endpoints define the unrounded construction.

k	n_k	k	n_k
0	4	16	429
1	4	17	448
2	13	18	464
3	23	19	504
4	37	20	524
5	59	21	578
6	82	22	549
7	115	23	527
8	137	24	595
9	167	25	602
10	167	26	584
11	239	27	473
12	277	28	340
13	292	29	261
14	315	30	205
15	375	31	179

The exact count sum is 9568. Empty bins are retained as zero coefficients; shifting the later coefficients left would create a different fingerprint.

The values supplied to completion

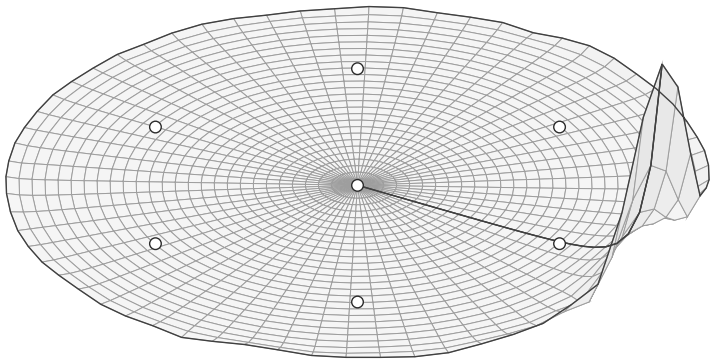
The observation nodes are $z_0 = 0$ and $z_{j+1} = 0.65e^{2\pi ij/6}$ for $j = 0, \ldots, 5$. The index in the following table identifies those nodes. These seven values are calculated after the entire histogram has been prepared.

Node	$\operatorname{Re} H(z_j)$	$\operatorname{Im} H(z_j)$
0	0.0004181	0.0000000
1	0.0063956	0.0000000
2	0.0000925	0.0001444
3	0.0003158	-0.0001063
4	0.0002995	0.0000000
5	0.0003158	0.0001063
6	0.0000925	-0.0001444

The displayed decimals are for reading, not a replacement for the exact counts. Horner evaluation of the coefficient record reconstructs the values at any desired arithmetic precision.

The prepared data defines a surface of values

Height = |H(z)|; H(z) = Σ $p_k z^k$. This is the constructed analytic fingerprint.



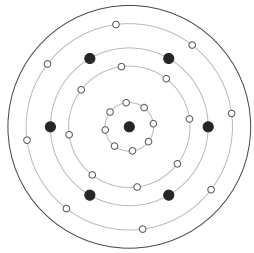
9,568 source records

32 nonnegative coefficients

Σ p_k = 1

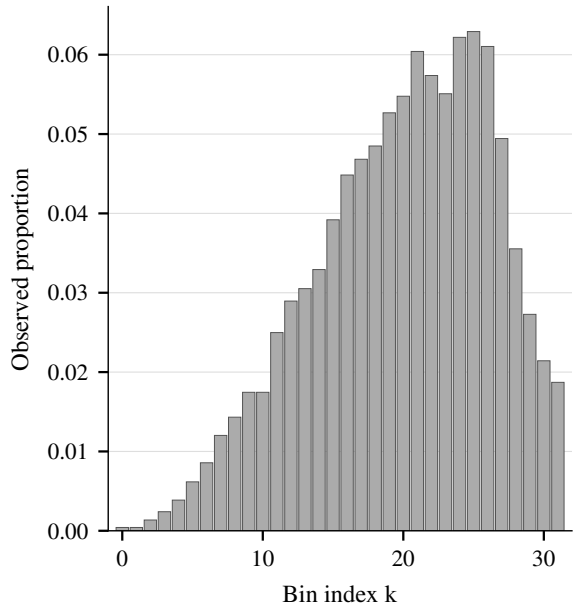
7 supplied function values

24 withheld queries in this field



Filled: supplied / open: withheld locations

Measured distribution



Completion by probe location

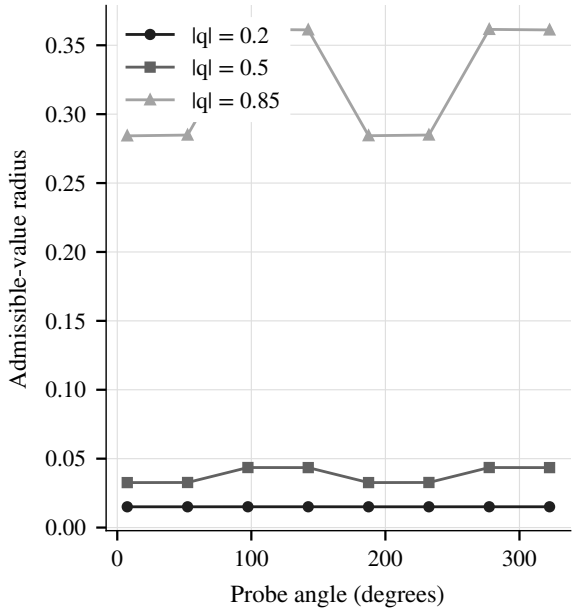


Figure H.16: Power plant: Relative humidity. The histogram and all retained query radii are computed from the same coefficient record. Connecting segments organize the eight sampled angles; they do not assert a bound between those angles.

The complete withheld-query register

For each radius $r = 0.2, 0.5, 0.85$, the eight directions are $q_{r,k} = r\exp(i(0.13 + 2\pi k/8))$, $k = 0, \ldots, 7$. The last column is $|H(q) - c(q)|/\rho(q)$, computed from the full-precision values. A ratio at most one records containment in the feasible disk. It is not a probability or a percentage prediction error.

r, k	$\mathrm{Re}\, H(q)$	$\mathrm{Im}\, H(q)$	$\rho(q)$	Ratio
0.2,0	0.0005787	0.0000369	0.0150699	0.02144
0.2,1	0.0004322	0.0001209	0.0150706	0.01747
0.2,2	0.0003659	0.0000553	0.0150895	0.01286
0.2,3	0.0003694	0.0000131	0.0150899	0.01046
0.2,4	0.0003740	-0.0000020	0.0150720	0.00996
0.2,5	0.0003660	-0.0000227	0.0150719	0.01101
0.2,6	0.0003751	-0.0000779	0.0150892	0.01417
0.2,7	0.0004832	-0.0001235	0.0150883	0.01917
0.5,0	0.0017932	0.0007568	0.0326348	0.04188
0.5,1	0.0001263	0.0002849	0.0327087	0.01748
0.5,2	0.0002510	0.0000126	0.0435678	0.00810
0.5,3	0.0003610	-0.0000318	0.0435185	0.00560
0.5,4	0.0003732	-0.0000047	0.0326456	0.00577
0.5,5	0.0003307	0.0000344	0.0327136	0.00581
0.5,6	0.0002059	-0.0000605	0.0435666	0.01028
0.5,7	0.0001271	-0.0006009	0.0435090	0.02404
0.85,0	-0.0038158	0.0477368	0.2842598	0.16097
0.85,1	0.0004595	-0.0006637	0.2848440	0.00675
0.85,2	0.0004725	0.0008112	0.3615180	0.00310
0.85,3	-0.0003035	-0.0000155	0.3611823	0.00198
0.85,4	0.0001208	-0.0008520	0.2843592	0.00390
0.85,5	0.0005984	-0.0000297	0.2848665	0.00082
0.85,6	-0.0001374	0.0016418	0.3615136	0.00728
0.85,7	-0.0012958	0.0017617	0.3611304	0.01378

The median radius is 0.0381113 and the largest is 0.3615180 at $r = 0.85$, $k = 2$. At that query the center is $-0.0005416 + 0.0003305i$ and the truth-to-center gap is 0.0011223. All 24 retained values satisfy the containment check.

What this field contributes

At probe radius 0.2, the eight admissible radii range from 0.0150699 to 0.0150899. At radius 0.5 they range from 0.0326348 to 0.0435678; at radius 0.85, from 0.2842598 to 0.3615180. These intervals separate the radial change from the variation across the eight directions.

The largest truth-to-center ratio among this field's 24 queries is 0.160969. The shared model and probe arrangement make the radii and containment ratios comparable across fields. Physical units belong to the histogram convention.

Power plant: Electrical output

Electrical output is an observed field, rather than a regression target here. Withheld queries are withheld evaluations of its prepared polynomial; they are not withheld plant-output measurements.

The source supplies 9,568 records, with observed minimum 420.26 and maximum 495.76 in MW. There are 31 occupied bins. The largest count is 577, in bin 8; its lower and upper edges are 439.135 and 441.494375. The bin index measures position in this declared interval system.

The complete coefficient record

Use $p_k = n_k/9568$ and $H(z) = \sum_{k=0}^{31} p_k z^k$. The edges follow $b_k = 420.26 + (2.359375)k$. Decimal edge parameters are displayed to nine significant digits; the saved source endpoints define the unrounded construction.

k	n_k	k	n_k
0	2	16	274
1	0	17	320
2	118	18	392
3	190	19	411
4	377	20	359
5	466	21	327
6	538	22	334
7	554	23	299
8	577	24	301
9	538	25	267
10	508	26	219
11	427	27	175
12	408	28	115
13	328	29	82
14	321	30	37
15	284	31	20

The exact count sum is 9568. Empty bins are retained as zero coefficients; shifting the later coefficients left would create a different fingerprint.

The values supplied to completion

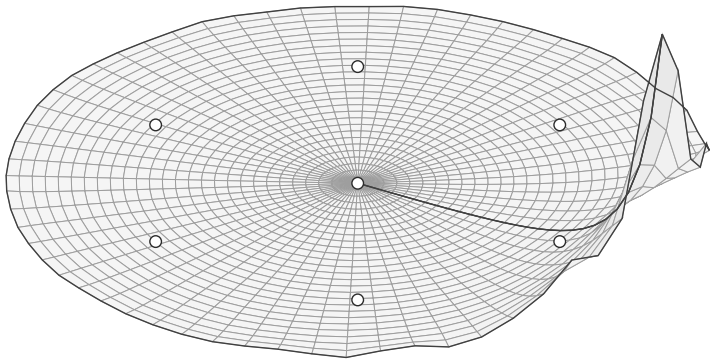
The observation nodes are $z_0 = 0$ and $z_{j+1} = 0.65e^{2\pi ij/6}$ for $j = 0, \ldots, 5$. The index in the following table identifies those nodes. These seven values are calculated after the entire histogram has been prepared.

Node	$\operatorname{Re} H(z_j)$	$\operatorname{Im} H(z_j)$
0	0.0002090	0.0000000
1	0.0354227	0.0000000
2	-0.0049362	-0.0031590
3	-0.0006420	-0.0021887
4	0.0039993	-0.0000000
5	-0.0006420	0.0021887
6	-0.0049362	0.0031590

The displayed decimals are for reading, not a replacement for the exact counts. Horner evaluation of the coefficient record reconstructs the values at any desired arithmetic precision.

The prepared data defines a surface of values

Height = $|H(z)|$; $H(z) = \Sigma\, p_k z^k$. This is the constructed analytic fingerprint.



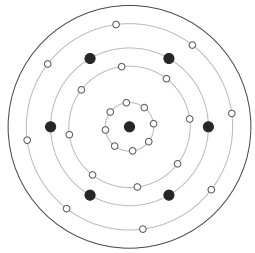
9,568 source records

32 nonnegative coefficients

$\Sigma\, p_k = 1$

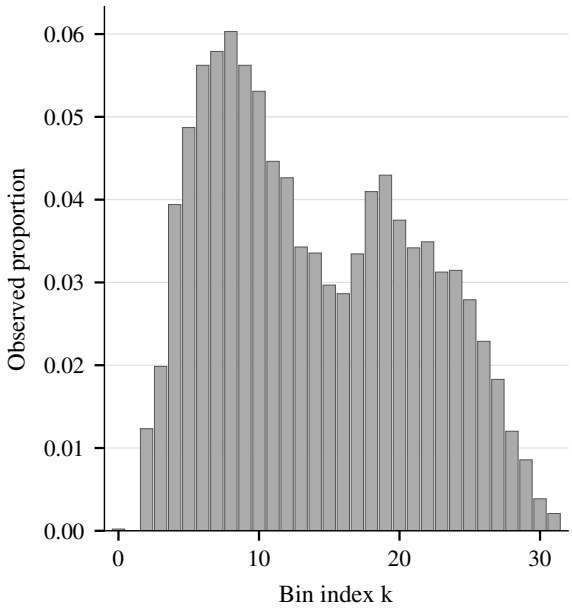
7 supplied function values

24 withheld queries in this field



Filled: supplied / open: withheld locations

Measured distribution



Completion by probe location

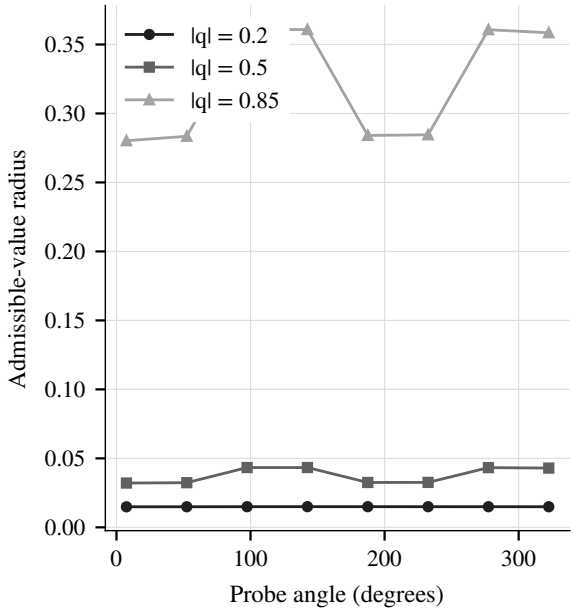


Figure H.17: Power plant: Electrical output. The histogram and all retained query radii are computed from the same coefficient record. Connecting segments organize the eight sampled angles; they do not assert a bound between those angles.

The complete withheld-query register

For each radius $r = 0.2, 0.5, 0.85$, the eight directions are $q_{r,k} = r\exp(i(0.13 + 2\pi k/8))$, $k = 0, \dots, 7$. The last column is $|H(q) - c(q)|/\rho(q)$, computed from the full-precision values. A ratio at most one records containment in the feasible disk. It is not a probability or a percentage prediction error.

r, k	$\operatorname{Re} H(q)$	$\operatorname{Im} H(q)$	$\rho(q)$	Ratio
0.2,0	0.0009029	0.0002313	0.0148926	0.09147
0.2,1	-0.0001179	0.0004888	0.0149165	0.07493
0.2,2	-0.0001639	-0.0002329	0.0149677	0.05468
0.2,3	0.0003554	-0.0003616	0.0149881	0.04355
0.2,4	0.0005833	0.0000904	0.0149752	0.04092
0.2,5	0.0001781	0.0003971	0.0149654	0.04620
0.2,6	-0.0002670	0.0000371	0.0149576	0.06052
0.2,7	0.0002020	-0.0006491	0.0149235	0.08215
0.5,0	0.0099126	0.0052501	0.0321053	0.16702
0.5,1	-0.0041345	0.0009299	0.0323732	0.07776
0.5,2	-0.0008554	-0.0013473	0.0433294	0.03398
0.5,3	0.0008225	-0.0017845	0.0433539	0.02231
0.5,4	0.0022527	0.0005503	0.0325307	0.01859
0.5,5	0.0000705	0.0017476	0.0325786	0.02489
0.5,6	-0.0014489	0.0012973	0.0432784	0.04262
0.5,7	-0.0039951	-0.0050151	0.0429621	0.10731
0.85,0	0.0854624	0.1269509	0.2802583	0.37624
0.85,1	-0.0087371	-0.0126671	0.2834408	0.03370
0.85,2	-0.0003062	-0.0006630	0.3609416	0.00848
0.85,3	0.0005926	-0.0052884	0.3608294	0.00968
0.85,4	0.0069984	0.0038203	0.2840763	0.00430
0.85,5	-0.0013795	0.0033783	0.2845560	0.00773
0.85,6	0.0014930	0.0019556	0.3607450	0.00817
0.85,7	-0.0292865	0.0108483	0.3584951	0.08736

The median radius is 0.0377704 and the largest is 0.3609416 at $r = 0.85$, $k = 2$. At that query the center is $-0.0028623 + 0.0010221i$ and the truth-to-center gap is 0.0030615. All 24 retained values satisfy the containment check.

What this field contributes

At probe radius 0.2, the eight admissible radii range from 0.0148926 to 0.0149881. At radius 0.5 they range from 0.0321053 to 0.0433539; at radius 0.85, from 0.2802583 to 0.3609416. These intervals separate the radial change from the variation across the eight directions.

The largest truth-to-center ratio among this field's 24 queries is 0.376245. The shared model and probe arrangement make the radii and containment ratios comparable across fields. Physical units belong to the histogram convention.

Notation guide

Notation is introduced where it is used. Reused symbols carry a local definition; a matching letter is not a claim that the objects are identical.

Symbol	Meaning
U, u	Latent entity set and a particular entity; context distinguishes this from coordinate domains.
$\mathcal{O}_i, E_i$	Observations in channel i and their encoder.
S	A semantic sameness structure; in the Davis statistic discussion, S instead denotes the scalar statistic and is explicitly reintroduced.
M, g	A smooth manifold and its Riemannian metric.
T_pM	Tangent space at p.
φ_i, U_i, V_i	A chart and its manifold and coordinate domains.
T_{ij}	A translator from description i to description j.
DT_x, J; $\sigma_{\min}(J)$	Jacobian and its least gain on unit input vectors, zero if it loses a direction.
G_x	Matrix of a metric in the x coordinates.
$L_g(\gamma)$	Length of a curve in the metric g.
d_g, d_M	The stated intrinsic distance, with its admissible path family understood.
$D(x, y)$	Directed translator-path cost.
$d_{\mathrm{sym}}, \tilde{d}$	The maximum of the two directed costs on a mutually reachable component.
X_T, d_T	The quotient by zero cost in both directions, and the induced symmetric distance.
$\mathcal{R}$	A declared operating region.
Φ, Ψ	Realization and reconstruction maps; inverses are used only when defined.
$\mathsf{SamTrans}(S)$, $\mathsf{SamGeom}(S)$	The stated categories of translator and geometric realizations.
$\mathcal{P}(L)$	Permitted paths up to the specified horizon.
ε, η	Explicit approximation or error bounds, whose units depend on their definition.
$\varepsilon_{\mathrm{loc}}$	Uniform relative chart-distortion allowance, dimensionless and less than one.
$\mathrm{Cyc}, \varepsilon_{\mathrm{cyc}}$	Pointwise cocycle residual and its uniform bound on the stated overlap.
$\hat{\varepsilon}_{\mathrm{rt},0.99}$, $\hat{\varepsilon}_{\mathrm{cyc},0.99}$	Empirical 99th percentiles of the round-trip and cocycle residuals.
a, b	Positive lower and upper length or distance comparison factors.
τ, τ_s, τ_d	A decision threshold, or lower and upper thresholds for abstention.
$\kappa_{\mathrm{hard}}, \kappa_{\mathrm{soft}}$	Configuration margins in the Davis geometric construction.
$E_{\mathrm{geom}}, E_{\mathrm{link}}$	Named geometry and linkage failure budgets, with event definitions supplied.
ξ, ζ	Calibration and abstention-failure quantities in the decision framework.
$q, -q$	Two unit quaternion representatives of one physical orientation.
$S^3, SO(3)$	The unit-quaternion sphere and the rotation group.
ρ	The quaternion covering map; in interpolation tables, a separately defined feasible-disk radius.
θ, d_{ch} (attitude)	Physical rotation angle and half the operator norm of the rotation-matrix difference. The quaternion chord uses a different ambient space.
Γ^k_{ij}	Christoffel coefficients of the Levi–Civita connection.

Symbol	Meaning
A, F	A connection form and its curvature, with $F = dA + A \wedge A$.
K	Gaussian curvature in the sphere example; capacity proposals require their own explicit curvature definition.
$\|\cdot\|_2, \|\cdot\|_F$	Euclidean/operator norm as determined by the argument, and Frobenius matrix norm.
$\|v\|_G$	The vector norm defined by a positive-definite matrix G.
$H(z), p_k, n_k$	Analytic histogram fingerprint, bin proportion and bin count.
$\mathbb{D}$	The open unit disk.
$c(q), \rho(q)$	Center and radius of a feasible-value disk at query q.
P	The Pick interpolation matrix in the analytic-completion chapter.
μ_y, σ_y^2, π_y	Conditional statistic mean, conditional variance and class prevalence on the stated population.
$I(p)$	Fisher information in the Bernoulli model.
$P_u = uu^T$	The rank-one orthogonal projector representing an unoriented line.
$\delta([u], [v])$	The shortest angle between two unoriented lines.
$\mathrm{Exp}_p, \mathrm{Retr}_p$	The exponential map and the explicitly chosen retraction at a point.
Σ_x	Coordinate covariance, transformed by $J\Sigma_x J^T$.
B (graph context)	An oriented incidence matrix; elsewhere B is explicitly defined as a small operator.
d_0, d_1	Linear compatibility maps with $d_1 d_0 = 0$.
$U_x(q)$	Observation update followed by transition for the specified hidden Markov model.
α, β (HMM)	Transition and emission-error probabilities; not the geometric angles used in other chapters.
$\Omega = E^T \dot{E}$	The skew-symmetric angular-rate matrix of an orthonormal moving frame.
$K_A = AA^* - A^*A$	The matrix self-commutator in the operator study; not a Riemann curvature tensor.
$T = \phi(A)$	The operator mapped from its elliptical numerical range to the unit disk.
G, Z (certificate)	The positive metric and positive separating witness in the retained rational certificate.
$\mathcal{T}_k$	The degree-k Chebyshev polynomial of the first kind.
J, ω (complex geometry)	The complex structure and compatible two-form; J also denotes a Jacobian elsewhere, where defined.
p_U	The one-sided zero-failure binomial confidence endpoint for the specified sampling model.
$\mathcal{P}_n$	The manifold of real symmetric positive-definite $n \times n$ matrices.
$d_{\mathcal{P}}$	The affine-invariant distance between positive-definite rulers.
μ, σ	Mean and positive standard deviation of the univariate normal family.
$I(\mu, \sigma)$	Fisher information matrix in the displayed calibration coordinates.
T_{BA}	Rigid coordinate map from frame A to frame B.
$D, \widehat{D}$	Reference and translated distances for the same observed pair in the transfer calculation.
Q_Δ	Grid rounding with width Δ and a declared tie rule.
U, P_+, P_- (signals)	Half-turn shift, invariant average and sign-changing half-difference; Chapter 14.
$\|f\|_2^2$ (signals)	Uniform angular mean of $\|f\|^2$; distinct from the finite-dimensional norms above.
α (annulus)	The closed angle form $(-y\,dx + x\,dy)/(x^2 + y^2)$; its integral around a loop records 2π times the winding number.

Framework glossary

Abstention. A decision to withhold a definite label because the supplied evidence or guarantee does not settle it. Its frequency is reported through coverage.

Admissible path. A route permitted by the model. Distances compared by an infimum must use the same path family unless the change of family is explicitly analyzed.

Affine-invariant metric. A metric whose value is preserved by the specified invertible congruence transformations. On positive-definite matrices, the displayed trace metric compares relative changes in a ruler.

AMC. Adaptive Multi-Vector Correlation: the telemetry representation work used as the recurring engineering example.

Atlas. A collection of compatible coordinate charts covering a manifold. It supplies several local maps of the same geometric object.

Benign evolution. A task-declared class of changes that preserves the relevant identity or configuration. It is a modeling choice that needs evidence.

Basepoint. A specified starting point kept fixed when comparing loops; fundamental groups and their induced homomorphisms use this choice.

Bi-Lipschitz bound. A two-sided multiplicative comparison of distances with strictly positive lower and finite upper constants. An additive slack gives a different, approximate statement.

Boundary mass. The probability assigned to observations close enough to a decision threshold that a bounded numerical error can change their labels.

Calibration. Establishing how an instrument reading relates to a reference; in probabilistic prediction, agreement between a reported probability or uncertainty claim and its behavior on the stated population.

Category. Objects together with composable maps, identities and an associative composition law. Numerical error bounds accompany maps rather than replace those laws.

Causal state. An equivalence class of observation histories with the same conditional law of every finite future observation sequence, under the specified stochastic model.

Chart. A homeomorphism from an open part of a manifold to a Euclidean open set, belonging to the specified smooth atlas.

Chartability. The stated local structure and compatibility conditions under which descriptions can be treated as coordinate patches. Finite diagnostics alone are not a global existence theorem.

Chordal distance. A comparison using a chord in an ambient representation. Its relation to intrinsic distance depends on the object and normalization.

Christoffel symbols. Coordinate coefficients of a connection. They account for the changing coordinate basis and are not by themselves tensor components.

Cocycle condition. Agreement between a direct transition and a two-step transition on a triple overlap.

Cohomology. An algebraic measure of closed data that are not produced by the preceding compatibility map. The spaces and maps must be specified; a generic numerical residual is not automatically a cohomology class.

Complex structure. A tangent-space operator J with $J^2 = -I$, together with the needed smoothness and integrability conditions when it defines complex coordinates.

Condition number. A ratio quantifying sensitivity or unequal scaling. State the matrix, norm, coordinates and units, or compare metrics through generalized eigenvalues.

Configuration. A task-relevant state or class assigned by the model. Configuration margins describe the separation required between designated classes.

Congruence. The matrix transformation $G \mapsto AGA^T$ for invertible A. Its physical interpretation depends on what G represents.

Connected partition. A partition of a graph vertex set in which every block induces a connected subgraph.

Connection. A rule for comparing tangent or fiber quantities along paths. Parallel transport is determined by this rule.

Coverage. The probability or measured fraction of cases on which a selective system gives a definite answer.

Covering map. A projection for which each sufficiently small base neighborhood lifts to disjoint copies mapped homeomorphically onto it.

Curvature. A geometric quantity describing how the local structure departs from a flat model. Specify sectional, Gaussian, scalar or connection curvature as appropriate.

Davis manifold. The framework combining an observation map, Riemannian model, allowed paths, distortion controls and configuration margins.

Davis system. A Davis manifold together with a statistic, decision rule, error accounting, operating horizon and abstention behavior.

Deck transformation. A symmetry of a covering space that leaves its projection unchanged. On the quaternion attitude cover, $q \mapsto -q$ exchanges the two representatives.

DHOOM. The GIGI record encoding used in the tested three-sensor round trip. The decoder restores the original fixture fields.

Diffeomorphism. A smooth bijection with a smooth inverse. It allows a smooth description to be transported both ways on the stated domains.

Directed path cost. An infimum over permitted routes that can differ when the endpoints are exchanged. It is not automatically a symmetric metric.

Distortion. A quantified change in lengths or distances under a map, with its norm, domain and constants stated.

Double cover. A covering map with exactly two representatives over every base point. Two related descriptions alone do not establish this structure.

Dual coordinates. Natural and expectation coordinates related by the gradient of a strictly convex exponential-family potential. Their one-to-one translation is distinct from a two-sheeted cover.

Encoder. A map from observations into a representation. Its information loss and geometric rank require separate examination.

Equivalence relation. A reflexive, symmetric and transitive relation. A fixed distance tolerance usually fails transitivity.

Evenly covered neighborhood. An open base neighborhood whose inverse image splits into disjoint copies, each mapped homeomorphically onto it.

Exponential family. A statistical family whose density has a declared exponential form with sufficient statistics, natural parameters and a normalization potential.

Exponential map. The map taking an initial tangent velocity to the endpoint of its geodesic after unit parameter time, where that geodesic is defined.

Fiber. The collection of points mapped to one base point. In the record example it identifies the values associated with one observation address.

Fingerprint. The explicitly constructed analytic polynomial of bin proportions in the GIGI measured-data study. It retains that binned distribution, not all raw records.

Fisher information. The expected outer product of the score of a regular statistical model. Where positive definite, it supplies a statistical Riemannian metric.

Fisher score. The parameter derivative of the log density in a regular statistical model. Its expected outer product defines the Fisher information metric.

Fourier mode. An angular oscillation $e^{ik\theta}$ with integer frequency index k. Under a half-turn it gains the factor $(-1)^k$, so precisely the even-index modes describe an unoriented axis.

Fréchet mean. A minimizer of the stated weighted sum of squared distances on a metric space. Existence and uniqueness require their own conditions.

Functor. An assignment of objects and maps preserving identities and composition. Approximate reconstruction needs separately quantified defects.

Gauge freedom. A change of an unobservable reference convention, such as a common clock origin or global rigid frame, that leaves the specified relative observations unchanged.

Generalized eigenvalue. An eigenvalue comparing two quadratic forms. Its square root supplies a sharp local length-ratio factor for positive-definite metrics.

Geodesic. A curve with zero covariant acceleration in the chosen metric. It is locally straight, but not necessarily globally minimizing over a long interval.

Geometric realization. A specified geometric model and map relating it to a representation system. Its existence, injectivity and distance bounds are separate hypotheses.

GIGI. The computational research developed through GIGI Builds, GIGI Thinks and the geometric-inference study used in the application chapters.

Harmonic conjugate. A function v paired with a harmonic u by the Cauchy–Riemann equations, making $u + iv$ holomorphic. Global existence also requires every period of $-u_y\,dx + u_x\,dy$ to vanish.

Hausdorff condition. The requirement that distinct points have disjoint neighborhoods. It is one of the global topological conditions needed in the manifold construction.

Hellinger distance. A distance between probability distributions formed from their square roots; this book states the normalization explicitly.

Hidden-state belief. A posterior or predicted probability vector over a specified hidden-state model. It is a representation of predictive information, not automatically a distinct causal state.

Holonomy. The transformation produced by transport around a closed path under a specified connection or comparison rule.

Hopf fibration. The phase quotient $S^3 \to \mathbb{CP}^1$ with circle fibers. It differs from the two-point quaternion cover of $SO(3)$.

ICARUS. The geometric flight-control project supplying the exact orientation model and the stated simulator comparisons.

Identity. Sameness of the task-defined underlying object or event. It is distinct from proximity of measured values.

Immersion. A smooth map with injective derivative at every point. It preserves local tangent rank but need not be globally injective.

In-regime region. The domain on which the stated assumptions and bounds are intended to hold.

Incidence matrix. A matrix representing oriented edge differences of node values. Its nullspace records freedoms such as a common clock origin.

Invariant. A quantity unchanged by a specified class of transformations. Name the transformations as well as the quantity.

Jacobi field. The infinitesimal separation field obtained by varying a family of geodesics. Its equation relates neighboring routes to curvature.

Jacobian. The matrix of first derivatives of a map in chosen coordinates. Its singular values describe local scaling and rank.

Kullback–Leibler divergence. A generally asymmetric discrepancy between probability distributions. Its local quadratic term yields the Fisher metric under the stated regularity assumptions.

Legendre transform. The convex dual potential obtained by optimizing the linear pairing minus the original convex function.

Lift. A path or point in a covering space that projects to a specified base path or point. A path lift is fixed by choosing its initial representative.

Lineage. The recorded relationship between an observation and later transformations or copies of it.

Lipschitz bound. A bound d(f(x),f(y)) no larger than a constant times d(x,y), with the metrics and domain specified.

Lyapunov metric. A positive-definite quadratic ruler that decreases under a specified system evolution, as established by an operator inequality.

Manifold. A Hausdorff, second-countable space locally described by Euclidean open sets, with the required compatibility for a smooth structure.

Margin. The separation available before an uncertainty or representation error can cross a specified decision boundary.

Matrix geometric mean. The midpoint of the displayed affine-invariant positive-definite-matrix geodesic. It is generally not an entrywise geometric mean.

Metric compatibility. Agreement between an evolving metric and a transport rule that preserves their quadratic lengths; the finite matrix version is derived explicitly in the text.

Metric tensor. A smoothly varying positive-definite inner product on tangent spaces; the local ruler from which curve lengths are calculated.

Morphisms. The structure-preserving maps allowed between the objects of a specified category. Their required properties must be explicit.

Nonnormal operator. An operator for which $AA^* \neq A^*A$. Eigenvalues alone can fail to describe transient norm growth.

Operator norm. The largest amplification of a unit vector under a linear map in the chosen input and output norms.

Orthogonal projection. The closest element of a closed subspace in a specified inner-product norm. For uniformly weighted circular signals, averaging opposite representatives projects onto the axis observables.

Parallel transport. Evolution of a tangent or fiber quantity along a path under the rule that its covariant derivative vanishes.

Period. The integral of a differential form around a closed path. A nonzero period obstructs recovering a single-valued potential by integration.

Pick matrix. The matrix whose positive semidefiniteness tests bounded analytic interpolation for the stated disk nodes and values.

Poincare metric. The hyperbolic ruler used on the unit disk, here normalized as four times the Euclidean squared differential divided by (1 minus squared radius) squared.

Posterior correctness. The probability that a target label is correct conditional on the actual alert or selection event.

Projector mean. The leading-eigenspace construction minimizing the stated sum of squared projector differences. It is not automatically a mean for squared intrinsic distance.

Pseudometric. A nonnegative symmetric distance satisfying the triangle inequality that may assign zero to distinct points.

Pullback metric. A ruler transported through a smooth map by applying its derivative to tangent vectors. It is positive definite exactly when the derivative is injective.

Quantization. A map from a continuous or finer-resolution value to a discrete grid. Its reconstruction allowance and tie convention are part of the comparison contract.

Quantile. A specified rank or interpolation summary of a distribution or sample. An empirical 95th percentile is not a uniform upper bound on unseen values.

Quaternion. A four-component algebraic representation whose unit elements double-cover three-dimensional rotations.

Quotient. A space formed by identifying points under a specified equivalence relation. What is identified determines what information is retained.

Rank. The dimension of the image of a linear map or derivative. Losing rank makes some tangent directions invisible.

Realizability. The existence of the specified geometric realization. Exact atlas data can provide a construction; approximate learned data require additional conditions.

Retraction. A local map from tangent vectors back to a manifold that agrees with the identity and its first derivative at zero; it need not equal the exponential map.

Relative distortion. A dimensionless comparison such as $|d_T/d_i - 1|$ for positive reference distance. Absolute error $|d_T - d_i|$ has distance units.

Round trip. A conversion followed by a return map. Exact inversion and bounded reconstruction error are different claims.

Saturation. Clipping a measurement at an instrument or representation limit. A finite reconstruction allowance requires a further bound on the possible input.

Schur complement. The reduced matrix expression obtained by eliminating an invertible block. Its positive-semidefinite condition yields the completion disk in the strictly feasible Pick calculation.

Second countability. The existence of a countable basis of open sets. It is part of the manifold conditions used in the atlas construction.

Section. An assignment choosing one point in each fiber over its domain, so projection returns the original base point. A record is a value of such an assignment.

Selective risk. Error conditional on the system answering. It must be reported with coverage.

Singular value. A nonnegative local scaling factor of a matrix, obtained from the square roots of the eigenvalues of its transpose times itself.

Smoothness. Differentiability to the stated order. A qualitative smoothness claim is not itself a numerical derivative bound.

Stein inequality. A discrete-time contraction inequality $T^*GT \preceq G$ in a positive-definite metric.

Sufficient statistic. A statistic retaining all information about a parameter in a specified statistical model. Sufficiency for one model does not imply sufficiency for every task.

Translator. A map between descriptions. It is an exact coordinate transition only when the appropriate domain, inverse and compatibility conditions hold.

Update commutator. The difference between two specified orders of applying update maps. It measures order sensitivity without by itself defining differential-geometric holonomy.

Winding number. The signed integer count of circuits around an excluded point. A logarithm retains that integer; a square root retains only its parity.

Witness. An explicit object or example that can be checked to verify a claim or exhibit an obstruction.

Sources and reproducibility

The classical results in this book are distinguished from the author's constructions and computational studies. The cited manuscripts describe the research program; the worked calculations in the text state their own hypotheses. An experiment supports the specified implementation, dataset and protocol.

The application records

AMC. The deposited manuscript is Davis's *Adaptive Embedding Format Switching and Semantic Translation in High-Volume Telemetry Systems* (2025), Zenodo record 17211403. The author's engineering account concerns telemetry work at NASA; the paper's reported evaluation uses public MODIS data. The coordinate, sparse-storage and quantization examples in this book are fully specified mathematical calculations, not additional measurements from that evaluation.

ICARUS. The orientation example uses the standalone v3 simulator and the benchmark account in *The Geometry of Flight*. A fresh 500-trial run on 9 November 2026 reproduced all 24 saved scenario-result rows at their reporting precision, excluding computational timing. The fixed-gain geometric baseline is the comparator implemented in that simulator. These records are simulated outcomes, not flight tests or energy measurements.

GIGI books and records. *GIGI Builds*, chapters 2 and 6, supplies the record and DHOOM material; *GIGI Thinks*, chapters 1, 3, 4 and 6, supplies the half-angle, translator, frame and overlap discussions. The complete author-supplied editions contain 509 and 520 pages. The three-record DHOOM example was replayed through the local converter on 9 November 2026 and decoded back to the original records. The frame and score-margin calculations are constructed worked examples.

Measured-data inference. The geometric-inference companion, research draft version 0.2 of 29 October 2026, supplies the wine and power-plant study. Its public companion links the specifications, data and verification record. The raw data are the UCI Wine Quality and Combined Cycle Power Plant datasets, distributed under CC BY 4.0. The manuscript uses 6,497 wine rows and the first 9,568-row power-plant worksheet, not duplicated shuffled worksheets.

All 408 completion queries over 17 fields were replayed on 9 November 2026 using the preserved executable identified by the research record. Final radii reproduced the saved results. Contradiction, same-size edit and restart checks passed, as did a separate planted identity-function example. An independently implemented checker verified 64 prefix disks at 70-digit precision, nesting and held-out containment, 16 derivative and harmonic-identity examples, and all 17 field totals. The intermediate prefix table uses the independently reconstructed values; the final radii also have the executable replay. Harmonic-recovery numbers are from the saved study and independent arithmetic checks.

How to locate the research

The Davis Geometric research portal, https://www.davisgeometric.com/research, and project pages below provide public entry points. Some developing manuscripts and software records are author-held research versions rather than separately deposited publications. Their titles and reviewed versions are identified here without inventing a DOI. The book does not treat the Field Equations program's conjectures as completed proofs.

References

[1] Samuel Eilenberg and Saunders Mac Lane. “General Theory of Natural Equivalences.” *Transactions of the American Mathematical Society* 58(2), 231–294, 1945. https://doi.org/10.1090/S0002-9947-1945-0013131-6. Archival copy: https://zenodo.org/records/3928610.

[2] British Museum. *Hieroglyphs: Unlocking Ancient Egypt*. Exhibition large-print guide, 2022. Rosetta Stone discovery and three-script inscription. https://www.britishmuseum.org/sites/default/files/2022-10/Hieroglyphs_unlocking_ancient_Egypt_large_print_guide_The_British_Museum.pdf.

[3] John M. Lee. *Introduction to Smooth Manifolds*. Second edition. Graduate Texts in Mathematics 218. Springer, 2013. Author’s page: https://sites.math.washington.edu/~lee/Books/ISM/.

[4] John M. Lee. *Introduction to Riemannian Manifolds*. Second edition. Graduate Texts in Mathematics 176. Springer, 2018. Author’s page: https://sites.math.washington.edu/~lee/Books/RM/.

[5] Allen Hatcher. *Algebraic Topology*. Cambridge University Press, 2002. Covering spaces: chapter 1. Author’s electronic edition: https://pi.math.cornell.edu/~hatcher/AT/ATpage.html.

[6] Joshua B. Tenenbaum, Vin de Silva and John C. Langford. “A Global Geometric Framework for Nonlinear Dimensionality Reduction.” *Science* 290(5500), 2319–2323, 2000. https://doi.org/10.1126/science.290.5500.2319.

[7] Jean-Yves Audibert and Alexandre B. Tsybakov. “Fast Learning Rates for Plug-in Classifiers.” *Annals of Statistics* 35(2), 608–633, 2007. https://doi.org/10.1214/009053606000001217. Author manuscript: https://arxiv.org/abs/0708.2321.

[8] Wassily Hoeffding. “Probability Inequalities for Sums of Bounded Random Variables.” *Journal of the American Statistical Association* 58(301), 13–30, 1963. https://doi.org/10.1080/01621459.1963.10500830.

[9] Nihat Ay, Jürgen Jost, Hông Vân Lê and Lorenz Schwachhöfer. “Information Geometry and Sufficient Statistics.” *Probability Theory and Related Fields* 162, 327–364, 2015. https://doi.org/10.1007/s00440-014-0574-8. Author manuscript: https://arxiv.org/abs/1207.6736.

[10] Georg Pick. “Über die Beschränkungen analytischer Funktionen, welche durch vorgegebene Funktionswerte bewirkt werden.” *Mathematische Annalen* 77, 7–23, 1916. https://doi.org/10.1007/BF01456817.

[11] Michel Crouzeix and César Palencia. “The Numerical Range Is a $(1+\sqrt{2})$-Spectral Set.” *SIAM Journal on Matrix Analysis and Applications* 38(2), 649–655, 2017. Author manuscript: https://arxiv.org/abs/1702.00668.

[12] Paulo Cortez and collaborators. *Wine Quality*. UCI Machine Learning Repository. https://doi.org/10.24432/C56S3T. Dataset, CC BY 4.0.

[13] Pinar Tufekci and Heysem Kaya. *Combined Cycle Power Plant*. UCI Machine Learning Repository. https://doi.org/10.24432/C5002N. Dataset, CC BY 4.0.

[14] Bee Rosa Davis. *Adaptive Embedding Format Switching and Semantic Translation in High-Volume Telemetry Systems*. 2025. https://doi.org/10.5281/zenodo.17211403. Reading guide: https://www.davisgeometric.com/research/papers/amc.

[15] Bee Rosa Davis. *PRISM reconciliation implementation*. Author-held source, revision `66c88ee0189b120c`. Matching logic in `prism/reconcile/matcher.py`, with the fixed-currency check covered by the repository tests.

[16] Bee Rosa Davis. *HERALD: High-resolution Early Recognition of Antigenic Landscape Divergence*. Academic draft v0.1, author-held LaTeX source examined 9 November 2026. Pullback metric, distortion regime and conditional decision bounds. Research portal: https://www.davisgeometric.com/research. Related author-held manuscript: *HERALD DMS Validation Study*, 19 November 2025. The constructed condition-number example in Chapter 21 is a separate mathematical fixture.

[17] Bee Rosa Davis. *The Davis Manifold*. Author's research manuscript, version examined 9 November 2026. Geometric construction, configuration margins, admissible paths, and finite-variance decision bounds. Research portal: https://www.davisgeometric.com/research.

[18] Bee Rosa Davis. *The Field Equations of Semantic Coherence*. Author's research manuscript, version examined 9 November 2026. Conjectural research program on completion, compatibility and capacity. Research portal: https://www.davisgeometric.com/research.

[19] Bee Rosa Davis. *The Double Cover Principle*, version 3. Author's research manuscript, examined 9 November 2026. Covering-map and observable-domain discussion. Research portal: https://www.davisgeometric.com/research.

[20] Bee Rosa Davis. *The Geometry of Flight*. Expanded second edition, 2026. Mathematical foundations and ICARUS benchmark discussion. Project and book reader: https://www.davisgeometric.com/icarus.

[21] Bee Davis. *GIGI Builds*. Author-supplied complete edition, 2026. Chapters 2 and 6; sections, records and DHOOM encoding.

[22] Bee Davis. *GIGI Thinks*. Author-supplied complete edition, 2026. Chapters 1, 3, 4 and 6; half-angle comparisons, translators, frames and chart overlaps.

[23] Bee Rosa Davis. Geometric inference research draft, version 0.2, 29 October 2026. Companion, specifications, data and verification: https://www.davisgeometric.com/research/papers/geometric-inference.html.

[24] Bee Rosa Davis. *Perfect Identity Does Not Exist*. Author's research manuscript, version examined 9 November 2026. Transported comparisons and identity. Research portal: https://www.davisgeometric.com/research.

[25] Bee Rosa Davis. *Causal States as Predictive Sections*. Author's research manuscript, version examined 9 November 2026. Predictive equivalence and ordered updates. Research portal: https://www.davisgeometric.com/research.

[26] Bee Rosa Davis. Marcella accumulated-rotation bridge and public-interface documentation, author-held software version examined 9 November 2026. Single-plane diagnostics and normalized traces are distinguished in that implementation. Research portal: https://www.davisgeometric.com/research.

[27] Bee Rosa Davis. Sameness, manifold and operator bridge research notes, 28 October 2026. Author-held review and recorded computational checks. The elementary norm-transport result used here is proved in Chapter 19; no universal constant-two result is inferred from it.

[28] Bee Rosa Davis. *Information Geometry of Semantic Fields*. Author's research manuscript, version examined 9 November 2026. Statistical families, Fisher geometry and compatible representations. Research portal: https://www.davisgeometric.com/research.

[29] Bee Rosa Davis. *The Cohomology of Completion*. Author's research manuscript, version examined 9 November 2026. Local constraints, restriction maps and global completion. Research portal: https://www.davisgeometric.com/research.

[30] Bee Rosa Davis. *The Subgraph Lattice Theorem and Geometric Partition Counting*. Zenodo v1, 2026, Def. 2.2 and Thm. 3.5. https://doi.org/10.5281/zenodo.19503065.

[31] Bee Rosa Davis. *Constructive Metric Bounds for the Crouzeix Conjecture*. Author's research manuscript, version examined 9 November 2026. Finite matrix families, coupled metrics and explicit certificates. Research portal: https://www.davisgeometric.com/research.

[32] Bee Rosa Davis. *The Kähler Substrate, version 0.4*. Author's research manuscript, version examined 9 November 2026. Metric, complex structure and geometric computation. Research portal: https://www.davisgeometric.com/research.

[33] Bee Rosa Davis. *Geodesic Computation, version 7.1*. Author's research manuscript, version examined 9 November 2026. Frame transport and geometric update rules. Research portal: https://www.davisgeometric.com/research.

[34] Xavier Pennec, Pierre Fillard and Nicholas Ayache. "A Riemannian Framework for Tensor Computing." *International Journal of Computer Vision* 66, 41–66, 2006. https://doi.org/10.1007/s11263-005-3222-z. Affine-invariant geometry of positive-definite matrices.

[35] Frank Nielsen. "A Simple Approximation Method for the Fisher–Rao Distance between Multivariate Normal Distributions." *Entropy* 25(4), 654, 2023. https://doi.org/10.3390/e25040654. Open manuscript: https://pmc.ncbi.nlm.nih.gov/articles/PMC10137715/. Univariate normal-family geometry and limits of multivariate closed forms.

[36] M. Ya. Vygodsky. *Spravochnik po vysshei matematike* [Handbook of Higher Mathematics]. Twelfth stereotype edition. Moscow: Nauka, 1977. First published in 1956. Line angles: §21, pp. 34–36; trigonometric orthogonality and Fourier coefficients: §§413–414, pp. 605–609; angle and log-radius differentials: §433, p. 638; path independence and exact differentials: §§475–476, pp. 696–700. Page references use the 1977 Russian edition. Scan: https://ikfia.ysn.ru/wp-content/uploads/2018/01/Vygodskij1977ru.pdf. Original-edition catalog: https://search.rsl.ru/ru/record/01005871902.

[37] Bee Rosa Davis. *Zadachnik*. Ongoing geometry reading, translation and construction notebook. https://www.davisgeometric.com/zadachnik/. The harmonic quotient and annular constructions in this book develop the cited classical calculations in the book's own framework.

[38] Oliver Knill. *The Good Will Hunting Problem*. Harvard Math 21b, updated May 2017. Handout with film frame.

[39] Bee Rosa Davis. *Zero Does Not Exist: A Geometric Foundation for the Natural Numbers*. Zenodo v1, 2026, §§5.2–5.3. https://doi.org/10.5281/zenodo.19477365.

[40] Ada Chan and Chris D. Godsil. Symmetry and Eigenvectors. In *Graph Symmetry*, 75–106, 1997. Author copy, §6.

[41] Frank Simon, Peter Tittmann and Martin Trinks. Counting Connected Set Partitions of Graphs. *Electron. J. Combin.* 18(1), P14, 2011, §§3–4.

[42] Gábor Horváth, Judit Kórándi and Csaba Szabó. Mathematics in Good Will Hunting II. *Teaching Math. Comput. Sci.* 11(1), 3–19, 2013, §2.

[43] Joseph Drapeau et al. Complete Equitable Decompositions. *Linear Algebra Appl.* 701, 112–137, 2024, Def. 2.1.

[44] B. L. Ho and R. E. Kalman. Effective Construction of Linear State-Variable Models from Input/Output Functions. *Regelungstechnik* 14, 545–548, 1966.

[45] *Good Will Hunting*. Directed by Gus Van Sant, written by Matt Damon and Ben Affleck. Miramax, 1997.

Closing reflections

When I started down this road, I did not know whether there was going to be anything at the end of it. I had an intuition—that the two ways I had been building detection systems, one through translators and one through geometry, were secretly the same story—but I did not know whether that intuition would survive contact with the details.

The picture that emerged is more modest and more precise than the intuition I started with:

- Modest, because the equivalence is approximate, scoped, and falsifiable. There are regimes where it holds and regimes where it does not. There is no grand unified manifold that all systems secretly inhabit whether they like it or not.
- Precise, because within its scope it tells you exactly how error budgets move, exactly how much slack you pay, and which geometric, threshold-margin and population quantities you must check.

We started this book in Chapter 1 with a rendering problem: how do we see hidden variables u through noisy observations? How do we decide when two observations show "the same" underlying entity?

We end with a responsibility problem: once we have chosen a notion of sameness—once we have defined benign evolution and drawn our equivalence classes—how do we keep our promises about that sameness when the coordinate systems change?

The geometry of sameness is the thread connecting these: it says that if we are careful about how we stitch our representations together, the answers to "are these the same?" should not evaporate when we swap encoders, upgrade translators, or move from manifolds to metric quotients. Guarantees should be invariant, up to explicit slack.

If there is one message I hope you take from this book, it is not "everything is geometry" or "everything is category theory." It is something quieter:

> ***When we talk about sameness in complex systems—when we link records, reconcile entities, follow identities across formats—we are making claims that can and should be audited across representations.***

The geometry of sameness is one way of doing that audit. It gives us a language for saying: *these two views of the world agree, up to this much slack, on this particular question.* It insists that guarantees should not evaporate when we change coordinate systems.

There is a more personal undercurrent to all of this, too. When I was young, sitting in the library with more questions than explanations, I had a sense that there were hidden shapes behind the things I was reading—shapes that kept showing up in different guises, even when the notation changed. I did not have words for that feeling then. I just knew that something was the same underneath.

I developed work across AMC, ICARUS, GIGI, HERALD and PRISM while navigating my own questions about identity and sameness. What does it mean to be "the same person across time, across contexts, across the ways others see you? What does it mean when a system decides two versions of you are—or are not—equivalent?

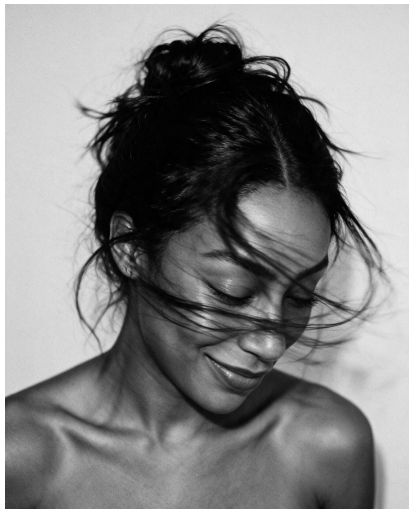
The math cannot answer those questions. But it can insist that the systems be honest about their assumptions, testable in their claims, and accountable when the representations change. That feels like something.

If you are building systems that decide when two things are "the same," I hope you will use these tools—to design better representations, to test your invariances, to argue for your guarantees honestly. If you are a mathematician, I hope some of the open questions here tempt you. If you are a student, I hope you feel, in your own way, what that kid in the library felt: that there are structures under the surface, and that you are allowed to touch them.

The geometry of sameness is not the only way to think about these problems, but it is one way that works, and now it has equations, diagnostics, and worked examples. What you do with it is up to you.

Thank you for walking through this with me.

About the Author



Bee Rosa Davis (she/her) is a Trinidadian–American cybersecurity and AI executive, mission control engineer, mathematician, novelist, and mother of three. For nearly three decades she has lived at the intersection of logic, large-scale systems, and human vulnerability—designing infrastructures that keep people safe in places where failure is not an option.

Bee's technical roots trace back to a B.A. in Logic and Analytics from Morehouse College and an M.S. in Digital Forensics, Counterintelligence, and CyberSecurity from Brown University. That double formation—formal logic on one side, forensic investigation on the other—runs straight through this book: proofs and error budgets live next to audits, incident reports, and logs.

She began her career on the sharp edge of production systems: first at IBM, then at E*TRADE in the early days of online finance, and later as one of the early engineers at Pandora, when the company was still a 50-person startup trying to make sense of music, attention, and ads at scale. At Pandora she helped architect the Machine Learning Digital Engagement Platform that powered hyper-personalized ad targeting and scaled to millions of listeners, an experience that taught her how fragile "sameness" is when money, behavior, and identity are all entangled.

From there, Bee's work moved steadily toward national-scale systems and safety-critical environments. She led anomaly detection and threat intelligence projects across U.S. defense and intelligence agencies, including work at the NSA, DISA, and naval and forward-base operations through Axiom88, designing systems like *Arc Angel* and later KRAKEN for real-time threat detection in the physical world.

At NASA, Bee served as a principal software and mission control engineer within the flight operations ecosystem, contributing to *Project Luna*, a mission control simulation effort that builds high-fidelity virtual test environments for aerospace hardware—from fuses to full launch stacks. She also helped design secure data pipelines for the James Webb Space Telescope, making sure terabytes of astrophysical data could move safely and reliably from deep space to human eyes. Those roles—where simulation, telemetry, and lives-on-the-line reliability meet—shaped the way she thinks about geometry, drift, and error budgets more than any math seminar ever could.

Along the way, Bee has led security and AI programs across healthcare, finance, education, and enterprise: Director of Security Engineering at Humana, CISO and Chief Privacy Officer roles in health analytics and ed-tech, Head of Information Security at Eaze, and Principal Adversarial Intelligence Engineer in IBM's X-Force Red. Her work ranges from quantum-resistant encryption pipelines and bias-aware data equity architectures to red-team platforms for modeling advanced adversaries.

Bee's research sits where this book lives: on the seam between representation, geometry and guarantees. Her work includes the Davis Manifold framework, the Field Equations of Semantic Coherence research program, AMC's telemetry representations, ICARUS's geometric control models, and GIGI's executable treatments of records and inference. HERALD, PRISM, KRAKEN and her work on algorithmic proximate harm explore related questions in their respective settings. *The Geometry of Sameness* develops the mathematical relationship between translator and geometric descriptions, with constructions, assumptions and experiments stated at their own level.

She also writes stories. Bee is the author of the queer sci-fi duology *Fly Into the Sun* and *A Pale Jewel*, both available on Amazon, which follow characters navigating identity, love, and danger in futures that feel uncomfortably close to the present. Her fiction, like her technical work, is obsessed with what it means to remain yourself while the world insists on misreading you.

The through-line underneath all of this is care. Bee is a Black trans woman, a lesbian, and a proud Trinidadian and American who wants her country—and its technologies—to live up to their promises. She is a sharp advocate for algorithmic protections for marginalized people, including in places as intimate and under-regulated as dating apps, where misclassification can translate into very real harm. Much of her work on algorithmic harm, proximate risk, and replay-exact audits is aimed at giving survivors and regulators something stronger than vibes: a mathematical, traceable story of what went wrong and who is responsible.

The most satisfying and demanding job of her life, though, has never appeared on a résumé: raising three children she had in her early twenties. They arrived exactly when the world might have expected wild, drifting years, and instead forced her to grow up early, ask better questions, and build systems—emotional, practical, and technical—that could actually hold weight. The way she thinks about benign evolution and identity over time owes as much to watching kids become themselves as it does to any geometry textbook.

Bee still writes code, still reads papers late at night, and still feels most at home in a quiet library, surrounded by questions bigger than she is. She publishes research and tools under the handle `@nurdymuny` and keeps trying to build systems—and stories—where people can be safely, stubbornly the same person across time, context, and representation.

If this book helps you see your own systems, or your own life, as a little more chartable and a little more honest about what it means to say "these two things are the same," then it has done the job she hoped it would do.

Index

abstention, 317, 526
admissible path, 526
admissible paths, 21, 71, 72, 74, 115, 120, 121, 124, 240, 432
affine-invariant metric, 526
AMC, 6, 526
applications, 352
approximate inverse, 193
atlas, 160, 526
  gluing construction, 168
audit, 185, 330, 341, 343, 344, 431

basepoint, 203, 526
benign evolution, 526
bi-lipschitz bound, 526
boundary mass, 526

cache
  dependency, 373
calibration, 526
  finite audit, 326
Cantelli inequality, 345
category, 187, 526
causal state, 526
chart, 6, 24, 129, 147, 151, 152, 331, 356, 358, 362, 420, 526
  conditioning, 33
  numerical conditioning, 465
  overlap, 32
chartability, 138, 153, 526
chordal distance, 526
Christoffel symbols, 83, 526
clock synchronization, 138
coarse graining, 388
cocycle condition, 123, 124, 153, 334, 526
cohomology, 138, 527
completion disk, 402
complex structure, 433, 527
condition number, 527
conditioning, 290
configuration, 527
congruence, 527
connected partition, 10, 527
connection, 256, 527
constraint rank, 433
covariance
  coordinate transformation, 160
coverage, 317, 527
covering map, 527
  evenly covered neighborhood, 232
Crouzeix research
  metric certificate, 306
curvature, 89, 527
  circumference, 90
  scale of approximation, 89
cylinder, 173

Davis manifold, 281, 527
Davis system, 281, 527
  construction, 282
decision guarantee
  proof, 326
deck transformation, 232, 527
DHOOM, 527
diffeomorphism, 527
directed path cost, 527
distortion, 332, 527
  composition, 16, 192
double cover, 220, 527
  measurement, 107
  square root, 406
dual coordinates, 59, 527

encoder, 528
equivalence relation, 2, 9, 528
error budget, 123, 124, 311–314, 333, 340, 356, 359, 362, 420
  composition, 193
error budget transfer, 311
evenly covered neighborhood, 528
evidence
  types, 364
exponential family, 59, 528
exponential map, 74, 528

fairness, 428, 430
fiber, 528
Field Equations, 419
fingerprint, 528
Fisher information, 48, 435, 528

Fisher metric
local divergence, 388
normal family, 65
Fisher score, 528
fixed-point quotient, 173
fourier mode, 528
Fourier modes
half-turn symmetry, 226
Frobenius norm, 268
Fréchet mean, 251, 528
functor, 164, 179, 192, 528

gauge freedom, 154, 528
Gaussian model, 65
generalized eigenvalue, 528
generalized eigenvalues, 290
geodesic, 47, 71, 72, 528
sphere calculation, 83
geodesic equation, 74
geometric realization, 528
GIGI, 528
inference, 380
preparation and query, 469
records, 368
gluing
manifold, 167
Good Will Hunting, 409

harmonic conjugate, 528
global obstruction, 406
Hausdorff condition, 168, 528
held-out validation, 469
Hellinger distance, 48, 529
hidden Markov model, 195
hidden-state belief, 529
histogram
sensitivity, 402
holonomy, 256, 529
flat connection, 256
Hopf fibration, 241, 529
hyperbolic geometry, 65

ICARUS, 529
identity, 360, 432, 529
versus similarity, 9
immersion, 529
implementation
geometric comparison, 464
in-regime region, 529
incidence matrix, 138, 529
information geometry, 435
information loss, 99
invariant, 9, 204, 420, 529
inverse map
derivative certificate, 336

Jacobi field, 90, 529
Jacobian, 335, 336, 529

Kahler geometry, 433
Kullback–Leibler divergence, 388, 529

laboratories
decision margins, 323
double cover, 247
GIGI completion, 415
metric transformation, 52
overlap consistency, 144
stereographic atlas, 39
least squares
registration, 154
Legendre transform, 59, 529
lift, 529
lineage, 2, 529
linkage condition, 282
lipschitz bound, 529
logarithm
covering space, 406
Lyapunov metric, 268, 529

manifold, 31, 45–47, 529
maps and globe, 6
margin, 529
matrix geometric mean, 297, 530
mean
quotient space, 251
metric, 24, 44–46, 71, 121, 122, 125, 126, 147, 185, 213, 353
coordinate transformation, 16
cross terms, 268
gluing, 168
space of metrics, 297
time dependent, 302
metric comparison, 290
metric compatibility, 302, 530
metric tensor, 44, 48, 530
morphism, 165, 182, 192, 208, 213
composition constants, 193
morphisms, 530

moving frame, 373

nonnormal operator, 268, 530

observability, 107
observable
  descent, 221
operator norm, 268, 530
orthogonal projection, 530
  axis observables, 226

paired experiment, 364
parallel transport, 256, 530
  covariance, 467
  sphere, 256
path horizon, 282
path length
  numerical integration, 466
path lifting, 221, 232
period, 530
  of a differential form, 406
Pick matrix, 530
poincare metric, 530
polar coordinates, 48
polygonal approximation, 130
positive-definite matrix, 297
posterior correctness, 530
predictive equivalence, 10
probability, 339
probability simplex, 402
projective line, 220, 221
projector mean, 251, 530
pseudometric, 530
pseudometric quotient, 130
pullback
  rank, 100
pullback metric, 99, 530

quadrature, 466
quantile, 530
  empirical versus uniform bound, 333
quantization, 530
quaternion, 238, 239, 530
  interpolation, 241
  sampled history, 468
quotient, 121, 530
  orientation, 220
quotient metric, 187

rank, 531
  local versus global, 107
rank deficiency, 99
realizability, 167, 531
relative distortion, 531
reproducibility, 344
retraction, 74, 531
rigid frame, 154
rotation, 268
  angular velocity, 468
  multiple planes, 268
round trip, 123, 124, 217, 333, 531

S^3, 239
saturation, 531
Schur complement, 402, 531
second countability, 531
section, 531
selective risk, 317, 531
semantic sameness, 2
singular value, 531
singular values, 100
smoothness, 531
$SO(3)$, 239
stein inequality, 531
stereographic projection, 32, 33
sufficient statistic, 48, 531

threshold, 313, 315, 339
  boundary mass, 311
threshold band, 326
tolerance
  nontransitivity, 2
translation and distance, 16
  complete circle construction, 187
translator, 20, 21, 23, 26, 27, 114, 115, 118, 120–122, 125, 126, 147, 160, 164, 179, 185, 205, 208, 342, 343, 354, 355, 362, 531
translator metric quotient, 114
transport
  matrix equation, 302

uncertainty ellipse, 467
union bound, 345
update commutator, 195, 531

winding number, 173, 406, 531
witness, 531

zero-failure bound, 345