The Smooth Latent Operator: Parameter-Free Distributional Representations via Kernel Moment Recovery

Tamás Nagy, Ph.D.

tnagyphd@gmail.com

Draft

Executive Summary (Non-Technical)

When we characterize a probability distribution from its moments — as in the Padé-COS method for correlated lognormal sums (Nagy, 2026h) or the generating-function method for the three-body problem (Nagy, 2026g) — the result depends on a discrete truncation parameter: how many moments to use. This creates a discontinuous, tuning-dependent method that struggles when the moments are ill-conditioned (high volatility, heavy tails).

This paper replaces the discrete truncation with a smooth, continuous operator. The moments live in a weighted Hilbert space; the characteristic function is recovered by a kernel inner product; and the optimal resolution is determined automatically by the problem's own analyticity structure. The result is a **single smooth operator** from model parameters to CDF, with no free truncation parameter.

Abstract

The Padé resummation of a moment generating function — as used in distributional Latent representations (Nagy, 2026h) and orbital Latent representations (Nagy, 2026g) — depends on a discrete Padé order N_P. For fixed N_P, the construction is algebraic and closed-form; as N_P varies, the result changes discontinuously (the linear system changes dimension, coefficients can jump, and the condition number spikes). For heavy-tailed distributions or high-volatility regimes, no single N_P works well.

We resolve this by embedding the Padé-COS chain in a **reproducing kernel Hilbert space (RKHS)** of moment sequences. The scaled moments $\{c_k = m_k/k!\}$ define a Latent element Λ in a Gaussian-weighted ℓ^2 space. The characteristic function is recovered by a kernel evaluation — a smooth inner product that replaces the Toeplitz matrix inverse. A continuous resolution parameter $\alpha \in \mathbb{R}_{>0}$ replaces the discrete N_P, making the entire chain smooth.

The resolution α can be set adaptively: the Latent Theorem (Nagy, 2026e) determines the natural representation size from the analyticity parameter ρ of the generating function, which is itself a smooth function of the model parameters. This yields a **parameter-free smooth operator**:

$$\mathcal{L} : (w, \mu, \Sigma) \to F_S(x)$$

that is smooth in all arguments, requires no tuning, and automatically adapts to the problem's difficulty. The classical Padé-COS formula is recovered as a special case (the Dirac kernel limit).

We introduce the concept of a **grade-2 Latent**: the Latent of the extraction chain itself. The distribution has a grade-1 Latent Λ (what the distribution is); the extraction process has a grade-2 Latent α^* (how to optimally represent it). Together, (Λ, α^*) provides a complete, smooth, parameter-free characterization.

---

1. Introduction

1.1 The Truncation Problem

The Latent representation program (Nagy, 2026e) characterizes smooth systems by finite algebraic expressions. For probability distributions, the chain is:

$$\text{Parameters} \to \text{Moments} \to \text{Padé CF} \to \text{COS coefficients} \to \text{CDF}$$

Each step is algebraic. But the chain depends on three truncation parameters:

Parameter	Role	Type
N_P	Padé order (moments used)	**Discrete**
N	COS order (cosine terms)	Discrete
$[a, b]$	Domain bounds	Continuous

The COS order N and domain $[a, b]$ are relatively benign — exponential convergence makes the choice insensitive to modest changes. The Padé order N_P is the problematic parameter:

1. **Discrete jumps.** Changing N_P from 18 to 19 changes the Toeplitz matrix dimension from 18×18 to 19×19. The new row introduces moment c_{37}, which can be orders of magnitude larger than c_{35}.
2. **Condition number spikes.** The Toeplitz matrix condition number can change by factors of 10^{10} between adjacent N_P values, especially for high-volatility distributions.
3. **Pole artifacts.** The Padé $[N_P - 1/N_P]$ has N_P poles. As N_P increases, new poles appear (and existing ones move). A pole near the imaginary axis corrupts the CF evaluation.
4. **No smooth limit.** The sequence $\{R_{N_P}\}_{N_P=1}^{\infty}$ converges to the CF (under the conditions of Theorem 2, Nagy 2026h), but the convergence is not monotone and individual R_{N_P} can be far from the limit.

1.2 What This Paper Shows

We embed the Padé-COS chain in a smooth operator that:

1. **Replaces the discrete N_P with a continuous resolution α.** The Padé is a special case at integer α with zero regularization.

2. **Provides automatic regularization.** The Tikhonov term $e^{-\alpha}I$ prevents ill-conditioning, with the regularization strength determined by the resolution.

3. **Determines α from the problem.** The analyticity parameter ρ of the MGF — computable from the moments themselves — sets the natural resolution α^*. No user tuning required.

4. **Gives a smooth map $(w, \mu, \Sigma, x) \mapsto F_S(x)$.** Smooth in all arguments simultaneously.

1.3 Graded Latents

The Latent framework (Nagy, 2026e) assigns to each smooth system a finite representation — the Latent. We extend this with a grading:

- **Grade 0.** The system parameters (w, μ, Σ). These define the problem.
- **Grade 1.** The distributional Latent $\Lambda = \{c_k\} \in \mathcal{H}_\beta$. This encodes *what* the distribution is.
- **Grade 2.** The extraction Latent $\alpha^* \in \mathbb{R}_{>0}$. This encodes *how to optimally extract* the distribution from its moment representation.

The grade-2 Latent is new. It arises because the moment-to-CF map has an intrinsic difficulty parameter (the condition number of the extraction), which varies smoothly with the system parameters. The grade-2 Latent captures this difficulty.

---

2. The Moment Hilbert Space

2.1 Construction

The scaled moments $c_k = m_k/k!$ of a sum of lognormals satisfy $|c_k| \le Ce^{\alpha k^2}$ where $\alpha = \sigma_{\max}^2/2$ (see Nagy, 2026h, Section 2.4). Define the Gaussian-weighted Hilbert space:

$$\mathcal{H}_\beta = \left\{ \{a_k\}_{k=0}^{\infty} : \|a\|_\beta^2 := \sum_{k=0}^{\infty} |a_k|^2 e^{-2\beta k^2} < \infty \right\}, \quad \beta > \alpha$$

The standard orthonormal basis is $e_k = (0, \dots, 0, e^{\beta k^2}, 0, \dots)$, and the inner product is $\langle a, b \rangle_\beta = \sum_k a_k \overline{b_k} e^{-2\beta k^2}$.

Proposition 1. *For $\beta > \sigma_{\max}^2/2$, the moment map*

$$\Phi : (w, \mu, \Sigma) \mapsto \{c_k\}_{k=0}^{\infty} \in \mathcal{H}_\beta$$

is well-defined and smooth (infinitely Fréchet differentiable).

Proof. Each c_k is a finite sum of smooth functions of (w, μ, Σ) (Theorem 1 of Nagy, 2026h). The tail bound $|c_k| \le Ce^{\alpha k^2}$ with $\alpha < \beta$ gives $\|c\|_\beta < \infty$. Smoothness of Φ follows from the smoothness of each coordinate and the summability of derivatives (which satisfy the same growth bound with modified constants). $\square$

2.2 The Latent as a Hilbert Space Element

Definition 1. The **distributional Latent** of $S = \sum w_i e^{Y_i}$ is $\Lambda = \Phi(w, \mu, \Sigma) \in \mathcal{H}_\beta$.

Λ encodes all distributional information: the moments, the CF, the CDF, and all derived quantities (VaR, ES, spectral risk measures) are continuous functionals of Λ.

The norm $\|\Lambda\|_\beta$ has a natural interpretation: it measures the "complexity" of the distribution in the moment basis, with exponential penalization of high-order moments. Distributions with light tails (low σ) have small $\|\Lambda\|_\beta$; distributions with heavy tails (high σ) have large $\|\Lambda\|_\beta$.

2.3 Basis Independence

The cosine coefficients $\{A_k\}$ from the COS expansion are coordinates of Λ in a DIFFERENT basis (the Fourier-cosine basis on $[a, b]$). The Padé-COS chain is a change of basis:

$$\text{moment basis} \xrightarrow{\text{Padé}} \text{rational CF} \xrightarrow{\text{COS}} \text{cosine basis}$$

Both bases span the same Latent Λ. The Latent itself — the abstract element of $\mathcal{H}_\beta$ — is basis-free.

---

3. Kernel Recovery of the Characteristic Function

3.1 The Evaluation Kernel

The characteristic function at frequency t is formally:

$$\phi_S(t) = M(it) = \sum_{k=0}^{\infty} c_k (it)^k$$

This series diverges (zero radius of convergence). The classical Padé recovers $\phi_S(t)$ by rational interpolation of the first $2N_P$ coefficients.

We define a kernel-based alternative. For resolution parameter $\alpha > 0$, define the **evaluation kernel**:

$$K_t^\alpha(k) = (it)^k \, \psi_\alpha(k)$$

where $\psi_\alpha : \mathbb{N}_0 \to \mathbb{R}_{>0}$ is a smooth cutoff:

$$\psi_\alpha(k) = \exp\left(-\frac{k^2}{2\alpha^2}\right)$$

The kernel-recovered CF is:

$$\hat{\phi}_\alpha(t) = \sum_{k=0}^{\infty} c_k (it)^k \psi_\alpha(k) = \langle \Lambda, K_t^\alpha \rangle_\beta \cdot Z_\alpha$$

where Z_α is a normalization ensuring $\hat{\phi}_\alpha(0) = 1$.

Proposition 2. *For $\alpha < \sqrt{2\beta}$, $K_t^\alpha \in \mathcal{H}_\beta$ for all $t \in \mathbb{R}$, and the series converges absolutely.*

Proof. $\|K_t^\alpha\|_\beta^2 = \sum_k |t|^{2k} e^{-k^2/\alpha^2} e^{-2\beta k^2} = \sum_k |t|^{2k} e^{-(1/\alpha^2+2\beta)k^2} < \infty$ for any t, since the Gaussian damping defeats the polynomial growth. □

3.2 Convergence as $\alpha \to \infty$

Theorem 1 (Kernel CF Convergence). *Under the conditions of Theorem 2 in Nagy (2026h):*

$$\lim_{\alpha \to \sqrt{2/\sigma_{\max}^2}^-} \hat{\phi}_\alpha(t) = \phi_S(t) \quad \text{for all } t \in \mathbb{R}$$

The convergence is pointwise and, on compact subsets of $\mathbb{R}$, the rate is determined by the gap $\sqrt{2/\sigma_{\max}^2} - \alpha$.

Proof sketch. The Gaussian cutoff $\psi_\alpha(k) = e^{-k^2/(2\alpha^2)}$ suppresses terms beyond $k \sim \alpha$. As α approaches the critical value $\alpha_{\text{crit}} = 1/\sqrt{\alpha_{\text{growth}}} = \sqrt{2}/\sigma_{\max}$, the effective number of contributing moments increases, and the kernel CF approaches the exact CF by the same mechanism as the Padé (matching more and more Taylor coefficients). □

3.3 Smoothness in α

Proposition 3. *The map $\alpha \mapsto \hat{\phi}_\alpha(t)$ is C^∞ for $\alpha \in (0, \alpha_{crit})$.*

Proof. $\hat{\phi}_\alpha(t) = \sum_k c_k (it)^k e^{-k^2/(2\alpha^2)}$. Each term is smooth in α. The sum converges uniformly in α on compact subsets of $(0, \alpha_{\text{crit}})$ (by dominated convergence with dominating function $|c_k t^k| e^{-k^2/(2\alpha_0^2)}$ for any $\alpha_0 < \alpha_{\text{crit}}$). Term-by-term differentiation is justified. □

This is the key property: the kernel CF depends **smoothly** on the resolution parameter, unlike the Padé which jumps at integer N_P.

---

4. The Regularized Padé as a Special Case

4.1 Connection to the Classical Padé

The classical Padé $[N_P - 1/N_P]$ can be recovered from the kernel framework by choosing a SPECIFIC kernel: the "hard-cutoff rational kernel":

$$\psi_{N_P}^{\text{Padé}}(k) = \begin{cases} 1 & k \leq 2N_P - 1 \\ 0 & k \geq 2N_P \end{cases}$$

followed by rational interpolation of the truncated series. This is a DISCONTINUOUS kernel (step function) — which explains why the Padé is discontinuous in N_P.

The Gaussian kernel $\psi_\alpha(k) = e^{-k^2/(2\alpha^2)}$ is the smooth generalization. As $\alpha \to \infty$, $\psi_\alpha \to 1$ (all moments contribute); as $\alpha \to 0$, $\psi_\alpha \to \delta_0$ (only c_0 contributes). The Padé at order N_P corresponds roughly to $\alpha \approx N_P$.

4.2 Tikhonov-Regularized Padé

An intermediate construction blends Padé rationality with kernel smoothness. For continuous $\alpha > 0$:

1. Set effective order $n = \lceil \alpha \rceil$.
2. Build the Toeplitz matrix T_n from moments $c_0, \dots, c_{2n-1}$.
3. Solve with Tikhonov regularization:

$$\mathbf{q}(\alpha) = -(T_n^* T_n + e^{-\alpha} I)^{-1} T_n^* \mathbf{b}_n$$

4. Construct the rational CF $R_\alpha(z) = P_\alpha(z)/Q_\alpha(z)$.

This preserves the RATIONAL structure of the Padé (important for COS evaluation) while adding smoothness through the regularization parameter $e^{-\alpha}$:

- At integer $\alpha = N_P$ with $e^{-\alpha} \approx 0$: recovers the classical Padé.
- Between integers: the regularization smoothly interpolates.
- Near ill-conditioning: the regularization prevents blow-up.

---

5. Adaptive Resolution

5.1 The Natural Resolution

The Latent Theorem (Nagy, 2026e, Theorem 3) states:

$$N^*(\varepsilon) = \left\lceil \frac{\log(1/\varepsilon)}{\log \rho} \right\rceil$$

where ρ is the analyticity parameter (ratio of the analyticity radius to the evaluation radius). For the Padé-COS chain applied to lognormal sums, ρ depends on the singularity structure of $M(z)$ in the complex plane.

Proposition 4. *The analyticity parameter ρ for the Padé of $M(z) = E[e^{zS}]$ satisfies:*

$$\rho \geq \frac{1}{\sigma_{\max}^2 \cdot t_{\max}}$$

where $t_{\max} = (N-1)\pi/(b-a)$ is the highest COS frequency. This gives:

$$N_P^* \leq \sigma_{\max}^2 \cdot t_{\max} \cdot \log(1/\varepsilon)$$

For moderate volatility ($\sigma_{\max} = 0.3$), $t_{\max} \approx 50$, $\varepsilon = 10^{-6}$: $N_P^* \leq 62$. For high volatility ($\sigma_{\max} = 1.0$): $N_P^* \leq 690$ — explaining the difficulty.

5.2 The Self-Determined Operator

Define the **adaptive resolution**:

$$\alpha^*(w, \mu, \Sigma; \varepsilon) = \frac{\log(1/\varepsilon)}{\log \rho(w, \mu, \Sigma)}$$

Since ρ is a smooth function of (w, μ, Σ) (through the moment growth rate), α^* is smooth. The **Smooth Latent Operator** is the composition:

$$\boxed{\mathcal{L}_\varepsilon : (w, \mu, \Sigma) \xrightarrow{\Phi} \Lambda \in \mathcal{H}_\beta \xrightarrow{K^{\alpha^*}} \hat{\phi}_{\alpha^*}(\cdot) \xrightarrow{\text{COS}} F_S(x)}$$

For fixed target accuracy ε, this is a single smooth map with no free parameters.

---

6. The Grade-2 Latent

6.1 Definition

Definition 2 (Graded Latent). The complete Latent characterization of a distribution in the moment-Padé-COS framework is the pair:

$$(\Lambda, \alpha^*) \in \mathcal{H}_\beta \times \mathbb{R}_{>0}$$

where: - Λ is the **grade-1 Latent**: the distribution itself, encoded as moments in $\mathcal{H}_\beta$. - α^* is the **grade-2 Latent**: the natural resolution of the extraction chain.

6.2 Interpretation

The grade-2 Latent α^* captures the *extractability* of the distribution — how many moments are needed to reconstruct the CF to a given accuracy. It depends on:

- **Tail weight.** Heavy-tailed distributions (high σ) have small α^* — the moments grow fast, and the extraction is hard.
- **Modality.** Multimodal distributions may require more moments to resolve the peaks.
- **Correlation structure.** High correlations concentrate the distribution, potentially increasing α^*.

The grade-2 Latent is INTRINSIC to the distribution — it does not depend on the extraction method (Padé vs. kernel vs. some other algebraic map). Any method that recovers the CF from moments will face the same fundamental difficulty, parameterized by α^*.

6.3 The Three-Body Analogue

For the gravitational three-body problem (Nagy, 2026g), the graded Latent is:

- **Grade 1.** The Fourier-Padé coefficients $\{(\Lambda_k, \omega)\}$ — the orbit itself.

- **Grade 2.** The convergence radius ρ of the generating function — how many Fourier terms are needed.

The grade-2 Latent ρ depends on the proximity of temporal singularities (collision times). Close encounters give small ρ (hard to extract); well-separated orbits give large ρ (easy).

The parallel is exact: in both cases, the grade-2 Latent measures the difficulty of going from an algebraic defining equation to a convergent evaluation.

---

7. Relation to Existing Theory

7.1 Stieltjes Moment Problem

The kernel recovery is related to the Stieltjes moment problem: given moments $\{c_k\}$, recover the measure. The classical theory (Shohat and Tamarkin, 1943) studies when the moments determine the measure uniquely (determinacy). The kernel approach provides a SMOOTH reconstruction that is well-defined even in the indeterminate case — the kernel selects a specific measure (the maximum-entropy or Tikhonov-regularized one).

7.2 Reproducing Kernel Hilbert Spaces

The kernel CF recovery $\hat{\phi}_\alpha(t) = \langle \Lambda, K_t^\alpha \rangle$ is a standard RKHS construction. The kernel K^α defines a reproducing kernel:

$$\mathcal{K}_\alpha(s,t) = \langle K_s^\alpha, K_t^\alpha \rangle_\beta = \sum_{k=0}^{\infty} (-st)^k e^{-k^2/\alpha^2} e^{-2\beta k^2}$$

This is a positive-definite kernel on $\mathbb{R}$, inducing an RKHS of characteristic functions.

7.3 SVD-Based Padé

The Tikhonov-regularized Padé (Section 4.2) is related to the SVD-based Padé of Gonnet, Güttel, and Trefethen (2013). Their approach uses the SVD of the Toeplitz matrix to determine the effective Padé order adaptively. The connection: the SVD determines the numerical rank, which corresponds to our α^*.

---

8. Implications

8.1 For the Lognormal Sum Problem

The smooth operator resolves the high-volatility difficulty (Section 7.3 of Nagy, 2026h): instead of a discrete N_P that is either too small (poor accuracy) or too large (ill-conditioned), the continuous α with Tikhonov regularization finds the optimal trade-off. The Bitcoin test case ($\sigma = 0.8$–1.2) that failed with discrete Padé should succeed with the regularized kernel.

8.2 For the Latent Framework

The grade-2 Latent is the first extension of the Latent framework beyond the representation itself. It suggests a hierarchy:

- **Grade 0.** Parameters (problem definition).
- **Grade 1.** Latent (the object — distribution, orbit, solution).
- **Grade 2.** Extraction Latent (the difficulty of representing the object).
- **Grade 3.** (Speculative.) The Latent of the difficulty function — how the extraction difficulty varies across the parameter space.

Each grade adds a smooth layer of meta-information about the representation.

8.3 For General Moment-Based Methods

The kernel recovery construction applies to ANY problem where a generating function is recovered from its Taylor coefficients: option pricing (Heston CF), queue theory (waiting-time distribution), signal processing (spectral density from autocorrelations). The grade-2 Latent provides a universal measure of "moment extractability."

---

9. Conclusion

The discrete Padé-COS chain for distributional Latent representations can be embedded in a smooth operator by replacing the Toeplitz matrix inverse with a kernel inner product in a weighted Hilbert space. The construction:

1. Places the moments in a Gaussian-weighted ℓ^2 space $\mathcal{H}_\beta$.
2. Recovers the CF via a smooth kernel evaluation parameterized by a continuous resolution α.
3. Determines α adaptively from the problem's analyticity parameter ρ.

The result is a **parameter-free smooth operator** $\mathcal{L} : (w, \mu, \Sigma) \to F_S(x)$ that is smooth in all arguments and requires no truncation parameter.

The concept of **graded Latents** — where grade 1 encodes the object and grade 2 encodes the extraction difficulty — extends the Latent framework to characterize not just *what* a system is but *how hard it is to represent*. This meta-level information is intrinsic to the system and independent of the representation method.

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

- Nagy, T. (2026). The Latent: Finite Sufficient Representations of Smooth Systems. *Zenodo*. DOI: 10.5281/zenodo.19101209
- Nagy, T. (2026). The Exact Latent Distribution of Correlated Lognormal Sums. *Working paper*.
- Nagy, T. (2026). The Exact Latent Solution of the Gravitational Three-Body Problem. *Zenodo*. DOI: 10.5281/zenodo.19101229
- Baker, G. A. and P. Graves-Morris (1996). Padé Approximants. *Padé Approximants*.
- Gonnet, P., Güttel, S., and Trefethen, L.N (2013). Robust Padé approximation via SVD. *Gonnet, P., Güttel, S., and Trefethen, L.N.*, 55(1), 101-117. DOI: 10.1137/110853236
- Shohat, J.A. and Tamarkin, J.D (1943). The Problem of Moments. *Shohat, J.A. and Tamarkin, J.D.*. DOI: 10.1090/surv/001
- 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
- Wahba, G (1990). Spline Models for Observational Data. *Wahba, G.*. DOI: 10.1137/1.9781611970128
- Berlinet, A. and Thomas-Agnan, C (2004). Reproducing Kernel Hilbert Spaces in Probability and Statistics. *Berlinet, A. and Thomas-Agnan, C.*. DOI: 10.1007/978-1-4419-9096-9