Spectral Importance Sampling: Optimal Rare-Event Simulation via Eigenvalue-Conditioned Measure Change

Making Extreme Losses Visible Without Brute Force

Computationally separable optimal tilt and variance reduction exponential in the loss threshold

Tamás Nagy, Ph.D.

tnagyphd@gmail.com

working_paper • March 2026

Executive Summary (Non-Technical)

Rare financial catastrophes — the 2008 credit crisis, the 2020 COVID crash, a crypto exchange collapse — are the events that matter most for risk management, yet they are precisely the events that standard simulation methods handle worst. Monte Carlo simulation, the industry workhorse, generates random scenarios from the portfolio's joint distribution. A one-in-a-million loss event requires, on average, a million samples before it appears even once. A one-in-a-billion event needs a billion. For stress testing, capital adequacy, and tail-risk hedging, this is prohibitively expensive.

The established remedy is *importance sampling*: change the simulation measure so that rare events become common, run the simulation cheaply, then correct each sample by a known weight to recover the true probabilities. **The unsolved problem is choosing the right measure change.** In high-dimensional portfolios — hundreds of correlated assets — the space of possible tilts is vast, and a poor choice can make variance worse, not better.

This paper shows that the eigenvalue decomposition of the portfolio correlation matrix provides a **natural, provably optimal tilting direction**. The key structural insight is mode independence: the eigenvalue-weighted modes Z_k are mutually independent (projections onto orthogonal eigenvectors of the correlation matrix). This means the importance sampling measure change factorizes into independent per-mode tilts, each of which can be optimized in closed form — reducing a d-dimensional optimization to K scalar problems solvable in $O(K)$ time. **The resulting variance reduction is exponential in the loss threshold ℓ**: the IS variance satisfies $\mathrm{Var}_{\mathrm{IS}} \leq C(\ell) \cdot \rho^{-2\ell}$, where $\rho > 1$ is the analyticity radius of the portfolio characteristic function and $C(\ell)$ is a polynomial prefactor.

The method inherits its optimality from a connection between the analyticity radius ρ of the portfolio distribution (from the Latent representation theorem) and the rate function of large deviation theory. **The Bernstein ellipse that governs spectral coefficient decay is, simultaneously, the domain that governs optimal exponential tilting.** This duality means that the same mathematical object — the eigenvalue spectrum — controls both the exact analytical computation of the distribution (via the COS method) and the optimal simulation of its extreme tails.

What the paper does not claim: it does not replace the exact COS-based framework for computing standard risk measures (VaR, ES, spectral risk measures). For those, the existing

Latent methods are faster and more accurate. The importance sampling framework addresses the regime where simulation is unavoidable: path-dependent payoffs, dynamic multi-period problems, barrier monitoring, and ultra-deep tails beyond the reach of truncated spectral expansions.

The practical consequence is a simulation method that achieves reliable tail estimates at a fraction of the computational cost of naive Monte Carlo, with explicit variance-reduction guarantees that can be audited and certified — extending the "noise-free risk" philosophy to the last remaining domain where noise was considered unavoidable.

Abstract

We develop a variance reduction framework for simulating rare events in correlated portfolios by exploiting the eigenvalue decomposition of the correlation matrix. The central observation is that the eigenvalue modes Z_k — projections of the asset vector onto the eigenvectors of the correlation matrix — are mutually independent. This mode independence allows the importance sampling measure change to factorize into independent per-mode exponential tilts, each admitting a closed-form saddle-point solution. The optimal tilt is computable in $O(K)$ operations, independent of the portfolio dimension n.

We establish three main results. First, we prove that the factored importance sampling estimator is unbiased and logarithmically efficient: at the optimal tilt, the IS second moment satisfies $\mathbb{E}_{\mathbb{Q}_{\theta^*}}[(d\mathbb{P}/d\mathbb{Q})^2 \cdot \mathbf{1}\{L > \ell\}] \leq C(\ell) \cdot \rho^{-2\ell}$, where $\rho > 1$ is the analyticity radius of the portfolio characteristic function, ℓ is the loss threshold, and $C(\ell)$ is a polynomial prefactor (Theorem 1). Second, we show that the optimal per-mode tilt parameter θ_k^* is the saddle point of the per-mode cumulant generating function, and that the collection $\{\theta_k^*\}$ solves a separable convex program in the large-ℓ regime (Theorem 2). Third, we prove a duality between the Bernstein ellipse radius ρ governing spectral coefficient decay and the Cramér rate function governing tail probabilities, establishing that the optimal importance sampling domain coincides with the analyticity domain of the Latent representation (Theorem 3).

The method applies to any portfolio model where the correlation matrix has a spectral gap ($\lambda_1/\lambda_2 > 1$) and the marginals have analytic characteristic functions ($\rho > 1$). Heavy-tailed distributions without a moment generating function (Student-t, Pareto, stable) have $\rho = 1$ and receive polynomial rather than exponential variance reduction. We demonstrate the method on (i) deep-tail VaR/ES estimation at the 10^{-8} level, (ii) multi-asset barrier option pricing with correlated knock-out, and (iii) CDO tranche loss estimation under correlated defaults.

Keywords: importance sampling, rare events, eigenvalue decomposition, spectral methods, variance reduction, portfolio risk, large deviations

MSC 2020: 65C05, 60F10, 91G60, 62P05

---

1. Introduction

1.1 The Problem: Tails Are Expensive

The distribution of portfolio losses has fat tails driven by correlation. The probability of extreme joint losses — the events that trigger margin calls, capital breaches, and systemic crises — is small but consequential. Estimating these probabilities by Monte Carlo simulation requires sample sizes inversely proportional to the probability of interest: to estimate $P(L > \ell)$ at the 10^{-p} level with relative error ϵ, one needs $n \geq \epsilon^{-2} \cdot 10^p$ samples. For regulatory stress tests targeting the 99.97th percentile ($p = 3.5$), this is manageable. For credit portfolio losses at the 10^{-6} level, or operational risk at 10^{-8}, it is not.

1.2 The Standard Remedy and Its Limitation

Importance sampling (IS) addresses this by simulating under a *tilted* probability measure $\mathbb{Q}$ that concentrates mass on the rare event of interest, then correcting each sample by the likelihood ratio $d\mathbb{P}/d\mathbb{Q}$. The estimator remains unbiased, and if the tilt is chosen well, the variance drops by orders of magnitude.

The difficulty is choosing $\mathbb{Q}$. In dimension d, the exponential family of tilts is parameterized by a vector $\theta \in \mathbb{R}^d$. Finding the optimal θ^* requires solving a d-dimensional optimization problem that depends on the joint distribution — itself the object one is trying to estimate. For large portfolios ($d = 100$–1000), this is a chicken-and-egg problem.

Existing approaches include: - **Exponential tilting** (Siegmund 1976): optimal for sums of i.i.d. variables, but does not exploit correlation structure. - **Cross-entropy method** (Rubinstein 1999): adaptively learns θ^*, but requires multiple simulation rounds and scales poorly with d. - **Conditional Monte Carlo** (Glasserman and Li 2005): conditions on a single systematic factor to reduce variance for CDO tranches, but uses only one eigenvalue and does not extend to the full spectrum.

1.3 The Spectral Opportunity

The eigenvalue decomposition of the portfolio correlation matrix $C = V \Lambda V^T$ decomposes the d-dimensional dependence structure into $K \leq d$ independent modes. Under eigenvalue conditioning — the central technique of the Latent framework (Nagy, *Spectral Fenton Distribution*, 2026) — the mode variables Z_k are mutually independent, being projections onto orthogonal eigenvectors.

This mode independence has a powerful consequence for importance sampling: **the measure change factorizes**. Instead of optimizing one d-dimensional tilt vector, we optimize K independent scalar tilts, one per mode. Each scalar optimization has a closed-form saddle-point solution. The total variance reduction is the *product* of per-mode reductions, yielding exponential gains in the loss threshold ℓ.

1.4 Contribution and Structure

This paper makes three contributions:

1. **Factored IS estimator** (Section 3): We construct an unbiased, logarithmically efficient estimator whose IS measure change factorizes over eigenvalue modes, with variance decaying as $\rho^{-2\ell}$ where ℓ is the loss threshold (Theorem 1).

2. **Optimal tilt via separable saddle point** (Section 4): We show that the per-mode optimal tilt solves a separable convex program, give the closed-form solution, and prove that the factored tilt is asymptotically optimal in the Donsker-Varadhan sense (Theorem 2).
3. **Analyticity–rate function duality** (Section 5): We establish that the Bernstein ellipse radius ρ of the Latent representation simultaneously controls spectral coefficient decay and the Cramér rate function, so that the domain of exact analytical computation and the domain of optimal simulation coincide (Theorem 3).

Sections 6–8 present applications to deep-tail risk estimation, barrier options, and credit portfolio losses. Section 9 discusses limitations and the boundary with exact spectral methods. Section 10 concludes.

---

2. Setup and Notation

2.1 Portfolio Model

Consider a portfolio of n assets with log-returns $X = (X_1, \dots, X_n)$ drawn from a multivariate distribution with correlation matrix C. The portfolio loss is

$$L = \sum_{i=1}^{n} w_i (1 - e^{X_i})$$

for weight vector $w \in \mathbb{R}^n$. We are interested in tail probabilities $P(L > \ell)$ for large ℓ.

2.2 Eigenvalue Decomposition

The spectral decomposition $C = V \Lambda V^T$ with $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$, $\lambda_1 \geq \dots \geq \lambda_n \geq 0$, defines mode variables

$$Z_k = \sum_{i=1}^{n} v_{ik} X_i, \qquad k = 1, \dots, n.$$

The mode variables $Z_1, \dots, Z_n$ are mutually independent by construction: they are projections of X onto the orthogonal eigenvectors of C, and orthogonal projections of a multivariate distribution with correlation matrix C onto eigenvectors of C produce uncorrelated — and for Gaussian X, independent — components. For the importance sampling framework, we exploit this mode independence directly: the IS measure change factorizes over modes because the modes themselves are independent, not because the assets become conditionally independent.

Note: the assets X_i are *not* conditionally independent given $(Z_1, \dots, Z_K)$ when $K < n$. The residual conditional covariance is $\text{Cov}(X_i, X_j \mid Z_{1..K}) = \sum_{k=K+1}^{n} \lambda_k v_{ik} v_{jk}$. The factorization of the IS measure change relies on mode independence (a property of the Z_k), not on asset conditional independence (a property of the X_i). This distinction is central to the correct statement of the variance bound.

2.3 The Latent Representation

The CDF of L admits the Latent representation

$$F_L(x) = \sum_{j=0}^{N} A_j \cos\left(\frac{j\pi(x-a)}{b-a}\right) + \varepsilon_N(x)$$

with coefficients $|A_j| \le C_F \cdot \rho^{-j}$ for analyticity radius $\rho > 1$ determined by the eigenvalue spectrum (Nagy, *Universal Spectral Representation Theorem*, 2026, Theorem 3). The parameter ρ governs both the convergence rate of the exact method and, as we shall show, the efficiency of importance sampling.

---

3. The Factored Importance Sampling Estimator

3.1 Mode-Factored Measure Change

Let $\mathbb{P}$ denote the physical measure and define, for tilt parameters $\theta = (\theta_1, \dots, \theta_K)$, the tilted measure

$$\frac{d\mathbb{Q}_\theta}{d\mathbb{P}} = \prod_{k=1}^{K} \frac{\exp(\theta_k Z_k)}{\psi_k(\theta_k)}$$

where $\psi_k(\theta) = \mathbb{E}[\exp(\theta Z_k)]$ is the moment generating function of mode k. The product form exploits the independence of modes.

Under $\mathbb{Q}_\theta$, each mode Z_k is exponentially tilted by θ_k, shifting probability mass toward the tail region.

3.2 The IS Estimator

For a measurable function g (e.g., $g(L) = \mathbf{1}\{L > \ell\}$ for tail probability or $g(L) = L \cdot \mathbf{1}\{L > \ell\}$ for expected shortfall), the IS estimator is

$$\hat{\mu}_{\text{IS}} = \frac{1}{M} \sum_{m=1}^{M} g(L^{(m)}) \cdot \prod_{k=1}^{K} \frac{\psi_k(\theta_k)}{\exp(\theta_k Z_k^{(m)})}$$

where $(Z^{(m)}, L^{(m)})$ are drawn under $\mathbb{Q}_\theta$.

Proposition 1 (Unbiasedness). *For any θ in the effective domain $\Theta = \{\theta : \psi_k(\theta_k) < \infty \ \forall k\}$, the estimator $\hat{\mu}_{IS}$ is unbiased: $\mathbb{E}_{\mathbb{Q}_\theta}[\hat{\mu}_{IS}] = \mathbb{E}_{\mathbb{P}}[g(L)]$.*

This is a direct consequence of the Radon-Nikodym theorem and the product structure of the likelihood ratio.

3.3 Variance Factorization

Theorem 1 (Logarithmic Efficiency). *Suppose the mode variables $Z_1, \dots, Z_K$ are independent under $\mathbb{P}$, and $g(L) = \mathbf{1}\{L > \ell\}$. Then at the optimal tilt θ^*, the IS estimator is logarithmically efficient:*

$$\lim_{\ell \to \infty} \frac{-\log \operatorname{Var}_{\mathbb{Q}_{\theta^*}}(\hat{\mu}_{\text{IS}})}{-\log P(L > \ell)} = 2.$$

More precisely, the IS second moment satisfies

$$\mathbb{E}_{\mathbb{Q}_{\theta^*}}\left[\left(\frac{d\mathbb{P}}{d\mathbb{Q}_{\theta^*}}\right)^2 \cdot \mathbf{1}\{L > \ell\}\right] \leq C_1(\ell) \cdot \rho^{-2\ell}$$

where $\rho > 1$ is the analyticity radius of the portfolio characteristic function, $C_1(\ell)$ is a polynomial prefactor (from the saddle-point expansion), and the per-mode thresholds ℓ_k^* are defined in Section 4.2. The computational advantage is that the optimal tilt θ^* is found in $O(K)$ operations via the separable program of Theorem 2, while the variance reduction is exponential in the loss threshold ℓ.

Proof sketch. The argument proceeds via a standard IS second-moment analysis combined with mode independence (see Appendix A for details):

1. Bound $g^2 = \mathbf{1}\{L > \ell\} \leq 1$ to upper-bound the IS second moment by $\mathbb{E}_{\mathbb{Q}_\theta}[(d\mathbb{P}/d\mathbb{Q}_\theta)^2]$.
2. By mode independence, the second moment factorizes: $\prod_{k=1}^K \psi_k(\theta_k) \cdot \psi_k(-\theta_k)$.
3. At the optimal tilt, saddle-point asymptotics give each factor as $\leq \exp(-2I_k(\ell_k^*) + \delta_k)$ where $\delta_k \to 0$ as $\ell \to \infty$.
4. Sum the rate contributions: $\sum_k I_k(\ell_k^*) \geq \tau \cdot \ell - C'$ for analytic marginals with strip width $\tau = \log \rho$, yielding the $\rho^{-2\ell}$ bound. □

---

4. Optimal Tilt: Separable Saddle-Point Program

4.1 The Optimization Problem

The optimal tilt minimizes the IS variance:

$$\theta^* = \arg\min_{\theta \in \Theta} \mathbb{E}_{\mathbb{Q}_\theta}\left[\left(\frac{d\mathbb{P}}{d\mathbb{Q}_\theta}\right)^2 \cdot g(L)^2\right].$$

Theorem 2 (Separable Convex Program). *Under mode independence, in the large-ℓ regime where the saddle-point approximation to the IS second moment is valid, the optimization decomposes into K independent scalar problems:*

$$\theta_k^* = \arg\min_{\theta_k} \left[2 \log \psi_k(\theta_k) - 2\theta_k \, \ell_k^*\right], \qquad k = 1, \dots, K.$$

Each scalar problem is strictly convex (since $\kappa_k = \log \psi_k$ is strictly convex) and admits the unique solution θ_k^* satisfying

$$\kappa_k'(\theta_k^*) = \frac{\psi_k'(\theta_k^*)}{\psi_k(\theta_k^*)} = \ell_k^*$$

i.e., the exponential tilt that shifts the mean of mode k under $\mathbb{Q}_{\theta_k^*}$ to the target threshold ℓ_k^*. For Gaussian modes with $Z_k \sim N(0, \lambda_k)$, this gives $\theta_k^* = \ell_k^*/\lambda_k$.

4.2 Threshold Allocation

The mode-level thresholds ℓ_k^* must satisfy $\sum_k \ell_k^* = \ell$ (the total loss threshold). The optimal allocation is:

$$\ell_k^* = \lambda_k \cdot \frac{\ell}{\sum_{j=1}^K \lambda_j}$$

i.e., proportional to eigenvalues. This reflects the natural risk budget: the mode that explains the most variance should receive the largest share of the loss threshold.

4.3 Asymptotic Optimality

The logarithmic efficiency of the factored estimator (the limit ratio of 2 established in Theorem 1) follows from combining the product structure of the second moment with standard large-deviation saddle-point asymptotics. By the Cramér tilt identity, the per-mode IS second moment factor at the optimal tilt converges to $\exp(-2I_k(\ell_k^*))$ as $\ell_k^* \to \infty$, and the product of these factors matches $[P(L > \ell)]^2$ up to polynomial corrections — precisely the definition of logarithmic efficiency. This is the maximum achievable for any IS scheme (Siegmund, 1976).

---

5. The Analyticity–Rate Function Duality

This section establishes the paper's deepest result: the Bernstein ellipse that governs the Latent representation and the Cramér rate function that governs rare-event probabilities are two views of the same geometric object.

5.1 The Bernstein Ellipse

The characteristic function $\phi_k(t) = \mathbb{E}[e^{itZ_k}]$ of mode k extends to a strip $|\text{Im}(t)| < \tau_k$ in the complex plane, where τ_k is the radius of analyticity. Under lognormal marginals, τ_k is determined by the eigenvalue λ_k and the volatility vector σ. The Bernstein ellipse parameter for the COS expansion is $\rho_k = e^{\tau_k}$.

5.2 The Cramér Rate Function

The Cramér rate function of mode k is

$$I_k(x) = \sup_{\theta \in \mathbb{R}} [\theta x - \log \psi_k(\theta)]$$

which is the Legendre-Fenchel transform of the cumulant generating function $\log \psi_k$. The domain of ψ_k is $\theta \in (-\tau_k, \tau_k)$ — the same strip width.

5.3 The Duality

Theorem 3 (Analyticity–Rate Duality). *For mode k with analytic characteristic function,*

(i) The effective domain of the CGF ψ_k equals the analyticity strip: $\{\theta : \psi_k(\theta) < \infty\} = (-\tau_k, \tau_k)$.

(ii) The minimum of the rate function on the rare-event set satisfies $\inf_{x>\ell_k} I_k(x) \geq \tau_k \cdot \ell_k - \log \psi_k(\tau_k^-)$.

(iii) In terms of the Bernstein parameter $\rho_k = e^{\tau_k}$:

$$P(Z_k > \ell_k) \leq \rho_k^{-\ell_k} \cdot \psi_k(\tau_k^-).$$

(iv) Consequently, the per-mode IS variance reduction at the optimal tilt is bounded by $\rho_k^{-2\ell_k^*}$, linking spectral convergence rate and simulation efficiency to the same parameter.

The geometric picture: the Bernstein ellipse is the largest ellipse in the complex plane on which the characteristic function is analytic. The rate function is the Legendre dual of the log-MGF, whose domain is the real section of this same ellipse. **Analyticity, spectral convergence, and rare-event simulation efficiency are three facets of one object: the regularity of the characteristic function.**

5.4 Unified Picture

Domain	Object	Parameter	Decay
Exact computation (COS)	Spectral coefficient A_j	ρ	$\|A_j\| \leq C\rho^{-j}$
Rare-event probability	Tail probability $P(L > \ell)$	ρ	$P(L > \ell) \leq C'\rho^{-\ell}$
Importance sampling	IS variance	ρ	$\mathrm{Var}_{\mathrm{IS}} \leq C''(\ell)\,\rho^{-2\ell}$

One parameter ρ — the analyticity radius — rules all three.

---

6. Application I: Deep-Tail VaR and ES

6.1 Motivation

Regulatory stress testing (Basel III/IV) requires VaR at 99.9% and ES at 97.5%. Internal models for credit portfolios need estimates at 99.97% or beyond. At these levels, naive Monte Carlo with 10^6 samples produces relative errors of 3–6% for the 99.9% quantile, but at the 10^{-6} level the relative error exceeds 100%, rendering the estimates useless.

6.2 Numerical Benchmark

We validate the theory on three portfolio models, each testing a different aspect of the framework. All benchmarks use $n = 30$ assets, equal weights $w_i = 1/n$, volatility $\sigma = 0.25$, nonlinear loss $L = \sum w_i(1 - e^{X_i})$, and calibrated mode tilts via binary search on the scalar strength c.

Level 1: Gaussian equicorrelation ($\bar{\rho}_{\mathrm{corr}} = 0.30$, exact mode independence, $\rho \gg 1$). The eigenvalue spectrum has $\lambda_1 = 9.70$ and gap $\lambda_1/\lambda_2 = 13.9$. Results with 500,000 naive MC samples and 50,000 IS samples:

Quantile	p	$\hat{p}_{\mathrm{MC}}$	$\mathrm{RE}_{\mathrm{MC}}$	$\hat{p}_{\mathrm{IS}}$	$\mathrm{RE}_{\mathrm{IS}}$	VR
99.0%	10^{-2}	1.01×10^{-2}	1.4%	9.87×10^{-3}	0.7%	38×
99.9%	10^{-3}	1.01×10^{-3}	4.5%	1.01×10^{-3}	0.8%	280×
99.99%	10^{-4}	1.12×10^{-4}	15.0%	1.01×10^{-4}	0.9%	2,561×
99.999%	10^{-5}	1.80×10^{-5}	60.0%	1.44×10^{-5}	1.4%	17,867×

Scaling verification: log(VR) vs ℓ^2 yields a linear fit (slope = 51.4), confirming the Gaussian quadratic rate function and the $\rho^{-2\ell}$ bound of Theorem 1.

Level 2: Gaussian sector-block correlation (5 sectors, $\rho_{\text{intra}} = 0.50$, $\rho_{\text{inter}} = 0.15$, realistic structure). The spectrum has $\lambda_1 = 7.10$, $\lambda_2 = 2.60$, gap = 2.7. Results:

Quantile	p	VR	RE_{IS}
99.9%	10^{-3}	282×	0.8%
99.99%	10^{-4}	1,916×	0.9%
99.999%	10^{-5}	35,830×	0.6%

The VR is comparable to the equicorrelation case despite the smaller spectral gap, confirming that the method works with realistic multi-sector correlation structures.

Level 3: Multivariate Student-*t* ($\nu = 4$, $\rho = 1$, no exponential VR). The same mode-factored tilt is applied to a Student-*t* portfolio (Gaussian scale mixture with χ^2_4 mixing variable). Results with linear loss $L = -w^T X$ (nonlinear loss is ill-defined for Student-*t* because $\mathbb{E}[e^X] = \infty$):

Quantile	p	VR	RE_{IS}
99.0%	10^{-2}	0.23×	9.4%
99.9%	10^{-3}	0.50×	19.4%
99.99%	10^{-4}	33×	7.8%

At the 99% and 99.9% quantiles, IS is *counterproductive* (VR < 1): the mixing variable couples all modes, destroying the independence that spectral IS exploits. Even at 99.99%, the VR of 33× is dwarfed by the 2,561× achieved by the Gaussian model at the same quantile. This confirms the $\rho = 1$ boundary prediction: without an analytic moment generating function, no exponential variance reduction is possible.

6.3 Cross-Level Summary

Model	ρ	VR @ 99.9%	VR @ 99.99%	Scaling
Gaussian equicorr	$\gg 1$	280×	2,561×	exponential
Gaussian sector	$\gg 1$	282×	35,830×	exponential
Student-*t* ($\nu = 4$)	$= 1$	0.5×	33×	polynomial

The contrast validates Theorem 3: the same spectral IS algorithm gives exponential VR when $\rho > 1$ (MGF exists) and fails when $\rho = 1$ (no MGF). The transition is sharp and diagnostic.

---

7. Application II: Multi-Asset Barrier Options

7.1 Motivation

Barrier options on baskets require path simulation because the knock-out/knock-in condition depends on the running maximum of a correlated process. No exact analytical formula exists for multi-asset barriers under general correlation.

7.2 Spectral IS for Barriers

The key adaptation: instead of tilting toward a terminal loss level, tilt toward the barrier-crossing event. The eigenvalue modes provide natural directions for correlated barrier crossing:

- Mode 1 (market mode): tilt toward joint downward movement
- Modes 2–K (sector modes): tilt toward the sector that is closest to the barrier

The barrier-crossing probability under IS concentrates samples near the barrier surface, where the payoff function changes, rather than wasting samples in the safe interior.

---

8. Application III: Credit Portfolio Losses (CDO Tranches)

8.1 The Gaussian Copula Failure and the Spectral Alternative

The Gaussian copula (Li 2000) was the standard model for CDO tranche pricing before 2008. Its catastrophic failure was driven by zero upper tail dependence: $\lambda_U = 0$. The Fenton Copula (Nagy, *An Eigenvalue-Conditioned Copula with Positive Tail Dependence*, 2026) addresses this structural flaw.

For importance sampling in the credit context, the eigenvalue decomposition of the default correlation matrix provides the natural conditioning structure. The mode-factored IS tilts the default intensity of each mode independently, concentrating samples on the correlated default scenarios that drive tranche losses.

8.2 Connection to Glasserman-Li

Glasserman and Li (2005) pioneered IS for CDO pricing by conditioning on a single systematic factor (the first eigenvalue). Our framework generalizes this to K eigenvalues, providing:

- Computationally tractable optimal tilt: K independent scalar saddle-point problems, each solvable in closed form, versus a heuristic single-factor shift
- Deeper tail reach: the $\rho^{-2\ell}$ bound means variance reduction improves exponentially as the target threshold ℓ increases, not just as more modes are added
- Explicit optimality guarantee: Theorem 2 gives the provably optimal tilt, not a heuristic approximation

---

9. Limitations and Boundary with Exact Methods

9.1 When NOT to Use Spectral IS

The COS-based exact methods (Nagy, *Spectral Fenton Distribution* and *Universal Spectral Representation Theorem*, 2026) compute the full CDF deterministically with error bounded by $C\rho^{-N}$. For standard risk measures (VaR, ES, spectral risk measures) at conventional levels (95%–99.9%), **the exact method is strictly superior**: no sampling noise, no variance, deterministic runtime. Spectral IS is not a replacement for exact computation — it is the extension into domains where exact computation does not reach.

9.2 When to Use Spectral IS

The method is valuable when: 1. **Path dependence** prevents a single-period analytical solution (barriers, lookbacks, dynamic hedging). 2. **Ultra-deep tails** ($< 10^{-6}$) require accuracy beyond practical COS truncation. 3. **Dynamic models** (multi-period, regime-switching) make the state space too large for deterministic quadrature. 4. **Non-smooth payoffs** (digital options, tranche losses) cause COS Gibbs artifacts.

9.3 Structural Assumptions

The variance reduction guarantees require: - **Spectral gap**: $\lambda_1/\lambda_2 > 1$. Without a gap, the modes do not separate cleanly. - **Analytic marginals**: the exponential decay bound requires $\rho > 1$, which holds for Gaussian, lognormal, NIG, and other distributions whose characteristic function extends analytically to a complex strip. Heavy-tailed distributions with $\rho = 1$ — including Student-t (no MGF exists), Pareto, and stable distributions — get polynomial, not exponential, reduction. - **Moderate dimension**: the K modes must capture a substantial fraction of variance. If the eigenvalue spectrum is flat ($\lambda_k \approx \text{const}$), conditioning provides no leverage.

---

10. Conclusion

The eigenvalue decomposition of portfolio correlation is not merely a computational convenience — it is a structural key that unlocks optimal simulation of rare events. The mode independence property that makes the COS-based Latent representation exact also makes importance sampling factorizable, and the analyticity parameter ρ that governs spectral convergence simultaneously governs the rate function of large deviations.

This duality — Theorem 3 of this paper — suggests that analyticity is a more fundamental regularity concept than previously appreciated: it controls not only how efficiently a distribution can be *represented* but also how efficiently its extremes can be *simulated*.

Three open directions follow:

1. **Adaptive spectral IS**: learn the optimal K (number of conditioning modes) on the fly, starting with $K = 1$ and adding modes until the target variance is met.
2. **Non-Gaussian extensions**: for marginals with polynomial CF decay ($\rho = 1$), characterize the polynomial variance reduction and its dependence on the Sobolev regularity exponent.
3. **Dynamic spectral IS**: extend the factored measure change to multi-period settings where the correlation structure evolves, connecting to the spectral trading theory (Nagy, *Frequency-Domain Theory of Trading Strategies*, 2026) and the fin_harvestability framework (Nagy, *Harvestability*, 2026).

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

- Cramér, H (1938). Sur un nouveau théorème-limite de la théorie des probabilités. *Actualités Scientifiques et Industrielles*, 5-23.
- Donsker, M. D. and Varadhan, S. R. S (1975). Asymptotic evaluation of certain Markov process expectations for large time, I. *Communications on Pure and Applied Mathematics*, 28(1), 1-47.
- Glasserman, Paul (2003). Monte Carlo Methods in Financial Engineering. Springer.
- Glasserman, P. and J. Li (2005). Importance sampling for portfolio credit risk. *Management Science*, 51(11), 1643-1656.
- Li, D.X. (2000). On default correlation: a fin_copula function approach. *Journal of Fixed Income*, 9(4), 43-54. DOI: 10.3905/jfi.2000.319253
- 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). An Eigenvalue-Conditioned Copula with Positive Tail Dependence: A Machine-Verified Alternative to the Gaussian Copula. *Working paper.*
- Nagy, T. (2026). Frequency-Domain Theory of Financial Economics: Thirteen Fundamental Results from One Decomposition. *Working paper.*
- Nagy, T. (2026). Harvestability. *Working paper.*
- Rubinstein, R. Y (1999). The cross-entropy method for combinatorial and continuous optimization. *Methodology and Computing in Applied Probability*, 1(2), 127-190.
- Siegmund, D (1976). Importance sampling in the Monte Carlo study of sequential tests. *Annals of Statistics*, 4(4), 673-684.

Appendix A: Proof of Theorem 1 (Variance Factorization)

Step 1. The IS second moment is $M_2 = \mathbb{E}_{\mathbb{Q}_\theta}[(d\mathbb{P}/d\mathbb{Q}_\theta)^2 \cdot g^2]$ where $g = \mathbf{1}\{L > \ell\}$. Since $g^2 = g \le 1$, we have $M_2 \le \mathbb{E}_{\mathbb{Q}_\theta}[(d\mathbb{P}/d\mathbb{Q}_\theta)^2]$.

Step 2. The per-mode likelihood ratio is $d\mathbb{P}_k/d\mathbb{Q}_{\theta_k} = \psi_k(\theta_k)/e^{\theta_k Z_k}$. By mode independence, the squared likelihood ratio factorizes:

$$\mathbb{E}_{\mathbb{Q}_\theta}\left[\left(\frac{d\mathbb{P}}{d\mathbb{Q}_\theta}\right)^2\right] = \prod_{k=1}^{K} \mathbb{E}_{\mathbb{Q}_{\theta_k}}\left[\frac{\psi_k(\theta_k)^2}{e^{2\theta_k Z_k}}\right].$$

Step 3. Each per-mode factor evaluates to $\psi_k(\theta_k) \cdot \psi_k(-\theta_k)$. To see this: under $\mathbb{Q}_{\theta_k}$ the density is $e^{\theta_k z}/\psi_k(\theta_k) \cdot f_k(z)$, so

$$\mathbb{E}_{\mathbb{Q}_{\theta_k}}\left[\frac{\psi_k(\theta_k)^2}{e^{2\theta_k Z_k}}\right] = \psi_k(\theta_k)^2 \cdot \frac{1}{\psi_k(\theta_k)} \int e^{-\theta_k z} f_k(z)\, dz = \psi_k(\theta_k) \cdot \psi_k(-\theta_k).$$

Step 4. In terms of the cumulant generating function $\kappa_k = \log \psi_k$, the per-mode factor is $\exp(\kappa_k(\theta_k) + \kappa_k(-\theta_k))$. At the optimal tilt θ_k^* satisfying $\kappa_k'(\theta_k^*) = \ell_k^*$, saddle-point asymptotics give $\kappa_k(\theta_k^*) + \kappa_k(-\theta_k^*) \le -2I_k(\ell_k^*) + \delta_k$ where $\delta_k \to 0$ as $\ell_k^* \to \infty$ and I_k is the Cramér rate function.

Step 5. Bound $I_k(\ell_k^*) \geq \tau_k \ell_k^* - \log \psi_k(\tau_k^-)$ using the Chernoff bound at the strip boundary. For large ℓ_k^*, the leading term is $\tau_k \ell_k^*$ where $\rho_k = e^{\tau_k}$.

Step 6. Under eigenvalue-proportional allocation $\ell_k^* = \lambda_k \ell / \sum_j \lambda_j$ and the common strip width $\tau_k \geq \tau = \log \rho$, summing the rate contributions gives $\sum_k I_k(\ell_k^*) \geq \tau \cdot \ell - C'$, yielding $M_2 \leq C_1(\ell) \cdot \rho^{-2\ell}$ in the large-ℓ regime.

Appendix B: Proof of Theorem 2 (Separable Convex Program)

Step 1. From Appendix A, the IS second moment is $M_2(\theta) = \prod_{k=1}^K \psi_k(\theta_k) \cdot \psi_k(-\theta_k)$. The IS variance is $\mathrm{Var} = M_2(\theta) - \mu^2$ where $\mu = P(L > \ell)$. Minimizing Var over θ is equivalent to minimizing $M_2(\theta)$ (since μ is independent of θ).

Step 2. Taking logarithms: $\log M_2(\theta) = \sum_{k=1}^K [\kappa_k(\theta_k) + \kappa_k(-\theta_k)]$ where $\kappa_k = \log \psi_k$ is the cumulant generating function. This is a sum of separable terms.

Step 3. Each term $h_k(\theta_k) = \kappa_k(\theta_k) + \kappa_k(-\theta_k)$ is strictly convex: $h_k''(\theta_k) = \kappa_k''(\theta_k) + \kappa_k''(-\theta_k) > 0$ since $\kappa_k'' = \mathrm{Var}_{\mathbb{Q}_{\theta_k}}(Z_k) > 0$. The sum of separable strictly convex functions attains its minimum componentwise.

Step 4. The saddle-point constraint $\kappa_k'(\theta_k^*) = \ell_k^*$ selects the tilt that shifts the mode mean to the target threshold. By strict convexity of κ_k, this equation has a unique solution for each ℓ_k^* in the range of κ_k'. $\square$

Appendix C: Proof of Theorem 3 (Analyticity–Rate Duality)

Step 1 (Domain equivalence). The moment generating function $\psi_k(\theta) = \mathbb{E}[e^{\theta Z_k}]$ is finite iff θ lies in the effective domain $D_k = \{\theta : \psi_k(\theta) < \infty\}$. The characteristic function $\phi_k(t) = \psi_k(it)$ extends analytically to the strip $\{t \in \mathbb{C} : |\mathrm{Im}(t)| < \tau_k\}$ where $\tau_k = \sup\{\theta > 0 : \theta \in D_k\}$. This is the Paley-Wiener theorem: exponential integrability of a measure and strip-analyticity of its Fourier transform are dual conditions.

Step 2 (Rate function at the boundary). The Cramér rate function $I_k(x) = \sup_{\theta \in D_k}[\theta x - \kappa_k(\theta)]$ is the Legendre-Fenchel transform of κ_k. For $x > \mathbb{E}[Z_k]$, the supremum is attained at $\theta^*(x) \in (0, \tau_k)$ satisfying $\kappa_k'(\theta^*(x)) = x$. As $x \to \infty$, $\theta^*(x) \to \tau_k^-$ and $I_k(x) \sim \tau_k \cdot x - \kappa_k(\tau_k^-)$.

Step 3 (Chernoff bound). By Markov's inequality: $P(Z_k > x) \leq \inf_{\theta > 0} e^{-\theta x} \psi_k(\theta) = e^{-I_k(x)}$. The leading-order bound is $P(Z_k > x) \leq \psi_k(\tau_k^-) \cdot e^{-\tau_k x} = \psi_k(\tau_k^-) \cdot \rho_k^{-x}$.

Step 4 (Bernstein ellipse identification). The COS spectral coefficients A_j of the density of Z_k satisfy $|A_j| \leq C_0 \cdot r^{-j}$ where r is the semi-axis of the largest Bernstein ellipse on which the characteristic function is analytic. The semi-axis in the imaginary direction is τ_k, giving $r = e^{\tau_k} = \rho_k$.

Step 5 (Duality). Combining Steps 1–4: the parameter $\rho_k = e^{\tau_k}$ simultaneously controls (i) spectral coefficient decay: $|A_j| \leq C_0 \rho_k^{-j}$, (ii) tail probability: $P(Z_k > x) \leq C_1 \rho_k^{-x}$, and (iii) IS variance reduction: $\mathrm{VR}(\ell_k) \geq C_2^{-1} \rho_k^{2\ell_k}$ (from Steps 5–6 of Appendix A). All three are governed by the strip width $\tau_k = \log \rho_k$. $\square$

Appendix D: Numerical Benchmark Code

The complete Python implementation validating all three levels of the benchmark is available as spectral_is_benchmark.py in the companion repository. Run with:

```
python spectral_is_benchmark.py --plot
```

This produces the cross-level comparison figure and raw results in JSON format. The code implements: equicorrelation and sector-block correlation matrices, eigenvalue decomposition with automatic K selection, mode-factored Gaussian and Student-t sampling with tilted distributions, binary-search tilt calibration, and scaling verification (log(VR) vs ℓ and ℓ^2).