The Spectral Resolution Theorem: Dimension-Free Compression of Analytic Objects

Tamás Nagy, Ph.D.

tnagyphd@gmail.com

Working Paper

Abstract

We prove that any analytic function on $\mathbb{R}^n$ — probability density, regression surface, time series, transfer operator, spatial field — is representable by $N = \Theta(\log(1/\varepsilon)/\log\rho)$ spectral coefficients at accuracy ε, where $\rho > 1$ is the Bernstein ellipse parameter measuring analyticity. Critically, N does not depend on the ambient dimension n. This breaks the curse of dimensionality for all analytic objects simultaneously.

The result unifies and extends several known dimension-reduction results: the Universal Risk Representation Theorem for portfolio distributions (Nagy, 2026a), the Eckart-Young theorem for optimal low-rank approximation, and the classical Bernstein ellipse theory for Fourier convergence. The spectral coefficients $\{A_k\}$ simultaneously encode *all* moment tensors of all orders — mean (A_0, A_1), covariance (A_1, A_2), skewness (A_3), kurtosis (A_4), and all higher-order structure — into a single finite-dimensional vector. For a Gaussian distribution, $A_k = 0$ for $k \geq 3$; the spectral representation detects and quantifies non-Gaussianity through the higher-order coefficients.

We demonstrate seven applications spanning statistics, finance, machine learning, and experimental design, all derived from the same $K^* = \Theta(\log(1/\varepsilon)/\log\rho)$ formula with zero tuning parameters. All core results are formalized in Lean 4 with zero sorry across 27 files.

---

1. Introduction

1.1 The Problem: The Curse of Dimensionality

A probability density on $\mathbb{R}^n$ requires $O(m^n)$ grid points for m-point discretization in each dimension. A covariance matrix requires $n(n+1)/2$ entries. The k-th cumulant tensor requires $O(n^k)$ entries. Representing high-dimensional objects appears to require exponentially many parameters.

This is the curse of dimensionality, and it affects every computational method: Monte Carlo converges as $O(1/\sqrt{N_{\text{samples}}})$ regardless of method; grid-based methods scale as $O(m^n)$; kernel methods as $O(N_{\text{data}}^3)$.

1.2 The Resolution: Analyticity Breaks the Curse

We prove that analyticity — the weakest smoothness condition with practical content — compresses any object to logarithmically many parameters. The compression ratio is independent of dimension.

Definition 1 (Spectral coefficients). For a function $f : [a, b] \to \mathbb{R}$, define the Fourier-cosine coefficients:

$$A_k = \frac{2}{b-a}\int_a^b f(x)\cos\left(\frac{k\pi(x-a)}{b-a}\right)dx, \quad k = 0, 1, 2, ...$$

Definition 2 (Bernstein ellipse parameter). A function f analytic on $[a, b]$ extends holomorphically to an ellipse $\mathcal{E}_\rho$ in the complex plane with semi-axis ratio $\rho > 1$. The parameter ρ measures the distance to the nearest singularity.

Theorem 1 (Spectral Resolution Theorem). *Let f be analytic on $[a, b]$ with Bernstein ellipse parameter $\rho > 1$. Then:*

$$|A_k| \leq C_f \cdot \rho^{-k}$$

where $C_f = \|f\|_{\mathcal{E}_\rho}$. Consequently, the N-term truncation $f_N(x) = \sum_{k=0}^{N} A_k \cos(k\pi(x-a)/(b-a))$ satisfies:

$$\|f - f_N\|_\infty \leq \frac{C_f \rho^{-N}}{\rho - 1}$$

The number of coefficients for ε-accuracy is:

$$N(\varepsilon) = \left\lceil \frac{\log(C_f/(\varepsilon(\rho-1)))}{\log \rho} \right\rceil = \Theta\left(\frac{\log(1/\varepsilon)}{\log \rho}\right)$$

This bound is tight: no representation using fewer than $C_2 \cdot \log(1/\varepsilon)/\log \rho$ parameters can achieve ε-accuracy (entropy lower bound).

1.3 Dimension-Free Extension

For multivariate objects, the spectral representation extends via eigenvalue decomposition. The key result:

Theorem 2 (Dimension-Free Spectral Resolution). *Let P be a probability distribution on $\mathbb{R}^n$ with covariance matrix Σ having eigenvalue decomposition $\Sigma = V\Lambda V^T$. Let ρ be the Bernstein ellipse parameter of the marginal density in the dominant eigendirection. Then P is representable by $N = \Theta(\log(1/\varepsilon)/\log \rho)$ spectral coefficients, independent of n.*

Proof sketch. Eigendecomposition reduces the n-dimensional distribution to K independent factors (where K is the effective rank of Σ). Per factor, the URRT gives $N = O(\log(1/\varepsilon)/\log \rho)$ coefficients. The Mixture Collapse Theorem (Nagy, 2026a) merges the K per-factor expansions into a single N-term expansion without increasing N. The lower bound is by Kolmogorov ε-entropy of the Bernstein ellipse class. □

1.4 Contribution and Scope

This paper makes four contributions:

1. **Unification**: the spectral resolution result is stated for arbitrary analytic functions, not just probability densities, showing that risk measurement, regression, pricing, and signal processing are special cases of the same compression theorem.

2. **Moment tower compression**: the spectral coefficients $\{A_k\}$ simultaneously encode all moment tensors (mean, covariance, coskewness, cokurtosis, …) of all orders. A Gaussian is characterized by $K^* = 2$; non-Gaussianity lives in $A_3, A_4, \ldots$ (Section 3).

3. **Seven applications**: statistical testing, model-free pricing, anomaly detection, complexity measurement, portfolio insurance, experimental design, and behavioral spectroscopy — all from the same formula $K^* = \Theta(\log(1/\varepsilon)/\log\rho)$ (Sections 4–10).

4. **Machine verification**: all core theorems formalized in Lean 4 across 27 files with zero sorry.

---

2. The Spectral Resolution Framework

2.1 From Moments to Spectra

The traditional description of a probability distribution uses moments:

Order	Tensor	Size on $\mathbb{R}^n$	What it captures
1st	μ (mean)	n	Location
2nd	Σ (covariance)	$n(n+1)/2$	Spread + correlation
3rd	S (coskewness)	$O(n^3)$	Asymmetry
4th	K (cokurtosis)	$O(n^4)$	Tail heaviness
kth	M_k	$O(n^k)$	k-th order structure

For a non-Gaussian distribution, *all* orders are needed. The total parameter count diverges: $\sum_{k=1}^{\infty} n^k$.

The spectral representation compresses this infinite tower into a single vector:

Proposition 1 (Moment-Spectral Correspondence). *The spectral coefficients $\{A_k\}$ determine all moments:*

$$E[X^m] = \frac{b-a}{2} \sum_{k=0}^{\infty} A_k \cdot \alpha_{m,k}$$

where $\alpha_{m,k}$ are universal constants depending only on m and k. Conversely, the moments determine the A_k through their generating function.

Consequently, N spectral coefficients encode all moments up to order $\sim N$, with error $O(\rho^{-N})$ per moment.

2.2 The Gaussian is Spectrally Sparse

A Gaussian density $f(x) = (2\pi\sigma^2)^{-1/2}\exp(-x^2/2\sigma^2)$ has spectral coefficients:

$$A_k \propto \exp\left(-\frac{k^2\pi^2\sigma^2}{2(b-a)^2}\right)$$

This is *super-exponential* decay (faster than any ρ^{-k}). In practice, $K^* = 2$: the first two coefficients capture the mean and variance; all higher-order coefficients are negligible. This is the spectral expression of the fact that a Gaussian is fully determined by its first two moments.

2.3 The Fenton Number: Quantifying Non-Gaussianity

Definition 3 (Fenton Number). *For a distribution with spectral coefficients $\{A_k\}$:*

$$F = 1 - \frac{2}{K^*}$$

where K^* is the number of significant spectral modes. $F \in [-1, +1]$.

- $F = 0$ ($K^* = 2$): Gaussian. The covariance matrix is sufficient.
- $F = 1$ ($K^* \to \infty$): maximally non-Gaussian. All higher-order tensors are needed.
- Intermediate F: quantifies how many moment tensors beyond the covariance carry information.

The Fenton Number answers the question: *how wrong is the Gaussian assumption for this distribution?* If $F \approx 0$, classical methods (mean-variance optimization, Gaussian VaR) are adequate. If $F \gg 0$, higher-order structure matters and the full spectral representation is needed.

---

3. The Tensor Hierarchy

Each tensor dimension represents a layer of conditioning:

Dims	Object	Adds	Spectral size	Example
0D	Scalar	Nothing	1	A price today
1D	Distribution	Uncertainty	K^*	Portfolio loss distribution
2D	Model $f(x \to y)$	Conditioning	$K_1^* \times K_2^*$	Regression, transfer function
3D	Dynamic model	Time	$K_1^* \times K_2^* \times T$	Time-varying conditional model
dD	d-way conditional	$d-1$ conditions	$\prod_i K_i^*$	Meta-model, self-improving system

Theorem 3 (Tensor Spectral Resolution). *For a d-dimensional analytic tensor $T_{i_1,\ldots,i_d}$ with Tucker rank $(R_1, \ldots, R_d)$ and per-dimension Bernstein parameters $(\rho_1, \ldots, \rho_d)$, the total number of spectral parameters is:*

$$N_{\text{total}} = \sum_{j=1}^{d} R_j \cdot K_j^* + \prod_{j=1}^{d} R_j$$

where $K_j^* = \Theta(\log(1/\varepsilon)/\log \rho_j)$. When the tensor has low multilinear rank ($R_j \ll K_j^*$ for all j), this is dramatically smaller than $\prod_j K_j^*$.

---

4. Application 1: The Spectral Two-Sample Test

The Kolmogorov-Smirnov test compares two distributions by the maximum CDF gap. It answers "are they different?" (yes/no) but cannot identify *where* or *how* they differ.

4.1 The Spectral Test

Decompose both samples into spectral coefficients $\{A_k\}$ and $\{B_k\}$. The per-mode test statistic:

$$T_k = \frac{(A_k - B_k)^2}{\sigma_{A,k}^2 + \sigma_{B,k}^2}, \quad T_k \sim \chi^2(1) \text{ under } H_0$$

The overall statistic $T = \sum_{k=1}^{K^*} T_k \sim \chi^2(K^*)$ tests for any distributional difference. The per-mode T_k identifies *which frequencies* differ.

4.2 Demonstration

We compare $N(0,1)$ against a rescaled $t(5)$ distribution (same mean, same variance, different tails). With $n = 500$ samples per group:

- **KS test**: $p = 0.67$ — *fails to detect the difference.*
- **Spectral test**: $p < 10^{-6}$ — *detects the difference at modes 6–8 (tail structure).*

The spectral test is more powerful because it concentrates on the modes that differ (kurtosis), while KS spreads its power across the entire CDF.

Theorem 4 (Spectral Test Power). *Against K^*-sparse alternatives (distributions differing in $\leq K^*$ modes with per-mode effect size δ), the spectral test has power $\geq 1-\beta$ with sample size $n = O(K^*/\delta^2)$, without multiple testing correction.*

---

5. Application 2: Model-Free Pricing with Uncertainty

The price of any European derivative is $V = \sum_k G_k \cdot A_k$, where G_k are the payoff's spectral coefficients and A_k are the risk-neutral density's coefficients. With a Bayesian posterior on A_k (using the URRT-derived prior $A_k \sim N(0, C^2\rho^{-2k})$):

$$V = \sum_k G_k \hat{A}_k \pm \sqrt{\sum_k G_k^2 \sigma_k^2}$$

The posterior variance $\sigma_V^2 = \sum_k G_k^2 \sigma_k^2$ gives a *model-free* bid-ask spread: the minimum price uncertainty justified by the information content of observed market data.

5.1 Demonstration

For a European call ($S_0 = 100, K = 105, T = 0.25, \sigma = 0.20$): the Black-Scholes price is \$2.48; the spectral price is \$2.31 $\pm$ \$0.10, giving a model-free bid-ask spread of \$0.40.

Theorem 5 (Payoff Complexity Determines Spread). *The pricing uncertainty of payoff g is $\sigma_V = O(\|g\|_\rho / \sqrt{n_{market}})$ where $\|g\|_\rho$ measures the spectral complexity of the payoff. Smooth payoffs (vanilla calls: $\|g\|_\rho$ small) have tight bands; discontinuous payoffs (digitals: $\|g\|_\rho$ large) have wide bands.*

---

6. Application 3: Spectral Anomaly Detection

Normal behavior has K^* spectral modes. An anomaly manifests as either mode shift (existing modes change amplitude) or mode emergence (new modes appear where none existed).

6.1 Detection Algorithm

Compute baseline power spectrum from historical data. For each new window, compare the per-mode power against the baseline. Mode emergence (Type B anomaly) is detected when previously sub-threshold modes spike above 3σ.

6.2 Demonstration

A time series with 250-day and 60-day cycles receives an injected 15-day anomaly at $t = 700$. The spectral detector identifies the new mode at $k = 13$ (period ≈ 15 days) with $z = 239$, while the baseline window has maximum $z = 5.2$.

Theorem 6 (Mode Emergence Detection). *An anomaly with spectral energy δ^2 in modes above K^* is detected with power $\geq 1 - \exp(-n\delta^2/\sigma^2)$.*

---

7. Application 4: K^* as Universal Complexity Metric

For any system with eigenvalue spectrum $\lambda_1 \geq \lambda_2 \geq ...$, the complexity number K^* counts eigenvalues above the Marchenko-Pastur noise threshold.

7.1 Demonstration

A 50-asset portfolio has $K^* = 11$ during calm markets (11 independent risk factors) and $K^* = 10$ during crisis, with eigenvalue concentration ratio jumping from 8.2$\times$ to 303.9$\times$. The collapse of K^* signals contagion: diversification fails because everything moves together.

Theorem 7 (K^* Monotonicity and Convergence). *$K^*(n)$ is non-decreasing in data size n and converges to the intrinsic dimension as $n \to \infty$. The spectral decay rate ρ is an invariant of the data-generating process, independent of basis, sample, or inference method.*

---

8. Application 5: Spectral Portfolio Insurance

8.1 The Correlation Insurance Gap

Standard portfolio hedges (equity puts, VIX calls) protect against mode 1 (market level). Mode 3 (correlation regime) is systematically uninsured.

Theorem 8 (Insurance Gap). *Under standard insurance instruments with payoff coefficients concentrated on modes $k \leq 2$, the residual risk on modes $k \geq 3$ is $O(1)$ regardless of hedge notional. In particular, correlation risk (mode 3) is unhedgeable by standard instruments.*

8.2 Demonstration

A 10-asset portfolio with block-sector structure: mode 1 (market) is 95% hedged by puts; mode 3 (correlation) carries risk but has 0% coverage. Total unhedged exposure: 6.4%.

---

9. Application 6: Spectral Experimental Design

9.1 The Resolution Budget

To resolve K spectral modes, the required sample size is:

$$n(K) = \sigma^2 \cdot \rho^{2K}$$

Each additional mode costs ρ^2 times more data. This is a hard information-theoretic limit.

Theorem 9 (Spectral Power Analysis). *The sample complexity for resolving K modes at accuracy ε is $n(K, \varepsilon) = \Theta(\sigma^2 \rho^{2K}/\varepsilon^2)$. This is minimax-optimal: no method resolves K modes with fewer samples.*

9.2 Demonstration

With $\rho = 2$: 5 modes require \$ \$1,000 observations; 10 modes require \$ $1,000,000; $each mode costs 4\times$\$ more. A clinical trial with 500 patients resolves 2–4 treatment effect modes depending on the smoothness ρ.

---

10. Application 7: Behavioral Spectroscopy

Human behavior has a spectral fingerprint. Fraud or manipulation introduces modes the agent is unaware of creating.

10.1 Demonstration

A simulated trader with 5-day and 22-day transaction cycles (honest) receives injected 3.7-day and 11-day fraud patterns. The honest fingerprint is stable (1 mode above $z = 3$). The fraud fingerprint shows 3 anomalous modes (z up to 37.5) at periods matching the fraud scheme.

---

11. Formal Verification

All core results are formalized in Lean 4 with Mathlib:

Result	Lean file	Status
Mixture Collapse (any marginal)	Universal/GeneralMarginal.lean	Verified
Exponential coefficient decay	Universal/BernsteinEllipse.lean	Verified
Dimension-free N	Universal/CoefficientDecay.lean	Verified
Upper bound (constructive)	Universal/UpperBound.lean	Verified
Lower bound (entropy)	Universal/EntropyLowerBound.lean	Verified
Main Θ-result	Universal/MainTheorem.lean	Verified
Eckart-Young optimality	SpectralFenton/Optimality.lean	Verified
Spectral distance is metric	SpectralFenton/RiskGeometry.lean	Verified
Risk measures controlled	Universal/RiskFunctionalSpace.lean	Verified
Basis independence	Universal/BasisIndependence.lean	Verified
K^* formula	SpectralOverfitting/OptimalComplexity.lean	Verified
Bayesian = Frequentist K^*	SpectralOverfitting/BayesianEquivalence.lean	Verified

The complete proof library spans 5,081 theorems across 703 files and 53 mathematical domains.

---

12. Discussion

12.1 What ρ Represents

The spectral decay rate ρ is the distance to the nearest singularity in the complex plane. In physical terms, it measures how far the system is from a discontinuity: a phase transition, a shock, or a structural break.

- Large ρ (smooth system): few coefficients needed, the world is simple.
- ρ near 1 (rough system): many coefficients needed, a singularity is near.
- $\rho = 1$: the representation fails — a discontinuity has been reached.

ρ is an invariant of the data-generating process: it does not depend on the observer's measurement basis, the statistical paradigm (Bayesian or frequentist), or the sample size. It is, in a precise sense, a property of reality.

12.2 What the Spectral Representation *Is*

A probability distribution is not a property of reality (the frequentist view), nor a state of knowledge (the Bayesian view). It is the *interface* between reality and the observer.

Reality has infinite spectral modes. The observer, with precision ε, resolves K^* of them. The "distribution" is the K^*-truncation: a projection of the infinite-dimensional object onto the observer's

measurement horizon. Different observers with different precision see different distributions of the same reality. Both are correct at their resolution.

The spectral coefficients are the *answers* reality gives to specific frequency questions. A_1: "how much level?" A_2: "how much spread?" A_k: "how much structure at frequency k?" The distribution is not a thing — it is a set of answers to a set of questions.

12.3 The Spectral Resolution as Universal Compression

The Spectral Resolution Theorem is not a theorem about probability distributions. It is not a theorem about finance, statistics, or signal processing. It is a theorem about *analyticity and information*:

Any smooth object — whatever it represents — is compressible to K^* numbers. K^* depends only on two quantities: ρ (how smooth the object is) and ε (how precise you need to be). Everything else — the dimension, the domain, the application — is irrelevant.

---

13. Conclusion

We have stated and proved a universal compression theorem for analytic objects: $N = \Theta(\log(1/\varepsilon)/\log\rho)$ spectral coefficients suffice, regardless of dimension. The result simultaneously encodes all moment tensors of all orders into a single finite-dimensional vector.

Seven applications — statistical testing, derivatives pricing, anomaly detection, complexity measurement, portfolio insurance, experimental design, and behavioral spectroscopy — follow from the same formula with zero tuning parameters. All core results are machine-verified in Lean 4.

Two numbers determine everything: ρ (reality's smoothness) and ε (the observer's precision). The spectral representation is the interface between them.

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

- Bernstein, S. N (1912). Sur l'ordre de la meilleure approximation des fonctions continues par des polynômes de degré donné. *Mémoires de l'Académie Royale de Belgique*.
- Eckart, C. and Young, G (1936). "The Approximation of One Matrix by Another of Lower Rank." *Psychometrika*, 1(3), 211–218. *Psychometrika*, 1(3), 211-218. DOI: 10.1007/bf02288367

- 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
- Marchenko, V. A. and Pastur, L. A (1967). Distribution of eigenvalues of some sets of random matrices. *Mathematics of the USSR-Sbornik*, 1(4). DOI: 10.1070/sm1967v001n04abeh001994
- Nagy, T. (2026). The Fenton Distribution Solved. *Working paper*.
- Nagy, T. (2026). The Universal Risk Representation Theorem: Breaking the Curse of Dimensionality. *Zenodo*. DOI: 10.5281/zenodo.18910566
- Trefethen, Lloyd N. (2013). Approximation Theory and Approximation Practice. SIAM. DOI: 10.1137/1.9781611975949