Universal Approximation Theorems for Spectral Decision Functionals

Dr. Tamás Nagy

tnagyphd@gmail.com

Draft

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.

Reader-Friendly Subtitle

Approximation guarantees for decision operators, not only predictors.

Technical Strapline

Density and rate theorems for spectral representations of risk, allocation, and control functionals.

Executive Summary (Non-Technical)

Most approximation theory in ML focuses on predicting values. In deployment, however, we often need robust decision functionals: pricing maps, risk maps, and allocation rules.

This paper moves approximation guarantees directly to that functional level in spectral coordinates. The target is both existence and quantitative rates.

If established, this creates a theorem-backed basis for reusable decision libraries with clear error and cost trade-offs.

The paper does not claim every functional is equally easy to approximate. Rate quality depends on regularity and basis alignment assumptions.

Abstract

We prove universal approximation results for a broad class of decision functionals represented in spectral coordinates. The theorem characterizes expressivity in terms of basis regularity, coefficient decay, and functional smoothness, and provides quantitative approximation rates. This establishes a common approximation backbone for risk, allocation, and control functionals.

1. Problem

Universal approximation theorems in ML (Cybenko 1989, Hornik 1991) guarantee that neural networks can approximate any continuous function. But in risk-aware deployment, we do not approximate functions — we approximate decision functionals: maps from probability distributions to actions, risk numbers, or prices.

The question: do spectral representations provide universal approximation for this functional class, and what are the quantitative rates? Density alone is not enough; we need explicit error-rate bounds as a function of spectral budget and regularity.

2. Setup

2.1 Decision Functional Space

Let $\mathcal{P}$ be a set of probability distributions. A decision functional is:

$$\Phi : \mathcal{P} \to \mathbb{R}$$

Canonical examples: - Risk measures: $\Phi(P) = \mathrm{CVaR}_\alpha(P)$, $\mathrm{VaR}_\alpha(P)$ - Pricing functionals: $\Phi(P) = \mathbb{E}^Q[X]$ for some pricing measure - Allocation maps: $\Phi(P) = w^*(P)$ solving Markowitz under distribution P

Definition 1 (Lipschitz Decision Functional). Φ is L-Lipschitz in spectral distance if:

$$|\Phi(P) - \Phi(Q)| \leq L \cdot d_R(P, Q) = L \cdot \|\mathbf{A}(P) - \mathbf{A}(Q)\|_2$$

2.2 Spectral Approximation Architecture

The K-mode spectral approximation of Φ is:

$$\Phi_K(P) = f(A_0(P), A_1(P), \dots, A_{K-1}(P))$$

where $f : \mathbb{R}^K \to \mathbb{R}$ is a function of the first K COS coefficients.

2.3 Regularity

Definition 2 (Analytic Coefficient Decay). Distribution P has spectral gap $\rho > 1$ if $|A_k(P)| \leq C\rho^{-k}$.

Connection to existing kernel: the URRT (Universal/MainTheorem.lean) proves that CDF approximation achieves geometric convergence $\|F - F_K\|_\infty \leq C\rho^{-K}$ under analytic regularity. The present paper lifts this to functionals of the distribution.

3. Main Theorem

Theorem Candidate 1 (Universal Approximation of Decision Functionals). For any L-Lipschitz decision functional Φ and any distribution family with spectral gap $\rho > 1$, there exists a function $f_K : \mathbb{R}^K \to \mathbb{R}$ such that:

$$|\Phi(P) - f_K(\mathbf{A}_K(P))| \leq L \cdot C \cdot \rho^{-K}$$

for all P in the family. The approximation is: - **Dense:** as $K \to \infty$, the error vanishes. - **Explicit-rate:** geometric in K with rate ρ^{-K}. - **Dimension-free:** the bound does not depend on the ambient dimension of the distribution.

Theorem Candidate 2 (Rate Optimality). For the class of all L-Lipschitz functionals, the rate ρ^{-K} is optimal up to constant factors: there exists a functional Φ^* with:

$$\inf_{f_K} |\Phi^*(P) - f_K(\mathbf{A}_K(P))| \geq c \cdot \rho^{-K}$$

Corollary (CVaR Approximation). CVaR_α is Lipschitz in spectral distance with $L \leq 1/(1-\alpha)$. Therefore:

$$|\mathrm{CVaR}_\alpha(P) - \mathrm{CVaR}_\alpha(P_K)| \leq \frac{C}{1-\alpha}\rho^{-K}$$

4. Proof Sketch

1. **CDF approximation.** From URRT, $\|F - F_K\|_\infty \leq C\rho^{-K}$.
2. **Lipschitz transfer.** For L-Lipschitz Φ: $|\Phi(P) - \Phi(P_K)| \leq L \cdot d_R(P, P_K) \leq L \cdot \|F - F_K\|_\infty \leq LC\rho^{-K}$.
3. **Function construction.** Set $f_K(a_0, \dots, a_{K-1}) = \Phi(\hat{P}_K)$ where $\hat{P}_K$ is the distribution with COS coefficients $(a_0, \dots, a_{K-1}, 0, \dots)$.
4. **Lower bound.** Construct a functional that depends precisely on mode K: $\Phi^*(P) = A_K(P)$. Any K-mode approximation must miss this entirely.

5. Empirics/Simulation

5.1 Risk Measure Approximation

- VaR, CVaR, spectral risk measures for lognormal, Student-t, and mixture distributions.
- Report error as a function of K.
- Validate geometric convergence rate.

5.2 Allocation Functional

- Markowitz weights as a functional of the joint distribution.
- Compare spectral approximation vs Monte Carlo.
- Report: approximation error, computation time, and stability.

5.3 Pricing Functional

- European option price as functional of the underlying distribution.
- Compare spectral pricing vs Black-Scholes (analytic) vs FFT.

6. Limits

- **Non-Lipschitz functionals:** VaR is discontinuous; the bound applies only via smoothed versions.
- **Basis mismatch:** if the distribution family does not have geometric decay, rates are subgeometric.
- **High-dimensional joint distributions:** the COS basis works naturally for marginals; joint distributions need tensor or copula extensions.

- **Model risk:** the approximation is only as good as the spectral representation of the input distribution.

7. Related Work

- **Universal approximation:** Cybenko (1989), Hornik (1991) — function approximation.
- **Operator learning:** Li et al. (2020) — Fourier neural operators; Chen-Chen (1995) — operator networks.
- **Functional approximation in finance:** Fang-Oosterlee (2008) COS method; our URRT paper.
- **Risk measure continuity:** Cont et al. (2010) — robustness of risk measures.

8. Cross-Paper Connections

- **Minimal Sufficient State (paper 4):** the approximation budget K here is the same object as K^* in paper 4. Paper 4 derives K^* from sufficiency; this paper provides the approximation rate at that budget. Together they close the loop: K^* modes are both necessary (paper 4) and sufficient at rate ρ^{-K^*} (this paper).
- **Nonlinear FTAP (paper 2):** the nonlinear pricing functional Π from paper 2 is a decision functional. This paper guarantees that Π can be approximated at rate ρ^{-K} when Π is Lipschitz in spectral distance.
- **Info Geometry Bridge (paper 6):** the Lipschitz constant L of a decision functional in spectral distance is bounded by L_F/c_1 via the metric sandwich. This means Fisher-information-optimal estimation rates transfer to decision-functional approximation rates.
- **Ergodic Control (paper 7):** the value function of the ergodic controller is itself a decision functional of the state distribution. This paper's rates apply to value-function approximation.

9. Extension to Sobolev-Regular Distributions

9.1 Beyond Analytic Regularity

The analytic regime ($|A_k| \leq C\rho^{-k}$, geometric decay) covers Gaussian, lognormal, and exponential families. But many distributions encountered in practice — mixtures with rough boundaries, empirical distributions, distributions with kinks — are only Sobolev-regular.

Definition 3 (Sobolev Spectral Regularity). Distribution P has Sobolev regularity $r > 0$ if $|A_k(P)| \leq Ck^{-r}$ (polynomial decay, not geometric).

9.2 Algebraic Approximation Rate

Theorem Candidate 3 (Sobolev-Class Approximation). For r-Sobolev distributions and L-Lipschitz decision functionals:

$$|\Phi(P) - f_K(\mathbf{A}_K(P))| \leq L \cdot C \cdot K^{-(r-1/2)}$$

provided $r > 1/2$ (so the COS series converges in L^∞).

Proof sketch: the CDF tail error is $\|F - F_K\|_\infty \le C \sum_{k>K} k^{-r} \le C' K^{-(r-1)}$. The Lipschitz transfer gives the decision error bound. The extra 1/2 comes from Parseval's inequality in the Sobolev norm.

9.3 Rate Comparison

Regularity	Coefficient decay	Approximation rate	Example families
Analytic	ρ^{-k}	ρ^{-K} (geometric)	Gaussian, lognormal, Student-t
Sobolev-r	k^{-r}	$K^{-(r-1/2)}$ (algebraic)	Mixtures, empirical, piecewise
L^2 only	no decay guarantee	no uniform rate	Cantor distributions

The gap between analytic and Sobolev rates is significant: for $K = 32$, analytic gives $\sim 10^{-14}$ error while Sobolev-3 gives $\sim 10^{-3}$. This motivates using analytic approximations (e.g., Gaussian mixtures) as surrogates when the raw distribution is only Sobolev.

10. Joint Distribution Handling via Tensor COS

10.1 Multivariate Extension

For a d-dimensional distribution P on $\mathbb{R}^d$, the COS coefficients become tensorial:

$$A_{\mathbf{k}} = A_{k_1,\dots,k_d}$$

with $\mathbf{k} \in \{0, \dots, N-1\}^d$. The total number of coefficients is N^d — the curse of dimensionality.

10.2 Sparse Tensor Approximation

Strategy 1 (Hyperbolic cross): Retain only multi-indices with $\prod_j (k_j + 1) \le M$. This gives $O(M(\log M)^{d-1})$ coefficients instead of M^d, preserving the geometric rate for product-type regularity.

Strategy 2 (Copula + marginals): Separate the joint distribution into marginals (1D COS, fast) and copula (dependence structure, lower-dimensional). The copula spectral decomposition operates in $[0,1]^d$ and has its own coefficient decay, typically faster due to the bounded domain.

10.3 Decision Functional Rate for Joint Distributions

Theorem Candidate 4 (Multivariate Decision Approximation). For separable d-dimensional distributions with analytic marginals and analytic copula, the decision functional approximation rate is:

$$|\Phi(P) - \Phi_K(P)| \le L \cdot d \cdot C \cdot \rho_{\min}^{-K}$$

where $\rho_{\min} = \min(\rho_1, \dots, \rho_d, \rho_{\text{copula}})$ and K is the per-dimension mode budget.

The factor d is the cost of dimensionality — linear, not exponential — under the copula+marginals decomposition. This is a key advantage over full tensor approaches.

11. Outlook

- **Decision libraries:** deploy pre-computed f_K as reusable decision modules with certified error bounds. The Sobolev extension (Section 9) expands the deployable distribution family.
- **Online decision update:** as new data updates $\mathbf{A}(P)$, the decision updates in $O(K)$.
- **Lean formalization:** Theorem 1 follows almost directly from existing URRT chain plus Lipschitz composition; target: LeanProofs/DecisionApproximation/UniversalTheorem.lean. The Sobolev rate (Theorem 3) requires extending the URRT to polynomial-decay regimes.
- **Bridge to Minimal Sufficient State:** the K^* from paper 4 is the same object as the decision-functional approximation budget here.
- **Multivariate priority:** the copula+marginals decomposition (Section 10) is the most promising path to high-dimensional deployment and should be the first experimental target.
- **Practical benchmark:** compare spectral decision approximation vs Monte Carlo for a 10-asset CVaR optimization, measuring both accuracy and wall-clock time.