GIGI: A Fiber-Bundle Database

Mutation-Consistent Geometric Inference and Bounded Analytic Completion

Bee Rosa Davis
Davis Geometric

October 29, 2026
Author-review draft, version 0.1

Abstract

Geometric inference over a database is only as reliable as its connection to the records it describes. A correct algorithm can still produce an invalid answer when its population is inconsistent, its cached model is stale, or its regularization is presented as measurement. This paper describes three implemented upgrades to GIGI, a fiber-bundle database: mutation-consistent statistical summaries; bounded holomorphic interpolation with admissible-value sets; and recovery of a harmonic conjugate from an explicitly represented real polynomial. The latter yields one complex object containing both harmonic components, a domain-wide magnitude bound, and analytic derivatives. Classical centered-moment identities, Nevanlinna–Pick interpolation, Schur recursion, and harmonic conjugacy supply the mathematical foundation. The contribution is their integration with explicit measurement contracts, record provenance, numerical refusals, and mutation-aware database behavior. A shared evaluation uses 16,065 public observations from two domains. Interpolation contains all 408 withheld oracle values, with median admissible radius 0.03267 against an unconstrained unit radius. Conjugate recovery agrees at 816 unseen evaluation points, with maximum normalized value-and-derivative error 2.28×10^{-14}. These are analytic distribution-fingerprint experiments, not claims that raw physical data are holomorphic. Mechanism-removal tests, independent oracles, and live restart controls establish scoped acceptance. The result is a database that can return a constrained mathematical object and the limits of its inference, rather than only a point estimate.

Keywords: fiber-bundle databases; online statistics; bounded holomorphic functions; harmonic conjugates; geometric inference; numerical provenance.

1 Introduction

A database answers questions about stored observations. A model answers questions about an assumed class of possible worlds. Combining the two creates a useful but demanding obligation: every inferred object must remain attached to the observations, assumptions, and revision that justify it.

GIGI organizes record identity and record value separately, using the language of a base space and its fibers. This supplies a common interface for storage, statistics, and geometric operators. The present work develops a specific consequence of that architecture: a collection can support an executable mathematical object whose validity conditions travel with its answer. The object is not a substitute for the records. It is a query-specific representation of what those records permit one to calculate.

Consider three failures that a plausible number can conceal. First, a record is corrected, but an old covariance is reused because the row count did not change. Second, a fit combines coordinate means computed from different missing-data populations. Third, a solver returns one smooth completion without revealing that many other completions are possible. The first is a storage-consistency error, the second an estimator-definition error, and the third an inference-contract error. All three matter to downstream reasoning.

1.1 Contributions and scope

This manuscript reports an implemented, locally validated research snapshot:

1. **Live-state numerical integrity.** Scalar statistics follow current records. Full covariance uses a single complete-case population. Incremental moment reuse is bounded, and cross-bundle caches distinguish both revision and bundle lifetime. Measured covariance is separated from solver adjustments.
2. **Bounded completion with retained alternatives.** Finite complex observations on the unit disk produce either a bounded analytic representation and marginal admissible-value disks, or a precise infeasible or unresolved result.
3. **A unified bounded harmonic pair.** A complete real polynomial is checked for harmonic compatibility. When compatible and sufficiently bounded, one returned holomorphic polynomial contains the real component, the recovered imaginary component, and two complex derivatives.
4. **Evidence at the database boundary.** General-purpose code is tested against independent numerical oracles, planted contradictions, destructive mechanism tests, mutation sequences, and freshly built live engines.

No new version of Welford's algorithm or the Pick theorem is claimed. Nor do these experiments establish a general replacement for neural training. The scientific contribution is the implemented composition of these methods, its explicitly limited contracts, and its falsifiable evaluation.

Snapshot status. The upgrade code is uncommitted and not deployed as of this manuscript's snapshot. The base repository commit is `bd47a0b6745ccb2f455c92310c28e7068ca61014`; that commit alone does not contain the reported changes. A source-and-evidence hash manifest accompanies the manuscript. Existing production services were not restarted for the gates. The paper is an author-review draft, not a claim of peer review or release certification.

2 From Records to Geometric Objects

2.1 Bundle semantics without an implicit smoothness assumption

Let B ("bee") be a set of keys and F_b ("eff sub bee") the allowed typed values at key b. Define the total space E ("ee") and projection π ("pie") by

$$E = \coprod_{b \in B} F_b, \qquad \pi : E \longrightarrow B. \tag{1}$$

A live collection defines a section σ ("SIG-muh") over its occupied keys: $\pi \circ \sigma$ is the identity there. More precisely, an individual record is a point $\sigma(b)$, or the section restricted to a singleton, rather

than a section over every possible key. Creating a record extends the occupied domain; replacing a record changes one section value; deleting it restricts the domain.

This formulation is a discrete family of typed fibers. With a discrete base and common fiber type it has the elementary product-bundle interpretation. It does not by itself endow arbitrary records with a smooth atlas, Riemannian metric, connection, or complex structure. A geometric verb must state which additional structure it constructs or accepts. The analytic verbs in this paper use an explicit complex coordinate $z = x + iy$ ("zee equals ex plus eye why"), with the ordinary complex structure on the unit disk.

2.2 Three different meanings of a geometric summary

It is important not to conflate the following objects:

1. *Measured summaries*: counts, means, centered moments, extrema, and spectra of a specified empirical covariance.
2. *Constructed models*: a regularized Gaussian metric, a chosen analytic interpolant, or a polynomial representation.
3. *Evidence about a model*: excluded counts, residuals, tolerances, admissible sets, source keys, and the revision used for the calculation.

For example, GIGI's implemented scalar dispersion statistic has the form

$$K_{\text{disp}} = \frac{1}{m} \sum_{j=1}^{m} \frac{\text{Var}(f_j)}{R_j^2}. \tag{2}$$

Here m is the number of eligible numerical fields and R_j is an observed range with a numerical floor. This statistic is not, merely by its name, the scalar curvature tensor contraction of a smooth metric. Likewise the ordinary response score $1/(1 + K_{\text{disp}})$ is not a calibrated probability that a record is true. The measured/model/evidence distinction makes such interpretations auditable.

2.3 Position in the Davis research program

Davis's *Field Equations of Semantic Coherence* develops a wider constraint-based geometric research program [9]. Its own mathematical status notice identifies its proposed relationships as conjectures under development. The relationship $C = \tau/K$ ("see equals tau over kay") motivates the question of how much completion is justified by constraints and geometric complexity. This paper neither assumes that relationship as a universal theorem nor identifies it with Equation (2)'s response score. Instead, it isolates an analytic class in which feasible completions can be characterized by established mathematics and checked in executable code.

3 Mutation-Consistent Statistical Geometry

3.1 The population is part of the estimator

For an ordered coordinate list J, let I_J be the live records with a finite numeric measurement in every requested coordinate. A full fit uses precisely I_J for both its mean and covariance. Missing coordinates exclude a record and increment a disclosed exclusion count; nonfinite numeric

measurements produce an actionable refusal. Fewer than two complete records cannot support the implemented sample covariance.

This rule differs from computing each mean on all available values of its field and then computing cross-products on complete rows. That mixed rule has no single empirical population. Complete-case consistency does not, however, establish that missingness is random or that the selected rows are representative.

3.2 Centered moments and incremental maintenance

Write n ("en") for complete-record count, μ ("myoo") for mean, and S ("ess") for centered scatter. For a new complete vector x, multivariate Welford accumulation uses

$$n' = n + 1, \qquad \delta = x - \mu, \tag{3}$$

$$\mu' = \mu + \frac{\delta}{n'}, \qquad S' = S + \frac{n}{n'}\delta\delta^\top. \tag{4}$$

The first observation initializes the mean and zero scatter. The outer product $\delta\delta^\top$ is read "delta times delta transpose." The sample covariance is $\widehat{\Sigma} = S/(n-1)$ ("Sigma hat"), whereas ordinary per-field population variance uses its centered sum divided by its own count. Centered updates avoid the raw-moment subtraction of two large, nearly equal numbers. These are established methods [1, 2], not GIGI-specific statistical inventions.

Proposition 3.1 (Consistent append identity)**.** *In exact arithmetic, Equation (4) maintains the mean and scatter of exactly the accepted complete vectors. An incomplete appended record changes neither quantity.*

Proof. If μ is the old mean, shifting each old residual by $-\delta/(n+1)$ contributes $n\delta\delta^\top/(n+1)^2$ because the old residuals sum to zero. The new vector's residual is $n\delta/(n+1)$, contributing $n^2\delta\delta^\top/(n+1)^2$. Their sum is $n\delta\delta^\top/(n+1)$, as claimed. Excluding an incomplete record leaves the accepted vector multiset unchanged. □

The implementation lazily caches up to four ordered coordinate sets per bundle, each at most 512 coordinates wide. Wider requests use an uncached path. Appends update warm complete-case moments in $O(d^2)$ arithmetic per accepted d-vector. Replacements and deletions invalidate those moments. Destructive edits currently rebuild scalar summaries, including exact live extrema; they are not advertised as constant-time operations. Extrema cannot in general be reconstructed from a count, mean, and centered sum alone.

3.3 Lifetime and revision

The conceptual identity of a reusable derived object is

$$(\text{bundle lifetime},\ \text{revision},\ \text{ordered fields},\ \text{policy}). \tag{5}$$

A revision distinguishes same-size edits. Lifetime distinguishes dropping a bundle and recreating its name at the same numeric revision. Mmap materialization preserves the source identity; the identity is process-local, not a durable content hash. Correctness still depends on every relevant mutation participating in revision and invalidation discipline. The tests exercise those paths; the key alone is not a proof of the entire engine.

Live-state summaries and ingestion-history anomaly scores also have different populations. An anomaly scored at arrival against an earlier baseline is a historical event, not a fresh score of the current database. The upgrade labels these populations rather than silently changing historical meaning.

3.4 Measured covariance is not regularized covariance

The full fit reports the measured covariance and its eigenvalues before applying the existing diagonal and spectral floors. Let Σ_1 be the diagonal-adjusted matrix, $Q\Lambda Q^\top$ its eigendecomposition, and λ_* the disclosed floor. The effective matrix is

$$\Sigma_{\text{eff}} = Q\,\mathrm{diag}(\max(\lambda_j, \lambda_*))Q^\top, \qquad G = \Sigma_{\text{eff}}^{-1}. \tag{6}$$

Cholesky inversion supplies the precision matrix G ("jee"). It defines the Gaussian quadratic form $(x-\mu)^\top G(x-\mu)$. The current policy retains an absolute floor of 0.03 in input-coordinate units, alongside relative floors. Consequently the solver is unit-dependent; this is disclosed, not corrected by renaming the output. Under an affine coordinate transformation $x' = Ax + b$, the *measured* covariance transforms as $A\widehat{\Sigma}A^\top$. An absolute floor need not obey that transformation.

Metadata names total, accepted, and excluded rows; estimator; coordinate policy; revision; lifetime; and moment-cache work. A legacy field named `eigenvalues_raw` is retained for compatibility but refers to the post-diagonal-floor spectrum. The truly measured spectrum is explicitly named in the added evidence.

4 Bounded Holomorphic Completion

4.1 The exact mathematical problem

Let $\mathbb{D} = \{z \in \mathbb{C} : |z| < 1\}$ ("the unit disk"). Given distinct nodes z_i, observed complex values w_i, and a positive bound M, seek

$$\mathcal{F} = \{f \text{ holomorphic on } \mathbb{D} : \|f\|_\infty \le M,\ f(z_i) = w_i \text{ for all } i\}. \tag{7}$$

The supremum norm is the largest magnitude over the domain, not just over sampled points. Holomorphic means complex differentiable throughout the open domain. Its real and imaginary parts satisfy the Cauchy–Riemann equations and are harmonic [5]. These facts are stronger than a small discrete graph-Laplacian energy.

Normalize $y_i = w_i/M$. The central classical result is the following.

Theorem 4.1 (Nevanlinna–Pick, classical)**.** *For distinct interior nodes, the set in Equation* (7) *is nonempty if and only if the Hermitian matrix*

$$P_{ij} = \frac{1 - y_i\overline{y_j}}{1 - z_i\overline{z_j}} \tag{8}$$

is positive semidefinite.

The bar means complex conjugation; "positive semidefinite" means $v^*Pv \ge 0$ for every complex vector v. The theorem and its generalizations are classical interpolation theory [4]. GIGI applies it to explicitly declared complex measurements; it does not infer that arbitrary data satisfy its hypotheses.

4.2 Construction and admissible values

At a node a with normalized value b, define the disk automorphism ϕ_a ("fie sub ay") and Schur reduction by

$$\phi_a(z) = \frac{z-a}{1-\bar{a}z}, \qquad g(z) = \frac{y(z)-b}{1-\bar{b}y(z)}\frac{1}{\phi_a(z)}. \tag{9}$$

The apparent singularity at a is removable for a compatible interpolant. Reverse the reduction using

$$y(z) = \frac{b+\phi_a(z)g(z)}{1+\bar{b}\phi_a(z)g(z)}. \tag{10}$$

Iterated reductions produce a finite representation with a remaining free Schur function in the nonunique case. The returned representative chooses that free function identically zero. It is a compatible choice, not an estimate known to equal the unknown source function.

At a single query point, composition of the fractional-linear maps gives $(At+B)/(Ct+D)$ for $|t| \le 1$. When $|D|^2 - |C|^2 > 0$, its image is a disk with

$$c = M\frac{B\bar{D} - A\bar{C}}{|D|^2 - |C|^2}, \qquad r = M\frac{|AD-BC|}{|D|^2-|C|^2}. \tag{11}$$

This follows by substituting the inverse fractional-linear map into $|t| \le 1$ and completing the square. Each value in this single-query disk is attainable by a suitable constant free terminal function. The joint set of answers at several queries is *not* generally the Cartesian product of their disks: different marginal choices may require incompatible functions.

Proposition 4.2 (Constraint refinement)**.** *For a fixed analytic class and bound, adding compatible observations cannot enlarge the exact admissible-value set at any query.*

Proof. Additional interpolation equalities intersect $\mathcal{F}$ with additional constraints. The resulting function set is a subset of $\mathcal{F}$. Evaluation at a fixed point preserves set inclusion. □

This is a deterministic information statement. An admissible disk is not a frequentist confidence interval, a posterior distribution, or a probability of truth. Contradictory observations instead make the feasible set empty.

4.3 Numerical contract

The implementation accepts 1–64 observations and at most 256 queries, with coordinate modulus at most $1 - 10^{-8}$. Equal-node equal-value observations are deduplicated; conflicting duplicates are infeasible. Nearly colliding distinct nodes, unsafe denominators, and borderline arithmetic are unresolved. An independent Pick diagnostic uses the real symmetric block representation of the Hermitian matrix; its eigenvalues occur twice. The tolerance is $10^{-10}\max(1, \rho(P))$, where $\rho(P)$ ("rho of pee") is the spectral radius. No ridge term is added to make an infeasible Pick matrix pass.

The Schur construction is independently checked against every original observation. The normalized interpolation-residual threshold is 10^{-8}. Near-unit Schur parameters are not rounded into an assertion of unique completion. The reported states are `feasible_family`, `feasible_unique`, `infeasible`, and `numerically_unresolved`. The latter two publish no function or partial query answer. All states remain finite-precision conclusions, not formal exact-arithmetic certificates.

Dense spectral diagnostics take $O(s^3)$ time and $O(s^2)$ storage for s observations. The Schur representation has $O(s)$ size and evaluation cost per query. "No gradient training" therefore does not mean "no computation."

5 One Object for Bounded Harmonics and Holomorphicity

The second analytic contract solves a different problem. It accepts a *complete real polynomial*, not partial complex samples, and reconstructs its harmonic conjugate. A missing monomial means a zero coefficient; it does not mean an unknown measurement.

5.1 Recovery from real coefficients

Write the supplied field as

$$u(x, y) = \sum_{p+q \le d} c_{pq} x^p y^q. \tag{12}$$

The Laplace operator Δ ("DEL-tuh") is $\Delta u = u_{xx} + u_{yy}$. On a simply connected domain, a harmonic real function admits a harmonic conjugate, unique up to an additive real constant in the imaginary component [5]. Fix that constant by specifying $v(0) = v_0$ ("vee at zero equals vee sub zero").

Proposition 5.1 (Polynomial recovery and compatibility)**.** *For a real polynomial u and fixed v_0, define*

$$a_0 = c_{00} + iv_0, \qquad a_n = c_{n0} - \frac{i}{n} c_{n-1,1} \quad (1 \le n \le d), \qquad f(z) = \sum_{n=0}^{d} a_n z^n. \tag{13}$$

There is a holomorphic polynomial with real part u and imaginary value v_0 at zero if and only if re-expanding $\operatorname{Re} f(x + iy)$ recovers every coefficient of u. When it exists, this polynomial is unique and u is harmonic.

Proof. In degree n, the coefficient of x^n in $\operatorname{Re}(a_n(x + iy)^n)$ is $\operatorname{Re} a_n$, and the coefficient of $x^{n-1}y$ is $-n \operatorname{Im} a_n$. Any possible polynomial must therefore have Equation (13)'s coefficients. Re-expansion checks all remaining monomials, so equality is both necessary and sufficient. A holomorphic polynomial's real part is harmonic by the Cauchy–Riemann equations. Uniqueness follows from the forced coefficients and fixed constant. Conversely, a harmonic real polynomial has a polynomial conjugate: $u_x - iu_y$ is a holomorphic polynomial and can be integrated termwise, with the constants fixed as above. □

The implementation checks *all* real coefficients, including omitted zeros. Reading only the x-axis coefficients would construct a polynomial but would not validate the supplied field. A coefficient-Laplacian residual is reported as a separate diagnostic. Nonzero re-expansion mismatch above $10^{-12} \max(1, \sum |c_{pq}|)$ yields `non_harmonic`; a smaller nonzero mismatch is unresolved. No near-harmonic projection is silently substituted. An exact binary64 match is numerical evidence, not a symbolic proof that rounding played no role.

5.2 The full complex bound is a separate obligation

Harmonicity alone does not imply boundedness of the whole complex function. Even a bounded real harmonic function on a disk need not have a bounded harmonic conjugate. For example, $h(z) = \log((1 + z)/(1 - z))$ is analytic on the disk; $\operatorname{Re}(ih)$ is bounded because the mapped right half-plane has bounded argument, but $\operatorname{Im}(ih)$ diverges as real z approaches 1. Also, a bounded *entire* holomorphic function is constant. The domain and the full-complex bound are essential.

For a polynomial f, the maximum-modulus principle reduces the disk supremum to its boundary [6]. A useful computable enclosure follows.

Proposition 5.2 (Global bound from a mesh and a derivative bound)**.** *Let* $A = \sum_{n=0}^{d} |a_n|$, $L = \sum_{n=1}^{d} n|a_n|$, *and* $m_N = \max_{0 \le k < N} |f(e^{2\pi ik/N})|$. *In exact arithmetic,*

$$\max\{m_N, \max_n |a_n|\} \;\le\; \|f\|_\infty \;\le\; \min\{A,\; m_N + \pi L/N\}. \tag{14}$$

Proof. The triangle inequality gives A. The angular derivative has magnitude at most L. Every boundary angle is within π/N of a mesh point, and integration along the intervening arc bounds the change in f by $\pi L/N$. Sampled magnitudes are lower bounds. Cauchy's coefficient formula gives $|a_n| \le \|f\|_\infty$. Finally apply the maximum-modulus principle. □

Production adds a disclosed floating allowance $128\epsilon(d+1)\max(1, A)$, where ϵ ("EP-sih-lon") is machine epsilon. This allowance is not directed-rounding interval arithmetic. If the triangle bound is inconclusive, the mesh doubles from 64 to 65,536. An upper bound within the requested M permits acceptance; a lower bound above M yields `bound_exceeded`; a remaining gap is unresolved. Sampled boundary values without the global correction are never sufficient.

5.3 Unified output and worked example

The accepted object contains ascending complex coefficients, the domain, bound, and conjugate normalization. Each query yields $f(z)$, $f'(z)$, and $f''(z)$ from the same simultaneous Horner recurrence. Thus $u = \operatorname{Re} f$ and $v = \operatorname{Im} f$ share one analytic representation, rather than two independently fitted surfaces. Both obey $|u|, |v| \le |f| \le M$.

For $u(x, y) = x^2 - y^2$ and $v_0 = 0$, the object is $f(z) = z^2$ and $v(x, y) = 2xy$. At $z = 0.2 + 0.3i$ it returns

$$f(z) = -0.05 + 0.12i, \qquad f'(z) = 0.4 + 0.6i, \qquad f''(z) = 2. \tag{15}$$

Changing the coefficient of y^2 to $+1$ must reject the field, not produce a smoother answer. Changing only v_0 can violate the complex bound while leaving the real polynomial unchanged. Both behaviors are gated.

The supported degree is at most 32, with 1–561 distinct monomials and at most 256 interior queries. Coefficient checking is $O(d^2)$; evaluation of values and both derivatives is $O(d)$ per query. Mesh work is $O(Nd)$ with the stated cap. Polynomial coefficients provide an executable representation, but retained provenance and response evidence still consume storage.

6 Engineering the Mathematical Contract

6.1 Read-only analytic endpoints

Both analytic endpoints are enabled by the `kahler` feature. They read one locked bundle snapshot and return source keys, field mapping, source measurements, count, revision, and process-local lifetime. No result cache is introduced. Existing heap/mmap access patterns are reused.

Contract	Input	Successful result
`holomorphic`	Partial complex observations; magnitude bound	Schur representative and marginal value disks
`harmonic_conjugate`	Complete real polynomial; imaginary center; magnitude bound	One bounded complex polynomial and two derivatives

The route prefix is `/v1/bundles/{name}/brain/`. Malformed or unsupported inputs receive 422 with an actionable reason; missing bundles receive 404. Well-formed but infeasible mathematics receives a result status, not a fabricated value. Unknown request keys are refused on these new endpoints. This stricter rule is not a claim about all legacy APIs.

The analytic measurement boundary requires distinct, unencrypted numeric fiber fields and finite measurements. Nulls, timestamps, ciphertext, and integers not exactly representable as binary64 are refused. It does not silently drop records or fill missing values. Polynomial powers must be nonnegative integers, and duplicate monomials are refused rather than implicitly summed.

6.2 Transport, durability, and provenance

Nested evidence exposed a DHOOM composite-value decoding defect, which the statistics work corrected. Updated JSON and DHOOM consumers preserve the same evidence; consumers with older DHOOM decoders should use JSON. Live restart testing also identified an ordinary HTTP update path bypassing the engine's write-ahead-log wrapper. That path now uses the wrapper.

The wrapper does not itself seal encrypted partial updates. Consequently this HTTP surface refuses patches touching encrypted fibers before journaling or mutation, including its versioned form. Unencrypted fields remain patchable. This is a scoped refusal, not certification of all engine-level encryption or mutation surfaces. The existing ignored tests include defects outside this work.

6.3 Acceptance gates as executable scientific controls

The workflow separates mathematical correctness, boundary correctness, and usefulness on data:

1. Write the specification and predeclare measurable acceptance criteria.
2. Obtain behavioral failures against a compiling stub, then implement.
3. Test planted examples, contradictions, ambiguous numerics, and input refusals.
4. Remove a load-bearing mechanism and require its dedicated gate to fail.
5. Restore it, run full regressions and example targets, then rebuild the server.
6. Verify the linked binary's identity and test a live isolated engine on real data, including mutations and restart.

The tests removed complete-case means, live-summary rebuilds, incremental append reuse, measured-spectrum reporting, metadata, lifetime qualification, WAL routing, composite decoding, and encrypted-update refusal in separate statistical controls. Analytic controls removed Pick negativity enforcement, query-radius restriction, harmonic coefficient validation, and the global boundary correction. Each corresponding gate failed; all mutants were restored.

A particularly discriminating bound fixture is $f(z) = 1 + e^{-i\pi/64}z$ with requested bound 1.9999. Its true boundary maximum is 2, between nodes of the initial 64-point mesh. A sampled-only test can accept it incorrectly. The global derivative correction prevents that error. The cancellation fixture $1 + z - z^2$, whose maximum is $\sqrt{5}$, also proves that the refined bound can accept 2.5 even when the coefficient triangle bound of 3 is inconclusive.

7 Real-Data Evaluation

7.1 Datasets and independent questions

The experiments use all 6,497 red and white Wine Quality records and all 9,568 original Combined Cycle Power Plant rows from UCI [7, 8]. The power workbook's other shuffled sheets are not counted as new evidence. Both sources are licensed CC BY 4.0; archive digests are retained.

The statistical full-fit experiment uses 12 wine coordinates, including quality, and four power input coordinates. The two analytic experiments use *every* numeric measured field: 12 wine and five power, including power output. Quality and power output are measured coordinates here, not prediction targets. Record identities are excluded from numerical field analysis.

The three experiments ask different questions. Statistics tests live-population integrity. Partial interpolation tests admissible-set narrowing. Conjugate recovery tests reconstruction from a complete real representation. Neither analytic experiment measures held-out physical prediction accuracy.

7.2 Experiment A: a changing empirical population

The shared workflow imports 80% of each dataset, warms a full fit, appends the remaining 20%, introduces a planted incomplete outlier, corrects records, deletes extrema, replaces duplicate keys, restores a complete case, and tests snapshot/restart and a same-size mmap-backed change. A same-name recreation control repeats the numeric revision with different values. Every stage compares returned records and numerical evidence against an independent centered two-pass oracle with accurate summation and NumPy eigensolvers.

All 24 stages passed. Maximum scaled discrepancies were below 1.2×10^{-12} against a declared 2×10^{-8} tolerance. A planted incomplete outlier would shift the mixed-population mean by 1.539 standard deviations on wine and 1.045 on power; the corrected complete-case mean moves by zero. These are controlled evaluations of the old formula, not a benchmark against an old deployed server.

Table 1: Structural reuse after append. These counts measure moment work, not total endpoint cost or general throughput.

Collection	Old rows scanned once	Appended rows	Old rows rescanned
Wine	5,197	1,300	0
Power	7,654	1,914	0

Warm moments add work to writes; the gate includes a matched append control without warm summaries. Mmap fits still materialize a temporary heap view and do not retain that moment cache across requests. No throughput claim is made for destructive rebuilds. The numerical evidence supports correct reuse, not free computation.

7.3 A common analytic fingerprint

Each measured scalar field is transformed by the same 32-bin histogram rule. For counts $h_0, \ldots, h_{31}$ summing to N, define

$$p_j = h_j/N, \qquad H(z) = \sum_{j=0}^{31} p_j z^j. \tag{16}$$

For $|z| \leq 1$, $|H(z)| \leq \sum p_j = 1$. This gives a justified bounded analytic representation of the *binned empirical distribution*. Original records, bin edges, counts, and source hashes are retained. The transformation is explicit and lossy: neither the physical field nor the raw measurement process is being declared holomorphic. The binning implementation belongs to the generic probe, not an automatic database feature.

7.4 Experiment B: partial complex observations

Seven interior evaluations per field, 119 in total, are supplied to the interpolator. Twenty-four other points per field are withheld from the endpoint. The independent oracle evaluates the full empirical polynomial and derives admissible disks from a Pick-matrix Schur complement, not production's Schur recursion. The predeclared usefulness criterion requires median radius at most 0.5, with no field dropped from analysis.

Table 2: Partial bounded completion: all 17 measured fields.

Measurement	Result
Withheld oracle values contained	408 / 408
Median marginal radius (unit-bound problem)	0.0326711382
Worst marginal radius	0.3615180139
Maximum independent disk disagreement	1.996×10^{-14}
Maximum normalized interpolation residual	2.270×10^{-16}
Maximum representative error against oracle	0.1792607945

The median radius is 96.73% below the unconstrained unit radius. This is set narrowing, not a 96.73% improvement in predictive accuracy. The nonzero representative error is important: the implementation retains ambiguity instead of presenting its chosen function as recovered truth. Contradiction/restoration, coherent same-size change, and restart controls pass.

7.5 Experiment C: recovery from real coefficients only

For each field, the probe expands the real part of $NH(x+iy) = \sum h_j(x+iy)^j$ into real monomial records. It sends those records, $v_0 = 0$, and bound $1.01N$ to the conjugate endpoint. No imaginary function values are supplied. Using integer counts instead of independently rounded probabilities preserves coefficient compatibility; this representation was specified before the live gate.

An independent direct-power oracle evaluates the complex polynomial and both derivatives at 48 unseen interior points per field. It also checks 2,048 boundary points per field. Production uses simultaneous Horner evaluation, and its bound comes from Equation (14), not the oracle mesh. The normalized error is the complex magnitude of the discrepancy divided by source count, maximized over values and the two derivatives.

Every field passes the required 10^{-8} error tolerance and at least 100-fold imaginary-RMSE improvement over assigning zero to the missing component. That baseline measures whether meaningful quadrature was recovered; it is not a competitive learning baseline. Coefficient reconstruction and normalized coefficient-Laplacian residuals are zero for all fields. Nonharmonic edit/restoration, same-size coherent scaling, imaginary-normalization bound violation, and restart controls all pass.

These empirical generating functions have nonnegative real analytic coefficients, so their real expansions contain redundant information about those coefficients. Recovery is algebraically

Table 3: Recovery from complete real polynomials: all fields retained.

Measurement	Result
Fields accepted at normalized bound 1.01	17 / 17
Unseen values agreeing, with both derivatives checked	816 / 816
Maximum normalized value/derivative error	2.275×10^{-14}
Worst recovered-imaginary RMSE	4.231×10^{-17}
Best zero-imaginary baseline RMSE	0.0052489440
Real monomial terms supplied	3,227
Complex coefficients returned	544

determined, not mysterious creation of unseen measurements. Planted mixed-complex-coefficient tests separately verify the conjugate sign and cases that this empirical family cannot exercise. All empirical bounds use the coefficient-triangle path; the harder cancellation and between-mesh-peak paths are covered by the independent planted fixtures.

7.6 Full field results

Table 4 exposes all fields rather than selected successes. Each radius median is over that field's 24 interpolation queries; the global median in Experiment B is over all 408 queries, not a median of these medians.

Table 4: Field-level evidence. r_{50} and $r_{\max}$ are marginal radii in Experiment B; $e_{\max}$ is normalized value/derivative error in Experiment C.

Domain	Field	r_{50}	$r_{\max}$	$e_{\max}$
Wine	Fixed acidity	0.03714	0.35963	1.12×10^{-14}
	Volatile acidity	0.03288	0.35606	1.42×10^{-15}
	Citric acid	0.03367	0.35445	1.78×10^{-15}
	Residual sugar	0.02624	0.30176	9.93×10^{-16}
	Chlorides	0.02547	0.32708	9.16×10^{-16}
	Free sulfur dioxide	0.02956	0.35152	1.99×10^{-15}
	Total sulfur dioxide	0.03710	0.35999	1.09×10^{-14}
	Density	0.03099	0.35309	2.67×10^{-15}
	pH	0.03798	0.36128	1.52×10^{-14}
	Sulphates	0.03274	0.35540	5.33×10^{-15}
	Alcohol	0.03709	0.35953	3.54×10^{-15}
	Quality	0.03805	0.36121	2.27×10^{-14}
Power	AT	0.03799	0.36133	1.10×10^{-14}
	V	0.03810	0.36148	5.40×10^{-15}
	AP	0.03811	0.36151	1.32×10^{-14}
	RH	0.03811	0.36152	1.50×10^{-14}
	PE	0.03777	0.36094	7.32×10^{-15}

7.7 Failures, numerical disclosures, and regression coverage

An initial live conjugate probe required bit-identical echoed query coordinates. It stopped on a 1.39×10^{-17} JSON round-trip difference. The failure receipt is retained. The harness was corrected to

report and enforce coordinate drift at most 10^{-15}; observed maximum drift was 5.56×10^{-17}. Neither the query set, solver, nor the predeclared 10^{-8} value/derivative threshold changed. The independent oracle still evaluates the originally requested points. This is a disclosed harness correction, not evidence of bit-exact transport.

The final full `kahler` regression run reports 2,092 passes, zero failures, and 23 pre-existing ignored items, including doctests. The default run reports 1,580 passes, zero failures, and 22 ignored items. All three explicit example targets compile and run in each configuration; they define no unit tests. Ignored items are not counted as successes, and other optional feature combinations are not certified. Earlier staged totals are not added to these final totals.

Largest observed debug-build request times were 17.02 ms for seven-observation interpolation and 31.21 ms for polynomial recovery. These are local observations, not throughput, production-latency, or cross-system benchmarks. Independent measurement correctness and the no-rescan receipt are stronger claims than those timings.

8 Compression, Training, and Downstream Reasoning

8.1 What is actually condensed?

"Compression" must specify what information remains recoverable. Three different reductions appear here:

1. A mean and covariance preserve particular second-order calculations, not record identity or arbitrary queries. Under a Gaussian modeling assumption, centered moments determine the fitted quadratic geometry without gradient descent.
2. A finite Schur representation describes a chosen compatible analytic function, while its free terminal and admissible disks disclose remaining freedom. Seven observations do not generally identify the original function.
3. A degree-d complex polynomial has $d+1$ complex coefficients. It encodes its real and imaginary harmonic components and derivatives together. Given the complete compatible real polynomial and normalization, this re-encoding is exact in the mathematical model.

The 3,227-to-544 count compares real monomial terms with complex coefficient entries, not bytes or equal-size scalars. A complex coefficient contains two real numbers, and responses retain provenance. Further, the empirical histogram was already a lossy compression of the source records. No lossless database compression ratio is established by that count.

8.2 Why some operations need no training loop

The methods exploit declared structure. Centered moments admit online algebra; bounded interpolation admits positivity tests and recursive construction; polynomial conjugacy admits coefficient recovery. They do not search a neural parameter space through repeated stochastic-gradient epochs.

This does not mean GIGI performs arbitrary learned tasks without training, nor that all machine learning requires iterative gradient descent. Closed-form and nonparametric learning methods also exist. The comparison is about the specific task and assumptions, not a universal hierarchy of technologies. When the needed representation is supplied, direct mathematical inference can remove an unnecessary optimization loop. Learning the representation, chart, semantic grounding, or noise

model is a separate problem.

8.3 A future Marcella contract

A small language model could use these objects through tools: request a calculation, read the evidence, and verbalize a result with its scope. The integration is not implemented in this work. A viable consumer must distinguish source measurement, chosen representative, admissible set, and refusal. It must not convert "numerically unresolved" into "probably true" or convert a marginal disk into a joint or probabilistic guarantee.

The analytic object can therefore support an external reasoning system, but does not give a language model general understanding or calibrated self-knowledge by itself. End-to-end acceptance would require source-grounded answers, faithful handling of refusals, and task-level evaluation against an appropriate baseline. The deferred index-filtered SEMANTIC request is likewise not delivered by these analytic endpoints.

9 Related Work and Mathematical Position

Online centered moments follow Welford and the multivariate moment literature [1, 2]. Incremental view maintenance has long treated the problem of keeping derived results current under changes; DBToaster develops higher-order delta processing for database views [3]. GIGI's bounded moment cache is a narrower mechanism, not a replacement for general incremental view maintenance or a demonstrated performance improvement over it.

The exact analytic feasibility question is the classical Nevanlinna–Pick problem [4]. Harmonic conjugacy and maximum-modulus reasoning are standard complex analysis [5, 6]. The recovery and global-mesh propositions here specialize those facts to explicit polynomial data and implementation checks. Their derivations make the contract inspectable; they are not presented as priority claims for new complex-analysis theorems.

The bundle viewpoint connects these methods at the level of record identities, typed measurements, and derived structure. It does not imply that all existing GIGI geometric verbs use the same metric, topology, or uncertainty semantics. For instance, the global SEMANTIC metadata describes a constructed two-dimensional complex with modulo-two coefficients, not persistence or semantic truth. A separate persistence verb elsewhere in the repository is not the subject of this evaluation. Broad database and AI comparisons would require new experiments rather than an expanded reading of the present gates.

10 Limitations and Next Research Gates

The main limits follow directly from what supplies the information.

- **Restricted analytic domain.** One scalar complex variable on a unit disk is supported. General manifolds, learned charts, multiply connected domains, and sheaf gluing are not implemented here. On a multiply connected domain, a harmonic field can have nonzero conjugate periods, obstructing a single-valued global conjugate. Local compatibility alone is insufficient.
- **Exact model, numerical computation.** Neither analytic endpoint fits a noisy physical

process. Binary64 diagnostics and an allowance are not directed-rounding certificates. Higher precision and interval enclosures would need dedicated near-boundary, ill-conditioned, and adversarial gates.

- **No automatic statistical validity.** Complete-case selection does not cure selection bias. Admissible sets condition on a declared function class and bound. Calibration against unknown populations remains separate.
- **Limited evaluation family.** The two public datasets exercise general-purpose code, but their analytic fingerprints share one construction. They do not demonstrate domain-independent intelligence, physical forecasting, or every possible complex coefficient pattern. Planted tests supplement, not replace, that limitation.
- **Computational and durability boundaries.** Dense eigensolvers, degree and query caps, destructive rescans, mmap materialization, legacy ignored defects, and decoder compatibility remain relevant. The work does not certify all production feature combinations or write surfaces.

Natural next gates are unit-aware regularization, explicit noisy-observation feasibility, joint multi-query admissibility, directed-rounding bounds, and period-aware chart gluing. Each requires a new mathematical contract, a counterexample that would expose failure, and a consequential real-data use case. More sophisticated mathematics is valuable only when it enlarges a clearly defined capability without obscuring its assumptions.

11 Conclusion

The implemented upgrades connect three levels of responsibility: keeping measured statistics attached to the current records, restricting completion to a declared mathematical class, and returning evidence about what remains undetermined. GIGI can now construct a bounded holomorphic completion from partial complex observations and recover a bounded harmonic pair from a complete compatible real polynomial. Both are database operations with source identity and refusal semantics, not standalone plausible-number generators.

The compelling result is not that geometry eliminates uncertainty. It is that uncertainty, compatibility, and boundedness can become explicit query results. This provides a concrete foundation for richer geometric inference and for downstream systems that must know the difference between a stored fact, a mathematically implied value, and a choice among alternatives.

A Reproducibility and Claim-to-Artifact Map

The source files, specifications, tests, and result reports are local repository artifacts. They must accompany a reproducible release; a link to the base commit alone is insufficient. The companion `verify_paper.py` verifies reported aggregate metrics, the field table, regression totals, and current artifact hashes. It does not replace replay of the acceptance probes.

Claim or mechanism	Repository evidence
Complete-case moments	`src/statistics.rs`; `tests/statistical_summaries.rs`

Claim or mechanism	Repository evidence
Live-state and cache discipline	`src/bundle.rs`, `src/mmap_bundle.rs`, `src/vector_cache.rs`, `src/morse_cache.rs`; statistics result report
Analytic completion	`src/geometry/holomorphic.rs`; `tests/bounded_holomorphic.rs`
Conjugate and full bound	`src/geometry/harmonic_conjugate.rs`; `tests/harmonic_conjugate.rs`
HTTP boundary	`src/bin/gigi_stream_handlers/holomorphic.rs`; `src/bin/gigi_stream_tests/`; `openapi.json`
Experiment A	`e2e/probes/statistical_summaries_real_data.py`; `docs/STATISTICAL_SUMMARIES_RESULTS.md`
Experiment B	`e2e/probes/bounded_holomorphic_real_data.py`; `docs/BOUNDED_HOLOMORPHIC_RESULTS.md`
Experiment C	`e2e/probes/harmonic_conjugate_real_data.py`; `docs/HARMONIC_CONJUGATE_RESULTS.md`
Machine-readable evidence	`output/statistical-summaries/`, `output/bounded-holomorphic/`, `output/harmonic-conjugate/`: reports and gate logs

Replay. From the repository root, use a current toolchain and the declared features:

```
cargo test -j 2
cargo test --features kahler -j 2
cargo test --examples -j 2
cargo test --examples --features kahler -j 2
cargo build --features kahler --bin gigi-stream
python e2e/probes/statistical_summaries_real_data.py \
  --binary target/debug/gigi-stream.exe
python e2e/probes/bounded_holomorphic_real_data.py \
  --binary target/debug/gigi-stream.exe
python e2e/probes/harmonic_conjugate_real_data.py \
  --binary target/debug/gigi-stream.exe
```

The displayed backslash continuations are for POSIX shells; in PowerShell put each Python invocation on one line. Python probes require NumPy and openpyxl. They download pinned public archives and start only their own hidden engine, with a temporary data directory on an available `127.0.0.1` port. Production cloud credentials and GIGI settings are stripped from the child environment only. A test run does not rebuild an already running server; explicit relinking and binary identity checks precede HTTP acceptance.

Pinned inputs.

Wine archive SHA-256:
`3ed56667f4b828242bd732d7d1dd7f2861e54432239d7fa63877014cbb0304d4`

Power archive SHA-256:
`cc7b2a4977c0a44e8221c91d9a7e5746b3c68186cff7e5c61c70af6432b98c7a`

Final conjugate live executable SHA-256:
`a2441626d439879346171ff4f58ab92ef0a730eabf0d1170566809c7875d34c4`

The three experiments have distinct build receipts. The final conjugate executable was linked at 2026-10-30 00:45:21 UTC, still October 29 in the local evaluation time zone. The statistical and interpolation reports retain their own executable hashes. All probe engines were stopped after acceptance.

B Notation and Pronunciation

Symbol	Spoken form	Meaning here
B, F_b, E	bee, eff sub bee, ee	Keys, allowed values at one key, total space
π, σ	pie, SIG-muh	Projection to keys; section assigning records
$\mu, \widehat{\Sigma}$	myoo, Sigma hat	Mean; measured sample covariance
δ, ϵ	DEL-tuh, EP-sih-lon	Centered increment; machine epsilon
$\mathbb{D}, \mathbb{C}, \mathbb{R}$	disk, complex numbers, reals	Domain and number systems
$z = x + iy$	zee equals ex plus eye why	Complex coordinate, with $i^2 = -1$
$\bar{z}$	zee bar; conjugate zee	$x - iy$
ϕ_a	fie sub ay	Disk automorphism centered at a
Δu	Laplacian of you	$u_{xx} + u_{yy}$
$\|f\|_\infty$	supremum norm of eff	Largest magnitude on the domain
u, v	you, vee	Real component; harmonic conjugate
f', f''	eff prime, eff double prime	First and second complex derivatives
τ, K	tau, kay	Tolerance and curvature symbols in Davis's program

C Publication and Companion-Site Boundary

Before public release, the author should approve terminology and authorship, archive the exact source snapshot and receipts, assign a public artifact location, and rerun the declared gates on that released snapshot. This draft does not claim an external mathematical review or public reproducibility archive already exists.

A companion site can demonstrate the paper's central distinction with three views: a correction that invalidates a numerical model; an observation that shrinks an admissible disk or creates a contradiction; and a real polynomial whose harmonic conjugate and complex bound appear in one object. Each view should expose assumptions and evidence next to the result. The all-field table, failure controls, and reproduction receipts belong alongside the headline results. Website deployment and Marcella integration remain separate deliverables; neither was performed for this manuscript.

References

[1] B. P. Welford. Note on a Method for Calculating Corrected Sums of Squares and Products. *Technometrics*, 4(3):419–420, 1962. doi:10.1080/00401706.1962.10490022.

[2] Philippe Pébay. *Formulas for Robust, One-Pass Parallel Computation of Covariances and Arbitrary-Order Statistical Moments*. Sandia National Laboratories, SAND2008-6212, 2008. `https://www.osti.gov/servlets/purl/1028931`.

[3] Yanif Ahmad, Oliver Kennedy, Christoph Koch, and Milos Nikolic. DBToaster: Higher-order Delta Processing for Dynamic, Frequently Fresh Views. *Proceedings of the VLDB Endowment*, 5(10):968–979, 2012. arXiv:1207.0137.

[4] Jim Agler and John E. McCarthy. *Nevanlinna–Pick Kernels and Localization*. Author manuscript, arXiv version deposited 2016; see Theorem 0.1 for the classical disk theorem. https://arxiv.org/abs/1610.01965.

[5] MIT OpenCourseWare. *18.112 Functions of a Complex Variable, Lecture 16: Harmonic Functions*. Fall 2008. Lecture notes.

[6] NIST Digital Library of Mathematical Functions. *Section 1.10(v): Maximum-Modulus Principle*. Accessed October 29, 2026. https://dlmf.nist.gov/1.10#v.

[7] P. Cortez, A. Cerdeira, F. Almeida, T. Matos, and J. Reis. *Wine Quality*. UCI Machine Learning Repository, 2009. CC BY 4.0. doi:10.24432/C56S3T.

[8] Pınar Tüfekci and Heysem Kaya. *Combined Cycle Power Plant*. UCI Machine Learning Repository, 2014. CC BY 4.0. doi:10.24432/C5002N.

[9] Bee Rosa Davis. *The Field Equations of Semantic Coherence: A Geometric Theory of Meaning, Curvature, and Reasoning in Transformer Architectures*. Research manuscript dated November 30, 2025; conjectural status as declared in the manuscript. Repository source: `theory/field_equations_semantic_coherence.tex`.