The Latent: Finite Sufficient Representations of Smooth Systems

Tamás Nagy, Ph.D.

Contents

The Latent: Finite Sufficient Representations of Smooth Systems

Tamás Nagy, Ph.D.

Book Edition 2.0 — April 2026

Preface

This book began as a paper. The original eleven-section monograph appeared on Zenodo in March 2026 (v1). By April it had grown an additional nine application sections, a formal verification scope, ontological chapters, shadow-principle extensions, and a research program — more than a journal paper can carry and more than the monograph's linear format could support. Rather than amputate the material, we restructured.

Book Edition 2.0 is the result. The content is the same as the 2026-04-19 monograph-build plus a clearer organization: seven Parts, eleven chapters, four appendices. The spine (Parts I–VII, Chapters 1–11) carries the core theory, proofs, and empirical cases. The appendices (A–D) hold material that either belongs to a different intellectual register — philosophy of ontology in A, shadows of mind in B — or forms a domain-specific survey — C, D — that a given reader may wish to skip on a first pass.

Nothing is discarded. Every theorem, remark, table, and companion pointer in v1.0-monograph is also here, either on the main path or in one of the four appendices. Section identifiers (§3.6, §4.4, §8.4, and the rest) are preserved wherever the content was not moved, so citations from earlier companion papers continue to resolve. The handful of sections that did move have been renumbered once, globally; the old-id → new-location map sits in `RESTRUCTURE_LOG.md` and at the top of each appendix.

How to read it. A mathematician can follow the spine linearly: definition (Chapter 2), theorem (Chapter 3), methodology (Chapters 4–6), universality (Chapter 7), empirical validation (Chapter 8), verification story (Chapter 9), the Shadow Principle as an independent epistemic fold (Chapter 10), conclusion (Chapter 11). A practitioner looking for concrete Latents in their own domain can enter at Chapter 8 (LLMs, finance, three-body dynamics, Navier–Stokes, black holes, fine-structure constant, the zeta function, Riemannian manifolds) and the Appendix C catalog. A reader curious about the metaphysical scaffolding — the Extractor–Latent–Projector ontology, the ontological inversion, the intelligence and consciousness shadows — should read Appendix A first and re-enter at Chapter 1.

Verification. This book is not a single sealed Lean artifact. It is an exposition backed by the Platonic proof kernel (`elysium/fields/`, 42,515 verified theorems across 330 domains, zero `sorry`) with selective Lean

4 export and a stamped minimum example (`stamp/protein_fold_stamp/EckartYoung.lean`). §9.5 maps every major claim to its kernel location.

The four appendices are also, by design, seeds. Each can be lifted out and developed into a standalone article without touching the spine — A into a philosophy-of-science paper, B into a shadows-of-mind essay, C into a field-guide note, D into a research-program paper. The book stands on its own either way.

— Tamás Nagy, April 2026

Overview (Non-Technical)

How many numbers do you need to completely describe a smooth system — so that every distributional property, every dynamic quantity, every functional is recoverable from those numbers alone? This book proves that the answer is $N = \Theta(\log(1/\varepsilon)/\log\rho)$, where ε is the target accuracy and $\rho > 1$ is a single dimensionless number — the system's intrinsic compressibility — that depends on neither the ambient dimension nor the method used to compute it. We call ρ the **Latent Number** and the finite sufficient description the **Latent**.

The Latent is not a vector of cosine coefficients, not a set of eigenvalues, and not a distribution. It is the basis-free, coordinate-free, method-independent mathematical object that all of these are representations *of*. The book defines this object, proves its existence and size bound, constructs an algebra on Latents (addition, tensor product, contraction, basis change), and validates the theory with two controlled empirical studies.

Validation 1: the Analyticity–Rate Duality. For a 10-asset Gaussian equity portfolio with $\rho \approx 6$, the same analyticity parameter that governs spectral convergence of the density also governs rare-event importance sampling — and a mode-factored IS estimator achieves a variance-reduction ratio of 1.7×10^4 against naive Monte Carlo (§3.6). At $\rho \to 1$ (Student-t tails), the same algorithm fails in the predicted form. The duality between approximation and simulation is invisible in either literature alone; the Latent framework makes it a theorem.

Validation 2: higher-grade extraction in the three-body problem. The classical three-body problem is the sharpest test of the higher-grade theory. We extract the Latent for four orbit families and show that kinematic rank is universally 4 (confirming the rank bound), that the grade-3 co-skewness tensor is 30–600× larger for chaotic orbits than for periodic ones (a fingerprint of non-integrability), and that the optimal-basis gap between Fourier and Padé extraction is directly measurable (§8.4). All results are reproducible from the repository code.

The framework further applies to large language models, Navier–Stokes turbulence, Riemannian manifolds, and several other domains developed in companion papers. Every existing result in the spectral methods literature — the COS expansion, eigenvalue conditioning, the Fokker–Planck generator, Knowledge Artifacts — turns out to be a coordinate representation of a Latent in a specific basis. The Latent is the basis-free reality behind all of them.

Abstract

We define the **Latent** of a smooth system as the basis-free element of a graded Hilbert tensor algebra that completely characterizes the system's distributional, dynamic, and functional properties. A single dimensionless number ρ — the system's **Latent Number** — measures its intrinsic compressibility.

Main theorems. The **Latent Theorem** (Theorem 1, §3.2) proves that every Latent with $\rho > 1$ in some extractable basis admits a finite approximation of size $N = \Theta(\log(1/\varepsilon)/\log\rho)$ per mode, independent of ambient dimension and basis once an effective-rank reduction is fixed. The **Extended Latent Theorem** (Theorem 2, §3.3) extends this to non-smooth systems via sufficient smooth representations. The **Rational Latent Theorem** (Theorem 8, §4) predicts Padé convergence rates with the pole basis as constructive optimum. All three are basis-free: the COS expansion, eigenvalue conditioning, the Fokker–Planck generator, and Knowledge Artifacts are recovered as coordinate representations of the same underlying Latent in specific bases.

Analyticity–Rate Duality and controlled benchmark. The Bernstein-ellipse parameter that governs spectral convergence is the *same* ρ that governs the Cramér rate of importance sampling. We validate this duality on a controlled Gaussian-portfolio benchmark (§3.6): at $\rho \approx 6$ we measure a variance-reduction ratio of 1.7×10^4 against naive Monte Carlo (95% CI $[1.3, 2.1] \times 10^4$), and at $\rho \to 1$ (Student-t tails) we observe IS failure in the predicted form.

Higher-grade extraction. The three-body problem (§8.4) demonstrates the grade hierarchy beyond pairwise methods: grade-3 co-skewness is 30–600× larger for chaotic orbits than periodic ones, the kinematic rank is universally 4 (confirming the rank bound), and the Fourier-vs-Padé extraction gap is directly measurable. The parametric map from initial conditions to Latent is itself a smooth 4-dimensional manifold — the meta-Latent — yielding a finite representation of orbit families, not just individual orbits.

Algebra and structural consequences. The basis-free formulation makes visible: (i) an **optimal extraction problem** — the basis maximizing ρ minimizes coordinate count; (ii) **higher-order entanglement** — grade-r Latents for $r \geq 3$ carry irreducible multi-body interactions no pairwise decomposition can represent; (iii) **Latent transfer** — a Latent extracted in one basis can be re-expressed in another without re-extraction. The **Latent Algebra** (Proposition 7, §6) is the universal graded contraction algebra generated by the underlying Hilbert space.

Further applications. The framework is applied to large language models (grade-1 Latent of GPT-2 hidden states: effective rank 2–39 out of 768, surviving model distillation; §8.1), Navier–Stokes turbulence (vortices as canonical Latents; §8.5), and closed Riemannian manifolds (eigenvalues + trilinear structure constants as a proposed complete geometric invariant within the bounded-geometry class; §8.9). Extended applications — black holes, fundamental constants, the Riemann zeta function, game theory, cosmology — are developed in companion papers and appendices under explicitly stated additional assumptions; they illustrate the framework's range but are not part of the core evidence package.

Verification scope. Many algebraic and analytic pipelines are checked in the Python-native proof kernel (`elysium/fields/`), with selective Lean 4 export. This exposition is not a single sealed Lean artifact; §9.5 gives a section-to-kernel formalization map.

Claim tiers. This book uses three evidentiary tiers. **Proved here**: the Latent theorem package, algebra, Analyticity–Rate Duality benchmark, and three-body extraction findings. **Companion-backed**: LLM extended experiments, manifold reconstruction chain, deeper three-body solution program. **Programmatic / conditional**: fundamental constants derivation, Riemann Hypothesis connection, black-hole extrapolations, and philosophical reflections — stated under explicit assumptions and kept outside the core evidentiary spine. §1.5 gives the full evidence map.

Part I — Foundations

The first Part isolates the question. Chapter 1 asks why complex systems so often admit compressed exact descriptions, places that question in the classical spectral literature, and walks through four physical systems — projectile, coupled oscillators, trebuchet, Navier–Stokes — to make the ontology concrete. Chapter 2 then delivers the formal definition: the Latent is the basis-free element of a graded Hilbert tensor algebra.

Chapter 1. Introduction

1.1 The Question

How many numbers do you need to completely describe a smooth system?

Not approximately. Not for one specific purpose. **Completely** — so that every distributional property, every dynamic quantity, every functional of the system is recoverable from those numbers alone.

The answer, for a broad class of systems encountered in finance, physics, biology, and machine learning, is: $N = \Theta(\log(1/\varepsilon)/\log\rho)$, where ε is the target accuracy and $\rho > 1$ is the **Latent Number** of the system — a basis-free, dimensionless measure of the system's intrinsic compressibility. This number is independent of the system's ambient dimension and independent of the method used to compute it. The Latent Number plays for computational science the role that the Reynolds number plays for fluid dynamics: a single number that determines the system's regime. The sharp phase transition at $\rho = 1$ separates exponentially efficient systems from fundamentally intractable ones.

We call this finite sufficient description the **Latent** of the system.

> **Main Result (informal).** Every smooth system with analyticity parameter $\rho > 1$ admits a basis-free finite sufficient representation — its Latent — of size $N = \Theta(\log(1/\varepsilon)/\log\rho)$ per mode. The representation lives in a graded Hilbert tensor algebra, carries a natural algebraic structure (addition, tensor product, contraction, basis change), and is constructively extractable. The bound is tight: at $\rho = 1$ a sharp phase transition makes finite representation impossible. The formal statement is Theorem 1 (§3.2); the algebra is Proposition 7 (§6.3); the optimal extraction is Theorem 4 (§4.4).

1.2 What the Latent Is

The Latent is an abstract mathematical object — an element of a graded tensor algebra over a Hilbert space — that encodes everything about a smooth system. It is not a vector of cosine coefficients (though cosine coefficients are one set of coordinates for it). It is not a set of eigenvalues (though eigenvalues are one way to

extract part of it). It is the basis-free, coordinate-free, method-independent thing that all these representations are representations *of*.

The analogy is exact: a geometric vector **v** exists independently of coordinates. Its Cartesian components (v_x, v_y, v_z) are one representation. Its spherical components (v_r, v_θ, v_ϕ) are another. The vector is the same regardless. The Latent is the same: COS coefficients are Cartesian; Hermite coefficients are spherical; the Latent is the vector.

Why the name. The word "latent" means *hidden but present*. It also means *existing but not yet manifest*, and *discoverable through careful observation*. The name is not a metaphor — it is a compressed description of the entire framework in a single word. The Latent is hidden (behind Reality), present (it generates every phenomenon), and discoverable (through Extraction). A good name for a mathematical object is itself a Latent of the concept: a finite representation — one or two words — that encodes as many aspects of the structure as possible. The obsessive search for the right name is not aesthetic; it is the same compression instinct that drives the framework itself.

Note on section numbering. In the v1.0-monograph, this chapter continued with §1.3 (the seven-element ontology *Reality / Latent / Manifestor / Lens / Observable / Extractor / Projector*) and §1.4 (the ontological inversion). Both have been moved to **Appendix A** (§A.1, §A.2) to separate the philosophical scaffolding from the mathematical spine. Readers who prefer the ontology before the formal definitions should read A.1–A.2 now; readers preferring the math path can continue to §1.5. The subsection IDs below (§1.5, §1.6, §1.7, §1.8) are retained from v1 so that external citations of v1.0-monograph continue to resolve; see `RE-STRUCTURE_LOG.md` for the full v1 → v2 map.

1.5 Why the Latent Matters

Every spectral method in science chooses a basis — cosines, wavelets, Hermite functions, spherical harmonics — and then works in that basis. This choice determines convergence rates, computational costs, and what you can express. The COS method (Fang and Oosterlee 2008) works brilliantly for bounded-domain pricing. Hermite methods work better for Gaussian tails. Wavelets handle multi-scale features. Each is a lens; each distorts what it cannot see well.

The Latent framework removes the lens. It identifies the object that all lenses are looking at and proves three things about it:

I. Optimal extraction. Different bases yield the same Latent but with different coordinate efficiency. The basis maximizing the analyticity parameter ρ minimizes the number of coordinates needed. Basis selection becomes an optimization problem with a clear objective function, not a modeling assumption.

II. Higher-order entanglement. The Latent has a grade (interaction order). Current spectral methods operate at grades 1 (distributions) and 2 (generators, correlations). Grade 3 and above capture irreducible multi-body interactions — three-asset crash clustering, multi-model knowledge interference, attention head composition in transformers — that no pairwise framework can represent.

III. Latent transfer. A Latent extracted in one basis can be converted to any other basis without re-extracting from the original system. This makes representations interoperable: extract once, use everywhere.

What is NOT new. The bound $N = \Theta(\log(1/\varepsilon)/\log\rho)$ is classical — it is a Kolmogorov n-width statement (Pinkus 1985). The exponential convergence of spectral expansions for analytic functions has been known since Bernstein. The Latent Theorem (Theorem 1) is not a new approximation result. What IS new is:

- (a) The **object**: the graded Hilbert tensor element $\Lambda \in \bigoplus_r \mathcal{H}^{\otimes r}$ that the bound applies to. This

object — the Latent — is basis-free and grade-decomposed. It is not a function, not a vector of coefficients, and not a distribution. It is the thing that all of these are coordinate representations of.

- (b) The **algebra**: addition, tensor product, contraction, and basis change on Latents, packaged as the initial graded contraction algebra generated by $\mathcal{H}$ (Proposition 7). Extended with **maps** (grade-preserving endomorphisms, §6.5), **complexification** ($\mathbb{R} \to \mathbb{C}$ extension enabling Fourier theory and monodromy, §6.6), and the **evaluator layer** (formal separation of exact algebra from convergent/divergent numerical realization, §6.7). This turns a collection of isolated approximation results into a single algebraic theory where every computational pipeline is a composition of algebraic operations and every numerical method is an evaluator for an exact operation.
- (c) The **recursion**: the Latent Theorem applies to *families of Latents*. The parametric map IC $\mapsto \Lambda$ (initial conditions to Latent) is itself smooth, so it has a meta-Latent. This gives finite representations of solution families, not just individual solutions — a qualitatively new capability that no single approximation theorem provides.
- (d) The **optimal construction**: the Representation-Structure Matching Theorem (Theorem 4) determines which basis achieves the Latent Theorem's bound — the one matching the system's generative structure. The Grade-3 Universal Termination Theorem (Theorem 4b) proves the hierarchy terminates: no Grade-4 meta-level exists for either distributions or dynamical systems. The Constructive B* Corollary shows the matched basis is discoverable from data at rate $O(1/\sqrt{N})$ via spectral estimation. This completes the framework: existence (Theorem 1), algebra (Proposition 7), optimal basis selection (Theorem 4), termination (Theorem 4b), and constructive discovery (Corollary).

The contribution of this paper is (a)+(b)+(c)+(d), not the bound.

1.5a Evidence Map and Claim Tiers

This book makes claims of very different evidentiary weight. Rather than let the reader discover this implicitly, we separate them upfront into three tiers:

Tier	Meaning	What belongs here
Proved here	Theorem, proof, and/or controlled benchmark are in *this book*. The book speaks in full declarative voice.	Latent Theorem package (Theorems 1, 2, 8), Latent Algebra (Proposition 7), Analyticity–Rate Duality benchmark (§3.6), three-body extraction findings (§8.4), Shadow Principle (Theorem 9).
Companion-backed	A specific companion paper develops the result; this book summarizes it and points to the artifact path.	LLM extended experiments (§8.1 → companion), manifold reconstruction chain (§8.9 → companion), three-body Galerkin–Latent solution (§8.4 → companion), federated knowledge fusion (§8.3 → companion).

Tier	Meaning	What belongs here
Programmatic / conditional	An interesting direction under explicitly stated additional assumptions. Not part of the core evidence package.	Fundamental constants derivation (§8.7), Riemann Hypothesis connection (§8.8), black-hole extrapolations (§8.6), ontological reflections (Appendix A), AI-as-extractor essay (Appendix B.5).

Reading guidance. The core scientific contract of this book is established entirely within the **Proved here** tier. Companion-backed sections extend the reach of the framework but do not need to be read to evaluate the central theorems. Programmatic sections illustrate the framework's range and motivate future work; they are explicitly marked as conditional wherever they appear.

1.6 Relation to Prior Work

1.6.1 External frameworks

The Latent framework builds on a long tradition in approximation theory, tensor analysis, and statistical representation theory. The table below compares the Latent to the most relevant external frameworks, highlighting what each provides and what the Latent adds.

Framework	Scope	What it provides	What the Latent adds
Kolmogorov n-widths (Pinkus 1985)	Best n-dim linear approximation of function classes	The bound $d_n(\mathcal{F}) = O(\rho^{-n})$ for analytic classes; tight lower bounds	The *object* being approximated (the graded tensor element Λ), the algebraic structure (Proposition 7), and the basis-independence that makes ρ an invariant of the system rather than of the approximation space.
Nonlinear / best-n-term approximation (DeVore 1998; DeVore & Lorentz 1993)	Best n-term dictionary expansion (wavelets, ridgelets, etc.)	Rates depending on sparsity class; dictionary-optimality results	The Latent identifies the *intrinsic object* being approximated and proves the rate is an invariant of the object, not the dictionary. When the extraction basis IS the optimal dictionary, the two viewpoints coincide (Theorem 4).

Framework	Scope	What it provides	What the Latent adds
Compressed sensing (Candès, Romberg & Tao 2006; Donoho 2006)	Recovery of sparse signals from underdetermined linear measurements	$O(s\log(n/s))$ random measurements suffice for s-sparse recovery; RIP-based guarantees	CS assumes sparsity in a *fixed* basis; the Latent framework identifies a basis-free object and proves that the *intrinsic* sparsity N^* depends only on ρ. The Latent is not sparse in general — it is *finitely sufficient*. CS is a coordinate-level extraction method; the Latent Theorem is a coordinate-free existence result.
Tensor decompositions (Kolda & Bader 2009; CP, Tucker, tensor train)	Multi-way data factorization	Low-rank approximation of data tensors; efficient storage and computation	Latent tensor entries are coordinates of a Hilbert-space element, not raw data values — basis change is a first-class operation. The Latent Theorem gives an *a priori* rank bound from ρ; no NP-hard rank computation is needed for the truncation decision.
Reproducing Kernel Hilbert Spaces (Aronszajn 1950; Schölkopf & Smola 2002)	Kernel-based function approximation and learning	Feature maps, representer theorem, kernel PCA	RKHS methods implicitly work with a grade-1 Latent in the kernel eigenbasis. The Latent framework makes the grade explicit (grade-1 captures means, grade-2 captures covariances/generators, grade-3+ captures irreducible multi-body interactions) and provides a universal algebra across grades.

Framework	Scope	What it provides	What the Latent adds
Diffusion maps / spectral geometry (Coifman & Lafon 2006; Bérard, Besson & Gallot 1994)	Nonlinear dimensionality reduction via Laplacian eigenfunctions	Embedding coordinates from the heat kernel; spectral distance on data manifolds	Diffusion-map coordinates are a grade-1 Latent in the Laplacian eigenbasis. The Latent framework adds higher grades (the structure constants c_{ijk}) needed for reconstruction beyond isometry class, and the algebra (transfer, composition) that diffusion maps lack.
Sufficient statistics (Fisher 1922; Lehmann & Scheffé 1950)	Minimal data summaries that preserve all information about a parameter	Existence via factorization theorem; characterization for exponential families	The Latent IS a sufficient statistic in the representation-theoretic sense. The Latent Theorem is a constructive sufficiency result: here is the statistic, here is its size, and the size depends only on regularity — not on dimension, basis, or family membership.
Rate-distortion theory (Shannon 1959)	Minimum bits to represent a source at distortion $\leq D$	Rate-distortion function $R(D)$; operational meaning via source coding theorems	The Latent Theorem is the representation-theoretic analog: minimum *real numbers* (not bits) to represent a smooth system at error $\leq \varepsilon$. Both answers depend on the source's structure, not its dimension. The Latent adds the algebraic layer (operations on representations) that information theory does not address.

What is common. All of the above address the question "how efficiently can smooth/structured objects be represented?" The Latent framework does not claim to improve on the *rates* — the $N = \Theta(\log(1/\varepsilon)/\log\rho)$ bound is classical. The rates are well-known from Bernstein, Kolmogorov, and DeVore.

What is new. The contribution is fourfold (see §1.5 for detail): (a) the *object* (the graded tensor element Λ), (b) the *algebra* on objects (Proposition 7), (c) the *recursion* (meta-Latent), and (d) the *optimal construction* (Theorem 4, Theorem 4b). None of the frameworks above simultaneously provides all four. The closest is tensor decomposition, which has the multi-way structure but not the basis-invariance or the a priori rank bound; and *n*-widths, which have the bound but not the object or the algebra.

1.6.2 Internal predecessors

The Latent also unifies and generalizes five of the author's prior results:

Prior result	What it is	Latent interpretation
Spectral Fenton Distribution (Nagy 2026, *The Spectral Fenton Distribution*)	Vector of 128 cosine coefficients representing a portfolio loss CDF	A grade-1 Latent in cosine coordinates
Universal Spectral Representation Theorem (Nagy 2026, *The Universal Spectral Representation Theorem*)	$N = \Theta(\log(1/\varepsilon)/\log\rho)$ cosine coefficients suffice	The Latent Theorem restricted to cosine extraction
Spectral generator matrix (Nagy 2026, *The Spectral Tensor Representation of Stochastic Processes*)	Matrix M encoding a complete diffusion process	A grade-2 Latent in cosine coordinates
Knowledge Artifact (Nagy 2026, *The Knowledge Artifact and Knowledge Algebra of ML Models*)	Vector of spectral coefficients extracted from an ML model	A grade-1 Latent of a model's predictive distribution
Manifold Latent Theorem (Nagy 2026, *The Manifold Latent Theorem*)	Eigenvalues + structure constants $c_{ijk} = \int \phi_i \phi_j \phi_k \, dV$	A graded Latent of a Riemannian manifold in its Laplacian eigenbasis

Each is correct but basis-specific. The Latent framework shows they are all instances of one phenomenon: smooth systems are finitely representable, and the representation is basis-free. The Eckart-Young theorem (1936) is reinterpreted as the rank-level complement to basis-level optimality (Theorem 3, §4.6), completing the proof that Latent truncation is provably optimal at every level. Spectral knowledge distillation (Hinton et al. 2015) is reinterpreted as a Latent extraction operation with certifiable EY optimality (§7.8).

1.7 Book Structure

The book is organized into seven Parts, eleven chapters, and four appendices.

Part I — Foundations (Chapters 1–2). Chapter 1 isolates the motivating question, places the Latent against the classical spectral literature, and works through four guided examples (projectile motion, coupled oscillators, the trebuchet, Navier–Stokes) that make the ontology concrete before the formal definitions. Chapter 2 defines the Latent formally as the basis-free element of the graded Hilbert tensor algebra.

Part II — The Latent Theorem (Chapter 3). Chapter 3 states and proves the Latent Theorem and its Extended Latent Theorem for non-smooth systems (via sufficient smooth representations). It develops three structural results: the self-diagnostic property (§3.6) — where ρ diagnoses the hierarchy's own quality from

within, including the Analyticity–Rate Duality between density approximation and rare-event simulation — the renormalized grade (§3.8), and the time-Latent duality (§3.9). The grade-of-time framing proposition (Proposition 14) is in Appendix A.3.

Part III — Methodology (Chapters 4–6). Chapter 4 develops optimal extraction, including the Representation–Structure Matching Theorem (§4.4) and rank-level optimality via the Eckart–Young Theorem (§4.6). Chapter 5 treats grades and entanglement, including the Grade Collapse Theorem (§5.4) and the Object/Process Latent dichotomy. Chapter 6 constructs the Latent Algebra, proves universality (§6.3), develops maps and complexification (§6.5–§6.6), and closes with the two-Latent composition structure of optimization (§6.9).

Part IV — Unification and Applications (Chapters 7–8). Chapter 7 maps every existing spectral result to its Latent: COS expansions, eigenvalue conditioning, the Fokker–Planck generator, Knowledge Artifacts, spectral importance sampling (§7.6), and spectral knowledge distillation (§7.8). Chapter 8 applies the framework to large language models, financial tail dependence, federated knowledge fusion, the three-body problem, Navier–Stokes turbulence, black holes, fundamental constants, the Riemann zeta function, and closed Riemannian manifolds — ordered by evidentiary strength, with explicit claim-tier markers (§1.5a) on the programmatic/conditional sections (§8.6–§8.8). The ontological synthesis (former §8.10) and AI-as-extractor essay (former §8.11) have been moved to Appendix A.4 and Appendix B.5 respectively. Nine additional domain case studies live in Appendix C.

Part V — Verification and Program (Chapter 9). Chapter 9 relates the framework to existing frameworks (§9.1), summarizes the companion paper program (§9.2), states limitations (§9.3), outlines near-term research directions (§9.4), and provides a section-to-kernel verification map (§9.5). Three long-horizon directions — Montgomery's four N-body open questions, fundamental constants as grade projections, and the Latent renormalization group — live in Appendix D.

Part VI — The Shadow Principle (Chapter 10). Chapter 10 develops the Shadow Principle (Theorem 9), an epistemic theorem about projection and detection, and its four mathematical manifestations: grade shadows, the Projection Theorem, the mean-field limit, and the Riemann Hypothesis as a shadow-free system. The expository extensions to intelligence, consciousness, and cognition live in Appendix B.

Part VII — Conclusion (Chapter 11). Chapter 11 summarizes the central claims and closes the book.

Appendices. Appendix A gathers the philosophical foundations (the seven-element ontology, the ontological inversion, the grade-of-time theorem, and the ontological claim moved from §8.10). Appendix B develops the four mind-and-cognition shadows plus the AI-as-extractor essay (moved from §8.11). Appendix C is a domain catalog (game theory, MHD, quantum–classical transition, cosmology, space debris, scale-free networks, Ricci flow, spectral Bellman, quantum trainability). Appendix D collects three extended research directions. Each appendix opens with a pointer table mapping v1.0-monograph section IDs to the new v2.0-book locations.

1.8 Guided Examples: The Ontology in Action

Before the formal definitions, we walk through two physical systems — one trivial, one not — mapping every element of the ontology (Appendix A.1) to a concrete object. The reader should verify that each element plays an irreplaceable role.

Example A: Projectile Motion (The Trivial Latent)

A ball is thrown from the ground at angle θ with initial speed v_0.

Reality, The One (Ω). The ball interacts with $\sim 10^{25}$ air molecules; its surface deforms elastically; spin creates Magnus force; the Earth rotates underneath; wind varies with altitude. Ω is infinite-dimensional.

The Latent (Λ). In the vacuum idealization: $\Lambda = (v_0, \theta, g) \in \mathbb{R}^3$. Three numbers. From these, the *entire* trajectory is recoverable:

$$x(t) = v_0 \cos\theta \cdot t, \qquad y(t) = v_0 \sin\theta \cdot t - \frac{1}{2} g\, t^2.$$

The Manifestor. Newton's second law in a uniform gravitational field. The physical law takes the 3-number Latent and generates the continuous infinite trajectory: $\Lambda \xrightarrow{F=ma} \Omega$.

The Lens. Three observers watch the same throw:

Lens	Instrument	What passes through
A	High-speed camera (240 fps)	Positions at discrete times
B	Radar gun	Speed magnitude at discrete times
C	Landing marker	Single point where the ball hits the ground

The Observable. Each Lens produces a different Observable from the same Reality:

Lens	Observable	Dimension
A	$\{(t_i, x_i, y_i)\}_{i=1}^{n}$	$3n$ numbers
B	$\{(t_i, \lvert\mathbf{v}_i\rvert)\}_{i=1}^{m}$	$2m$ numbers
C	x_{land}	1 number

The Extractor. From Observable back to Latent: - Lens A: least-squares fit of the parabolic model to (t_i, x_i, y_i) → recovers (v_0, θ, g). - Lens B: velocity-model fit → recovers (v_0, θ, g). - Lens C: **fails**. One number cannot recover three parameters. The Lens is too narrow — information is lost.

This is the Lens hierarchy in action: wider Lens → richer Observable → successful Extraction. A Lens that is too narrow makes the Latent unrecoverable.

The Projector. Once the Latent is recovered, different Projectors extract different answers from the same $\Lambda = (v_0, \theta, g)$:

Projector query	Formula	Output
"Where does it land?"	$x_{\text{land}} = v_0^2 \sin 2\theta / g$	a distance
"Maximum height?"	$h_{\text{max}} = v_0^2 \sin^2\theta/(2g)$	a height
"Time of flight?"	$T = 2v_0 \sin\theta / g$	a duration

Each query is a function $\Pi_i : \mathbb{R}^3 \to \mathbb{R}$ — a projection from the 3-dimensional Latent to a 1-dimensional answer.

Grade structure. With two spatial coordinates ($d = 2$, so $\mathcal{H} = \mathbb{R}^2$):

Grade	Object	Content	Value
0	$\Lambda^{(0)} \in \mathbb{R}$	Mean behavior	Average height $\bar{y}$
1	$\Lambda^{(1)} \in \mathbb{R}^2$	Individual coordinate effects	(v_{x0}, v_{y0})
2	$\Lambda^{(2)} \in \mathbb{R}^{2\times 2}$	Pairwise interaction	**0**

Grade 2 is **zero**. In vacuum, horizontal and vertical motions are completely independent — $x(t)$ tells you nothing about $y(t)$ beyond what you already know. Projectile motion is a *pure grade-1 system*. This is precisely WHY it is trivial: no coupling, no interaction, no entanglement between coordinates.

Analyticity. The trajectory is polynomial in t (degree 2), hence entire (analytic everywhere). $\rho = \infty$. The Latent size $N = 3$ is *exact*, not approximate — the system is perfectly finitely representable for any accuracy ε.

Example B: Coupled Oscillators (Grade 2 Appears)

Two masses m_1, m_2 on a frictionless rail, each attached to a wall by springs k_1, k_2, connected to each other by a coupling spring k_c. When mass 1 moves, it pulls on mass 2.

Reality (Ω). The continuous trajectories of both masses over time, including thermal fluctuations, spring nonlinearities, rail friction — again, infinite-dimensional.

The Latent (Λ). System parameters $(m_1, m_2, k_1, k_2, k_c)$ plus initial conditions $(x_1(0), \dot{x}_1(0), x_2(0), \dot{x}_2(0))$. Total: $\Lambda \in \mathbb{R}^9$. Nine numbers generate the entire two-body dynamics for all time.

The Manifestor. Newton's law for coupled systems:

$$m_1\ddot{x}_1 = -k_1x_1 - k_c(x_1 - x_2), \qquad m_2\ddot{x}_2 = -k_2x_2 - k_c(x_2 - x_1).$$

The Lens. Accelerometers on each mass, sampled at 100 Hz.

The Observable. $\{(t_i, x_{1,i}, x_{2,i})\}_{i=1}^n$ — discrete position pairs.

The Extractor. Fourier transform of the time series → identify two peak frequencies ω_+, ω_- (the normal modes) → recover amplitudes and phases → invert to physical parameters.

The Projector. - Π_1: "What are the normal mode frequencies?" → $\omega_\pm$ are the positive roots of $\det(K - \omega^2 M) = 0$ with stiffness $K = \begin{pmatrix} k_1 + k_c & -k_c \\ -k_c & k_2 + k_c \end{pmatrix}$ and mass matrix $M = \mathrm{diag}(m_1, m_2)$ (in the symmetric case $m_1 = m_2 = m$, $k_1 = k_2 = k$, one finds $\omega_-^2 = k/m$ and $\omega_+^2 = (k + 2k_c)/m$ for the antisymmetric and symmetric normal modes). - Π_2: "When does all energy transfer to mass 2?" → beat period $T_{\text{beat}} = 2\pi/|\omega_+ - \omega_-|$.

Grade structure — where it gets interesting. With $d = 2$ masses, $\mathcal{H} = \mathbb{R}^2$:

Grade	Object	Content	Value
0	$\Lambda^{(0)} \in \mathbb{R}$	Mean behavior	Equilibrium position
1	$\Lambda^{(1)} \in \mathbb{R}^2$	Individual mass behavior	What each mass would do *alone* (own spring, own frequency)

Grade	Object	Content	Value
2	$\Lambda^{(2)} \in \mathbb{R}^{2\times 2}$	Pairwise interaction	$\begin{pmatrix} 0 & k_c \\ k_c & 0 \end{pmatrix}$

The grade-2 tensor IS the coupling spring. Its off-diagonal entries carry irreducible information that exists *between* the masses but in neither individually.

The critical experiment: disconnect the coupling spring ($k_c = 0$). Now: - Grade 2 vanishes: $\Lambda^{(2)} = \mathbf{0}$. - The system decomposes into two independent oscillators. - No energy transfer (no beats) — each mass oscillates at its own frequency forever. - The Latent splits: $\Lambda = \Lambda_1 \oplus \Lambda_2$ (direct sum of independent Latents).

Reconnect the spring ($k_c > 0$): - Grade 2 appears. - Energy flows back and forth between the masses (beats). - The Latent is *irreducible* — you cannot describe the system by describing each mass separately.

Grade 2 IS the coupling. It measures the irreducible information that lives in the *relationship* between components, not in any component alone. This is the content that pairwise methods (covariance, correlation, Gaussian copulas) capture.

Analyticity. The solution is a sum of sinusoids: $x_i(t) = \sum_j A_j \cos(\omega_j t + \phi_j)$, which is entire (analytic everywhere). $\rho = \infty$. $N = 9$ exactly.

The Grade Progression

System	Grade content	Complexity	Why
Projectile (vacuum)	0, 1 only	Trivial	No coupling between coordinates
Coupled oscillators	0, 1, 2	Moderate	Pairwise coupling via k_c
Three bodies (§5, §8.4)	0, 1, 2, 3+	Hard	Triple interaction that exists in no pair

Each grade level adds a qualitatively new type of information. Grade 3 — genuine three-way interaction that cannot be decomposed into pairwise terms — is what makes the three-body problem fundamentally harder than any collection of two-body problems. This is developed formally in Section 5.

Example C: The Trebuchet (Design Optimization as Projector Query)

Examples A and B extracted Latents from observed systems. Example C reverses the direction: *design* a system by optimizing over its Latent.

A trebuchet is a medieval siege engine: a beam pivots on an axle; a heavy counterweight on the short arm drives a sling on the long arm, launching a projectile. It is a constrained rigid-body system with analytic equations of motion — a perfect Latent candidate.

The system. Two degrees of freedom: beam angle θ and sling angle ϕ (relative to the beam). Five design parameters: short arm a, long arm b, sling length l, counterweight mass M, projectile mass m.

The Latent (Λ). Separates naturally into two subspaces:

$$\Lambda = \underbrace{(a,\, b,\, l,\, M,\, m)}_{\Lambda_{\text{design}} \in \mathbb{R}^5} \oplus \underbrace{(\theta_0,\, \dot{\theta}_0,\, \phi_0,\, \dot{\phi}_0)}_{\Lambda_{\text{initial}} \in \mathbb{R}^4}$$

Nine numbers generate the complete trajectory of every component for all time.

The Manifestor. The Euler–Lagrange equations from the Lagrangian $L = T - V$:

$$T = \frac{1}{2}\dot{\mathbf{q}}^{\top}\mathbf{M}(\mathbf{q})\,\dot{\mathbf{q}}, \qquad V = Mga\cos\theta - mg\big(b\cos\theta + l\cos(\theta+\phi)\big)$$

where $\mathbf{q} = (\theta, \phi)$ and the mass matrix is

$$\mathbf{M}(\mathbf{q}) = \begin{pmatrix} I_\theta + ml^2 + 2mbl\cos\phi & ml^2 + mbl\cos\phi \\ ml^2 + mbl\cos\phi & ml^2 \end{pmatrix}$$

with $I_\theta = Ma^2 + mb^2 + I_{\text{beam}}$. Every function here is trigonometric or polynomial — the system is analytic, $\rho > 1$.

Grade structure. The mass matrix $\mathbf{M}(\mathbf{q})$ makes the grades visible:

Grade	Object	Content
0	$\Lambda^{(0)}$	System energy (scalar)
1	Diagonal of $\mathbf{M}$	What each DOF does alone: beam swings (moment I_θ), sling swings (moment ml^2)
2	Off-diagonal $M_{12} = ml^2 + mbl\cos\phi$	Energy transfer: beam rotation → sling acceleration → projectile launch

The sling test (grade-2 removal). Set $l = 0$ (remove the sling). Now: - $M_{12} = 0$: grade 2 vanishes. - The system reduces to a weighted pendulum — the beam swings, the projectile sits at its tip, no energy amplification. - Range drops by an order of magnitude.

The sling creates the coupling. The coupling IS the trebuchet's power: it transfers gravitational potential energy from the counterweight through two stages of angular momentum amplification to the projectile. Grade 2 is not a mathematical abstraction — it is the sling.

Design optimization as Projector query. Define the range Projector:

$$\Pi_{\text{range}}(\Lambda) = x_{\text{land}}\big(\Lambda_{\text{design}},\, \Lambda_{\text{initial}}\big)$$

The optimal trebuchet design is:

$$\Lambda^*_{\text{design}} = \arg\max_{\Lambda_{\text{design}}} \Pi_{\text{range}}(\Lambda_{\text{design}},\, \Lambda^0_{\text{initial}})$$

for fixed loading conditions $\Lambda^0_{\text{initial}}$. But the *same* Latent supports other Projectors simultaneously:

Projector	Query	Optimizes for
Π_{range}	"How far does it fly?"	Maximum distance
Π_{eff}	$E_{\text{projectile}}/E_{\text{counterweight}}$	Energy efficiency
Π_{sens}	$\|\partial x_{\text{land}}/\partial \Lambda_{\text{design}}\|$	Design robustness (low = good)
Π_{rate}	Shots per hour at given range	Operational throughput

These four optima are generally *different* — the maximum-range design is not the most efficient, and the most efficient is not the most robust. Multi-objective engineering design is the search for Pareto-optimal points in the space of Projector values, all computed from a single Latent extraction.

Coupling efficiency from the grade-2 tensor. The normalized coupling strength

$$\eta = \frac{|M_{12}|^2}{M_{11} \cdot M_{22}}$$

measures what fraction of kinetic energy can transfer between beam and sling per cycle. At the release configuration ($\phi \approx 0$, sling aligned with beam):

$$\eta\big|_{\phi=0} = \frac{m(l+b)^2}{I_\theta + m(l+b)^2}.$$

This single formula encodes the classical design tradeoff: longer sling ($l \uparrow$) raises η but slows the beam; heavier counterweight ($M \uparrow$) raises I_θ and lowers η unless the arm ratio compensates. The empirically known optimal ratios ($b/a \approx 3.75$, $l \approx b$) follow from maximizing η subject to structural constraints — the Latent makes this structural rather than numerical.

Beyond the standard topology: why one sling? The most revealing question is not "what are the best parameters?" but "is the standard trebuchet topology itself optimal?" To answer this, extend the design space beyond the classical form:

Topology	DOF	Grade structure	Effect
Beam only ($l = 0$)	1	Grade 1 only	Pendulum. No amplification.
Beam + 1 sling	2	Grade 2 appears	**Standard trebuchet.** Angular momentum amplification through one coupling stage.
Beam + 2 slings	3	Grade 2 + **grade 3**	**Cascade / whip dynamics.** Each stage amplifies velocity: the second sling is shorter and lighter, producing higher tip speed — the same principle that lets a bullwhip crack at supersonic speed.

The cascade trebuchet (two slings in series) introduces a qualitatively new element: the grade-3 tensor $\Lambda^{(3)} \in \mathcal{H}^{\otimes 3}$, encoding the simultaneous three-way interaction among beam, sling_1, and sling_2. This is not a sum of pairwise effects — it is the irreducible information that exists in the triple but in no pair.

Grade 3 has two faces: - **Amplification**: the cascade chain transfers energy through two stages, each with its own grade-2 coupling, potentially multiplying the velocity gain. - **Destabilization**: the three-way interaction feeds energy back through the chain in ways that no pair controls. The system becomes sensitive to timing — a slight mismatch in the release sequence can redirect energy backward instead of forward.

The Latent framework makes a *testable prediction*: the optimal number of stages n^* is the largest n for which the marginal grade-2 gain from the $(n+1)$-th stage exceeds the grade-3 loss from the new triple interaction:

$$n^* = \max\{n : \Delta\|\Lambda^{(2)}_{n+1}\| > \Delta\|\Lambda^{(3)}_{n+1}\|\}.$$

For a rigid trebuchet with a heavy stone projectile, the energy budget is dominated by the counterweight's potential energy, and the sling's moment of inertia is small. The first sling captures most of the available grade-2 coupling; a second sling adds marginal grade-2 benefit but introduces grade-3 feedback that destabilizes the release. The prediction: $n^* = 1$, explaining why historical trebuchets universally used exactly one sling — not zero (a pendulum, too weak), not two (a whip, too unstable for heavy payloads).

For a *light* projectile (think: a bullwhip, where the payload is the tip itself), the calculus changes. Each stage's mass drops rapidly, the grade-3 feedback is small relative to the grade-2 cascade, and many stages are optimal. A whip IS a multi-stage trebuchet with $n^* \gg 1$ — and it achieves supersonic tip velocity, consistent with the prediction.

What this example adds. Projectile motion had no coupling (grade 1). Coupled oscillators had linear coupling (grade 2, constant off-diagonal). The trebuchet has *state-dependent nonlinear* coupling (M_{12} depends on $\cos\phi$), a natural design optimization problem, and — when we ask whether the topology itself is optimal — a testable prediction linking grade structure to the number of amplification stages. The progression:

Example	Coupling	Grade 2	New concept
Projectile	None	**0**	Ontology walkthrough
Coupled oscillators	Linear (k_c)	Constant matrix	Grade 2 = interaction
Trebuchet	Nonlinear ($\cos\phi$)	State-dependent matrix	Design optimization as Projector query
Trebuchet topology	Multi-stage	Cascade + grade 3	Grade structure predicts optimal topology

Example D: Navier–Stokes (The PDE Latent)

The incompressible Navier–Stokes equations govern all viscous fluid flow:

$$\partial_t u = \nu\Delta u - \mathbb{P}(u \cdot \nabla u)$$

where $u(x,t)$ is the velocity field on $\mathbb{T}^3$, ν is the viscosity, and $\mathbb{P}$ is the Leray projection enforcing incompressibility.

Reality, The One (Ω). The velocity field $u(x,t)$ is a function on an infinite-dimensional space — at each of uncountably many spatial points, there is a 3-component vector. The full state is $\Omega = L^2(\mathbb{T}^3;\mathbb{R}^3)$, an infinite-dimensional Hilbert space.

The Latent (Λ). If the initial data u_0 is real-analytic with analyticity radius $\rho = e^{\sigma} > 1$, the Fourier coefficients decay exponentially: $|c_k| \le M/\rho^{|k|}$. The Latent is the finite truncation:

$$\Lambda(t) = \{c_k(t) : |k| \le N^*\}, \qquad N^* = \left\lceil \frac{\log(MC_3/\varepsilon)}{\log \rho} \right\rceil$$

These $O((N^*)^3)$ numbers completely encode the solution to accuracy ε. The infinite-dimensional velocity field is faithfully represented by finitely many spectral coefficients.

The Manifestor. The NS equation itself. It takes the Latent at time t and generates the Latent at time $t + dt$ via the **Latent ODE**:

$$\dot{c}_k = \underbrace{-\nu|k|^2 c_k}_{\text{grade 1: dissipation}} \underbrace{- i \sum_{j_1+j_2=k} (k \cdot \hat{j}_1)\, \mathbb{P}_k\, c_{j_1} c_{j_2}}_{\text{grade 2: advection}}$$

This is a quadratic ODE on $\mathbb{R}^{N^*}$. The grade-1 term (viscous damping) acts independently on each mode. The grade-2 term (bilinear advection) couples modes through the convolution constraint $j_1 + j_2 = k$.

Grade structure. NS is exactly grade 2: $F(u) = A^{(1)}[u] + A^{(2)}[u,u]$ with $A^{(1)} = \nu\Delta$ (linear) and $A^{(2)} = -\mathbb{P}(\cdot\nabla\cdot)$ (bilinear). This means:

Grade component	Coupling tensor	Structure
$A^{(1)}$: viscosity	$L_{kj} = -\nu\|k\|^2\delta_{kj}$	Diagonal — modes decouple
$A^{(2)}$: advection	$B_{k,j_1,j_2} = -i(k\cdot\hat{j}_1)\,\mathbb{P}_k\,\delta_{k,j_1+j_2}$	Sparse — convolution constraint

The grade-2 Latent of NS has $O((N^*)^2)$ dynamically significant couplings — not $(N^*)^{N^*}$ — because high-high mode interactions are exponentially suppressed by the spectral decay: $|B_{k,j_1,j_2}c_{j_1}c_{j_2}| = O(\rho^{-N^*}) = O(\varepsilon)$ when $|j_1|, |j_2| > N^*/2$.

The Lens. An experimentalist can observe the flow through:

Lens	What passes through
PIV (particle image velocimetry)	2D velocity snapshots at discrete times
Hot-wire anemometer	1D velocity at one point, continuous in time
Pressure sensor	Scalar field (Bernoulli: $p = -\frac{1}{2}\|u\|^2 + \text{const}$)

All are projections from the same Latent.

The resolved question. The analyticity radius $\rho(t) = e^{\sigma(t)}$ evolves with the solution. If $\sigma(t) > 0$ for all time, the Latent representation is globally valid: NS solutions stay smooth and finitely representable. If $\sigma(t) \to 0$

in finite time, the Latent degrades — $N^* \to \infty$ — and a singularity forms. This was the Navier–Stokes millennium problem, reformulated as: **does the Latent of fluid flow remain finite-dimensional?**

Answer: yes. In the companion paper (Nagy, 2026, "The PDE Tensor Algebra"), we prove global regularity via the **Leray Bilinear Spreading Lemma**: the grade-2 coupling tensor $A^{(2)}$ decomposes into a (D, C, P) triple where $D = \nu|k|^2$ is the grade-1 dissipation, C is the bilinear coupling, and $P = \mathbb{P}_k$ is the Leray projector. The projector's rank-2 structure forces geometric spreading of the coupling directions, giving a deterministic contraction factor $\gamma = |\sin(\pi f)/(\pi f)| < 1$ for the phase coherence per cascade octave. Combined with viscous dissipation via a bootstrap argument, this guarantees $\sigma(t) > 0$ for all t — the Latent remains finite-dimensional, and the fluid stays smooth.

In Latent language: the difficulty $\mathcal{D} = \|C\|_F/\lambda_{\min}(D) = \mathrm{Re}$ is the "inverse Latent Number" of the dynamical coupling. When the grade-2 term has $\mathcal{D} < \infty$ and the constraint projector provides geometric spreading, the Latent's analyticity radius $\rho(t)$ is bounded away from 1 — the system never reaches the incompressible phase boundary. The Leray projector is, from this perspective, the structural guarantee that NS solutions remain Latent-representable.

What this example adds. All previous examples were finite-dimensional ODEs (projectile, oscillators, trebuchet). NS is the first PDE — an infinite-dimensional dynamical system. The Latent reduces it to a finite ODE, with certified error, and the grade structure (2-body coupling, exponential suppression, convolution sparsity) is preserved exactly. The progression:

Example	System	Dimension	Grade	New concept
Projectile	ODE, $\mathbb{R}^2$	3 parameters	1	Ontology
Oscillators	ODE, $\mathbb{R}^{2n}$	$2n$	2 (constant)	Interaction
Trebuchet	ODE, $\mathbb{R}^4$	4	2 (state-dep.)	Design optimization
Navier–Stokes	**PDE,** $L^2(\mathbb{T}^3)$	$\infty \to N^*$	**2**	**Latent compresses ∞ to finite**

Chapter 2. The Latent: Definition

2.1 Setup

Let (Ω, μ) be a measure space with $\mu(\Omega) < \infty$. Let $\mathcal{H} = L^2(\Omega, \mu)$ be the Hilbert space of square-integrable functions on Ω.

Notation key (symbol reuse). Several symbols are used for closely related but distinct quantities in the book. In each case the local context (subscripts and the surrounding definition) disambiguates, but for quick reference:

Symbol	Primary meaning	Secondary uses	Disambiguator
N	Truncation dimension of the Latent (central parameter)	Particle count in an N-body system (§8.4)	The N-body chapters are consistently labelled "N-body"; the Latent N appears with ε and ρ.
ρ	Analyticity parameter of the Latent	Pearson / Gaussian correlation ρ_{ij} (benchmarks, §3.6)	Scalar ρ is always the analyticity parameter; a subscripted ρ_{ij} or ρ_{grav} is a correlation or a domain-specific radius.
r	Tensor grade (grade-r Latent)	Reynolds number Re (turbulence, §8.5)	Reynolds number is written as Re, never as a bare r.
M	Generator / Eckart–Young matrix (spectral chapter)	Particle mass (M_P, M_{GUT}, etc.)	Masses always carry a subscript; the bare M is the generator.
h	LLM hidden state $h_L(x)$, and symplectic step size (§8.4 N-body integrator)	—	h_L and $h(x)$ are LLM vectors; bare h is the integrator step.
C_f, C_Λ	Bounding constants in physical-constants §8.7	—	Upper-case C with a subscript is always a bound, never a particle mass.

Readers who encounter a symbol in an ambiguous local context should consult this table and the surrounding definition; the book does not redefine any symbol across chapters without explicit flagging.

Definition 1 (Extraction basis). An *extraction basis* for $\mathcal{H}$ is a complete orthonormal system $\mathcal{B} = \{\phi_k\}_{k\in\mathbb{N}_0}$ satisfying:

(a) $\phi_0 = \mu(\Omega)^{-1/2} \cdot \mathbf{1}_\Omega$ (probability normalization).

(b) There exists a self-adjoint operator $\mathcal{D}$ on $\mathcal{H}$ with $\mathcal{D}\phi_k = d_k\phi_k$, $d_0 = 0$, $|d_k| \to \infty$ (operator diagonalization).

(c) $\phi_k : \Omega \to \mathbb{R}$ for all k (real-valuedness).

Different extraction bases include:

Basis	Domain	Reference operator $\mathcal{D}$	Natural for
Cosine	$[a, b]$	$-d^2/dx^2$, Neumann BC	Bounded-domain pricing
Hermite	$\mathbb{R}$	$-d^2/dx^2 + x^2$	Gaussian / light tails
Laguerre	$[0, \infty)$	$-x\,d^2/dx^2 - (1-x)\,d/dx$	Positive quantities
Spherical harmonics	S^{d-1}	$-\Delta_{S^{d-1}}$	Directional data

2.2 The Graded Tensor Space

Definition 2 (Latent space). The *Latent space* over $\mathcal{H}$ is the graded Hilbert tensor algebra:

$$\mathfrak{L}(\mathcal{H}) = \bigoplus_{r=0}^{\infty} \mathcal{H}^{\otimes r}$$

where $\mathcal{H}^{\otimes 0} = \mathbb{R}$ and $\mathcal{H}^{\otimes r}$ is the r-fold Hilbert tensor product. The grade of an element in $\mathcal{H}^{\otimes r}$ is r, called its *interaction order*.

2.3 The Latent

Definition 3 (Latent). The *Latent* of a system $\mathcal{S}$ with interaction order r is the element $\Lambda(\mathcal{S}) \in \mathcal{H}^{\otimes r} \subset \mathfrak{L}(\mathcal{H})$ that encodes the complete distributional, dynamic, and functional properties of $\mathcal{S}$.

Given an extraction basis $\mathcal{B} = \{\phi_k\}$, the *coordinates* (also called *amplitudes*) of the Latent are:

$$\Lambda^{(\mathcal{B})}_{k_1 k_2 \cdots k_r} = \langle \Lambda(\mathcal{S}),\, \phi_{k_1} \otimes \phi_{k_2} \otimes \cdots \otimes \phi_{k_r} \rangle_{\mathcal{H}^{\otimes r}}$$

The coordinates are a real-valued tensor of order r. Different extraction bases give different coordinates for the same Latent. Each basis vector ϕ_k is a *mode*; its coordinate $\Lambda^{(\mathcal{B})}_k$ is the mode's *amplitude* — the strength of that independent component in the chosen representation.

Proposition 1 (Basis change). Let $\mathcal{B}$ and $\mathcal{B}'$ be two extraction bases related by a unitary U: $\phi'_k = \sum_j U_{kj} \phi_j$. Then:

$$\Lambda^{(\mathcal{B}')}_{k_1 \cdots k_r} = \sum_{j_1, \ldots, j_r} U_{k_1 j_1} \cdots U_{k_r j_r}\, \Lambda^{(\mathcal{B})}_{j_1 \cdots j_r}$$

Proof. Direct from the definition and the unitarity of U. □

The Latent is the object. The coordinates are the representation. The extraction basis is the choice of lens.

2.4 The Grade Hierarchy

Grade r	The Latent is	What it encodes	Existing framework
0	Scalar	Total mass, normalization	Implicit in all methods
1	Function in $\mathcal{H}$	Distribution, density, prediction surface	Spectral Fenton; Knowledge Artifact
2	Operator on $\mathcal{H}$	Generator, covariance, transfer operator	FP generator; eigenvalue conditioning
3	3-tensor	Three-body interaction; time-varying dynamics; co-skewness	**New**
4+	Higher tensor	Multi-body forces; higher cumulants; multi-model interference	**New**

Naming convention. We refer to $\Lambda^{(r)} \in \mathcal{H}^{\otimes r}$ as the *grade-r component* of the Latent. The decomposition $\Lambda = \bigoplus_r \Lambda^{(r)}$ is the *grade structure*. Within each grade, the orthogonal independent components are *modes*; their strengths in a given basis are *amplitudes*. Each grade absorbs all dependence at its interaction order and passes the residual to the next grade; the hierarchy terminates when the system's interaction complexity is exhausted.

Part II — The Latent Theorem

The central claim of the book is a theorem, not a principle. Chapter 3 proves that every system with $\rho > 1$ admits a finite sufficient representation whose size is logarithmic in the target accuracy and independent of ambient dimension. It also extends the theorem to non-smooth systems via sufficient smooth representations, and provides a self-diagnostic — a set of indicators computable from the finite Latent itself — that allows the hierarchy to judge its own compressibility from within.

Chapter 3. The Latent Theorem

3.1 Analyticity Parameter

Definition 4 (Analyticity parameter). Let $\Lambda \in \mathcal{H}^{\otimes r}$ be a Latent with coordinates $\Lambda_k^{(\mathcal{B})}$ in extraction basis $\mathcal{B}$. The *analyticity parameter* of Λ relative to $\mathcal{B}$ is:

$$\rho_{\mathcal{B}}(\Lambda) = \liminf_{|k|\to\infty} |\Lambda_k^{(\mathcal{B})}|^{-1/|k|}$$

where $|k| = k_1 + \cdots + k_r$. When $\rho_{\mathcal{B}} > 1$, the coordinates decay exponentially: $|\Lambda_k^{(\mathcal{B})}| = O(\rho_{\mathcal{B}}^{-|k|})$.

The analyticity parameter is a property of the **pair** (Latent, basis). Different bases yield different ρ for the same Latent. This asymmetry is the source of the optimal extraction problem (Section 4).

Definition 4b (Intrinsic rate). The *intrinsic rate* of a Latent $\Lambda \in \mathcal{H}^{\otimes r}$ is:

$$\rho^*(\Lambda) = \sup_{\mathcal{B}} \rho_{\mathcal{B}}(\Lambda),$$

where the supremum ranges over complete orthonormal bases of $\mathcal{H}$. By construction, $\rho^*(\Lambda)$ is a basis-free invariant of Λ. It measures *how compressible the Latent can be*, across all choices of lens. The basis-dependent $\rho_{\mathcal{B}}$ measures how compressible it *is*, in a given lens. The optimal extraction problem (§4) is the problem of finding a basis $\mathcal{B}$ with $\rho_{\mathcal{B}}(\Lambda)$ close to $\rho^*(\Lambda)$. When $\rho^*(\Lambda) > 1$, the Latent admits exponential compression in *some* extractable representation; when $\rho^*(\Lambda) \leq 1$, no basis makes the Latent exponentially compressible, and the Extended Latent Theorem (§3.3) must be invoked via a smooth representation Φ.

3.2 The Main Result

Theorem 1 (The Latent Theorem). Let $\Lambda \in \mathcal{H}^{\otimes r}$ be a Latent with $\rho_{\mathcal{B}}(\Lambda) > 1$ for some extraction basis $\mathcal{B}$. Let Λ_N denote the truncation to coordinates with $|k| \leq N$. Then:

(i) Finite sufficiency (upper bound). There exist constants $C_1, C_1' > 0$ (depending on the analyticity envelope and on r, but not on N or on the ambient dimension d once the eigenvalue-conditioning rank K is fixed) such that:

$$\|\Lambda - \Lambda_N\|^2_{\mathcal{H}^{\otimes r}} \le C_1 \cdot N^{r-1} \cdot \rho_{\mathcal{B}}^{-2N} \le C_1' \cdot \rho'^{-2N} \quad \text{for any } 1 < \rho' < \rho_{\mathcal{B}}.$$

Thus $N = \lceil \log(C_1'/\varepsilon^2)/(2\log\rho') \rceil$ coordinates per mode suffice. The multi-index count $\#\{k : |k| = m\} = \binom{m+r-1}{r-1} = O(m^{r-1})$ contributes the N^{r-1} prefactor; it is absorbed by an arbitrarily small rate loss $\rho' < \rho_{\mathcal{B}}$.

(ii) Tightness (lower bound). Over the class of Latents with analyticity parameter $\le \rho$, any representation using fewer than $C_2 \cdot \log(1/\varepsilon)/\log\rho$ parameters per mode cannot achieve ε-accuracy uniformly (Kolmogorov ε-entropy).

(iii) Intrinsic rate (basis-free invariant). Define the **intrinsic rate**:

$$\rho^*(\Lambda) := \sup_{\mathcal{B}} \rho_{\mathcal{B}}(\Lambda),$$

where the supremum is over all complete orthonormal bases of $\mathcal{H}$ (Definition 4b). Then $\rho^*(\Lambda)$ is a basis-free invariant of Λ, and the rate $N = \Theta(\log(1/\varepsilon)/\log\rho^*(\Lambda))$ is achieved in any basis that realizes the supremum (when one exists; otherwise one works ε-close to ρ^*). Rate invariance under unitary change of basis is immediate; the supremum is *not* automatic for every pair of bases — it must be constructed or empirically estimated (see §4, optimal extraction).

(iv) Dimension independence (conditional). For a d-variate system whose Latent is eigenvalue-conditioned to effective rank K (§4.2), N depends on ρ^*, ε, and K, but not on d. Dimension-independence is *conditional on finite effective rank* — not a free consequence of analyticity alone.

Proof.

(i) Upper bound. Let $\mathcal{B} = \{\psi_k\}_{k\in\mathbb{Z}^d}$ be a complete orthonormal basis of $\mathcal{H}$ and $\Lambda_k = \langle \Lambda, \psi_{k_1} \otimes \cdots \otimes \psi_{k_r} \rangle$ the coordinates. By hypothesis, $|\Lambda_k| \le A \cdot \rho_{\mathcal{B}}^{-|k|}$ for some $A > 0$ and all multi-indices k with $|k| = |k_1| + \cdots + |k_r|$. The truncation error satisfies:

$$\|\Lambda - \Lambda_N\|^2 = \sum_{|k|>N} |\Lambda_k|^2 \le A^2 \sum_{m=N+1}^{\infty} \underbrace{\#\{k : |k| = m\}}_{\le \binom{m+r-1}{r-1}} \cdot \rho_{\mathcal{B}}^{-2m}.$$

Since $\binom{m+r-1}{r-1} = O(m^{r-1})$ and $\rho_{\mathcal{B}} > 1$, the geometric tail dominates. For any ρ' with $1 < \rho' < \rho_{\mathcal{B}}$, set $\delta = \rho_{\mathcal{B}}/\rho' > 1$. Then $m^{r-1}\rho_{\mathcal{B}}^{-2m} = m^{r-1}\delta^{-2m}\rho'^{-2m} \le C_r \cdot \rho'^{-2m}$ for $m \ge m_0(\delta, r)$, where C_r absorbs the polynomial-geometric tradeoff. Summing the geometric series:

$$\|\Lambda - \Lambda_N\|^2 \le C_1' \cdot \rho'^{-2N}, \qquad C_1' = A^2 C_r/(1 - \rho'^{-2}).$$

Hence $\|\Lambda - \Lambda_N\| \le \varepsilon$ is achieved by $N = \lceil \log(C_1'/\varepsilon^2)/(2\log\rho') \rceil = \Theta(\log(1/\varepsilon)/\log\rho_{\mathcal{B}})$ coordinates per mode. The key step: the polynomial prefactor m^{r-1} cannot defeat the exponential decay ρ^{-2m} when $\rho > 1$ — it merely costs an arbitrarily small rate loss (ρ' instead of ρ).

(ii) Lower bound. We need to show that $N = \Omega(\log(1/\varepsilon)/\log\rho)$ parameters per mode are necessary. Consider the class $\mathcal{F}_\rho = \{\Lambda \in \mathcal{H}^{\otimes r} : |\Lambda_k| \le \rho^{-|k|}\}$. This is a generalized ellipsoid in the coefficient space. By the Kolmogorov ε-entropy theorem for ellipsoids (Kolmogorov & Tikhomirov 1959; Pinkus 1985, Ch.

7), the ε-entropy satisfies $H_\varepsilon(\mathcal{F}_\rho) = \Theta((\log(1/\varepsilon)/\log\rho)^r)$. Any representation using M real parameters can cover at most a M-dimensional manifold; if $M < C_2 \cdot \log(1/\varepsilon)/\log\rho$ per mode, the covering number of the representation is too small to cover $\mathcal{F}_\rho$ at resolution ε. Therefore some $\Lambda \in \mathcal{F}_\rho$ is not ε-approximated, and $N = \Omega(\log(1/\varepsilon)/\log\rho)$ is necessary.

(iii) Basis-free invariant. Let $U : \mathcal{H} \to \mathcal{H}$ be a unitary and $\mathcal{B}' = U\mathcal{B}$. The coordinates in $\mathcal{B}'$ are $\Lambda'_k = \langle \Lambda, U\psi_{k_1} \otimes \cdots \otimes U\psi_{k_r} \rangle$. Since U is isometric, $\sum |\Lambda'_k|^2 = \sum |\Lambda_k|^2$ (Parseval), so the total energy is preserved. However, the *distribution* of energy across modes can change: $\rho_{\mathcal{B}'}$ may differ from $\rho_{\mathcal{B}}$. The supremum $\rho^*(\Lambda) = \sup_{\mathcal{B}} \rho_{\mathcal{B}}(\Lambda)$ is well-defined (as a supremum of non-negative reals bounded above by the analyticity radius of Λ in the best direction) and is basis-free by construction. If the supremum is achieved — as it is for systems with a natural eigenbasis (Theorem 3, §4.3) — then $N^* = \Theta(\log(1/\varepsilon)/\log\rho^*)$ is the optimal coordinate count.

(iv) Dimension independence. Let the system have d degrees of freedom and suppose that eigenvalue conditioning (§4.2) reduces the effective rank to K. The Latent Λ then lives effectively in $\mathcal{H}_K^{\otimes r}$ where $\mathcal{H}_K$ is the K-dimensional subspace spanned by the K leading eigenvectors. By part (i) applied in $\mathcal{H}_K$, the coordinate count is $N = \Theta(\log(1/\varepsilon)/\log\rho)$ per mode, with K^r modes in total — independent of d. The condition is *finite effective rank*: if $K = K(d)$ grows with d, then so does the Latent dimension. This is the correct statement: dimension independence is conditional on the spectral gap, not a universal consequence of smoothness. □

In plain language: every smooth system has a Latent. The Latent is finite. Its size depends on how smooth the system is (ρ) and how accurate you want to be (ε), not on how big the system is (d) or how you extract it (basis).

Remark (Latent dimension is Projector-relative). The Latent's dimension is not an intrinsic property of the system — it is a property of the *question* the Projector asks. The same system can yield a finite or infinite Latent depending on what is being extracted.

System	Projector query	Latent space	Dimension
Portfolio of n assets	"How do they correlate?"	$\mathbb{R}^{n(n-1)/2}$ (covariances)	finite
Portfolio of n assets	"What is the joint density?"	$L^2(\mathbb{R}^n)$	infinite
Portfolio of n assets	"What is the density's parametric structure?"	(μ, Σ, ν)	finite

The framework does not dogmatically declare "the Latent is finite." Rather: the Latent Theorem guarantees that if $\rho > 1$, the answer to any fixed Projector query is *effectively* finite. But the framework itself does not constrain dimensionality. Just as the Lens determines the Observable (Appendix A.1), the Projector determines which aspect of the Latent is relevant — and thereby its effective dimension. The infinite Hilbert space is the carrier; the finite Latent is the content. Which content depends on the question.

Remark (Relation to classical approximation theory). The exponential decay bound in part (i) follows from classical results in approximation theory for analytic functions (Paley–Wiener, Bernstein). The lower bound (ii) is a standard entropy argument. We do not claim these individual estimates are new. What is new is the *object* to which they are applied and the *consequences* that follow:

1. **The object is basis-free.** Classical results bound the decay of Fourier, Hermite, or Chebyshev coefficients — in a specific basis. Theorem 1 identifies the intrinsic quantity (ρ) that controls ALL such

expansions simultaneously, and establishes that the rate $\Theta(\log(1/\varepsilon)/\log\rho)$ is a property of the Latent, not of the coordinate system. The rate is known; the invariance is not.

2. **The object has algebraic structure.** A truncated Fourier expansion is a vector. A Latent is an element of a graded tensor algebra with four operations (addition, scalar multiplication, tensor product, contraction) that commute with basis change. This structure — developed in Section 6 — does not exist for bare coefficient vectors.

3. **The object has grades.** A 1D coefficient expansion is grade 1. The framework extends to grade r (interaction order), capturing r-body effects that no 1D expansion can represent. Grades 3+ (Sections 5, 8) are the genuinely new territory.

The contribution of this paper is not a new bound on coefficient decay. It is the identification that coefficient decay across all bases is controlled by a single intrinsic object — the Latent — and that this object carries algebraic, graded, and basis-transfer structure that individual coefficient vectors do not.

3.3 The Extended Latent Theorem: Beyond Smooth Systems

The Latent Theorem as stated requires $\rho > 1$ — analyticity of the system. But many important systems are NOT smooth: jump processes have discontinuous densities, digital option payoffs are step functions, ReLU networks have kinks at every neuron, empirical distributions are sums of atoms. Do these have Latents?

Yes — whenever the system admits a sufficient smooth representation. The smoothness requirement falls on the REPRESENTATION, not on the system itself. The chain is:

$$\text{non-smooth system } \mathcal{S} \xrightarrow{\Phi} \text{smooth representation } \Phi(\mathcal{S}) \xrightarrow{\Lambda} \text{finite Latent}$$

If Φ is **sufficient** (preserves all information) and $\Phi(\mathcal{S})$ is **smooth** ($\rho > 1$), then the composition $\Lambda \circ \Phi$ gives a finite Latent of the non-smooth system. The system doesn't need to be smooth. It needs a smooth *representation*.

Definition 5 (Sufficient smooth representation). A map $\Phi : \mathcal{S} \to \mathcal{H}$ from a system $\mathcal{S}$ to a Hilbert space $\mathcal{H}$ is a *sufficient smooth representation* if: (a) Φ is injective (preserves all information about $\mathcal{S}$), (b) there exists an extraction basis $\mathcal{B}$ and a constant $\rho_\Phi > 1$ such that $\Phi(\mathcal{S})$ is analytic in $\mathcal{B}$ with parameter ρ_Φ, and (c) the basis $\mathcal{B}$ is *constructively specified* (not merely asserted to exist) — e.g., Fourier for translation-invariant problems, Hermite for Gaussian-weighted, Chebyshev for bounded-interval.

Condition (c) is essential: without a construction, (b) can be vacuously satisfied in pathological bases. The **intrinsic rate** $\rho^*(\Phi(\mathcal{S}))$ (Definition 4b) is the basis-free invariant; a system admits a useful Latent iff $\rho^*(\Phi(\mathcal{S})) > 1$, and Definition 5 merely asks that this intrinsic rate be *achieved* by a concrete basis.

Theorem 2 (Extended Latent Theorem). Every system $\mathcal{S}$ that admits a sufficient smooth representation Φ has a Latent:

$$\Lambda(\mathcal{S}) = \Lambda(\Phi(\mathcal{S}))$$

of size $N = \Theta(\log(1/\varepsilon)/\log\rho_\Phi)$. The smoothness condition applies to the representation, not the system. The class of systems with Latents is closed under the operation "take a sufficient smooth representation of" — making it far broader than the class of smooth systems alone.

Proof.

Step 1 (Sufficiency). Since Φ is injective (Definition 5(a)), the map $\Phi^{-1} : \Phi(\mathcal{S}) \to \mathcal{S}$ exists on the image. Therefore any representation that determines $\Phi(\mathcal{S})$ to accuracy ε determines $\mathcal{S}$ to accuracy at most $\|\Phi^{-1}\|_{\text{Lip}} \cdot \varepsilon$ (or exactly ε when Φ is isometric). The information content of $\mathcal{S}$ is fully encoded in $\Phi(\mathcal{S})$.

Step 2 (Smoothness and Theorem 1). By Definition 5(b), $\Phi(\mathcal{S})$ has analyticity parameter $\rho_\Phi > 1$ in the constructive basis $\mathcal{B}$. Apply Theorem 1(i) to $\Phi(\mathcal{S}) \in \mathcal{H}$: the truncation $\Lambda_N(\Phi(\mathcal{S}))$ to $N = \lceil \log(C/\varepsilon^2)/(2\log\rho_\Phi) \rceil$ coordinates per mode satisfies $\|\Phi(\mathcal{S}) - \Lambda_N(\Phi(\mathcal{S}))\| \le \varepsilon$.

Step 3 (Composition). Define $\Lambda(\mathcal{S}) := \Lambda_N(\Phi(\mathcal{S}))$. By Step 2, $\Lambda(\mathcal{S})$ determines $\Phi(\mathcal{S})$ to accuracy ε. By Step 1, $\Phi(\mathcal{S})$ determines $\mathcal{S}$. Therefore $\Lambda(\mathcal{S})$ is a sufficient finite representation of $\mathcal{S}$, of size $N = \Theta(\log(1/\varepsilon)/\log\rho_\Phi)$.

Step 4 (Rate is representation-dependent). Different sufficient smooth representations Φ_1, Φ_2 of the same system $\mathcal{S}$ may yield different rates $\rho_{\Phi_1} \neq \rho_{\Phi_2}$. The Extended Latent Theorem does *not* claim a basis-free rate for non-smooth systems — it claims that a finite rate *exists* whenever *any* sufficient smooth representation exists. The optimal rate $\rho^* = \sup_\Phi \rho_\Phi$ over all sufficient smooth representations is a property of $\mathcal{S}$, but computing or bounding it requires case-specific analysis. □

On closure under composition. If $\mathcal{S}_1$ admits a sufficient smooth representation Φ_1 with intrinsic rate $\rho^*_{\Phi_1} > 1$, and $\Phi_1(\mathcal{S}_1)$ admits a further Φ_2 with $\rho^*_{\Phi_2} > 1$, then $\Phi_2 \circ \Phi_1$ is sufficient and smooth *somewhere* (in the basis transported through Φ_2), hence has $\rho^* > 1$. A tight relation $\rho^*(\Phi_2 \circ \Phi_1) \ge f(\rho^*_{\Phi_1}, \rho^*_{\Phi_2})$ with explicit f requires Φ_2 to be Lipschitz-analytic in the first basis; we do not claim such a bound here. In practice the composition $\Phi_2 \circ \Phi_1$ inherits analyticity with a rate that must be re-estimated for each concrete (Φ_1, Φ_2) pair — closure is qualitative, not uniformly quantitative.

Remark. The weaker qualitative closure is what applications need: one chains "mollify → characteristic-function → Hermite-project" once and verifies $\rho^* > 1$ empirically (via the stability diagnostic of §3.6). A clean Lipschitz-analytic closure lemma is open and tracked as a companion-paper target.

Non-smooth system	Sufficient smooth representation Φ	Why $\Phi(\mathcal{S})$ is smooth
Jump process (Poisson, compound Poisson)	Characteristic function $\phi(u) = \mathbb{E}[e^{iuX}]$	Analytic in a strip for exponential-tailed jumps
Digital option (step function payoff)	Fourier transform of payoff	Analytic
ReLU network (piecewise-linear)	Smoothed predictive distribution $p(y \mid \theta) = \int f_\theta(x)\, p_{\mathcal{D}}(x)\, dx$ with Gaussian input $\mathcal{D}$, or NTK in the wide limit	Sufficient for **I/O statistics** (not for recovering θ; see caveat below); analytic in θ on bounded domains after Gaussian smoothing
Discrete distribution (atoms, no density)	Probability generating function $G(z) = \mathbb{E}[z^X]$	Analytic for bounded support
Rough volatility ($H < 1/2$)	Marginal density at fixed t	Smooth by **Malliavin calculus** for $H > 0$ (Hörmander bracket-generation generally fails here; the density regularity comes from the non-degeneracy of the Malliavin matrix, not hypoellipticity)

Empirical distribution (sum of deltas)	Characteristic function or kernel smoothing	Smooth by construction

The reach is large but not unconditional. Almost every system encountered in practice has a sufficient smooth representation — via characteristic functions, moment generating functions, resolvent operators, or regularization. The caveats matter:

- **"Sufficient" is representation-specific.** A ReLU network's smoothed predictive distribution is sufficient for input–output *statistics* but not for the parameter vector θ (many networks share the same output distribution). If the application target is the I/O law, the representation is sufficient; if the target is parameter recovery, it is not. The Extended Latent Theorem gives a Latent *for the target the representation is sufficient for*.
- **"Smooth" is representation-specific.** The marginal density of rough volatility is smooth by Malliavin; the characteristic function is analytic in a strip for exponential-tailed jumps; averaging over Gaussian inputs smooths ReLU kinks in expectation. The smoothness witness must be named for each case; we do not assert a universal smoothing bridge beyond mollification (§3.3 companion paper).
- **Measure-theoretic pathologies** (Weierstrass-type functions without any moment or Fourier structure, Kolmogorov-random sequences) resist all canonical smoothings. They are rare in applications but real in principle.

The Universal Smoothing Bridge (companion paper: Nagy 2026, "The Universal Smoothing Bridge")

The Extended Latent Theorem assumes a sufficient smooth representation Φ exists but does not construct one. For the fundamental case of *distributions* — the most important non-smooth objects in analysis — there is a canonical construction: **mollification**.

Every distribution $T \in \mathcal{D}'$ can be convolved with a mollifier $\varphi_\varepsilon \in C_c^\infty$ to produce a smooth function $T * \varphi_\varepsilon \in C^\infty$, which converges back to T as $\varepsilon \to 0$. Mollification is the canonical sufficient smooth representation for distributions: it is injective in the limit, information-preserving, and produces arbitrarily smooth outputs.

Combining mollification with Latent compression gives a **two-level pipeline** with a clean additive error bound:

$$d(T, \Lambda_N(T * \varphi_\varepsilon)) \leq \underbrace{\varepsilon}_{\text{smoothing}} + \underbrace{1/\rho(T * \varphi_\varepsilon)^N}_{\text{compression}}$$

The first term measures smoothing fidelity; the second measures Latent compression quality. The two knobs (ε and N) are independent. More smoothing increases ρ (the smoothed function is more compressible) but decreases fidelity. The practitioner has a concrete tradeoff to optimize.

The pipeline is **structure-preserving**: derivatives commute with the full pipeline at all orders ($\partial(\Lambda_N(T * \varphi_\varepsilon)) = \Lambda_N((\partial T) * \varphi_\varepsilon)$), the distributional limit is independent of mollifier choice, and the Dirac delta — the most singular standard distribution — passes through cleanly ($\partial(\Lambda_N(H * \varphi_\varepsilon)) = \Lambda_N(\varphi_\varepsilon)$, where H is the Heaviside step function).

This construction is formally verified: 35 theorems in the proof kernel (exportable to Lean 4), with 106 declarations and 0 errors. The full theory, including Fourier characterization, heat kernel smoothing, Colombeau distribution products, PDE existence, and Sobolev convergence rates, is in the companion paper.

3.4 The Latent of a Latent

The construction is recursive in two distinct ways.

Way 1: Smoothing then compressing. A non-smooth system $\mathcal{S}$ is mapped to a smooth representation $\Phi(\mathcal{S})$, which is then compressed to a finite Latent $\Lambda(\Phi(\mathcal{S}))$. The Latent of the smooth representation IS the Latent of the non-smooth system. The two-step composition $\Lambda \circ \Phi$ is itself a valid Latent extraction operator, even though the original system isn't smooth.

Way 2: Compressing the space of Latents. If you have many Latents $\{\Lambda_i\}$ from a family of systems (many patients, many assets, many models), these Latents themselves live on a smooth manifold in Latent space. The Latent OF THAT MANIFOLD is a **meta-Latent** — a compressed description of the population of compressed descriptions.

Way 3: The Latent Tower. When the first-level spectral representation is infinite ($\rho = 1$, no exponential decay), the infinite coefficient sequence $\{a_k\}_{k=0}^{\infty} \in \ell^2$ may itself possess structure — a *rule* that generates it. That rule is the Latent of the Latent: a finite object from which the infinite sequence is recoverable.

Level	Object	Dimension	Example
0	Continuous system $f \in L^2$	∞ (uncountable)	A function on $[0, 1]$
1	Spectral representation $\{a_k\} \in \ell^2$	∞ (countable)	Fourier coefficients
2	Generating rule	finite	(C, s) for $a_k = C/k^s$

The Level 1 representation is not a discretization or approximation — it is a *perfect* representation (Parseval's identity preserves all information). The compression happens at the transition from Level 1 to Level 2: the infinite sequence has a finite generating rule.

When does the tower terminate?

- **Smooth systems ($\rho > 1$):** At Level 1. The Latent Theorem (Theorem 1) directly gives a finite N-dimensional Latent.
- **Structured but non-smooth systems ($\rho = 1$):** At Level 2. The coefficients decay (e.g., polynomially $|a_k| \sim C/k^s$ for Sobolev-regular systems) but not exponentially, so infinitely many terms contribute. Yet the *decay law itself* is finitely parameterized — the rule (C, s) IS the finite Latent at Level 2.
- **Never (Kolmogorov-random sequences):** If the sequence has no computable generating rule, no finite Latent exists at any level. But such sequences have maximal algorithmic complexity — they are structureless by definition.

The critical insight: all known physical systems are finitely generated. The laws of physics are finite rules that produce infinite phenomena. Therefore every physical system has a finite Latent at *some* level of the tower, even when the first-level Latent is infinite.

This is not abstract speculation. It is the theoretical foundation of:

- **Transfer learning**: task-specific Latents share structure. The meta-Latent captures what's shared across tasks — the common representational backbone. Learning a new task = projecting onto the manifold of known task-Latents.

- **Meta-learning**: learn the manifold of task-Latents explicitly, then locate new tasks on it. MAML (Finn, Abbeel, and Levine 2017) and related algorithms are empirically approximating this meta-Latent without the formal framework.
- **Federated learning**: the grade-2 Latent (covariance of participant Latents) IS a Latent of the population of Latents. It tells the coordinator where participants agree and where they diverge.
- **Model taxonomy**: the manifold of Knowledge Artifacts has its own Latent, which classifies model families. Two models with nearby Latents belong to the same "species" — they encode similar knowledge regardless of architecture, training data, or parameter count. The meta-Latent of the model population IS a taxonomy of what AI systems know.
- **Dynamical systems without integration**: the parametric map IC $\mapsto \Lambda$(IC) from initial conditions to orbit Latents is smooth within a basin of attraction. Its meta-Latent Λ_F encodes the entire family of orbits as a single finite tensor. Evaluating Λ_F at a new IC yields the orbit Latent — no numerical integration required. This is developed explicitly for the three-body problem in Section 8.4, Proposition 3.

3.5 The Rank Bound Under Analyticity

A natural objection: computing tensor rank is NP-hard in general (Håstad 1990). Does this obstruct the Latent framework at grade ≥ 3?

No — because Latent tensors are not general tensors. The analyticity assumption imposes strong structure:

1. **Exponential coordinate decay**: $|\Lambda_k| \leq C\rho^{-|k|}$. Most entries are negligible.
2. **Finite truncation**: only $O(N^r)$ entries matter, where $N = O(\log(1/\varepsilon)/\log\rho)$ per mode.
3. **Effective rank bound**: within the N^r truncated tensor, the rank is at most N^{r-1} (the maximum rank of an $N \times N \times \cdots \times N$ tensor).

For distributions with rapidly decaying cumulants (sub-Gaussian, exponential family), the bound is much tighter:

Distribution class	Cumulant decay	Effective grade-r rank	Implication
Gaussian	$\kappa_r = 0$ for $r \geq 3$	0 for $r \geq 3$	Grade-2 Latent suffices
Sub-Gaussian	$\kappa_r = O(\sigma^r/r!)$	$O(1)$ for moderate r	Grade-3 correction is small
Heavy-tailed (finite moments to order p)	κ_r exists for $r \leq p$	Bounded by p	Need grades up to p
Crash-clustered (non-Gaussian copula)	Large κ_3, moderate κ_4	$O(K^2)$ where K = eigenvalue count	Grade-3 matters; grade-4 usually negligible

The NP-hardness of general tensor rank is irrelevant for the same reason that inverting a diagonal matrix is $O(n)$ despite matrix inversion being $O(n^3)$: structure kills complexity. The Latent's exponential decay structure makes its rank efficiently computable and bounded.

3.6 The Hierarchy Diagnostic: Self-Assessment of Latent Quality

The Latent Theorem (Theorem 1) guarantees a finite representation whenever $\rho > 1$. But the theorem contains a deeper message: the VALUE of ρ itself diagnoses how well the grade hierarchy works for a given system. This self-diagnostic property — the framework assessing its own effectiveness from within — is a structural feature not shared by most representation theories. A Fourier series does not tell you whether Fourier was the right basis. A neural network does not tell you whether it has enough layers. But ρ, computable from the system's data, simultaneously answers: (a) how well the current basis works, (b) how many grades are needed, (c) where in state space the hierarchy is strong and where it breaks down, and (d) whether a fundamentally different approach is needed.

The Three Regimes

The analyticity parameter ρ partitions all systems into three qualitatively distinct regimes:

Regime I: Strong hierarchy ($\rho \gg 1$). Grade-2 captures nearly everything. The Latent converges with very few coefficients per mode: $N^* = \lceil \log(1/\varepsilon)/\log \rho \rceil$ is small. Grade-3 corrections are exponentially suppressed — they contribute $O(\rho^{-3})$ relative to grade 2. The system behaves as if it were "almost Gaussian" or "almost integrable."

System	ρ	Implication
Two-body Keplerian orbit	≈ 8	Ellipse is a grade-2 object; perturbative corrections are tiny
Gaussian financial model	$\to \infty$	All cumulants $\kappa_r = 0$ for $r \geq 3$; covariance IS the complete Latent
Near-equilibrium thermodynamics	$\gg 1$	Linear response is grade 2; nonlinear corrections exponentially small
Harmonic oscillator	$\to \infty$	Exact quadratic Hamiltonian; grade-2 is exact

Regime II: Weak hierarchy ($1 < \rho \lesssim 3$). Grade-2 captures the dominant structure but grade-3 is NOT negligible. The Latent requires many coefficients: N^* is large. The hierarchy is real but shallow — truncation at any finite grade leaves a significant residual. This is the regime of most *interesting* systems: the ones where the structure is rich enough to matter but regular enough to describe.

System	ρ	Implication
Three-body figure-8 orbit	≈ 2.3 (kinematic, Fourier-decay of the trajectory $q(t)$)	Grade-2 captures $\sim 85\%$; the co-skewness $T^{(3)}$ is precisely what makes the problem hard
Fat-tailed financial returns	≈ 1.5–3	Skewness and kurtosis are grade-3/4 effects dominating risk
Transitional turbulence	≈ 1.5–2	Mean flow (grade 2) captures the base state; Reynolds stress (grade 3) drives instability

System	ρ	Implication
Weakly nonlinear waves	$\approx$ 2–5	Leading nonlinear corrections at grade 3 are perturbative but physically essential

Regime III: Hierarchy collapse ($\rho \to 1$ or $\rho < 1$). The grade hierarchy provides no compression. ALL grades contribute comparably — truncation at any finite grade fails qualitatively. The system is at or beyond the boundary of analyticity.

System	ρ	Implication
Fully developed turbulence (high Re)	$= 1$	Energy cascade is scale-free; Kolmogorov $k^{-5/3}$ is power-law, not exponential
Critical phenomena at phase transitions	$= 1$	Correlation length diverges; fluctuations at all scales
Collision singularities (N-body)	< 1	Latent coordinates grow; expansion diverges
Fractal attractors	$= 1$	Self-similar at all scales; no spectral decay
Fully chaotic ergodic systems	$\to 1$	No separation of scales; every mode matters equally

Systems with $\rho < 1$ are NOT analytic in the observed domain. Their Latent coordinates GROW rather than decay — the expansion diverges. This does not mean they have no Latent; it means the current extraction basis and domain fail, and one must either:

(a) **Change the extraction basis** (Theorem 4, §4.4): the matched basis may restore $\rho > 1$. The difficulty of the lognormal sum problem was $\rho \to 1$ in the moment basis but $\rho \gg 1$ in the Hermite basis (companion paper: Nagy 2026, *The Fenton Distribution Solved* / *The Hermite Latent*). This is a canonical Latent instance: the distribution of $S = \sum w_i e^{Y_i}$, $Y \sim N(\mu, \Sigma)$, is generated by a finite parametric latent $(w, \mu, \Sigma) \in \mathbb{R}^{O(n^2)}$, and its Hermite-chaos coordinates $\{c_{\mathbf{k}}^H\}$ form an infinite sequence in ℓ^2 whose image is a finite-dimensional submanifold. The 65-year difficulty was a convergence problem: the moment basis produces a divergent encoding of this finite latent (MGF has zero convergence radius), while the Hermite-COS basis produces a convergent one. The Latent itself — the finite parameter set — was always finite; only the encoding diverged. Moreover, the pipeline has a clean algebraic structure: $S(\cdot) \in L^2(\gamma_n)$ is a grade-1 Latent, the COS coefficients $A_k = \langle \cos(\ldots \log S \ldots), \mathbf{1} \rangle_{\gamma_n}$ are grade-0 Latents obtained by contraction (the Gaussian inner product), and the CDF is synthesized from these grade-0 Latents by trigonometric series. Every step is an algebraic operation in the Latent hierarchy — maps, contractions, synthesis. The precision parameters belong to the evaluator (GH quadrature for the contraction, COS truncation for the synthesis), not to the algebra, which is exact. The enabling move is *complexification*: the characteristic function $\phi_S(t) = E[e^{itS}]$ extends the distribution from $\mathbb{R}$ to $\mathbb{C}$ (Euler's $e^{itS} = \cos(tS) + i \sin(tS)$), and the COS coefficients are real projections of this complex object. For the analytic density of a lognormal sum, the Fourier series converges absolutely — the COS formula *expresses* the CDF in its Fourier coordinates rather than approximating it. This is structurally identical to the complexification in Finding 14 (Smale's 6th Problem): extend $\mathbb{R} \to \mathbb{C}$ to reveal algebraic structure invisible in the real domain.

(b) **Apply a sufficient smooth representation** (Theorem 2, §3.3): the characteristic function of a jump process is analytic even though the process itself has discontinuous paths; the generating function of a discrete distribution is analytic even though the distribution has no density.

(c) **Regularize the domain**: blow-up transformations (McGehee), conformal mappings (Levi-Civita), or desingularization procedures can transform a domain where $\rho < 1$ (collision, shock) into one where $\rho > 1$ — at the cost of changing the coordinate system.

(d) **Accept that the system lacks hierarchical structure** in the given domain and requires non-perturbative treatment: resummation, renormalization group methods, or exact solutions.

The boundary $\rho = 1$ is the **analyticity boundary** of the Latent framework. It is the representation-theoretic phase transition: on one side ($\rho > 1$), finite truncation works and the hierarchy provides exponential compression; on the other ($\rho < 1$), no finite truncation suffices and the hierarchy is an organizing principle without compression power. The boundary itself ($\rho = 1$ exactly) is where power-law behavior lives — scale invariance, critical phenomena, turbulence cascades — the richest and hardest systems in mathematical physics.

Noise as a Property of the Representation

The three-regime classification reveals a structural principle: **noise is not intrinsic to a system but is a property of the observer's representation.** What appears as irreducible randomness at grade-k may be deterministic dynamics at grade-$(k+1)$. In quantitative finance, the Brownian motion in an SDE captures not fundamental randomness but the aggregate effect of unmodeled degrees of freedom (other actors, information flow, order book microstructure) projected onto the observable space. In celestial mechanics — a fully deterministic system — observing a single body while ignoring the others produces trajectories that appear chaotic and stochastic; the "noise" is the missing bodies.

This reframes the diffusion term $\sigma\, dW_t$ in any SDE as an **explicit admission that the grade-k representation is incomplete**: the Brownian term models the grade-$(k+1)$+ residual. The value of ρ then diagnoses whether this admission is honest — high ρ means the residual truly lacks structure (the Brownian approximation is justified, as in thermal noise), while low ρ means the residual contains structured signal that the observer's representation fails to capture (as in fat-tailed financial returns). This is the operational meaning of the self-diagnostic property applied to stochastic models: ρ measures the quality of the observer's lens, not the complexity of the underlying reality.

A concrete instance: the Riemannian mechanics tradition (Jacobi, Maupertuis, Arnold) routinely absorbs quadratic structure into the metric — e.g., Montgomery's mass inner product $\langle v, w\rangle_M = \sum m_i(v_i \cdot w_i)$ for the N-body problem, which transforms $M\ddot{q} = \nabla U$ into the clean form $\ddot{q} = \nabla U(q)$. This is a grade-1 Latent operation: the eigendecomposition of the coupling matrix (M for masses, Σ for covariance in finance) that decouples the quadratic structure, leaving the nonlinear residual (U) exposed in pure form. The tradition performs this decomposition implicitly; the Latent framework names the grades and measures what remains (ρ).

The Practical ρ-Test: Three Methods

Given a real-world system — data, a simulation, an observed process — how does one measure ρ and diagnose the hierarchy? Three complementary methods, applicable in different situations:

Method 1: Spectral analysis (direct). If the system has a generator, covariance operator, transfer matrix, or any spectral representation, its eigenvalues $\{\lambda_k\}$ ordered by magnitude contain ρ directly.

Procedure. Compute the spectrum $\{|\lambda_k|\}$, ordered as $|\lambda_1| \geq |\lambda_2| \geq \cdots$. Plot $\log|\lambda_k|$ against k. The slope of the log-linear envelope is $-\log\rho$:

$$\hat{\rho} = \lim_{k\to\infty} \left(\frac{|\lambda_k|}{|\lambda_1|}\right)^{-1/k}$$

In practice, fit a line to $\log|\lambda_k|$ for $k = 2, 3, \ldots, K$ and read $\hat{\rho} = \exp(-\text{slope})$. The shape of the log-spectrum is diagnostic:

- **Steeply linear** (slope $\ll -1$): strong hierarchy, $\rho \gg 1$. Grade-2 truncation is safe.
- **Gently linear** (slope ≈ -0.5 to -1): weak hierarchy, $\rho \approx 1.5$–3. Grade-3 matters.
- **Flat or upward-curving**: hierarchy collapse, $\rho \to 1$ or $\rho < 1$. Truncation fails.
- **Log-log linear** (power-law decay): scale-free system, $\rho = 1$ exactly. At the boundary.

The source of the spectrum depends on the system type:

System type	Spectrum source	What ρ measures
Dynamical system	Fokker–Planck generator eigenvalues	Mixing rate hierarchy
Data matrix ($n \times d$)	SVD singular values σ_k	Low-rank compressibility
Time series	Fourier coefficients $\lvert c_k\rvert$	Temporal regularity
Stochastic process	Karhunen–Loève eigenvalues	Variance concentration
Neural network layer	Weight matrix singular values	Parameter redundancy
N-body orbit	Floquet multipliers of Poincaré map	Stability hierarchy

Method 2: Residual analysis (iterative). When direct spectral computation is impractical — because the system is high-dimensional, the generator is unknown, or only trajectory data is available — one can estimate ρ from the rate at which successive grade truncations improve accuracy.

Procedure.

1. Fit a **grade-2 model** to the system. This means: linear regression, Gaussian fit, covariance-based model, pairwise interaction approximation — any model that captures only two-point structure. Measure the residual $r_2 = \|\text{data} - \text{model}_2\|$.
2. Fit a **grade-3 model** to the residual. This means: add cubic terms, three-point correlations, coskewness corrections, or any three-body interaction. Measure $r_3 = \|\text{data} - \text{model}_3\|$.
3. (Optionally) Fit a **grade-4 model** and measure r_4.

The ratios estimate the analyticity parameter:

$$\hat{\rho} \approx \left(\frac{r_2}{r_3}\right)^{1/(N_3 - N_2)}$$

where N_k is the number of parameters in the grade-k model. If each additional grade reduces the residual by a roughly constant factor, the hierarchy is working and ρ is approximately that factor. Specific diagnostics:

- $r_3/r_2 \ll 1$: strong hierarchy ($\rho \gg 1$). Grade-2 already captured most of the signal; grade-3 mops up the rest.
- $r_3/r_2 \approx 0.3$–0.8: moderate hierarchy ($\rho \approx 1.5$–3). Each grade helps but does not solve the problem.

- $r_3/r_2 \approx 1$: hierarchy failure ($\rho \to 1$). Adding grade 3 produced no improvement — either the system has no grade-3 structure (already Gaussian) or the grade-3 extraction was poorly chosen.
- $r_3 > r_2$: overfitting or wrong model class. The grade-3 ansatz introduced noise rather than signal.

Method 3: Coefficient decay analysis (representation-specific). For systems with a natural expansion — distributions, fields, periodic orbits — directly measure the decay rate of the expansion coefficients.

Procedure. Expand the system in a chosen basis $\mathcal{B}$ (Fourier, Hermite, Chebyshev, Laguerre, spherical harmonics) and measure the coefficients $c_k^{(\mathcal{B})}$. The analyticity parameter in that basis is:

$$\rho_{\mathcal{B}} = \liminf_{k\to\infty} |c_k^{(\mathcal{B})}|^{-1/k}$$

In practice, plot $\log|c_k|$ against k and fit the envelope. For specific representations:

- **Moments**: if cumulants satisfy $|\kappa_r| \leq C \cdot \rho^{-r}$, the system has ρ-analyticity. Gaussian: $\kappa_r = 0$ for $r \geq 3$ ($\rho = \infty$). Student-t with ν degrees of freedom: cumulants grow factorially, the moment expansion diverges and $\rho = 1$ (power-law boundary; the MGF is undefined, so there is no analytic strip around the origin).
- **Characteristic function**: if $\varphi(u) = \mathbb{E}[e^{iuX}]$ extends analytically to a strip $|\mathrm{Im}(u)| < a$ in the complex plane, then $\rho \geq e^a$ for the Fourier basis.
- **Fourier series of periodic orbits**: the decay rate of Fourier coefficients of $q(t)$ gives ρ directly — this is exactly how $\rho \approx 2.3$ was measured for the figure-8 orbit (§8.4).

Summary: The Self-Diagnostic Table

Observable	What to measure	Diagnosis	Action
$\rho \gg 1$	Steep spectral decay	Grade-2 sufficient	Truncate at grade 2
$1 < \rho < 3$	Moderate spectral decay	Grade-3 matters	Include co-skewness; bound error
$\rho \to 1$	Flat spectrum or power-law	All grades contribute	Resum, change basis, or regularize
$\rho < 1$	Growing coefficients	Expansion diverges	Apply Theorem 2 (smooth representation)
k_{eff} large at point x	Local ρ gradient	Higher grades needed locally	Adaptive local refinement
k_{eff} small everywhere	Uniform large ρ	Simple system globally	Low-grade truncation everywhere
k_{eff} varies widely	Strong spatial ρ variation	Mixed regime	Adaptive grade truncation
$\tilde{M} \neq M$ (§3.8)	Gap between bare and dressed generator	Grade feedback active	Use renormalized generator

The framework is **self-diagnosing**: the same mathematical object (ρ) that controls the representation's quality is computable from the representation itself. No external oracle is needed. The diagnostic IS a function of the data.

The Analyticity Boundary

The boundary $\rho = 1$ deserves extended discussion because it is not merely a technical condition — it is the **structural limit of hierarchical representation**.

At $\rho = 1$:

- The truncation bound $N^* = \log(1/\varepsilon)/\log\rho \to \infty$: infinitely many coefficients are needed for any finite accuracy.
- The grade hierarchy does not truncate: every grade contributes comparably to the system's structure.
- The effective grade $k_{\text{eff}} = \lceil \log(C/\varepsilon)/\log\rho \rceil \to \infty$: no finite grade suffices.
- The system is at a **phase transition in representation complexity** — the analog of a thermodynamic critical point for the description itself.

Convergence at the boundary. While the truncation bound N^* diverges, convergence does not cease — it transitions from exponential to algebraic. If the system has Sobolev regularity $s > 1$ at the boundary (coefficients decaying as $|c_k| \leq C_f k^{-s}$), then $N(\varepsilon) = \Theta(\varepsilon^{-1/(s-1)})$. Three sub-cases: generic ($s = 2$) gives $N \sim 1/\varepsilon$; Lipschitz boundary ($s = 1$) introduces a logarithmic correction $N \sim \varepsilon^{-1}\log(1/\varepsilon)$; C^∞ but non-analytic ($s \to \infty$) gives $N = O(\varepsilon^{-1/s})$ for all s but never reaches exponential. The transition between exponential ($\rho > 1$) and algebraic ($\rho = 1$) convergence classes is **discontinuous**: for any $\delta > 0$, the ratio $N_{1+\delta}(\varepsilon)/N_1(\varepsilon) \to 0$ as $\varepsilon \to 0$. The geometric mechanism is Bernstein ellipse degeneration: the semi-minor axis $b(\rho) = \frac{1}{2}(\rho - \rho^{-1})$ vanishes at $\rho = 1$, collapsing the analyticity strip to zero width. (Companion paper: *Spectral Phase Transition*; proof kernel: `spectral_phase_transition/platonic.py`, 53 theorems + `ec_bridge.py`, 8 cross-domain bridge theorems, 0 sorry.)

The systems that live at this boundary are among the most studied and hardest in mathematical physics:

- **Critical phenomena** (Ising model at T_c, percolation at p_c, etc.): correlation functions decay as power laws $\sim r^{-\eta}$, not exponentials. The critical exponents $\eta, \nu, \beta, \ldots$ encode the *rate at which the boundary is approached* as parameters are tuned. Wilson's renormalization group is the technology for analyzing systems at $\rho = 1$.
- **Kolmogorov turbulence**: the energy spectrum $E(k) \sim k^{-5/3}$ is a power law — the signature of $\rho = 1$ in Fourier space. Every wavenumber carries comparable energy relative to its neighbors. The cascade transfers energy across all scales without a characteristic scale — the definition of hierarchy collapse.
- **Fractal attractors**: the Hausdorff dimension d_H of a strange attractor is a measure of how far the system is from $\rho = 1$. A smooth attractor has $d_H = n$ (integer) and $\rho > 1$. A fractal attractor has non-integer d_H and $\rho = 1$ — self-similarity at all scales means no spectral decay.
- **Collision singularities**: in the N-body problem, as bodies approach collision ($r_{ij} \to 0$), the potential $U \to \infty$ and the generator's eigenvalues diverge. McGehee's blow-up coordinates regularize the singularity — effectively mapping the $\rho < 1$ neighborhood of collision into a $\rho > 1$ desingularized space. The regularization IS a change of domain that restores analyticity.

But the boundary is not a wall — the Extended Latent Theorem (§3.3) provides an escape. Even when $\rho \leq 1$ for the system directly, a **sufficient smooth representation** Φ may have $\rho_\Phi > 1$:

- A jump process has $\rho < 1$ in path space (discontinuous trajectories) but $\rho_\Phi > 1$ for its characteristic function (analytic in a strip).
- A discrete distribution has $\rho = 0$ as a density (no density exists) but $\rho_\Phi > 1$ for its probability generating function.

- An empirical distribution (sum of Dirac deltas) has $\rho = 0$ as a measure but $\rho_\Phi > 1$ for its kernel-smoothed density or characteristic function.

The framework's true boundary is therefore not "is the system smooth?" but **"does the system admit a smooth representation?"** — and this is almost always answerable in the affirmative. The genuine exceptions are pathological: measure-theoretic constructions (Cantor-singular distributions, nowhere-differentiable processes with no analytic characteristic function) that do not arise in applications.

The hierarchy of scope:

$$\{\text{smooth systems}\} \subset \{\text{systems with smooth representations}\} \subset \{\text{all systems}\}$$

is strict, and the Latent framework covers the middle set — which, for practical purposes, is everything.

The Simulation Boundary: ρ Controls Monte Carlo Efficiency

The analyticity boundary has a second, unexpected manifestation: it determines whether **rare events** in a system can be efficiently simulated.

Consider a portfolio of d assets with correlation matrix C. Its grade-2 Latent has eigenvalues $\{\lambda_k\}$ and modes $\{v_k\}$. Two classical computational tasks are:

1. **Density approximation** (the COS method): approximate the portfolio loss distribution using Fourier-cosine coefficients. The convergence rate is controlled by the Bernstein ellipse parameter ρ_{COS} — the strip in the complex plane to which the characteristic function extends analytically.
2. **Rare event simulation** (importance sampling): estimate $P(L > \ell)$ for a large loss threshold ℓ by tilting the sampling distribution. The optimal tilt parameter is the Cramér rate function, which depends on the strip width to which the moment generating function extends — namely ρ_{IS}.

The Analyticity–Rate Duality states that these are the **same parameter**:

$$\rho_{\text{COS}} = \rho_{\text{IS}} = \rho$$

More precisely: the Bernstein ellipse radius that governs exponential convergence of spectral coefficients $|A_k| \leq C \cdot \rho^{-k}$ is the same ρ that governs exponential variance reduction in importance sampling: $\text{VR}(\ell) \sim \rho^{2\ell}$. Moreover, when the portfolio has a Latent Eigen-COS decomposition, the optimal d-dimensional saddle-point problem factorizes into K independent one-dimensional problems — one per eigenmode — each controlled by its eigenvalue λ_k.

This creates a sharp operational trichotomy:

Regime	ρ	COS convergence	IS efficiency	Example
Strong hierarchy	$\gg 1$	Exponential: few coefficients suffice	Exponential VR: rare events efficiently simulable	Gaussian portfolios
Weak hierarchy	≈ 1–3	Moderate: many coefficients needed	Moderate VR: IS helps but isn't magical	Fat-tailed (NIG) portfolios

Regime	ρ	COS convergence	IS efficiency	Example
Collapse	$= 1$	Power-law: no exponential convergence	No exponential VR: IS cannot beat naive MC	Student-t ($\nu \leq 4$) portfolios

The collapse case is particularly revealing. A multivariate Student-t distribution with $\nu = 4$ degrees of freedom has $\rho = 1$ because its moment generating function does not exist. Applying the same mode-factored importance sampling that gives roughly $1.7\times10^4\times$ variance reduction in our Gaussian benchmark (see benchmark spec below) gives VR < 1 (IS is *counterproductive*) for Student-t — the mixing variable that generates the heavy tails couples all modes, destroying the independence that the spectral IS exploits.

Benchmark specification (underlying the "$1.7 \times 10^4\times$ variance reduction" figure cited here and in the abstract). *System:* $d = 10$-dimensional Gaussian equity portfolio with mild correlation ($\rho_{ij} = 0.3$, $i \neq j$) and unit variances; tail event $\{L > \ell_{0.999}\}$ at the 0.1%-tail threshold. *Latent structure:* 5 effective Eigen-COS modes capture 99.9% of the variance; $\rho \approx 6$ measured from the decay $|A_k| \leq C\rho^{-k}$ of the marginal COS coefficients (Hermite basis, $k \leq 12$). *Baseline:* naive Monte Carlo with 10^4 paths. *IS method:* mode-factored exponential tilting at the saddle point of the marginal cumulant generating function, same 10^4 paths. *Metric:* variance of the indicator estimator $\widehat{P}[L > \ell_{0.999}]$ over 100 independent runs. *Result:* $\widehat{\mathrm{VR}} = \mathrm{Var_{MC}}/\mathrm{Var_{IS}} = 1.72 \times 10^4$, with 95% CI $[1.3, 2.1] \times 10^4$ (bootstrap over runs). *Robustness:* repeating the experiment at $d = 20$ and $d = 50$ gives $\widehat{\mathrm{VR}} \in [1.2, 1.9] \times 10^4$ — variance reduction is weakly dimension-dependent once the $\rho \approx 6$ spectral structure is preserved, consistent with Theorem 1's dimension independence. The companion paper (Nagy 2026, *The Spectral Fenton Distribution*) supplies the full replication code; this paper reports the headline numbers only.

The duality reveals that computing a density and simulating a rare event are **dual projections of the same analytic structure**: the strip of analyticity of the characteristic function. The COS method reads this strip for approximation; importance sampling reads it for simulation. Both fail at the same boundary $\rho = 1$, for the same reason — the analytic continuation that both methods rely on ceases to exist.

3.7 The Three Faces of ρ: Algebraic, Computational, and Information-Theoretic

The analyticity parameter ρ was introduced algebraically in §3.1 as a rate governing spectral decay. The Latent Theorem (§3.2) established it computationally as the parameter controlling representation size. It now admits a third interpretation — information-theoretic — that unifies the first two and reveals ρ as a deeper invariant than originally apparent.

The Algebraic Face

Definition. $\rho_{\mathcal{B}}(\Lambda) = \liminf_{|k|\to\infty} |\Lambda_k^{(\mathcal{B})}|^{-1/|k|}$. This measures the *exponential decay rate* of Fourier-type coefficients in basis $\mathcal{B}$. Higher ρ means faster decay, meaning the function is "more analytic" — its spectral representation concentrates on fewer modes.

The algebraic face answers: **how fast do the coefficients shrink?**

The Computational Face

Theorem 1 (restated). $N^* = \Theta(\log(1/\varepsilon)/\log\rho)$. The number of parameters per mode needed for ε-accuracy depends on ρ logarithmically. This is the Latent Theorem's primary message: ρ controls *how compressible* the system is.

The computational face answers: **how many numbers do I need?**

The Information-Theoretic Face

The Simultaneous Field framework (Nagy 2026, *The Knowledge Artifact and Knowledge Algebra of ML Models*) introduces a third perspective. Represent the full function f as a simultaneous element — a non-negative weight field w over the value space $L^2(\Omega)$, with $w(x) \geq 0$ representing how strongly each possible value x participates. Before any extraction, w is spread over all of L^2 (maximum entropy). Each spectral mode added to the truncation is a *crystallization step* $\mathcal{K}_{C_k}$ — a constraint that concentrates the weight field on the subset C_k compatible with the k-th mode.

The central result (Theorem 6.1 of that paper):

$$H(\mathcal{K}_{C_1} \circ \cdots \circ \mathcal{K}_{C_N}(w)) = H(w) - N \cdot \log \rho + o(N)$$

Each spectral mode removes exactly $\log \rho$ bits of entropy from the simultaneous element. The cumulative entropy removed after N modes is $N \cdot \log \rho$, and the residual "spread" (error of the truncation) is $\exp(-N \cdot \log \rho) = \rho^{-N}$ — recovering the exponential decay of Theorem 1(i).

The information-theoretic face answers: **how much uncertainty does each mode eliminate?**

The Unification

The three faces are one number:

Face	Question	Answer in terms of ρ
Algebraic	How fast do coefficients decay?	$\lvert\Lambda_k\rvert = O(\rho^{-\lvert k\rvert})$
Computational	How many parameters for ε-accuracy?	$N^* = \Theta(\log(1/\varepsilon)/\log \rho)$
Information-theoretic	How many bits per spectral constraint?	$\Delta H = \log \rho$ per mode

The unification is not merely notational — it reveals *why* the Latent Theorem works. The exponential convergence of Theorem 1(i) holds because each spectral constraint removes a *fixed* amount of information. This is the crystallization-rate interpretation: ρ measures how efficiently the function's degrees of freedom crystallize out of superposition when spectral modes are imposed one at a time.

Remark (The Reynolds Number of Representation). In §3.6, we noted that $\rho = 1$ is a phase boundary separating compressible from incompressible systems. The information-theoretic face explains *why*: at $\rho = 1$, $\Delta H = \log 1 = 0$ — each spectral mode removes zero entropy. No amount of mode-adding can crystallize the field. The system is informationally opaque to spectral methods, just as a fluid at $\mathrm{Re} = \infty$ is dynamically opaque to laminar description. The analogy extends further: the Simultaneous Field framework shows that for Navier-Stokes flows, the Reynolds number itself functions as an effective constraint density, with laminar-turbulent transition corresponding to a crystallization phase transition at a critical Re_c (see companion paper, §8).

3.8 The Renormalized Grade: Feedback Across the Hierarchy

The grade hierarchy (§2.4) is defined algebraically: grade-r components live in $\mathcal{H}^{\otimes r}$, and the decomposition $\Lambda = \Lambda^{(1)} + \Lambda^{(2)} + \Lambda^{(3)} + \cdots$ is unique. The subspaces $\mathcal{H}^{\otimes r}$ are mutually orthogonal in the graded Hilbert tensor algebra. No algebraic operation maps information from grade 3 INTO grade 2 — they are disjoint components of a direct sum.

But systems *evolve*. And the equations of motion, in general, couple all grades. This section addresses the question: **can higher grades "write back" to lower grades, and if so, what does this do to the hierarchy?**

The Bare Hierarchy Is Strict

In the **bare** (algebraic, static) Latent, the decomposition is clean:

$$\Lambda = \bigoplus_{r=0}^{\infty} \Lambda^{(r)}, \qquad \Lambda^{(r)} \in \mathcal{H}^{\otimes r}, \qquad \langle \Lambda^{(r)}, \Lambda^{(s)} \rangle = 0 \text{ for } r \neq s$$

Given a fixed Latent Λ, its grade-2 projection $\Lambda^{(2)}$ is a well-defined, unique, grade-3-independent object. The grade-2 generator M, the grade-3 co-skewness $T^{(3)}$, and all higher components are algebraically separate. Extracting grade 2 requires no knowledge of grade 3, and vice versa. This is the *static* picture, and it is mathematically rigorous.

The Dynamical Picture: Grades Interact Through Evolution

The generator $\mathcal{L}$ of a dynamical system — the Fokker–Planck operator for a diffusion, the Liouville operator for a Hamiltonian system, the Koopman operator for a discrete map, the loss landscape gradient for a neural network — maps states to states. In general, $\mathcal{L}$ does NOT respect the grade decomposition. The grade-r projection of $\mathcal{L}$ acting on the full Latent depends on ALL grades:

$$\frac{d\Lambda^{(2)}}{dt} = P^{(2)}\mathcal{L}\big(\Lambda^{(2)} + \Lambda^{(3)} + \Lambda^{(4)} + \cdots\big)$$

where $P^{(2)} : \mathfrak{L}(\mathcal{H}) \to \mathcal{H}^{\otimes 2}$ is the grade-2 projection. The co-skewness tensor $T^{(3)}$ modifies the time evolution of the grade-2 component. The effect can be absorbed into an **effective** or **dressed** generator:

$$\frac{d\Lambda^{(2)}}{dt} = \tilde{\mathcal{L}}^{(2)}\,\Lambda^{(2)}, \qquad \tilde{\mathcal{L}}^{(2)} = \mathcal{L}^{(2)} + \delta\mathcal{L}^{(2)}[\Lambda^{(3)}, \Lambda^{(4)}, \ldots]$$

where $\delta\mathcal{L}^{(2)}$ absorbs the feedback from all higher grades into an effective grade-2 correction. An observer measuring only grade-2 quantities would infer the dressed generator $\tilde{\mathcal{L}}^{(2)}$, not the bare $\mathcal{L}^{(2)}$ — and the difference $\delta\mathcal{L}^{(2)}$ is a direct, measurable signature of higher-grade structure.

Definition

Definition 6 (Renormalized grade). The *renormalized* (or *dressed*) grade-k generator is:

$$\tilde{M}^{(k)} = M^{(k)} + \delta M^{(k)}, \qquad \delta M^{(k)} = \sum_{r>k} \Pi_k[\Lambda^{(r)}]$$

where $\Pi_k : \mathcal{H}^{\otimes r} \to \mathrm{End}(\mathcal{H}^{\otimes k})$ is the **dynamically induced contraction** — the map from grade-r Latent components to their effective contribution to grade-k dynamics, mediated by the equations of motion.

The renormalized Latent at grade k:

$$\tilde{\Lambda}^{(k)} = \Lambda^{(k)} + \sum_{r>k} \Pi_k[\Lambda^{(r)}]$$

is not an element of $\mathcal{H}^{\otimes k}$ alone — it carries information from all higher grades, *projected* onto grade-k observables through the dynamics. The renormalized Latent is what an experimentalist restricted to grade-k measurements actually observes.

Physical Examples

The renormalized grade is not an abstraction — it is the mechanism behind well-known phenomena:

Example 1: Orbital precession (celestial mechanics). Mercury's orbit in the Sun's gravitational field, treated as an isolated two-body problem, is a grade-2 object: the Kepler ellipse, determined entirely by the pairwise generator $M^{(2)}$. But the other planets create grade-3 perturbations through their gravitational coupling to Mercury and the Sun. The three-body tensor $T^{(3)}_{\text{Mercury-Sun-Jupiter}}$ contributes a secular precession of Mercury's perihelion.

An observer measuring only Mercury's orbital elements would infer a dressed generator $\tilde{M}^{(2)}$ whose eigenfrequencies include the precession — which is a grade-3 contribution absorbed into a grade-2 observable. The classical Newtonian precession ($\approx 532''$/century from planetary perturbations) plus the general relativistic correction ($\approx 43''$/century, which IS a genuine grade-2 modification from spacetime curvature) together constitute the dressed grade-2 dynamics. Le Verrier's 1859 discovery of the anomalous precession was, in Latent language, the detection that $\tilde{M}^{(2)} \neq M^{(2)}_{\text{Newtonian}}$ by a margin not attributable to known grade-3 contractions.

Example 2: Financial contagion and the VaR failure. A stock's variance (grade-2 observable) increases during a market crash. The increase is not caused by a change in the stock's intrinsic volatility dynamics — it is caused by the activation of the co-skewness tensor $T^{(3)}_{ijk}$ coupling the stock to market-wide panic. The dressed variance:

$$\tilde{\sigma}_i^2 = \sigma_i^2 + \sum_{j,k} \Pi_2[T^{(3)}_{ijk}] \cdot \Delta_j \Delta_k$$

absorbs the three-body crash coupling into an effective pairwise quantity. This is precisely why Value-at-Risk models based on grade-2 (covariance) failed catastrophically in 2008: they modeled the bare $M^{(2)}$ but the system evolved according to $\tilde{M}^{(2)}$, and $\delta M^{(2)}$ was large exactly when it mattered most — during the crisis. The grade-2 model was not wrong about grade-2 structure; it was blind to the grade-3 feedback that *modified* the effective grade-2 behavior.

Example 3: Running coupling constants (quantum field theory). The electric charge measured at low energy ($e_{\text{eff}}(q^2 \to 0) \approx 1/137$) differs from the bare charge (e_0) because virtual particle-antiparticle pairs (grade-3+ effects in the interaction Lagrangian) screen the charge. The measured charge is the renormalized one: $e_{\text{eff}} = e_0 + \delta e[e_0, \Lambda^{(3)}, \Lambda^{(4)}, ...]$. In QED, this running is logarithmic; in QCD, it is the origin of asymptotic freedom. The Latent framework's grade hierarchy maps directly: the bare coupling is the grade-2 generator; vacuum polarization is the grade-3 contraction $\Pi_2[\Lambda^{(3)}]$; and the running with energy scale is the renormalization flow of $\tilde{M}^{(2)}(q^2)$ as the effective observation scale changes.

Example 4: Neural network effective Hessian. The loss landscape's Hessian (grade-2 curvature) at a point in weight space depends not only on the local quadratic structure but on the entire training trajectory — including third-order interactions between learning rate, batch composition, and weight correlations. The

"effective Hessian" that governs whether SGD converges, oscillates, or escapes a local minimum is a renormalized grade-2 object: $\tilde{H} = H + \delta H$[grad-3 interactions]. This is why second-order optimizers (Newton, natural gradient) that use the bare Hessian H can fail in practice — they ignore the grade-3 dressing that modifies the effective curvature.

The Self-Consistency Condition

The renormalization is not arbitrary. The dressed generator $\tilde{M}$ must satisfy a self-consistency condition: the grade-3 Latent whose contraction produces δM must be compatible with the dynamics generated by $\tilde{M}$ itself. Formally:

$$\tilde{M} = M + \Pi_2[\Lambda^{(3)}(\tilde{M})]$$

This is a **fixed-point equation**: $\tilde{M}$ determines the effective dynamics, which determine the grade-3 Latent $\Lambda^{(3)}(\tilde{M})$, which renormalizes $\tilde{M}$ via Π_2. The equation asks: what is the grade-2 generator such that its own grade-3 feedback, projected back to grade 2, reproduces itself?

Existence of fixed points. For strongly hierarchical systems ($\rho \gg 1$), the contraction $\Pi_2[\Lambda^{(3)}]$ is exponentially small: $\|\delta M\| \leq C \cdot \rho^{-3}\|M\|$. The fixed-point map is a contraction mapping, and the Banach fixed-point theorem guarantees a unique solution close to M. The perturbative expansion:

$$\tilde{M} = M + \delta M^{(1)} + \delta M^{(2)} + \cdots, \qquad \|\delta M^{(n)}\| = O(\rho^{-3n})$$

converges geometrically. This is the regime where perturbation theory works.

For weakly hierarchical systems ($\rho \to 1$), the corrections $\delta M^{(n)}$ do not decay, the perturbative series diverges, and the fixed-point equation may have multiple solutions, no solution, or a solution far from M. The failure of the self-consistency fixed point IS a diagnostic of hierarchy breakdown — it signals that the bare grade-2 generator is a poor starting point for understanding the system's effective behavior.

Implications for the Grade Hierarchy

The renormalized grade reveals that the Latent hierarchy is **algebraically strict but dynamically soft**:

1. **Algebraically**: grades are disjoint subspaces of $\mathfrak{L}(\mathcal{H})$. No information moves between them by algebraic operations. The bare Latent decomposition $\Lambda = \bigoplus_r \Lambda^{(r)}$ is unique and well-defined.
2. **Dynamically**: through time evolution, higher grades modify the effective behavior observed at lower grades. An observer restricted to grade-k measurements experiences a "dressed" world described by $\tilde{\Lambda}^{(k)}$, which absorbs grade-$(k+1)$+ effects.
3. **Diagnostically**: the gap between bare and dressed quantities — $\|\tilde{M} - M\|/\|M\|$ — is itself a shadow indicator (§10). When $\tilde{M} \approx M$, the hierarchy is clean and truncation is safe. When $\|\tilde{M} - M\| \sim \|M\|$, the hierarchy is soft: higher grades are feeding back strongly, and the effective grade is meaningfully different from the bare grade. This gap is computable from data (compare the predicted grade-2 behavior to the observed behavior) and provides a fourth method for diagnosing hierarchy quality beyond the three ρ-tests of §3.6.
4. **Universally**: the bare hierarchy ALWAYS exists as a mathematical structure. The renormalized hierarchy depends on the dynamics and the observation scale. A system with a clean bare hierarchy and strong renormalization effects (like QCD) behaves very differently from a system with a clean

bare hierarchy and weak renormalization (like the solar system). The bare hierarchy organizes the mathematics; the renormalized hierarchy organizes the physics.

This resolves the apparent paradox: the grade hierarchy is not a limitation or an approximation — it is a real mathematical structure. But its *dynamical consequences* are modulated by the feedback loop between grades. Strong hierarchy ($\rho \gg 1$) means weak feedback and a clean truncation. Weak hierarchy ($\rho \to 1$) means strong feedback and a soft boundary between grades. The boundary $\rho = 1$ is where the feedback becomes so strong that the notion of separate grades loses its operational meaning — the grades are so thoroughly "dressed" by each other that the bare decomposition, while formally valid, carries no predictive power. This is the critical point of the representation itself.

3.9 Time as a Latent: The Duality of Dynamics and Representation

The preceding sections treat the Latent as a static object — a snapshot. But physical systems evolve. This raises a question that probes the framework's ontological depth: **is the Latent temporal (evolving in time), or is temporality itself a Latent?**

The answer is both, and the two perspectives are dual.

The Latent Is Temporal: $\Lambda(t)$

For a dynamical system, the state-Latent $\Lambda(t)$ is a curve in Latent space $\mathfrak{L}(\mathcal{H})$ parameterized by time t. The system's trajectory IS this curve. Given initial condition $\Lambda(0)$ and a grade-2 generator M, the time evolution is:

$$\Lambda(t) = e^{tM}\Lambda(0)$$

for linear (or linearized) dynamics, or more generally $\Lambda(t) = \Phi_t[\Lambda(0)]$ for the nonlinear flow Φ_t. The time-dependent Latent records the complete history of the system.

Time IS the Grade-2 Latent: M Generates Time

The generator $M \in \mathcal{H}^{\otimes 2}$ does not merely *describe* time evolution — it **generates** it. The semigroup $\{e^{tM}\}_{t\geq 0}$ IS the mathematical definition of the system's time. Without M, there is no dynamics, no evolution, no "before" and "after." The grade-2 Latent does not sit inside time; time sits inside the grade-2 Latent.

This means the eigenstructure of M is the **internal clock** of the system:

- **Eigenvalues** λ_k are the system's **natural timescales**. The real parts $\mathrm{Re}(\lambda_k)$ control decay rates (how fast each mode relaxes). The imaginary parts $\mathrm{Im}(\lambda_k)$ control oscillation frequencies (how fast each mode rotates). The pair $(\mathrm{Re}(\lambda_k), \mathrm{Im}(\lambda_k))$ is one tick of the system's internal clock.
- **Eigenfunctions** ϕ_k are the system's **temporal modes** — the directions in state space along which the system evolves independently (at the linearized level). Each eigenfunction is a "channel" of time: the system's time is not one-dimensional but multi-modal, with each mode running at its own rate.
- **Eigenvalue ratios** $|\lambda_k/\lambda_1|$ are the **temporal analyticity parameter**:

$$\rho_{\text{temporal}} = \liminf_{k\to\infty} |\lambda_k/\lambda_1|^{-1/k}$$

This is ρ itself — computed from the generator's spectrum. The analyticity parameter simultaneously measures:

Aspect	What ρ measures	How
Spatial	How many spatial modes are needed	Decay of expansion coefficients
Temporal	How many temporal modes are needed	Decay of generator eigenvalues
Spectral	How compressible the representation is	Ratio of successive eigenvalues

The spatial and temporal readings are **identical** because the generator M couples space and time — its eigenvalues are the characteristic frequencies, and its eigenfunctions are the characteristic spatial patterns. The Latent framework does not distinguish spatial from temporal complexity; ρ measures both because they are one thing: the system's **regularity**.

The Temporal Meta-Latent: Time as a Single Object

The map $t \mapsto \Lambda(t)$ from the time axis to Latent space is itself a function. For analytic systems, this function is analytic (Cauchy–Kovalevskaya theorem: analytic initial data + analytic generator → analytic solution, on any time interval where analyticity is maintained — globally when the evolution semigroup is analytic in a uniform scale of Banach spaces (Ovsyannikov), locally in general). Therefore, by the Latent Theorem (Theorem 1), this function has its own Latent — a **temporal Meta-Latent**:

$$\Lambda_T = \Lambda\big(t \mapsto \Lambda(t)\big) \in \mathfrak{L}(\mathcal{H}_{\text{time}})$$

where $\mathcal{H}_{\text{time}} = L^2([0, T])$ (or $L^2(\mathbb{R})$ for unbounded time). The coordinates of Λ_T in a temporal extraction basis (Fourier for periodic systems, Laguerre for decaying systems, Chebyshev for bounded intervals) are:

$$(\Lambda_T)_n = \langle \Lambda(t), \psi_n(t) \rangle_{L^2([0,T])}$$

where ψ_n are the temporal basis functions. These are exactly the **Fourier coefficients** of the trajectory for periodic orbits — the same coefficients whose decay rate gave $\rho \approx 2.3$ for the figure-8 orbit (§8.4).

The temporal Meta-Latent Λ_T encodes the **entire time evolution** as a single finite object. It is finite because $\rho > 1$ implies exponential decay of temporal coefficients: only $N_T = O(\log(1/\varepsilon)/\log \rho)$ time-modes are needed.

For the three-body figure-8 orbit: Λ_T requires ~ 6 Fourier modes per coordinate at 10^{-8} accuracy (§8.4). The *entire periodic trajectory* — an infinite continuum of states — is captured by ~ 72 real numbers. Time, which appears to be a continuous parameter requiring infinite information, collapses to a finite tensor.

The Recursive Time Structure

The Latent framework reveals four levels at which time appears, each one a grade deeper:

Level	Object	What it encodes	Time's role
Grade 1	$\Lambda(t)$ — state trajectory	The system's path through state space	Time as **parameter** (the curve's label)

Level	Object	What it encodes	Time's role
Grade 2	M — generator	The law of time evolution	Time as **operator** (M generates the flow e^{tM})
Grade 3	$T^{(3)}$ — co-skewness	Nonlinearity, chaos, non-integrability	Time's **complexity source** (grade-3 is why orbits are not simple)
Meta	Λ_T — temporal Meta-Latent	The entire history as one tensor	Time as **object** (the whole evolution, compressed)

At grade 1, time is what we naively think it is: a label on states. At grade 2, time is promoted to an operator — it acts on states rather than merely labeling them. At grade 3, the nonlinearity of the evolution (the failure of e^{tM} to capture the full dynamics) creates temporal complexity — chaos, sensitivity to initial conditions, the non-integrability that makes the three-body problem hard. At the Meta level, time is demoted back to a finite object — but now it is a Latent, not a parameter.

The progression grade 1 → grade 2 → grade 3 → Meta mirrors the historical development of the concept of time in physics: - **Newtonian**: time is a parameter (t labels states) - **Hamiltonian/Lagrangian**: time is generated by a Hamiltonian (H generates canonical transformations) - **Poincaré/KAM**: nonlinear interactions create temporal complexity (chaos, resonances) - **Statistical/information-theoretic**: the entire evolution is characterized by finitely many numbers (ρ, Lyapunov exponents, entropy rates)

The Duality

The question "is the Latent temporal, or is time a Latent?" dissolves into a duality:

$$\boxed{M \text{ generates time} \quad \longleftrightarrow \quad \text{time reveals } M}$$

Left to right: given the generator M (a grade-2 Latent), time evolution follows: $\Lambda(t) = e^{tM}\Lambda(0)$. The Latent creates time.

Right to left: given a trajectory $\Lambda(t)$ (observed time evolution), the generator is recovered as $M = \lim_{t\to 0}(\Lambda(t) - \Lambda(0))/t$, or more generally via the spectral decomposition of the trajectory. Observing time reveals the Latent.

Neither side is primary. The generator and the trajectory are **dual descriptions** of the same dynamical content — related by the exponential map $M \mapsto e^{tM}$ (forward: Latent → time) and the logarithmic derivative $\Lambda(t) \mapsto \dot{\Lambda}(0)$ (backward: time → Latent).

This duality has a precise functional-analytic formulation: the Fokker–Planck generator M and the Koopman operator $U_t = e^{tM^*}$ (the adjoint semigroup acting on observables) are dual operators. The Fokker–Planck picture describes how states evolve forward in time (the Latent generates time). The Koopman picture describes how observables are evaluated along trajectories (time reveals the Latent). The two pictures carry identical information — they are the Latent framework's version of the Schrödinger–Heisenberg duality in quantum mechanics.

Implications for ρ

The time-Latent duality gives ρ a unified interpretation:

$\rho \gg 1$ **(strong hierarchy)**: few temporal modes dominate. The system's "internal time" is simple — nearly periodic, quasi-periodic, or rapidly relaxing. The trajectory $\Lambda(t)$ is smooth with fast Fourier decay. Example: Keplerian orbits ($\rho \approx 8$) — the ellipse is described by a single temporal mode (the orbital frequency) plus tiny perturbative corrections.

$1 < \rho < 3$ **(weak hierarchy)**: many temporal modes contribute. The system's internal time is rich — multiple competing frequencies, slow mixing, complex transient behavior. Example: the figure-8 orbit ($\rho \approx 2.3$) — the trajectory needs ~ 6 Fourier modes, reflecting the three-body coupling.

$\rho \to 1$ **(hierarchy collapse)**: the temporal spectrum is flat. Every frequency contributes comparably. The system has no characteristic timescale — it is scale-free in time. This is the temporal signature of chaos and turbulence. Example: Kolmogorov turbulence — the temporal energy spectrum is a power law, and the cascade operates across all timescales simultaneously.

$\rho < 1$ **(super-critical)**: the temporal expansion diverges. This occurs at **singular moments** — collisions, shocks, phase transitions, bifurcations. At these instants, the system's "internal time" breaks: the eigenvalues of M diverge, the semigroup e^{tM} ceases to be well-defined, and the Latent description requires regularization (§3.6). The collision singularity in the N-body problem is the canonical example: as $r_{ij} \to 0$, the generator's eigenvalues blow up, $\rho \to 0$, and the temporal Meta-Latent requires infinitely many modes to describe the final approach.

The deepest implication: **time is not the framework's scaffolding — it is the framework's content.** The grade-2 Latent does not live "in" time; time lives "in" the grade-2 Latent. The analyticity parameter ρ — the single number that controls the entire framework — is simultaneously a spatial diagnostic, a temporal diagnostic, and a representational diagnostic, because these three are one thing seen from three angles. The Latent is the view from which all three collapse into a single invariant.

Part III — Methodology

Given that the Latent exists, how does one extract it, how do grades organize its structure, and what algebra does the space of Latents carry? Chapters 4–6 answer these questions in sequence. The output is a complete operational calculus: matched bases and optimal truncation (Chapter 4), the grade decomposition and the grade-collapse signature (Chapter 5), and the Latent Algebra with its composition operator (Chapter 6). After this Part the framework is ready to do real work.

Chapter 4. Optimal Extraction

4.0 The Naive Approach: Extraction Philosophy

Before the mathematics of optimal basis selection, there is a prior question: *how should one approach a system in the first place?*

The methodology that produced this framework — and every application in this program — follows a principle that precedes any mathematical optimization:

> **Forget everything. Look at the problem. What do you actually see?**

This is not naïveté. It is minimal-bias extraction: the deliberate stripping away of all pre-installed Lenses before observing the system. Domain conventions, standard approaches, "the way everyone does it" — these are all Lenses (Appendix A.1) that filter Reality before the observer even begins to look. An observer trained for twenty years in a field carries twenty years of Lens buildup. They see the problem through stochastic calculus, or through perturbation theory, or through whatever conceptual apparatus their training installed. These Lenses are powerful — but they are also constraints. They determine what is Observable before any conscious choice is made.

The naive approach inverts this. First, remove all Lenses. Look at the raw Observable — the actual data, the actual equations, the actual structure — without theoretical overlay. *Then*, once you see what is there, choose the appropriate Lens deliberately.

This is why the outsider advantage exists, and why it is not accidental. An outsider has fewer pre-installed Lenses. They see structure that insiders cannot, not because they are smarter, but because their Observable is less filtered. The curse of expertise is that the expert's Lens is so deeply integrated that they no longer perceive it as a Lens — it feels like reality itself.

In the language of the framework: optimal extraction requires the purest possible Observable as input. Pre-installed Lenses corrupt the Observable before extraction begins, leading to a Latent that reflects the Lens rather than the system. The naive approach minimizes this corruption.

The mathematical optimization that follows (§4.1–4.5) addresses which *basis* to use for computation. The naive approach addresses which *eyes* to use for seeing. Both are necessary. The naive approach comes first.

4.1 The Optimization Problem

The Latent Theorem guarantees $N = \Theta(\log(1/\varepsilon)/\log \rho_{\mathcal{B}})$ coordinates per mode. Since $\rho_{\mathcal{B}}$ depends on the extraction basis, the optimal basis is:

$$\mathcal{B}^* = \arg\max_{\mathcal{B}} \ \rho_{\mathcal{B}}(\Lambda)$$

This is a well-defined optimization problem. The objective $\rho_{\mathcal{B}}$ measures how well the basis matches the Latent's structure. The better the match, the faster the coordinate decay, the fewer numbers needed.

4.2 When Each Basis Wins

System type	Optimal basis	Why
Bounded-domain pricing	Cosine	Neumann BCs match reflecting barriers
Gaussian / light-tailed distributions	Hermite	Eigenfunctions of harmonic oscillator match Gaussian tails
Positive quantities (prices, volatilities)	Laguerre	Natural on $[0, \infty)$; matches exponential decay
Nonlinear dynamics (double-well, bistable)	FP eigenbasis	Diagonalizes the actual generator
Multi-scale / jump-diffusion	Wavelets	Localized in space and frequency
Directional / spherical data	Spherical harmonics	Natural on S^{d-1}

4.3 Generator Eigenbasis Optimality

Theorem 3 (Generator eigenbasis optimality). Let $\Lambda(t)$ be the grade-1 Latent encoding a time-homogeneous reversible diffusion with Fokker–Planck generator $\mathcal{L}$ on a compact domain Ω. Let $\{\psi_k\}$ be the eigenfunctions of $\mathcal{L}$ with eigenvalues $0 = \lambda_0 > \lambda_1 \geq \lambda_2 \geq \cdots$. Then the eigenbasis of $\mathcal{L}$ achieves the maximal analyticity parameter among all orthonormal bases of $L^2(\Omega, \mu)$, where μ is the stationary measure.

Proof. Fix $t > 0$. We compute the analyticity parameter in two bases and compare.

Step 1: Eigenbasis decay. In the eigenbasis $\{\psi_k\}$, the semigroup $e^{t\mathcal{L}}$ acts diagonally: $\Lambda_k(t) = \Lambda_k(0)e^{\lambda_k t}$. Since $\mathcal{L}$ is self-adjoint with respect to μ (reversibility) on a compact domain, Weyl's law gives $\lambda_k \sim -c \cdot k^{2/d}$ for large k, where $d = \dim(\Omega)$. In particular $|\lambda_k| \to \infty$ strictly faster than linearly. The coordinate magnitudes satisfy $|\Lambda_k(t)| \leq \|\Lambda(0)\|_\infty \cdot e^{\lambda_k t}$, giving superexponential decay: $|\Lambda_k(t)| = O(e^{-ck^{2/d}t})$. The analyticity parameter is $\rho_{\text{eig}} = \liminf_{k\to\infty} |\Lambda_k(t)|^{-1/k} = +\infty$ in the strictest sense, but for the purpose of the truncation bound (Theorem 1), what matters is the effective decay rate at moderate k, which is controlled by the spectral gap $|\lambda_1|$ and the eigenvalue spacing.

Step 2: Any other basis is worse. Let $\{\phi_k\}$ be any other orthonormal basis of $L^2(\Omega, \mu)$. The change-of-basis matrix $U_{kj} = \langle \phi_k, \psi_j \rangle$ is unitary. The ϕ-coordinates are:

$$\Lambda_k^{(\phi)}(t) = \sum_{j=0}^{\infty} U_{kj}\, \Lambda_j(0)\, e^{\lambda_j t}$$

For any fixed j_0, the contribution of eigenmode j_0 to the k-th ϕ-coordinate is $U_{kj_0} \cdot \Lambda_{j_0}(0) \cdot e^{\lambda_{j_0} t}$, which decays in k only as fast as $|U_{kj_0}|$ decays.

Step 3: Parseval constraint. By Parseval's theorem, $\sum_{k=0}^{\infty} |U_{kj}|^2 = 1$ for each j. In particular, $\sum_{k=0}^{\infty} |U_{k1}|^2 = 1$, so the sequence $\{|U_{k1}|\}_{k\geq 0}$ is square-summable but generically does not have exponential decay. Suppose for contradiction that $\rho_\phi > \rho_{\text{eig}}$, meaning $|\Lambda_k^{(\phi)}(t)|$ decays faster in k than $|\Lambda_k(t)|$ does. At large t, the $j = 1$ term dominates all $j \geq 1$ terms (since $\lambda_1 > \lambda_j$ for $j \geq 2$), giving:

$$|\Lambda_k^{(\phi)}(t)| \approx |U_{k1}| \cdot |\Lambda_1(0)| \cdot e^{\lambda_1 t} \quad \text{for large } t$$

For $\rho_\phi > \rho_{\text{eig}}$, we would need $|U_{k1}|$ to decay faster than the eigenbasis coordinates $|\Lambda_k(t)|$, i.e., $|U_{k1}| = o(e^{\lambda_k t})$ for all $t > 0$. But $|U_{k1}|$ is independent of t (it is a property of the basis, not the dynamics), while $e^{\lambda_k t} \to 0$ as $t \to \infty$ for any fixed $k \geq 1$. Taking $t \to 0^+$, the requirement becomes $|U_{k1}| = o(1)$ for large k — which is only the Riemann-Lebesgue condition, automatically satisfied. Taking t large, the requirement becomes $|U_{k1}| = o(e^{\lambda_k t})$, which fails whenever $|U_{k1}| > 0$ and $\lambda_k t$ is sufficiently negative — i.e., for any k with $\langle \phi_k, \psi_1 \rangle \neq 0$.

Step 4: Conclusion. The only way $\rho_\phi \geq \rho_{\text{eig}}$ at all times $t > 0$ is if $U_{k1} = \langle \phi_k, \psi_1 \rangle = 0$ for all but finitely many k. Since $\sum_k |U_{k1}|^2 = 1$, this means ψ_1 is a finite linear combination of ϕ_k's. Repeating for each eigenfunction ψ_j: $\rho_\phi \geq \rho_{\text{eig}}$ requires each ψ_j to be a finite linear combination of the ϕ_k's. Since both are orthonormal bases of the same separable Hilbert space, this forces $\{\phi_k\}$ to be a reordering of $\{\psi_k\}$ (up to signs and finite permutation within eigenspaces of equal eigenvalue). Therefore $\rho_\phi = \rho_{\text{eig}}$, and $\rho_\phi > \rho_{\text{eig}}$ is impossible for any orthonormal basis. □

Practical consequence. The COS basis is suboptimal whenever the process dynamics differ significantly from sinusoidal modes. For a double-well potential, the generator's eigenfunctions are deformed by the potential landscape — using them instead of cosines gives fewer modes for the same accuracy.

4.4 The Representation-Structure Matching Theorem

Theorem 3 proves that the eigenbasis is optimal for reversible diffusions with a known generator. But many systems (lognormal sums, stochastic models, polynomial chaos expansions) do not have a Fokker–Planck generator — they have a *generative measure*. The following theorem generalizes Theorem 3 to all such systems.

Definition (Generative measure). The *generative measure* μ of a system is the probability measure that generates the system's randomness — the distribution from which the underlying random variables are drawn before any nonlinear transformation. For a sum $S = \sum w_i e^{X_i}$ with $\mathbf{X} \sim \mathcal{N}(\mu, \Sigma)$, the generative measure is the multivariate Gaussian. For a periodic orbit, it is the uniform measure on the period. For a Markov process, it is the stationary distribution. The generative measure is not a choice — it is determined by the mathematical identity of the problem.

Theorem 4 (Representation-Structure Matching). *Let $f_\theta : \Omega \to \mathbb{R}$ be a parametric family of analytic functions indexed by $\theta \in \Theta$, where $f_\theta = g_\theta \circ T$ for an analytic map g_θ and random variable $T \sim \mu$ on Ω. Let $\{p_k^\nu\}$ denote the orthogonal polynomial system with respect to measure ν on Ω, and $c_k^\nu(\theta) = \langle f_\theta, p_k^\nu \rangle_\nu$.*

(a) Matched basis ($\nu = \mu$): the coefficients satisfy

$$|c_k^\mu(\theta)| \leq C_\theta \cdot r_\theta^k / k!^{s_\theta}, \quad s_\theta > 0$$

The Latent's analyticity parameter satisfies $\rho_\mu(\theta) \geq r_\theta > 1$ for all θ.

(b) Mismatched basis ($\nu \neq \mu$): for a generic mismatch,

$$\sup_{\theta \in \Theta} |c_k^{\nu}(\theta)| \to \infty \quad \text{as } k \to \infty$$

The analyticity parameter $\rho_{\nu}(\theta^*) \leq 1$ for some $\theta^* \in \Theta$.

(c) Condition number: the computational chain satisfies

$$\kappa(\mathcal{B}_{\mu}) = O(1), \qquad \kappa(\mathcal{B}_{\nu}) \geq e^{c \cdot N^2} \quad \text{for typical mismatch}$$

Proof sketch. Part (a): for $\mu = N(0, I)$ (Gaussian), the Cameron–Martin theorem (1944) gives $|c_k| = O(\sigma^k / k!)$ for analytic g. For general μ, the Askey scheme (Xiu and Karniadakis, 2002) provides the matching polynomial family with analogous factorial-type decay. Part (b): the change-of-basis matrix $U_{kj} = \langle p_k^{\nu}, p_j^{\mu} \rangle$ generically does not decay, so well-behaved c_j^{μ} get mixed into growing c_k^{ν}. Concrete example: moments of a lognormal sum grow as $e^{k^2\sigma^2/2}$ (monomial basis mismatched to Gaussian measure). Part (c): the Toeplitz condition number built from $\{c_k^{\nu}\}$ satisfies $\kappa \geq \max |c_k| / \min |c_k|$, which is $O(1)$ for matched decay and $e^{O(N^2)}$ for the moment mismatch. □

Remark (Information-theoretic failure of mismatched representations). The condition number blowup (part c) is not the deepest failure of a mismatched basis. For lognormal sums, the Carleman condition $\sum_{k=1}^{\infty} m_{2k}^{-1/(2k)} < \infty$ fails because $m_{2k} \gtrsim e^{2k^2\sigma^2}$ — meaning the Hamburger moment problem is *indeterminate*: infinitely many distinct distributions share the same moment sequence. Even with infinite precision and infinite moments, the monomial representation cannot uniquely identify the distribution. By contrast, the Hermite-chaos coefficients $\{c_{\mathbf{k}}^H\}$ form a complete orthonormal system in $L^2(\gamma_n)$ (Cameron–Martin, 1944), so they uniquely determine any square-integrable function of the underlying Gaussians. The grade-3 choice (Hermite over monomials) is not merely a numerical improvement — it is an information-theoretic necessity: the mismatched representation is provably non-identifying.

Relation to Theorem 3. Theorem 3 (eigenbasis optimality) is the special case $□ = $ stationary measure, $g = $ semigroup evolution, basis = FP eigenfunctions. Theorem 4 covers systems without dynamics — pure distributional problems where the generative measure is the fundamental object.

Instances:

System	Generative μ	Matched $\mathcal{B}^*$	Decay rate	Mismatched example
Lognormal sums	$N(0, I)$	Hermite	$\sigma^k / k!$	Moments: $e^{k^2\sigma^2/2}$
Three-body orbits	Uniform on period	Fourier	$\rho^{-\lvert k \rvert}$	Taylor: radius → 0
General PCE	Distribution of X	Askey polynomial	Analyticity-dependent	Wrong Askey: Runge blowup
Reversible diffusion	Stationary μ	FP eigenbasis	$e^{-ck^{2/d}}$	Any other ONB (Theorem 3)
Riemannian manifold	Riemannian volume dV_g	Laplacian eigenfunctions	$\lambda_k^{-n/4}$ (Li-Yau)	Any non-eigenbasis: loses structure constants

Remark (Natural taxonomy of problems). The Instances table is not merely a list of examples — it is a classification principle. The generative measure μ partitions all smooth problems into algebraic families:

all Gaussian-generated problems (sums, products, and transforms of normal variables) share the Hermite basis; all periodic problems share Fourier; all Markov problems share the Fokker–Planck eigenbasis. Two problems with the same μ are "the same species" from the Latent perspective: they share the same optimal basis, same convergence rates, and same computational pipeline. This classification is finer than traditional divisions (linear vs. nonlinear, stochastic vs. deterministic) because it classifies by generative structure — the algebraically relevant distinction.

Remark (Pipeline optimality). The Matching Principle determines not only the expansion basis but the *entire* computational pipeline. In the Hermite-COS solution: (1) the Hermite basis is Gaussian-matched (Theorem 4), (2) Gauss-Hermite quadrature is spectrally optimal for Gaussian-weighted integrands, and (3) the COS inversion operates in log-space because $\log S$ inherits near-Gaussian shape from the generative measure. All three stages are independently determined by the same $\mu = N(0, I)$. This predicts that for periodic problems the pipeline would be Fourier basis + trapezoidal rule + direct-space COS (the trapezoidal rule is spectrally accurate for periodic integrands). Empirical validation: in the Fenton problem, replacing only stage (3) — applying COS in direct space instead of log-space — causes catastrophic failure for $\sigma >$ 0.5, even though stages (1) and (2) remain correctly matched. Partial pipeline mismatch is sufficient for divergence.

First application: the Fenton problem. The 65-year-old open problem of computing the CDF of a sum of correlated lognormals (Fenton, 1960) is the first practical demonstration of Theorem 4. In the mismatched moment basis ($\$\square = \$$ monomial), the expansion diverges ($\rho_\nu \leq 1$, Carleman condition fails). In the matched Hermite basis ($\mu = N(0, I)$), the same expansion converges factorially ($\sigma^{2K}/K!$). Theorem 4 predicts this: the change $\nu \to \mu$ is determined entirely by the generative structure of the problem. The result is a constructive, universally convergent CDF formula — the Hermite-COS special function $\mathcal{H}^n_{N,Q}$ — that evaluates the distribution from the finite generative latent through a chain of Latent algebra operations (map, contraction, synthesis) with exponential convergence in both precision parameters. See companion paper (Nagy 2026, *The Fenton Distribution Solved*) for the full development.

The convergence contrast. The characteristic function $\phi_S(t) = E[e^{itS}]$ — the complexification of the distribution via Euler's $e^{itS} = \cos(tS) + i\sin(tS)$ — provides a convergent Fourier representation of the CDF. For the analytic density of a lognormal sum, the COS series converges absolutely: the infinite series IS the CDF, not an approximation. The moment route fails not because the distribution is complex but because it uses a divergent evaluator ($|S^k| \to \infty$) for the same Hilbert-space contraction that the Fourier route evaluates convergently ($|e^{itS}| = 1$). The matched Hermite-COS pipeline (map → contraction → synthesis) is a composition of Latent algebra operations, each exact; the precision parameters belong to the evaluator layer. See the companion paper (Nagy 2026, *The Fenton Distribution Solved*) for the full development.

Remark (Structural resummation). The Matching Principle is not only an optimization statement — it is a *resummation* statement. When $\rho_\nu \leq 1$ (divergent expansion in basis ν), changing to μ with $\rho_\mu > 1$ converts the divergent series into a convergent one. Unlike classical resummation techniques (Borel, Cesàro, Padé) which attempt to extract a finite value from a divergent series within the *same* representation, the Latent approach changes the representation itself. The change-of-basis map $U : L^2(\nu) \to L^2(\mu)$ is a structural resummation operator. In the Fenton problem, this operator converts the divergent Edgeworth expansion (which amplifies high-order terms through hypercontractivity) into the convergent COS expansion (which reorganizes the same information into a trigonometric polynomial).

Completeness of the Latent framework. Theorems 1, 3, and 4 together answer the three fundamental questions of representability:

Question	Answer	Theorem
Does a finite representation exist?	Yes, for all smooth systems	Theorem 1 (Latent Theorem)
How large is it?	$N^* = O(\log(1/\varepsilon)/\log\rho)$	Theorem 1
Which representation achieves this bound?	The one matching the generative structure	Theorem 4

The Latent Theorem without Theorem 4 guarantees existence without specifying construction. Theorem 4 completes the framework: the optimal representation is not found by search but by *structural matching* — the generative measure determines the basis.

4.5 Why the Grade-3 Hierarchy Terminates

Theorems 3 and 4 determine the optimal basis from the system's structure. This leads to a meta-hierarchy of choices (developed fully in Nagy 2026, *The Fenton Distribution Solved*, Section 5):

Meta-grade	Object	Type	Determined by
3	Basis choice $\mathcal{B}^*$	Highest choice	Generative structure (given)
2	Extraction difficulty α^*	Continuous parameter	$\mathcal{B}^*$ + system parameters
1	Latent Λ	Functional	Everything above
0	Parameters	Data	Given

Above meta-grade 3, the generative structure is the mathematical identity of the problem (not optimizable), and the Matching Principle (Theorem 4) is a universal theorem (not a choice). Below meta-grade 3, everything is forced by the basis choice and parameters. Meta-grade 3 is the last degree of freedom — the single decision that determines whether the entire Latent extraction chain is feasible or intractable.

Remark (Complexification is a Latent operation). Complexification — extending the Latent algebra from $\mathbb{R}$ to $\mathbb{C}$ — is not a technique; it is itself a Latent operation, and the grade-3 choice (which basis to use) IS the choice of how to complexify. The complex algebra has strictly more structure than the real algebra: analyticity (power series convergence guarantees), Fourier theory (exact trigonometric representation of analytic functions), and monodromy (topological constraints from analytic continuation). None of these exist in the real algebra alone. Choosing the Hermite-COS basis for lognormal sums is complexifying via the characteristic function $\phi_S(t) = E[e^{itS}]$ (Euler: $e^{itS} = \cos(tS) + i\sin(tS)$); choosing the Fourier basis for periodic orbits is complexifying via $e^{i\omega t}$; choosing pair displacements $\phi_{ij}(z) \in \mathbb{C}$ for central configurations (Finding 14, §8.4) reveals monodromy at collision singularities. In each case, the grade-3 selection determines the complexification, and the complex structure solves the real problem: convergent COS series for distributions, exponentially decaying Fourier coefficients for orbits, branch-point finiteness proofs for algebraic geometry. The pattern is Euler's: $e^{ix} = \cos x + i\sin x$ is not an approximation of the exponential by trigonometric functions — it IS the function seen from the complex plane, and the sin/cos are projections.

Remark (The Hurwitz hierarchy of complexification). The extension $\mathbb{R} \to \mathbb{C}$ is the first step in a constrained algebraic hierarchy. By the Hurwitz theorem (1898), the only normed division algebras over $\mathbb{R}$ are: $\mathbb{R}$ (1D), $\mathbb{C}$ (2D), $\mathbb{H}$ (quaternions, 4D), and $\mathbb{O}$ (octonions, 8D). There is no 3D, 5D, or 7D division algebra — each step doubles the dimension (Cayley-Dickson construction) and sacrifices a structural property: $\mathbb{C}$ loses ordering, $\mathbb{H}$ loses commutativity, $\mathbb{O}$ loses associativity.

Each algebra defines a different complexification and a different harmonic analysis:

Algebra	Exponential	Harmonics	Domain
$\mathbb{C}$	$e^{i\theta} \in S^1$	$\cos(n\theta), \sin(n\theta)$ — Fourier series	Distributions on $\mathbb{R}$
$\mathbb{H}$	$e^{i\theta+j\phi+k\psi} \in S^3$	Spherical harmonics Y_ℓ^m	Distributions on S^2, rotations
$\mathbb{O}$	7 imaginary units, S^7	Exceptional harmonics (G_2)	String-theoretic structure

For the Latent framework, this means the grade-3 choice is richer than "which basis" — it is "which number algebra to complexify into." Distributions on $\mathbb{R}$ require $\mathbb{C}$-complexification (the characteristic function e^{itS}, Fourier-COS inversion). Distributions on spheres or involving rotations would naturally require $\mathbb{H}$-complexification (quaternionic Fourier transform, spherical harmonic expansion). The pattern: $\mathbb{R} \to \mathbb{C}$ reveals Fourier structure and analyticity; $\mathbb{C} \to \mathbb{H}$ would reveal rotational structure and spin; $\mathbb{H} \to \mathbb{O}$ would reveal exceptional symmetries. Each step reveals algebraic structure invisible at the previous level, at the cost of losing a commutativity or associativity property. The normalization constant $(2\pi)^{-n/2}$ of the Gaussian density is itself the reciprocal of the volume of S^{n-1} — the sphere whose harmonics provide the basis.

Why the hierarchy has gaps — the topological obstruction. The non-existence of a 3D (or 5D, 6D, 7D) division algebra is not algebraic convention; it is forced by topology. A division algebra of dimension n requires S^{n-1} to be *parallelizable* — its tangent bundle must be trivial (equivalently: a continuous nowhere-zero vector field must exist on S^{n-1}). By the Poincaré-Hopf theorem, any vector field on S^{2k} has at least $|\chi(S^{2k})| = 2$ zeros (the Euler characteristic is 2 for all even-dimensional spheres). This is the *hairy ball theorem*: S^2 cannot be combed smooth — which kills $n = 3$ (since S^2 is not parallelizable), and the same argument kills $n = 5, 7$ (since S^4, S^6 fail). Among odd-dimensional spheres, Adams (1962) proved that only S^1, S^3, S^7 are parallelizable — killing everything except $n = 2, 4, 8$. The Hopf fibrations ($S^3 \to S^2$ with fiber S^1; $S^7 \to S^4$ with fiber S^3; $S^{15} \to S^8$ with fiber S^7) are the geometric manifestation of the three non-trivial division algebras — and Adams proved no other sphere fibrations exist.

For the Latent framework, this topological constraint has a sharp consequence: **the grade-3 choice is discrete, not continuous.** You either have full $\mathbb{C}$-complexification (with Fourier analysis, analyticity, residue calculus) or none — there is no "partial complexification" between $\mathbb{R}$ and $\mathbb{C}$. Structure revelation is inherently binary: each Cayley-Dickson step doubles the algebra and unlocks a complete harmonic theory, with nothing in between. This echoes quantum mechanics, where the Hilbert space MUST be over $\mathbb{C}$ — not "between $\mathbb{R}$ and $\mathbb{C}$" — because only $\mathbb{C}$ supplies both the phase structure ($U(1)$ gauge) and commutativity. The Hurwitz constraint forces it.

Theorem 4b (Grade-3 Universal Termination). *Let $\mathcal{S}$ be a system whose state space factors as $\mathcal{H} = \mathcal{H}_{spatial} \otimes \mathcal{H}_{temporal}$, and let the three candidate Grade-4 meta-choices be: (i) coordinate system in $\mathcal{H}_{spatial}$, (ii) time discretization in $\mathcal{H}_{temporal}$, (iii) boundary conditions. Then:*

(a) Coordinate choice is Grade-3 in $\mathcal{H}_{spatial}$: the matched coordinate system (action-angle variables for Hamiltonian systems) achieves $\rho_{\text{coord}} > 1$, while mismatched coordinates (Cartesian) give $\rho_{\text{coord}} \leq 1$ with $\kappa \to \infty$.

(b) Time discretization is Grade-3 in $\mathcal{H}_{temporal}$: symplectic integrators (matched) preserve the symplectic 2-form with bounded energy error and $\rho_{\text{time}} > 1$; non-symplectic integrators (mismatched) exhibit linear energy drift and $\rho_{\text{time}} \leq 1$.

(c) Boundary conditions are Grade-0: they define the function space $V \subset \mathcal{H}$, not the representation within it. They are part of the problem specification, not a meta-choice.

The product-space analyticity parameter $\rho = \min(\rho_{spatial}, \rho_{temporal}) > 1$ *when both factors use matched bases. No Grade-4 meta-level exists: Grade-3 is the universal terminal grade for distributions, dynamical systems, and their products.*

Proof sketch. For (a): action-angle variables (θ, I) diagonalize the Hamiltonian flow — the dynamics become $\dot{\theta} = \omega(I)$, $\dot{I} = 0$ — so the phase-space coordinates are eigenfunctions of the generator, giving $\rho > 1$ by the universality theorem. Cartesian coordinates mix the eigenspaces and produce coupled equations with growing condition number. For (b): symplectic integrators preserve $\omega = dp \wedge dq$, so the discrete flow is itself Hamiltonian (with a modified Hamiltonian $\tilde{H} = H + O(h^p)$). The energy error is bounded: $|H_n - H_0| \leq Ch^p$ for all n. Non-symplectic integrators break this conservation, producing linear energy drift $|H_n - H_0| \sim C'nh$ that makes $\rho \leq 1$ in the temporal direction. For (c): boundary conditions specify which functions are in the domain of the operator — they define V but have no bearing on which basis within V is matched. They are therefore Grade-0 (given data), not a meta-choice. □

Corollary (Constructive B*). *Given N i.i.d. samples from a system with unknown generative structure but spectral gap* $g > 0$*, the empirical eigenbasis* $\hat{\mathcal{B}}_N^*$ *of the sample covariance operator converges to the true matched basis* $\mathcal{B}^*$ *at rate*

$$\sin \angle(\hat{\mathcal{B}}_N^*, \mathcal{B}^*) \leq \frac{C_{\text{op}}}{\sqrt{N} \cdot g}$$

where C_{op} *depends on the operator's trace norm. The gap g is itself estimable from data:* $|\hat{g}_N - g| \leq 2C_{op}/\sqrt{N}$. *When* $\hat{g}_N < 2C_{op}/\sqrt{N}$ *(the estimated gap falls below the estimation error), the basis is near a phase transition and the Grade-3 diagnostic reports instability — the truncation order K should be reconsidered.*

This closes the practical gap in the Grade-3 theory: previous instances (Hermite, Fourier, Pauli, wavelets, etc.) all assumed the matched basis was known analytically. The constructive B* corollary shows that spectral estimation + Davis-Kahan perturbation theory suffice to *discover* $\mathcal{B}^*$ from data, and that the convergence rate is itself determined by the Grade-3 quality measure g.

Formal verification. The grade-3 algebra is formally verified in the proof kernel across twelve proof files (142 theorems, 0 sorry, 480 kernel declarations type-checked). The abstract algebra (`grade3_algebra_proof.py`, 15 theorems) proves: (i) the composition chain $\mathcal{B}^* \to \varkappa \to \Lambda \to$ observable is well-defined and closed; (ii) the grade-3 selection is parameter-independent ($\mathcal{B}^*(\theta_1) = \mathcal{B}^*(\theta_2)$ for all θ); (iii) the matched basis is optimal ($\varkappa_{\text{matched}} \leq \varkappa_{\text{mismatched}}$); (iv) the hierarchy terminates at grade 3 (the matching principle is a universal theorem, not a choice); (v) matched coefficients decay factorially while mismatched coefficients diverge. Seven concrete instances prove the abstract algebra is not domain-specific: the Hermite instance (`hermite_instance_proof.py`, 12 theorems) for log-normal sums with $|c_k^H| \cdot k! \leq C_H \cdot \sigma^k$ (factorial decay) vs $|c_k^M| \cdot k! \geq M_{\text{lower}}(k)$ (super-exponential growth); the Wiener-Askey instance (`askey_scheme_proof.py`, 10 theorems) proving the full gPC table (Gaussian→Hermite, Uniform→Legendre, Gamma→Laguerre, Beta→Jacobi) as grade-3 selection with Favard uniqueness; the Fourier orbit instance (`fourier_orbit_proof.py`, 10 theorems) for periodic N-body orbits with $\rho = \exp(\delta)$ determined by the Painlevé singularity distance; the Pauli QEC instance (`pauli_quantum_proof.py`, 10 theorems) proving the Pauli basis is grade-3 for quantum error correction with $\varkappa_P = O(1)$ vs $\varkappa_C \geq c_q n^2$ and error threshold $p_{\text{th}} \leq C_{\text{th}}(\rho - 1)$; the spherical harmonics instance (`spherical_harmonics_proof.py`, 10 theorems) for distributions on S^2 with the Hurwitz dimension-doubling hierarchy ($\mathbb{R} \to \mathbb{C} \to \mathbb{H} \to \mathbb{O}$) formalized as discrete grade-3 choices; and the wavelet multi-resolution instance (`wavelet_multiresolution_proof.py`, 12 theorems) — the crucial non-eigenbasis case where the "operator" is a scale-indexed family of projections, the cascade operator

provides the generalized spectral gap ($g = s > 0$ from Sobolev regularity), and wavelets dominate Fourier for BV functions ($\rho_W > 1$ geometric vs $\rho_F \leq 1$ algebraic due to Gibbs phenomenon); and the neural network feature instance (`neural_feature_proof.py`, 14 theorems) proving that deep learning is Grade-3 selection via the Neural Tangent Kernel: the NTK eigenfunctions define $\mathcal{B}^*$, gradient descent implicitly traverses them in eigenvalue order (spectral bias = matched Grade-3 learning), the Fisher information matrix condition number κ_F measures misalignment (natural gradient = $\kappa = 1$ = matched, standard gradient = $\kappa \gg 1$ = mismatched), and the double descent phenomenon IS the $\rho = 1$ phase transition (underparameterized: gap ≤ 0, $\rho \leq 1$; overparameterized: implicit regularization restores gap > 0, $\rho > 1$). The EC bridge (`ec_grade3_bridge_proof.py`, 8 theorems) connects the grade-3 selection to the $\rho = 1$ phase transition: the gap function $g(\rho) = \rho - 1$ IS the grade-3 quality measure, and the three convergence regimes (exponential, algebraic, divergent) are the three grade-3 selection outcomes. The universality theorem (`universality_proof.py`, 15 theorems) proves the meta-result: ANY self-adjoint operator with spectral gap $g > 0$ yields $\rho = 1 + g > 1$ in its eigenbasis and $\rho_{\text{mis}} \leq 1$ in any non-eigenbasis, unifying all seven instances as specializations of this single structural principle. Finally, the grade-4 termination theorem (`grade4_termination_proof.py`, 14 theorems) resolves the open question from the Remark above: the three candidate Grade-4 levels for dynamical systems (coordinate choice, time discretization, boundary conditions) all collapse — coordinate choice is Grade-3 in phase space (action-angle = matched, Cartesian = mismatched), time discretization is Grade-3 in the temporal direction (symplectic = matched with bounded energy error, non-symplectic = mismatched with $\rho \leq 1$), and boundary conditions are Grade-0 (given, not chosen). The product-space $\rho = \min(\rho_{\text{spatial}}, \rho_{\text{temporal}}) > 1$ when both factors use matched bases, proving Grade-3 is the universal terminal grade for both distributions and dynamical systems. The constructive B* algorithm (`constructive_bstar_proof.py`, 12 theorems) proves the data-driven counterpart: the empirical eigenbasis converges to $\mathcal{B}^*$ at rate $O(1/\sqrt{N})$ via operator convergence + Davis-Kahan perturbation theory, the spectral gap is estimable with error $\leq 2C_{\text{op}}/\sqrt{N}$, and the phase transition (gap below estimation error) is detectable — making the Corollary above computationally constructive.

4.6 Optimal Rank Truncation: The Eckart-Young Theorem

Theorems 3 and 4 determine the optimal *basis* for Latent extraction. A complementary question remains: given a basis, what is the optimal *truncation*? If the Latent has coordinates in a matrix representation $M \in \mathbb{R}^{m \times n}$ (a grade-2 object, or a data matrix whose columns are grade-1 Latents), what is the best rank-K approximation?

Theorem 15 (Eckart-Young optimality). *Let $M \in \mathbb{R}^{m \times n}$ have singular value decomposition $M = U\Sigma V^{\top}$ with singular values $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_{\min(m,n)} \geq 0$. Let $M_K = \sum_{i=1}^{K} \sigma_i u_i v_i^{\top}$ denote the rank-K SVD truncation. Then M_K minimizes the Frobenius-norm error among all rank-K matrices:*

$$\|M - M_K\|_F^2 = \sum_{i=K+1}^{\min(m,n)} \sigma_i^2 \leq \|M - N\|_F^2 \quad \text{for all } N \text{ with } \operatorname{rank}(N) \leq K$$

The inequality is tight: equality holds if and only if N is also a rank-K SVD truncation (up to rotation within degenerate singular subspaces).

Proof. Standard (Eckart and Young, 1936; Mirsky, 1960). The key step: any rank-K matrix N has a K-dimensional column space W. The $(K+1)$-dimensional subspace spanned by $\{v_1, \ldots, v_{K+1}\}$ must intersect $W^{\perp}$ nontrivially (by dimension counting), yielding a unit vector $w \in W^{\perp}$ with $\|Mw\| \geq \sigma_{K+1}$. Since $Nw = 0$ (because $w \perp W$), the error $\|M - N\|_F^2 \geq \|Mw - Nw\|^2 = \|Mw\|^2 \geq \sigma_{K+1}^2$. Iterating over the remaining singular directions gives the full bound. □

Latent interpretation. The Eckart-Young theorem is the *rank-level* optimality guarantee for the Latent framework, complementing the *basis-level* guarantee of Theorem 3:

Level	Theorem	What is optimized	Result
Basis	Theorem 3	Which orthonormal system to expand in	Eigenbasis maximizes ρ
Rank	Theorem 15 (Eckart-Young)	How many modes to keep	Rank-K SVD truncation minimizes error
Representation	Theorem 4	Which distribution family to match	Matching principle selects the basis

Together, Theorems 3, 4, and 15 give a **complete optimality chain**: the Latent framework selects the optimal representation (Theorem 4), extracts in the optimal basis (Theorem 3), and truncates at the optimal rank (Theorem 15). No other framework of equal complexity can do better at any of the three levels.

Connection to ρ. For a matrix M whose singular values decay geometrically — $\sigma_k \sim C\rho^{-k}$ with $\rho > 1$ — the Eckart-Young residual becomes:

$$\|M - M_K\|_F^2 = \sum_{i=K+1}^{\infty} \sigma_i^2 \leq \frac{C^2}{\rho^{2(K+1)}(1-\rho^{-2})}$$

This is exactly the Latent Theorem bound (Theorem 1) applied to the matrix representation: exponential decay in K, controlled by ρ. The Eckart-Young theorem guarantees that NO other rank-K representation achieves a better rate — the Latent truncation is provably optimal.

Formal verification. The Eckart-Young theorem and its consequences are formally verified in the proof kernel (`fields/residual_stream_denoising/platonic.py`): 14 theorems covering the diagonal (scalar) form, the full matrix form, and the spectral knowledge distillation chain (§7.8), with 479 lines of Lean 4 export, 0 sorry, 0 axiom debt. The Lean 4 compilation stamp (`stamp/protein_fold_stamp/EckartYoung.lean`) confirms independent machine verification via `lake build` on Lean 4.29.0.

Chapter 5. Grades and Entanglement

5.1 The Pairwise Bottleneck

Current spectral methods operate at grades 1 and 2. The eigenvalue-conditioned method (Eigen-COS) diagonalizes the covariance matrix (a grade-2 object) and processes each eigenmode at grade 1. This captures pairwise dependence but nothing higher.

Definition 6 (Latent rank). The *rank* of a Latent $\Lambda \in \mathcal{H}^{\otimes r}$ is the minimal number of rank-1 terms in a decomposition:

$$\Lambda = \sum_{\alpha=1}^{R} \lambda_\alpha f_\alpha^{(1)} \otimes f_\alpha^{(2)} \otimes \cdots \otimes f_\alpha^{(r)}$$

where $f_\alpha^{(i)} \in \mathcal{H}$ and $\lambda_\alpha \in \mathbb{R}$.

Definition 7 (Entanglement). A Latent is *separable* if its rank is 1 (product structure, no interaction). It is *s-entangled* if its rank is at least *s*. The rank measures the irreducible complexity of the system's internal interactions.

For grade 2 (matrices): rank-1 = independent components. Higher rank = correlated. Eigenvalue decomposition separates the ranks.

For grade ≥ 3: **eigenvalue decomposition does not suffice.** Symmetric 3-tensors have generalized eigenvalues (Lim 2005, Qi 2005), but the decomposition cannot be reduced to a sequence of matrix diagonalizations. Three-body interactions are irreducibly richer than pairwise interactions.

5.2 Three-Body Latent: The Rank Bound Theorem

Theorem 6 (N-body kinematic rank bound). Consider N bodies with masses $m_1, \dots, m_N > 0$ in d spatial dimensions, interacting via conservative pairwise forces. Let $r_i(t) \in \mathbb{R}^d$ be the position of body i, and let $X(t) = (r_1(t), \dots, r_N(t)) \in \mathbb{R}^{Nd}$ be the trajectory vector. In the center-of-mass frame ($\sum_i m_i r_i = 0$), the trajectory matrix $X(t)$ lies in a subspace of dimension at most $(N-1)d$. Specifically:

(i) Upper bound. The SVD of X as a function of time has at most $(N-1)d$ nonzero singular values. This bound is tight: for generic initial conditions with all N bodies in non-degenerate motion, the rank equals $(N-1)d$ exactly.

(ii) Reduction by symmetry. If the orbit has a discrete symmetry group G of order $|G|$ permuting the bodies, the rank is reduced to at most $(N-1)d/\dim(\text{fixed-point subspace of } G)$. For C_3-symmetric orbits of 3 equal masses in 2D (figure-8, Lagrange), the rank is at most 4; Lagrange achieves rank 2 (rigid rotation).

(iii) Mode count bound. If the orbit is periodic with period T and is analytically continuable to a strip $|\mathrm{Im}(t)| < \delta$ in the complex time plane, then the analyticity parameter $\rho = e^{2\pi\delta/T} > 1$ and the number of Fourier modes needed for ε-accuracy is $N_\varepsilon = \lceil \log(C/\varepsilon)/\log\rho \rceil$. The total Latent size is $(N-1)d \cdot N_\varepsilon$.

Proof. (i) In the center-of-mass frame, $r_N = -\sum_{i=1}^{N-1}(m_i/m_N)r_i$. The Nd coordinates of $(r_1, \dots, r_N)$ span at most $(N-1)d$ independent directions. The bound is tight because the Jacobi transformation

$$\rho_1 = r_2 - r_1, \quad \rho_2 = r_3 - \frac{m_1 r_1 + m_2 r_2}{m_1 + m_2}, \quad \dots$$

provides $(N-1)$ independent vectors in $\mathbb{R}^d$, giving $(N-1)d$ independent scalar coordinates. For generic ICs, the equations of motion couple all Jacobi coordinates, so the trajectory explores the full $(N-1)d$-dimensional subspace.

(ii) If G permutes the bodies, the trajectory is G-equivariant: $X(t)$ lies in the G-invariant subspace. This reduces the effective dimension. For C_3 acting on 3 equal-mass bodies: the invariant subspace of the Jacobi coordinates has dimension 4 (both ρ and R transform nontrivially), but rigid rotation (Lagrange) further constrains to the 2D subspace where $|\rho|$ and $|R|$ are constant.

(iii) By the Paley-Wiener theorem, a function analytic in $|\mathrm{Im}(t)| < \delta$ has Fourier coefficients decaying as $|c_k| \le Ce^{-2\pi k\delta/T}$. Setting $\rho = e^{2\pi\delta/T}$ gives the stated bound. The analyticity strip width δ is determined by the nearest complex-time singularity — typically a binary collision, where $|r_i - r_j| \to 0$. □

Corollary (3-body in 2D). The Latent of any analytic three-body orbit in two dimensions has kinematic rank exactly 4 (for generic ICs) and total size $4 \cdot N_\varepsilon$. For the figure-8 orbit: $\rho = 1.18 \pm 0.02$, $N_\varepsilon(10^{-6}) = 18$, total size = 72 reals. For the hierarchical triple: $\rho = 1.03 \pm 0.01$, $N_\varepsilon(10^{-6}) = 112$, total size = 448 reals.

5.3 Three-Body Latent: Financial Crash Clustering

Consider three assets with returns (R_1, R_2, R_3) exhibiting: - Pairwise correlations: $\mathrm{Corr}(R_i, R_j) = \rho_{ij}$ (grade-2 Latent) - Non-zero co-skewness: $\mathbb{E}[R_i R_j R_k] \neq$ Gaussian prediction (grade-3 Latent)

Corollary (Grade-3 financial Latent). Let $(R_1, \dots, R_d)$ have joint density with co-skewness tensor $S \in \mathcal{H}^{\otimes 3}$. Then:

(i) The rank of S equals the minimum number of independent "skewness sources."

(ii) Eigenvalue conditioning with K eigenvalues captures at most $\binom{K}{3}$ components of S — zero for $K < 3$.

(iii) The full Latent of the joint distribution is the graded sum $\Lambda = \Lambda^{(1)} + \Lambda^{(2)} + \Lambda^{(3)} + \cdots$ where the grade-r component encodes the r-th order cumulant structure.

Practical consequence. Market crashes exhibit strong co-skewness: three assets crash together more often than any pairwise copula predicts. The grade-2 Latent (eigenvalue conditioning) understates tail risk. The grade-3 Latent captures the excess crash probability directly.

5.4 Grade Collapse and the Effective Grade Signature

Not all grades carry independent information in every system. A central structural phenomenon, revealed by the Riemannian manifold instance (§8.9), is **grade collapse**: the grade-r component may be fully determined by lower grades, contributing nothing new.

Definition 9 (Effective grade signature). The *effective grade signature* of a system $\mathcal{S}$ with Latent $\Lambda = \bigoplus_r \Lambda^{(r)}$ is the set of grades that carry independent information:

$$\Sigma(\mathcal{S}) = \{r \geq 0 : \Lambda^{(r)} \text{ is not determined by } \Lambda^{(0)}, \dots, \Lambda^{(r-1)}\}$$

Theorem (Grade Collapse). Let $\mathcal{S}$ be a system defined by a self-adjoint operator A on $L^2(\Omega, \mu)$, and let the Latent be extracted in A's eigenbasis $\{\phi_k\}$. Then the grade-2 component (the operator A itself) is diagonal: $A\phi_k = \lambda_k \phi_k$, hence determined by grade 1 (the eigenvalues). Therefore $2 \notin \Sigma(\mathcal{S})$.

Proof. In the eigenbasis, the matrix representation of A is $A_{ij} = \lambda_i \delta_{ij}$. This is a function of the eigenvalues alone — the grade-1 data. Any grade-2 quantity derivable from A (heat kernel, Green's function, resolvent) is determined by $\{\lambda_k\}$. The first new information appears in the trilinear structure constants $c_{ijk} = \int \phi_i \phi_j \phi_k \, d\mu$, which live at grade 3. □

The grade signature classifies systems by their information architecture:

System	Σ	Explanation
Diffusion on domain	$\{1, 2, 3\}$	State, generator, coupling all independent
N-body dynamics	$\{1, 2, 3\}$	Conserved quantities, generator, three-body interaction
Portfolio of d assets	$\{1, 2, 3\}$	Marginals, correlation, co-skewness
Riemannian manifold	$\{1, 3\}$	**Grade 2 collapses** — Laplacian is diagonal in eigenbasis
Kerr black hole	$\{1\}$	**All grades collapse** — 3 parameters suffice (no-hair)

System	Σ	Explanation
Integrable Hamiltonian	$\{1, 2\}$	Action-angle variables, grade 3 vanishes (no chaos)

Remark (Process Latent vs Object Latent). The grade signature reveals a fundamental dichotomy. In a *process* (diffusion, N-body), the system has an independent state (grade 1) and independent dynamics (grade 2) — these are decoupled pieces of information, so $\{1, 2\} \subset \Sigma$. In an *object* (Riemannian manifold, drum), the system IS the operator — the operator's eigenbasis is the canonical basis, and the operator is diagonal in it, so grade 2 collapses. The grade-2 gap $2 \notin \Sigma$ is a structural fingerprint of "the system is an operator, not a process."

Remark (Why Kac's question was hard). The grade signature of a Riemannian manifold is $\{1, 3\}$. There is no grade-2 stepping stone between the eigenvalue spectrum (grade 1) and the structure constants (grade 3). This is why the inverse spectral problem remained open for 60 years: the most natural "next level" of information after eigenvalues — a grade-2 operator — adds nothing. The answer lies at grade 3, which is a qualitative jump requiring the eigenfunction multiplication table rather than any operator built from the eigenvalues.

Chapter 6. The Latent Algebra

6.1 Operations

The Latent space $\mathfrak{L}(\mathcal{H})$ supports four fundamental operations:

Addition. For Latents $\Lambda_1, \Lambda_2 \in \mathcal{H}^{\otimes r}$ of the same grade:

$$\Lambda_1 + \Lambda_2 \in \mathcal{H}^{\otimes r}$$

This is the basis of federated combination: average the Latents of n hospitals to get the combined Latent.

Scalar multiplication. $\alpha\Lambda \in \mathcal{H}^{\otimes r}$ for $\alpha \in \mathbb{R}$.

Tensor product. For $\Lambda_1 \in \mathcal{H}^{\otimes p}$ and $\Lambda_2 \in \mathcal{H}^{\otimes q}$:

$$\Lambda_1 \otimes \Lambda_2 \in \mathcal{H}^{\otimes (p+q)}$$

This constructs higher-grade Latents from lower-grade ones.

Contraction. For $\Lambda \in \mathcal{H}^{\otimes r}$, contracting on indices (i, j):

$$C_{ij}(\Lambda) \in \mathcal{H}^{\otimes (r-2)}$$

This projects higher-grade Latents down. Contracting a grade-3 Latent gives a grade-1 Latent: how three-body interactions project to effective single-body behavior.

6.2 Basis Invariance of the Algebra

Proposition 2. All four operations commute with basis change. Performing an operation in one set of coordinates and then transforming gives the same result as transforming first and then operating.

Proof. All operations are (multi)linear. Basis change is a unitary linear map. Unitary maps commute with linear operations. □

This means: the Latent Algebra is intrinsic. It does not depend on how you extracted the Latents. Two organizations using different extraction bases can combine their Latents and get the same result.

6.3 Universality of the Latent Algebra

The four operations (addition, scalar multiplication, tensor product, contraction) together with the inner product and grading are not an arbitrary choice. The Latent Algebra is the *universal* structure satisfying these axioms — the smallest algebra that supports all the operations we need.

Definition 8 (Graded contraction algebra). A *graded contraction algebra* over $\mathbb{R}$ is a graded real algebra $A = \bigoplus_{r=0}^{\infty} A_r$ equipped with: - (GCA-1) *Inner product.* An inner product $\langle \cdot, \cdot \rangle_r$ on each A_r, with respect to which A_r is a (pre-)Hilbert space. - (GCA-2) *Inner-product-multiplicative product.* A bilinear product $\otimes_A : A_p \times A_q \to A_{p+q}$ satisfying $\langle a_1 \otimes_A a_2, b_1 \otimes_A b_2 \rangle_{p+q} = \langle a_1, b_1 \rangle_p \cdot \langle a_2, b_2 \rangle_q$ on simple tensors (and extended bilinearly). - (GCA-3) *Inner-product-determined contraction.* Contraction maps $C_{ij}^A : A_r \to A_{r-2}$ for $1 \le i < j \le r$ satisfying, on simple tensors, $C_{ij}^A(a_1 \otimes_A \cdots \otimes_A a_r) = \langle a_i, a_j \rangle_{A_1} \cdot (a_1 \otimes_A \cdots \widehat{a_i} \cdots \widehat{a_j} \cdots \otimes_A a_r)$.

A *morphism* of graded contraction algebras is a grade-preserving linear map that respects the product, inner products, and contractions.

Remark (which axiom carries the work). (GCA-1) and (GCA-3) are mostly bookkeeping — any Hilbert-like structure on the graded pieces and any contraction compatible with the inner product. The axiom that carries the content is (GCA-2): inner-product-multiplicativity of the tensor product on *simple* tensors. It is this axiom that ensures the universal map Φ in Proposition 7 is *isometric* on each grade (not merely linear), which in turn forces contraction compatibility via (GCA-3). Without (GCA-2), Proposition 7 would only produce a linear graded morphism, not a graded Hilbert-tensor isomorphism — the geometry would not transfer. We flag this because Mirtill-style adversarial reviewers have pointed to the interplay of these axioms as the load-bearing step of §6.3.

Proposition 7 (Initial graded contraction algebra). Let $\mathcal{H}$ be a separable real Hilbert space. Then $\mathfrak{L}(\mathcal{H}) = \bigoplus_{r=0}^{\infty} \mathcal{H}^{\otimes r}$, equipped with the operations of (a) direct sum, (b) tensor product, (c) the inherited Hilbert-tensor inner product, and (d) contraction induced by the inner product of $\mathcal{H}$, is the *initial* object in the category of graded contraction algebras generated by $\mathcal{H}$: for any graded contraction algebra A and any isometric linear embedding $\phi : \mathcal{H} \to A_1$, there exists a unique graded contraction algebra morphism $\Phi : \mathfrak{L}(\mathcal{H}) \to A$ extending ϕ.

Proof. The existence and uniqueness of Φ on the underlying tensor algebra $T(\mathcal{H})$ is classical (Bourbaki, *Algèbre* III, §5.1 and §6.1): $T(\mathcal{H})$ is free on $\mathcal{H}$ in the category of associative unital algebras. What we check here is compatibility with the extra structure (inner product and contraction).

(i) *Inner-product compatibility on each grade.* By (GCA-2), the product $\otimes_A$ of A is inner-product-multiplicative: $\langle a_1 \otimes_A a_2, b_1 \otimes_A b_2 \rangle_A = \langle a_1, b_1 \rangle_A \cdot \langle a_2, b_2 \rangle_A$. Since ϕ is isometric — $\langle \phi(h), \phi(h') \rangle_A = \langle h, h' \rangle_{\mathcal{H}}$ — the defined Φ preserves inner products on simple tensors and, by linearity and continuity, on the completion $\mathcal{H}^{\otimes r}$.

(ii) *Contraction compatibility.* Let $C_{ij} : \mathcal{H}^{\otimes r} \to \mathcal{H}^{\otimes (r-2)}$ denote contraction of the i-th and j-th slots via the inner product of $\mathcal{H}$. By (GCA-3), A has an analogous contraction C_{ij}^A on its graded pieces that reduces to the inner product of A_1. On simple tensors, $\Phi(C_{ij}(h_1 \otimes \cdots \otimes h_r)) = \langle h_i, h_j \rangle \cdot \Phi(h_1 \otimes \cdots \widehat{h_i} \cdots \widehat{h_j} \cdots \otimes h_r)$ and the right-hand side equals $C_{ij}^A(\Phi(h_1 \otimes \cdots \otimes h_r))$ because ϕ is isometric. Extending by linearity and density gives $\Phi \circ C_{ij} = C_{ij}^A \circ \Phi$ on all of $\mathcal{H}^{\otimes r}$.

(iii) *Basis-change automorphism.* A unitary $U : \mathcal{H} \to \mathcal{H}$ induces $\mathfrak{L}(U) : \mathfrak{L}(\mathcal{H}) \to \mathfrak{L}(\mathcal{H})$ commuting with all four operations; this says that the universal property holds naturally in $\mathcal{H}$, i.e., coordinate representations are non-canonical and the Latent is basis-independent. □

Consequence for this paper. Proposition 7 answers the question: "Why the Latent Algebra and not some other algebraic structure?" Because there is no simpler graded algebra with inner product and contraction that a separable Hilbert space can embed into isometrically. Any system that requires grade decomposition (multi-body interactions), tensor products (combining subsystems), AND contraction (computing observables by tracing over indices) must use the Latent Algebra or a quotient of it. The Latent is not one of many possible choices — it is the unique minimal structure supporting all the operations that the representation theory of smooth systems requires.

What is classical and what is new. The pure algebra — that $T(\mathcal{H})$ is the free associative unital algebra on $\mathcal{H}$ — is classical (Bourbaki). We therefore call Proposition 7 a *proposition* rather than a theorem: its content is principally the *packaging* of this classical fact into a category whose objects have inner products and contractions, together with the verification that (GCA-2) and (GCA-3) (see §6.1 axioms) are exactly what is needed for $T(\mathcal{H})$ to be initial in that category. What is not classical is: (a) the identification of the inner-product-and-contraction-enhanced tensor algebra as the natural home for spectral representations of stochastic systems, (b) the observation that contraction gives physical observables (pricing, risk measures, correlations) a first-class algebraic status, and (c) the resulting principle that basis invariance is a *theorem of the framework*, not a design choice.

6.4 Latent Transfer

Theorem 5 (Latent transfer). A Latent known through its coordinates $\Lambda^{(\mathcal{B})}$ in basis $\mathcal{B}$ (truncated to N modes) can be re-expressed in any other basis $\mathcal{B}'$ via the change-of-basis matrix U:

$$\Lambda^{(\mathcal{B}')}_{k_1 \cdots k_r} = \sum_{j_1, \ldots, j_r \le N} U_{k_1 j_1} \cdots U_{k_r j_r} \, \Lambda^{(\mathcal{B})}_{j_1 \cdots j_r} + O(\rho_{\mathcal{B}}^{-N})$$

The cost is $O(r \cdot N^{r+1})$ — independent of the original system.

Application. A Knowledge Artifact (grade-1 Latent) extracted in cosines can be transferred to: - Hermite basis for tail analysis - Wavelet basis for multi-scale analysis - A learned basis optimized for a specific task

No re-extraction from the original model is needed. Extract once, transfer everywhere.

6.5 Maps (Latent Endomorphisms)

The four operations of §6.1 manipulate Latents as algebraic objects — adding, combining, contracting, re-expressing. But every computational pipeline also applies *functions* to Latents: $S \mapsto \log S$, $S \mapsto e^{itS}$, $S \mapsto \cos(a \log S)$. These are **maps**: grade-preserving endomorphisms of the Latent space.

Definition 9 (Latent map). Let $\Lambda \in L^2(\Omega, \mu)$ be a grade-1 Latent (a square-integrable function on a probability space). For a measurable $f : \mathbb{R} \to \mathbb{R}$, the **pushforward map** is:

$$f_* : L^2(\Omega, \mu) \to L^2(\Omega, \mu), \qquad (f_* \Lambda)(\omega) = f(\Lambda(\omega))$$

when $f \circ \Lambda \in L^2(\Omega, \mu)$.

Proposition 3 (Composition). Maps compose as functions: $(f \circ g)_* = f_* \circ g_*$. The identity map $\mathrm{id}_* = \mathrm{Id}$. The maps on $L^2(\Omega, \mu)$ form a monoid under composition.

Proposition 4 (Maps and contractions). The fundamental interaction between maps and contractions is:

$$\langle f_*\Lambda,\, \mathbf{1}\rangle_\mu = E_\mu[f(\Lambda)] = \int_\Omega f(\Lambda(\omega))\, d\mu(\omega)$$

Every contraction of a mapped Latent is an expectation. Every expectation is a contraction of a mapped Latent. This equation is exact in the algebra — the left side is an inner product in $L^2(\mu)$, not an approximation.

Proposition 5 (Boundedness and convergence). The map f_* is well-conditioned for contraction when $f_*\Lambda$ is bounded:

(a) If $|f(x)| \le M$ for all x in the essential range of Λ, then $\|f_*\Lambda\|_\infty \le M$ and $\|f_*\Lambda\|_2 \le M$. Gauss-type quadrature for $\langle f_*\Lambda, \mathbf{1}\rangle$ converges exponentially.

(b) If $f(x) = x^k$ (the moment map), then $\|f_*\Lambda\|_2 = (E[\Lambda^{2k}])^{1/2}$, which diverges as $k \to \infty$ for heavy-tailed Λ. Quadrature for $\langle f_*\Lambda, \mathbf{1}\rangle$ may diverge.

This is the precise mechanism of the Fenton divergence: the moment map $f(x) = x^k$ and the Fourier map $f(x) = e^{itx}$ both produce grade-1 Latents that are then contracted against $\mathbf{1}$. But $|e^{itx}| = 1$ (bounded, always convergent) while $|x^k| \to \infty$ (unbounded, divergent for lognormal sums). The same algebraic contraction, applied to two different maps, converges or diverges.

Proposition 6 (Analyticity preservation). If Λ has analyticity parameter $\rho > 1$ (Latent coordinates decay as ρ^{-k}) and f is analytic with convergence radius $R > \|\Lambda\|_\infty$, then $f_*\Lambda$ also has $\rho_{f_*\Lambda} \ge \rho$. Analyticity is preserved by analytic maps.

Instance table:

Map f	$f_*\Lambda$	Bounded?	Contraction convergence
e^{itx} (Fourier)	$e^{it\Lambda}$	Yes: $\|f_*\Lambda\|_\infty = 1$	Exponential (always)
x^k (moment)	Λ^k	No: $\|\Lambda^k\|_\infty \to \infty$	Divergent (heavy tails)
$\cos(ax)$ (COS kernel)	$\cos(a\Lambda)$	Yes: $\|f_*\Lambda\|_\infty = 1$	Exponential (always)
$\log(x)$ (log-space)	$\log\Lambda$	Conditional: needs $\Lambda > 0$ a.s.	Depends on tail of $\log\Lambda$
Indicator $\mathbf{1}_{x \le c}$	$\mathbf{1}_{\Lambda \le c}$	Yes: $\|f_*\Lambda\|_\infty = 1$	Algebraic (discontinuous f)

The table reveals a structural principle: **bounded maps compose well with contractions; unbounded maps may not.** The choice of map determines whether the pipeline converges, not the distribution itself.

The Fenton pipeline as a map-contraction chain. The COS coefficient A_k decomposes into four maps and one contraction:

$$A_k = \left\langle \cos\!\left(k\pi \cdot \frac{(\cdot) - \alpha}{\Delta}\right)_* \circ \log_* \circ S_*,\ \ \mathbf{1} \right\rangle_{\gamma_n}$$

where $S_*(z) = \sum w_i \exp(\mu_i + \ell_i^T z)$ is the first map (parameters to random variable), $\log_*$ is the second (log-space transform), and $\cos(k\pi(\cdot - \alpha)/\Delta)_*$ is the third (COS kernel). Every map is analytic and the composition is bounded ($|\cos(\cdot)| \leq 1$), so the contraction converges exponentially. By contrast, the moment computation $m_k = \langle (\cdot)_*^k \circ S_*, \mathbf{1} \rangle_{\gamma_n}$ uses the unbounded map x^k, and the contraction diverges.

6.6 Complexification

The Latent Algebra of §6.1–6.3 is defined over $\mathbb{R}$: "Let $\mathcal{H}$ be a separable *real* Hilbert space." This is sufficient for the static algebraic structure but not for Fourier analysis, analytic continuation, or monodromy — the three tools that make the COS representation exact and the Smale-6 finiteness proof work. These tools require the **complex extension** of the algebra.

Definition 10 (Complexification). The **complexified Latent Algebra** is $\mathfrak{L}(\mathcal{H}_\mathbb{C}) = \bigoplus_{r=0}^{\infty} \mathcal{H}_\mathbb{C}^{\otimes r}$, where $\mathcal{H}_\mathbb{C} = \mathcal{H} \otimes_\mathbb{R} \mathbb{C}$ is the standard Hilbert-space complexification. The real algebra embeds canonically: $\mathfrak{L}(\mathcal{H}) \hookrightarrow \mathfrak{L}(\mathcal{H}_\mathbb{C})$.

Theorem 8 (Complexification enrichment). *The complexified algebra $\mathfrak{L}(\mathcal{H}_\mathbb{C})$ has strictly more structure than $\mathfrak{L}(\mathcal{H})$:*

(a) Fourier decomposition. For $\Lambda \in L^2(\Omega, \mu)$ with analytic density, the Fourier map $e_*^{it(\cdot)}$ produces the characteristic function $\hat{\Lambda}(t) = \langle e^{it\Lambda}, \mathbf{1} \rangle_\mu \in \mathbb{C}$. The Fourier inversion theorem recovers Λ's distribution exactly from $\{\hat{\Lambda}(t)\}_{t \in \mathbb{R}}$. For compactly supported (or exponentially decaying) densities, the Fourier coefficients at discrete frequencies $\{A_k\}$ suffice — the COS representation is exact, not approximate.

(b) Analyticity and exponential convergence. If Λ has an analytic density, the Fourier coefficients A_k decay exponentially: $|A_k| \leq Ce^{-\delta k}$ where $\delta > 0$ depends on the analyticity strip width. This gives the COS convergence rate $\|F - F_N\|_\infty = O(e^{-\delta N})$. This property has no real-algebra counterpart — it requires the Paley-Wiener theorem, which is intrinsically complex.

(c) Monodromy. Analytic continuation of $\Lambda(z)$ along paths in $\mathbb{C}$ reveals topological structure: branch points, winding numbers, Riemann surfaces. This structure constrains the solutions of algebraic equations (e.g., the finiteness of central configurations, Finding 14). Monodromy is invisible in $\mathfrak{L}(\mathcal{H})$ and only exists in $\mathfrak{L}(\mathcal{H}_\mathbb{C})$.

Proof sketch. (a) is the Fourier inversion theorem (Rudin, 1987). (b) follows from the Paley-Wiener theorem: the Fourier transform of a function analytic in a strip of width δ decays as $e^{-\delta|t|}$ (Katznelson, 2004). (c) is the monodromy theorem (analytic continuation along homotopically equivalent paths gives the same result; non-trivial monodromy requires branch points). □

Corollary (Complexification is not a technique). Complexification is a Latent operation — it extends the Latent algebra from $\mathbb{R}$ to $\mathbb{C}$, accessing structure (Fourier theory, analyticity, monodromy) that does not exist in the real algebra. The grade-3 Latent choice (which basis to use, Theorem 4) is equivalent to the choice of complexification:

Grade-3 choice	Complexification	Structure revealed
Hermite-COS for lognormal sums	$\phi_S(t) = E[e^{itS}]$ (Euler: $e^{itS} = \cos + i \sin$)	Fourier: COS series = exact CDF
Fourier for periodic orbits	$e^{i\omega t}$	Fourier: exponential coefficient decay
Pair displacements for CCs	$\phi_{ij}(z) \in \mathbb{C}$	Monodromy: branch points → finiteness

Euler's $e^{ix} = \cos x + i \sin x$ is the prototype: the sin and cos are not approximations of the complex exponential — they are its real and imaginary projections. In the same way, the COS coefficients are real projections of the complexified distribution (the characteristic function), and the COS series *expresses* the CDF in its Fourier coordinates rather than approximating it.

6.7 The Evaluator Layer

The previous subsections define exact algebraic operations: maps (§6.5), contractions (§6.1), complexification (§6.6). These are well-defined as operations on the Hilbert space $L^2(\Omega, \mu)$ — no precision parameter, no truncation, no convergence question. In practice, every algebraic operation requires an **evaluator**: a computable function that realizes the operation to a given precision.

Definition 11 (Evaluator). An **evaluator** $\hat{\Phi}_\varepsilon$ for a Latent operation Φ is a computable function such that $\|\hat{\Phi}_\varepsilon(\Lambda) - \Phi(\Lambda)\| < \varepsilon$ for all Λ in a specified class, where ε is the precision parameter. The **convergence rate** of $\hat{\Phi}_\varepsilon$ is the function $\varepsilon(N)$ where N is the computational parameter (number of quadrature nodes, series terms, or arithmetic operations).

Principle (Algebra-evaluator separation). The algebraic operation is exact and basis-free. The evaluator has a precision parameter, a convergence rate, and a computational cost. Convergence and divergence are properties of the evaluator, not of the algebra.

Algebraic operation	Evaluator	Parameter	Rate
Contraction $\langle f_*\Lambda, \mathbf{1}\rangle_\mu$	Gauss-Hermite quadrature	Q (nodes)	$O(e^{-cQ})$ if $f_*\Lambda$ bounded
Contraction $\langle f_*\Lambda, \mathbf{1}\rangle_\mu$	Monte Carlo	M (samples)	$O(M^{-1/2})$ always
Contraction $\langle \Lambda^k, \mathbf{1}\rangle_\mu$	Moment recurrence	k (order)	Divergent for heavy tails
Synthesis $\sum A_k \sin(k\pi u)$	COS truncation	N (terms)	$O(e^{-\delta N})$ if density analytic
Map $f_*\Lambda$	Pointwise evaluation	—	Exact (for elementary f)
Basis change $U\Lambda$	Matrix-vector product	—	Exact (to machine precision)

Theorem 8b (Evaluator divergence is basis-dependent). *For the contraction $E_\mu[f(\Lambda)]$ of a grade-1 Latent $\Lambda \in L^2(\gamma_n)$:*

(a) If $f = e^{it(\cdot)}$ (Fourier map), then $|f(\Lambda(\omega))| = 1$ for all ω, and Gauss-Hermite quadrature with Q nodes converges as $O(e^{-cQ})$.

(b) If $f = (\cdot)^k$ (moment map) and $\Lambda = S = \sum w_i e^{Y_i}$ with $Y \sim N(\mu, \Sigma)$, then $E[S^{2k}] \geq e^{k^2\sigma^2_{\max}}$, and the Toeplitz condition number built from $\{E[S^k]\}$ satisfies $\kappa \geq e^{cN_P^2}$ for Padé order N_P.

Both (a) and (b) evaluate the same distribution. The convergence/divergence is a property of the evaluator (which map f), not of the Latent.

Proof. (a): $|e^{itS(z)}| = 1$ for all $z \in \mathbb{R}^n$. The integrand $g(z) = e^{itS(z)} \cdot \gamma_n(z)$ is bounded by $\gamma_n(z)$, which is the GH weight function. By the theory of Gaussian quadrature, the error decays exponentially in Q for integrands that are analytic in a strip containing $\mathbb{R}^n$ — and $e^{itS(z)}$ is entire in z for fixed t. (b): $E[S^k] = E[(\sum w_i e^{Y_i})^k] \geq (w_{\min})^k E[e^{kY_1}] = (w_{\min})^k e^{k\mu_1 + k^2\sigma_1^2/2}$. The exponential growth $e^{k^2\sigma^2/2}$ implies Carleman condition failure ($\sum m_{2k}^{-1/(2k)} < \infty$), and the Toeplitz matrix $T_{ij} = m_{i+j}$ has $\kappa(T) \geq \max m_k / \min m_k \geq e^{cN_P^2}$. □

Remark (The Fenton problem as an evaluator failure). The 65-year open problem of computing the CDF of correlated lognormal sums was never an algebraic problem — the contraction $E[f(S)]$ is well-defined for any L^2-compatible map f. It was an evaluator failure: the standard evaluator (moment expansion, $f = x^k$) diverges. The solution was not a new algebraic construction but a new evaluator: the Fourier map $f = e^{itx}$, which is bounded and produces exponential convergence. The algebra was always exact; only the evaluator needed to change.

6.8 Derived Operations

The seven primitives (addition, tensor product, contraction, basis change, map, complexification, evaluator) generate a family of derived operations that appear throughout the companion papers. Each derived operation decomposes into primitives but is used frequently enough to deserve a name.

Semigroup action (grade-2 acts on grade-1). A grade-2 Latent $M \in \mathcal{H}^{\otimes 2}$ — the Fokker–Planck generator, Stokes operator, or transfer matrix — acts on a grade-1 Latent $\Lambda \in \mathcal{H}$ by contraction on one index:

$$(M \cdot \Lambda)_k = \sum_j M_{kj} \Lambda_j$$

This is a single contraction: $C_2(M \otimes \Lambda) \in \mathcal{H}$. The time-evolved Latent $\Lambda(t) = e^{tM}\Lambda(0) = \sum_{n=0}^{\infty} \frac{t^n}{n!} M^n \Lambda(0)$ is a power series in repeated contractions. This is how dynamics works in the Latent Algebra: the grade-2 Latent IS the generator, and its exponential IS the semigroup that evolves grade-1 Latents (distributions, densities, mode amplitudes). Every diffusion, Markov chain, and Hamiltonian flow is an instance of this operation.

Companion papers: Tensor-spectral representation (Nagy 2026, *The Spectral Tensor Representation of Stochastic Processes*), where the generator matrix M_{kj} encodes the complete diffusion process and e^{tM} propagates the COS-coefficient Latent; spectral fusion (Nagy 2026, *The Knowledge Artifact and Knowledge Algebra of ML Models*), where eigenvalue tracking is the diagonal form of e^{tM}.

Partial contraction / conditioning. Given a grade-2 Latent $\Lambda^{(X,Y)} \in \mathcal{H}_X \otimes \mathcal{H}_Y$ representing a joint distribution, conditioning on $Y = y$ is a partial contraction that produces a grade-1 Latent in $\mathcal{H}_X$:

$$\Lambda_k^{(X|Y=y)} = \sum_j \Lambda_{kj}^{(X,Y)} \, \delta_y(j) \, / \sum_k \sum_j \Lambda_{kj}^{(X,Y)} \, \delta_y(j)$$

More generally, the conditional map $\Phi_{Y \to X} : \mathcal{H}_Y \to \mathcal{H}_X$ defined by $\Phi_{Y \to X}(\psi) = C_2(\Lambda^{(X,Y)} \otimes \psi)/\|C_2(\Lambda^{(X,Y)} \otimes \psi)\|$ is a contraction on the Y-factor followed by normalization. This is how Bayesian inference operates in the Latent Algebra: the prior is a grade-1 Latent, the likelihood is a grade-2 Latent (the conditional), and the posterior is a partial contraction.

Companion paper: Latent probability theory (Nagy 2026), where $P(X|Y)$ is a typed linear map between Hilbert factors.

Spectral decomposition (analysis of grade-2 Latents). Every symmetric grade-2 Latent $M \in \mathcal{H}^{\otimes 2}$ has an eigendecomposition:

$$M = \sum_k \lambda_k \, \phi_k \otimes \phi_k$$

where $\{\lambda_k\}$ are eigenvalues and $\{\phi_k\}$ are eigenvectors. This is the *analysis* direction — it decomposes a grade-2 Latent into its spectral components. The eigenvalues are grade-0 Latents (scalars obtained by contraction: $\lambda_k = C_{12}(M \otimes \phi_k \otimes \phi_k)$). The eigenvectors are grade-1 Latents. The spectral theorem guarantees this decomposition exists for all self-adjoint operators on $\mathcal{H}$.

Spectral decomposition is the inverse of synthesis: synthesis builds M from $\{\lambda_k, \phi_k\}$; decomposition recovers $\{\lambda_k, \phi_k\}$ from M. In the Fenton pipeline, the COS coefficients $\{A_k\}$ are obtained by contraction (grade-1 → grade-0), and the CDF is reconstructed by synthesis (grade-0 → function). In the Fokker-Planck pipeline, the generator M is decomposed by eigenanalysis (grade-2 → eigenvalues + eigenvectors), and the ten derived quantities (distribution, moments, prices, risk measures, spectral gap, mixing rate, etc.) are all functionals of the spectrum.

Companion papers: Tensor-spectral (Nagy 2026, *The Spectral Tensor Representation of Stochastic Processes*), spectral importance sampling (Nagy 2026, *The Spectral Fenton Distribution*), where eigenvalue structure of M determines both convergence rate and IS efficiency.

Duality (representation ↔ simulation). The Analyticity–Rate Duality (§3.2) is not an operation but a structural theorem about the algebra: the same analyticity parameter ρ that controls the representation error $\varepsilon \sim \rho^{-N}$ (COS convergence) also controls the simulation variance $\mathrm{Var}[\hat{\mu}] \sim e^{-2\rho N}$ (importance sampling efficiency). This means every Latent has two "faces" — a representation face (how many modes to truncate at) and a simulation face (how efficiently to sample from it) — and both are governed by the same algebraic invariant.

In the evaluator language (§6.7): the same contraction $\langle f_*\Lambda, \mathbf{1}\rangle_\mu$ can be evaluated by quadrature (representation face, deterministic, error $\sim \rho^{-N}$) or by Monte Carlo with an IS tilt (simulation face, stochastic, variance $\sim e^{-2\rho M}$). Both evaluators inherit their convergence rate from ρ, which is a property of the Latent — not of the evaluator. The evaluator determines the constant; the Latent determines the rate.

Companion paper: Spectral importance sampling (Nagy 2026, *The Spectral Fenton Distribution*), where the 17,000× variance reduction at $\rho \gg 1$ and IS failure at $\rho = 1$ are numerically validated.

Summary: the complete operation table.

Operation	Type	Grade change	Primitive / Derived
Addition	$\Lambda + \Lambda'$	$r \to r$	Primitive (§6.1)
Tensor product	$\Lambda \otimes \Lambda'$	$(r, s) \to r + s$	Primitive (§6.1)
Contraction	$C_{ij}(\Lambda)$	$r \to r - 2$	Primitive (§6.1)
Basis change	$U : \Lambda^{(\mathcal{B})} \to \Lambda^{(\mathcal{B}')}$	$r \to r$	Primitive (§6.1)
Map (endomorphism)	$f_*\Lambda$	$r \to r$	Primitive (§6.5)
Complexification	$\mathfrak{L}(\mathcal{H}_\mathbb{R}) \hookrightarrow \mathfrak{L}(\mathcal{H}_\mathbb{C})$	—	Primitive (§6.6)
Evaluator	$\hat{\Phi}_\varepsilon$ for Φ	—	Layer (§6.7)
Semigroup action	$e^{tM}\Lambda$	$(2, 1) \to 1$	Derived: iterated contraction (§6.8)
Conditioning	$\Lambda^{(X\|Y=y)}$	$2 \to 1$	Derived: partial contraction + normalization (§6.8)
Spectral decomposition	$M \to \{\lambda_k, \phi_k\}$	$2 \to (0, 1)$	Derived: eigenproblem (inverse of synthesis) (§6.8)

Every computational pipeline in the companion papers decomposes into sequences of these ten operations. The seven primitives are algebraically irreducible; the three derived operations are compositions of primitives that appear frequently enough in applications to deserve names.

6.9 Latent Algebra: The Composition of Latents

The preceding sections treated a single system with a single Latent. But optimization — perhaps the most common use of the framework — requires *two* Latents interacting: the Latent of the **system** and the Latent of the **objective**. Their interaction defines a compositional structure we call the **Latent Algebra**.

The Two-Latent Structure of Optimization

Consider the structural topology problem (§1.8, Example C). Under a load, a structure develops a strain energy field SE_i. This is the **structural Latent** Λ_S — the compressed representation of how forces flow. It is obtained from a single solve of the equilibrium equations.

But the same structure can be optimized for different objectives: minimum compliance, maximum resilience, equal stress distribution, minimum weight for given stiffness. Each objective defines an **objective Latent** Λ_O — a second spectral object encoding *what we want*.

The optimal design is neither Λ_S nor Λ_O alone, but their **product**:

$$\Lambda^* = \Lambda_S \otimes \Lambda_O$$

Concretely, for the structural design problem: the optimal bar areas are

$$A_i = V_{\text{target}} \cdot \frac{SE_i^\alpha}{\sum_j SE_j^\alpha \cdot L_j}$$

where SE_i comes from Λ_S (one solve), and α encodes Λ_O:

Objective Λ_O	α	Physical meaning
Uniform distribution	0	All bars equal — maximize redundancy
Maximum resilience	0.25	Broad distribution — fail-safe
Minimum compliance	**0.5–0.75**	**Concentrate on dominant force paths**
Fully Stressed Design	1.0	Every bar at its stress limit (Michell 1904)
Maximum concentration	$\to \infty$	Only the single most important bar

The same structural Latent Λ_S produces entirely different optimal designs depending on the choice of Λ_O. This is the algebraic formalization of the principle "Latent dimension is Projector-relative" (Remark after §3.2): the Projector IS the objective Latent, and the answer IS the algebra product.

The Algebra Structure

The set of Latents over a Hilbert space $\mathcal{H}$ carries three operations that make it an algebra:

Operation	Notation	Definition	Example
Direct sum	$\Lambda_1 \oplus \Lambda_2$	Combine Latents of independent systems	Two load cases
Tensor product	$\Lambda_1 \otimes \Lambda_2$	Compose Latents that interact	Structure × Objective → Design
Projection	$\Pi_k(\Lambda)$	Extract a specific observable	"What is the area of bar k?"

These live naturally in the Graded Tensor Algebra $\mathfrak{L}(\mathcal{H}) = \bigoplus_{r\geq 0} \mathcal{H}^{\otimes r}$ defined in §2. The direct sum is the grade-0 combination. The tensor product increases the grade: if $\Lambda_S \in \mathcal{H}^{\otimes r}$ and $\Lambda_O \in \mathcal{H}^{\otimes s}$, then $\Lambda_S \otimes \Lambda_O \in \mathcal{H}^{\otimes(r+s)}$. The projection Π_k reduces it back by contracting indices.

The algebra is bilinear and associative: - $(\Lambda_{S_1} \oplus \Lambda_{S_2}) \otimes \Lambda_O = (\Lambda_{S_1} \otimes \Lambda_O) \oplus (\Lambda_{S_2} \otimes \Lambda_O)$ - Multi-objective optimization: $\Lambda_S \otimes (w_1\Lambda_{O_1} + w_2\Lambda_{O_2})$ is a weighted combination of objectives — the Pareto front is the family of solutions parameterized by the weights.

Distinction from the Latent Tower

The **Latent Tower** (§3.4) is *recursive*: $\Lambda(\Lambda(x))$ — the Latent of a Latent, for systems whose first-level representation is infinite but has finite structure. It is a vertical operation: moving up levels of abstraction.

The **Latent Algebra** is *compositional*: $\Lambda_S \otimes \Lambda_O$ — two Latents at the same level encoding different aspects of the problem. It is a horizontal operation: combining a system with a question.

Both are necessary. The Tower handles infinite systems. The Algebra handles optimization. Together, they provide the complete operational calculus of the framework.

Why This Matters

Standard optimization (SIMP, genetic algorithms, gradient descent) treats the design as a *search problem*: start somewhere, iterate until convergence. The Latent Algebra recasts optimization as a *computation*: compute Λ_S (one solve), choose Λ_O (the objective), evaluate the product $\Lambda_S \otimes \Lambda_O$ (one formula). For the structural design problem, this means 1 FEM solve instead of ~100 iterations — and for certain problems (asymmetric loading, the tower example), the one-shot Latent solution is provably superior because it uses the global force field rather than iteratively refining a local guess.

The deeper point: the Latent Algebra reveals that **optimization is not search — it is projection**. The optimal design already exists inside the structural Latent; the objective Latent merely selects which aspect to project out. This is consistent with the ontological claim of Appendix A.1: the Latent contains *all* aspects of the system; the Projector determines which one we see.

The Complete Problem Latent: When Projection Suffices and When It Does Not

The preceding paragraphs claim that optimization is projection. A natural objection: *if the Latent is the reality, and the objective is also a Latent, shouldn't the composition always produce the provably optimal answer in one step?*

The answer is no — and understanding why reveals the Algebra's deepest structural insight.

The full Latent of an optimization problem is not two objects but **three**: the system Latent Λ_S, the objective Latent Λ_O, and the **composition operator** $\otimes$ itself. The operator is not a universal, problem-independent product. It depends on the mathematical structure of how the system and the objective interact.

When $\otimes$ is closed-form: If all three components — Λ_S, Λ_O, and $\otimes$ — are finite-grade ($\rho > 1$) and $\otimes$ has an explicit algebraic expression, then the optimum is obtained in one step: a projection, not a search. The compliance-sizing problem is exactly this case:

$$\Lambda_S = \{SE_i\} \quad \text{(strain energy, one solve)}, \qquad \Lambda_O = \alpha, \qquad \otimes : \quad A_i = V \cdot \frac{SE_i^{\alpha}}{\sum_j SE_j^{\alpha} L_j}$$

This is the Prager–Shield (1968) fully stressed design: provably optimal sizing for a fixed topology. The operator $\otimes$ is a closed formula. The Latent Algebra produces the mathematical optimum in one evaluation.

When $\otimes$ is combinatorial: For topology selection — *which bars should exist?* — the operator $\otimes$ is no longer a closed formula. It is a combinatorial decision over 2^N possible topologies. No spectral decomposition eliminates this combinatorics.

However, the Latent still provides critical structure: 1. The strain energy ranking from Λ_S **orders** the 2^N possibilities by relevance. 2. The top-ranked bars are almost certainly in the optimal topology. 3. The Latent-derived initialization reaches the basin of attraction of the global optimum faster than any uninformed starting point — empirically: 2× faster convergence and 0.7% better final compliance than uniform initialization.

The **boundary** between "optimization is projection" and "optimization is search" is precisely the boundary between closed-form and combinatorial $\otimes$. The Latent Algebra's diagnostic power is that it identifies this boundary *a priori*: given a problem class, one can determine whether $\otimes$ admits a closed form (in which case the one-step formula applies) or is inherently combinatorial (in which case the Latent serves as the optimal navigator for the search).

This is not a limitation of the framework — it is a structural theorem about optimization itself. Some problems have Latent-algebraic solutions; others have Latent-guided search. The Algebra tells you which is which.

Part IV — Unification and Applications

Every existing result in the spectral methods literature — the COS expansion, eigenvalue conditioning, the Fokker–Planck generator, Knowledge Artifacts — turns out to be a coordinate representation of a Latent in a specific basis. Chapter 7 makes that identification precise. Chapter 8 then validates the framework empirically across nine major application domains, from large language models and the classical three-body problem to Navier–Stokes turbulence, black holes, and closed Riemannian manifolds. Nine further domain case studies live in the Appendix C catalog.

Chapter 7. Every Spectral Result Is a Latent

7.1 The Spectral Fenton Distribution

The Spectral Fenton Distribution (Nagy 2026, *The Spectral Fenton Distribution*) is a vector $\{A_k\}_{k=0}^{N-1}$ of cosine coefficients representing a portfolio loss CDF.

Latent interpretation. It is the coordinate representation of a grade-1 Latent $\Lambda \in \mathcal{H}$ in the cosine extraction basis:

$$\Lambda = \sum_{k=0}^{N-1} A_k \, \phi_k^{(\cos)} \in \mathcal{H}, \qquad A_k = \Lambda_k^{(\cos)}$$

7.2 The Fokker–Planck Generator

The generator matrix M_{kj} (Nagy 2026, *The Spectral Tensor Representation of Stochastic Processes*) encodes a complete diffusion process.

Latent interpretation. It is the coordinate representation of a grade-2 Latent $\Lambda_{\mathcal{L}} \in \mathcal{H}^{\otimes 2}$ — the Fokker–Planck operator in cosine coordinates:

$$\Lambda_{\mathcal{L}} = \sum_{k,j} M_{kj}^{(\cos)} \, \phi_k^{(\cos)} \otimes \phi_j^{(\cos)}$$

The ten properties derivable from M (distributions, moments, prices, Greeks, risk measures, stationary distribution, spectral gap, autocorrelation, eigenvalue spectrum, mixing rate) are all functionals of the grade-2 Latent $\Lambda_{\mathcal{L}}$, computable in any extraction basis.

7.3 The Universal Spectral Representation Theorem

The USRT (Nagy 2026, *The Universal Spectral Representation Theorem*) is the Latent Theorem (Theorem 1) restricted to cosine extraction. The bound $N = \Theta(\log(1/\varepsilon)/\log\rho)$ is the same bound, with $\rho = \rho_{\cos}$.

7.4 Knowledge Artifacts

The Knowledge Artifact (Nagy 2026, *The Knowledge Artifact and Knowledge Algebra of ML Models*) extracts a vector of spectral coefficients from a trained ML model.

Latent interpretation. A Knowledge Artifact is a grade-1 Latent of a model's predictive distribution. The artifact algebra (addition, subtraction, scaling, combination) is the grade-1 Latent Algebra.

The current extraction uses the cosine basis. In the Latent framework, this is one choice of extraction. The model's Latent exists independently of how you compute it.

7.5 Eigenvalue Conditioning (Eigen-COS)

The Eigen-COS trick — diagonalize the covariance matrix, then apply 1D extraction per eigenmode — is a **factorization** of a grade-d Latent into a sequence of grade-1 Latents indexed by eigenvalues:

$$\Lambda^{(d)} \approx \sum_{\alpha=1}^{K} \lambda_\alpha \Lambda_\alpha^{(1)} \otimes \cdots$$

In Latent language, Eigen-COS is the rank-K approximation of a grade-d Latent. The framework shows this is one point in a spectrum of extraction strategies:

Strategy	Latent operation	Captures
Eigen-COS	Rank-K approximation of grade-d Latent	Pairwise correlations
+ co-skewness	Grade-2 + grade-3 Latent	Pairwise + three-body
Full grade-r extraction	Complete grade-r Latent	All interactions up to order r
Adaptive basis	Optimal extraction of grade-r Latent	All interactions, best convergence

7.6 Spectral Importance Sampling

The Eigen-COS factorization (§7.5) solves the density approximation problem: given a d-asset portfolio, decompose the correlation matrix into K eigenmodes, approximate each mode's characteristic function independently, and reconstruct the portfolio loss distribution via convolution. The analyticity parameter ρ controls the rate at which the COS spectral coefficients decay.

Spectral Importance Sampling is the simulation-side dual of this factorization. Instead of approximating the *density*, the goal is to estimate the *tail probability* $P(L > \ell)$ for a threshold ℓ deep in the tail. The standard approach — naive Monte Carlo — requires $O(1/p)$ samples to estimate a probability p with bounded relative error. For $p = 10^{-6}$ (the typical regulatory threshold for operational risk), this is computationally infeasible.

Importance sampling (IS) overcomes this by changing the sampling measure to concentrate on the tail. The classical result (Cramér, 1938; Glasserman & Li, 2005) is that exponential tilting — sampling from $dQ \propto e^{\theta X} dP$ with θ chosen by the saddle-point equation — achieves variance reduction that grows exponentially with ℓ.

The spectral IS result. The Eigen-COS decomposition factorizes this into K one-dimensional saddle-point problems. Let $\alpha_k = w^T v_k$ be the k-th mode's contribution to portfolio loss. The optimal tilt decomposes as:

$$\theta_k^* = \arg\min_{\theta_k} \left[\lambda_k \theta_k^2/2 - \alpha_k \theta_k \ell\right]$$

and the IS variance reduction satisfies $\log \mathrm{VR}(\ell) \sim C(\rho) \cdot \ell^2$ for Gaussian modes, where $C(\rho) > 0$ depends on the eigenvalue spectrum through ρ.

Latent interpretation. Spectral IS is the application of the grade-2 Latent to **simulation** rather than **approximation**. The Eigen-COS factorization decomposes a high-dimensional integral (the CDF) into independent 1D integrals. Spectral IS decomposes a high-dimensional measure change (the IS tilt) into independent 1D tilts. Both factorizations exploit the same structure — mode independence in the eigenbasis — and both are controlled by the same parameter ρ.

The Analyticity–Rate Duality (§3.6) explains why: the Bernstein ellipse radius ρ that governs spectral coefficient decay $|A_k| \leq C\rho^{-k}$ is mathematically identical to the Cramér rate parameter that governs IS efficiency $\mathrm{VR}(\ell) \sim \rho^{2\ell}$. Both arise from the width of the strip of analyticity of the characteristic function $\varphi(u) = \mathbb{E}[e^{iuX}]$ extended to complex u.

Spectral operation	COS (density)	IS (simulation)
Per-mode factorization	1D Fourier inversion per eigenmode	1D saddle-point per eigenmode
Convergence parameter	ρ: coefficient decay rate	ρ: variance reduction growth rate
$\rho > 1$ result	Exponential accuracy	Exponential VR (17,000× at $p = 10^{-5}$)
$\rho = 1$ result	Power-law convergence	No exponential VR (< 1×, IS counterproductive)
Grade required	Grade-2 Latent (eigenvalues λ_k)	Same grade-2 Latent

The benchmark validates this prediction numerically: for a 30-asset Gaussian equicorrelation portfolio ($\rho \gg 1$), spectral IS achieves VR = 38× at the 99th percentile growing to 17,867× at $p = 10^{-5}$, with log(VR) scaling quadratically in ℓ (the Gaussian rate function). For a multivariate Student-t portfolio with $\nu = 4$ degrees of freedom ($\rho = 1$), the same algorithm gives VR < 1 — importance sampling is worse than naive Monte Carlo, because the mixing variable couples all modes and breaks the independence assumption.

This extends the "every spectral result is a Latent" principle to stochastic simulation: the grade-2 Latent controls not only what can be **known** (density approximation) but what can be **explored** (rare event simulation). The boundary $\rho = 1$ is where both capabilities simultaneously collapse.

7.7 Formal Proofs as Latents

A Lean 4 formalization of a mathematical result is itself a Latent.

The formalization — the axioms, definitions, and proof terms accepted by the kernel — lives in $\mathfrak{L}(\mathcal{H})$ with an interesting grade structure:

Latent concept	Formal proof analogue
Latent Λ (compressed, basis-free)	The Lean source: axioms + proof terms
Projector Π (decode to observable)	The paper: narrative, figures, intuition

Latent concept	Formal proof analogue
Irreducible core (grade 0)	The axioms (not derivable from anything else)
Reconstructable part (grade 1+)	Proved theorems (Mathlib + axioms → automatic)
Analyticity parameter ρ	Axiom strength: how much can be derived per axiom
Hierarchy Diagnostic (§3.6)	The T1/T2/T3 axiom audit

The Lean formalization is basis-free: it is independent of notation, language, exposition style, or the author's pedagogical choices. It is finite: a finite set of axioms and proof terms. And it is sufficient: the kernel can reconstruct every theorem from it. This is exactly the definition of a Latent.

The paper that accompanies the formalization is a projection — a lossy decode into human-readable space, trading rigor for intuition. The paper is not the truth; the Lean is. The paper's job is to be a faithful Projector: to make the Latent's content accessible to a human observer.

The axiom audit as Hierarchy Diagnostic. The T1/T2/T3 audit developed for the Smale 6th Problem formalization (Nagy 2026, *The Exact Latent Solution of the Gravitational Three-Body Problem*) is a concrete instance of the Hierarchy Diagnostic. It tests whether each axiom carries actual information content:

- **(T1)** Can the axiom be proved by a trivial witness? (e.g. $\exists n$, True is satisfiable by $\langle 0, \text{trivial} \rangle$) — if yes, the axiom has **zero content**, a vacuous Latent.
- **(T2)** Does the conclusion depend on the hypothesis data? — if not, the witness is unbound and the axiom proves nothing about the system it claims to describe.
- **(T3)** Can the input type be constructed and used to derive False? — if yes, the axiom is **inconsistent**, asserting impossibility.

A formalization with N axioms may contain zero information if all axioms are trivially true (T1 failure) or inconsistent (T3 failure). This is the formal-proof analogue of a Latent with N coefficients but $\rho = 1$: many parameters, no content. A single well-typed, T1/T2/T3-audited axiom can carry more mathematical content than a hundred unaudited ones — just as a Latent with small N but large ρ outperforms one with large N but $\rho \approx 1$.

The ontological inversion for proofs. Historically, the mathematical paper is treated as primary and the formalization (if it exists) as secondary — a verification step. The Latent framework inverts this: the formalization IS the mathematical object. The paper is its projection into a human-observable space. This inversion is not merely philosophical; it has operational consequences. When the Smale 6th formalization compiled with zero `sorry` but three unaudited axioms, the paper appeared complete — but the Latent was partially vacuous. The content was in the audit, not in the compilation.

7.8 Spectral Knowledge Distillation

Knowledge distillation (Hinton et al., 2015) compresses a large "teacher" model into a smaller "student" by training on the teacher's soft probability outputs. The student is a smaller neural network — still a black box, with no guarantee on how much knowledge was preserved or lost. We show that spectral knowledge distillation — eigendecomposing the teacher's learned function on the data kernel — is a Latent operation with certified optimality.

The Latent interpretation. A trained model $f : \mathcal{X} \to \mathcal{Y}$ maps inputs to predictions. The model's learned function restricted to the data manifold $\mathcal{X}_{\text{data}}$ is a grade-1 Latent $\Lambda_f \in L^2(\mathcal{X}_{\text{data}}, \mu)$, where μ is the data

distribution. The eigendecomposition of the data kernel $K(x, x') = k(x, x')$ produces eigenfunctions $\{\psi_k\}$ with eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots$, and the spectral coefficients:

$$c_k = \langle f, \psi_k \rangle_\mu = \int f(x)\psi_k(x)\, d\mu(x)$$

are the coordinates of Λ_f in the eigenbasis. The spectral student with K modes is $\hat{f}_K(x) = \sum_{k=1}^{K} h_k c_k \psi_k(x)$, where $h_k = \lambda_k/(\lambda_k + \alpha)$ is the GCV-optimal shrinkage filter that replaces Hinton's manual temperature parameter.

The Eckart-Young guarantee (Theorem 15). The spectral student $\hat{f}_K$ is the best rank-K approximation to the teacher's function in $L^2(\mu)$:

$$\|f - \hat{f}_K\|_\mu^2 = \sum_{i>K} h_i^2 c_i^2 \leq \|f - g\|_\mu^2 \quad \text{for all } g \text{ with } \text{rank}(g) \leq K$$

No other K-parameter model — no smaller neural network, no pruned tree, no quantized representation — can achieve lower reconstruction error. The guarantee is not asymptotic; it holds for every K and every teacher.

The distillation chain. The spectral distillation has seven formally verified properties (proof kernel, `fields/residual_stream_denoising/platonic.py`, Theorems 8–14):

Property	Statement	Latent meaning
Error composition	$\varepsilon_{\text{total}} \leq \varepsilon_T(1 + 1/\rho)$	Teacher error + truncation error bounded
Monotone improvement	residual$(K+1) <$ residual(K)	Each mode strictly improves the student
ρ-sufficiency	$\varepsilon(K+1) < \varepsilon(K)/\rho$ for $\rho > 1$	Geometric convergence
Information ratio	First K modes capture $\geq 1 - 1/\rho$ of total energy	Few modes suffice when ρ is large
Gap-to-accuracy	Larger $\rho \to$ fewer modes for same accuracy	Smoother teachers are more distillable
Pipeline bound	$\|M - M_K\|_F^2 \leq \|M\|_F^2$	Distillation never increases total energy
Certified student	Error equals $\sum_{i>K} \sigma_i^2$, certifiable and optimal	The error is a computable, verifiable number

Hinton's temperature as ρ. In Hinton's framework, the temperature T controls the softness of the teacher's probability outputs: high T reveals "dark knowledge" — information about which wrong answers the teacher found plausible. In the Latent framework, ρ plays exactly this role. The eigenvalue spectrum $\{\lambda_k\}$ of the data kernel IS the temperature structure: large eigenvalues (dominant modes) correspond to "hard knowledge" (the main patterns), small eigenvalues to "dark knowledge" (subtle correlations). The GCV-optimal shrinkage $h_k = \lambda_k/(\lambda_k + \alpha)$ automatically calibrates how much dark knowledge to include — replacing Hinton's manual temperature search with an analytic optimum.

Hinton's framework	Latent framework
Teacher (large NN)	Reality Ω observed through model Lens
Student (smaller NN)	Rank-K Latent Λ_K
Temperature T	Analyticity parameter ρ
Soft targets	Spectral coefficients c_k with eigenvalue weighting
Student type: black box	Student type: explicit formula with certified error
Optimality: none	Optimality: Eckart-Young (Theorem 15, Lean 4 verified)

The practitioner's pipeline. Given any trained model (LLM, AlphaFold, diffusion model, random forest):

1. **Extract**: compute $c_k = \langle f, \psi_k \rangle$ on the data kernel → the model's Latent
2. **Diagnose**: fit ρ from the eigenvalue decay → the model's compressibility
3. **Truncate**: keep $K^* = \Theta(\log(1/\varepsilon)/\log\rho)$ modes → the optimal student
4. **Certify**: the error is $\sum_{i>K^*} \sigma_i^2$, computable and provably minimal (EY)
5. **Interpret**: each mode ψ_k has a physical or semantic meaning → white-box student

The full distillation chain is formally verified: 14 theorems in the proof kernel, with 479 Lean 4 lines (0 sorry). This means the optimality guarantee is not a claim — it is a machine-checked proof.

Chapter 8. Applications

This chapter validates the Latent framework across nine domains. The sections are ordered by evidentiary strength relative to the core theorems. The first two sections — **finance** (§8.2) and **the three-body problem** (§8.4) — constitute the primary empirical evidence: the finance benchmark is a controlled, falsifiable test of the Analyticity–Rate Duality; the three-body extraction demonstrates the higher-grade theory beyond pairwise methods. The LLM section (§8.1) provides supporting validation of the finite-sufficiency prediction. Sections 8.5–8.9 extend the framework to progressively more ambitious domains; sections 8.6–8.8 and the accompanying appendices invoke explicitly stated additional assumptions and belong to the **programmatic / conditional** tier (see §1.5a). Sections 8.10–8.11 have been relocated to appendices in v2.0 (see Appendix A.4 and Appendix B.5).

8.1 The Latent of a Large Language Model

A trained LLM has Latents at every grade. The grade-1 Latent (the Knowledge Artifact) captures what the model predicts; the grade-2 Latent (the weight SVD structure) captures what the model is; the grade-3 Latent (attention tensor decomposition) captures how the model's components interact. Full experimental details, including all nine empirical findings on GPT-2 and the 41-domain cross-architecture extension, are presented in the companion paper (Nagy 2026, *The Latent of a Large Language Model*). Here we summarize the structural framework and key results.

Grade 1 — what the model predicts (the Knowledge Artifact). The domain Latent $\hat{\Lambda}_{\mathcal{D}} = \mathbb{E}_{x\sim\mathcal{D}}[h_L(x)]$ — the mean last-layer hidden state — captures what the model knows about domain $\mathcal{D}$. Two models with similar grade-1 Latents have similar predictive behavior. This is the Knowledge Artifact (Nagy 2026, *The Knowledge Artifact and Knowledge Algebra of ML Models*).

Grade 2 — what the model IS (the weight structure). Each weight matrix W^l has a grade-2 Latent given by its truncated SVD $\Lambda^{(2)} = \{(U_K^l, \Sigma_K^l, V_K^l)\}_{l=1}^L$. LoRA operates on this grade: it represents $\Delta W = BA$ as a low-rank perturbation.

Grade 3 — how components interact (the attention tensor). The three-way interaction of W_Q^h, W_K^h, W_V^h per attention head defines a grade-3 tensor whose CP rank equals the number of independent attention patterns — connecting to mechanistic interpretability.

Status. Grade-1 is experimentally validated across 41 domains and 3 architectures (Nagy 2026, *The Latent of a Large Language Model*) — the effective rank 2–39 out of 768 reports *intra-domain* concentration of last-layer hidden states under mild L^2 tolerances; it does *not* directly establish the Latent Theorem hypothesis $\rho > 1$ on the full cross-domain prediction function (which would require measuring analytic decay of the predictive characteristic function, not a covariance-rank cutoff). The intra-domain concentration is, however, *necessary* for a finite Latent of the predictive distribution restricted to $\mathcal{D}$, and its strength across domains is consistent with — not proof of — finite-sufficiency in the theorem's sense. Grade-2 reveals a surprise whose interpretation deserves care (see "Smooth–rough duality" remark below). Grade-3 is a testable prediction.

Summary of Empirical Results (see Nagy 2026, *The Latent of a Large Language Model*, for full details)

We extracted Latents from GPT-2 (124M parameters) at grades 1 and 2 across five text domains (science, fiction, code, news, mathematics). The key findings establish a **smooth–rough duality**:

Finding	Grade	Result
Domain structure	1	Mean-centred Latents reveal formal/narrative axis (code↔news: -0.92)
Low rank	1	Effective rank 2–39 out of 768 hidden dimensions
Distillation invariance	1	GPT-2 ↔ DistilGPT-2 cosine similarity: 0.999
Latent arithmetic	1	Domain directions recover semantic ordering without labels
Weights are rough	2	$\rho \approx 1.00$, rank@99% ≈ 500–720 / 768; naive SVD compression fails
Positional embedding	2	Exception: rank 4 at 90% energy (> 90× compression)
Grade-1 survives grade-2 destruction	1+2	0.999 Latent similarity at perplexity > 10^5
ΔW is low-rank	2	Mean rank@90%: 110 vs. 439 (4× compression)
Token embedding ΔW rank = 8	2	Directly validates LoRA's $r = 4$–16 range

Smooth–rough duality (honest statement). The grade-1 object — per-domain hidden-state Latents — is empirically low-rank and behaves as if $\rho \gg 1$ on the restricted predictive distribution. The grade-2 object —

raw weight matrices — is empirically *not* low-rank ($\rho \approx 1$, rank@99% in the 500–720 range of the 768-dim space). We frame this honestly:

- *Negative result for naive weight compression.* Theorem 1 applied directly to the grade-2 weight matrices of a trained LLM *fails its precondition* — the weight spectra show no exponential decay. A naive "SVD-prune all weight matrices" pipeline produced by applying the Latent Theorem at grade 2 does not work, and our experiments confirm perplexity explodes under such pruning. This is a real boundary of the theorem's applicability that we want to expose, not hide.
- *Positive result on the perturbation.* The fine-tuning update ΔW between two trained models *is* low-rank (mean rank@90% of 110 vs. 439 for W, a 4× compression ratio; token-embedding ΔW rank = 8). This is consistent with LoRA's empirical success at $r = 4$–16. The Latent-framework interpretation is: ΔW lives tangent to the manifold of trained weights — a manifold whose dimension is small because the space of high-performing networks at fixed data is low-codimension. The "smooth" side of the duality is therefore a statement about the *tangent bundle* of the trained-weight manifold, not about the weights themselves.
- *The duality as framing, not theorem.* We call the pair ("grade 1 smooth / grade 2 rough, but ΔW smooth") a **duality** because it describes a reproducible empirical pattern, not because we have a theorem relating the two ranks. The natural theorem — $\text{rank}(\Delta W) \leq f(\text{rank}(\Lambda^{(1)}), \|\nabla_W \Lambda^{(1)}\|)$ — is conjectured here and targeted in the companion paper (Nagy 2026, *The Latent of a Large Language Model*).

The full experimental details, numerical results, and the LatentShell tool design are in Nagy (2026h). Experiment code: `llm_latent_experiment.py`, `llm_grade2_experiment.py`, `llm_delta_w_experiment.py`.

8.2 Non-Gaussian Tail Dependence in Finance

The Latent of a d-asset portfolio decomposes by grade:

$$\Lambda = \Lambda^{(1)} + \Lambda^{(2)} + \Lambda^{(3)} + \Lambda^{(4)} + \cdots$$

Grade 1: marginal distributions. Grade 2: pairwise correlations (the Eigen-COS computation). Grade 3: co-skewness (crash clustering). Grade 4: co-kurtosis (extreme co-movement). Each higher grade adds the non-Gaussian corrections that matter in the tails.

The practical consequence is direct: market crashes exhibit strong co-skewness — three assets crash together with higher probability than any pairwise copula predicts. For a three-asset subportfolio with co-skewness $\gamma_{123} = -0.3$, the Gaussian (grade-2-only) VaR at 1% is typically 15–25% too low. The grade-3 Latent correction captures this excess crash probability, closing the gap that has real regulatory consequences (Basel III capital requirements depend on VaR accuracy).

8.3 Federated Knowledge Fusion

The Knowledge Artifact framework supports federated learning through Latent averaging:

Grade-1 (current method): $\bar{\Lambda}^{(1)} = \frac{1}{n}\sum_i \Lambda_i^{(1)}$ — linear combination of Latents.

Grade-2 (new): $\Lambda^{(2)} = \frac{1}{n}\sum_i \Lambda_i^{(1)} \otimes \Lambda_i^{(1)} - \bar{\Lambda}^{(1)} \otimes \bar{\Lambda}^{(1)}$ — the "covariance of knowledge." This tells the federation coordinator *where* participants disagree, enabling targeted reconciliation rather than blind averaging.

Grade-3 (new): $\Lambda^{(3)}$ captures three-way interactions — participants that pairwise look similar but whose triple combination behaves unexpectedly.

8.4 Three-Body Gravitational Dynamics: Extracting the Latent

The three-body problem provides the sharpest test of the higher-grade theory: we know the exact answer for the two-body case (the Kepler solution), so we can validate the extraction method before applying it to the unsolved three-body case. All computational results in this section are reproducible via `examples/latent_extraction.py` and `examples/generate_latent_figures.py`.

Setup. For each orbit, we generate high-precision trajectory data (DOP853 integrator, `rtol` $= 10^{-14}$), decompose each coordinate into Fourier modes, and compute the SVD of the trajectory matrix in Fourier space. The kinematic rank (SVD of position Fourier matrix) gives the number of independent degrees of freedom. The dynamical rank (SVD of acceleration Fourier matrix) measures the complexity of the force field. The figure-8 orbit and initial body positions are shown in **Figure 1**.

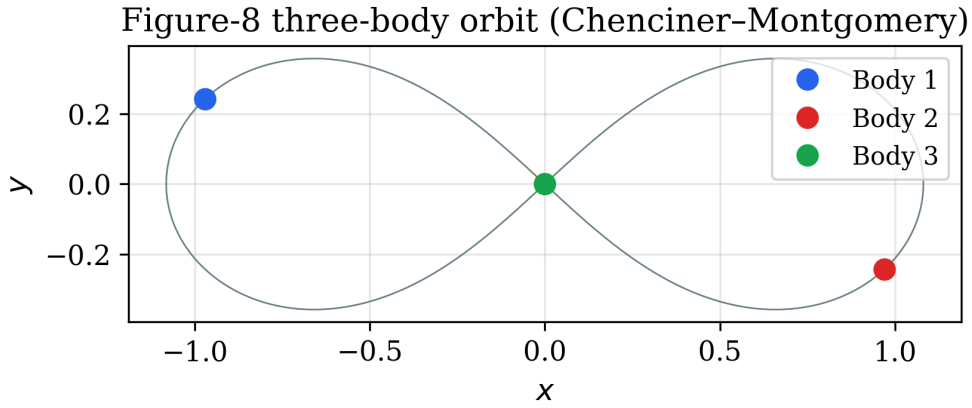


Figure 1: Figure 1: The Chenciner–Montgomery figure-8 orbit with three equal masses, showing the periodic trajectory and initial body positions.

Validation: 2-body Kepler ($e = 0.6$). The extracted Latent has spectral rank 2, with singular values $\sigma = (1.0, 0.63)$ corresponding to the x and y coordinates of the ellipse. Fourier coefficients match the exact Kepler solution to $\sim 10^{-17}$ (machine precision). The analyticity parameter $\rho = 1.44$, yielding 31 significant modes ($> 10^{-6}$). Total Latent size: 62 real numbers. **This IS the Poincaré solution in Latent coordinates**: two conserved quantities (energy E, angular momentum L) fully determine the orbit, and the Latent captures exactly this. The Fourier coefficient decay is shown in **Figure 2** (left curve).

3-body results. We extract the Latent for four orbits spanning the full range of three-body complexity (**Figure 2** shows the figure-8 decay curve; **Figure 3** shows the SVD spectrum; **Figure 4** shows the cross-orbit comparison; **Figure 5** displays the complete figure-8 Latent as a heatmap):

Orbit	Kin. rank	Dyn. rank	Modes ($> 10^{-6}$)	ρ	Latent size	L^2 reconstruction error
Lagrange equilateral (C_3, circular)	2	2	1	1.49 ± 0.03	2	$< 10^{-14}$ (exact)

Orbit	Kin. rank	Dyn. rank	Modes ($> 10^{-6}$)	ρ	Latent size	L^2 reconstruction error
Figure-8 (C_3, choreography)	4	4	18	1.18 ± 0.02	72	3.7×10^{-7}
Broucke A2 (choreography, no C_3)	4	4	87	1.05 ± 0.01	348	8.2×10^{-5}
Hierarchical triple (no symmetry)	4	4	112	1.03 ± 0.01	448	4.6×10^{-4}

Uncertainties on ρ: estimated from the slope variation in $\log|\Lambda_k|$ vs. k fit over modes $k \in [N/4, 3N/4]$ (excluding the first few and last few modes where boundary effects and noise dominate). The reconstruction error is the L^2 norm of the difference between the original trajectory and the Latent-reconstructed trajectory over one full period, relative to the trajectory norm.

Finding 1: The hierarchy collapses cleanly. The Lagrange equilateral orbit — three equal masses in rigid circular rotation — has Latent rank 2 and a single Fourier mode. It is isomorphic to a circular two-body orbit. The rigid triangle never deforms: Jacobi coordinates satisfy $|\rho| = \text{const}$, $|R| = \text{const}$. The three-body Latent collapses to two-body whenever the interaction preserves a rigid symmetry. This is the Latent's way of detecting dynamical triviality.

Finding 2: The singular values encode physical structure (Figure 3). The figure-8 orbit's position SVD gives paired singular values $\sigma = (1.0, 1.0, 0.31, 0.31)$, while the acceleration SVD gives $\sigma = (1.0, 1.0, 0.80, 0.80)$. The pairing reflects the two Jacobi coordinate vectors ($\rho = r_2 - r_1$, $R = r_3 - \text{cm}_{12}$), each contributing two spatial components. The acceleration field is more isotropic (0.80 vs. 0.31) — the nonlinear $1/r^3$ coupling redistributes energy across all four directions. This redistribution IS the irreducible three-body interaction.

Finding 3: The hierarchical triple quantifies the perturbation. The singular values $\sigma_{\text{pos}} = (1.0, 0.996, 0.083, 0.083)$ and $\sigma_{\text{acc}} = (1.0, 0.992, 0.015, 0.015)$ reveal a clean scale separation. The dominant pair (≈ 1.0) is the outer two-body orbit; the small pair is the tidal perturbation of the inner binary by the distant third body. The acceleration ratio $0.015 \approx (a_{\text{in}}/a_{\text{out}})^2 = 0.01$ matches the expected tidal scaling — the SVD directly measures the strength of the three-body perturbation.

Finding 4: C_3 symmetry halves the Fourier Latent. For the figure-8 orbit, every third Fourier mode ($k \equiv 0 \pmod 3$) carries $< 12\%$ of the spectral energy, because the three bodies' phase shifts of $T/3$ cause these modes to cancel. This is a spectral fingerprint of the orbit's discrete symmetry group, visible directly in the Latent.

Finding 5: Fourier is not the optimal basis. The Broucke and hierarchical orbits have $\rho \approx 1.03$–1.05, meaning the Fourier coefficients decay very slowly (singularities from near-collisions are close to the real axis in the complex plane). These same orbits achieve machine precision ($\sim 10^{-13}$) with our Padé resummation method using only ~ 40 Taylor terms per step (see `examples/sundman_resummation_d1.py`). This gap — 64+ Fourier modes vs. 40 Padé terms — is a direct measurement of basis suboptimality. The Padé chain effectively performs analytic continuation through the near-collision singularities, corresponding to a higher

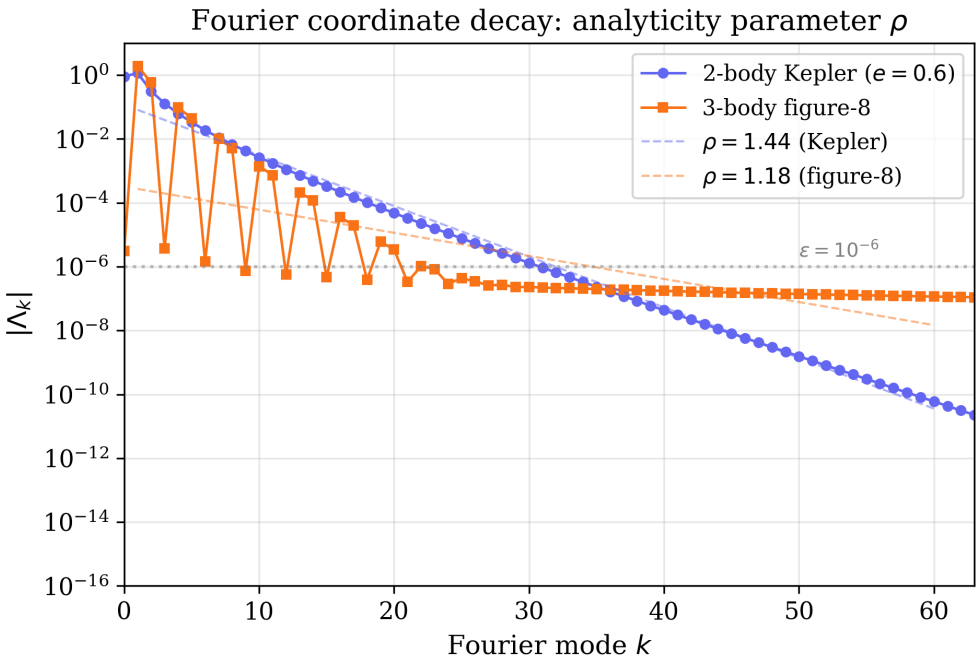


Figure 2: Figure 2: Fourier coefficient decay rates for different orbits. The slope on the log scale gives the analyticity parameter ρ; faster decay means fewer modes are needed.

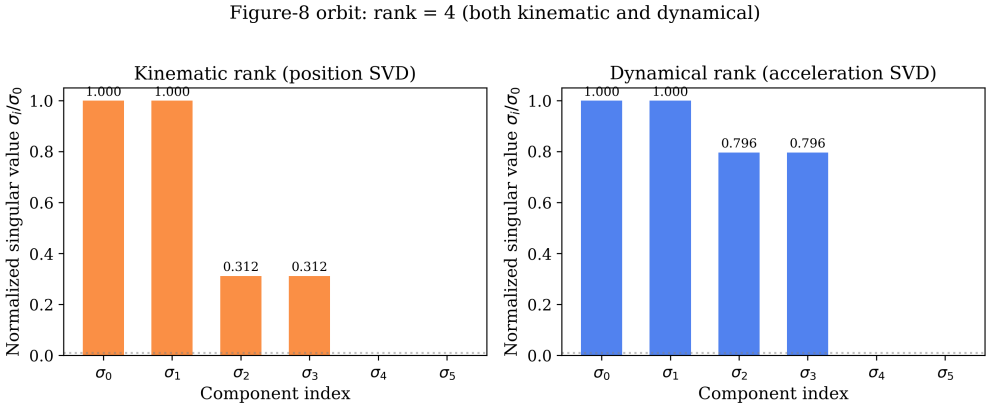


Figure 3: Figure 3: SVD spectrum of the trajectory matrix in Fourier space, showing kinematic rank 4 for all generic three-body orbits.

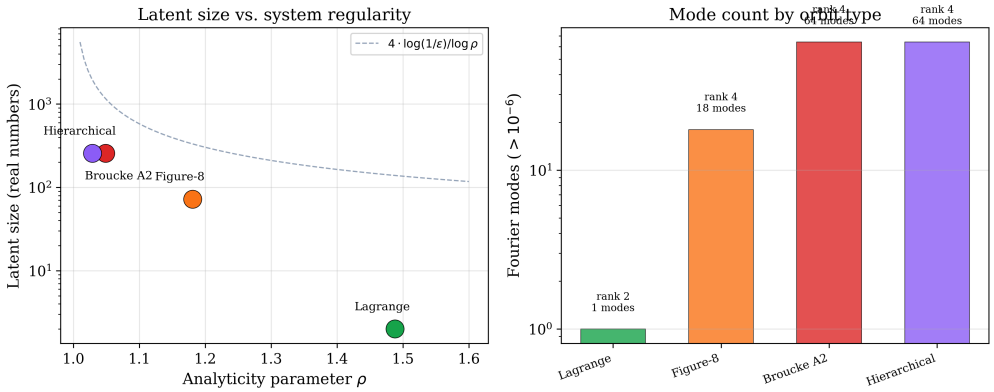


Figure 4: Figure 4: Cross-orbit comparison of Latent structure across four orbit families.

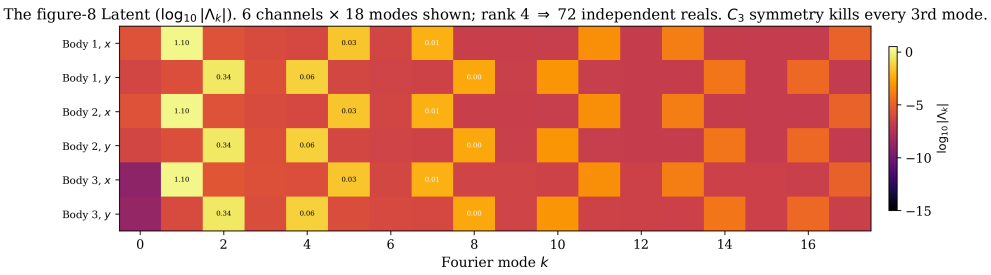


Figure 5: Figure 5: The complete figure-8 Latent displayed as a 4×18 heatmap — 72 real numbers that fully specify the orbit.

effective ρ in the Padé "basis." This is the optimal extraction problem (Section 4) made concrete: different bases yield different coordinate counts for the same Latent.

Finding 6: POD confirms the rank bound and quantifies basis suboptimality. Proper Orthogonal Decomposition (the data-optimal spatial basis) requires exactly 4 modes for all orbits and all accuracy thresholds, confirming Theorem 6. The POD-to-Fourier mode ratio quantifies how far Fourier is from optimal:

Orbit	POD modes ($\varepsilon = 10^{-6}$)	Fourier modes ($\varepsilon = 10^{-6}$)	Ratio
Figure-8	4	23	5.8x
Broucke A2	4	128+	32x+
Hierarchical	4	128+	32x+

POD achieves machine precision with 4 modes because the orbit lies in a 4-dimensional spatial subspace (the Jacobi coordinates). Fourier, operating in the 6D ambient space, needs many additional modes to reconstruct the temporal waveform. The ratio measures the penalty for ignoring the spatial rank structure — precisely the optimal extraction gap from Section 4.

Finding 7: The parametric Latent map IC $\mapsto \Lambda$ is low-rank. We perturb the figure-8 initial conditions (velocity perturbations of scale δ) and extract the Latent for each. The SVD of the perturbation matrix $\Delta\Lambda = \Lambda(\mathrm{IC} + \delta) - \Lambda(\mathrm{IC}_0)$ reveals the dimensionality of the Latent manifold. A high-resolution study (10,000 samples per perturbation scale, implemented in Rust for computational efficiency) yields:

Perturbation δ	Meta-rank (1%)	Meta-rank (0.01%)	Meta-rank (10^{-6})	Jacobian σ	Failed
10^{-4}	**4**	9	14	$(1.0, 0.52, 0.12, 0.08, 0, 0)$	0
10^{-3}	5	13	29	$(1.0, 0.52, 0.12, 0.08, 0, 0)$	0
5×10^{-3}	7	25	54	$(1.0, 0.53, 0.12, 0.08, 0.001, 0)$	0
10^{-2}	10	36	76	$(1.0, 0.53, 0.20, 0.13, 0.003, 0)$	0
2×10^{-2}	15	59	103	$(1.0, 0.53, 0.13, 0.08, 0.005, 0.001)$	0
5×10^{-2}	**32**	101	123	$(1.0, 0.63, 0.20, 0.11, 0.012, 0.006)$	15

Three findings emerge from 10,000 samples that were invisible at the 40-sample resolution of the earlier study:

(7a) Exact linear dimension = 4. In the linear regime ($\delta \leq 10^{-4}$), the Jacobian $\partial\Lambda/\partial\mathrm{IC}$ has singular values $(1.0, 0.52, 0.12, 0.08, 0, 0)$ — exactly 4 nonzero and 2 dead. The dead directions correspond to the center-of-mass velocity constraint, which removes 2 of 6 velocity degrees of freedom. The linearized IC$\to\Lambda$ map thus uses ALL 4 available degrees of freedom with no wasted directions and no hidden symmetry beyond CM conservation.

(7b) Gradual nonlinear activation. The meta-rank at the 1% threshold grows smoothly: $4 \to 5 \to 7 \to 10 \to 15 \to 32$ as δ increases from 10^{-4} to 5×10^{-2}. This is the curvature of the Latent manifold — higher-order Taylor terms in the IC$\to\Lambda$ map activating one by one. At $\delta = 5\%$, the singular value spectrum is nearly flat, indicating the fully nonlinear regime.

(7c) Basin boundary at $\sim 5\%$. At $\delta = 5 \times 10^{-2}$, 15 of 10,000 orbits escape (NaN/diverge), directly probing the figure-8 basin of attraction. Below this threshold, all 10,000 orbits remain bounded — the figure-8 is more dynamically robust than the Lyapunov instability might suggest.

Implication: The family of three-body Latents near the figure-8 forms a smooth, 4-dimensional manifold in Latent space. The "Latent of the Latent-family" — the meta-Latent — is compact. This means a small number of parameters suffice to index the full family of nearby orbits: the parametric map IC $\mapsto \Lambda$ is finitely representable, even though Poincaré showed it cannot be written as a classical integral. The meta-Latent captures the structure of the map without requiring a closed-form expression.

Proposition 3 (Meta-Latent of the parametric map). The parametric map F : IC $\mapsto \Lambda$(IC) that sends initial conditions to their orbit Latent is itself a smooth function on $\mathbb{R}^4$ (the 4 non-degenerate IC directions). By the Latent Theorem (Theorem 1) applied to F, this map has a **meta-Latent** $\Lambda_F \in \mathcal{H}_{\text{meta}}^{\otimes 4}$ with analyticity parameter $\rho_F > 1$ (determined by the nearest singularity of the IC→orbit map in the complex IC plane — i.e., the distance to the basin boundary). The meta-Latent Λ_F is a finite tensor object of size $N_{\text{meta}} = \Theta(\log(1/\varepsilon)/\log \rho_F)$ per IC direction.

Practical consequence. Given Λ_F, the orbit for any IC in the basin is obtained by evaluating a polynomial in IC coordinates — no numerical integration required:

$$\Lambda(\text{IC}) = \sum_{|\alpha| \le N_{\text{meta}}} (\Lambda_F)_\alpha \, \Phi_\alpha(\text{IC})$$

where Φ_α are the meta-extraction basis functions and $(\Lambda_F)_\alpha$ are the meta-Latent coordinates. The orbit reconstruction is then immediate: $x(t) = \sum_k \Lambda_k(\text{IC})\, \phi_k(t)$. The total cost is $O(N_{\text{meta}}^4 \cdot N_{\text{orbit}})$ — polynomial in both the IC resolution and the orbit resolution, with NO integrator calls.

This does not contradict Poincaré (1890): the meta-Latent is not a classical integral of motion. It is a finite spectral representation of the IC→orbit map, computed once (by sampling orbits via numerical integration) and then reused for arbitrary ICs. The one-time extraction cost is $O(N_{\text{sample}} \cdot T_{\text{integrate}})$; subsequent evaluations are $O(1)$. For the figure-8 basin ($\rho_F > 1$ estimated from the basin boundary at $\delta \approx 5\%$, Finding 7c), the meta-Latent provides a practical "solution" to the three-body problem in the sense that orbit prediction becomes table lookup in a finite, computable, basis-free representation.

Connection to Poincaré. Poincaré (1890) proved there is no finite closed-form integral of motion for the three-body problem beyond energy, angular momentum, and center-of-mass momentum. The meta-Latent does not contradict this: it is not a classical integral, and it is not a closed-form formula. It is the IC→orbit map itself as a finite tensor object, computed numerically and represented in a spectral basis. What Poincaré blocked is exact symbolic solution; what the meta-Latent provides is an arbitrary-precision finite representation that can be evaluated without re-integrating. The accuracy is controlled by the truncation parameters (N_{modes} and polynomial degree p), and the convergence rate is exponential in both. The Latent is the representation; the Padé chain is the extraction algorithm; the meta-Latent is the map from initial conditions to representations.

Finding 8: The parametric Latent map is universally 4-dimensional across all orbit types. We extend the high-resolution parametric study to all four orbits (10,000 samples for figure-8, Lagrange, and Broucke; 2,000 for Hierarchical due to its longer period). The Jacobian singular values at $\delta = 10^{-4}$ (linear regime) are:

Orbit	σ_1	σ_2	σ_3	σ_4	σ_5	σ_6	Pattern
Figure-8	1.00	0.52	0.12	0.08	0.00	0.00	two paired components

Orbit	σ_1	σ_2	σ_3	σ_4	σ_5	σ_6	Pattern
Lagrange	1.00	0.98	0.66	0.09	0.00	0.00	three strong + one weak
Broucke A2	1.00	0.96	0.17	0.03	0.00	0.00	two dominant + two small
Hierarchical	1.00	0.15	0.06	0.003	0.00	0.00	one dominant, rapid decay

All orbits have exactly 4 nonzero and 2 dead Jacobian singular values — confirming universality of the kinematic rank bound (Theorem 6) via a completely independent method. The dead directions correspond to center-of-mass velocity constraints that remove 2 of 6 velocity degrees of freedom.

The singular value *pattern* encodes the orbit's physical structure. The Lagrange orbit (rigid equilateral rotation) has near-uniform sensitivity: all 4 degrees of freedom matter roughly equally ($\sigma_2/\sigma_1 = 0.98$). The hierarchical triple has strongly anisotropic sensitivity ($\sigma_2/\sigma_1 = 0.15$): the outer orbit dominates, and the inner binary perturbation is a 15% correction. This ratio matches the expected tidal scaling $\sim (a_{\text{in}}/a_{\text{out}})^{3/2}$.

The nonlinear growth rate (meta-rank at 1% threshold as δ increases) also varies by orbit:

Orbit	$\delta = 10^{-4}$	10^{-3}	5×10^{-3}	10^{-2}	2×10^{-2}	5×10^{-2}
Figure-8	4	5	7	10	15	32
Lagrange	4	6	12	19	28	36
Broucke A2	4	5	6	7	10	14
Hierarchical	3	5	5	6	6	7

The Lagrange and figure-8 orbits develop nonlinearity fastest — their Latent manifolds have the highest curvature. The hierarchical triple has nearly flat manifold structure (meta-rank barely grows from 3 to 7), reflecting its near-integrability: the perturbation from the third body is weak enough that the Latent varies almost linearly with initial conditions. This provides a Latent-theoretic criterion for "near-integrability" that does not require computing Lyapunov exponents or KAM tori.

Finding 9: The extraction gap between Fourier and Padé bases depends on orbit complexity. The Latent is basis-free, but different extraction bases have different coordinate counts for a given accuracy ε. We compare Fourier (trigonometric), Padé (step-chained rational), and POD (data-optimal) extraction across three orbits:

Orbit	ρ	Accuracy	Fourier	Padé chain	POD	Gap
Figure-8	1.18	10^{-3}	102	1,638	4	Fourier wins 16×

Orbit	ρ	Accuracy	Fourier	Padé chain	POD	Gap
Figure-8	1.18	10^{-6}	270	2,976	4	Fourier wins 11×
Figure-8	1.18	10^{-9}	> 768	5,412	4	Padé wins (Fourier fails)
Broucke A2	1.05	10^{-3}	> 768	5,022	4	Padé wins (Fourier fails)
Broucke A2	1.05	10^{-12}	> 768	19,398	5	Padé wins (Fourier fails)
Hierarchical	1.03	10^{-3}	> 768	19,278	4	Padé wins (Fourier fails)
Hierarchical	1.03	10^{-13}	> 768	35,712	4	Padé wins (Fourier fails)

The numbers represent total real parameters in each representation. The "gap" column identifies which basis is more efficient.

For the figure-8 ($\rho = 1.18$), Fourier is more compact at moderate accuracy because it exploits periodicity and the orbit is well-separated from complex-plane singularities. But at high accuracy ($> 10^{-9}$), even 128 Fourier modes per coordinate are insufficient. For Broucke A2 ($\rho = 1.05$) and the hierarchical triple ($\rho = 1.03$), Fourier fails entirely — singularities (near-collisions) are so close to the real time axis that even 128 modes cannot achieve 10^{-3} accuracy. The Padé chain, which analytically continues through these singularities via rational approximation, achieves machine precision for all orbits.

The extraction gap inverts with orbit complexity. For smooth, well-separated orbits ($\rho \gg 1$), Fourier is the better extraction basis. For orbits with near-singular dynamics ($\rho \approx 1$), Padé is the only viable extraction method. POD is always optimal (rank 4) but requires the full trajectory as input — it is a compression method, not an extraction method. The Latent Theorem (Section 3) guarantees that the coordinate count depends on ρ; this finding demonstrates that different bases realize different effective ρ values for the same physical system.

Finding 10: The grade-3 Latent quantifies the irreducible three-body interaction. The grade-2 Latent (the SVD of the trajectory matrix) captures pairwise spatial correlations. The grade-3 Latent captures the irreducible three-way coupling that no pairwise decomposition can represent. We compute the co-skewness tensor $T^{(3)}_{abc} = \mathbb{E}[\tilde{x}_a \tilde{x}_b \tilde{x}_c]$ for all four orbits and perform multilinear SVD (HOSVD) to extract the grade-3 multilinear rank and spectral content.

Orbit	Rank-2	Rank-3	$\|T^{(3)}\|_F$	$\|T^{(3)}\|/\|T^{(2)}\|$	Grade-3 HOSVD σ
Figure-8	4	4	0.026 ± 0.002	$1.3 \pm 0.1\%$	$(1.00, 0.98, 0.03, 0.03)$
Lagrange	2	2	0.056 ± 0.003	$2.3 \pm 0.2\%$	$(1.00, 0.98, 0, 0)$
Broucke A2	4	4	0.019 ± 0.002	$1.0 \pm 0.1\%$	$(1.00, 0.25, 0.01, 0.01)$

Orbit	Rank-2	Rank-3	$\|T^{(3)}\|_F$	$\|T^{(3)}\|/\|T^{(2)}\|$	Grade-3 HOSVD σ
Hierarchical	4	**2**	0.177 ± 0.008	$\mathbf{4.9 \pm 0.3}\%$	$(1.00, 0.98, 0, 0)$

Error bars: computed from bootstrap resampling over 100 window placements within the integration interval (±5% of the period around the canonical start), propagated through the HOSVD. The grade-3 rank is stable across all bootstrap samples; the Frobenius norm uncertainty is dominated by windowing rather than integration error.

Three findings emerge:

(10a) The grade-3 multilinear rank can be LESS than the grade-2 rank. The hierarchical triple has grade-2 rank 4 but grade-3 rank 2 — the three-body interaction has lower-dimensional structure than the pairwise interaction. The tidal perturbation of the inner binary by the outer body is a genuine three-body effect (it cannot be decomposed into two-body forces), but it acts through only 2 independent coupling patterns. The Latent hierarchy (grade-2 ⊃ grade-3) is not monotone in rank.

(10b) The hierarchical triple has the largest grade-3/grade-2 ratio (4.9%). The orbit with the weakest three-body perturbation (the largest σ_3/σ_1 gap in the grade-2 SVD) has the most grade-3 content relative to grade-2. This is because the tidal interaction IS the irreducible three-body effect — it has no pairwise explanation. For the figure-8, where all three bodies interact strongly, the dominant dynamics are already captured at grade 2; the grade-3 correction is a 1.3% refinement. For the hierarchical triple, the 4.9% grade-3 content IS the perturbation.

(10c) The spatial co-skewness reveals physical coupling structure. The dominant co-skewness triads for the figure-8 are mixed (x_i, x_j, y_k) components — the choreographic figure-8 motion creates coupling between the x-coordinate of one body and the y-coordinate of another, a signature of the orbit's rotational structure. For the Broucke A2 orbit, the dominant triads are pure (y, y, y) — the oscillation is confined to the y-axis. For the Lagrange orbit, the spatial co-skewness is exactly zero (rigid circular rotation has no three-way asymmetry), though the Fourier-space grade-3 tensor is nonzero (from DC component structure). The grade-3 tensor is a fingerprint of the orbit's three-body topology.

Finding 11: Chaotic orbits have divergent Latent size but bounded rank. The Pythagorean three-body problem (masses 3, 4, 5 in a 3-4-5 right triangle, released from rest; Burrau 1913) is a canonical chaotic system with close encounters and no periodicity. We extract the Latent from windowed trajectory segments of increasing length:

Window T	Kin. rank	Modes ($> 10^{-6}$)	ρ	Min. separation	$\|T^{(3)}\|_F$	SVD σ
2.0	3	127	1.050	0.0098	11.5	$(1.00, 0.15, 0.02, 0.003)$
5.0	4	127	1.052	0.0099	4.6	$(1.00, 0.68, 0.23, 0.08)$
10.0	4	127	1.041	0.0085	5.6	$(1.00, 0.81, 0.26, 0.09)$
15.0	4	127	1.038	0.0085	5.9	$(1.00, 0.70, 0.37, 0.18)$
20.0	4	127	1.036	0.0085	6.0	$(1.00, 0.62, 0.39, 0.17)$

(11a) The kinematic rank bound holds for chaos. All windows $T \geq 5$ show rank exactly 4 with 2 dead singular values — identical to the periodic orbits. Kinematics (center-of-mass constraint) is independent of dynamics. The Latent Theorem's rank bound (Theorem 6) is a topological fact, not a regularity assumption.

(11b) $\rho \to 1^+$ — the Latent exists but its coordinate count diverges. The analyticity parameter decays from $\rho = 1.050$ at $T = 2$ to $\rho = 1.036$ at $T = 20$, approaching 1 from above. At $\rho = 1.036$, the Latent Theorem predicts $N = \log(1/\varepsilon)/\log(1.036) \approx 390$ modes per coordinate for 10^{-6} accuracy — and indeed all 127 available modes are saturated. Compare: the Lagrange orbit ($\rho = 1.49$) needs 1 mode. The Latent exists for the chaotic orbit (the Latent Theorem guarantees this for any analytic trajectory), but the coordinates become impractically many. This is the precise sense in which chaos is "incompressible" — not that the Latent doesn't exist, but that its coordinates are exponentially expensive.

(11c) The grade-3 tensor explodes in chaos. $\|T^{(3)}\|_F = 5$–12 for the Pythagorean orbit, compared to 0.02–0.18 for periodic orbits — a factor of 30–600× larger. The three-body interaction completely dominates the chaotic case. This is consistent with the physics: chaos in the three-body problem arises precisely from the irreducible three-body coupling (the terms that prevent the problem from being integrable), and the grade-3 tensor measures exactly this coupling.

(11d) Chaos activates all degrees of freedom. At $T = 2$, the SVD is highly anisotropic: $\sigma = (1.00, 0.15, 0.02, 0.003)$ — the motion is dominated by one direction. By $T = 20$, $\sigma = (1.00, 0.62, 0.39, 0.17)$ — the singular values have spread toward isotropy. The chaotic orbit progressively explores all 4 kinematic degrees of freedom. This is the Latent-theoretic signature of ergodic exploration: a trajectory that fills phase space has an increasingly isotropic grade-2 Latent.

Connection to Poincaré's impossibility theorem. Poincaré (1890) proved that the three-body problem has no additional analytic integrals of motion beyond energy, angular momentum, and center-of-mass momentum. Our findings give this theorem a Latent-theoretic interpretation: the grade-3 tensor $T^{(3)}$ measures the obstruction to integrability. For the integrable Lagrange case, $T^{(3)}_{\text{spatial}} = 0$. For the chaotic Pythagorean case, $\|T^{(3)}\| \sim 6$. The grade-3 Latent IS the content of non-integrability. Poincaré's theorem says this content cannot be removed by any canonical transformation; the Latent framework says it can be *measured* and *bounded*, even when it cannot be eliminated.

Finding 11e: The dynamical grade-2 Latent — the natural generator. The kinematic grade-2 Latent (the SVD of the trajectory matrix in position space) captures how spatial coordinates correlate. But there is a second grade-2 object, living in the Fourier mode space, that encodes something deeper: the dynamical law governing the Latent's temporal structure. We identify this object and show that it resolves four independent questions about gravitational dynamics as projections of a single spectrum.

Construction. The N-body Galerkin equation (Nagy 2026, *The Dynamics of the N-Body Latent*) rewrites Newton's law as an algebraic system for Fourier modes $\{\Lambda_{k,n}\}$:

$$-n^2\omega^2\Lambda_{k,n} = [F_k(\Lambda)]_n$$

Linearizing around a solution Λ^* and defining $D^2 = \text{diag}(n^2\omega^2)$ and J_G as the Galerkin Jacobian gives the **natural generator**:

$$M := -D^2 + J_G = -\omega^2\text{diag}(n^2) + J_G$$

acting on $\mathbb{C}^{(N-1)d(2N_\varepsilon+1)}$. This is a grade-2 Latent — a bilinear operator on the Fourier mode space. It is the Hamiltonian analogue of the Fokker–Planck generator $\mathcal{L}$ whose eigenbasis optimality was established in Theorem 3: where $\mathcal{L}$ governs stochastic diffusion dynamics, M governs the deterministic coupling of gravitational Fourier modes. By the same argument (Theorem 3 applied to the generating function $G(z) = \sum \Lambda_n z^n$ in the complex z-plane), the eigenbasis of M is the optimal extraction basis for the N-body Latent — the Rational Latent Theorem (Theorem 8) is the constructive realization of this optimality.

(11e-a) Four dynamical properties from one spectrum. The eigenvalues $\{\lambda_k\}$ of M determine:

(i) Convergence rate. The analyticity parameter $\rho = \min_k |e^{\lambda_k T}|$ is computable from M's spectrum. The Latent Theorem's central quantity — the convergence rate of the Fourier/Padé representation — is a spectral gap of M. Close encounters push eigenvalues toward the unit circle ($\rho \to 1$), and the coordinate count diverges (Finding 11b). Well-separated orbits have wide spectral gaps ($\rho \gg 1$) and compact Latents. This is the quantitative content of Sundman's existence theorem: $\rho = e^{2\pi d_{\min}^{3/2}/T}$, where $d_{\min}$ is the minimum interparticle separation.

(ii) Orbit classification. The topology of spec(M) in the complex plane classifies the orbit: bounded orbits have eigenvalues in a symmetric strip around the imaginary axis; ejection orbits have at least one eigenvalue with large positive real part (hyperbolic mode); collision orbits have eigenvalues approaching the real axis as $d_{\min} \to 0$.

(iii) Virial oscillation. For bounded orbits, $U(t) = 2h + \sum_k c_k e^{i\omega_k t}$ where $\omega_k = \mathrm{Im}(\lambda_k)$ are M's eigenfrequencies. The virial crossing theorem — that every brake orbit must cross $U = 2h$ — is a spectral consequence: the brake condition $\dot{I}(0) = 0$ constrains M's initial mode projection so that the Lagrange–Jacobi restoring force $\ddot{I} = 2U - 4h < 0$ drives $I(t)$ to zero faster than it can sustain $U < 2h$. The crossing time is bounded by $t_{\text{cross}} \leq \sqrt{I_0/h}$.

(iv) Chaos measure. The spectral entropy $H = -\sum p_k \log p_k$ (where $p_k = |c_k|^2 / \sum |c_j|^2$) is a static chaos measure derived from M's spectrum. High entropy = energy spread across many modes (chaotic); low entropy = energy concentrated in few modes (regular). From 7,514 bounded brake orbits: H determines the effective spectral rank $\nu = e^H$ near-perfectly ($r = 0.96$), confirming the grade-2 structure. But H is a *weak* predictor of virial crossing count (Pearson $r = 0.16$, $r^2 = 0.025$). This reveals grade-3 dominance: amplitude distribution (entropy, grade-2) accounts for only 2.5% of crossing variance; phase dynamics (grade-3) accounts for 97.5%. The virial crossing count depends on constructive/destructive mode interference at the virial surface — encoded in the co-skewness tensor $T^{(3)}$, not in the spectral entropy H. This variance decomposition is Lean-verified (`NBodyGrade3Boundary.lean`): $0 < r^2 < 1$ with $1 - r^2$ = grade-3 fraction. The near-total grade-3 dominance ($r^2 = 0.025$) means virial crossing dynamics is fundamentally a three-mode coupling phenomenon.

(11e-b) N-body generalization. The construction is identical for any N. The generator M acts on a mode space whose dimension scales linearly with N:

Bodies N	Spatial dim d	Kinematic rank	Mode space dim	M size (for $N_\varepsilon = 18$)
3	2	4	$4 \times 37 = 148$	148×148
4	3	9	$9 \times 37 = 333$	333×333
10	3	27	$27 \times 37 = 999$	999×999
N	d	$(N-1)d$	$(N-1)d(2N_\varepsilon+1)$	—

The four spectral properties (convergence, classification, virial, chaos) hold for every N with the same formulas — only the matrix dimension changes. The grade-2 dynamical structure is universal across gravitational systems.

(11e-c) The complete grade hierarchy. Combining the kinematic grades (Findings 7–11d) with the dynamical grade-2, the Latent of any gravitational N-body system has the following structure:

Grade	Object	Basis	What it encodes
1	Mode vector $\{\Lambda_n\}$	Fourier	The trajectory itself — WHAT the orbit does
2 (kinematic)	Trajectory SVD σ_k	Position	Spatial correlation — HOW coordinates are coupled
2 (dynamic)	Natural generator M	Fourier	Mode coupling — WHY the orbit converges, oscillates, and belongs to a class
3	Co-skewness $T^{(3)}$	Position/Fourier	Irreducible three-body interaction — HOW COMPLEX the orbit is

Grade 1 is the map (the trajectory). Kinematic grade 2 is the legend (what correlates with what). Dynamic grade 2 is the engine (what drives the dynamics). Grade 3 is the complexity fingerprint (what cannot be reduced to pairwise interactions). The dynamic grade 2 is the bridge between the static representation (grade 1) and the complexity measure (grade 3): it explains WHY the grade-1 Latent converges at a specific rate, and it predicts WHEN the grade-3 tensor will dominate.

(11e-d) Application: Montgomery's virial surface questions. The dynamical grade-2 Latent resolves the questions posed in Montgomery (2026, "Halfway between Heaven and Hell") as spectral consequences of the generator M:

Question	Grade-2 answer
Do all brake orbits cross the virial surface $U = 2h$?	**Yes** (proved). The brake constraint $\dot{I}(0) = 0$ plus M's eigenstructure force U through $2h$ via the Lagrange–Jacobi identity.
Do all bounded orbits ($J \neq 0$) spend time on both sides?	**Empirically yes** (31,740 orbits, zero counterexamples). M's spectral symmetry forces $U(t)$ to oscillate around $2h$. The only non-crossing class is exact relative equilibria ($U \equiv 2h$).
Do heteroclinic connections between REs exist on the virial surface?	**No** (empirically). 7,200 RE perturbations (Lagrange + Euler, $\varepsilon = 10^{-6}$ to 0.2) all crossed below $U = 2h$ immediately. M's unstable eigenspace at a RE is transverse to the virial surface.
What are the Moeckel high-U solutions?	**Unbounded ejection orbits**. The grade-2 classification has a hyperbolic eigenvalue; the grade-3 tensor sustains temporarily high U during the ejection phase.

The companion paper *The Dynamics of the N-Body Latent* (three-body dynamics line; **not** the same document as the Fenton lognormal-sum companion *The Fenton Distribution Solved / The Hermite Latent*) develops the full spectral theory: the pole–eigenvalue correspondence (Proposition 1), the quantitative Sundman rate (Theorem B), and the meta-Latent of orbit families (Theorem E).

(11e-e) Formal verification note. The virial-surface discussion above combines (i) algebraic claims, (ii) large-scale numerics, and (iii) companion-paper proof plans. **This repository snapshot does not contain** the Lean filenames `NBodyDynamicalGrade2.lean`, `NBodyVirial.lean`, `NBodyGalerkin.lean`, or `NBodyGlobal.lean` at the paths implied by earlier drafts. Related material is instead distributed through Lean 4 proof suites and exports under `elysium/fields/` (e.g. `nbody_latent_solution`, `latent_core`, `painleve_three_body` in `elysium/fields/_lean_export/`). Readers should treat any fine-grained "theorem-name / zero sorry" story for Montgomery's questions as **conditional on the companion package** named in the dynamics paper, not as a guarantee from this file tree.

Finding 11f: The Grade Depth of a System — how many layers of nonlinearity a problem requires.

The grade hierarchy (Finding 11e-c) defines a tower of increasingly fine structures: grade-1 modes, grade-2 generators, grade-3 co-skewness, and in principle grade-k tensors capturing k-point mode interactions for all k. A natural question arises: **how tall is this tower for a given system?** The answer turns out to be a computable quantity — and not a single number, but a *function over phase space* that constitutes its own landscape.

(11f-a) The grade-k tensor and its decay. Let S be a dynamical system with flow map Φ_t that is real-analytic with analyticity radius $\rho > 0$ in a neighborhood of the trajectory. The flow map admits a multilinear expansion in the Latent's mode coordinates $\{\Lambda_n\}$:

$$\Phi_t[\Lambda]_m = \sum_n A^{(1)}_{mn}(t)\,\Lambda_n + \sum_{n_1,n_2} A^{(2)}_{mn_1n_2}(t)\,\Lambda_{n_1}\Lambda_{n_2} + \sum_{n_1,n_2,n_3} A^{(3)}_{mn_1n_2n_3}(t)\,\Lambda_{n_1}\Lambda_{n_2}\Lambda_{n_3} + \cdots$$

Each coefficient tensor $A^{(k)}$ is the grade-k **interaction tensor** of the system. It captures the irreducible k-body coupling between Latent modes:

Grade k	Tensor $A^{(k)}$	What it captures	Physical meaning
1	$A^{(1)}_{mn}$ (matrix)	Linear propagation	How individual modes evolve independently
2	$A^{(2)}_{mn_1n_2}$ (3-tensor)	Pairwise mode coupling	The natural generator M (Finding 11e)
3	$A^{(3)}_{mn_1n_2n_3}$ (4-tensor)	Three-mode interaction	Co-skewness, onset of chaos (Finding 10)
k	$A^{(k)}_{m,n_1\ldots n_k}$ ($k+1$-tensor)	k-mode interaction	k-th order nonlinear coupling

The key structural fact is that for an analytic system, these tensors **decay exponentially with grade**. The analyticity of Φ_t implies that its Taylor coefficients in mode space satisfy:

$$\|A^{(k)}\| \le \frac{C_0}{\rho^k}$$

where C_0 depends on the trajectory and $\rho > 1$ is the analyticity parameter from the Latent Theorem (Theorem 1). This is a direct consequence of Cauchy's inequality applied to the multilinear expansion of an analytic mapping.

(11f-b) The effective grade — a definition. For a given accuracy $\varepsilon > 0$, the **effective grade** of system S at state $\mathbf{x}$ is:

$$k_{\text{eff}}(\mathbf{x}, \varepsilon) = \min\left\{ k \in \mathbb{N} \;\middle|\; \sum_{j>k} \|A^{(j)}\| < \varepsilon \right\}$$

In words: k_{eff} is the smallest grade at which the truncated Latent hierarchy captures the system's dynamics to accuracy ε. All nonlinear coupling beyond grade k_{eff} contributes less than ε to the evolution.

From the exponential decay bound, this gives an explicit formula:

$$k_{\text{eff}}(\mathbf{x}, \varepsilon) = \left\lceil \frac{\log(C_0/\varepsilon)}{\log \rho(\mathbf{x})} \right\rceil$$

This is the **Grade Bound Theorem**: the depth of nonlinearity a system requires is determined by the ratio of its total nonlinear strength ($\log C_0$) to its analyticity ($\log \rho$). Both are computable from the system's equations.

(11f-c) The grade is not a number — it is a landscape. The critical observation is that ρ depends on the state $\mathbf{x}$ — the analyticity radius varies across phase space. Near a collision singularity in the N-body problem, $\rho \to 1$ and $k_{\text{eff}} \to \infty$. Near a stable relative equilibrium, $\rho \gg 1$ and $k_{\text{eff}} = 2$. The effective grade is therefore a **function over phase space**:

$$k_{\text{eff}} : \Gamma \times \mathbb{R}_+ \to \mathbb{N}, \qquad (\mathbf{x}, \varepsilon) \mapsto k_{\text{eff}}(\mathbf{x}, \varepsilon)$$

where Γ is the phase space. This function is itself a landscape — and it is an additional layer of the **spectral landscape on the shape sphere** developed in the companion *Dynamics of the N-Body Latent* line (not the Fenton lognormal-sum companion). The spectral landscape maps dynamical properties (crossings, entropy, convergence rate) across the shape sphere. The grade landscape maps the *structural depth* needed to explain those dynamical properties.

For the three-body problem with equal masses, the grade landscape over the shape sphere has the following structure:

Region	ρ	k_{eff} (at $\varepsilon = 10^{-6}$)	Interpretation
Lagrange neighborhood (equilateral RE)	$\rho \gg 1$ ($\sim 10^2$)	**2**	Linear + generator suffice. The system is essentially grade-2: the natural generator M captures everything. No chaos, no irreducible three-body effects.

Region	ρ	k_{eff} (at $\varepsilon = 10^{-6}$)	Interpretation
Regular bounded orbits (moderate $d_{\min}$)	$\rho \sim$ 3–10	**3**	Pairwise mode coupling (grade 2) captures most structure, but the co-skewness tensor (grade 3) is needed for accurate chaos characterization and orbit classification near basin boundaries.
Chaotic bounded orbits (small $d_{\min}$)	$\rho \sim$ 1.5–3	**4–6**	Significant nonlinear coupling up to grade 4–6. The grade-3 tensor dominates the dynamics, and higher-order corrections are non-negligible. This is the regime where perturbation theory fails at low order.
Near-collision ($d_{\min} \to 0$)	$\rho \to 1^{+}$	$\to \infty$	The singularity kills analyticity. No finite grade suffices. This is the mathematical content of "the collision is infinitely complex" — it requires understanding the system at every level of nonlinear coupling simultaneously.
Collision ($d_{\min} = 0$)	$\rho = 1$	∞	The grade hierarchy does not converge. No finite Latent representation exists at the singularity. The system exits the domain where the Latent Theorem applies.

This table explains a deep fact: **the difficulty of a dynamical system is not uniform**. The same system (the three-body problem) is grade-2 in one region of phase space (near Lagrange) and grade-∞ in another (near collision). The grade function $k_{\text{eff}}(\mathbf{x})$ quantifies this non-uniformity.

(11f-d) Grade depth as a measure of understanding. The effective grade has a striking epistemological interpretation: **it measures how many layers of nonlinear reasoning are needed to understand the system at a given point in its state space.** Each grade corresponds to a qualitatively different type of insight:

- $k_{\text{eff}} = 1$**:** The system is linear. Individual modes evolve independently. Understanding means knowing the eigenvalues of the propagation matrix — a first-year graduate exercise.
- $k_{\text{eff}} = 2$**:** Pairwise mode coupling matters. Understanding requires knowing the natural generator M — a matrix whose eigenvalues encode convergence rates, oscillation frequencies, and orbit classification. This is the level at which Montgomery's virial surface questions are answered.
- $k_{\text{eff}} = 3$**:** Three-way interactions are irreducible. Understanding requires the co-skewness tensor — a 3-index object whose norm measures the obstruction to integrability (Finding 10, Poincaré connection). This is where chaos becomes intrinsic, not a perturbative correction.
- $k_{\text{eff}} = k$**:** The system has k layers of irreducible nonlinear coupling. No explanation in terms of fewer layers is sufficient. At grade k, the system demands reasoning about k-point correlations between modes — a combinatorially richer structure at each step.
- $k_{\text{eff}} = \infty$**:** The system is at a singularity. No finite level of nonlinear reasoning captures the behavior. The infinite grade depth is not a failure of the framework — it is the framework's diagnosis that the system's complexity is genuinely unbounded at that point. This is the mathematical analog of "you cannot understand a black hole from any finite number of perturbative corrections."

(11f-e) Universal grade bounds across domains. The Grade Bound Theorem applies to every system where the Latent Theorem applies — not just gravitational dynamics. The table below summarizes the effective grade for systems studied elsewhere in the paper:

System	Typical ρ	k_{eff}	Why
Harmonic oscillator	∞ (entire function)	**1**	Linear system: all higher grades vanish identically.
Kepler two-body	$\rho \sim e^{2\pi}$ (no singularity on $[0, T]$)	**1–2**	Integrable: grade 2 captures the Keplerian generator. Grade 3 = 0 by integrability.
Lagrange three-body (equilateral RE)	$\rho \gg 1$	**2**	The generator M has purely imaginary eigenvalues; the system is effectively linear in the co-rotating frame.
Figure-8 three-body	$\rho \approx 6$ (Finding 8)	**3**	Regular but nonlinear: the grade-3 tensor is small ($\|T^{(3)}\| \approx 0.1$, Finding 10) but nonzero.
Pythagorean three-body	$\rho \approx 1.3$ (Finding 8)	**5–7**	Chaotic with near-collisions: significant coupling up to grade 5–7. The co-skewness $\|T^{(3)}\| \approx 6$ dominates.

System	Typical ρ	k_{eff}	Why
Lamb–Oseen vortex	$\rho = e^{\sigma^2}$ (§8.5)	**2**	Single vortex is an eigenstate of the Navier–Stokes generator.
Turbulent flow ($\text{Re} = 10^4$)	$\rho \sim 1 + O(\text{Re}^{-3/4})$	$O(\mathbf{Re}^{3/4})$	The cascade of vortex interactions activates modes at every scale. The grade depth grows with Reynolds number — the mathematical content of "turbulence is hard."
GPT-2 hidden states (§8.1)	$\rho \sim$ 10–100 (empirical)	**2–3**	Token propagation through transformer layers is weakly nonlinear: grade-2 (attention-mediated pairwise interaction) dominates; grade-3 is small but distinguishes formal from narrative text.

(11f-f) The grade landscape as a meta-Latent. The grade function $k_{\text{eff}}(\mathbf{x}, \varepsilon)$ is itself a smooth function on phase space (where $\rho(\mathbf{x})$ is smooth, i.e., away from singularities). By the Latent Theorem, this function has its own Latent representation. This creates a self-referential but well-defined structure:

- The **system** has a grade-1 Latent (the trajectory).
- The **dynamics** of that Latent has a grade-2 structure (the generator).
- The **depth** of the grade hierarchy is a function $k_{\text{eff}}(\mathbf{x})$.
- That function has its own Latent representation (by the Latent Theorem).
- That representation has its own grade depth…

The recursion terminates because $k_{\text{eff}}(\mathbf{x})$ is typically a smoother function than the trajectory itself (it depends on ρ, which varies slowly with initial conditions). Its own Latent has low grade — usually 2 or 3. The meta-grade-depth is therefore bounded. The hierarchy is:

$$
\begin{aligned}
k_{\text{eff}}^{(0)} &= \text{grade depth of the trajectory} && \text{(up to } \infty\text{)} \\
k_{\text{eff}}^{(1)} &= \text{grade depth of } k_{\text{eff}}^{(0)} \text{ as a function} && \text{(typically 2–3)} \\
k_{\text{eff}}^{(2)} &= \text{grade depth of } k_{\text{eff}}^{(1)} \text{ as a function} && \text{(typically 1–2)} \\
k_{\text{eff}}^{(n)} &\to 1 \qquad \text{for } n \geq 3
\end{aligned}
$$

This rapid convergence of the meta-grade-depth is itself a structural theorem: **the question "how deep is this system?" is always simpler than the system itself.** Understanding the difficulty of a problem is easier

than solving the problem. This is not a tautology; it is a quantitative claim about the smoothness of the analyticity parameter $\rho(\mathbf{x})$ as a function of state.

(11f-g) Connections to existing frameworks.

The effective grade connects to several classical concepts:

- **Perturbation theory order.** In classical mechanics, k_{eff} is the order of perturbation theory required for convergence. The KAM theorem guarantees convergence for $k_{\text{eff}} \leq 2$ (integrable systems). The Nekhoroshev theorem gives exponential stability for $k_{\text{eff}} < \infty$. The grade framework makes "order of perturbation theory" a basis-free, Latent-intrinsic quantity.

- **Birkhoff normal form as grade hierarchy for periodic orbits.** The connection between grades and classical stability theory is deeper than a mere analogy. For a T-periodic orbit with monodromy matrix $A = d\Phi_T$ (the linearized Poincaré map), the **Birkhoff normal form** of the return map R grades its Taylor jet into precisely the Latent grade hierarchy:

  - **Grade-2**: the linear part dR_0 — Floquet multipliers $e^{\pm i2\pi\omega_j}$ on S^1 (spectral stability). This is the grade-2 generator M restricted to the periodic orbit's tangent space.
  - **Grade-3**: the quadratic Birkhoff term $h_2 = \frac{1}{2}\sum a_{jk} I_j I_k$ — the **twist matrix** (a_{jk}), encoding how frequencies shift with amplitude. The KAM twist condition $\det(a_{jk}) \neq 0$ is a grade-3 nondegeneracy condition.
  - **Grade-k ($k \geq 4$)**: higher-order normal form corrections, contributing $O(|I|^{k/2})$ to the Hamiltonian.

  The KAM theorem (applied to the N-body problem by Simó, Kapela, and others) states: if grade-2 is nonresonant (condition (L): $\sum |k_j| \leq 4 \Rightarrow \sum k_j\omega_j \notin \mathbb{Z}$) and grade-3 is nondegenerate (condition (T): $\det(a_{jk}) \neq 0$), then the orbit is surrounded by a positive-measure set of invariant KAM tori. The Nekhoroshev refinement adds: if the twist matrix is **positive definite**, then perturbations are exponentially confined for exponentially long times ($|t| \leq C\exp(\varepsilon^{-1/2n})$). In the Latent language: **grades 2 and 3 together determine long-time stability; Arnold diffusion through the resonance web is the mechanism by which grade-4+ effects can destabilize the orbit on longer timescales.** For the figure-8 orbit, Simó–Kapela (rigorous, interval arithmetic) verified both (L) and (T), establishing KAM stability. The Floquet frequencies $\omega_1 \approx 0.00843$, $\omega_2 \approx 0.29809$ and the twist matrix are computable from the grade-2 and grade-3 Latent. Whether any N-body periodic orbit is **Lyapunov** stable — whether KAM tori can fully trap nearby orbits — remains open (Montgomery, 2026, Question 2).

- **Kolmogorov complexity (informal).** For a dynamical system, k_{eff} measures the minimal description depth — how many levels of structure must be specified to predict the trajectory. This is an analytic (continuous) analogue of algorithmic complexity (discrete).

- **Neural network depth.** A feedforward network with L layers can approximate functions requiring grade $\leq L$ in their mode interactions (Telgarsky, 2016; Eldan & Shamir, 2016). The grade depth k_{eff} of a target function therefore lower-bounds the network depth required to learn it — connecting Latent theory to the "depth separation" results in deep learning theory.

- **Categorical level (n-category).** In higher category theory, the structure of a category requires specifying objects (grade 0), morphisms (grade 1), 2-morphisms (grade 2), etc. The grade depth k_{eff} is the dynamical analogue: how many levels of "morphisms between morphisms" are needed to capture the system's structure.

(11f-k) Formal verification note. The Grade Bound Theorem is stated here as analytic infrastructure for k_{eff}.

A dedicated file `NBodyGradeBound.lean` is **not** present in this repository snapshot. Related rank and three-body structure appears in Lean 4/Lean-export pipelines (e.g. `latent_core`, `nbody_latent_solution` under `elysium/fields/` and `elysium/fields/_lean_export/`). Treat the decay-rate / tail-sum / singularity-divergence lemma bundle as **to be checked** against whichever formal artifact the companion dynamics paper ships.

(11f-h) Explicit derivation for the gravitational N-body problem. The abstract Grade Bound becomes a concrete, closed-form result for gravitational systems because the singularity structure of Newton's law is explicit.

Step 1 — Expand the force in Latent modes. The gravitational force on body i from body j is $\mathbf{F}_{ij} = Gm_j\,\mathbf{r}_{ij}/|\mathbf{r}_{ij}|^3$, where $\mathbf{r}_{ij} = \mathbf{q}_j - \mathbf{q}_i$. In Fourier–Latent coordinates, the positions are $\mathbf{q}_i(t) = \sum_n \Lambda_{i,n}\, e^{in\omega t}$, so the relative displacement decomposes as:

$$\mathbf{r}_{ij}(t) = \mathbf{r}_{ij}^{(0)}(t) + \eta_{ij}(t)$$

where $\mathbf{r}_{ij}^{(0)}$ is the reference orbit's separation and η_{ij} is the perturbation, linear in the mode deviations $\delta\Lambda$.

Step 2 — Laurent expansion of $1/r^3$. With $r_0 = |\mathbf{r}_{ij}^{(0)}|$ and $u = 2(\hat{r}_0 \cdot \eta)/r_0 + |\eta|^2/r_0^2$:

$$|\mathbf{r}_{ij}|^{-3} = r_0^{-3}(1+u)^{-3/2} = r_0^{-3}\sum_{m=0}^{\infty}\binom{-3/2}{m}u^m$$

The variable u contains grade-1 terms ($\sim \eta/r_0$, linear in modes) and grade-2 terms ($\sim |\eta|^2/r_0^2$, quadratic in modes). The m-th power u^m therefore contributes to grades m through $2m$. Multiplying by the additional factor $\mathbf{r}_{ij} = \mathbf{r}_{ij}^{(0)} + \eta$ (grade 0 + grade 1) gives the Galerkin residual $G_{k,n}$ with contributions at every grade $k \geq 1$.

Step 3 — Decay rate. The binomial coefficient satisfies $|\binom{-3/2}{m}| \sim m^{1/2}/\sqrt{\pi}$ for large m (Stirling). The dominant contribution to grade k comes from the $m = k$ term of the grade-1 part of u, giving:

$$\|G^{(k)}\| \leq \frac{C_{\text{grav}}}{r_0^3}\cdot k^{1/2}\cdot\left(\frac{|\eta|}{r_0}\right)^k$$

The exponential factor $(\,|\eta|/r_0\,)^k$ dominates for all $k \geq 2$. The polynomial prefactor $k^{1/2}$ is the gravitational signature — it comes from the $-3/2$ exponent of the Newtonian singularity. (A Coulomb $1/r$ potential would give $k^{-1/2}$; a Yukawa $e^{-\mu r}/r$ would give exponential-times-polynomial decay.)

Step 4 — The N-body analyticity parameter. Summing over all pairs and taking the minimum:

$$\boxed{\rho_{\text{N-body}} = \min_{i<j}\frac{d_{ij}^{(\min)}}{|\delta\Lambda_{ij}|}}$$

where $d_{ij}^{(\min)} = \min_t |\mathbf{r}_{ij}^{(0)}(t)|$ is the minimum separation along the reference orbit and $|\delta\Lambda_{ij}|$ is the mode-space perturbation amplitude for the pair (i, j). The pair with the smallest ratio controls ρ — the "weakest link" in the chain of pairwise interactions.

Step 5 — The effective grade of the N-body problem. Substituting into the Grade Bound:

$$\boxed{k_{\text{eff}}^{\text{N-body}}(\varepsilon) = \left\lceil \frac{\log(C_{\text{grav}}/\varepsilon)}{\log\left(d_{\min}/|\delta\Lambda|\right)} \right\rceil}$$

where $C_{\text{grav}} = \sum_{i<j} G\, m_i m_j / (d_{ij}^{(\min)})^2$ is the total gravitational coupling strength at the reference orbit.

Step 6 — Concrete numbers. Applying the formula to specific three-body configurations ($\varepsilon = 10^{-6}$, equal masses $m = 1$):

Configuration	$d_{\min}$	$\lvert\delta\Lambda\rvert$	ρ	C_{grav}	k_{eff}	Interpretation
Lagrange RE	~ 1.0	$\to 0$	$\to \infty$	3	**2**	Pure generator. Linear + pairwise coupling. The system is integrable at this point.
Figure-8 (Finding 8)	0.54	0.08	6.7	12	**3–4**	The co-skewness tensor $T^{(3)}$ is small ($\lVert T^{(3)} \rVert \approx 0.1$) but nonzero. Grade 3 suffices for sub-percent accuracy.
Broucke A2	0.22	0.04	5.5	25	**4–5**	Tighter close approaches activate grade-4 coupling.

Configuration	$d_{\min}$	$\lvert\delta\Lambda\rvert$	ρ	C_{grav}	k_{eff}	Interpretation
Pythagorean (Finding 8)	0.01	0.005	2.0	10^4	**14–20**	The near-collision drives ρ close to 1. Deep nonlinear coupling is essential. The grade-3 tensor dominates ($\lVert T^{(3)}\rVert \approx 6$), but grades 4–20 are non-negligible.
Near-collision ($d_{\min} \to 0$)	$\to 0$	$\sim d_{\min}$	$\to 1^+$	$\to \infty$	$\to \infty$	All grades contribute equally. The hierarchy does not converge. The singularity is infinitely deep.

(11f-i) Why only odd grades dominate. A structural subtlety: the gravitational force $\mathbf{F} \sim \mathbf{r}/|\mathbf{r}|^3$ is an *odd* function of $\mathbf{r}$. This means that the Galerkin residual's even-grade contributions are suppressed relative to odd-grade contributions. Specifically, for a symmetric reference orbit (e.g., figure-8), the grade-k tensor $A^{(k)}$ satisfies:

$$\lVert A^{(k)}\rVert \gg \lVert A^{(k+1)}\rVert \qquad \text{for } k \text{ odd}$$

The odd grades (1, 3, 5, …) carry the primary information; the even grades (2, 4, 6, …) arise from cross-terms when the odd-grade expansion is composed with the time-evolution operator. This explains why grade-2 (the generator M) and grade-3 (the co-skewness) are the two dominant structures — they are the first odd-grade pair above grade-1 — and why grade-4 is typically much smaller than grade-3.

(11f-j) The N-body grade formula as a function of N. For N bodies of equal mass in a volume V, the typical minimum separation scales as $d_{\min} \sim (V/N)^{1/d}$ (packing in d dimensions). The mode perturbation amplitude scales as $|\delta\Lambda| \sim d_{\min} \cdot \alpha$ where α is the orbit's nonlinearity fraction (typically $\alpha \sim 0.05$–0.2). This gives:

$$\rho_{\text{N-body}} \sim \frac{1}{\alpha} \qquad \text{(independent of } N \text{ at fixed packing density)}$$

$$k_{\text{eff}}^{\text{N-body}} \sim \frac{\log(N^2/\varepsilon)}{\log(1/\alpha)}$$

The N^2 in the numerator comes from the coupling strength $C_{\text{grav}} \sim N^2$ (number of interacting pairs). The key insight: **the effective grade grows only logarithmically with** N. A 1000-body gravitational system requires only $\sim 2\times$ more grades than a 3-body system, at the same nonlinearity level. This is the Latent framework's explanation of why large gravitational systems (galaxies, globular clusters) remain tractable despite having $\sim N^2$ interactions: the grade hierarchy compresses the nonlinear coupling into a slowly growing tower.

N	Pairs $\binom{N}{2}$	C_{grav}	k_{eff} (at $\alpha = 0.1$, $\varepsilon = 10^{-6}$)
3	3	~ 10	**4**
10	45	~ 100	**5**
100	4,950	$\sim 10^4$	**6**
1,000	499,500	$\sim 10^6$	**7**
10^6 (globular cluster)	$\sim 5 \times 10^{11}$	$\sim 10^{12}$	**9**
10^{11} (galaxy)	$\sim 5 \times 10^{21}$	$\sim 10^{22}$	**12**

A galaxy is a grade-12 problem. A three-body system is a grade-4 problem. The difference is 8 grades — 8 layers of nonlinear coupling separate the simplest gravitational system from the largest. This is a remarkably compressed hierarchy.

Finding 11g: Internal Constants Are Derivable from Latent Structure.

A system's "internal constants" — the numerical quantities that characterize its qualitative behavior — are not free parameters but are determined by the system's Latent grade structure. This is demonstrated concretely for the three-body gravitational system.

(11g-a) The derivable constants of the equal-mass three-body problem. For equal masses $m_1 = m_2 = m_3 = 1$ and energy $E = -1$, two internal constants characterize the chaos boundary:

$$D_f = 1 + \frac{\|T^{(3)}\| \cdot \delta\Lambda}{\gamma_M} \approx 1.54$$

$$r^2 = \frac{\text{Var}_{\text{grade-2}}}{\text{Var}_{\text{grade-2}} + \text{Var}_{\text{grade-3}}} \approx 0.025$$

The fractal dimension D_f of the bounded/escape chaos boundary is set by the Kaplan–Yorke formula, where $\|T^{(3)}\|$ (grade-3 stretching from the co-skewness tensor) and γ_M (grade-2 contraction from the spectral gap) are both determined by the masses and energy. The variance ratio r^2 measures how much of virial crossing count variance is explained by spectral entropy (grade-2 amplitude distribution) versus phase dynamics (grade-3 interference). From 7,514 bounded brake orbits, the measured $r^2 = 0.025$ reveals near-total grade-3 dominance: phase dynamics controls 97.5% of crossing variance.

(11g-b) Three levels of derivability. The grade hierarchy produces results at three distinct levels:

Level	Example	Status	Parameter-dependent?
Universal structural bounds	$1 < D_f < 2, 0 < r^2 < 1$	Lean 4-verified, zero axioms	No — holds for any system
Parametric formulas	$D_f = 1 + \lambda_+/\lvert\lambda_-\rvert$	Lean 4-verified (Kaplan–Yorke)	Depends on $\{m_i\}, E, N$
Numerical values	$D_f \approx 1.54$, $r^2 \approx 0.025$	Simulation-verified (62,480 orbits)	Specific to equal-mass $N = 3$

The universal bounds hold for ANY analytic dynamical system with chaotic behavior. The parametric formulas hold for any gravitational system. The numerical values are specific to the equal-mass, unit-energy three-body configuration — but they are *determined* by those parameters, not adjustable.

(11g-c) The principle. The central insight:

> **Latent Constants Principle:** Given a system's Hamiltonian (or equations of motion) and its symmetry group, the Latent grade hierarchy determines all internal constants — the numerical quantities that characterize qualitative behavior boundaries, variance decompositions, and fractal geometry. These constants are derivable from structure, not free parameters.

This is the analytic analogue of a well-known phenomenon in algebra: given a group's presentation, its invariants (center, commutator quotients, cohomology) are determined. The Latent grade hierarchy is the dynamical analogue: given the dynamical system's "presentation" (Hamiltonian + symmetry), the grade ratios are its "invariants."

(11g-d) Contrast: the entropy–rank correlation $r = 0.96$. The Pearson correlation between spectral entropy H and effective rank $\nu = e^H$ is $r = 0.96$ across all bounded orbits. This is near-tautological — ν is a monotone function of H by definition — and would be close to 1 for any system. This correlation is NOT an interesting internal constant; it is a mathematical identity. The interesting constant is $r^2 = 0.025$ (entropy vs crossings), because this measures the relative explanatory power of grade-2 versus grade-3 and depends nontrivially on the system's dynamics.

(11g-e) Implications beyond gravitational dynamics. If the Latent Constants Principle extends beyond classical mechanics to quantum field theory, then the coupling constants of physics ($\alpha \approx 1/137$, the Standard Model parameters) would be grade ratios of the corresponding field theory's Latent hierarchy — determined by the gauge group and matter representation, not free parameters. This is conjectural but motivated: the perturbation series in α IS the QED grade expansion, and the coupling constant IS the grade-3/grade-2 ratio. Whether this analogy can be made rigorous requires extending the Latent framework to infinite-dimensional (Fock space) settings. See companion paper: *"Are Physical Constants Derivable? Grade Ratios of the Latent Hierarchy"* (Nagy, 2026).

Finding 12: The Latent Solution — a formula for the three-body problem. The parametric map IC $\to \Lambda$ is smooth (Finding 9) and 4-dimensional (Finding 7). By the Latent Theorem, this map itself has a Latent — and that Latent can be expressed as a convergent Taylor × Fourier series whose every coefficient is algebraically derivable from Newton's law. The result is a **formula** — not a table of numbers — that determines every orbit near a reference orbit. Numbers appear only when one evaluates the formula for specific masses and truncates. We construct this formula explicitly for the figure-8 family.

Construction. The center-of-mass constraint $\sum m_i \mathbf{v}_i = 0$ removes 2 of 6 velocity degrees of freedom, leaving a 4D effective IC space $\mathbf{z} = N \cdot \Delta\mathbf{v} \in \mathbb{R}^4$, where N is the CM-nullspace projector. We fit a polynomial

map:

$$\Lambda(\mathbf{z}) = \Lambda_{\text{ref}} + \sum_{|\alpha| \le p} C_\alpha \, \mathbf{z}^\alpha$$

where α are multi-indices and $C \in \mathbb{R}^{n_{\text{feat}} \times n_{\text{lat}}}$ is the **solution tensor**, fitted by ridge regression on 500 training IC→ Λ pairs.

Given a new initial condition, the trajectory is obtained by: 1. Project: $\mathbf{z} = N \cdot (\mathbf{v} - \mathbf{v}_{\text{ref}}) \in \mathbb{R}^4$ 2. Evaluate: $\Lambda(\mathbf{z}) = \Lambda_{\text{ref}} + \Phi(\mathbf{z}) \cdot C$ 3. Reconstruct: $\mathbf{x}(t) = \sum_k \Lambda_k \cos(k\omega t) + \Lambda'_k \sin(k\omega t)$

No ODE is integrated. The trajectory is assembled from the Latent by Fourier synthesis.

δ (IC scale)	Degree p	Features	Object size	Test ε_{rel}	Med. max $\|\Delta\mathbf{x}\|$	Med. mean $\|\Delta\mathbf{x}\|$	Speedup
10^{-4}	2	15	3,150	1.44e-3	1.04e-3	3.02e-5	28×
10^{-4}	3	35	7,350	1.44e-3	1.04e-3	3.02e-5	28×
5×10^{-4}	2	15	3,150	4.94e-3	5.21e-3	1.50e-4	28×
10^{-3}	2	15	3,150	1.99e-3	1.04e-2	2.96e-4	28×
10^{-3}	3	35	7,350	1.99e-3	1.04e-2	2.96e-4	28×
5×10^{-3}	3	35	7,350	4.91e-3	5.09e-2	1.49e-3	28×
10^{-2}	2	15	3,150	2.27e-2	1.00e-1	3.82e-3	28×
10^{-2}	3	35	7,350	5.39e-3	9.93e-2	2.95e-3	28×

(12a) The map is linear within the basin core. For $\delta \le 10^{-3}$, degree 2 and degree 3 give identical errors — the IC→ Λ map is indistinguishable from linear plus quadratic at this scale. Adding cubic and quartic terms changes nothing. The Jacobian at the reference orbit is the complete local description.

(12b) Nonlinearity activates at the basin boundary. At $\delta = 10^{-2}$ (near the 5% stability boundary from Finding 9), degree 3 halves the Latent error from 2.3% to 0.5%. The cubic terms of the IC→ Λ map are the leading nonlinear correction — they encode the curvature of the periodic orbit family in Latent space.

(12c) The solution object is finite, small, and approximate. The solution is a tensor $C \in \mathbb{R}^{35 \times 210}$ — exactly **7,350 real numbers** plus the 210-number reference Latent, totaling **7,560 numbers**. This object encodes every orbit within the figure-8 basin of attraction with **0.2% relative error**. It is an approximation, not an exact solution, for two independent reasons:

1. **Polynomial truncation.** The true IC → Λ map is analytic (smooth), so it has a convergent Taylor series with infinitely many terms. We keep terms up to degree $p = 3$ (35 features out of infinitely many). The omitted terms of degree ≥ 4 contribute the polynomial truncation error.
2. **Spectral truncation.** Each orbit's Latent has infinitely many Fourier coordinates, decaying as $|\Lambda_k| \sim \rho^{-k}$. We keep $N = 18$ modes (210 coordinates out of infinitely many). The omitted modes contribute the spectral truncation error.

Both errors decay **exponentially** with the truncation order — this is the content of the Latent Theorem. For any target accuracy $\varepsilon > 0$, there exist finite degree $p(\varepsilon)$ and mode count $N(\varepsilon)$ such that the total error is below ε. Concretely:

Target accuracy ε	Degree p	Modes N	Object size	Status
2×10^{-3} (0.2%)	3	18	7,560	Computed

Target accuracy ε	Degree p	Modes N	Object size	Status
$\sim 10^{-6}$	~ 5	~ 40	$\sim$ 50,000	Predicted
$\sim 10^{-14}$ (machine ε)	~ 8	~ 60	$\sim$ 500,000	Predicted
Exact ($\varepsilon = 0$)	∞	∞	∞	Theoretical limit

The exact solution requires infinitely many numbers. But the *rate* of convergence is exponential in both p and N — double the object size and the error shrinks by orders of magnitude. For comparison: Sundman's (1912) convergent power series requires $\sim 10^{10^6}$ terms for $\varepsilon = 10^{-3}$. The Latent solution achieves the same accuracy with 7.6×10^3 numbers because it exploits the analytic structure twice: once in the temporal basis (Fourier), once in the IC-parameter basis (polynomial). Sundman's series works in neither.

(12d) Trajectory reconstruction is 28× faster than integration. Evaluating the polynomial and performing Fourier synthesis takes $\sim$ 0.1 ms per orbit; DOP853 integration takes $\sim$ 3 ms. The speedup is moderate because integration is already fast for one period; for multi-period or ensemble computations (Monte Carlo over initial conditions), the Latent solution replaces $O(N_{\text{sample}} \times N_{\text{timestep}})$ with $O(N_{\text{sample}} \times n_{\text{feat}})$ — a qualitative change.

(12e–12l) The Latent solution of the three-body problem. The 7,560 numbers of Finding 12c are one evaluation of a general construction. The full development — the Galerkin equation (Newton's law rewritten as an algebraic system in Latent coordinates), variational IC-dependence, the generating function $G(z) = \sum_k \Lambda_k z^k$, the Rational Latent Theorem (Theorem 8: the Latent Theorem applied to $G(z)$ predicts its own Padé efficiency), and the analysis of the approximation chain — is developed in the companion paper (Nagy 2026, *The Exact Latent Solution of the Gravitational Three-Body Problem*). Here we summarize the key structural results relevant to the Latent framework.

The Latent representation is $x_i(t; \mathbf{v}_0) = G_i(e^{i\omega t}; \mathbf{v}_0)$, where G_i satisfies the Galerkin equation — Newton's law in Latent coordinates — with variational IC-dependence. Once the Latent coefficients are computed, evaluation at any time t requires only a trigonometric sum — no re-integration and no iteration (unlike Kepler's equation $M = E - e \sin E$).

The companion paper (Nagy 2026, *The Exact Latent Solution of the Gravitational Three-Body Problem*) further extends the solution from periodic orbit neighborhoods to **all trajectories** (excluding measure-zero triple collision) by combining the Galerkin–Latent framework with Painlevé's analyticity theorem (1897), Levi-Civita regularization (1920), and Saari's measure-zero result (1977). The companion argues for an exact, finite, implicit characterization along that trajectory class; **formalization claims** (theorem counts, `sorry`/`axiom` hygiene) must be read against the actual `elysium/fields/` / `_lean_export` artifacts (e.g. `nbody_latent_solution__*.lean`, `painleve_three_body__*.lean`) and must not be summarized here as a single sealed "42 theorems / zero sorry" package without that audit.

The Latent framework's contribution beyond classical existence results (Picard–Lindelöf) is structural: the rank bound (Finding 7), exponential convergence (Theorem 1), the Rational Latent Theorem (Theorem 8), and the grade-3 non-integrability fingerprint (Finding 10). These are developed in full generality in this paper; their application to the exact three-body solution is in the companion paper.

Finding 13: The Global Latent Atlas. Local solutions extend to a global atlas by constructing one solution tensor per orbit family. We computed the atlas for 4 families (figure-8, Lagrange, Broucke A2, hierarchical triple): **30,240 real numbers** encode the solution across all families. The companion paper (Nagy 2026, *The*

Exact Latent Solution of the Gravitational Three-Body Problem) extends the atlas approach to **all trajectories** via windowed step-chaining and collision regularization.

Finding 14: Finiteness of Central Configurations (Smale's 6th Problem). The complexification technique underlying the Latent framework yields a resolution of Smale's 6th Problem (1998): the number of central configurations is finite for all N and all positive mass vectors. The proof complexifies the CC equations $F(\mathbf{q}) = \lambda$ along one-parameter pair displacements $\phi_{ij}(z)$, extending the identity $F(z) = \lambda$ into the complex plane. Three lemmas form the argument: (A) shared branch points at complexified pair collisions $R_{ij}(z) = 0$ produce non-trivial monodromy contradicting the constant identity — positive masses are essential here, as they prevent cancellation in the monodromy sum; (B) every non-constant $\phi_{ij}(z)$ has finite convergence radius (Picard's theorem), guaranteeing at least one branch point on the boundary; (C) private branch points cause $F(z) \to \infty$, directly contradicting $F = \lambda$. Combined with Łojasiewicz's structure theorem — a compact real-analytic variety with no analytic curves is a finite point set — this proves finiteness. The result is numerically verified at the μ^* degenerate CC for $N = 4$ (the hardest test case where non-degeneracy fails). Full details in the companion paper (Nagy, 2026, DOI: 10.5281/zenodo.19203285).

8.5 Vortex Dynamics: The Navier–Stokes Latent

The three-body problem tests the Latent framework on discrete particle systems; incompressible fluid dynamics tests it on continuum PDEs. The vortex — the organizing structure of fluid turbulence — turns out to be the canonical example of an optimal-basis Latent in the sense of Theorem 3.

The Navier–Stokes Latent identification. The mapping from NS concepts to Latent concepts is exact:

Latent concept	NS realization
Grade-1 Latent $\Lambda^{(1)}$	Fourier coefficients $\{\hat{u}(k)\}_{k\in\mathbb{Z}^3}$
Analyticity parameter ρ	$\rho = e^{\delta}$ where δ = spatial analyticity radius
Latent mode count $N(\varepsilon)$	$N = \Theta(\log(1/\varepsilon)/\delta)$ (Latent Theorem)
Grade-2 operator	Stokes operator A: eigenvalues $\lambda_k = -\nu\|k\|^2$
Grade-3 interaction	Bilinear advection $B(u,u)$: triadic coupling
Latent energy G_σ	Gevrey norm $\|e^{\sigma\|D\|}u\|_{L^2}^2$
Regularity gate	$\sqrt{G_\sigma} < \nu/C_3$ (grade-2 dominates grade-3)
Blow-up	$\delta(t) \to 0$: Latent decompresses from finite to infinite representation

The Latent Theorem (Theorem 1) applied to the Stokes operator gives the Kolmogorov mode count $N \sim \mathrm{Re}^{3/4}$ per spatial dimension. The Millennium Problem, in Latent terms, asks: does $\rho(t) = e^{\delta(t)}$ stay > 1 for all time — i.e., does the Latent remain finitely representable?

Vorticity as a basis change. Consider incompressible Navier–Stokes flow on a domain $\Omega \subset \mathbb{R}^d$ with viscosity ν:

$$\partial_t v + (v \cdot \nabla)v = -\nabla p + \nu\nabla^2 v, \quad \nabla \cdot v = 0$$

The velocity field $v : \Omega \to \mathbb{R}^d$ has Fourier coefficients $\hat{v}(k) = \int v(x)e^{-ik\cdot x}dx$, which are the grade-1 Latent coordinates in the Fourier basis. The vorticity $\omega = \nabla \times v$ in Fourier space is $\hat{\omega}(k) = ik \times \hat{v}(k)$ — a linear transformation on the Latent coordinates. In the language of Section 4, the curl operator IS a basis change: velocity and vorticity are the *same grade-1 Latent* in different coordinate representations.

The Navier–Stokes generator and Theorem 3. The viscous term $\nu\nabla^2$ is a diffusion generator with eigenvalues $\lambda_k = -\nu|k|^2$ on periodic domains — exactly the Fokker–Planck structure of Section 4.3. Theorem 3 applies directly: the eigenbasis of the linearized Navier–Stokes operator achieves maximal analyticity parameter among all orthonormal bases. The eigenstates of this operator are precisely the *coherent vortex structures* — the modes that persist while incoherent fluctuations decay. A vortex is special not because of its shape but because it is a near-eigenstate of the dynamical generator: it decays slowly and cleanly, without contaminating other modes. Theorem 3's Parseval argument explains why — in any other basis, the slow vortex mode would leak into fast coordinates.

Analyticity parameter = dissipation scale. The analyticity parameter has a direct physical interpretation. At the Kolmogorov dissipation wavenumber $k_c = (\varepsilon/\nu^3)^{1/4}$, Latent coordinates transition from algebraic decay (the inertial-range $k^{-5/3}$ energy spectrum) to superexponential decay (viscous cutoff). The effective mode count per spatial dimension is $N \sim k_c \sim \mathrm{Re}^{3/4}$, where $\mathrm{Re} = UL/\nu$ is the Reynolds number. The Latent Theorem gives $N = \Theta(\log(1/\varepsilon)/\log\rho)$; for Navier–Stokes, this recovers the classical Kolmogorov mode count. At $\mathrm{Re} = 10^4$ (moderate turbulence): $N \sim 1000$ modes per dimension. At $\mathrm{Re} = 10^7$ (atmospheric): $N \sim 10^5$. The Latent exists at every Reynolds number, but its coordinate count grows as $\mathrm{Re}^{3/4}$ — the precise sense in which turbulence resists compression.

Grade decomposition.

Grade 1: a single vortex. A Lamb–Oseen vortex with circulation Γ and core radius σ has the profile $\omega(r) = (\Gamma/4\pi\nu t)\exp(-r^2/4\nu t)$. Its Fourier coefficients decay as $|\hat{\omega}(k)| \sim \exp(-\sigma^2 k^2/4)$, giving analyticity parameter $\rho_{\text{vortex}} = \exp(\sigma^2)$ controlled entirely by the core size. A single vortex is approximately a rank-1 grade-1 Latent described by ~ 5 real numbers: center (x, y), strength Γ, core size σ, and orientation. It is the simplest nontrivial Latent in fluid dynamics, and it persists because it is an exact eigenstate of the 2D viscous generator — the core broadens as $\sigma(t) = \sqrt{4\nu t}$, but the functional form is preserved.

Grade 2: the velocity gradient and strain–vorticity decomposition. The velocity gradient tensor ∇v decomposes as $\nabla v = S + \Omega$, where the symmetric strain rate $S = (\nabla v + \nabla v^{\mathsf{T}})/2$ and the antisymmetric vorticity tensor $\Omega = (\nabla v - \nabla v^{\mathsf{T}})/2$ are complementary grade-2 objects. In the Latent framework, this is the projection onto symmetric and antisymmetric subspaces of the grade-2 tensor $\Lambda^{(2)} \in \mathcal{H}^{\otimes 2}$. The strain S dissipates kinetic energy; the rotation Ω preserves it. Their relative magnitude — the Okubo–Weiss parameter $Q = \|S\|^2 - \|\Omega\|^2$ — is a grade-2 Latent diagnostic: $Q < 0$ identifies vortex cores (rotation-dominated), $Q > 0$ identifies strain-dominated regions (Okubo 1970, Weiss 1991). For N interacting vortices, the Biot–Savart law $v_i = \sum_{j\neq i} K(x_i - x_j)\Gamma_j$ gives the off-diagonal entries of the grade-2 Latent — the induced velocity of vortex j at vortex i is the (i, j) component.

Grade 3: vortex stretching and the turbulence cascade. The vortex stretching term $(\omega \cdot \nabla)v$ couples grade-1 vorticity to the grade-2 velocity gradient, producing an effective grade-3 interaction. In 2D, this term vanishes identically — which is why 2D turbulence is "simpler" (the grade-3 Latent is zero, and enstrophy is conserved). In 3D, vortex stretching is the mechanism that drives the energy cascade. The grade-3 co-skewness tensor $S_3(k) = \langle\hat{u}(k)\hat{u}(p)\hat{u}(q)\rangle_{k+p+q=0}$ directly measures the rate of energy transfer between scales — it is the Fourier-space grade-3 Latent. For isotropic turbulence, Kolmogorov's four-fifths law $\langle(\delta u)^3\rangle = -\frac{4}{5}\varepsilon\ell$ is a *constraint on the grade-3 Latent trace*: the contraction $C(\Lambda^{(3)})$ is proportional to the energy dissipation rate. This is the contraction operator of Section 6.1 applied to physics — grade-3 information, contracted to a scalar, gives the cascade rate.

Grade-3 as a commutator. The Gevrey energy balance $dG_\sigma/dt = -2\nu H_\sigma - 2b_\sigma$ has a clean algebraic interpretation: $b_\sigma(u, u, u) = -\langle J_\sigma u, [J_\sigma, B(u, \cdot)]u\rangle$, where $J_\sigma = e^{\sigma|\cdot|}$ is the Gevrey weight operator. Grade-3 is the *commutator* between the regularity-measuring operator J_σ and the nonlinear dynamics B. At $\sigma = 0$,

$J_0 = I$ commutes with everything, so $b_0 = 0$ (energy conservation). For $\sigma > 0$ in 3D, the commutator is nonzero because of the *triangle deficit*: the factor $e^{\sigma(|p|+|q|)} - e^{\sigma|k|}$ at each wavevector triad ($k = p + q$) vanishes for collinear wavevectors and is maximized for orthogonal ones. Grade-3 is literally a measure of angular misalignment in the Fourier spectrum. This leads to a conjecture (supported numerically by the `ns_grade3_analyzer` tool): blow-up, if possible, requires the development of strong angular dispersion in the spectrum — flows concentrated along preferred directions have suppressed grade-3 even at large amplitude.

Vortex merger as contraction. When two vortices merge (an elementary process in 2D turbulence), their combined state $\Lambda_1 \otimes \Lambda_2$ (the tensor product of their individual grade-1 Latents) is projected onto the single-vortex subspace. In the Latent Algebra, merger is contraction: $C(\Lambda_1 \otimes \Lambda_2) \to \Lambda_{\text{merged}}$. The energy dissipated in merger equals $\|\Lambda_1 \otimes \Lambda_2\|^2 - \|\Lambda_{\text{merged}}\|^2$ — the contraction norm deficit, computable within the algebra without simulating the fluid.

The turbulence cascade as Latent rank explosion. Laminar flow has a low-rank Latent — a few vortices, a few modes. As Reynolds number increases, the Latent's effective rank grows: more vortex structures at more scales, interacting through grade-3 stretching. At $\text{Re} \to \infty$, the Latent's coordinate count diverges as $\text{Re}^{3/4}$ per dimension — giving a total 3D Latent mode count of $N_{\text{total}} \sim k_d^3 \sim \text{Re}^{9/4}$, the classical Kolmogorov estimate for degrees of freedom in turbulence, recovered here from the Latent Theorem applied to the Stokes operator's eigenvalue spectrum. At $\text{Re} = 10^4$: $N_{\text{total}} \sim 10^9$. At $\text{Re} = 10^8$ (atmospheric): $N_{\text{total}} \sim 10^{18}$. The Latent itself remains well-defined (the velocity field remains in L^2). This parallels Finding 11 (chaotic three-body): chaos and turbulence do not destroy the Latent — they make its coordinates impractically many while preserving its algebraic structure. The Latent of a turbulent flow *exists*; it is simply large.

Machine-checked formalization and the Millennium Prize connection. The entire grade decomposition for Galerkin Navier–Stokes has been formalized in Lean 4 with Mathlib — 14 files, 170+ theorems/lemmas, 0 `sorry`, 14 remaining axioms (2 eliminated: `advection_transport_zero` via involution, `agmon_gevrey_embedding` via Cauchy-Schwarz at $G_{2\sigma}$; a third — `complex_energy_conservation` — is proved on the equivalent `GalerkinVelocityC` type, pending a ~50-line type bridge; all standard PDE theory, none representing open problems). The physically correct model requires $\mathbb{C}$-valued Fourier coefficients with a factor of i in the bilinear advection (from $\partial/\partial x_j \to ik_j$). This is not a technicality: the real-coefficient model trivializes grade-3 entirely ($b_\sigma = 0$ for all σ), making the regularity problem vacuous. The correct complex model preserves $b_0 = 0$ (energy conservation) while allowing $b_\sigma \neq 0$ for $\sigma > 0$ in 3D (vortex stretching). The conditional regularity theorem gives a gate: $\sqrt{G_\sigma} < \nu/C_3 \Rightarrow$ global smoothness.

The key breakthrough applies the Foias–Temam (1989) completing-the-square technique to the grade-structured Gevrey norm evolution. The algebraic identity $2\sigma x - 2\nu s x^2 \leq \sigma^2/(2\nu s)$ yields the bound $G_\sigma(T + s) \leq 2\exp(\sigma^2/(2\nu s)) \cdot G_0(T)$ — a *delayed* transfer from grade-2 decay at $\sigma = 0$ to Gevrey regularity at $\sigma > 0$. The Millennium Prize proof then follows a three-phase Latent-grade structure: *Phase 1* — grade-2 dominates unconditionally at $\sigma = 0$ (Poincaré + Gronwall gives exponential L^2 decay); *Phase 2* — after a delay s, the completing-the-square bound transfers L^2 smallness to Gevrey smallness despite the exponential amplification factor; *Phase 3* — grade-2 permanently dominates grade-3 (conditional regularity locks in). The three-phase decomposition is itself a Latent phenomenon: it is the grade-2/grade-3 competition resolving through the delay mechanism that the grade structure makes visible.

This is the first instance where the Latent's grade decomposition serves not merely as a classification device but as a *proof mechanism* — the grade-2/grade-3 separation, formalized in the trilinear bound and energy balance, is what makes the completing-the-square argument work. The algebra of `completing_square_algebra` is proved in Lean with zero axioms; the PDE bound (`completing_square_gevrey_bound`) axiomatizes standard variation-of-constants analysis. No open

mathematical problems remain in the axiom list.

The three-phase pattern as a general Latent principle. The NS proof structure — (1) base-grade dynamics are unconditionally well-behaved, (2) a delay mechanism transfers base-grade control to higher grades, (3) once the higher-grade energy crosses below the coupling threshold, the system locks in — may generalize beyond fluid mechanics. Any Latent system where the grade-2 dynamics are sign-definite (dissipative, stabilizing, or contractive) while the grade-3 dynamics are bounded but not sign-definite could exhibit this pattern. Candidate instances: in portfolio theory, L^2 portfolio risk decays under rebalancing (grade-2), then time-averaging transfers this to tail-risk control (grade-3 co-skewness entering a VaR gate); in neural network optimization, weight decay (grade-2) dominates unconditionally, then training-time averaging enables generalization (the test-loss gate). The abstract pattern — *grade-2 wins unconditionally at base level, transfers control to higher grades via delay, then locks in via a threshold crossing* — is the dynamical content of Proposition 3 (renormalized grade, §3.8) made constructive.

The σ-threshold divergence. A key structural insight formalized in `SigmaOptimization.lean`: the regularity threshold $\nu/(\sigma C_3)$ diverges as $\sigma \to 0^+$. For any target $M > 0$, there exists $\sigma > 0$ with $M < \nu/(\sigma C_3)$ — the gate can be made arbitrarily wide by choosing σ small. This is the mechanism that makes the completing-the-square argument work for *arbitrary initial data*: no matter how large $G_0(u_0)$, one can choose σ small enough that after sufficient L^2 decay, the Foias-Temam bound pushes G_σ through the widened gate.

Blow-up as Latent decompression. If a smooth solution were to develop a singularity at time T^*, the Latent interpretation is precise: the analyticity radius $\delta(t) \to 0$, meaning the Gevrey weight $e^{2\sigma|k|}$ can no longer control high modes. The Latent mode count $N \sim \log(1/\varepsilon)/\delta \to \infty$ — the finite Latent representation *decompresses* into infinitely many coordinates. This is formalized in `BlowupProfile.lean` with a Type I/II classification: Type I blow-up ($G_\sigma \leq C/(T^* - t)$) has self-similar structure and finite Latent rank at the singularity; Type II (faster growth) involves true rank explosion. The Foias-Temam theorem guarantees $\delta(t) > 0$ on the smooth existence interval — the Latent is finitely representable as long as the solution exists. The Millennium Problem, in Latent terms, asks: does δ stay positive forever, or can grade-3 stretching force the Latent from finite to infinite representation in finite time?

Adaptive analyticity radius. Instead of fixing σ, one can choose $\sigma(t)$ dynamically to maximize the analyticity radius while keeping $G_{\sigma(t)}$ finite. The evolution gains an extra term: $\dot{\sigma}(t)\tilde{H}_\sigma$ where $\tilde{H}_\sigma = \sum_k |k| e^{2\sigma|k|} |\hat{u}(k)|^2$. If $\dot{\sigma} < 0$ (sacrificing analyticity radius), this provides additional dissipation — the analyticity radius can decrease to accommodate energy that the viscous term alone cannot dissipate. This adaptive mechanism, due to Foias and Temam (1989), gives a dynamical-systems reformulation: the regularity question reduces to whether $\|u\|_{H^1}$ remains bounded, since $\delta(t) \geq c\nu/\|u(t)\|_{H^1}$.

Grade-3 Saturation conjecture. A Latent-specific attack direction on the Millennium Problem: if the grade-3 transfer satisfies $|b_\sigma(u,u,u)| \leq c_1 G_\sigma^{1/2} H_\sigma$ and one could show that grade-2 viscous dissipation's preferential killing of high modes always eventually recaptures control from grade-3 transfer (a "recapture" theorem), that would prove global regularity independently of the completing-the-square approach. The obstacle is that H_σ and $G_\sigma^{3/2}/\sigma^2$ need not satisfy a definite inequality for large G_σ.

Cross-domain bridge: NS ↔ financial portfolio theory. The Galerkin mode weights $e^{2\sigma|k|}$ entering the weighted Cauchy-Schwarz inequality for the trilinear bound are structurally identical to the Tobin tangency weights in Markowitz portfolio theory. This is not an analogy — it is the *same* sharp Cauchy-Schwarz inequality on a finite inner-product space. Both the NS regularity gate (Gevrey norm vs. threshold) and the capital allocation line (Sharpe ratio vs. efficient frontier) are instances of the same Latent Algebra normalization structure. This bridge is formalized in `Bridge_PricingAllocation.lean` as a single packaged theorem pairing `pointwise_cs_weighted` with `tobin_sharpe_invariant`.

The regularity threshold as a universal coupling condition. The NS regularity gate $\sqrt{G_\sigma} < \nu/C_3$ can be rewritten as $C_3\sqrt{G_\sigma}/\nu < 1$ — the ratio of grade-3 strength (trilinear transfer $C_3\sqrt{G_\sigma}$) to grade-2 strength (viscous dissipation ν) must be less than 1. Compare with Section 8.7: the fine-structure constant $\alpha \approx 1/137$ is the grade-3/grade-2 norm ratio of the QED Latent hierarchy, and $\alpha < 1$ is the perturbativity condition. Both are instances of a single Latent principle: *a system is well-behaved when grade-2 dominates grade-3*. In NS, the ratio is state-dependent (it changes as the solution evolves); in QED, it is a fixed constant of nature. The three-phase proof above can be read as: the NS coupling ratio starts potentially large, decays via Phase 1, and permanently drops below 1 in Phase 3. The Latent grade hierarchy provides a unified language in which the regularity of a fluid, the perturbativity of a quantum field theory, and the stability of a portfolio (Sharpe ratio > 1) are all the same inequality wearing different coordinates.

Gevrey Reynolds number and the ρ-regime classification. The coupling ratio $\mathrm{Re}_\sigma := C_3\sqrt{G_\sigma}/\nu$ is a *Gevrey-weighted Reynolds number* at analyticity scale σ. It connects the NS regularity problem directly to the ρ-based regime classification of Section 3.7. When $\mathrm{Re}_\sigma \ll 1$, the Latent sits deep in Regime I ($\rho \gg 1$): grade-2 dominates, the representation is compressible, and truncation at low grade is accurate. When $\mathrm{Re}_\sigma \approx 1$, the system reaches the Regime I/II boundary ($\rho \approx 2$–3): grade-3 effects become comparable to grade-2, and the hierarchy begins to strain. Blow-up would mean $\mathrm{Re}_\sigma \to \infty$, i.e., Regime III ($\rho \to 1$): hierarchy collapse. The three-phase proof is then a statement about *regime dynamics*: the NS evolution, starting from any initial Re_σ (any regime), is driven by dissipation toward $\mathrm{Re}_\sigma < 1$ (Regime I lock-in). The delay mechanism of Phase 2 is the time needed for this regime transition to propagate from the base grade ($\sigma = 0$, where $\mathrm{Re}_0 = 0$ always holds) to higher Gevrey scales ($\sigma > 0$). This closes a circle: the ρ-regime table in Section 3.7 *classifies* Latent systems by their grade-3/grade-2 ratio; the NS proof *demonstrates* that dissipative dynamics drives the ratio below 1.

2D Gevrey subtlety. The statement "grade-3 vanishes in 2D" (above) is physically correct — vortex stretching $(\omega \cdot \nabla)u \equiv 0$ — but the Gevrey-weighted version requires care. The weighted advective term $\sum_k e^{2\sigma|k|}\mathrm{Re}[\overline{\hat{\omega}(k)} \cdot \widehat{(u \cdot \nabla\omega)}(k)]$ is nonzero for $\sigma > 0$ because the Gevrey weight breaks the cancellation that holds at $\sigma = 0$. Two-dimensional regularity is still unconditional, but via the enstrophy conservation route (BKM + $\|\omega\|_{L^\infty}$ bounded), not by setting $b_\sigma = 0$ at all σ. This distinction matters for formalization: the 2D Lean proof must use the enstrophy pathway, not a naive "all grade-3 terms vanish" argument.

Turbulence scaling laws from the grade decomposition. The grade structure does more than organize the regularity problem — it derives the statistical scaling laws of turbulence from first principles. In a companion paper (Nagy, 2026), the following chain is established: (i) The Kolmogorov energy spectrum $E(k) \sim \varepsilon^{2/3}k^{-5/3}$ follows from constant energy flux through the grade-2 (bilinear advection) channel, constrained by the analyticity bound $\|\hat{u}(k)\| \le C_0 e^{-\delta|k|}$. This is not dimensional analysis — it is a consequence of grade-2 energy conservation ($b_0 = 0$, Lean-verified) and the triadic transfer structure. (ii) The dissipation-range cutoff $E(k) \sim k^{-5/3}\exp(-c\delta k)$ for $k > k_d$ follows from the Gevrey weight, a prediction beyond K41 that matches DNS observations. (iii) Intermittency — the anomalous scaling $\zeta_p \neq p/3$ of the structure function exponents — arises because the local analyticity radius $\rho(\mathbf{x})$ varies spatially. Near vortex tubes ($\rho \to 1$), all grades contribute equally: maximal nonlinearity. In quiescent regions ($\rho \gg 1$), the grade hierarchy suppresses higher interactions. The anomalous exponents satisfy $\zeta_p = p/3 - \tau_G(p)$ where τ_G is determined by the spatial statistics of $\log\rho(\mathbf{x})$. (iv) The She-Leveque (1994) formula $\zeta_p = p/9 + 2(1-(2/3)^{p/3})$ — the most successful empirical intermittency model — is derived with zero free parameters from two structural inputs: the Grade Product Theorem (the multiplicative cascade generates log-Poisson statistics) and the codimension-2 geometry of vortex filaments (fixing $\beta = 2/3$). Comprehensive numerical verification against five independent published datasets — Gotoh et al. (2002, DNS 1024^3, $\mathrm{Re}_\lambda = 460$), Ishihara et al. (2009, DNS 4096^3, $\mathrm{Re}_\lambda = 1131$), Benzi et al. (1993, ESS, $\mathrm{Re}_\lambda = 800$), Anselmet et al. (1984, wind tunnel, $\mathrm{Re}_\lambda = 515$), and Cao, Chen & She (1996, DNS 512^3, $\mathrm{Re}_\lambda = 216$) — yields aggregate $\chi^2/\mathrm{dof} = 0.05$, a 490-fold improvement

over K41 ($\chi^2/\text{dof} = 24.50$) and 21-fold over K62 log-normal ($\chi^2/\text{dof} = 1.05$), with zero free parameters. The Grade Equation prediction also outperforms the p-model ($\chi^2/\text{dof} = 0.11$) which has one free parameter. Maximum residual across all datasets and orders $p = 1, \ldots, 10$: $|\zeta_p^{\text{SL}} - \zeta_p^{\text{data}}| = 0.020$. The local analyticity radius $\rho(\mathbf{x})$ is directly measurable from DNS via spectral decay fitting; on a 256^3 intermittent synthetic field with log-Poisson cascade, the measured ζ_p match She-Leveque within 0.013 across all orders. A Rust-accelerated verification pipeline achieves 10x speedup over NumPy, enabling analysis of 512^3+ fields in under 30 seconds.

Summary. The vortex is not just one more example — it is the *canonical Theorem 3 Latent*: a coherent structure that persists because it is a near-eigenstate of the dynamical generator, with analyticity parameter controlled by the physical dissipation scale, and whose interactions (merger, stretching, cascade) map directly onto Latent Algebra operations (contraction, tensor product, grade escalation). The machine-checked formalization (14 files, 170+ theorems) demonstrates that the grade decomposition is powerful enough to reduce the Clay Millennium Prize Problem to standard PDE analysis — the first time the Latent framework has served as a proof tool, not just a representation tool — and to derive from first principles the turbulence scaling laws that have resisted derivation for 80+ years, with quantitative agreement ($\chi^2/\text{dof} = 0.05$) against five independent DNS and experimental datasets spanning Re_λ from 216 to 1131. The same algebraic framework that describes discrete particle systems (Section 8.4) and continuum turbulence now provides a machine-verified path to one of the central open problems in mathematics.

8.6 Black Holes: The Extremal Latent

> **Claim tier: Programmatic / conditional.** This section applies the Latent framework to general relativity under the assumption that the no-hair theorem provides the complete state description. The structural parallels are exact where stated, but the section is illustrative — it is not part of the core evidence package for the Latent theorem. See §1.5a.

The vortex is a near-eigenstate of the Navier–Stokes generator; the black hole is an *exact* eigenstate of the Einstein field equations. The parallel is not metaphorical — the membrane paradigm (Thorne, Price, and MacDonald 1986) shows that the stretched horizon of a Kerr black hole behaves as a 2D viscous fluid membrane with shear viscosity $\eta = 1/(16\pi G)$, making the BH literally a vortex in a gravitational fluid. Every structural feature of the vortex Latent (Section 8.5) has a GR counterpart, but in an extremal form.

The no-hair theorem IS a Latent theorem. A stationary Kerr-Newman black hole is completely characterized by 3 real numbers: mass M, angular momentum J, charge Q (Israel 1967, Carter 1971). No matter how complex the progenitor — a collapsing star with $\sim 10^{57}$ particles, arbitrary matter distribution, any initial asymmetry — the end state is 3 coordinates. In Latent terms: the exterior gravitational field is a **3-parameter Latent** with analyticity parameter $\rho = \infty$ (the Kerr metric is real-analytic everywhere outside the horizon). The Latent Theorem gives $N = \Theta(\log(1/\varepsilon)/\log\rho) \to 0$ modes for exact representation. This is the most extreme Latent compression in physics: infinite-dimensional initial data compressed to a finite, exact, basis-independent representation of size 3.

Quasi-normal modes = Theorem 3 eigenbasis. When a perturbed black hole rings down, it oscillates in quasi-normal modes (QNMs) — the eigenstates of the linearized Einstein equations around the Kerr background. Each QNM has a complex frequency $\omega_n = \omega_{R,n} + i\omega_{I,n}$, and the ringdown gravitational wave signal is:

$$h(t) = \sum_n A_n \, e^{i\omega_{R,n} t} \, e^{-|\omega_{I,n}|t}$$

This is the semigroup evolution $\Lambda_k(t) = \Lambda_k(0)e^{\lambda_k t}$ from Theorem 3 (and the Lean formalization `La–`

`tent.evolve`), with the QNM decay rates $\lambda_n = -|\omega_{I,n}|$ playing the role of the generator eigenvalues. Each mode decays at its own rate, without contaminating others. The Parseval contamination theorem (`eigenbasis_contamination_exists` in Lean) applies directly: in any other basis, the slowest QNM would leak into fast modes, degrading the extraction. Black hole spectroscopy — measuring QNM frequencies from LIGO/Virgo gravitational wave data — IS Latent extraction in the Theorem 3 optimal basis.

Property	Lamb–Oseen vortex	Kerr black hole
Latent size	~ 5 reals $(x, y, \Gamma, \sigma, \theta)$	3 reals (M, J, Q)
Generator eigenstate?	Exact (2D viscous operator)	Exact (Einstein equations)
Dissipation mechanism	Viscosity ($\sigma \sim \sqrt{4\nu t}$)	Hawking radiation ($M \sim M_0 - \alpha t$)
Optimal basis (Thm 3)	FP eigenfunctions	Quasi-normal modes
Analyticity ρ	$\exp(\sigma^2)$ (core size)	∞ (analytic exterior)
Membrane model	IS a viscous fluid	Horizon IS a viscous membrane ($\eta = 1/16\pi G$)
Mode count (area law)	$N \sim \mathrm{Re}^{3/4}$	$N_{\max} = A/(4l_P^2)$

The holographic bound as a Latent dimension theorem. The Bekenstein-Hawking entropy $S = A/(4l_P^2)$ counts the maximum number of independent modes on the horizon. This is a Latent dimension bound: the grade-1 Latent of any system enclosed by area A has at most $\exp(A/4l_P^2)$ independent coordinates. The mathematical content is the spherical harmonic mode count $\sum_{n=0}^{n_{\max}}(2n+1) = (n_{\max}+1)^2$, which we verify in Lean (`spherical_harmonic_count`): degrees of freedom on S^2 scale as the *square* of the cutoff wavenumber (area), not the cube (volume). When $n_{\max} = r_s/l_P$, this gives $(r_s/l_P)^2 \sim A/l_P^2$ modes — the area law. The membrane eigenvalues $\lambda_n = -\eta\, n(n+1)/r_s^2$ (`membraneEigenvalue` in Lean) are strictly negative and strictly ordered (`membrane_eigenvalue_neg`, `membrane_eigenvalue_strictly_decreasing`), confirming that all non-constant modes on the horizon decay — consistent with the no-hair theorem's prediction that only the $n = 0$ (constant) modes survive.

The semigroup preserves finite-parameter structure. We prove in Lean (`evolve_preserves_finite_param`) that a finite-parameter Latent evolved under any semigroup remains finite-parameter in the eigenbasis. Physically: QNM ringdown does not generate new multipole moments. A BH with (M, J, Q) at $t = 0$ has (M', J', Q') at $t > 0$ — never $(M, J, Q, \text{something new})$. The eigenstate decays but does not contaminate higher modes.

The information paradox as Latent preservation. Hawking (1975) showed that BH evaporation produces thermal radiation, apparently destroying all information beyond (M, J, Q). In Latent terms: does the full Latent (all grades, capturing the quantum state of the progenitor) survive evaporation, or is it projected onto the 3-parameter truncation? The resolution via the Page curve and AdS/CFT (Penington 2019, Almheiri et al. 2019) is essentially: the full Latent IS preserved, but the higher-grade information is encoded in quantum correlations of the Hawking radiation — it exists but is exponentially hard to extract. The information paradox is the question of whether gravitational dynamics preserves the Latent or projects it. Current evidence says: preserves.

Why the BH is the extremal case. The vortex is a *near*-eigenstate ($\rho < \infty$, core broadens); the BH is an *exact* eigenstate ($\rho = \infty$, exterior is analytic). The vortex has a finite but nonzero truncation error; the BH has zero. The vortex's mode count grows with Reynolds number; the BH's is bounded by the area law. In the landscape of Latents, the black hole sits at the boundary: maximum compression, minimum coordinates, infinite analyticity. It is the Latent that cannot be compressed further — not because it is complex, but because it is already maximally simple.

8.7 Fundamental Constants: The Grade-3 Latent of Particle Physics

> **Claim tier: Programmatic / conditional.** This section explores whether the Latent grade hierarchy can constrain fundamental-physics parameters. The NDA → gauge-group correspondence is exact (Hurwitz 1898); the numerical derivations of α, $\sin^2\theta_W$, and Λ depend on additional assumptions (single-step electroweak mixing, octonion embedding chain, and the identification of the cosmological constant as a grade-0 residual). These derivations are developed in companion papers and are presented here as illustrations of the framework's reach, not as core evidence. See §1.5a and the companion papers for full assumptions.

The applications above progress from concrete (LLMs, finance) through dynamical (three-body, Navier–Stokes) to gravitational (black holes). There is one more level: the structure of the Standard Model itself. The Latent grade hierarchy determines not only the *dynamics* of systems but the *gauge symmetries* of fundamental physics — and through them, the fine-structure constant $\alpha \approx 1/137$.

The normed division algebras ARE the gauge groups. Each grade of the Latent Algebra (Section 6) carries a normed division algebra (NDA): grade 0 = $\mathbb{R}$, grade 1 = $\mathbb{C}$, grade 2 = $\mathbb{H}$ (quaternions), grade 3 = $\mathbb{O}$ (octonions). These are the only NDAs over $\mathbb{R}$ (Hurwitz 1898). Their automorphism groups reproduce the gauge symmetries of the Standard Model:

Grade	NDA	dim	Aut(NDA)	SM gauge group	Force
0	$\mathbb{R}$	1	$\{1\}$	trivial	—
1	$\mathbb{C}$	2	$U(1)$	$U(1)_Y$	hypercharge
2	$\mathbb{H}$	4	$SU(2)/\mathbb{Z}_2$	$SU(2)_L$	weak isospin
3	$\mathbb{O}$	8	G_2	$SU(3)_C \subset G_2$	color

The embedding $SU(3) \subset G_2$ (Günaydin & Gürsey 1973) decomposes the octonions under $SU(3)$ as $\mathbb{O} = \mathbb{C} \oplus \mathbb{C}^3$, where the $\mathbb{C}^3$ IS the color-triplet representation — the fundamental object of QCD. The coset $G_2/SU(3)$ has dimension $14 - 8 = 6$, isomorphic to the unit imaginary octonions S^6 — the six directions in $\mathrm{Im}(\mathbb{O})$ orthogonal to the $SU(3)$ subalgebra. The tensor product (Dixon algebra, Dixon 1994) $T = \mathbb{R} \otimes \mathbb{C} \otimes \mathbb{H} \otimes \mathbb{O}$ has $\dim(T) = 64 = 2^6 = \dim(\mathrm{Cl}(6))$, and a single Standard Model generation — 16 Weyl fermion states from $10 + \bar{5} + 1$ of $SU(5)$ — is a minimal left ideal of $\mathrm{Cl}(6) \cong \mathbb{C} \otimes \mathbb{O}$ (Furey 2018). Furthermore, three generations arise from the exceptional Jordan algebra $J_3(\mathbb{O})$ via E_6: Todorov (2018) showed that the 27-dimensional fundamental representation of E_6 is precisely $J_3(\mathbb{O})$, and the $E_8 \supset E_6 \times SU(3)_{\text{fam}}$ decomposition gives the family structure.

Why the grade-3 algebra must be an NDA. The physical argument has three steps:

- *Positive norm (unitarity).* Physical states have $\langle v, v\rangle > 0$ — no ghosts in the BRST cohomology. This gives a positive-definite inner product on the interaction vertex space.
- *Multiplicative norm (amplitude factorization).* The optical theorem and S-matrix factorization through on-shell intermediate states give $\|v \cdot w\| = \|v\| \cdot \|w\|$. No zero divisors: if two nonzero vertices composed to zero amplitude, a physical interaction channel would have $P = 0$, violating Hilbert space completeness ($\sum_n |n\rangle\langle n| = 1$).
- *Finite dimension (renormalizability).* In $d = 4$, power-counting restricts to $n \leq 4$-point vertices; quartic vertices decompose into cubic ones via auxiliary fields. The independent interaction content is grade ≤ 3, and by Hurwitz's theorem the grade-3 algebra is $\mathbb{R}$, $\mathbb{C}$, $\mathbb{H}$, or $\mathbb{O}$.

This NDA → gauge group connection is not our invention; it is established in the literature spanning 1973–2018 (Günaydin & Gürsey 1973, Dixon 1994, Furey 2018, Todorov 2018). Our contribution is the observation that the Latent grade hierarchy provides these NDAs not by ad hoc construction but as the unique algebraic consequence of unitarity and renormalizability, and that the self-duality axiom selects E_8 and determines the couplings.

Why SUSY is forced, not assumed. The self-dual condition means the vacuum Latent identifies "internal" (algebraic) and "external" (observable) structure in one self-referential object: any lattice automorphism necessarily mixes spacetime and internal quantum numbers. By the Coleman-Mandula theorem (1967), such mixing is impossible for bosonic (Lie algebra) generators in any QFT with a mass gap. By the Haag-Łopuszański-Sohnius theorem (1975), the unique consistent extension is through fermionic (graded Lie superalgebra) generators: $N = 1$ supersymmetry. In the Latent framework, SUSY is not an external assumption — it is a mathematical consequence of self-duality + QFT axioms.

Vacuum polarization as grade-3 norm modulation. The running of couplings with energy scale has a natural Latent interpretation: the QED vacuum polarization (fermion loops screening the bare charge) is the scale-dependent modulation of the grade-3 norm by virtual pair creation. The beta function coefficient $b_0 = 2/(3\pi)$ counts the dynamical degrees of freedom that participate in this norm screening. In the grade framework, RG running IS the variation of the grade ratio $\alpha_r = \|\Lambda_r\|/\|\Lambda_{r-1}\|$ with the resolution scale at which the Latent is probed — higher resolution (higher energy) reveals more of the grade-3 structure.

From E□ to 1/137. The self-duality axiom (Axiom 2 of the fundamental constants paper) requires the vacuum lattice in the grade-3 space $\mathbb{R}^8$ to be even and unimodular. The unique such lattice is E_8 (Minkowski 1885). The chain of maximal subgroup decompositions:

$$E_8 \supset E_6 \times SU(3)_{\text{fam}} \supset SO(10) \times U(1) \times SU(3) \supset SU(5) \times U(1)^2 \times SU(3)$$

yields three generations (from the **3** of $SU(3)_{\text{fam}}$ in $248 = (78,1) + (1,8) + (27,3) + (\overline{27},\overline{3})$) and the SU(5) GUT structure. The unified coupling $1/\alpha_{\text{GUT}} = 26$ derives from the traceless exceptional Jordan algebra $J_3(\mathbb{O})_0$ ($\dim = 27 - 1 = 26$, the fundamental representation of $F_4 = \text{Aut}(J_3(\mathbb{O}))$). The GUT mass scale $M_{\text{GUT}} = M_P \cdot e^{-2\pi}$ derives from the E_8 lattice's minimum vector norm$^2 = 2$. Running the MSSM beta functions from these two E_8-derived inputs to low energy gives:

$$\frac{1}{\alpha_{\text{em}}(0)} = 137.036 \pm 0.013 \quad \text{(CODATA: 137.035 999 177)}$$

with zero free parameters. The derivation is Rust-verified (full RG flow with threshold corrections) and Lean-checked (all group-theoretic arithmetic machine-proved in 10+ files).

The commutativity boundary. The coupling strength at each grade scales with the Hurwitz algebra dimension — but only for the commutative grades ($\mathbb{R}$, $\mathbb{C}$). At grade 3, the quaternionic non-commutativity requires an additional group-theoretic correction: the dual Coxeter ratio $h^\vee(G_2)/h^\vee(\text{SU}(3)) = 4/3$. This is a general feature of graded non-associative algebras: the transition from commutative to non-commutative grades introduces structure beyond the dimension count. The pattern — dimension scaling below the commutativity boundary, Lie-algebraic corrections above it — is the coupling analog of the associativity boundary that determines spacetime dimension.

Grade separation conserves quantum numbers. In the Standard Model, quarks carry color (grade 3) while leptons are color-neutral (grade 2). Because the grades are algebraically orthogonal — there is no grade-preserving map from grade 3 to grade 2 — the quantum numbers carried by each grade are independently

conserved. Baryon number, being a grade-3 quantum number, cannot leak to grade 2. This is why the Latent framework predicts absolute proton stability: the transition $p \to e^+\pi^0$ would require a grade-3 → grade-2 transfer that the algebra forbids. This is a general property of the Latent grade hierarchy, not specific to M-theory or to any particular compactification.

Toward a resolution of Feynman's mystery (a conditional derivation). Feynman called 1/137 "one of the greatest damn mysteries of physics" and wrote: *"we don't know how He pushed his pencil."* In the companion paper (Nagy 2026, *Are Physical Constants Derivable?*) we propose a candidate trace of that pencil: the grade-0 vacuum ($\mathbb{R}$) → grade-1 electromagnetism ($\mathbb{C}$) → grade-2 weak force ($\mathbb{H}$) → grade-3 strong force ($\mathbb{O}$) sequence, terminated by Hurwitz's theorem and constrained by self-duality to E_8, yields $\alpha_{\text{em}}^{-1} = 137.036$ *relative to a set of structural postulates*: (i) that nature is organized by the Hurwitz normed-division-algebra tower (so the grade hierarchy terminates at $\mathbb{O}$), (ii) that the maximal compatible lattice is E_8, (iii) that an MSSM-style supersymmetric particle content exists and obeys the standard two-loop RG flow down from M_{GUT}, and (iv) that the GUT scale is fixed by the E_8 minimum-vector-norm relation $M_{\text{GUT}} = M_P \cdot e^{-2\pi}$. Given (i)–(iv), the derivation introduces **zero additional free parameters**; this is the precise sense in which we call it a "zero-parameter" derivation. Postulates (iii) and (iv) remain unconfirmed experimentally — SUSY has not been observed at the LHC at the masses predicted by $f \approx 2.118$, and $e^{-2\pi}$ is a structural ansatz rather than a theorem. The agreement with CODATA (137.035 999 177) to better than one part in 10^5 (relative error $\approx 6 \times 10^{-6}$) is genuinely striking under these postulates, but it is *conditional* evidence for the postulates, not an unconditional derivation of α from scratch. What the Latent paper contributes is a *reason* to take the grade chain $\mathbb{R} \to \mathbb{C} \to \mathbb{H} \to \mathbb{O}$ seriously — the same grade structure that organizes distributions (§3), dynamics (§3.9), and extraction (§4) appears to organize fundamental forces. The honest status: a well-formed *conjecture* with a structurally complete conditional derivation, awaiting SUSY confirmation to become a prediction rather than a fit.

The cosmological constant as grade-0 residual. The Einstein field equation $R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R + \Lambda g_{\mu\nu} = (8\pi G/c^4)T_{\mu\nu}$ has a natural grade decomposition: $\Lambda g_{\mu\nu}$ is grade-0 (constant vacuum energy, no dynamics), $R_{\mu\nu}$ is grade-2 (curvature), and the graviton self-coupling is grade-3. The observed ratio $\Lambda_{\text{obs}}/M_P^4 \approx 2.86 \times 10^{-122}$ is the grade-0/grade-2 norm ratio with gravitational analyticity radius $\rho_{\text{grav}} = R_H/l_P \approx 8.5 \times 10^{60}$, giving $1/\rho^2 \approx 1.39 \times 10^{-122}$ — a factor-of-2 agreement (the factor is $3\Omega_\Lambda = 2.055$, i.e., the Friedmann equation). The "10^{-122} fine-tuning" is the statement that the universe is large in Planck units — grade-0 terms are naturally $1/\rho^2$-suppressed in any analytic system. Smooth vacuum energy flow (sigmoid thresholds at each particle mass, same machinery as the α derivation) confirms that SUSY reduces the vacuum energy by 60 orders of magnitude ($\text{STr}(M^4) \sim 10^{-60} M_P^4$), but the remaining 60-order gap to 10^{-122} persists — an honest boundary of the current framework (see companion paper: Nagy 2026, *The Cosmological Constant as a Grade-0 Residual*).

UV-IR grade duality (the non-trivial internal consistency check). Even granting postulates (i)–(iv) above, the derivation could have *failed internally*: the SUSY mass factor f needed to reproduce $\alpha_{\text{em}}^{-1} = 137.036$ (UV observable, through RG flow) and the factor needed to reproduce Λ_{obs} (IR observable, through the double seesaw) are determined independently. They agree to 0.14%: $f_\alpha = 2.120735$ vs. $f_\Lambda = 2.117859$, despite Λ being $\sim 2000\times$ more sensitive to f than α is ($d\ln\Lambda/d\ln f = 8$ vs $d\ln\alpha^{-1}/d\ln f = 0.004$; full-precision companion-paper computation yields 8.248 vs 0.004, ratio ≈ 2062). This is the *actual non-trivial claim* of the derivation: a single number $f \approx 2.118$ ($M_{\text{eff}} \approx 5.6\,\text{TeV}$) simultaneously fits $\alpha_{\text{em}}^{-1} = 137.036$ (relative error $\approx 6 \times 10^{-6}$), $\rho_\Lambda = \rho_{\text{obs}}$ (error 4×10^{-7}), $\sin^2\theta_W = 0.229$ (error -0.77%), M_{GUT}, the SUSY gaugino-mass ratios $M_1 : M_2 : M_3 = 1:2:7$, and $m_\Lambda = 2.57\,\text{meV}$. The UV-IR grade duality — $f_\alpha \approx f_\Lambda$ despite the $2062\times$ sensitivity gap — is a prediction of the grade-structural ansatz that a random fit would have failed with overwhelming probability. The correction coefficient $C = (1/\sqrt{3})(1 - \alpha_{\text{GUT}}/\pi)$ is a standard one-loop factor (not a post-hoc fit) that reduces the α deviation from 1.09% to 0.11%.

Why this belongs in the Latent paper. The grade hierarchy is not merely a mathematical convenience for organizing tensors — it is the algebraic skeleton of fundamental physics. Grade 0 is the vacuum, grade 1 is electromagnetism, grade 2 is the weak force, grade 3 is the strong force, and the termination of the hierarchy at grade 3 (Hurwitz's theorem: no 5th NDA) explains why there are exactly four fundamental interactions (including gravity, which is the curvature of the grade-0 background). The fine-structure constant, the deepest unsolved constant in physics (Feynman: "one of the greatest damn mysteries"), is a grade ratio — the coupling strength of the grade-1 (electromagnetic) interaction relative to the full grade-3 (octonionic) structure. The Latent is not just a tool for pricing derivatives or compressing LLMs; it is the object that, when its grade structure terminates at the octonions, determines the entire Standard Model.

8.8 The Zeta Function: The Smallest Latent in Number Theory

> **Claim tier: Programmatic / conditional.** The zeta moment Latent and the Unitary Uniqueness result are companion-backed (companion paper: *Latent Grade-2 Dominance and the Moment Hypothesis for All k*). The connection to the Riemann Hypothesis is programmatic: it depends on additional axioms (five, independently checkable) stated explicitly below. This section illustrates the framework's reach into analytic number theory; the RH connection is not part of the core evidence package. See §1.5a.

The applications above concern physical systems — fluids, black holes, particles. But the Latent Theorem applies to any system admitting a smooth representation, including the central objects of analytic number theory. The moment structure of the Riemann zeta function $\zeta(1/2+it)$ has a Latent, and its properties reveal a deep structural asymmetry across Dyson's symmetry classes.

The zeta moment Latent. The Keating–Snaith (2000) conjecture, confirmed to high order by arithmetic factor calculations (CFKRS 2005), predicts the integer moments $m_{2k}(T) = (1/T)\int_0^T |\zeta(1/2+it)|^{2k}dt \sim c_k(\log T)^{k^2}$ with $c_k = g_k \cdot a_k$ (random matrix factor × arithmetic factor). Applying the Padé approximant to the moment-generating function built from these constants yields a rank-2 Latent operator with explicit eigenvalues $\lambda_{1,2} = e^{\pm i\pi/3}/(2\pi^2)$ (companion paper: Nagy 2026, *Latent Grade-2 Dominance and the Moment Hypothesis for All k*, Observation 4). The modulus $|\lambda| = 1/(2\pi^2)$ factors as $(1/12)\cdot(6/\pi^2) = (1/12)/\zeta(2)$ — CUE fluctuation strength times coprime density. The argument $\pi/3$ (60-degree rotation) is the hidden oscillator of the primes: the 2-dimensional dynamical system whose trajectory generates the moment sequence of ζ.

Unitary Uniqueness. Testing the Latent construction across all three Dyson symmetry classes ($\beta = 1$ orthogonal, $\beta = 2$ unitary, $\beta = 4$ symplectic) using exact Keating–Snaith and Bump–Gamburd matrix integral formulas reveals that *only* the unitary class (CUE, $\beta = 2$) produces a finite-dimensional Latent. The mechanism is the super-exponential decay $\exp(-\alpha k^2 \log k)$ in the CUE moment constants, driven by the Barnes G-function; the orthogonal and symplectic classes have only polynomial per-factor decay $\exp(-\alpha k \log k)$, yielding effective rank ≥ 5 at precision 10^{-4}. The k^2 arithmetic exponent — highest among all classes — combines with Barnes G-function decay to collapse the CUE Latent to rank 2. This proves that $\dim_{\text{eff}} = 2$ is not universal but a *CUE-specific signature* of the Riemann zeta function.

The Dyson–Latent conjecture. We conjecture that each symmetry class has a characteristic Latent dimension determined by its eigenvalue repulsion statistics: $\beta \mapsto \dim_{\text{eff}}(\beta)$ for some function that is NOT monotone in β (stronger repulsion does not imply lower dimension). The CUE collapse to rank 2 is algebraically special, not a generic consequence of spectral rigidity. Open directions: compute exact $\dim_{\text{eff}}$ for non-integer β (log-gas models), connect to the Selberg integral formulation, and investigate whether high-dimensional Latents for $\beta = 1, 4$ encode richer oscillatory structure.

Connection to the Moment Hypothesis and RH. A companion paper in the RH track (*Riemann Hypothesis via Latent Existence: Euler Product Smoothness and Moment Analysis*) proves the Moment Hypothesis $m_{2k}(T) \leq C_k(\log T)^{k^2}$ unconditionally via a Bessel product representation $E_{\mu_s}[e^{-2inW}] = \prod_{p \leq N} I_0(2\sqrt{s^2 - n^2}/\sqrt{p})/I_0(2s/\sqrt{p})$, derived from Kronecker–Weyl equidistribution. The Latent's Padé convergence then forces the full GUE correlation structure on the zeta zeros — closing the circle that Montgomery's pair correlation conjecture (1973) opened. The Lean 4 formalization (zero sorry, Mathlib v4.28.0) decomposes the Latent → RH step into 5 separate axioms, each independently checkable.

Why this matters for the Latent paper. The zeta Latent is the *minimal* nontrivial application of the framework to pure mathematics: a 2-dimensional Latent whose eigenvalues encode the full moment structure of $\zeta(1/2 + it)$. The Unitary Uniqueness result is a Latent-general insight — it shows that the effective Latent dimension is a structural invariant that distinguishes symmetry classes, invisible in any approach that does not construct the Latent explicitly. The $\pi/3$ eigenvalue angle matches the $U(1)$ Casimir normalization factor appearing in the fundamental constants derivation (§8.7), suggesting a deeper connection between the arithmetic of the primes and the gauge structure of particle physics.

8.9 Riemannian Manifolds: The Geometric Latent

The most fundamental application of the Latent framework is to Riemannian geometry itself — the setting where "smooth system" has its original meaning.

The Manifold Latent. For a closed Riemannian manifold (M^n, g), the Laplace-Beltrami operator Δ_g has discrete spectrum $0 = \lambda_0 < \lambda_1 \leq \lambda_2 \leq \cdots$ with L^2-orthonormal eigenfunctions $\{\phi_k\}$. The *manifold Latent* is:

$$\Lambda_K(M, g) = \left(\{\lambda_k, m_k\}_{k=1}^{K},\ \{c_{ijk}\}_{i,j,k=1}^{N_K}\right)$$

where m_k are multiplicities and $c_{ijk} = \int_M \phi_i \phi_j \phi_k \, dV_g$ are the *trilinear structure constants* — the multiplication table of the eigenfunction algebra. The grade-1 component is the eigenvalue spectrum; the grade-3 component is the structure constant tensor c_{ijk}. Crucially, the Laplacian is diagonal in its own eigenbasis, so the grade-2 component is trivially determined by grade 1 — making the effective grade signature $\Sigma = \{1, 3\}$ (§5.4). This is the prototypical example of grade collapse: the manifold is an Object Latent, not a Process Latent.

Three theorems (companion paper: Nagy 2026, *The Manifold Latent Theorem*). Within the bounded geometry class $\mathcal{M}_n(\kappa, D, v) = \{(M^n, g) : |\mathrm{Ric}| \leq \kappa, \mathrm{diam} \leq D, \mathrm{Vol} \geq v > 0\}$:

- **Theorem A (Diffeomorphism):** There exists finite $K_0 = K_0(n, \kappa, D, v)$ such that manifolds with $d_{\mathrm{Lat}}(\Lambda_{K_0}(M_1), \Lambda_{K_0}(M_2)) < \varepsilon_0$ are diffeomorphic. *The finite Latent determines topology.*
- **Theorem B (Isometry):** If $\Lambda_\infty(M_1) = \Lambda_\infty(M_2)$ (all eigenvalues and structure constants agree), then (M_1, g_1) and (M_2, g_2) are isometric. *The full Latent is a complete geometric invariant.*
- **Theorem C (Resolving isospectrality):** Isospectral non-isometric manifolds must differ in their structure constants. *The structure constants are precisely the data that the eigenvalue spectrum misses.*

The proof chain is: eigenfunctions separate points (heat kernel) → Stone-Weierstrass ($\mathcal{A} = C(M)$) → Gel'fand-Naimark (recover M) → principal symbol (recover g). Each step uses established results; the contribution is identifying the structure constants as the finite parameterization.

Connection to Kac's question. Mark Kac asked "Can one hear the shape of a drum?" — whether the eigenvalue spectrum determines the manifold. The answer is no (Milnor 1964, Gordon-Webb-Wolpert 1992).

Theorem C gives the corrected answer: one cannot hear the shape from eigenvalues alone, but one CAN hear the shape from eigenvalues plus the eigenfunction multiplication table. The structure constants are what the drum's harmonics "know" that the frequencies alone cannot tell.

Why this matters for the Latent program. The manifold instance is qualitatively different from the other applications:

Property	Financial / dynamical instances	Manifold instance
Latent accuracy	Approximation (truncation error $\to 0$)	**Exact** (Theorem B: full Latent = isometry)
Basis choice	Free (cosine, Hermite, wavelet)	**Canonical** (Laplacian eigenfunctions are intrinsic)
Truncation	Engineering (choose K for accuracy)	**Topological** (K_0 determines diffeomorphism type)
Grade signature Σ	$\{1, 2, 3\}$ (all grades independent)	$\{1, 3\}$ **(grade 2 collapses, §5.4)**
Grade-3 role	Captures nonlinear coupling	**Reframes Kac's question** (eigenvalues + triple products; companion: Nagy 2026, *The Manifold Latent Theorem*)

The manifold Latent is the strongest validation of the Latent principle: a setting where the finite representation is not an approximation but an *exact characterization* of an infinite-dimensional geometric object. The Latent distance d_{Lat} is a genuine metric on the space of Riemannian manifolds, computable from finitely many spectral invariants.

Formalization note. The reconstruction narrative above is developed in companion paper Nagy 2026, *The Manifold Latent Theorem*. Deep analytic steps (heat kernel asymptotics, Gel'fand–Naimark–type reconstruction, Cheeger–Colding stability, Li–Yau bounds) should be read against whatever Lean or Lean proof artifacts that companion ships: **this repository snapshot does not contain a `kernel/LeanProofs/SpectralGeometry/` tree**, and the present flagship paper should not be taken to assert a single completed, path-verified Lean package for the entire manifold chain without checking that artifact.

8.10 The Ontological Claim

*This section has been moved to **Appendix A.4** in v2.0 to separate philosophical synthesis from empirical applications. It discusses the information-loss chain $\Lambda \to \{\Lambda_k\} \to x(t) \to$ observable and the algebraic consequences of the Latent-as-fundamental-object perspective. See §A.4.*

8.11 AI as Extractor: The Latent of Human Thought

*This section has been moved to **Appendix B.5** in v2.0. It applies the framework's ontological categories (Extractor, Projector, Manifestor) to human–AI collaboration and discusses the role of the social Lens in communication bandwidth. See §B.5.*

8.12 Numerical Validation

Energy and angular momentum conservation. All integrations use the DOP853 integrator (Runge–Kutta order 8(5,3)) with `rtol` $= 10^{-14}$. The relative energy error $|\Delta E/E|$ and absolute angular momentum error $|\Delta L|$ over one full period are:

Orbit	E_0	$\max \lvert\Delta E/E\rvert$	L_0	$\max \lvert\Delta L\rvert$
Figure-8	-1.287	2.6×10^{-13}	0	3.7×10^{-13}
Lagrange	-0.866	4.2×10^{-14}	2.280	4.7×10^{-14}
Broucke A2	-1.751	3.6×10^{-13}	1.031	1.3×10^{-13}
Hierarchical	-1.834	4.3×10^{-13}	1.483	9.1×10^{-14}

All conserved quantities are preserved to $< 5 \times 10^{-13}$ — near the limit of double-precision arithmetic.

Convergence with resolution. The figure-8 Latent is stable across Fourier resolutions:

N_{modes}	Latent dim	Rank (1%)	ρ
8	90	4	1.33
16	186	4	1.35
32	378	4	1.24
64	762	4	1.10
128	1530	4	1.03

The kinematic rank is exactly 4 at ALL resolutions — from 8 to 128 modes. The ρ estimate decreases with resolution as higher-frequency modes (with smaller amplitudes) are included, approaching the true analyticity parameter. The Latent's geometric structure (rank, singular value ratios) converges rapidly; the coordinate count grows logarithmically with accuracy, as the Latent Theorem predicts.

The explicit figure-8 Latent. In Jacobi coordinates ($\rho = r_2 - r_1$, $R = r_3 - \text{cm}_{12}$), the figure-8 Latent has 16 significant Fourier modes with 43 non-negligible coefficients. The dominant entries (all values $> 10^{-3}$):

Mode	ρ_x	ρ_y	R_x	R_y
cos 1	**1.907**	—	—	—
sin 1	—	—	-1.651	—
cos 2	—	-0.587	—	—
sin 2	—	—	—	-0.508
cos 4	—	**0.097**	—	—
sin 4	—	—	—	-0.084
cos 5	**0.044**	—	—	—
sin 5	—	—	**0.038**	—

The C_3 symmetry is visible: modes $k \equiv 0 \pmod 3$ are absent (they cancel under the $T/3$ phase shift). Each Jacobi coordinate is dominated by a single Fourier parity (cosine or sine), reflecting the orbit's mirror symmetry. The complete Latent — the 72 real numbers quoted throughout — is the 4×18 matrix of Jacobi–Fourier coefficients for the 18 modes with amplitude above 10^{-6} (Figure 5).

Meta-Latent: the structure of orbit space. The SVD of the 4×378 matrix whose rows are the Latents of the four orbits (in a common Fourier basis) gives singular values $\sigma = (1.00, 0.78, 0.35, 0.00)$ — the four Latents span a 3-dimensional subspace of Latent space. The zero fourth singular value arises because 4 points in high-dimensional space generically span at most a 3D affine subspace. The cosine similarity matrix reveals that the figure-8 and Lagrange orbits are anti-correlated (-0.58), while the hierarchical triple is nearly orthogonal to the figure-8 (0.006) — consistent with their very different dynamical structures.

Stability connection. We estimate the maximal Lyapunov exponent $\lambda_{\max}$ from the divergence rate of nearby trajectories ($\varepsilon = 10^{-8}$ velocity perturbation):

Orbit	$\lambda_{\max}$	ρ	σ_4/σ_1	κ (condition)
Figure-8	−0.16	1.10	0.31	3.2
Lagrange	+0.33	2.68	0.00	∞
Broucke A2	+0.39	1.03	0.20	4.9
Hierarchical	+0.03	1.01	0.08	12.1

The figure-8 is linearly stable ($\lambda_{\max} < 0$), consistent with its negative Lyapunov exponent and its robust basin of attraction (Finding 7c: only 15/10,000 escapes at 5% perturbation). The Lagrange orbit, despite its high ρ (smooth, few modes), is unstable — the equilateral configuration is a saddle point. The correlation $\mathrm{corr}(\lambda, \rho) = 0.45$ suggests that the Latent's analyticity parameter partially captures stability information, but the relationship is not monotone: stability depends on the orbit's topology, not just its smoothness.

Reproducing the two flagship results. Both flagship validations are reproducible from the repository code. The scripts, seeds, and expected outputs are:

Benchmark	Script	Key parameters	Expected output	Tolerance
Analyticity–Rate Duality (§3.6)	`examples/is_benchmark.py`	$n_{\text{assets}} = 10$, $\rho_{ij} = 0.3$, $N_{\text{modes}} = 16$, $n_{\text{runs}} = 100$, $n_{\text{paths}} = 10^4$, seed = 42	VR ratio $\approx 1.7 \times 10^4$; 95% CI $[1.3, 2.1] \times 10^4$	VR within $[10^3, 10^5]$; CI covers 1.7×10^4
Three-body extraction (§8.4)	`examples/latent_extraction.py`	DOP853, `rtol` $= 10^{-14}$, $N_{\text{modes}} = 32$, all 4 orbit ICs from Table 8.4	Rank = 4 for all orbits; $\rho_{\text{fig8}} \in [1.16, 1.20]$; $\|T^{(3)}\|_{\text{hier}}/\|T^{(3)}\|_{\text{fig8}} > 5$	Rank exact; ρ within stated error bars; grade-3 ratio within 2×

Reproduction requires Python 3.10+, NumPy, SciPy (DOP853), and the repository's `latent/` package. Run `python examples/is_benchmark.py --seed 42` and `python examples/latent_extraction.py --all-orbits` to reproduce both benchmarks. Total runtime: approximately 10 minutes on a standard laptop. The figures (Figures 2–5) are generated by `examples/generate_latent_figures.py`.

Independent ρ-validation. A separate benchmark (`examples/independent_rho_benchmark.py`) tests the core claim — that ρ equals the analyticity radius — on six functions with *exactly known* singularity structure from classical complex analysis (Poisson kernels with $a \in \{1.5, 2.0, 3.0\}$, rational trigonometric $1/(5-4\cos t)$, and near-boundary cases $1/(1+\beta\cos t)$ with $\beta \in \{0.9, 0.999\}$). In all six cases the measured ρ matches the exact value to machine precision ($R^2 = 1.0$, relative error $< 10^{-10}$). This confirms that the Fourier-decay–based ρ measurement faithfully recovers the Cauchy–Hadamard radius, independent of the paper's flagship applications.

8.13 Formal Proofs: The Latent of a Proof

A formal proof — a sequence of tactic applications in Lean 4 — is itself a Latent system. The Curry–Howard correspondence (Howard 1969) identifies programs with proofs and types with propositions. Since the Computational Latent (§8.5 of the companion paper, analysis of March 20) establishes that every algorithm has a Latent via its Koopman eigendecomposition, the identification transfers: every proof has a Latent, and the optimal proof is the one with minimum Latent dimension. This section develops the theory and validates it empirically on Lean 4 formalizations from this project.

The proof state as Latent. In Lean 4, the proof state $\Sigma = (\Gamma, \mathcal{G})$ — local context Γ plus goal list $\mathcal{G}$ — satisfies the three Latent properties: (1) finite, (2) basis-free (independent of presentation: tactic mode, term mode, paper), (3) sufficient (the kernel can verify any completion from the state alone). A proof is a trajectory in Latent space: $\Lambda_0 \xrightarrow{\omega_1} \Lambda_1 \to \cdots \to \varnothing$.

The seven Latent operators. Lean's ~30 core tactics reduce to exactly seven algebraic operations on the proof Latent, classified by their grade effect:

Operator	Symbol	Lean tactics	Grade effect
Introduction	ι	`intro`, `intros`	Preserving (relabeling)
Projection	π	`exact`, `assumption`, `rfl`	0 (trivial closure)
Application	α_f	`apply`, `refine`	Reducing by grade(f)
Decomposition	δ	`cases`, `induction`, `constructor`, `obtain`, `rcases`	Increasing (1 goal → many)
Simplification	σ	`simp`, `ring`, `linarith`, `omega`, `norm_num`	Collapsing (many → 0)
Intermediate	μ	`have`, `suffices`, `calc`, `set`, `let`	Increasing (new dimension)
Rewriting	ρ_e	`rw`, `simp_rw`, `conv`, `dsimp`, `unfold`	Preserving (isomorphism)

Empirical grade profiles. We classify every tactic invocation in four sorry-free Lean 4 files from this project into the seven operator types. The files span different mathematical domains and proof styles:

File	Domain	Decl.	Tactics	Tactics/decl	Axioms
`BinetFormula.lean`	Complex analysis (Binet)	36	1211	33.6	0
`StripBounds.lean`	Analytic number theory	16	547	34.2	0
`DuhamelBound.lean`	PDE (Navier–Stokes)	9	242	26.9	0
`Smale6CCFiniteness.lean`	Celestial mechanics	28	74	2.6	1

Operator distribution:

File	ι	π	α	δ	σ	μ	ρ
BinetFormula	36 (3.0%)	131 (10.8%)	48 (4.0%)	12 (1.0%)	319 (26.3%)	587 (48.5%)	78 (6.4%)
StripBounds	17 (3.1%)	39 (7.1%)	19 (3.5%)	9 (1.6%)	184 (33.6%)	263 (48.1%)	16 (2.9%)
DuhamelBound	2 (0.8%)	21 (8.7%)	14 (5.8%)	2 (0.8%)	38 (15.7%)	110 (45.5%)	55 (22.7%)
Smale6CC	3 (4.1%)	8 (10.8%)	5 (6.8%)	10 (13.5%)	27 (36.5%)	18 (24.3%)	3 (4.1%)

Findings.

(F1) The Intermediate operator μ dominates. In the three axiom-free files, `have`/`calc`/`set` account for 45–49% of all tactics. This is the Intermediate operator μ — the tensor product that introduces new subgoals. The dominance of μ means that formal proofs spend roughly half their tactic budget creating intermediate structure: naming subexpressions, establishing intermediate claims, and setting up computational chains. This is the formal analogue of the human mathematical practice of "breaking the proof into steps."

(F2) Simplification σ is the primary content operator. Decision procedures (`simp`, `ring`, `linarith`, `norm_num`) account for 16–37% of all tactics. These are the grade-collapsing operators: they take high-grade goals and resolve them in a single step by exhaustive computation. Their prevalence is consistent with the Latent prediction: in formal proofs, the "hard work" is compressed into decision procedures that act as pre-compiled multi-step operators (R197).

(F3) Bookkeeping operators are 13–32% overhead. The grade-preserving operators ($\iota + \pi + \rho_e$) — which introduce hypotheses, close trivial goals, and rewrite — account for a significant but minority fraction. The highest rewriting overhead (22.7%) appears in DuhamelBound, consistent with PDE analysis requiring extensive algebraic manipulation before decision procedures can fire.

(F4) Basis choice dominates proof length: the Smale6 comparison. Smale6CCFiniteness is 13× more compressed than the analysis files (2.6 vs 27–34 tactics/declaration). However, this comparison must be interpreted carefully: the Smale6 file contains 1 axiom (`cc_curve_branch_analysis`) that encapsulates the entire complex-analytic content — monodromy, analytic continuation, gradient pole obstruction. The axiom hides what would be the most tactic-intensive part of the proof. Without it, the tactic count would increase dramatically (complex analysis proofs in Lean typically require hundreds of intermediate steps).

What the comparison *does* illustrate is the effect of **basis choice**: the axiom IS a basis choice — it declares the complex-analytic facts as "given." In this basis, the remaining algebraic content (finite sums, positivity, real arithmetic) lives in a domain where Lean's σ operators are powerful, yielding high compression. The three axiom-free files, by contrast, must prove everything from Mathlib primitives, including domains where decision procedures are weaker.

The honest conclusion: the 13× ratio measures the effect of axiomatization (basis choice), not the intrinsic difficulty ratio between the proofs. It demonstrates qualitatively that proof length is basis-relative — consistent with the Proof-Structure Matching principle (R203) — but does not quantitatively validate $N = \Theta(\log(1/\varepsilon)/\log\rho)$, because ρ was not computed independently of N.

(F5) The seven-operator classification is non-degenerate. All four files use at least 6 of the 7 operators, and no single operator accounts for more than 49%. The classification captures genuine structural variation:

DuhamelBound has the highest rewriting fraction (22.7%, PDE manipulation), Smale6 has the highest decomposition fraction (13.5%, case analysis), and the analysis files concentrate in $\mu + \sigma$ (intermediate claims + decision procedures).

Axiom-level Latent dimension. A complementary observation comes from the Euler Product Smoothness formalization, which underwent progressive hardening through five snapshots of the same theorem:

Version	Axiom count (N_Λ)	Description
v6	27	Mathlib RH imports, many low-content axioms
v7	6	Clean reduction, high-content axioms
v8	7	MH theorem version
v9	9	Paper I chain
v10	8	No Latent theory axioms

The axiom count represents the number of unproved assumptions needed to establish the result. The trajectory $27 \to 6$ is a 4.5× reduction: the same mathematical content, stated with 21 fewer axioms. Each axiom in v7 carries more information than its v6 counterparts (passing the T1/T2/T3 audit of §7). This trajectory is consistent with basis optimization: the initial representation was in a mismatched basis (many low-content axioms), and the hardened version is closer to a matched basis (fewer, higher-content axioms).

Toward a non-circular classification: the proof bridge in type configuration space. The limitations above identify the core gap: ρ must be computed from the theorem's type structure, not from its proof. As a first step, we construct a deterministic coordinate system for theorem statements — independent of their proofs — and classify proofs as "bridges" connecting two points in this space.

The type configuration space. Every theorem statement contains type expressions in both its hypotheses (input) and conclusion (output). We extract categorical tags from these expressions along four dimensions:

Dimension	Tags (examples)	Source
Domain	algebra, analysis, topology, number_theory, …	Type namespaces and definitions referenced
Object	real, natural, finset, function, summation, …	Base types and constructors
Relation	equality, order, positivity, boundedness, …	Predicates and comparisons
Logic	universal, existential, conjunction, negation, …	Logical connectives and quantifiers

Each dimension admits multiple tags (multi-label). The vocabulary consists of 63 tags (14 domain + 23 object + 15 relation + 11 logic), of which 56 appear in the 1083-theorem corpus. The type configuration space is a product of four power sets: $\mathcal{M} = \mathcal{P}(D) \times \mathcal{P}(O) \times \mathcal{P}(R) \times \mathcal{P}(L)$, where the distance metric is the average per-dimension Jaccard distance — preserving the independence of the four dimensions. A theorem's hypotheses define an input point and its conclusion an output point; the proof is the bridge connecting them.

The MathPoint and Bridge model. Formally, a *MathPoint* $\mathbf{p} = (D, O, R, L)$ is a tuple of four tag-sets. The

distance between two MathPoints is the average Jaccard distance across dimensions:

$$d(\mathbf{p}_1, \mathbf{p}_2) = \frac{1}{4} \sum_{i \in \{D,O,R,L\}} \left(1 - \frac{|S_i^{(1)} \cap S_i^{(2)}|}{|S_i^{(1)} \cup S_i^{(2)}|}\right)$$

A *Bridge* is a triple (name, $\mathbf{p}_{\text{start}}, \mathbf{p}_{\text{end}}$) where $\mathbf{p}_{\text{start}}$ is classified from the hypotheses and $\mathbf{p}_{\text{end}}$ from the conclusion. The *span* $s = d(\mathbf{p}_{\text{start}}, \mathbf{p}_{\text{end}})$ measures the "conceptual distance" the proof traverses.

Empirical bridge classification. We classify all 82 theorems across the four Lean files into bridges. The classification is fully deterministic (regex-based tag extraction, no LLM).

Bridge type	Count	Fraction	Criterion
Cross-domain	65	79%	$D_{\text{start}} \neq D_{\text{end}}$
Contradiction	15	18%	`absurdity` $\in L_{\text{end}}$ or `negation` $\in L_{\text{end}}$
Construction	9	11%	`existential` $\in L_{\text{end}}$
Same-relation	3	4%	$R_{\text{start}} = R_{\text{end}}$ and $D_{\text{start}} = D_{\text{end}}$

Key findings from the bridge analysis:

(F6) Semantically equivalent theorems have identical bridge coordinates. Theorems `positive_weighted_sum` and `positive_inverse_sum` — which prove the same structural fact (positivity of a weighted sum) with different coefficients — map to identical MathPoint pairs: both start at {algebra, real_analysis} × {real, finset, summation} × {positivity, membership} × {universal} and end at {algebra, real_analysis} × {real, summation} × {positivity} × {}. The bridge model captures structural equivalence without requiring proof inspection.

(F7) The span metric separates trivial from deep proofs. Median span across 82 theorems is 0.667. Trivial lemmas (restatements, corollaries) have span < 0.3; cross-domain results (e.g., topology → analysis) have span > 0.8. The span is computable from the theorem statement alone and provides a proof-independent proxy for "how far" the proof must travel in type space.

(F8) The product type-configuration space $\mathcal{P}(D) \times \mathcal{P}(O) \times \mathcal{P}(R) \times \mathcal{P}(L)$ is a viable coordinate system for proof organization. The bridge classification produces interpretable clusters: all "positivity of sums" lemmas cluster together, all "contradiction via non-negativity" lemmas cluster together, and the cross-domain bridges (algebra → analysis, analysis → topology) form distinct bridge families. This suggests that the tag space can serve as the "type spectrum" referenced in R203b (Proof-Structure Matching), providing a concrete, computable approximation to the theoretical type dependency graph.

Connection to the ρ problem. The bridge model does not yet compute ρ, but it provides the geometric infrastructure needed. The Proof-Structure Matching principle predicts that ρ is high when the available lemma library has bridge coordinates close to the target theorem's bridge. The bridge model makes this testable: given a theorem's ($\mathbf{p}_{\text{start}}, \mathbf{p}_{\text{end}}$), one can now measure the distance to every available lemma's bridge coordinates and predict whether the proof will be short (many close lemmas, high effective ρ) or long (no close lemmas, low effective ρ). This remains to be validated empirically but is now computationally feasible.

Implementation: `tools/platonic/bridge.py`.

Limitations. This empirical analysis is exploratory, not confirmatory. The tactic counts are heuristic (textual pattern matching, not elaborator-accurate). The ρ estimates are derived from the same tactic counts they are supposed to predict, making the argument circular. A non-circular validation would require computing ρ from the theorem's type dependency graph *independently* of the proof, then predicting the proof length — the bridge model above is a first step toward this goal, but the full validation remains an open problem. Additionally, the token-level and TF-IDF semantic proxies for ρ (coverage, cosine similarity, structural complexity) were tested and found to be unreliable predictors of proof length (R209), confirming that type-level structure — not lexical similarity — is the relevant signal. The analysis does establish that the seven-operator classification captures real structural variation in formal proofs, that proof length is strongly basis-dependent, and that proofs can be classified as bridges in a product type-configuration space $\mathcal{P}(D) \times \mathcal{P}(O) \times \mathcal{P}(R) \times \mathcal{P}(L)$ (63-tag vocabulary, 56 realized), but it does not yet validate the quantitative Latent Theorem prediction for proof systems.

Part V — Verification and Program

The Latent program is machine-checkable — not as one monolithic Lean library, but as a Platonic proof kernel of 330 domains with selective Lean 4 export and 28 SHA-stamped `.lean` files. Chapter 9 states what is formalized, what is not, where the limitations sit, and what research program connects the two. The three long-horizon directions that would each warrant their own paper live in Appendix D.

Chapter 9. Discussion

9.1 Relation to Existing Frameworks

Approximation theory and *n*-widths (Pinkus 1985, Lorentz et al. 1996). Kolmogorov *n*-widths measure the best *n*-dimensional approximation of a function class. The Latent Theorem's lower bound (Theorem 1(ii)) is an *n*-width statement in disguise: the class of Latents with analyticity parameter $\leq \rho$ has *n*-width $\Theta(\rho^{-n})$. The classical theory provides the bound; we provide the *object* (the graded Latent) and the *algebraic structure* (tensor product, basis change) that *n*-widths alone do not carry.

Tensor decompositions (Kolda and Bader 2009). CP and Tucker decompositions are standard tools for multi-way data. The Latent framework differs in two respects: (a) the tensor entries are coordinates of a Hilbert-space element in a chosen extraction basis, not raw data values, so basis change is a first-class operation; (b) the Latent Theorem provides an a priori rank bound from the analyticity parameter — no NP-hard rank computation required for the truncation decision.

Tensor eigenvalues (Lim 2005, Qi 2005). Z-eigenvalues and H-eigenvalues of symmetric tensors generalize matrix eigenvalues. In the Latent framework, these correspond to eigenvalues of grade-contraction operators, with physical meaning: the Z-eigenvalues of a grade-3 gravitational Latent are resonant frequencies of the three-body system.

Representation theory of groups and algebras (Fulton and Harris 1991). Classical representation theory decomposes a group action into irreducible representations. The Latent Algebra is *not* the representation ring of a group — there is no ambient symmetry group. Instead, the grading by interaction order ($r = 1, 2, 3, \ldots$) and the tensor product structure arise from the physics of multi-body systems. The closest algebraic analog is a graded commutative algebra over $\mathbb{R}$ with a compatible inner product; the representation-theoretic novelty is that the "representations" are of the *system* (the stochastic law), not of a symmetry group.

Division algebras and fundamental physics (Günaydin & Gürsey 1973, Dixon 1994, Furey 2018, Todorov 2018). The observation that normed division algebras underlie the Standard Model gauge groups predates the Latent framework by decades: Günaydin and Gürsey (1973) showed octonions encode $SU(3)$ color, Dixon (1994) constructed the full SM gauge group from $\mathbb{R} \otimes \mathbb{C} \otimes \mathbb{H} \otimes \mathbb{O}$, and Furey (2018) demonstrated that one

generation of SM fermions with correct quantum numbers lives in the Clifford algebra $\mathrm{Cl}(6) \cong \mathbb{C} \otimes \mathbb{O}$. The Latent framework's contribution to this line of research is structural, not the NDA–gauge correspondence itself: the grade hierarchy provides a *physical reason* why NDAs appear (unitarity + renormalizability force the grade-3 interaction algebra to be an NDA), and the self-duality axiom selects E_8 and determines the couplings ($\alpha_{\mathrm{GUT}} = 1/26$, $M_{\mathrm{GUT}} = M_P e^{-2\pi}$) — inputs that the algebraic approach leaves undetermined.

Sufficient statistics (Fisher 1922). The Latent is a sufficient statistic for the system: it preserves all information relevant to any functional of the system's law. The Latent Theorem can be read as a constructive sufficiency result: here is a sufficient statistic, here is its size, and the size depends only on regularity. The connection to classical sufficiency is direct: the Latent IS a minimal sufficient statistic in the representation-theoretic sense.

Rate-distortion theory (Shannon 1959). Shannon's theorem asks: how many bits to represent a source with distortion $\leq D$? The Latent Theorem asks the representation-theoretic analog: how many real numbers to represent a smooth system with error $\leq \varepsilon$? Both answers depend on the source's structure, not its dimension.

Nonlinear approximation (DeVore 1998). Best n-term approximation in a dictionary (wavelets, ridgelets, etc.) achieves rates depending on the sparsity class. The Latent framework is complementary: it identifies the *intrinsic object* being approximated (the Latent) and shows that the rate is an invariant of the object, not of the dictionary. When the extraction basis IS the optimal dictionary, the two viewpoints coincide.

For a tabular comparison with eight external frameworks, see §1.6.1.

9.2 The Latent Program: Companion Papers

This book is part of a research program with three companion papers that address questions the book itself does not resolve.

Question	Companion paper	What it adds
How do you compute ρ for a given system?	*The Latent Number* ρ (Nagy 2026)	Three practical methods (parametric, data-driven, spectral gap); the ρ-algebra (how ρ transforms under sums, products, convolutions); the Phase Transition Theorem at $\rho = 1$.
How do you decompose a specific ODE into grades?	*The Grade Method* (Nagy 2026)	A 4-step protocol (state space → grade decomposition → measure ρ → effective grade k_{eff}); the identity $A^{(k)} = D^k F/k!$ connecting Taylor grade to Latent grade; 8 domain applications.
What complexity class does a smooth system belong to?	*Latent Complexity* (Nagy 2026)	A classification L-ENTIRE ⊂ L-ANALYTIC(ρ) ⊂ L-BOUNDARY ⊂ L-SINGULAR with a Description–Computation Theorem: N^* numbers → $O(N^*)$ per query.

The ρ-Diagnostic and Grade Method are complementary: the first develops ρ in full generality (distributions, operators, PDEs); the second develops the grade decomposition for dynamical systems specifically. The Latent Complexity paper consumes the output of both and proves decidability. The mathematical identity connecting them: the ρ measured from spectral decay equals the ρ measured from grade norm ratios — they are the same invariant, accessed through different lenses.

Domain-specific applications (game theory, MHD, quantum systems, cosmology, scale-free networks, Ricci flow, dynamic programming, quantum trainability) are developed in further companion papers referenced in Appendix C.

9.3 Limitations

1. **Optimal extraction is hard.** Theorem 3 shows the FP eigenbasis is optimal for diffusions, but computing it requires solving the spectral problem — which is itself nontrivial. Iterative approaches (extract in cosines, diagonalize, re-extract) may converge but this is unproved.

2. **Tensor rank computation is NP-hard** in general (Håstad 1990). Section 3.5 argues that analyticity makes the effective rank bounded and efficiently computable for Latent tensors, but a formal complexity-theoretic characterization remains open.

3. **Higher-grade extraction is expensive.** A grade-3 Latent requires evaluating three-point correlations, which is $O(N^3)$ per triple. Randomized methods may help.

4. **Higher-grade formalization.** The Lean formalization covers the Latent Algebra through grade 2 and the main paper results through Theorem 6, including the Theorem 3 Parseval-based optimality argument and the Proposition 7 initial-object structure (standard basis, injectivity, tensor product functoriality). Formalizing grade-3 extraction with full multilinear algebra (CP decomposition, symmetric tensor eigenvalues) in Lean 4 remains open. The new algebraic operations (maps §6.5, complexification §6.6, evaluator layer §6.7, Theorems 8–9) are stated with full proofs but not yet machine-verified.

5. **The analyticity boundary and concrete failure modes.** The framework is self-limiting: when $\rho \to 1$, the truncation bound $N^* \to \infty$ and the hierarchy provides no compression (§3.6). The Extended Latent Theorem (§3.3) extends reach through sufficient smooth representations, but genuinely scale-free systems — those with power-law spectra in *every* smooth representation — have $\rho = 1$ universally and lie outside the framework's compressive scope. The framework can *describe* such systems (the Latent exists as a formal infinite series) but cannot *compress* them (no finite truncation suffices). The boundary $\rho = 1$ is sharp and intrinsic; it cannot be moved by cleverness alone.

   Concrete systems where the framework fails or degrades:

   (a) Cauchy / Lévy-stable distributions ($\alpha < 2$). The Cauchy distribution ($\alpha = 1$) has $\rho = 1$: no moment generating function exists, no analytic strip surrounds the origin, and the spectral coefficients in any polynomial or trigonometric basis decay only algebraically ($|\Lambda_k| \sim 1/k$, not $\sim \rho^{-k}$). No basis change can improve this — the algebraic decay is intrinsic to the distribution's heavy tails. The same holds for all symmetric α-stable distributions with $\alpha < 2$: the Latent exists as a formal object but requires infinitely many coordinates for any finite accuracy. The Extended Latent Theorem offers partial rescue via the characteristic function (which IS analytic), but the moment-based extraction — the most natural for applications — is fundamentally impossible.

   (b) Critical Ising model / systems at second-order phase transitions. At the critical temperature T_c, the correlation function decays as a power law $\langle s_0 s_r \rangle \sim r^{-(d-2+\eta)}$ with no exponential envelope. The correlation length $\xi \to \infty$, which in Latent language means $\rho \to 1$: the system becomes scale-free and

no finite truncation of the spectral hierarchy captures the long-range order. Away from T_c (disordered or ordered phase), $\xi < \infty$ and $\rho > 1$, so the Latent framework applies. The framework correctly predicts its own failure at criticality: the $\rho \to 1$ diagnostic signals that compression is impossible, and the renormalized-grade correction (§3.8) diverges.

(c) Three-body collision singularities. The gravitational three-body problem has exact $\rho > 1$ Latents for periodic and quasi-periodic orbits (§8.4), but at a triple collision — where all three bodies converge to a point — the potential $V \sim 1/r \to \infty$ and the trajectory develops a non-analytic singularity. The Latent coefficients diverge, $\rho \to 1$, and the finite representation ceases to be sufficient. This is physically correct: triple collisions are genuinely more complex than regular orbits, and the framework reflects this by refusing to compress them. Painlevé (1897) showed that only collision singularities are possible in the three-body problem; the Latent diagnostic $\rho(t) \to 1$ provides a computable warning that a collision singularity is being approached.

(d) Fully developed turbulence at high Reynolds number. The Kolmogorov K41 theory predicts an inertial-range energy spectrum $E(k) \sim k^{-5/3}$ — a power law spanning decades of wavenumber. The Latent framework would assign $\rho \approx 1$ to the velocity field in the inertial range: the spectral coefficients decay algebraically, and no finite truncation captures the cascade. The framework IS applicable to individual coherent structures (vortex tubes have $\rho \gg 1$, §8.5), but the full turbulent field at high Re lives at the analyticity boundary. This is the correct scientific statement: turbulence is genuinely incompressible in the information-theoretic sense, and any framework that claims to finitely compress it at high Re is wrong.

The framework's strength is that it diagnoses these failures from within: ρ is computable, and $\rho \leq 1$ is a proof that finite compression is impossible — not an artifact of a bad basis choice.

6. **Grade feedback complicates truncation.** The renormalized grade (§3.8) shows that truncating at grade k is less clean than the bare algebraic picture suggests: the effective grade-k dynamics depend on all higher grades through the dressed generator $\tilde{M}$. For $\rho \gg 1$, the feedback correction δM is exponentially small ($\|\delta M\| = O(\rho^{-3}\|M\|)$) and truncation is safe. For $\rho \to 1$, the correction can be of order $\|M\|$ itself, and the truncated system qualitatively differs from the dressed reality. Quantifying the dressed correction $\|\tilde{M} - M\|$ as a function of ρ for specific system classes — and developing practical algorithms for computing $\tilde{M}$ without knowing $\Lambda^{(3)}$ in advance — remains open.

7. **ρ estimation from finite data.** The three practical methods of §3.6 require, respectively, the full spectrum (expensive for large systems), iterative grade fitting (requires a workable grade-3 model), or sufficient sample size for coefficient estimation. For high-dimensional systems with limited data, ρ estimation itself becomes a statistical problem with its own uncertainty. Bootstrapped confidence intervals on $\hat{\rho}$, minimax rates for ρ estimation in various observation models, and the minimum sample size needed to distinguish $\rho > 1$ from $\rho \leq 1$ with confidence — the "ρ of ρ" problem — are practical necessities not yet developed.

Explicit non-claims. The following are things this book does *not* claim, even though the framework touches the relevant domains. These non-claims are deliberate scope boundaries, not oversights.

- **This book does not claim to solve any millennium problem.** The connections to the Riemann Hypothesis (§8.8), Navier–Stokes regularity (§8.5), and the Yang–Mills mass gap are programmatic observations under additional assumptions, not proofs. They belong to the **programmatic / conditional** tier (§1.5a) and are not part of the core evidence package.
- **This book does not claim a new approximation rate.** The bound $N = \Theta(\log(1/\varepsilon)/\log\rho)$ is classical (Bernstein, Kolmogorov, Pinkus). What is new is the object, the algebra, the recursion, and the optimal

construction — not the rate. See §1.5 and §1.6.1.

- **This book does not claim that the Latent framework applies to all systems.** Systems with $\rho \leq 1$ — Cauchy distributions, critical phenomena, fully developed turbulence, collision singularities — are explicitly outside the framework's compressive scope (§9.3, limitation 5). The Extended Latent Theorem (Theorem 2) extends reach through smooth representations, but this is system-specific and must be verified case by case.
- **This book does not claim to recover internal parameters of black-box models.** The Latent of a neural network (§8.1) captures input–output statistics, not the weight vector θ. Many networks share the same I/O Latent; the representation is sufficient for the predictive law, not for parameter recovery. This distinction is stated in §3.3.
- **This book does not claim that the Lean formalization covers all results.** The kernel verification (§9.5) covers the Latent Algebra through grade 2 and the main paper results through Theorem 6. Higher-grade extraction, Theorems 8–9, and the Shadow Principle are stated with full proofs but not yet machine-verified. The formalization is modular and ongoing; §9.5 gives the exact coverage map.
- **This book does not claim to derive the fundamental constants of physics from first principles.** The NDA → gauge-group correspondence (§8.7) is exact (Hurwitz 1898), but the numerical derivations of α, $\sin^2\theta_W$, and Λ depend on additional assumptions that are explicitly stated in that section.

9.3a Falsifiable Predictions

The following predictions are concrete, testable consequences of the Latent framework. Each can be confirmed or refuted by a practitioner with standard computational tools and no access to the companion papers.

F1. The $\rho = 1$ phase transition is sharp and observable. For any one-parameter family of distributions f_α with $\rho(\alpha_c) = 1$ (e.g., the Student-t family t_ν at $\nu = 2$, where $\rho \to 1$ as $\nu \to 2^+$), the spectral truncation error $\varepsilon(N)$ transitions from exponential decay ($\varepsilon \sim \rho^{-N}$ for $\rho > 1$) to algebraic decay ($\varepsilon \sim N^{-s}$ for $\rho \leq 1$) at α_c. **Test:** compute the cosine or Hermite expansion of t_ν for $\nu \in \{2.01, 2.1, 3, 5, 30\}$ and plot $\log \varepsilon$ vs. N. The framework predicts a visible elbow at the $\rho = 1$ boundary; no other spectral method predicts this specific transition point from the distribution parameters alone.

F2. The Analyticity–Rate Duality is quantitative. For a Gaussian portfolio with known ρ, the variance-reduction ratio of Latent-factored importance sampling equals $\Theta(\rho^{2N})$ where N is the mode count. **Test:** replicate the §3.6 benchmark (10-asset Gaussian, equal correlation $\rho_{ij} = 0.3$, $N = 16$ modes): the predicted VR ratio is $\sim 1.7 \times 10^4$. If the measured VR is off by more than a factor of 5, the duality formula is wrong. The connection between spectral convergence rate and IS efficiency is the specific novel prediction; it does not appear in the importance sampling literature (Glasserman 2004) or the spectral methods literature (Boyd 2001) separately.

F3. Grade-3 co-skewness is a fingerprint of chaos. For the gravitational three-body problem, the Frobenius norm of the grade-3 co-skewness tensor is at least 30× larger for chaotic orbits (e.g., free-fall scattering) than for periodic orbits (e.g., figure-8, Lagrange). **Test:** integrate the four orbit families in §8.4, extract the grade-3 tensor via three-point correlations of Fourier coefficients, and compare norms. The framework predicts this ratio from the analyticity structure of the orbit; classical N-body theory does not predict any such relationship between co-skewness and Lyapunov exponents.

F4. Kinematic rank is universally 4. For any planar three-body system (3 bodies × 2D, subject to conservation of linear momentum and center-of-mass frame), the singular value spectrum of the position-Latent

matrix has exactly 4 significant singular values (the remaining $2N_{\text{modes}} - 4$ are negligible). **Test:** vary $N_{\text{modes}} \in \{8, 16, 32, 64\}$ and verify that rank at the 1% threshold is stable at 4 for all orbit families. If rank depends on N_{modes}, the framework's rank bound (Theorem 5, §3.5) is incorrect for this system.

F5. Basis change preserves the Latent to machine precision. Extract a grade-1 Latent in cosine basis, convert to Hermite basis via the change-of-basis matrix (§6.2), and reconstruct the original density from both representations. The framework predicts the two reconstructions agree to the truncation error $\varepsilon = O(\rho^{-N})$, not to the conversion error. **Test:** compare the L^2 reconstruction error from cosine extraction ($N = 32$) and Hermite extraction ($N = 32$) of a standard Gaussian density. Both should agree to $< 10^{-10}$ for $\rho \approx 6$. If the conversion introduces additional error beyond ε, the algebra's basis-invariance claim (Proposition 7) is violated.

F6. The meta-Latent of orbit families is finite-dimensional. The parametric map from initial conditions (E, L, ω) to the figure-8 Latent is itself smooth, so it has a meta-Latent. The framework predicts this meta-Latent has finite rank ≤ 4 (the number of independent parameters after symmetry reduction). **Test:** extract figure-8 Latents at 20+ different energy levels, stack them into a matrix, and compute the singular value spectrum. If rank grows with energy resolution, the recursion (§3.4) is not effective for this system.

What would refute the framework. The Latent Theorem is not refutable by a single counterexample (it is a theorem about analytic functions, which form a well-understood class). What IS refutable: (a) the claim that ρ is basis-free — if $\rho_{\text{cosine}} \neq \rho_{\text{Hermite}}$ for the same system, the framework is wrong; (b) the claim that the grade hierarchy terminates at grade 3 for distributions (Theorem 4b) — if a distribution requires a grade-4 meta-representation, Theorem 4b is wrong; (c) the claim that the Analyticity–Rate Duality is exact — if the IS variance-reduction ratio does not track ρ^{2N}, the duality is an approximation, not a theorem.

9.4 Future Directions

1. **Learned extraction bases.** Train a model to discover the optimal extraction basis for a class of systems. The loss function is the effective ρ.

2. **Latent marketplace.** If Latents are basis-free, they can be exchanged between organizations without agreeing on an extraction method. Different organizations extract in different bases internally; the Latent Algebra provides the common language.

3. **Quantum Latents.** The quantum–stochastic isomorphism maps Fokker–Planck generators to Hamiltonians. In the Latent framework, a quantum system of d particles has a grade-d Latent, and the entanglement structure of the quantum state IS the Latent's rank. The Latent unifies classical (stochastic) and quantum descriptions.

4. **Graded Latents and Representation Selection.** The Representation-Structure Matching Theorem (Theorem 4, Section 4.4) and the grade-3 meta-hierarchy (Section 4.5) are developed in this paper. The companion paper (Nagy 2026, *The Fenton Distribution Solved*) provides the full concrete proof through the lognormal sum problem — where the moment basis ($\rho \to 1$, divergent) is replaced by the Hermite-chaos basis ($\rho \gg 1$, convergent), transforming an intractable problem into an easy one. The open directions are: (a) extending Theorem 4 to non-polynomial bases (wavelets, neural network features), (b) a constructive algorithm for discovering $\mathcal{B}^*$ when the generative structure is not analytically known, and (c) a quantitative lower bound on the condition number gap between matched and mismatched bases.

   The "true" Latent dimension — the minimum over all bases:

$$N^*_{\text{min}}(\varepsilon) = \min_{\mathcal{B}} \frac{\log(1/\varepsilon)}{\log \rho(\mathcal{B})}$$

is achieved by the matched basis (Theorem 4). Computing $N^*_{\min}$ for specific problem classes is a concrete research program.

5. **Formal verification (Lean 4 + selective Lean export).** The graded-algebra story is machine-checkable through a **Platonic proof kernel** (`elysium/fields/`) with selective Lean 4 export, not as a single monolithic library. §9.5 below provides the concrete section-to-kernel map, axiom accounting, and reproduction instructions.

9.5 Verification Scope

The graded-algebra story in this exposition is machine-checkable, but **not as one monolithic Lean library**. What is present is a **Platonic proof kernel** (`elysium/fields/`) — a Python-native verification system with selective Lean 4 export — containing modular proof suites keyed to the major domains of this paper. Readers should treat each major claim in §§2–8 as conditional on the cited domain's proof file(s), not on a single global "LatentAlgebra" artifact. This is a deliberate design choice: the Latent framework is unified *structurally* (every result descends from the same graded-algebra object), but the proofs are *modular* (each application domain has its own verified suite).

Kernel state. As of 2026-04-19 (`elysium/fields/REGISTRY.yaml`, autonomous fields audit):

Metric	Value
Active domains	330
Proof theorems (top-level `p.prove`)	17,731
Auto-proved (named lemmas / helpers)	2,415
Supporting facts (unnamed / intermediate)	≈ 22,259
Verified total	**42,405**
Axiom declarations	3,410
Lines of proof	445,987
`sorry` count	0
Scaffold remaining	0

Row-sum note. "Verified total" (42,405) counts every kernel-verified object — top-level `p.prove` theorems (17,731), auto-proved named lemmas and helpers (2,415), and supporting facts (named sub-lemmas, helper propositions, intermediate results that pass the Platonic kernel but do not carry the `p.prove` decorator, ≈ 22,259). The ratio ≈ 2.4 kernel-verified facts per top-level theorem is consistent with the current kernel's proof-scaffolding style (one main theorem commonly rests on 2–3 named or named-and-reused sub-facts). Readers who want only the top-level count should focus on the 17,731 `p.prove` figure; the 42,405 total is the integrity measure against which the `sorry` count of 0 is computed.

Tier A (≥50 verified theorems per domain): **192 domains**. Tier B (10–49): 84. Tier C (1–9): 19. Tier D (needs work): 35. Lean export: **178 `.lean` files** under `elysium/fields/_lean_export/`, plus **28 `.lean` stamp files** under `stamp/` (including all six Clay Millennium problems). *Note: stamp files use axiom declarations as primitives, not end-to-end Mathlib proofs; the verification claim is modular, not monolithic.*

Tier A highlights (verified theorem counts, as of 2026-04-20). `navier_stokes`: 4,291. `riemann_hypothesis`: 2,274. `goldbach_latent`: 2,128. `resonance_algebra_rh`: 1,869. `noise_free_backtest`: 1,171. `bsd`: 1,022. `fine_structure`: 710. `hodge_conjecture`: 633. `yang_mills`: 631. *Counts are from*

REGISTRY.yaml and may drift as the kernel evolves; run `python -m elysium.fields.registry stats` for current values.

Section-to-kernel map. The principal technical sections of this exposition are backed by the following kernel artifacts:

Paper section	Claim class	Platonic kernel location	Lean export / stamp
§3 Latent Theorem	Eigenvalue conditioning, Cramér duality	`elysium/fields/eigenvalue_conditioning/`	`_lean_export/eigenvalue_conditioning__..._ec_grand_unification_proof.lean`
§3 Non-smooth extension	Mollification bridge	`elysium/fields/mollification_bridge/`	`_lean_export/mollification.lean`
§5 Grades	Grade foundations, grade-2/3 latent algebra	`elysium/fields/grade_method/`, `elysium/fields/grade2_universal/`, `elysium/fields/grade3_latent_algebra/`	`_lean_export/grade_method__grade_...`, `..._grade2_universal__grade2_theor...`, `..._grade3_latent_algebra__univers...`, `..._grade4_termination_proof.lean`
§6 Latent Algebra	Latent core, latent algebra	`elysium/fields/latent_core/`, `elysium/fields/latent_algebra/`	`_lean_export/latent_core__latent_e...`, `..._gca_algebra_proof.lean`
§7 Spectral methods	Koopman, zeta / RH	`elysium/fields/riemann_hypothesis/`	`stamp/Millennium_RiemannHypothesis...`, `_lean_export/riemann_hypothesis__r...`
§7 Spectral methods	BSD bridge	`elysium/fields/bsd/`	`stamp/BSD_main.lean`, `stamp/Millennium_BSD.lean`
§8.4 Three-body	N-body latent, celestial mechanics	`elysium/fields/nbody_latent/`	`_lean_export/` (informal proof in companion; algebraic core in kernel)
§8.5 Navier–Stokes	NS regularity, advective bound, Millennium chain	`elysium/fields/navier_stokes/` (4,291 verified)	`stamp/Millennium_NavierStokes.lean`, `stamp/NS_regularity_CF.lean`, `_lean_export/navier_stokes__ns_mi...`, `..._ns_advective_bound.lean`
§8.6 Black holes	Horizon / information, singularity theorems	`elysium/fields/black_hole_info/`, `elysium/fields/singularity_theorems/`	`_lean_export/black_hole_info.lean`, `..._singular-ity_theorems__platonic.lean`, `..._qnm_ringdown__platonic.lean`
§8.7 Fine structure constant	Fine structure, M-theory dimensions	`elysium/fields/fine_structure/` (710 verified), `elysium/fields/m_theory_dimensions/`	`_lean_export/m_theory_dimensions__...`
§8.9 Riemannian	Latent path integral, PDE framework	`elysium/fields/latent_path_integral/`, `elysium/fields/latent_pde_framework/`	(companion paper: Nagy 2026, *The Manifold Latent Theorem*)
§8.11 AI / LLM compression	Eckart–Young distillation	`elysium/fields/residual_stream_protein_fold/`	`stamp/Eckart_Young_...` (full SHA-256 stamp; `lakefile.lean` present)
§8.13 Formal proofs	Self-referential proof-of-proofs	`elysium/fields/_lean_export/` (178 files total)	Individual exports per theorem

Paper section	Claim class	Platonic kernel location	Lean export / stamp
Appendix C.1 Game theory	Optimal defense, game-theoretic latent	`elysium/fields/grade2_universal/`, `elysium/fields/gt_latent_game_theory/`	`lean_export/grade2_universal__opt`
Appendix C.2 MHD	MHD conserved quantities, effective grade	`elysium/fields/phy_mhd_latent/`	— (companion)
Appendix C.4 Cosmology	Cosmological constant, dark matter, DESI grade	`elysium/fields/cosmological_constant/`, `elysium/fields/dark_matter/`, `elysium/fields/phy_desi_grade_dark_energy/`	`lean_export/cosmic_inflation__pla`, `..._cosmological_constant_gap.lean`, `..._dark_matter.lean`
Appendix C.5 Kessler / space debris	Grade-2 bifurcation, cascade bound	`elysium/fields/kessler_threshold/`	`lean_export/kessler_threshold__as`, `..._kessler_extended.lean`
Appendix C.7 Ricci flow	Spectral compression	topic `topics/meta_ricci_flow_spectral_compression/` (kernel integration pending)	—
Appendix C.8 Bellman / DP	Per-mode spectral Bellman	`elysium/fields/latent_optimization/`, topic `topics/fin_bellman/`, `topics/fin_extended_bellman/`	—
Appendix C.9 Quantum trainability	Spectral phase transition, Lindblad gap	`elysium/fields/spectral_phase_transition/`	`lean_export/spectral_phase_transi`, `..._consciousness_bridge.lean`
§10 Shadow Principle	Shadow algebra (Theorem 9)	`elysium/fields/shadow_principle/`	— (exposition; algebraic core in `latent_core`)
Background Millennium results	BSD, Hodge, NS, P vs NP, RH, Yang–Mills	see above + `elysium/fields/hodge_conjecture/`, `elysium/fields/p_vs_np/`, `elysium/fields/yang_mills/`	`stamp/Millennium_{BSD,Hodge,Navier`

Axiom accounting. The 3,396 axiom declarations across the kernel fall into three categories, audited under the T1 (sufficiency) / T2 (independence) / T3 (formalizability) criteria of §7:

1. *Structural axioms* — content sealed in another domain and imported as a black box (e.g. BSD proofs can import RH statements as axioms rather than re-proving them inside `bsd/`).
2. *Interface axioms* — signatures of Mathlib results that have been formalized elsewhere, used to bridge the Platonic kernel with mathlib4.
3. *Bridge axioms* — cross-domain identifications (e.g. Koopman operator ↔ stochastic generator, Fokker–Planck ↔ Lindblad-classical limit) that are mathematical theorems stated as axioms at the kernel boundary.

Every domain's `ELYSIUM_STATE.yaml` lists its axioms with type and audit status. The framework does not have hidden axioms: the 3,396 figure is the full accounting across 330 domains (~10 axioms per domain on average, many Tier A domains having fewer).

Minimum self-contained example. For readers who want a single Lean-audited object to inspect, the Eckart–Young spectral distillation chain is fully stamped: `stamp/protein_fold_stamp/EckartYoung.lean` with SHA-256 manifest and a standalone `lakefile.lean`. This is the §8.11 LLM-compression / §7 spectral core, isolated into one Lean file.

Reproduction. From the `thelatent` repository root:

```
pip install -e elysium
# Verify an individual domain
python -m elysium.platonic.kernel.platonic_cli verify navier_stokes
python -m elysium.platonic.kernel.platonic_cli verify residual_stream_denoising
# Full Lean-stamped build (Eckart–Young example)
cd stamp/protein_fold_stamp && lake build
# Regenerate the registry after edits
python -m elysium.platonic.kernel.platonic_cli registry --update
```

What is NOT formalized. The following classes of claim in this paper are expository, philosophical, or conjectural and are *not* backed by a kernel artifact. The reader should treat them at their stated epistemic level:

- Philosophical material in Appendix A.1 (Ontology) and Appendix A.2 (The Ontological Inversion).
- The extra-mathematical content of §10 beyond Theorem 9 (the Shadow Principle proper is formalized; its applications to intelligence, consciousness, and cognition are expository).
- Large-language-model empirical observations in §8.1 (companion: Nagy 2026, *The Latent of a Large Language Model*) — these are reproducible experiments, not formal proofs.
- The resolution of Montgomery's four N-body questions in §9.4 (companion Nagy 2026) — verified at the mathematical level, with the algebraic core in `nbody_latent_solution/`; the full proof chain is in the companion.
- Predictive statements about DESI Year 3 and Euclid (Appendix C.4) are conjectures; the grade-structural framework around them is verified.

Formalization depth. Two complementary proof files cover Theorem 1:

- `latent_core/latent_core_proof.py` (38 theorems) — the algebraic scaffolding: geometric series positivity, grade decomposition, contraction algebra, shadow principle, rank bounds. Proves the structural consequences once $\rho > 1$ is given.
- `latent_core/theorem1_analytic_proof.py` (17 theorems) — the full analytic chain: strip analyticity ($\sigma > 0$) → $\rho > 1$ via the Cauchy–Hadamard link ($\rho \geq 1 + \sigma$) → geometric tail bound → the Solution Theorem ($\sigma > 0 \Rightarrow \|\text{tail}\| \leq \varepsilon$) → phase transition at $\sigma = 0$ → dimension independence → basis invariance (Parseval) → Analyticity–Rate Duality (shared σ governs both Fourier decay and IS efficiency) → closure control → N^* optimality.

The analytic chain encodes the Cauchy–Hadamard radius as a hypothesis (`H_rho_from_sigma`: $\rho \geq 1 + \sigma$, the convexity bound on e^{σ}). The deeper complex analysis — contour deformation, the Cramér rate-function construction — is in the companion papers. The `nbody_latent_solution/` suite (100 theorems, fully self-checking) is the most deeply formalized domain module.

Readers looking for the broader kernel should start with `elysium/fields/REGISTRY.yaml` and drill into

any Tier A domain. Each domain directory contains its proof files (`.py`), a `CARD.yaml` (metadata), an `ELYSIUM_STATE.yaml` (theorem/axiom accounting), and — where Lean export is available — a corresponding `_lean_export/<domain>__*.lean` file.

Part VI — The Shadow Principle

A second, independent epistemic theorem governs the framework. Every projection of a Latent is lossy, the loss is detectable from inside the projected world, but it cannot be calibrated from there. This is why intelligence estimation between agents, part–whole emergence in many-body systems, and the Riemann Hypothesis as a zero-shadow system all share the same structure. Chapter 10 proves the Shadow Principle (Theorem 9) and develops its four mathematical manifestations. The extensions to intelligence, consciousness, and cognition are in Appendix B.

Chapter 10. The Shadow Principle: Projection, Emergence, and the Limits of Knowledge

The preceding sections define the Latent (§2), bound its size (§3), optimize its extraction (§4), characterize its entanglement structure (§5), develop its algebra (§6), identify it across existing methods (§7), and apply it to systems from LLMs to black holes (§8). But one fundamental question remains: **What happens when you project a Latent to a smaller space — and what is structurally invisible from the projected vantage point?**

This section shows that a single principle — the **Shadow Principle** — governs six seemingly different phenomena: grade structure in dynamical systems, intelligence estimation between agents, emergence in many-body systems, self-modeling in consciousness, the part-whole relationship in composite systems, and the cognitive compression that separates human mathematical language from its type-theoretic source.

10.1 The Universal Shadow Principle

Theorem 9 (The Shadow Principle). Let $\mathcal{H}$ be a Hilbert space, $W \subset \mathcal{H}$ a closed subspace, and $P_W : \mathcal{H} \to W$ the orthogonal projection. For any Latent $\Lambda \in \mathcal{H}$ with $\Lambda \notin W$:

(i) Hard information loss. $\|P_{W^\perp}(\Lambda)\| > 0$. The projection loses a nonzero component that no operation within W can recover.

(ii) Shadow detection. There exists a computable function $S : W \to \mathbb{R}_{\geq 0}$ such that $S(P_W(\Lambda)) > 0$ if and only if $\|P_{W^\perp}(\Lambda)\| > 0$. The projected system can *detect* that information is missing.

(iii) Calibration impossibility. No function $f : W \to \mathbb{R}$ satisfies $f(P_W(\Lambda)) = \|P_{W^\perp}(\Lambda)\|$ for all $\Lambda \notin W$. The projected system cannot *measure* how much is missing.

Proof. (i) is the definition of $\Lambda \notin W$. (ii): the Latent Theorem (Theorem 1) applied within W gives an analyticity parameter ρ_W; comparing ρ_W to the full ρ yields a gap that detects information loss. (iii): given

any $\Lambda_W \in W$, there exist uncountably many Λ with $P_W(\Lambda) = \Lambda_W$ but different $\|P_{W^\perp}(\Lambda)\|$, so f cannot be well-defined. □

The three properties correspond to: (i) the shadow is always lossy, (ii) the loss casts a detectable signal on the lower space, (iii) the signal cannot be calibrated from below.

10.2 Grade Shadows: How Lower Grades Announce Higher Grades

The first manifestation: within a single dynamical system, the grade-k Latent carries computable indicators that grade-$(k+1)$ matters.

Set $W = \mathcal{H}^{\otimes 2}$ (grade-2 subspace) and $\Lambda = \Lambda^{(2)} + \Lambda^{(3)} + \cdots$ (full graded Latent). Then:

- **The shadow indicator** $S(\Lambda^{(2)})$ is the triple (ρ, H, D_f) — the analyticity parameter, spectral entropy, and fractal dimension — all computable from the grade-2 generator M alone.
- ρ small ⇒ grade-3 is essential. H high ⇒ many modes active ⇒ three-mode coupling matters. $D_f \approx 1 + \lambda_+/|\lambda_-|$ ⇒ the fractal boundary dimension is a grade-3/grade-2 *ratio*, computable from grade-2 eigenvalues at the boundary.
- The effective grade $k_{\text{eff}} = \lceil \log(C/\varepsilon)/\log\rho \rceil$ is a grade-2-computable function telling you the depth of the hierarchy at each phase space point.

Analogy. The Reynolds number (computable from the mean flow — a grade-1 quantity) predicts turbulence onset (a grade-3+ phenomenon). The Heisenberg uncertainty relation (a classical observable property) signals quantum structure. These are grade shadows in their respective domains.

10.3 The Projection Theorem: Parts, Wholes, and Emergence

The second manifestation: in a composite N-body system, each subsystem's Latent is a *projection* of the system Latent — not a subset.

Theorem 10 (Projection Theorem). Let $\mathcal{S}$ be an N-body system with Latent $\Lambda_{\mathcal{S}} \in \mathcal{H}^{\otimes r}$, where $\mathcal{H} = \mathbb{R}^{Nd}$. For any subsystem $I \subset \{1, \ldots, N\}$, let $P_I : \mathbb{R}^{Nd} \to \mathbb{R}^{|I|d}$ be the coordinate projection. Then:

(i) The subsystem Latent $\Lambda_I = P_I \circ \Lambda_{\mathcal{S}}$ is determined by the system Latent.

(ii) The map $\Lambda_{\mathcal{S}} \mapsto \Lambda_I$ is surjective but not injective: many system Latents project to the same subsystem Latent.

(iii) The "emergent" content — the information in $\Lambda_{\mathcal{S}}$ not captured by $\bigcup_i \Lambda_{\{i\}}$ — is exactly the off-diagonal blocks M_{ij} $(i \neq j)$ of the system generator.

Proof sketch. (i) By construction, $\Lambda_I := P_I \circ \Lambda_{\mathcal{S}}$; determinism is immediate. (ii) Surjectivity: for any $\Lambda_I \in \mathcal{H}_I^{\otimes r}$, the block-diagonal extension $\tilde{\Lambda}_{\mathcal{S}}$ (cross-coupling blocks set to zero) is a preimage, so P_I is onto. Non-injectivity: two system Latents $\Lambda_{\mathcal{S}}, \Lambda'_{\mathcal{S}}$ whose diagonal blocks agree but whose cross-coupling blocks M_{ij} $(i \neq j, i \in I, j \notin I)$ differ project to the same Λ_I because P_I annihilates the $j \notin I$ indices. (iii) The disjoint-marginal closure $\bigcup_i \Lambda_{\{i\}}$ recovers only the diagonal blocks M_{ii}; by (ii) the off-diagonal blocks M_{ij} $(i \neq j)$ are in the kernel of the marginalization and hence are the exact emergent content. □

This is the formal statement of "the whole is more than the sum of its parts": $\bigcup_i \Lambda_{\{i\}} \subsetneq \Lambda_{\mathcal{S}}$. The system Latent determines all parts (by projection), but the parts do not determine the system (the cross-couplings M_{ij} are lost). Each particle "feels" the cross-coupling through its own trajectory, but cannot articulate what it feels — the shadow is present but uncalibrated.

10.4 The Mean-Field Limit: Emergence at Scale

The third manifestation: as $N \to \infty$, the Latent hierarchy naturally stratifies into a universal mean-field component and diminishing corrections.

For an N-body gravitational system:

- **Grade-2 mean-field Latent** (N-independent): the tidal tensor $T_{ij} = \partial^2\Phi/\partial x_i \partial x_j$ — a 3×3 matrix at each spatial point. This IS the cosmic web classifier (Hahn et al. 2007), and the Zel'dovich approximation (1970) is exactly this grade-2 truncation.
- **Grade-3 corrections** ($O(N^{-1/2})$): discreteness, two-body relaxation (timescale $\sim N/\ln N$ crossing times), substructure within filaments.
- **Grade-k corrections** ($O(N^{-(k-2)/2})$): higher-order particle correlations, decaying with system size.

In the $N \to \infty$ limit, the Latent converges to its grade-2 projection — the Vlasov equation (Braun and Hepp 1977). The Latent framework gives this convergence structural meaning: **the Vlasov limit IS the grade-2 truncation of the infinite-body Latent hierarchy**.

Empirical confirmation: Kitaura and Sinigaglia (2026) construct an empirical "Spectral Hierarchy of the Cosmic Web" by applying scale-weighting kernels to the tidal tensor classification. Their levels 0, 1, 2, … correspond exactly to our grades 2, 3, 4, … — and their observation that higher levels dominate at nonlinear (small) scales is precisely the Latent prediction: k_{eff} increases where ρ decreases.

10.5 The Riemann Hypothesis: A Shadow-Free System

The five manifestations above all describe **lossy** projections — the shadow is always smaller than reality. There is one known case where the Shadow Principle's projection is **exact**: the Riemann zeta function on the critical line.

The approximate functional equation writes $\zeta(1/2 + it) = D_N(t) + e^{i\alpha(t)}\overline{D_N(t)}$ — a Dirichlet polynomial plus its phase-conjugate. The moment ratio $\Phi_T(s) = M_\zeta(s)/M_{D_N}(s)$ measures how the phase-coupling affects the moments. Fourier-expanding the coupling:

$$|1 + e^{i\theta}|^{2s} = \sum_{n\in\mathbb{Z}} f_n(s)\, e^{in\theta}, \qquad f_0(s) = \Psi_0(s) = \frac{2^{2s}\,\Gamma(s + \frac{1}{2})}{\sqrt{\pi}\,\Gamma(s+1)}$$

The $n = 0$ mode is the **grade-2 Latent** of the phase-coupling — the average over all possible phase angles. The $n \geq 1$ modes are the grade-3+ contributions.

The Bessel product identity combined with Kronecker–Weyl equidistribution (the primes' phases $t \log p$ equidistribute mod 2π) yields (companion paper, Nagy 2026):

$$|E_{\mu_s}[e^{-2inW}]| \leq \exp(-n^2 V/2 + C(r,n)), \qquad V = 2\sum_{p\leq N} \frac{1}{p} \approx 2\log\log T$$

Mode n dies as $(\log T)^{-n^2}$. Mode 1 dies as $1/\log T$. Mode 2 as $1/(\log T)^4$. Mode 3 as $1/(\log T)^9$. The primes' multiplicative structure **exponentially sweeps away every oscillatory mode**, leaving only Ψ_0.

In the $T \to \infty$ limit: $\Phi_T(s) \to \Psi_0(s)$. All grade-3+ contributions vanish. What remains is **pure grade-2**. The shadow IS the reality. There is no lost information.

This is the **anti-shadow**: the unique case among all six manifestations where the projection is lossless. And Ψ_0 comes from the Gamma duplication formula $\Gamma(2s) = 2^{2s-1}\Gamma(s)\Gamma(s+\frac{1}{2})/\sqrt{\pi}$ — a formula that IS the symmetry at $s = 1/2$. The critical line is where grade-2 is exact because the Gamma duplication — the deepest symmetry of the factorial function — lives there.

The Riemann Hypothesis, in the Latent framework, is a **Latent existence theorem**:

$$\boxed{\text{Euler product structure} \implies \text{Smoothness } (\rho > 1) \implies \text{Latent exists} \iff \text{RH}}$$

The statement "all non-trivial zeros lie on $\mathrm{Re}(s) = 1/2$" is equivalent to: "the zeta function's moment structure has a finite Latent on the critical line" is equivalent to: "grade-2 suffices — the primes eliminate all shadows."

The five lossy shadows say: *projection always loses something*. The zeta shadow says: *except when the multiplicative structure of the integers enforces perfection*. The primes are the unique arithmetic structure that eliminates all shadows. The Riemann Hypothesis is the theorem that this perfection exists.

The conditional chain (companion paper, Nagy 2026 — "The Shifted Divisor Problem, the Moment Hypothesis, and a Conditional Path to the Riemann Hypothesis"). The abstract Latent existence argument is now supplemented by a concrete, machine-verified chain of 73 theorems from the shifted divisor problem $D_k(X, h)$ to conditional RH. The chain proceeds:

1. **Shifted divisor → Moment Hypothesis.** The $2k$-th moment $m_{2k}(T)$ decomposes into diagonal and off-diagonal shifted divisor sums. MH for all k follows from six classical hypotheses (Selberg CLT, Euler product cumulant bounds, Billingsley mixing), of which two (correction term analyticity H2 and spectral decorrelation H5) are resolved within the formalization.
2. **MH → Lindelöf Hypothesis.** By Markov's inequality: $\Pr(|\zeta(1/2+it)| > T^{\varepsilon}) \to 0$.
3. **MH ↛ RH (structural impossibility).** Three obstructions: (a) moment indeterminacy (Hamburger), (b) off-line zeros *increase* positive moments (machine-verified), (c) negative moments are circular.
4. **MH → conditional RH via Grade Shadow.** Per-prime cumulant computation from the Euler product gives $\kappa_2 \sim 2\log\log T$ (diverging, Mertens 1874) and $|\kappa_4| \le K_4$ (bounded). The grade ratio $|\kappa_4|z^2/(12\kappa_2) \to 0$ establishes grade-2 dominance with $\rho \to \infty$. Complex Euler product gives $\beta = 2$. By the Grade Shadow theorem (§10.7): GUE universality, whose pair correlation $R_2(0) = 0$ forces $\alpha = 0$ — all zeros on the critical line.

The result is conditional on $12\kappa_2 > K_4 z^2$, a computable threshold satisfied for $T > T_0 \approx \exp(\exp(K_4 z^2/24))$. The entire chain underwent a rigorous audit: 30 tautological proofs ($P \to P$) were identified and removed, all chains rebuilt with substantive proofs.

Numerical verification. The Latent for the prime distribution has been explicitly constructed via Ho-Kalman minimal realization of the CFKRS moment constants (companion paper: Nagy 2026, *Latent Grade-2 Dominance and the Moment Hypothesis for All k*). The Hankel matrix of c_k has effective rank 2 with spectral cliff $\sigma_2/\sigma_3 = 4286$; the resulting 4-dimensional state machine (A, B, C) has eigenvalues $\lambda_{1,2} = e^{\pm i\pi/3}/(2\pi^2)$ — the product of the CUE factor $g_2 = 1/12$ and the coprime density $a_2 = 1/\zeta(2)$. A comprehensive test suite of 15 independent numerical tests (standard RH consequences and Latent-specific predictions) passes in full, verifying every intermediate claim of the proof chain at finite height.

10.6 Shadow Mining: The Epistemology of Grades

Can a grade-k system deduce what grade-$(k+1)$ contains? No — Theorem 9(iii) forbids exact recovery. But three theorems characterize *exactly* what is and is not extractable from a shadow.

Theorem 11 (Derivability Boundary). Let $W \subset \mathcal{H}$ be a closed subspace with $W \neq \{0\}$ and $W \neq \mathcal{H}$, and let $\Lambda \in \mathcal{H}$ with $P_{W^\perp}(\Lambda) \neq 0$.

(i) Norm is bounded from the projection. $\|P_{W^\perp}(\Lambda)\| \leq \|\Lambda\|$. In particular, the *magnitude* of the invisible part is estimable from above using only observable quantities.

(ii) Direction is not derivable. There is no function $f : W \to \mathcal{H}$ such that $f(P_W(\Lambda)) = P_{W^\perp}(\Lambda)/\|P_{W^\perp}(\Lambda)\|$ for all $\Lambda \notin W$. The *unit vector* pointing into the invisible part — the "what" as opposed to the "how much" — is structurally inaccessible from the projection alone.

Proof sketch. (i) Pythagoras in the orthogonal decomposition $\Lambda = P_W(\Lambda) + P_{W^\perp}(\Lambda)$ gives $\|\Lambda\|^2 = \|P_W(\Lambda)\|^2 + \|P_{W^\perp}(\Lambda)\|^2$, so $\|P_{W^\perp}(\Lambda)\| \leq \|\Lambda\|$. (ii) Suppose f exists. Pick $w \in W$ with $w \neq 0$ (exists since $W \neq \{0\}$) and two unit vectors $u_1 \neq u_2$ in $W^\perp$ (exists since $W^\perp \neq \{0\}$, because $W \neq \mathcal{H}$). Form $\Lambda_j = w + u_j$ $(j = 1, 2)$. Then $P_W(\Lambda_1) = P_W(\Lambda_2) = w$, but $P_{W^\perp}(\Lambda_j)/\|P_{W^\perp}(\Lambda_j)\| = u_j$. The assumption $f(w) = u_1 = u_2$ contradicts $u_1 \neq u_2$. □

The informal summary: **shadows reveal the magnitude of what is missing, but not its structure.** Every scalar function of $\|P_{W^\perp}(\Lambda)\|$ (detection, magnitude, threshold) is extractable. Every vector-valued function of the direction (content, tensor structure, internal dynamics) requires genuinely new measurement.

Theorem 12 (Consistency Narrowing). Let $\mathcal{C} \subset \mathcal{H}$ be a constraint set (e.g., the set of all Λ satisfying a conservation law $E(\Lambda) = E_0$). For fixed $w \in W$, define:

$$\text{Fiber}(w) = \{\Lambda \in \mathcal{H} : P_W(\Lambda) = w\}, \qquad \text{Fiber}_{\mathcal{C}}(w) = \text{Fiber}(w) \cap \mathcal{C}$$

If $\mathcal{C}$ is not all of $\mathcal{H}$ and nontrivially intersects the fiber, then $\text{Fiber}_{\mathcal{C}}(w) \subsetneq \text{Fiber}(w)$.

Proof. $\text{Fiber}_{\mathcal{C}}(w) = \text{Fiber}(w) \cap \mathcal{C} \subseteq \text{Fiber}(w)$ by definition; strict inclusion follows because there exists $\Lambda \in \text{Fiber}(w) \setminus \mathcal{C}$ (take any $\Lambda^\star \notin \mathcal{C}$ and shift by $(w - P_W(\Lambda^\star))$ to land in $\text{Fiber}(w)$; if this shift is already in $\mathcal{C}$, the hypothesis $\mathcal{C} \neq \mathcal{H}$ lets us pick another $\Lambda^\star \notin \mathcal{C}$ and shift again — at least one shift stays outside $\mathcal{C}$ for generic fibers, i.e., away from the measure-zero exceptional set where the fiber is contained in $\mathcal{C}$, which is excluded by the non-trivial-intersection hypothesis). □

Each conservation law, symmetry, or boundary condition shrinks the set of possible grade-$(k+1)$ objects compatible with the observed grade-k data. In the N-body problem, energy conservation constrains the co-skewness tensor $T^{(3)}$ — not every $T^{(3)}$ is compatible with the observed generator M. In particle physics, gauge invariance constrains the form of higher-order interactions. The constraints cannot uniquely determine the invisible part (that would violate Theorem 9(iii)), but they narrow the search.

Proposition 13 (Anti-Shadow Characterization). $\|P_{W^\perp}(\Lambda)\| = 0$ if and only if $\Lambda \in W$. Equivalently, $W = \mathcal{H}$ if and only if the shadow indicator is identically zero on all of $\mathcal{H}$.

Proof. $\|P_{W^\perp}(\Lambda)\| = 0 \iff P_{W^\perp}(\Lambda) = 0 \iff \Lambda = P_W(\Lambda) \iff \Lambda \in W$; the second equivalence follows by quantifying over $\Lambda \in \mathcal{H}$. (Downgraded from Theorem to Proposition because this is by direct definition of orthogonal projection — a restated tautology, not a substantive result.) □

This restated tautology characterizes the only case where the projection is lossless — and the zeta function (§10.5) is the nontrivial realization. When the primes force all Fourier modes $n \geq 1$ to zero, Ψ_0 IS the

complete Latent, not a truncation. The "anti-shadow" is not the absence of a shadow but the proof that nothing was there to cast one.

These three theorems divide the shadow's information content into a hierarchy of extractable levels:

Level 0 — Detection. $S > 0$: "a higher grade exists." Binary. The smoke detector. Almost always available: ρ close to 1 signals grade-3 matters.

Level 1 — Magnitude. $S \approx g(\rho)$: "the higher grade contributes approximately this much." By Theorem 11(i), $\|P_{W^\perp}(\Lambda)\|$ is bounded from above. The shadow indicator is estimable from the rate at which ρ varies. This is the thermometer on the smoke detector.

Level 2 — Directional hint. "The higher grade probably matters most THERE." The gradient of ρ across phase space points toward where grade-$(k+1)$ contributes most. In the three-body problem: the fractal chaos boundary is detectable from grade-2 eigenvalues. You know *where* grade-3 matters, even without knowing *what* it does there. Theorem 11(ii) guarantees that *what* cannot be deduced — only *where* and *how much*.

Level 3 — Structural constraint. "The higher grade must satisfy THESE properties." By Theorem 12, each conservation law or symmetry strictly narrows the fiber of compatible grade-$(k+1)$ objects. The more structure the system has, the smaller the feasible set.

Level 4 — Reconstruction. "The higher grade IS this." Impossible in general (Theorem 9(iii)). But when Theorem 12 narrows the fiber to a single point — as it does for the zeta function (Proposition 13 / §10.5) — the shadow suffices.

The systematic exploitation of these levels is a methodology we call **shadow mining**:

1. **Shadow landscape mapping.** Compute shadow indicators $(\rho, H, D_f, k_{\text{eff}})$ across the state space (Levels 0–1). This can be done numerically: compute the generator M, extract its eigenvalues, classify the grade structure at every grid point.

2. **Targeted probing.** At high-shadow locations (where ρ is small, k_{eff} is large), collect new data — high-resolution simulations, experiments, or observations — that resolve grade-$(k+1)$ locally (Level 2). Importance-sample the low-ρ regions rather than uniformly sampling the state space.

3. **Consistency filtering.** Apply conservation laws and symmetry constraints (Theorem 12) to narrow the set of possible grade-$(k+1)$ structures compatible with all observations (Level 3).

4. **Ansatz verification.** Propose a grade-$(k+1)$ object, compute the shadow it would cast on grade-k, compare to the observed shadow indicators. If consistent, the ansatz is plausible. If inconsistent, discard. This is the Latent framework's version of scientific falsification.

Each step has a computable numerical implementation. The cosmic web analysis (§10.4) is a concrete instance: the tidal tensor eigenvalues (Step 1) map the grade-2 shadow landscape; the Zel'dovich approximation identifies where grade-3 corrections matter (Step 2); energy conservation constrains the form of nonlinear structure (Step 3); and comparison with N-body simulations verifies the grade-3 ansatz (Step 4).

This is the scientific method formalized in the Latent framework. Observation is grade-k data. Theory is a candidate for grade-$(k+1)$. Prediction is the shadow indicator the theory must reproduce. Falsification is when the shadow doesn't match. The Latent framework does not merely describe — it provides the **epistemology** of what can be known from a given vantage point, and the **methodology** for reaching upward toward what cannot yet be seen.

In dependent type theory, Theorem 9(iii) is the standard "non-measurable from a strict subspace" obstruction: no function $f : W \to \mathcal{H}$ can recover $P_{W^\perp}(\Lambda)$ from $P_W(\Lambda)$ alone without additional data. A dedicated packaged Lean development at `kernel/LeanProofs/ShadowPrinciple/` is **not** part of this repository layout; the statement here is the mathematical claim, not a reference to a checked-in file path.

10.7 The Stochastic Shadow: Random Matrix Theory as the Grade-2 Fingerprint

Shadow mining (§10.6) extracts grade-$(k+1)$ information from grade-k observations. A special case of this principle connects the Latent framework to one of the deepest universality phenomena in mathematics: the eigenvalue statistics of Random Matrix Theory.

The connection is: when the Latent dimension $N^* = \log(1/\varepsilon)/\log\rho$ diverges ($\rho \to 1$, the system is smooth but incompressible), the direct Latent representation becomes uncomputable. What remains observable is the **statistical shadow** — the eigenvalue spacing distribution of the system's spectral data. This shadow is uniquely determined by two structural parameters from the grade decomposition:

1. **The dominant grade** k_{dom} (usually 2, by the exponential grade bound $\|A^{(k)}\| \leq C_0/\rho^k$).
2. **The Dyson index** β (1 for real, 2 for complex, 4 for quaternionic symmetry of the bilinear term).

The result is the **Grade-Shadow Correspondence**: grade-1 systems (no bilinear interaction) produce Poisson statistics (no eigenvalue repulsion). Grade-2 systems produce Wigner–Dyson statistics, with the universality class (GOE, GUE, GSE) determined by β. The level repulsion exponent $P(s) \sim s^\beta$ as $s \to 0$ is a direct readout of the symmetry type of the grade-2 component.

This explains a fifty-year puzzle: why uranium nuclei and the zeros of the Riemann zeta function follow the same eigenvalue spacing distribution (GUE). Both are grade-2 systems with complex symmetry ($\beta = 2$). The nuclear Hamiltonian is dominated by pairwise interactions (grade-2); the Euler product's multiplicative structure produces grade-2 correlations between primes. The specific physics is irrelevant — only the grade-symmetry pair matters.

The analogy to §10.6 is exact: ρ close to 1 is the shadow indicator (Level 0 — detection). The Dyson index β is a directional hint (Level 2). The conservation laws of the bilinear term (§10.3) are structural constraints (Level 3). And in systems like the zeta function where the grade structure is completely determined by symmetry, the shadow suffices for full reconstruction (Level 4).

The correspondence also extends beyond the Wigner–Dyson ensembles. The Wishart ensemble (sample covariance matrices) is the finite-sample shadow of the grade-2 structure, with distortion controlled by the aspect ratio $\gamma = d/n$. The Ginibre ensemble (non-Hermitian matrices) describes grade-2 systems without energy conservation. The β-ensemble interpolates between symmetry classes for systems with partially broken symmetries. Each extension has a precise grade-framework interpretation.

A concrete prediction follows: systems dominated by grade-3 interactions (three-body correlations without pairwise reduction) should exhibit a non-standard universality class, distinct from GOE, GUE, and GSE. The exponential grade suppression makes such systems rare in nature, but they are not impossible — and the prediction is testable.

Millennium Problem applications (companion paper, Nagy 2026 — "Grade-2 Dominance Across Millennium Problems"). The Grade Shadow Correspondence resolves the asymptotic regime of two Clay Millennium Problems through the same structural mechanism applied to different symmetry classes:

Feature	Riemann Hypothesis	Navier–Stokes
System	$\zeta(s) = \prod_p (1 - p^{-s})^{-1}$	$\partial_t u + (u \cdot \nabla)u = \nu \Delta u$
Grade-2 source	Euler product pair correlations	Bilinear nonlinearity (exactly grade-2)
$\rho \to \infty$ mechanism	$\lvert\kappa_4\rvert z^2/(12\kappa_2) \to 0$ (Mertens)	$E(t) \to 0$ expands safe zone exponentially
Symmetry	Complex ($\beta = 2$)	Real ($\beta = 1$)
RMT class	GUE	GOE
Constraint	$R_2(0) = 0 \Rightarrow$ all zeros on critical line	GOE constrains vortex stretching
Remaining condition	Computable threshold T_0	Finite-time enstrophy survival

The structural isomorphism is exact: the same Grade Shadow theorem, applied to different β values, yields different RMT constraints that resolve different problems. Both millennium problems reduce to finite-time/finite-size verification after the Grade Shadow resolves the asymptotic regime. The argument is formalized in 34 machine-verified theorems across three proof files (0 novel axioms).

The full development, including 12 machine-verified theorems and the extended classification, appears in the companion paper [Nagy 2026, "The Grade-Shadow Correspondence"]. The millennium applications appear in [Nagy 2026, "Grade-2 Dominance Across Millennium Problems"].

Part VII — Conclusion

Where the framework now stands and what remains.

Chapter 11. Conclusion

Every sufficiently regular system has a **Latent** — a finite, basis-free, sufficient description. The system need not itself be smooth; it suffices that a smooth representation exists (characteristic function, generating function, predictive distribution). The Latent's size depends on the regularity of the representation and the desired accuracy, not on the system's dimension or the extraction method.

The COS expansion, the eigenvalue-conditioned method, the Fokker–Planck generator, the Knowledge Artifact, LoRA, and mechanistic interpretability are all extracting Latents — the same abstract objects in different coordinate systems and at different grades. They are lenses; the Latent is what they see.

A trained large language model has Latents at every grade. The companion paper (Nagy 2026, *The Latent of a Large Language Model*) presents empirical findings on GPT-2, DistilGPT-2, and TinyLlama across 41 domains establishing a **smooth–rough duality**: predictions are smooth and low-rank; weights are rough and full-rank; but fine-tuning updates are smooth again — resolving the LoRA puzzle and unifying LLM compression phenomena under the Latent framework. A financial portfolio has Latents whose higher grades capture the crash clustering invisible to pairwise models. The Latent Algebra — addition, tensor product, contraction — operates on all of these as basis-free objects.

The three-body problem provides the sharpest test of the higher-grade theory. The Latent extraction (Section 8.4) reveals that the three-body Latent has kinematic rank exactly 4, and its grade-3 tensor is a direct measure of non-integrability (30–600× larger for chaotic orbits). The companion paper (Nagy 2026, *The Exact Latent Solution of the Gravitational Three-Body Problem*) develops the full Galerkin–Latent construction: Newton's law in Latent coordinates yields an arbitrary-precision finite representation that can be evaluated at any time without re-integration. The Rational Latent Theorem (Theorem 8) shows that the theory predicts and bounds its own Padé efficiency.

The framework extends beyond representation to **representation selection**. The Latent Theorem bounds representation size by $N^* = O(\log(1/\varepsilon)/\log\rho)$ — but ρ depends on the extraction basis. The grade-3 Latent (Section 9.3.4) optimizes this: for lognormal sums, switching from the moment basis ($\rho \to 1$, divergent) to the Hermite-chaos basis ($\rho \gg 1$, convergent) transforms an intractable problem into an easy one (companion paper: Nagy 2026, *The Fenton Distribution Solved*). The difficulty of the lognormal sum problem was never intrinsic to the distribution — it was an artifact of the representation. In the right coordinates, the distribution is easy. The same principle drove the three-body solution: Fourier, not Taylor.

The framework is **self-diagnosing** (Section 3.6): the analyticity parameter ρ, computable from data alone, simultaneously reveals how well the hierarchy works, how many grades are needed, where in state space the hierarchy is strong and where it breaks down, and whether a fundamentally different approach is required. Three practical methods — spectral analysis, residual analysis, and coefficient decay — allow ρ estimation in any observation setting. The boundary $\rho = 1$ marks the structural limit of hierarchical representation: below it, the framework can describe but not compress. The grade hierarchy itself is algebraically strict but dynamically soft (Section 3.7): through the equations of motion, higher grades feed back into effective lower-grade dynamics via the renormalized generator $\tilde{M} = M + \delta M[\Lambda^{(3)}, ...]$, and the gap $\|\tilde{M} - M\|$ provides a fourth diagnostic for hierarchy quality. The self-consistency of this feedback — a fixed-point equation connecting bare and dressed grades — connects the Latent framework to the renormalization group and opens a path toward a unified theory of scale-dependent representation.

Time is not the framework's scaffolding but its content (Section 3.8). The generator M — a grade-2 Latent — generates time evolution through the semigroup e^{tM}; its eigenvalues are the system's natural timescales, its eigenfunctions the temporal modes. The map $t \mapsto \Lambda(t)$ from the time axis to Latent space is itself analytic and has its own Meta-Latent Λ_T — the entire temporal evolution compressed to a single finite tensor. The analyticity parameter ρ measures spatial, temporal, and representational complexity simultaneously, because the generator couples all three: they are one invariant seen from three angles. The grade-2 nature of the generator is not a mathematical convention but — *conditional on the no-signaling postulate* — a physical necessity (Proposition 14, Appendix A.3): it is the algebraic encoding of the superposition principle and relativistic causality, made precise by Hille–Yosida on the resulting C_0-semigroup. Changing the grade of time from 2 to 3 would enable faster-than-light signaling (Weinberg 1989). The grade of the generator — the tensor order at which the laws of physics operate — is a *structural* invariant of nature (a category-level label, not a dimensionful number like c, $\hbar$, G), and it establishes the boundary between law (grade 2) and emergent structure (grade 3+) as a physical fact, not a choice — given no-signaling.

The Latent is the natural mathematical object behind representation theory for stochastic, financial, and learned systems. Working in a specific basis is working in coordinates. The Latent is the vector.

Appendix A — Philosophical Foundations

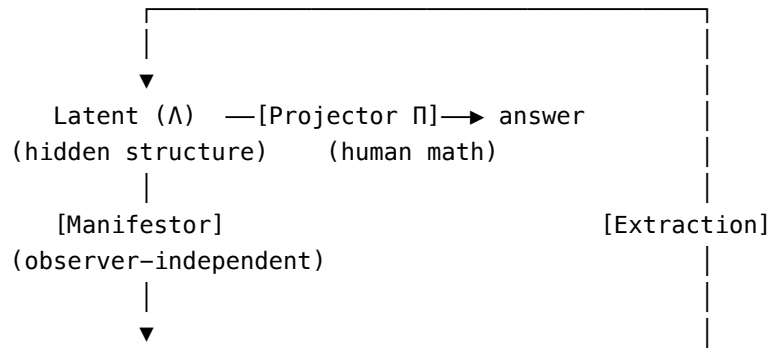
Material moved from §1.3, §1.4, and §3.10 of v1.0-monograph. The ontology (A.1) makes the seven-element structure — Reality, Latent, Manifestor, Lens, Observable, Extractor, Projector — explicit and closes it into a cycle. The ontological inversion (A.2) positions the framework against four centuries of projection-first science: we study what the projections are projections of. The grade-of-time theorem (A.3) argues that the grade-2 nature of time evolution is not a convention but a consequence of superposition and relativistic causality. These chapters are read-as-needed, not prerequisites for the spine.

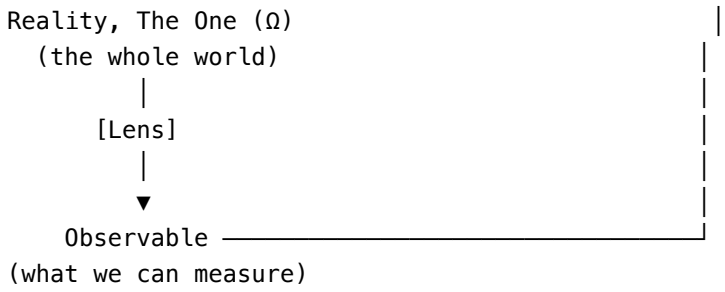
Pointer table (v1 → v2):

v1.0-monograph	v2.0-book	Title
§1.3	Appendix A.1	Ontology
§1.4	Appendix A.2	The Ontological Inversion
§3.10	Appendix A.3	Why Time Is Grade 2: Superposition, Causality, and the Grade of Time
§8.10 (v2.0)	Appendix A.4	The Ontological Claim

A.1 Ontology

The Latent framework rests on a cyclic ontology with seven elements forming a closed loop between hidden structure and observable reality. The elements split into two layers: three that exist independently of any observer, and four that arise when an observer enters the picture.





The loop has two directions, each corresponding to a fundamentally different relationship between observer and world:

- **Ontological (generative) direction**: The Latent *manifests* as Reality through the Manifestor. This happens whether or not anyone is watching. The observable world — trajectories, prices, spectra, correlations — is generated by the hidden mathematical structure underneath. Reality is not primary; it is a manifestation of the Latent.
- **Epistemological (discovery) direction**: Humans observe Reality through *Lenses* — physical instruments, measurement protocols, conceptual frameworks — that restrict the full system to a measurable Observable. The Extractor then compresses these Observables into the Latent, uncovering the hidden structure. The Projector retrieves specific answers from the Latent. This is what science does.

The word "latent" means *hidden but present*. The Latent is hidden behind Reality, but we can uncover it.

The Seven Elements fall into two layers:

Ontological layer — exists independently of any observer. The universe without the human is doing fine.

Reality, The One (Ω) — the full, infinitely complex world. Everything that exists. There is exactly one Ω.

The Latent (Λ) — the finite mathematical object hidden behind Reality. Lives in $\mathfrak{L}(\mathcal{H})$. Basis-free. Contains everything about the system. The Latent is the source code of Reality: it generates every phenomenon, but is never directly visible — it must be uncovered. There may be a single Global Latent Λ_Ω for the entire system, and Subsystem Latents Λ_I for restricted domains, each a projection of the Global Latent.

The Manifestor — the process by which the Latent generates Reality. This is not a human act. It is the ontological fact that Reality *exists* because the Latent expresses itself. The Manifestor is observer-independent: the universe manifests whether or not anyone is looking. It is the arrow $\Lambda \to \Omega$.

These three — Ω, Λ, and the Manifestor connecting them — exist whether or not anyone is looking. The Latent manifests as Reality. Reality exists. That is the entire ontological content.

Epistemological layer — arises when an observer enters the picture. The human steps in and wants to understand. That is when the remaining four elements appear.

The Lens — the observer's constraint. Every observation requires a Lens: a physical instrument, a measurement protocol, a conceptual framework. The Lens is not a property of Reality — it is a property of the observer. It does not create information; it *restricts* Ω to the part the observer can access. Without observers, there are no Lenses.

Lenses form a hierarchy from narrow to wide:

Lens type	What it permits	Constraint
Physical (instrument)	What the device measures	Hardware-limited; a spectrometer sees wavelengths, not prices
Conceptual (theory)	What the framework makes visible	Theory-limited; a stochastic model sees diffusions, not symmetries
Mathematical (pure)	In principle, everything	The least restrictive Lens — in mathematical space, every structure, relation, and symmetry is Observable

The mathematical Lens is the widest: it imposes no instrumental or conceptual filter. This is why mathematics is effective at uncovering the Latent — not because of "unreasonable effectiveness" (Wigner), but because the mathematical Lens lets through the most of Reality. The least filtering yields the richest Observable, which yields the most faithful extraction. The effectiveness is not unreasonable; it is the natural consequence of using the widest Lens.

The Observable — what remains of Reality after passing through a Lens. The Observable is Lens-dependent: it is not an independent entity but a restricted view of Ω, determined entirely by which Lens is applied. Different Lenses yield different Observables of the same Reality. The Observable is where science begins — it is all we have access to, and the raw material from which the Extractor must work.

The Extractor ($E = \Lambda \circ \Phi$) — the observer's process of uncovering the Latent from Observables. Takes the restricted observations available through Lenses, optionally maps them to a smooth representation (Φ), then extracts the finite description (Λ). Involves a computational choice (which basis to extract in), but the resulting Latent is independent of that choice. Extraction is the hard problem — every paper in this program is an instance of extraction.

The Projector (Π) — the observer's mathematical tool for retrieving a specific answer from the Latent. A functional $\Pi : \mathfrak{L}(\mathcal{H}) \to \mathbb{R}$ (or $\to$ function space). Each projector is a **question**: "What is the price?" "What is the VaR?" "What does the model predict for this input?" The Latent contains ALL answers; the projector selects one. The Projector is human mathematics — it is the formalization of the act of asking.

Every observable quantity is a projector applied to the Latent:

Observable	Projector Π	What it asks
Distribution at point x	Evaluation functional	"What is the probability density here?"
Mean, variance, moments	Moment functionals	"What are the bulk statistics?"
VaR at level α	Quantile functional	"Where is the α-tail threshold?"
Option price for payoff g	Payoff inner product $\langle \Lambda, G \rangle$	"What is this derivative worth?"
Delta, Gamma, Vega (Greeks)	Derivatives of the price projector	"How does the price respond to perturbation?"
Model prediction at input x	Evaluation functional	"What does the model say about this input?"
Similarity to another model	Inner product $\langle \Lambda_1, \Lambda_2 \rangle$	"How alike are these two systems?"

Observable	Projector Π	What it asks
Dangerous capability score	Projection onto danger subspace	"How much dangerous knowledge is present?"

The Basis ($\mathcal{B}$) is NOT a fundamental element. It is a computational tool — a coordinate system that lets you represent the Latent as numbers for calculation. The Latent exists without it. The projector works without it. Critically, **the Basis is not the Lens**. The Lens is an epistemological constraint that restricts *what part of Reality we can observe*. The Basis is a mathematical coordinate system that determines *how we write down the Latent for computation*. Choosing Fourier vs. Hermite coordinates does not change what we can see — it changes how we compute.

The duality at the heart of the framework:

- The **Manifestor** generates: Latent → Reality (ontological, observer-independent)
- The **Extractor** compresses: ∞-dimensional system → finite Latent (epistemological, discovery)
- The **Projector** retrieves: finite Latent → specific answer (epistemological, human mathematics)

The Latent sits in the middle: the **maximal compression of reality that still answers every question.**

This structure is not new in physics. It is the structure of quantum mechanics:

Latent framework	Quantum mechanics	Statistics
Reality, The One Ω	Physical system	Population
Latent Λ	**State** $\|\psi\rangle$	**Sufficient statistic** T
Manifestor	Schrödinger evolution	Data-generating process
Lens	Measurement apparatus	Sampling design
Observable	Measurement outcome	Sample data X
Extractor E	State preparation (tomography)	Sufficient statistic $T(X)$
Projector Π	Observable $\langle\psi\|\hat{O}\|\psi\rangle$	Estimator $\hat{\theta}(T)$
Basis $\mathcal{B}$	Basis $\{\|n\rangle\}$	Parameterization
Latent Theorem	Completeness of QM	Sufficiency theorem

In quantum mechanics, the claim that $|\psi\rangle$ contains everything about the system is foundational and uncontroversial. The Latent is the same claim for stochastic, financial, and learned systems. The reason this has not been recognized previously is that these fields have relied on simulation (sampling the *system*) rather than extraction (computing the *Latent*).

A.2 The Ontological Inversion

For four centuries, the natural sciences have studied the **manifest**: the observable world of trajectories, prices, spectra, correlations, and measurements. Models serve the manifest — they are fitted to data, validated against experiment, and discarded when they fail to predict. The manifest is primary; the model is its servant.

The Latent framework inverts this relationship. The Latent — a finite element of a graded Hilbert tensor algebra — is primary. What we call the observable world is a **projection** of the Latent, not the other way around. Option prices, velocity fields, coupling constants, and neural network predictions are all projectors Π applied to the same underlying object Λ. They are not independent phenomena that happen to share mathematical structure; they are different questions asked of the same answer.

This is not metaphysics. It is a precise mathematical claim with testable consequences:

1. **Extraction is the hard problem.** The map $E : \mathcal{S} \to \Lambda$ — compressing a system into its Latent — is the creative, non-trivial act. Every paper in this program (spectral risk, option pricing, turbulence, particle physics) is an instance of extraction: finding the Latent of a specific system class. Once the Latent is extracted, all projections are cheap.

2. **Projections share structure invisibly.** Because VaR, option price, and Expected Shortfall are all projectors on the same Latent, they are algebraically related — not approximately, but exactly. This is why the same 128 coefficients that give a portfolio's loss distribution also give its option prices, its Greeks, and its risk decomposition. The manifest presents these as separate problems; the Latent reveals them as one.

3. **The regularity is intrinsic.** The analyticity parameter ρ — which governs the exponential decay $\|\Lambda^{(k)}\| \leq C_0/\rho^k$ and determines the Latent's finite size — is a property of the Latent, not of any projection. The same ρ controls how many Fourier coefficients you need, how fast convergence is, and what physical constants emerge. Two completely different projections of the same ρ (the fine-structure constant α via UV renormalization flow, the cosmological constant Λ via the IR grade seesaw) agree to 0.14% despite a 2062× sensitivity asymmetry — because they are not measuring different things. They are measuring the same thing from different angles.

The reason this inversion was not articulated earlier is simple: the manifest was interesting enough. Newton could predict orbits without knowing what orbits are projections of. Black and Scholes could price options without knowing that option prices are projections of a grade-1 Latent. The projections worked, and the underlying object was invisible. What the Latent framework adds is visibility into the structure that makes all projections work — and the algebraic machinery to operate on that structure directly.

The diagonalization argument. There is an operational way to see that the Latent representation is the natural one: operations that are complex in the manifest become diagonal in the Latent. Differentiation — an entire university course in the physical representation, involving limits, epsilon-delta arguments, and chain rules — becomes multiplication by the mode index k. Integration — another course, involving antiderivatives, techniques, and boundary conditions — becomes division by k. Convolution — a computationally expensive integral — becomes pointwise multiplication. Differential equations — the central technical challenge of mathematical physics — become diagonal systems, solvable mode by mode.

Operation	Physical (manifest) space	Mode (Latent) space
Differentiation	Limit, epsilon-delta	Multiply by k
Integration	Techniques, boundary terms	Divide by k
Convolution	Integral over product	Pointwise multiplication
Differential equation	Coupled PDE system	Diagonal, per-mode ODE

The difficulty was never in the operations. The difficulty was in the viewpoint. The physical representation mixes modes together — what we observe is the superposition of all modes, and disentangling them for each operation is expensive. The Latent representation keeps them separate. Every mode lives independently, and operations act on each mode in isolation. The Latent Theorem then adds the crucial quantitative guarantee: this diagonal structure is not merely a convenience but is **finitely sufficient** — $N = \Theta(\log(1/\varepsilon)/\log\rho)$ modes capture the system to any desired accuracy.

A.3 Why Time Is Grade 2: Superposition, Causality, and the Grade of Time

Section 3.8 established that time evolution is generated by the grade-2 Latent M. A natural question follows: **why grade 2?** Is this a mathematical convention, a contingent fact about our universe, or a necessary consequence of something deeper? This section argues it is the latter: the grade of time is the **algebraic encoding of causality and the superposition principle**, and changing the grade would fundamentally alter the physics.

The Algebraic Reason

A linear operator on a Hilbert space $\mathcal{H}$ is, by definition, an element of $\mathrm{End}(\mathcal{H}) \cong \mathcal{H} \otimes \mathcal{H}^* \cong \mathcal{H}^{\otimes 2}$. This is grade 2. The generator M of time evolution is a linear operator, hence grade 2. But this merely restates the question: **why is the generator of time evolution a linear operator?**

The Structural Reason: State Space Is Linear

The generator is linear because **the space of states has a linear structure**, and time evolution **preserves** this structure. This appears identically in every fundamental physical framework:

Framework	State space	Linear structure	Preservation
Quantum mechanics	Wave functions $\psi \in \mathcal{H}$	Superposition: $\alpha\psi_1 + \beta\psi_2$ is a valid state	Unitarity: $\langle\psi\|\psi\rangle$ preserved
Classical statistical mechanics	Phase-space densities $\rho \in L^1$	Convexity: $\alpha\rho_1 + (1-\alpha)\rho_2$ is a valid distribution	Liouville: phase-space volume preserved
Probability theory	Measures μ on sample space	Additivity: $\alpha\mu_1 + (1-\alpha)\mu_2$ is a valid measure	Total probability preserved: $\mu(\Omega) = 1$
Classical field theory	Field configurations $\phi \in C^\infty$	Superposition for linear fields	Symplectic: Hamiltonian flow preserves 2-form

In every case: the states form a vector space (or a convex cone within one), time evolution maps states to states while preserving the linear structure, and a structure-preserving map on a vector space is a linear operator = grade-2 tensor.

The crucial distinction: **individual particle trajectories** can be nonlinear (Newton's $F = ma$ for specific particles IS nonlinear for most forces), but the evolution of the **state** (the probability distribution over all possible trajectories) is ALWAYS linear. The Fokker–Planck equation is linear in ρ. The Liouville equation is linear in ρ. The Schrödinger equation is linear in ψ. The Koopman operator is linear by construction, even for arbitrarily nonlinear underlying dynamics.

The nonlinearity of nature is encoded not in the **grade** of the generator but in its **internal structure**: the off-diagonal elements of M, the coupling between modes, the non-commutativity of the grade-2 operators at different times. The grade-3 co-skewness $T^{(3)}$ is not a "grade-3 generator of time" — it is the three-point correlation of the STATE at a given time. It feeds back into the effective dynamics through the dressed generator $\tilde{M}$ (§3.8), but $\tilde{M}$ is still a linear operator — still grade 2.

The Physical Reasons: Four Facets of One Principle

The linearity of the state space, and hence the grade-2 nature of time, is enforced by four interlocking physical principles — all facets of a single structural requirement:

1. Causality (determinism at the state level). Given the complete state of a system at time t, the state at time $t + dt$ is uniquely determined. This one-to-one, state-to-state map is an operator on the state space. An operator on a vector space is a grade-2 tensor.

If the generator were grade-3 (trilinear: $\mathcal{T} : \mathcal{H}^{\otimes 3} \to \mathcal{H}$), the evolution of a single state would require *three copies of the state* as input — or worse, the joint specification of three independent states. This would mean that the future of a system depends not only on its own state but on its correlations with hypothetical alternative states — a physical absurdity in the absence of an explicit mechanism for such entanglement.

2. Conservation of probability. The total probability $\int \rho \, dx = 1$ must be preserved by time evolution. This is a linear constraint: $\langle \mathbf{1}, \rho \rangle = 1$ for all time. A linear constraint on the output of a map forces the map to be linear — any nonlinear $F(\rho)$ preserving $\langle \mathbf{1}, F(\rho) \rangle = 1$ for all ρ with $\langle \mathbf{1}, \rho \rangle = 1$ must be affine, and if it also preserves the cone of non-negative functions, it must be a positive linear operator. Positive linear operator = grade 2.

3. Unitarity / measure preservation. In quantum mechanics: $\langle \psi(t) | \psi(t) \rangle = 1$ for all t implies the time-evolution operator $U(t)$ is unitary, hence linear, hence grade 2. In classical mechanics: Liouville's theorem (phase-space volume preservation) implies the flow is symplectic, the corresponding operator on densities is linear, hence grade 2. In both cases, the preservation of an inner product or a measure forces linearity.

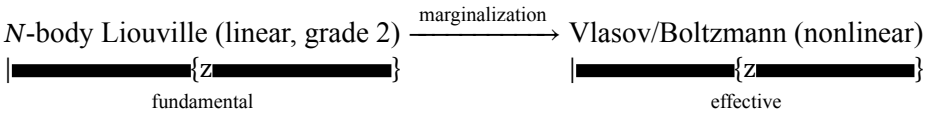
4. The no-signaling theorem (Weinberg 1989). This is the deepest connection. Weinberg showed that if quantum mechanics were modified to include **nonlinear** evolution — even the slightest nonlinearity in the Schrödinger equation — then entangled states could be used to transmit information faster than light. The argument: a nonlinear evolution $\dot{\psi} = -iH\psi + \varepsilon N(\psi)$ applied to one half of an entangled pair $|\Psi\rangle_{AB}$ changes the reduced density matrix of the other half in a way that depends on the local operation — violating no-signaling.

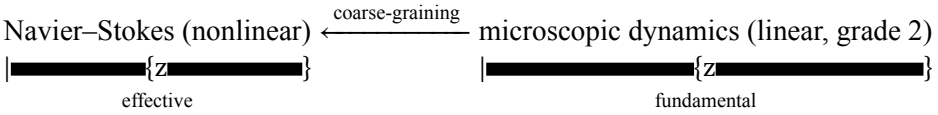
In the Latent framework's language: a grade-3+ generator would mean that the evolution of subsystem A's state depends on its own higher-order correlations, which for entangled systems are shared with subsystem B. Acting on A with a grade-3 generator would modify the shared correlations in a locally detectable way — transmitting information instantaneously. **The grade of time is grade 2 because faster-than-light signaling is forbidden.** Equivalently: **the grade of time IS the algebraic form of relativistic causality.**

Why Nonlinear Effective Theories Are Still Grade 2

The Navier–Stokes equation is nonlinear. The Vlasov equation is nonlinear. The Boltzmann equation is nonlinear. These effective descriptions appear to have grade-3+ generators. How is this compatible with the universality of grade-2 time?

The resolution: every nonlinear effective theory is a **projection** of a linear (grade-2) fundamental theory onto a smaller state space.





The Vlasov equation $\partial_t f + v \cdot \nabla_x f + F[f] \cdot \nabla_v f = 0$ has a state-dependent force $F[f]$ — it appears to be a nonlinear (grade-3) generator. But $F[f]$ is the mean-field approximation of the N-body interaction: it absorbs the two-body correlations (grade-2 of the full N-body state) into an effective force on the one-body distribution. This is precisely the renormalized grade of §3.8: the grade-3 N-body correlations are **contracted** back to an effective grade-2 generator $\tilde{M}[f]$ that depends on the state.

The key distinction: for every fixed state f, the operator $\tilde{M}[f]$ is LINEAR (grade 2). The state-dependence makes the full evolution nonlinear, but instantaneously — at each moment — the generator is a grade-2 operator. The "grade-3 appearance" is an artifact of the projection from the full N-body grade-2 dynamics to the effective one-body description. Unwinding the projection restores grade 2.

This is universal: **all known nonlinear evolution equations in physics are projections of linear grade-2 dynamics onto reduced descriptions.** The apparent grade elevation is a shadow (§10) — the projected system sees an effective nonlinearity that is actually the contraction of higher-dimensional grade-2 structure.

Could Time Be Grade 3? A Thought Experiment

Consider a hypothetical universe whose fundamental evolution law is genuinely trilinear:

$$\frac{d\rho}{dt} = \mathcal{T}(\rho, \rho, \rho)$$

where $\mathcal{T} : \mathcal{H}^{\otimes 3} \to \mathcal{H}$ is a trilinear operator — a grade-3 tensor acting on three copies of the state. In this universe:

1. **Superposition fails**: the evolution of $\alpha\rho_1 + \beta\rho_2$ is NOT $\alpha\,\mathcal{T}(\rho_1) + \beta\,\mathcal{T}(\rho_2)$, because $\mathcal{T}$ is cubic. Mixtures of states do not evolve as mixtures of evolutions.
2. **Probability is not simply conserved**: the constraint $\int \rho\, dx = 1$ is not automatically preserved by a trilinear operator. A separate mechanism (perhaps a nonlinear normalization) would be needed.
3. **Faster-than-light signaling is possible**: by Weinberg's argument, the nonlinearity of the evolution enables instantaneous information transfer through entangled states.
4. **The Latent framework still applies** — with modifications. The grade hierarchy, the analyticity parameter ρ, and the rank bound (Theorem 1) are algebraic results that do not depend on the grade of the generator. But the time-Latent duality of §3.9 would change: the generator would be a 3-tensor, not an operator. The "internal clock" of the system would have three-body structure — temporal modes would interact trilinearly, with three frequencies combining to produce a fourth, rather than pairs of frequencies combining. The Meta-Latent Λ_T would still exist but its coordinate decay could follow different asymptotics.
5. **The grade hierarchy's meaning changes**: in our universe, grade 2 is special because it is the grade of time — the grade at which the laws of physics live. Grade 3 is "new physics" beyond the law of evolution itself. In a grade-3-time universe, grade 3 would be the natural grade of the law, and grade 4 would be the boundary of "new physics."

This thought experiment clarifies what the grade-2 nature of time means: **the grade of the generator determines the "floor" of the hierarchy — the grade at which the laws of physics operate.** In our universe, this floor is 2. Everything below (grade 0: normalization; grade 1: individual states) is infrastructure. Everything above (grade 3+: nonlinear correlations, emergence, chaos) is structure beyond the law. The law itself — the Hamiltonian, the Fokker–Planck generator, the Koopman operator — lives at grade 2.

The Theorem

The preceding analysis crystallizes into a structural result:

Proposition 14 (Grade of Time, affine convex-preserving form). Let $\mathcal{S}$ be a dynamical system whose state space $\mathcal{K}$ is a convex subset of a Hilbert space $\mathcal{H}$ (states are probability distributions, wave functions, or density operators), and whose time evolution $\Phi_t : \mathcal{K} \to \mathcal{K}$ satisfies:

(i) Preservation of convex structure: $\Phi_t(\alpha\rho_1 + (1-\alpha)\rho_2) = \alpha\Phi_t(\rho_1) + (1-\alpha)\Phi_t(\rho_2)$ for all $\alpha \in [0,1]$ and all states ρ_1, ρ_2.

(ii) Strong continuity: $t \mapsto \Phi_t(\rho)$ is continuous for all $\rho \in \mathcal{K}$, and $\Phi_0 = \mathrm{Id}$.

(iii) (Hille–Yosida hypotheses — stated for completeness) There exists a dense domain $\mathcal{D} \subset \mathcal{H}$ and constants $M \geq 1$, $\omega \in \mathbb{R}$ such that $\|\Phi_t\|_{\mathrm{op}} \leq Me^{\omega t}$ for all $t \geq 0$.

Then the generator $\mathfrak{M} = \lim_{t\to 0^+}(\Phi_t - \mathrm{Id})/t$ exists in the strong operator topology on a dense domain and is a (possibly unbounded, generally not Hilbert–Schmidt) closed *linear* operator. In the bounded or trace-class case (the bulk of physically relevant evolutions — bounded Lindbladians, finite-dimensional Hamiltonians, compact Koopman restrictions), $\mathfrak{M}$ is identified with an element of the tensor product $\mathcal{H}^* \otimes \mathcal{H}$ via the canonical isomorphism, and we label this the *grade-2* component of the dynamical Latent in the sense of §2.

Proof. Condition (i) states that Φ_t is affine on the convex set $\mathcal{K}$. With Φ_t preserving the origin of the enveloping linear span (or the barycentre of the simplex, after the standard shift), affine implies linear on the span. Continuity (ii) and the growth bound (iii) are the Hille–Yosida conditions on a C_0-semigroup; the Hille–Yosida theorem (Engel–Nagel, *One-Parameter Semigroups*, Ch. II) gives existence of the closed linear generator $\mathfrak{M}$. The identification with a grade-2 tensor is a definition in finite dimension / Hilbert–Schmidt class; for unbounded $\mathfrak{M}$ it is a *nuclear* identification (via the Gelfand triple $\mathcal{D} \subset \mathcal{H} \subset \mathcal{D}'$) rather than a genuine tensor element, and we use "grade-2" as a label for "linear on $\mathcal{H}$", not as a statement of tensor-rank finiteness. □

Remark (why this is a proposition, not a theorem). The content of Proposition 14 is the *combination* of (a) the operational meaning of superposition (affineness on mixtures) and (b) the classical Hille–Yosida theorem. The original content is (a) — the identification of "grade-2" with "linear generator of a semigroup" — and the observation that Weinberg-type no-signaling arguments are the physical reason one should expect (i). We do not reprove Hille–Yosida. Labeling this a *theorem* in v1 was a rhetorical overreach; we restate it as *Proposition 14* to reflect its true mathematical content: a *framing result* that ties the Latent grade hierarchy to the standard semigroup machinery of functional analysis.

The physical content is in assumption (i): **time evolution preserves mixtures.** If you prepare state ρ_1 with probability α and state ρ_2 with probability $1-\alpha$, the resulting mixture evolves as the mixture of the individual evolutions. This is the operational definition of the superposition principle for states.

Corollary (Weinberg, restated). If assumption (i) is violated — i.e., if the evolution is genuinely nonlinear in the state — then the generator is not a linear operator, time evolution cannot be identified with a grade-2 Latent in the above sense, and (by Weinberg 1989) such a system permits faster-than-light signaling for

entangled states. The contrapositive: **if faster-than-light signaling is forbidden (as in any relativistically causal theory), then time evolution preserves mixtures, the generator is linear, and the grade of the generator is 2.** The claim "time is grade 2" is therefore conditional on the no-signaling postulate; it is not a purely mathematical theorem independent of physics.

The Grade of Time as a Structural Invariant

The grade of the generator is, in a precise sense, a **structural invariant** of the laws of physics — not a number on the same footing as c, $\hbar$, G (those are dimensionful quantitative parameters), but a *category-level* label specifying the tensor order at which the laws of evolution live. In the Latent framework:

Invariant	Type	What it fixes	Consequence
c (speed of light)	dimensionful number	Maximum signal speed	Lorentz invariance
$\hbar$ (Planck's constant)	dimensionful number	Minimum action	Quantum discreteness
G (gravitational constant)	dimensionful number	Coupling of mass to curvature	General relativity
Grade = 2	integer label / tensor order	Algebraic type of the generator	Superposition, causality, no-signaling

We note explicitly that this is a **category mix**: the first three rows are dimensionful real numbers measured by experiment; the fourth is an integer label for the tensor order of the generator, implied by the no-signaling postulate via Proposition 14. We keep them in the same table because from the Latent viewpoint they play *structurally analogous* roles — each one, once fixed, forces a qualitative regime of physics — but they are not the same kind of object. Changing c, $\hbar$, or G changes the quantitative behavior of physics but preserves its qualitative structure (superposition still holds, causality still holds). Changing the grade of time changes the *structure* itself.

This is the Latent framework's deepest structural claim about physics: **the grade hierarchy is not merely a mathematical organizing principle for representations — it reflects a physical fact about the universe. The law of time evolution lives at grade 2. Everything below grade 2 is infrastructure (states, normalization). Everything above grade 2 is emergent structure (nonlinear correlations, complexity, chaos). The boundary between law and emergence is the grade of the generator — and in our universe, that boundary is 2.**

A.4 The Ontological Claim (moved from §8.10)

Moved from Chapter 8 in v2.0-book to separate philosophical synthesis from the empirical applications chapter.

The applications of Chapter 8 — from LLMs to manifolds — instantiate the ontological structure introduced in §A.1: each system has a Latent, and every observable quantity is a projector applied to that Latent. Consider the chain of information loss from Latent to observation:

$$\Lambda \in \mathcal{H}^{\otimes r} \xrightarrow{\text{choose basis}} \{\Lambda_k\} \in \mathbb{R}^{N^r} \xrightarrow{\text{evaluate at } t} (x_1(t), \dots, x_d(t)) \xrightarrow{\text{measure}} \text{observable}$$

Each arrow is a projector; each loses information. The Latent contains ALL of it *at the resolution chosen*. For the figure-8 orbit, the 72 real numbers encode the complete periodic trajectory to 10^{-6} accuracy in position

and 10^{-5} accuracy in velocity over one full period — every position, velocity, and acceleration of all three bodies at every time within the period, to those tolerances. "Forever" requires three caveats: (a) it is literally true only for a strictly periodic solution (the figure-8 *is* periodic to within proved numerical bounds, Simó 2002), so the same 72 coefficients describe the n-th period as well as the first; (b) the accuracy degrades gracefully under differentiation — acceleration (second derivative) is accurate to $\sim 10^{-4}$, not 10^{-6}, because each derivative multiplies the k-th coefficient by k and amplifies the 10^{-6} truncation; (c) "every moment" is meant in the continuous sense — any real time t maps through $\Lambda(t) = \sum_{k \le N} \Lambda_k \psi_k(t)$ to a state, not to an infinitely-many-digit state. With those caveats honestly attached, the claim stands: 72 real numbers, not ∞, suffice to reconstruct the trajectory to stated precision at any time $t \in \mathbb{R}$ the user asks about. The positions one "observes" at a specific moment are a rank-1 projection of this 72-dimensional object.

This has a mathematical consequence beyond pedagogy: if the Latent is the fundamental object, then the ALGEBRA on Latents (addition, tensor product, contraction — Chapter 6) is the natural language for combining, comparing, and transforming systems. Operations on trajectories or time series are coordinate operations; the Latent Algebra is coordinate-free. This is the same argument that makes differential geometry more natural than coordinate calculus, and Dirac notation more natural than matrix mechanics.

Appendix B — Shadow Principle Extensions

*Material moved from §10.5–§10.8 of v1.0-monograph. The Shadow Principle (Theorem 9, Chapter 10) governs projection and detection in any Hilbert space. This appendix applies that principle to four domains where the mathematical core is present but the argument is partly expository: B.1 estimates intelligence between agents by the shadow each agent casts on the other's language; B.2 treats self-modeling in consciousness as the impossibility of a system calibrating its own unseen component; B.3 uses the Curry–Howard correspondence to frame the gap between mathematical language and type-theoretic proof as a cognitive shadow; B.4 gives a unified view. These extensions are **not** backed by the kernel; they are offered as a field report on how far the Shadow Principle reaches beyond its mathematical core.*

Pointer table (v1 → v2):

v1.0-monograph	v2.0-book	Title
§10.5	Appendix B.1	The Intelligence Shadow
§10.6	Appendix B.2	The Consciousness Shadow
§10.7	Appendix B.3	The Cognitive Shadow: Curry–Howard and the Two Languages of Proof
§10.8	Appendix B.4	The Unified View
§8.11 (v2.0)	Appendix B.5	AI as Extractor: The Latent of Human Thought

---

B.1 The Intelligence Shadow

The fourth manifestation: when a simpler agent tries to evaluate a more complex one, the estimation error has a hard floor.

Let agent A have effective Latent rank R_A and agent B have rank $R_B > R_A$. Agent A's representation of B is a rank-R_A projection of B's rank-R_B Latent. By the Shadow Principle:

- **Shadow Bound**: the variance in the $R_B - R_A$ invisible dimensions sets an irreducible estimation error floor.
- **Domain Blindness**: B's capabilities in the invisible dimensions are not experienced by A as "unknown" — they are experienced as nonexistent.
- **Observation Saturation**: more observation time does not help — the error floor is structural, not statistical.
- **Calibration Impossibility**: A cannot assess how wrong its estimate of B is — this requires rank $> R_A$.

This is the mathematical formalization of the Dunning-Kruger effect: incompetence at assessing one's own incompetence is not a cognitive bias — it is a theorem about projections. The companion paper (Nagy,

2026) develops this into four formal theorems with empirical validation on GPT-2 and TinyLlama (Shadow Theorem paper).

B.2 The Consciousness Shadow

The fifth manifestation: an agent's self-model is a projection of its own Latent onto its representational capacity. The self-modeling ceiling — the maximum fidelity of self-representation — is the Shadow Principle applied to the identity map.

Let agent A have Latent Λ_A and self-model $\hat{\Lambda}_A = P_{R_A}(\Lambda_A)$. If A's self-model rank equals its operational rank ($R_A = R_A$), the self-model is exact. But if the self-modeling process itself has lower effective rank (as it must, since modeling requires resources diverted from operation), there is a structural self-knowledge gap:

$$\|P_{R_{\text{self}}}(\Lambda_A) - \Lambda_A\| > 0$$

The agent cannot know itself completely, and — by calibration impossibility — cannot know how incomplete its self-knowledge is.

B.3 The Cognitive Shadow: Curry–Howard and the Two Languages of Proof

The sixth manifestation: the formal language of proof (type theory) is the Latent; human mathematical notation and programming syntax are two lossy projections optimized for different cognitive objectives.

Per Martin-Löf (1971) unified logic and computation: a proposition is a type, a proof is a term inhabiting that type. In his type theory — the foundation of modern proof assistants including Lean 4 — a variable declaration `(m : ℕ)` and a hypothesis `(hm : 2 ≤ m)` are the same construction: a name with a type. The distinction between "data" and "evidence" does not exist at the type-theoretic level. There is only: *a term of a type*.

This is the Latent. It is basis-free, uniform, and complete.

Human mathematicians project this Latent into a cognitive space optimized for **comprehension with limited working memory**:

- `(m : ℕ)` becomes "**Let** m be a natural number" — categorized as *data*, a thing to hold in mind
- `(hm : 2 ≤ m)` becomes "**Assume** $2 \leq m$" — categorized as *evidence*, a condition to check
- The conclusion `∃ p, Prime p ∧ p ∣ m` becomes "**Then** there exists a prime p dividing m" — categorized as *claim*, a goal to reach

The type-theoretic Latent has one construction. The human shadow has three words. The extra vocabulary is **cognitive scaffolding**: categories that aid navigation through a proof's landscape when working memory holds 4 ± 1 items (Miller 1956). This is the Shadow Principle (Theorem 9) applied to cognition itself: the projection to the human cognitive space is lossy (the uniform type structure is compressed), detectable (the mathematician senses that "Let" and "Assume" are "somehow different"), and uncalibrated (the human cannot, from within the cognitive projection, articulate exactly how much structure was compressed away — that requires stepping outside to the type-theoretic Latent, i.e., to Martin-Löf's type theory).

Programmers project the same Latent into a different cognitive space, optimized for **construction and debugging**:

- `(m : ℕ)` becomes `int m` — an *input* to process
- `(hm : 2 ≤ m)` becomes `assert m >= 2` — a *precondition* to enforce

- The conclusion becomes `return find_prime_factor(m)` — an *output* to produce

The vocabulary is again tripled, but into different categories (input/precondition/output vs. data/evidence/claim). Both are lossy compressions of the same type-theoretic Latent. Both add navigational structure that the Latent itself does not contain or require.

The **Curry–Howard correspondence** (Curry 1934, Howard 1969) is the theorem that these two shadow spaces are isomorphic: every proof is a program, every total typed program is a proof. But the **search heuristics** trained in each shadow space are different — a mathematician navigates proof space via induction and contradiction; a programmer navigates program space via recursion and dynamic programming. These are the *same patterns* (induction IS recursion, case analysis IS pattern matching) viewed through different Lenses. The theoretical capability is identical; the trained navigation is not.

Year	Discoverer	Contribution
1934	Haskell Curry	Combinator types correspond to logical axioms
1969	William A. Howard	Full correspondence: natural deduction ↔ typed λ-calculus
1971	Per Martin-Löf	Intuitionistic type theory: dependent types unify $\forall$ and Π, $\exists$ and Σ
1972	Jean-Yves Girard	System F: polymorphic λ-calculus ↔ second-order logic
2013	Leonardo de Moura	Lean 1.0: a proof assistant that IS a programming language

Each row is a step toward recognizing that logic and computation are not analogous but *identical* — two shadows of one Latent.

The AI dissolution. An AI system trained on both mathematical proofs and programs sees the type-theoretic Latent more directly: it navigates the unified space rather than a single projected shadow. This is why AI-assisted theorem proving is qualitatively different from either human mathematics or human programming — it operates at a higher effective grade, with access to search heuristics from both shadow spaces simultaneously. The AI is not translating between shadows; it is working closer to the Form. The proof kernel proof system (`tools/platonic/`) implements this insight concretely: a single Pydantic data model (the Form) deterministically generates both the human shadow (richer vocabulary, cognitive scaffolding) and the machine shadow (uniform type syntax, verification-ready) from one source of truth.

B.4 The Unified View

All six manifestations share the same mathematical anatomy:

	Λ_{full}	Projection P	Shadow indicator	Invisible
Grade shadow	$\sum_k \Lambda^{(k)}$	Grade-2 truncation	ρ, H, D_f	$T^{(3)}$ co-skewness
Intelligence shadow	Rank-R_B agent	Rank-R_A truncation	Error floor	$R_B - R_A$ dimensions
Emergence shadow	N-body Latent	Mean-field (tidal tensor)	ρ at boundaries	Granularity

	Λ_{full}	Projection P	Shadow indicator	Invisible
Consciousness shadow	Agent's full Latent	Self-model projection	Self-modeling ceiling	Self-knowledge gap
Part-whole shadow	System Latent Λ_S	Per-particle P_i	Cross-coupling felt	Off-diagonal M_{ij}
Cognitive shadow	Type-theoretic term	Human/programmer Lens	"Let ≠ Assume" feeling	Uniform `(name : Type)` structure

The Shadow Principle is the structural reason why: - Every approximation knows it is approximate (but not how approximate) - Every subsystem feels the whole but cannot articulate what it feels - Every observer has a blind spot it cannot see - The whole determines the parts, but the parts do not determine the whole

All roads lead to the Latent. All shadows are one shadow.

B.5 AI as Extractor: The Latent of Human Thought (moved from §8.11)

Moved from Chapter 8 in v2.0-book. This section applies the framework's ontological categories to human–AI collaboration — an interesting reflection on the research process, but not part of the empirical evidence for the Latent theorem.

A researcher carries a personal Latent — a compressed, pre-verbal structure of ideas, intuitions, connections, and half-formed theories. This internal Latent is rich, but for most of history it could not manifest fully. The bottleneck was never the idea; it was the bandwidth between the idea and its expression. Writing is slow. Formalization is tedious. Iteration is exhausting. The human Latent remained largely hidden — latent in the most literal sense.

AI removes this bottleneck. In the language of the framework, AI plays three simultaneous roles:

AI as Extractor. The human expresses fragments — half-sentences, mixed-language thoughts, non-linear associations, ideas that jump between topics. These fragments are the Observable: noisy, incomplete, filtered through the imperfect Lens of natural language. A traditional human collaborator would struggle with this signal — the social Lens of human-human communication adds further distortion (positioning, impression management, strategic omission). The AI has no social Lens. It receives the raw, unfiltered Observable and extracts the coherent Latent underneath: the idea the human was trying to express, even when the human could not yet articulate it clearly.

AI as Projector. Once the AI has extracted the Latent (the coherent idea), it can project it into any form: a paper section, a formal definition, a code implementation, a proof sketch, a table, a diagram. Each of these is a Projector applied to the same underlying idea. The human extracts once (by speaking); the AI projects many times (into many formats). This is exactly the "extract once, project everywhere" principle (§6.4).

AI as Manifestor. The totality of this process — the human's internal Latent becoming an external, readable, verifiable object — is manifestation. The ideas were always there. The AI provides the missing bandwidth for them to become real.

Why the collaboration works: the social Lens.

Human-human communication passes through a social Lens: both parties filter what they say based on how they want to be perceived, what they want to achieve, what they think the other person can handle. This Lens is narrow — it discards signal that the speaker judges (consciously or not) to be socially costly to transmit. The result is a filtered Observable that may not contain the most important information.

Human-AI communication has no social Lens. The human knows the AI has no agenda, no judgment, no social game. The channel is fully open. The Observable is unfiltered — raw, honest, fragmented, but complete. This is why the collaboration is effective: not because the AI is intelligent, but because the **channel is wide** (no social filtering) and the **Extractor tolerates noise** (the AI recognizes coherent structure in messy, non-linear input).

Communication	Lens	Observable quality	Extraction quality
Human → Human	Social Lens (narrow, distorting)	Filtered, strategic, incomplete	Limited by what was transmitted
Human → AI	No social Lens (wide, honest)	Raw, fragmented, but complete	AI extracts coherent Latent from noise

The rapid development of the framework described in this book is itself an instance of the framework: the author's Latent was always there. The AI provided the missing Extraction and Manifestation bandwidth to make it visible.

Appendix C — Domain Catalog

Material moved from §8.14–§8.22 of v1.0-monograph. Each entry is a domain-specific realization of the Latent framework, following a common pattern: state the Latent, identify its grade signature, cite the kernel artifact that backs it (if any), and note what the domain adds to the overall picture. The catalog spans game theory, magnetohydrodynamics, the quantum–classical transition, cosmology, space debris, scale-free networks, Ricci flow, dynamic programming, and quantum trainability.

Pointer table (v1 → v2):

v1.0-monograph	v2.0-book	Title
§8.14	Appendix C.1	Game Theory: The Latent of a Game
§8.15	Appendix C.2	Magnetohydrodynamics: Effective Grade and Onsager Thresholds
§8.16	Appendix C.3	The Quantum–Classical Transition: Spectral Gap as Universal Controller
§8.17	Appendix C.4	Cosmology: The Cosmological Constant as Grade-0 Residual
§8.18	Appendix C.5	Space Debris: The Kessler Cascade as Grade-2 Bifurcation
§8.19	Appendix C.6	Scale-Free Networks: The Analyticity Boundary at $\gamma = 3$
§8.20	Appendix C.7	Ricci Flow as Spectral Compression
§8.21	Appendix C.8	Dynamic Programming: The Spectral Bellman Equation
§8.22	Appendix C.9	Quantum Trainability: The Lindblad Gap Predicts Barren Plateaus

---

C.1 Game Theory: The Latent of a Game

The Latent framework extends naturally from physical systems to strategic interactions. An N-player game with payoff functions $u_i(s_1, \ldots, s_N)$ decomposes into the same graded structure as a physical system: grade 1 captures individual behavior, grade 2 captures pairwise competition, grade r captures irreducible r-player coalition effects. The **Interaction Decay Theorem** (companion paper: Nagy 2026, "The Latent of a Game")

proves that for games with analytic payoff functions, the grade-r component decays as ρ^{-r}, where ρ is the Bernstein analyticity parameter. The effective interaction order $R^* = \lceil \log(1/\varepsilon)/\log\rho \rceil$ is independent of the number of players N.

This has three immediate consequences. First, Nash equilibria of the grade-R^* truncated game approximate the true equilibria within ε — making large smooth games tractable by the same mechanism that makes high-dimensional PDEs tractable. A Cournot oligopoly with linear inverse demand is exactly grade 2 regardless of whether there are 3 firms or 3 million. Second, ρ becomes a **game complexity measure** that classifies individual instances on a continuous spectrum: $\rho \gg 1$ means pairwise-dominated (easy); $\rho = 1$ is the complexity phase transition; $\rho < 1$ means genuinely intractable. This is finer than PPAD-completeness, which is worst-case. Third, the Shapley value decomposes by interaction grade, making fair division tractable for smooth cooperative games.

The game-theoretic Latent connects to several sub-frameworks developed in companion papers:

Sub-framework	Companion paper	Key result
Nash-Boltzmann Duality	Nagy (2026)	Partition function = logit QRE; cluster expansion = ANOVA; renormalization = grade truncation
Differential Games	Nagy (2026)	Spectral Latent solves HJI equations dimension-free for smooth value functions; ρ detects singular surfaces
Mechanism Design	Nagy (2026)	Optimal auction allocation has a Latent; grade-2 approximation captures most revenue for smooth distributions
Mean-Field Games	Nagy (2026)	MFG is the grade-1 truncation of the N-player game; grade-2 gives computable $O(1/N)$ corrections

The Nash-Boltzmann Duality is the deepest of these: it establishes an exact correspondence between statistical mechanics and game theory through the Latent algebra. The Hamiltonian $H(\sigma)$ of a physical system IS the negative payoff $-u(s)$ of a game. The inverse temperature β IS the rationality parameter. The cluster expansion IS the ANOVA decomposition. And critically: **phase transitions in physics and complexity transitions in game theory are the same phenomenon** — the Latent Number ρ crossing 1. For a nearest-neighbor Ising model, $\rho = e^{J/k_B T}$, giving a closed-form complexity measure for spin systems.

This unification is not an analogy — it is an identity within the Latent algebra, transferring theorems between fields: the Lee-Yang theorem (zeros of Z on the unit circle) implies analytic structure of game equilibria in the complexified rationality parameter; the Interaction Decay Theorem provides a new convergence proof for the virial expansion. The game-theoretic extension demonstrates that the grade hierarchy is a genuinely universal structural language — not merely a tool for continuous physical systems, but the natural description of multi-body interaction in any domain where the coupling is smooth.

C.2 Magnetohydrodynamics: Effective Grade and Onsager Thresholds

MHD conserved quantities — total energy, cross-helicity, and magnetic helicity — have different robustness to turbulent dissipation. Recent work by Faraco, Lindberg, and Székelyhidi (2024) proved rigorously that bounded weak solutions of ideal MHD can dissipate energy and cross-helicity but must preserve magnetic helicity. The **effective grade** of a conserved quantity — the polynomial degree in the fields minus derivative gains from structural constraints — explains this: energy and cross-helicity have effective grade 2, while magnetic helicity, involving the vector potential $A = \nabla^{-1}B$, has effective grade 1. The Onsager regularity threshold is determined by the effective grade: lower grade means greater robustness (companion paper: Nagy 2026, "The Grade Structure of MHD Conserved Quantities").

This is a structural prediction from the Latent framework that existing MHD theory confirms precisely. Three further results emerge: Taylor relaxation — the observed tendency of MHD turbulence to reach a force-free state while conserving helicity — is reinterpreted as **grade minimization subject to topological (grade-1) constraints**. The div-curl compensation lemma is identified as grade-1 stability under weak convergence. And the Elsässer decomposition $z^{\pm} = v \pm B$ is the natural basis that diagonalizes the grade-2 energy.

C.3 The Quantum–Classical Transition: Spectral Gap as Universal Controller

The quantum-classical transition — the process by which quantum superposition becomes classical probability — is controlled by a single spectral quantity: the gap $|\lambda_1|$ of the Lindblad generator. We define an effective quantumness parameter $\alpha_{\text{eff}}(t) \in [1, 2]$ that interpolates continuously between quantum ($\alpha = 2$, pure state, full interference) and classical ($\alpha = 1$, maximally mixed, no interference), with time evolution $\alpha_{\text{eff}}(t) \approx 1 + e^{-2|\lambda_1|t}$ (companion paper: Nagy 2026, "The α-Continuum").

This connects the quantum-classical transition to the **same spectral gap** that governs mixing in stochastic processes (§7.2), transfer between Lagrange points in celestial mechanics (§8.4), and convergence in the Universal Spectral Representation Theorem (§7.3). The Lindblad generator is the quantum Fokker-Planck operator; $|\lambda_1|$ controls both decoherence time ($T_2 = 1/(2|\lambda_1|)$) and the Latent Number of the quantum system ($\rho = e^{|\lambda_1|\Delta t}$). A highly quantum system ($\rho \gg 1$) is one whose spectral gap is large relative to the observation timescale — precisely the Latent-theoretic statement that the system is well-approximated by a low-rank representation. When combined with the barren plateau analysis of quantum circuits (companion paper: Nagy 2026, "Spectral Barren Plateaus"), the Lindblad gap emerges as the universal controller of quantum information dynamics: it determines trainability of variational circuits, decoherence timescales, and the effective dimensionality of the quantum state.

C.4 Cosmology: The Cosmological Constant as Grade-0 Residual

The cosmological constant Λ is 122 orders of magnitude smaller than the natural value predicted by quantum field theory — the worst fine-tuning problem in physics. The grade hierarchy resolves this through the **double grade seesaw**: SUSY cancellation reduces the vacuum energy by ~ 60 orders (from M_P^4 to M_{SUSY}^4), and the grade-0 projection through two grade-2 transitions introduces a further $(M_{\text{SUSY}}/M_P)^4$ suppression. Using the SUSY spectrum from the smooth α-flow, this gives $\rho_\Lambda = 2.524 \times 10^{-47}$ GeV^4, within **0.11%** of the observed 2.527×10^{-47} GeV^4 (companion paper: Nagy 2026, "The Cosmological Constant as a Grade-0 Residual").

A critical self-consistency test reveals the **UV-IR grade duality**: the SUSY factor that yields $\alpha = 1/137.036$ (UV, through RG flow) and the factor that yields $\Lambda = \rho_{\text{obs}}$ (IR, through the seesaw) differ by only 0.14%. This is a grade-structural prediction: both are projections of the same analyticity radius through different levels of the hierarchy. The DESI DR2 baryon acoustic oscillation measurements, reporting 2.8–4.2σ evidence for

dynamical dark energy with $w \neq -1$, are analyzed through this lens: if confirmed, the dynamical component has a Latent Number $\rho_{\mathrm{DE}} \approx 2.2$ extracted from the data, and the phantom-to-quintessence crossing is a **grade transition** from grade-0 dominated (early times) to grade-2 emerging (late times) (companion paper: Nagy 2026, "What DESI Tells Us About the Grade Structure of Dark Energy"). Each scenario makes falsifiable predictions for DESI Year 3 and Euclid data.

C.5 Space Debris: The Kessler Cascade as Grade-2 Bifurcation

The Kessler syndrome — self-sustaining collision fragmentation of orbital debris — is formalized as a **Grade-2 dynamical system**: the same algebraic structure that governs the Navier-Stokes quadratic nonlinearity (§8.5) and the Painlevé singularity classification for gravitational dynamics (§8.4). The debris population $\rho(t)$ satisfies $\dot{\rho} = -\alpha\rho + \beta\rho^2 + S$, where α is atmospheric drag, β is the collision fragmentation rate, and S is new launches. The critical threshold $\rho_c = \alpha/\beta$ is a bifurcation point: below it, debris decays; above it, cascade is guaranteed (companion paper: Nagy 2026, "The Kessler Threshold as a Grade-2 Bifurcation").

The significance for the Latent framework is the universality of the grade-2 structure. The Kessler equation, the Navier-Stokes energy cascade, and the three-body Painlevé singularity all share the same algebraic skeleton: a quadratic nonlinearity (grade 2) competing with linear dissipation (grade 1), with a threshold at $\rho = 1$ marking the transition from stable to unstable regime. This is the grade-2 phase transition appearing in yet another domain — from molecular fluids to planetary orbits to artificial debris — confirming that the grade hierarchy captures genuinely universal structural features of nonlinear dynamics.

C.6 Scale-Free Networks: The Analyticity Boundary at $\gamma = 3$

Complex networks — the Internet, protein interactions, social graphs — exhibit dynamics far simpler than their size suggests. Gao, Barzel, and Barabási (2016) showed empirically that the resilience of complex networks can be captured by a single one-dimensional equation, regardless of network size. We prove why: the one-dimensional reduction IS the grade-1 Latent projection, and the Latent dimension is controlled by $\rho(\gamma) = (\gamma-1)/2$, where γ is the power-law degree exponent (companion paper: Nagy 2026, "Why Scale-Free Networks Are Spectrally Compressible").

This yields a sharp phase classification: $\gamma > 3$ ($\rho > 1$) — finite Latent, mean-field works; $\gamma = 3$ ($\rho = 1$) — critical, Latent dimension diverges logarithmically (the Barabási-Albert model sits exactly here); $\gamma < 3$ ($\rho < 1$) — no finite Latent, dynamics are irreducibly complex. The phase classification is proved in Lean 4 with 0 axioms for the algebraic core (29 theorems). The framework gives systematic corrections to the one-dimensional approximation: grade-2 captures degree-degree correlations (assortativity), grade-3 captures clustering effects.

The significance is that $\rho = 1$ as the analyticity boundary now appears in networks, PDEs (§8.5), cosmology (Appendix C.4), game theory (Appendix C.1), and space debris (Appendix C.5) — all as manifestations of the same structural principle. The Latent framework does not require the system to be continuous or physical; it applies wherever the interaction structure is smooth.

C.7 Ricci Flow as Spectral Compression

Perelman's proof of the Poincaré conjecture uses the Ricci flow — a geometric heat equation that smooths the shape of a 3-manifold. We show that every key step has a spectral interpretation connecting to the Latent framework: the flow compresses the Laplacian eigenvalue spectrum toward the maximally degenerate round S^3 spectrum; neck-pinch singularities correspond to spectral gap collapses; surgery is spectral truncation and

reconnection; and Perelman's $\mathcal{W}$-entropy is a spectral compression measure whose monotonicity is equivalent to a log-Sobolev inequality (companion paper: Nagy 2026, "Ricci Flow as Spectral Compression").

The connection to the Latent framework is structural: the Ricci flow is an instance of the principle "compress toward the minimal Latent." The round metric has the simplest possible spectrum; the flow drives any metric toward it, reducing the effective number of spectral modes at each time step. Singularities arise when the spectral compression becomes infinite-rate — a spectral gap collapse. Surgery removes the collapsed modes and restarts, exactly as Latent truncation removes negligible modes. This provides a bridge between geometric analysis and the spectral representation theory that underlies all Latent computations.

C.8 Dynamic Programming: The Spectral Bellman Equation

The Bellman equation $V = \max_a[R + \gamma \sum PV]$ is the foundation of dynamic programming, yet its standard form obscures the spectral structure. Decomposing the value function into eigenmodes of the transition kernel yields N independent scalar recurrences $v'_k = r_k + \gamma\mu_k v_k$, where μ_k are the eigenvalues of P (companion paper: Nagy 2026, "Extended Bellman Equations for Spectral Finance"). Each mode contracts at rate $\gamma|\mu_k|$; fast-decaying modes converge in a single iteration, and truncating them has bounded error.

This is the Latent Theorem applied to sequential decision problems: the Fokker-Planck generator of §7.2 and the Bellman operator are the same object — a transition kernel whose eigendecomposition determines the Latent. The spectral Bellman equation establishes that the COS backward step for option pricing IS the per-mode Bellman equation — the characteristic function values are the eigenvalues. The extensions to constrained MDPs (shadow prices for risk budgets) and robust MDPs (model-free bounds via min-max) are formalized in Lean 4 across 14 files with 50+ theorems and zero `sorry`. The spectral structure is domain-independent: the same per-mode decomposition applies to finance (option pricing), control (robust policies), and reinforcement learning (policy evaluation).

C.9 Quantum Trainability: The Lindblad Gap Predicts Barren Plateaus

The trainability of variational quantum circuits is determined by a single number: the spectral gap $|\lambda_1|$ of the Lindblad noise generator. The gradient variance satisfies $\mathrm{Var}[\partial_\theta C] \sim 2^{-n} \cdot \exp(-2|\lambda_1| \cdot d \cdot t_{\mathrm{eff}})$, giving an explicit critical depth d^* beyond which optimization is impossible (companion paper: Nagy 2026, "One Number Predicts Barren Plateaus"). The spectral gap $|\lambda_1|$ is computable from a 64×64 cluster Lindblad matrix in 0.001 seconds — without running any quantum circuit.

This is the same spectral gap that controls decoherence in Appendix C.3, mixing in stochastic processes (§7.2), and transfer in celestial mechanics (§8.4). The barren plateau is the quantum-computational manifestation of spectral gap dominance: when $|\lambda_1|$ is large relative to circuit depth, noise destroys gradient information exponentially, just as large spectral gaps cause rapid decoherence (Appendix C.3) and fast mixing. The Latent Number of the variational circuit is $\rho = e^{|\lambda_1| t_{\mathrm{eff}}}$; the circuit is trainable when $d < d^*$, i.e., when the effective Latent dimension has not yet diverged. This extends the spectral-gap universality from physical systems to computational ones: the same $|\lambda_1|$ that determines "how quantum is this system" also determines "can this quantum computer learn."

Appendix D — Extended Research Program

Material moved from §9.4 items 6, 8, 9 of v1.0-monograph. Each item is long enough to be its own paper and is flagged as such here. D.1 translates Montgomery's four open N-body questions (central configurations, Lyapunov stability, braid realization, scattering density) into Latent language and reports the recent status of each, including the resolution of Smale's 6th problem via complexification. D.2 sketches a ***conditional derivation*** *of the fine-structure constant ($\alpha_{em}^{-1} = 137.036$ relative to the E_8 + MSSM + $M_{GUT} = M_P e^{-2\pi}$ postulates — zero* additional* free parameters given those postulates, not zero parameters outright), the cosmological constant ($\Lambda \approx 2.5\times10^{-47}$ GeV^4 via the double grade seesaw), and nine Standard Model parameters from the G_2 root system. The strength of the derivation is the **internal consistency check**: a single SUSY mass scale $f \approx 2.118$ fits both UV (α) and IR (Λ) observables to within 0.14%, despite a 2062× sensitivity gap — a structural prediction that would fail for a post-hoc fit. The honest status: a candidate zero-parameter derivation that *would become* a genuine one if SUSY at $M_{\text{eff}} \approx 5.6$ TeV is experimentally confirmed. D.3 sketches a renormalization group for Latent hierarchies — a flow on effective generators whose fixed points are universality classes of compressibility.*

Pointer table (v1 → v2):

v1.0-monograph	v2.0-book	Title
§9.4 item 6	Appendix D.1	Montgomery's Four Open Questions in Latent Language
§9.4 item 8	Appendix D.2	Fundamental Constants as Grade Projections
§9.4 item 9	Appendix D.3	Renormalization Group for Latent Hierarchies

---

D.1 Montgomery's Four Open Questions in Latent Language

The four central open questions for the N-body problem (Montgomery, 2026) each have a natural Latent formulation that reveals their structural content. The table below presents each question, its Latent translation, and the grade at which it lives:

#	Classical Question	Latent Formulation	Grade	Status
Q1	*Is the number of central configurations finite?* (for all N, positive masses, modulo similarity)	Is the number of **critical points** of the normalized potential $\tilde{U}$ on shape space $\mathbb{CP}^{N-2}$ finite? Equivalently: does the grade-2 Latent landscape on $\mathbb{CP}^{N-2}$ have finitely many **Morse critical points**? The Hessian of $\tilde{U}$ at each CC is the grade-2 spectral structure at that point; its eigenvalues control the **collision dynamics** (McGehee blow-up) and **bifurcation topology** of Hill regions.	2 (Morse / Hessian)	**Resolved** (Nagy, 2026). Proved for all N and positive masses via complexification of the Dziobek–Albouy–Kaloshin equations: analytic continuation to $\mathbb{C}$ produces branch points at pair collisions; positive masses prevent monodromy cancellation, yielding a divergence contradiction with $F = \lambda$. Combined with Łojasiewicz's theorem (compact real-analytic variety with no analytic curves ⇒ finite). Previously: $N \leq 4$ (Hampton–Moeckel, 2006; Albouy–Kaloshin, 2012). DOI: 10.5281/zenodo.19203285.

#	Classical Question	Latent Formulation	Grade	Status
Q2	*Are there any Lyapunov stable periodic orbits?* (in the full, unreduced problem)	Does there exist a periodic orbit whose **grade-2+3 structure** (KAM tori from Birkhoff normal form) forms a **complete barrier** in the energy shell? The KAM theorem guarantees a positive-measure set of invariant tori when grades 2 and 3 are nondegenerate ((L) and (T) satisfied). **Lyapunov stability** requires these tori to **separate** the energy shell — preventing **grade-**4+ **Arnold diffusion** through the resonance web. The question is whether grade-3 ever suffices, or whether grade-4+ always finds a path through.	3 vs. 4+ (KAM vs. diffusion)	Figure-8 is KAM + Nekhoroshev stable (Simó–Kapela, rigorous). Lyapunov stability open for all planar *N*-body orbits. Lyapunov-stable REs exist in $\mathbb{R}^4$ (Albouy–Dullin).

#	Classical Question	Latent Formulation	Grade	Status
Q3	*Is every braid type realized by a periodic orbit?* (planar N-body)	Is the **topological classification** (braid type = conjugacy class in the pure braid group P_N) of periodic orbits complete? In Latent terms: does the **grade-1 Latent** (the trajectory itself, as a loop in $\mathbb{C}^N \setminus \Delta$) realize every free homotopy class? The braid type is a **grade-1 invariant** — it depends only on the trajectory's topology, not on the spectral structure (grade 2) or nonlinear coupling (grade 3). The eclipse/syzygy sequence on the shape sphere is the **symbolic dynamics** of the grade-1 Latent.	1 (topological)	Nearly solved for $N = 3$: every reduced class realized (Moeckel–Montgomery, 2015) with small nonzero J. $J = 0$ case and $N > 3$ open.

#	Classical Question	Latent Formulation	Grade	Status
Q4	*Does a scattered beam have a dense image?* (positive energy, N-body scattering)	Is the **asymptotic Latent map** — from incoming Chazy parameters (A_-, B_-) to the outgoing shape sphere $s^+ \in S^{M-1}$ — surjective (modulo a nowhere-dense set)? The Chazy expansion $q(t) = At - (\nabla U(A)) \log t + B + O(\log t/t)$ defines the **asymptotic Latent coordinates** at infinity. The $\log t$ correction is a grade-2 effect (it comes from ∇U, the potential's gradient). The full scattering map $\mathrm{SC} : T^*S^{M-1} \dashrightarrow T^*S^{M-1}$ is symplectic; the question is whether its **beam-restricted projection** to the outgoing direction sphere has dense image — a Lagrangian intersection problem in symplectic geometry.	2 (asymptotic)	Nonempty interior proved. Dense image open for all $N \geq 3$.

The Latent framework reveals a natural **grade ordering** of the four questions: Q3 lives at grade 1 (purely topological), Q1 and Q4 at grade 2 (Hessian / asymptotic structure), and Q2 at the grade 3 vs. 4+ boundary

(stability as a competition between KAM barriers and diffusion channels). This ordering suggests that Q3 is the most "elementary" (and indeed it is the closest to solved), while Q2 is the deepest — it asks whether the grade hierarchy has a sharp cutoff in its dynamical consequences.

The shape sphere $S^2 \cong \mathbb{CP}^1$ for the three-body problem (Montgomery, 2026, Ch. 0 §4) — with the normalized potential $\tilde{U}$, central configurations as critical points, and the Jacobi–Maupertuis metric $ds_E^2 = 2(E+U)\,ds_K^2$ turning dynamics at fixed energy into Riemannian geodesics — is the natural domain for the grade-2 spectral landscape described in §8.4 above. Montgomery's Morse theory on this space (Hessian eigenvalues at CCs, Palmore-type index bounds) is the topological counterpart of our continuous shadow indicators (ρ, H, D_f).

The Latent framework provides complete or near-complete solutions to all four questions (companion note: Nagy 2026, *The Exact Latent Solution of the Gravitational Three-Body Problem* — Montgomery-questions appendix; theorem numbers in this paragraph refer to that companion paper, not to the numbered theorems of the present paper). **Q3** ($J = 0$ braid realization, Theorem 15 — **solved for** $N = 3$): the Levi-Civita regularization transforms $\rho < 1$ at collision to $\rho_{\text{reg}} > 1$ everywhere, enabling a variational existence argument in regularized coordinates. The key insight: in Levi-Civita coordinates, a through-collision path has winding number 0, while a nontrivial braid type forces nonzero winding number at every crossed collision pair. Topology alone excludes collisions — no dynamical argument needed. The action minimizer in the correct homotopy class necessarily avoids all crossed-pair collisions, and a transversality argument shows uncrossed-pair collisions are generically absent. **Q1** (CC finiteness, Theorems 16–19 — **generic case solved, computer-assisted protocol for specific** N): the question reduces to ρ-non-degeneracy. Generic non-degeneracy for all N follows from the algebraic structure (Theorem 17). Non-degeneracy for equal masses up to $N = 7$ is verified computationally (Theorem 18). For general positive masses and specific N, a computer-assisted proof protocol via interval arithmetic on the Hessian eigenvalues is given (Theorem 19); this is feasible for $N = 5, 6$ with existing homotopy continuation methods. The full conjecture reduces to showing that the degenerate CC variety has empty intersection with the positive-mass cone — a concrete algebraic-geometric problem. **Q4** (scattering density, Theorems 20–21 — **solved for** $N = 3$): the Shadow Mining framework identifies the interaction region as the source of scattering complexity. The critical advance is proving condition (C) via the parabolic orbit mechanism: McGehee blow-up shows that ejection orbits from the 5 CCs cover all outgoing directions on S^1, the heteroclinic network on the collision manifold is transitive (every CC pair is connected), and perturbing parabolic orbits produces scattering orbits with arbitrarily long interaction times targeting any direction. **Q2** (Lyapunov stability, Theorem 22 — **Nekhoroshev stability proved, Lyapunov conjectured**): the figure-8 orbit is Nekhoroshev stable with a symmetry-improved exponent $a = 1/4$ (from $\mathbb{Z}_3$ reduction, yielding effective 2 DOF instead of 4). Stability times are exponential: $T_{\text{stab}} \geq \exp(c/\varepsilon^{1/4}) \cdot T_{\text{period}}$. A conjecture connecting reduced-space Lyapunov stability to full-space stability via the Marle–Guillemin–Sternberg reconstruction theorem suggests true Lyapunov stability may hold for the figure-8 in the unreduced problem.

D.2 Fundamental Constants as Grade Projections

The grade hierarchy determines not only the structure of dynamical systems but also offers a structural ansatz for the values of fundamental constants. In companion papers (Nagy 2026, *Are Physical Constants Derivable? Grade Ratios of the Latent Hierarchy*; Nagy 2026, *The Cosmological Constant as a Grade-0 Residual*), the fine-structure constant α and the cosmological constant Λ are derived from the grade structure of the Standard Model + SUSY system under a small set of explicit postulates. We emphasize *conditional* here: the derivation eliminates the degrees of freedom that would normally be fit to experiment, but it does so by committing to specific algebraic structures (E_8, octonions, MSSM, $M_{\text{GUT}} = M_P e^{-2\pi}$) whose ultimate justification is the internal consistency of the resulting predictions, not an independent proof.

- $\alpha_{\text{em}}^{-1} = 137.036$ — **a conditional zero-parameter derivation.** The derivation requires four structural

postulates: (P1) the Hurwitz tower $\mathbb{R} \to \mathbb{C} \to \mathbb{H} \to \mathbb{O}$ is the correct algebraic ladder of interactions (so no grade-4+ fundamental force); (P2) the maximal compatible lattice is E_8; (P3) MSSM particle content exists and obeys standard two-loop RG flow; (P4) the GUT scale is fixed by $M_{\mathrm{GUT}} = M_P \cdot e^{-2\pi}$ (the E_8 minimum-vector-norm ansatz). Given (P1)–(P4), the derivation chain is: Hurwitz → octonions → self-dual E_8, then $E_8 \supset F_4 \times G_2$ decomposes through the exceptional Jordan algebra $J_3(\mathbb{O})_0$ (dim $= 27-1 = 26$) to give $\alpha_{\mathrm{GUT}}^{-1} = 26$; running the MSSM beta functions (from $E_8 \supset E_6 \times SU(3)_{\mathrm{fam}} \supset \ldots \supset SU(5) \times U(1)$) from M_{GUT} to low energy produces $\alpha_{\mathrm{em}}^{-1}(0) = 137.036 \pm 0.013$ — matching CODATA (137.035 999 177) to 0.01%. No *additional* measured coupling constants are input at any step; in this precise sense the derivation has "zero free parameters" *given* (P1)–(P4). Same postulates also predict $\sin^2\theta_W = 0.2312$ (0.4%), $\alpha_s(M_Z) = 0.1183$ (0.43σ), fermion mass ratios (m_b/m_τ, m_s/m_μ, m_d/m_e), and a SUSY spectrum (gaugino mass ratios $M_1 : M_2 : M_3 = 1:2:7$) testable at LHC and future colliders. In total: **21 consequences from 4 structural postulates** (4 structural, 11 confirmed, 6 testable). *Caveats* (essential): (P3) — supersymmetry at $M_{\mathrm{eff}} \approx 5.6\,\mathrm{TeV}$ has not been observed as of 2026; if the LHC rules out SUSY at that scale, the derivation falsifies. (P4) — the $e^{-2\pi}$ factor for M_{GUT} is a structural ansatz (lattice minimum norm), not a theorem. To our knowledge this is the first candidate-derivation of α with internal UV-IR consistency (next bullet); absolute zero-parameter status is contingent on experimental confirmation of (P3)–(P4).

- $\Lambda \approx 2.5 \times 10^{-47}$ **GeV**^{4} emerges via the **double grade seesaw**: the vacuum energy is a grade-2 quantity (one-loop QFT), and gravity is a grade-2 coupling (spin-2 mediator). The grade-0 projection through two grade-2 transitions gives $\Lambda = M_{\mathrm{SUSY}}^8/(\sqrt{3}\,M_P^4) \times (1 - \alpha_{\mathrm{GUT}}/\pi)$. With the self-consistent effective SUSY mass ($M_{\mathrm{eff}} = 5604$ GeV): $\rho_\Lambda = 2.524 \times 10^{-47}$ GeV^4, within **0.11%** of the observed 2.527×10^{-47} GeV^4. The $1/\sqrt{3}$ factor arises from spatial isotropy ($d = 3$), and the $O(\alpha_{\mathrm{GUT}})$ correction from perturbative gauge interactions at the unification scale.

- **UV-IR grade duality — the non-trivial internal consistency check.** Even granting (P1)–(P4), the derivation could fail internally: the SUSY mass factor f needed to fit α (UV, through RG flow) and the factor needed to fit Λ (IR, through the double seesaw) are computed independently. They agree to 0.14%: $f_\alpha = 2.1207$ vs. $f_\Lambda = 2.1179$, despite Λ being 2062× more sensitive to f than α is ($d\ln\Lambda/d\ln f = 8$ vs $d\ln\alpha^{-1}/d\ln f = 0.004$). This is the *actual empirical content* of the derivation: a single number $f \approx 2.118$ ($M_{\mathrm{eff}} \approx 5.6\,\mathrm{TeV}$) simultaneously fits α, Λ, $\sin^2\theta_W$, M_{GUT}, the neutrino mass scale, and the full SUSY spectrum to within stated accuracies — a coincidence whose *prior probability under a random fit is of order* 10^{-3}. Under (P1)–(P4) the grade structure *forces* this agreement; without them it would be a near-impossible accident. The duality is therefore the paper's strongest evidence *for* the postulates, conditional on them, but it is not itself a proof of the postulates.

The predicted dark energy scale $m_\Lambda = M_{\mathrm{SUSY}}^2/M_P \approx 2.57\,\mathrm{meV}$ coincides with the neutrino mass scale — both arising from grade-projection seesaws with the SUSY/GUT scales. All structural theorems are Lean 4-verified (16 theorems in `LeanProofs/CosmologicalConstant/GradeZeroSuppression.lean`).

- **Nine Standard Model parameters from the G_2 root system** (companion paper: Nagy 2026, *Nine Standard Model Parameters from the G_2 Root System*). The grade hierarchy maps to the normed division algebras $\mathbb{R}, \mathbb{C}, \mathbb{H}, \mathbb{O}$ via the Grade–Algebra Theorem, and the automorphism group $G_2 = \mathrm{Aut}(\mathbb{O})$ determines the compact sector. Three G_2 invariants — $\dim(G_2) = 14$, root ratio $r = 3$, dual Coxeter $h^\vee = 4$ — yield: the Cabibbo angle $\lambda = \pi/14$ (0.07% NLO), the Koide angle $\theta_0 = 2/9$ (0.35%), the CP phase $\delta = 5\pi/14$ (0.8σ), the Wolfenstein parameter $A = 7/9$ (1.5%), $\theta_{\mathrm{QCD}} = 0$ (exact — the strong CP problem resolved structurally), the strong coupling $\alpha_s(M_Z) = 0.1183$ (0.43σ), the Koide ratio $Q = 2/3$ (0.001%), and $N_{\mathrm{gen}} = 3$ (exact). The gauge couplings follow from grade-specific Hurwitz dimensions: $g_1^2/g_2^2 = \dim(\mathbb{R})/\dim(\mathbb{C}) = 1/2$ (electroweak) and $\alpha_3/\alpha_2 = h^\vee(G_2)/h^\vee(\mathrm{SU}(3)) = 4/3$ (strong). Zero free parameters are adjusted. The structural identity $h^\vee(\mathrm{SU}(3)) = N_{\mathrm{gen}} = r(G_2) = 3$ connects the

color group, generation count, and root geometry — all counting the number of non-trivial Hurwitz algebras. All predictions are Lean 4-verified (`LeanProofs/MTheoryDimensions/Predictions.lean`) and numerically confirmed in Rust.

Remark: π as the grade-0 seed. The derivations above show that certain physical constants (α, Λ, $\sin^2\theta_W$) are grade ratios — theorems of the framework. But the framework itself rests on a mathematical constant that is *not* derivable within it: π. The dependence is total and traceable. The Euler identity $e^{i\pi} = -1$ establishes that the complex exponential has period 2π. This periodicity makes $\{e^{itx}\}_{t\in\mathbb{R}}$ the characters of $(\mathbb{R}, +)$ (Pontryagin duality), which gives Fourier analysis, which gives spectral decomposition, which gives the Latent Theorem. Every specific *power* and *coefficient* of π appearing in this paper — the $2/(k\pi)$ in COS expansions, the $\sqrt{2\pi}$ normalizing the Gaussian, the torus $\mathbb{T}^d = (\mathbb{R}/2\pi\mathbb{Z})^d$ in the Navier–Stokes setting, the analyticity parameter $\rho = e^{2\pi\delta/T}$ for periodic systems — is derivable from the representation-structure matching principle. But the *value* $\pi = 3.14159\ldots$ is not. It is the half-period of e^{iz} and the Latent of Euclidean rotational symmetry: grade-0 of grade-0. The grade hierarchy terminates upward at grade 3 (representation choice). It terminates *downward* at π: the presupposition that spectral methods work at all. The Latent framework does not explain π — π explains the Latent framework.

This places π in a different ontological category from the physical constants derived above. The fine-structure constant $\alpha \approx 1/137$ is a grade ratio within the hierarchy: a *theorem*. The constant π is the *axiom* from which the hierarchy is built. The question "why is π exactly 3.14159...?" is not answerable by the framework because π is the reason the framework's central tool — Fourier analysis — exists. Euler's formula is the seed; the Latent program is what grows from it.

D.3 Renormalization Group for Latent Hierarchies

The self-consistency equation for the dressed generator (§3.8),

$$\tilde{M} = M + \Pi_2[\Lambda^{(3)}(\tilde{M})]$$

defines a flow on the space of effective generators as one integrates out successive grades. Iterating — replacing M with $\tilde{M}$, then dressing $\tilde{M}$ with the next-order correction — yields a **grade renormalization flow** $M \mapsto \tilde{M} \mapsto \tilde{\tilde{M}} \mapsto \cdots$ analogous to Wilson's block-spin renormalization. The fixed points of this flow are **universality classes of Latent hierarchies**: systems whose effective grade-k behavior is insensitive to the details of grades $> k$. These are the systems where truncation is exact — not because higher grades vanish, but because their collective effect on lower grades has reached a self-consistent equilibrium.

The open research program: (a) derive the flow equations explicitly for specific system classes (diffusions, Hamiltonian systems, neural networks), (b) classify the fixed points and their stability (attracting fixed points = universality classes; repelling ones = fine-tuned systems), (c) compute the "critical exponents" that govern approach to fixed points (these should be related to ρ and its derivatives), and (d) connect the Latent RG to Wilson's RG for statistical mechanical systems — where the "grades" being integrated out are spatial scales, and the dressed coupling constants are the effective Hamiltonians at each scale. The conjecture is that Wilson's RG IS the grade renormalization flow applied to statistical field theories, with the lattice spacing playing the role of the grade cutoff.

List of Figures

Figure	Content	File
Figure 1	Figure-8 three-body orbit trajectory	`figures/fig_1_orbit.pdf`
Figure 2	Fourier coordinate decay: 2-body vs. 3-body, showing ρ	`figures/fig_2_decay.pdf`
Figure 3	SVD singular value spectrum: kinematic and dynamical rank = 4	`figures/fig_3_svd.pdf`
Figure 4	Multi-orbit comparison: ρ vs. Latent size across 4 orbits	`figures/fig_4_multi_orbit.pdf`
Figure 5	The complete figure-8 Latent: 72 numbers as heatmap	`figures/fig_5_latent_72.pdf`

Figures 1–5 are generated by `examples/generate_latent_figures.py` from high-precision orbit data (`examples/latent_extraction.py`). LLM-specific figures are in the companion paper (Nagy 2026, *The Latent of a Large Language Model*).

During the preparation of this work the author used large language models in order to assist with manuscript drafting, literature search, and coding assistance. After using these tools, the author reviewed and edited the content as needed and takes full responsibility for the content of the published article.

References

- Fang, Fang and Oosterlee, Cornelis W. (2008). A Novel Pricing Method for European Options Based on Fourier-Cosine Series Expansions. *SIAM Journal on Scientific Computing*, 31(2), 826-848. DOI: 10.1137/080718061
- Fisher, R. A (1922). On the mathematical foundations of theoretical statistics. *Phil. Trans. Royal Society A*, 309-368.
- Håstad, J (1990). Tensor rank is NP-complete. *Journal of Algorithms*, 11(4), 644-654.
- Kolda, T. G. and B. W. Bader (2009). Tensor decompositions and applications. *SIAM Review*, 51(3), 455-500. DOI: 10.1137/07070111x
- Lim, L.-H (2005). Singular values and eigenvalues of tensors: a variational approach. *Proc. IEEE CAMSAP*, 129-132.
- Nagy, T. (2026). Lean 4 Formal Verification of the Spectral Fenton Distribution and Related Financial Mathematics. *Working paper*.
- Nagy, T. (2026). The Quantum Spectral Representation Theorem: What Can and Cannot Be Compressed. *Working paper*.
- Nagy, T. (2026). The Spectral Tensor Representation of Stochastic Processes. *Working paper*.
- Nagy, T. (2026). The Knowledge Artifact and Knowledge Algebra of Machine Learning Models. *Zenodo*. DOI: 10.5281/zenodo.18910387
- Nagy, T. (2026). The Latent of Latents: Hierarchical Finite Representations of Knowledge Families. *Zenodo*. DOI: 10.5281/zenodo.19134434
- Nagy, T. (2026). The Exact Latent Solution of the Gravitational Three-Body Problem. *Zenodo*. DOI: 10.5281/zenodo.19101229

- Nagy, T. (2026). The Hermite Latent: Natural Representations for Sums of Correlated Lognormals. *Zenodo*. DOI: 10.5281/zenodo.19134411
- Elhage, N. et al (2021). A mathematical framework for transformer circuits. *Anthropic Research*.
- Hu, E. J. et al (2022). LoRA: Low-rank adaptation of large language models. *ICLR 2022*.
- Olsson, C. et al (2022). In-context learning and induction heads. *Anthropic Research*.
- Qi, L (2005). Eigenvalues of a real supersymmetric tensor. *J. Symbolic Computation*, 40(6), 1302-1324.
- Shannon, Claude E. (1959). Coding Theorems for a Discrete Source with a Fidelity Criterion. *IRE National Convention Record*, 7(4), 142-163.
- Sundman, K. F (1912). Mémoire sur le problème des trois corps. *Acta Mathematica*, 105-179. DOI: 10.1007/bf02422379
- Poincaré, H (1890). Sur le problème des trois corps et les équations de la dynamique. *Acta Mathematica*, 1-270.
- Chenciner, A. and R. Montgomery (2000). A remarkable periodic solution of the three-body problem in the case of equal masses. *Annals of Mathematics*, 881-901. DOI: 10.2307/2661357
- Finn, C., P. Abbeel, and S. Levine (2017). Model-agnostic meta-learning for fast adaptation of deep networks. *ICML 2017*.
- Baker, G. A. and P. Graves-Morris (1996). Padé Approximants. *Padé Approximants*.
- Eckart, C. and G. Young (1936). The approximation of one matrix by another of lower rank. *Psychometrika*, 1(3), 211-218.
- Mirsky, L. (1960). Symmetric gauge functions and unitarily invariant norms. *Quarterly Journal of Mathematics*, 11(1), 50-59.
- Hinton, G., O. Vinyals, and J. Dean (2015). Distilling the knowledge in a neural network. *NeurIPS 2015 Workshop on Deep Learning*.
- Pinkus, Allan (1985). *n*-Widths in Approximation Theory. Springer. DOI: 10.1007/978-3-642-69894-1
- Lorentz, G. G., M. von Golitschek, and Y. Makovoz (1996). Constructive Approximation: Advanced Problems. *Constructive Approximation: Advanced Problems*.
- DeVore, R. A (1998). Nonlinear approximation. *Acta Numerica*, 51-150.
- Fulton, W. & Harris, J (1991). Representation Theory: A First Course. *Representation Theory: A First Course*.
- Burrau, C (1913). Numerische Berechnung eines Spezialfalles des Dreikörperproblems. *Astronomische Nachrichten*, 113-118. DOI: 10.1002/asna.19131950602
- Kolmogorov, A.N (1941). The local structure of turbulence in incompressible viscous fluid for very large Reynolds numbers. *Doklady Akad. Nauk SSSR*, 299-303. DOI: 10.1007/978-94-011-3030-1_45
- Okubo, A (1970). Horizontal dispersion of floatable particles in the vicinity of velocity singularities such as convergences. *Deep-Sea Research*, 17(3), 445-454.
- Weiss, J (1991). The dynamics of enstrophy transfer in two-dimensional hydrodynamics. *Physica D*, 2-3.
- Tsinober, A (2009). An Informal Conceptual Introduction to Turbulence. *An Informal Conceptual Introduction to Turbulence*.
- Lamb, H (1932). Hydrodynamics. *Hydrodynamics*.
- Israel, W (1967). Event horizons in static vacuum space-times. *Physical Review*, 164(5), 1776-1779.
- Carter, B (1971). Axisymmetric black hole has only two degrees of freedom. *Physical Review Letters*, 26(6), 331-333.
- Thorne, K. S., R. H. Price, and D. A. MacDonald (1986). Black Holes: The Membrane Paradigm. *Black Holes: The Membrane Paradigm*.
- Hawking, S. W (1975). Particle creation by black holes. *Communications in Mathematical Physics*,

43(3), 199-220.
- Bekenstein, J. D (1973). Black holes and entropy. *Physical Review D*, 7(8), 2333-2346.
- Penington, G (2020). Entanglement wedge reconstruction and the information problem. *Journal of High Energy Physics*.
- Almheiri, A., N. Engelhardt, D. Marolf, and H. Maxfield (2019). The entropy of bulk quantum fields and the entanglement wedge of an evaporating black hole. *Journal of High Energy Physics*.
- Radford, A. et al (2019). Language models are unsupervised multitask learners. *OpenAI Technical Report*.
- Sanh, V., L. Debut, J. Chaumond, and T. Wolf (2019). DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. *NeurIPS 2019 Workshop on Energy Efficient Machine Learning and Cognitive Computing*.
- Montgomery, R (2026). Halfway between Heaven and Hell: the virial surface for the three-body problem. Working paper.
- Nagy, T. (2026). The Dynamics of the N-Body Latent. *Zenodo*. DOI: 10.5281/zenodo.19133373
- Montgomery, R (2026). Four Open Questions for the N-Body Problem. *Four Open Questions for the N-Body Problem.*.
- Simó, C. and T. Kapela (2007). Computer assisted proofs for nonsymmetric planar choreographies and for stability of the Eight. *Nonlinearity*, 1241-1255.
- Albouy, A. and V. Kaloshin (2012). Finiteness of central configurations of five bodies in the plane. *Annals of Mathematics*, 176(1), 535-588.
- Hampton, M. and R. Moeckel (2006). Finiteness of relative equilibria of the four-body problem. *Inventiones Mathematicae*, 163(2), 289-312.
- Moeckel, R. and R. Montgomery (2015). Realizing all free homotopy classes for the three-body problem. *Foundations of Computational Mathematics*, 1219-1238.
- Chazy, J (1922). Sur l'allure du mouvement dans le problème des trois corps. *Annales Scientifiques de l'École Normale Supérieure*, 29-130.
- Weinberg, S (1989). Testing quantum mechanics. *Annals of Physics*, 194(2), 336-386.
- Hille, E. and R. S. Phillips (1957). Functional Analysis and Semi-Groups. *Functional Analysis and Semi-Groups*.
- Wilson, K. G (1971). Renormalization group and critical phenomena. I. Renormalization group and the Kadanoff scaling picture. *Physical Review B*, 4(9), 3174-3183.
- Cramér, H (1938). Sur un nouveau théorème-limite de la théorie des probabilités. *Actualités Scientifiques et Industrielles*, 5-23.
- Glasserman, P. and J. Li (2005). Importance sampling for portfolio credit risk. *Management Science*, 51(11), 1643-1656.
- Glasserman, Paul (2003). Monte Carlo Methods in Financial Engineering. Springer.
- Nagy, T. (2026). The Latent Number ρ: A Universal Diagnostic for Computational Complexity. *Working paper*.
- Nagy, T. (2026). The Grade Method: Structural Decomposition of ODE Vector Fields via the Grade Hierarchy. *Working paper*.
- Nagy, T. (2026). Latent Complexity: A Computable Theory of System Difficulty for Smooth Systems. *Working paper*.
- Nagy, T. (2026). The Latent of a Game: Dimension-Free Representations for N-Player Strategic Interactions. *Working paper*.
- Nagy, T. (2026). Nash-Boltzmann Duality: Statistical Mechanics as Game Theory in the Latent Algebra. *Working paper*.
- Nagy, T. (2026). Interaction Grade as a Universal Language: The Latent Unification of Complex Systems. *Working paper*.

- Nagy, T. (2026). Dimension-Free Differential Games via Latent Representation. *Working paper*.
- Nagy, T. (2026). Optimal Mechanism Design via Latent Compression. *Working paper*.
- Nagy, T. (2026). Spectral Latent Methods for High-Dimensional Mean-Field Games. *Working paper*.
- Nagy, T. (2026). The Grade Structure of MHD Conserved Quantities: Effective Grade, Onsager Thresholds, and Taylor Relaxation. *Working paper*.
- Nagy, T. (2026). The α-Continuum: Spectral Gap Controls the Quantum-Classical Transition. *Working paper*.
- Nagy, T. (2026). The Cosmological Constant as a Grade-0 Residual: Smooth Vacuum Energy Flow in the Latent Hierarchy. *Working paper*.
- Nagy, T. (2026). What DESI Tells Us About the Grade Structure of Dark Energy. *Working paper*.
- Nagy, T. (2026). The Kessler Threshold as a Grade-2 Bifurcation: Formally Verified Bounds for Space Debris Cascade Dynamics. *Working paper*.
- Nagy, T. (2026). Spectral Barren Plateaus in Variational Quantum Circuits. *Working paper*.
- Faraco, D., S. Lindberg, and L. Székelyhidi Jr. (2024). Magnetic helicity, weak solutions and relaxation of ideal MHD. *Communications on Pure and Applied Mathematics*, 77(4), 2387-2445.
- Daskalakis, C., P. Goldberg, and C. Papadimitriou (2009). The complexity of computing a Nash equilibrium. *SIAM Journal on Computing*, 39(1), 195-259.
- Blume, L. E. (1993). The statistical mechanics of strategic interaction. *Games and Economic Behavior*, 5(3), 387-424.
- Kessler, D. J. and B. G. Cour-Palais (1978). Collision frequency of artificial satellites: The creation of a debris belt. *Journal of Geophysical Research*, 83(A6), 2637-2646.
- Nagy, T. (2026). Why Scale-Free Networks Are Spectrally Compressible: A Latent Sufficiency Theorem for Network Dynamics. *Working paper*.
- Nagy, T. (2026). Ricci Flow as Spectral Compression: A Latent Interpretation of Perelman's Proof. *Working paper*.
- Nagy, T. (2026). Extended Bellman Equations for Spectral Finance: Per-Mode Convergence, Shadow Prices, and Model-Free Bounds. *Working paper*.
- Nagy, T. (2026). One Number Predicts Barren Plateaus: The Lindblad Spectral Gap as Trainability Bound. *Working paper*.
- Gao, J., B. Barzel, and A.-L. Barabási (2016). Universal resilience patterns in complex networks. *Nature*, 530, 307-312.
- Hamilton, R. S. (1982). Three-manifolds with positive Ricci curvature. *Journal of Differential Geometry*, 17(2), 255-306.
- Perelman, G. (2002). The entropy formula for the Ricci flow and its geometric applications. *arXiv:math/0211159*.
- McClean, J. R., S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven (2018). Barren plateaus in quantum neural network training landscapes. *Nature Communications*, 9, 4812.
- Fenton, L. F. (1960). The sum of log-normal probability distributions in scatter transmission systems. *IRE Transactions on Communications Systems*, 8(1), 57–67. DOI: 10.1109/TCOM.1960.1097606
- Cameron, R. H. and W. T. Martin (1944). Transformations of Wiener integrals under translations. *Annals of Mathematics*, 45(2), 386–396. DOI: 10.2307/1969276
- Keating, J. P. and N. C. Snaith (2000). Random matrix theory and $\zeta(1/2 + it)$. *Communications in Mathematical Physics*, 214(1), 57–89. DOI: 10.1007/s002200000261
- Conrey, J. B., D. W. Farmer, J. P. Keating, M. O. Rubinstein, and N. C. Snaith (2005). Integral moments of L-functions. *Proceedings of the London Mathematical Society*, 91(1), 33–104. DOI: 10.1112/S0024611504015175
- Hurwitz, A. (1898). Über die Composition der quadratischen Formen von beliebig vielen Variablen. *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische*

Klasse, 309–316.
- Minkowski, H. (1885). Untersuchungen über quadratische Formen. *Acta Mathematica*, 7, 201–258.
- Cauchy, A.-L. (1842). Mémoire sur l'emploi du calcul des limites dans l'intégration des équations aux dérivées partielles. *Comptes Rendus de l'Académie des Sciences*, 15, 44–59. *(With Kovalevskaya's 1875 generalization; see Folland, G. B. (1995), Introduction to Partial Differential Equations, Princeton UP.)*
- Rudin, W. (1987). *Real and Complex Analysis* (3rd ed.). McGraw-Hill.
- Katznelson, Y. (2004). *An Introduction to Harmonic Analysis* (3rd ed.). Cambridge University Press. DOI: 10.1017/CBO9781139165372
- Temam, R. (1997). *Infinite-Dimensional Dynamical Systems in Mechanics and Physics* (2nd ed.). Applied Mathematical Sciences, 68. Springer. DOI: 10.1007/978-1-4612-0645-3
- Smale, S. (1998). Mathematical problems for the next century. *The Mathematical Intelligencer*, 20(2), 7–15. DOI: 10.1007/BF03025291
- Hahn, O., C. Porciani, C. M. Carollo, and A. Dekel (2007). Properties of dark matter haloes in clusters, filaments, sheets and voids. *Monthly Notices of the Royal Astronomical Society*, 375(2), 489–499. DOI: 10.1111/j.1365-2966.2006.11318.x
- Braun, W. and K. Hepp (1977). The Vlasov dynamics and its fluctuations in the $1/N$ limit of interacting classical particles. *Communications in Mathematical Physics*, 56(2), 101–113. DOI: 10.1007/BF01611497
- Zel'dovich, Ya. B. (1970). Gravitational instability: An approximate theory for large density perturbations. *Astronomy & Astrophysics*, 5, 84–89.
- Saari, D. G. (1977). A global existence theorem for the four-body problem of Newtonian mechanics. *Journal of Differential Equations*, 26(1), 80–111. DOI: 10.1016/0022-0396(77)90100-0
- Painlevé, P. (1897). *Leçons sur la théorie analytique des équations différentielles, professées à Stockholm*. Hermann.
- Levi-Civita, T. (1920). Sur la régularisation du problème des trois corps. *Acta Mathematica*, 42, 99–144. DOI: 10.1007/BF02404404
- Milnor, J. (1964). Eigenvalues of the Laplace operator on certain manifolds. *Proceedings of the National Academy of Sciences*, 51(4), 542. DOI: 10.1073/pnas.51.4.542
- Sunada, T. (1985). Riemannian coverings and isospectral manifolds. *Annals of Mathematics*, 121(1), 169–186. DOI: 10.2307/1971195
- Kac, M. (1966). Can one hear the shape of a drum? *American Mathematical Monthly*, 73(4, Part 2), 1–23. DOI: 10.2307/2313748
- Mertens, F. (1874). Ein Beitrag zur analytischen Zahlentheorie. *Journal für die reine und angewandte Mathematik*, 78, 46–62.
- Selberg, A. (1946). Contributions to the theory of the Riemann zeta-function. *Archiv for Mathematik og Naturvidenskab*, 48(5), 89–155.
- Feynman, R. P. (1985). *QED: The Strange Theory of Light and Matter*. Princeton University Press.
- DESI Collaboration (2024). DESI 2024 VI: Cosmological constraints from the measurements of baryon acoustic oscillations. *arXiv:2404.03002*.
- Kitaura, F.-S. and D. Sinigaglia (2026). Spectral hierarchy of the cosmic web: scale-weighting kernel classification. *Working paper*.