Spectral Ergodic Control with Provable Regret Guarantees

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

One framework for long-run stability and finite-time decision quality.

Technical Strapline

Joint ergodic and regret guarantees via mode-wise contraction and uncertainty propagation.

Executive Summary (Non-Technical)

Control systems need two things at once: stable long-run behavior and strong short-run performance. Most methods emphasize one and weaken the other.

This paper proposes a spectral control framework where both can be analyzed in the same state representation. The same decomposition supports ergodic guarantees and finite-time regret bounds.

The practical implication is better policy design for systems where reliability and adaptivity both matter.

The paper does not claim universal optimality. It provides explicit policy classes and guarantee regions.

Abstract

Ergodic control and online regret analysis are usually developed with different tools and assumptions. We present a spectral ergodic control framework where long-run average control objectives and finite-time regret can be analyzed jointly. Our main results provide policy classes with explicit stability and regret bounds derived from mode-wise contraction and uncertainty propagation.

1. Problem

Control theory traditionally splits into two camps: ergodic control focuses on long-run average cost under stationarity, while online learning focuses on finite-time regret against a best-in-hindsight comparator. Practitioners need both — a trading strategy must be stable over years while remaining competitive in each quarter.

We propose a spectral decomposition that lets both guarantees coexist: each spectral mode gets its own contraction guarantee (ergodic), and the aggregate policy gets a finite-time regret bound that decomposes across modes.

2. Setup

2.1 Controlled Stochastic Process

Let the system evolve in spectral coordinates:

$$dc_k(t) = \left(-\frac{c_k(t)}{\tau_k} + u_k(t)\right) dt + \sigma_k \, dW_k(t)$$

where c_k is mode k, τ_k is the mean-reversion timescale, $u_k(t)$ is the control action on mode k, and σ_k is mode volatility.

2.2 Cost Structure

The per-period cost is:

$$\ell(c, u) = \sum_{k=1}^{K} [q_k c_k^2 + r_k u_k^2]$$

The ergodic cost is $J_{\text{erg}}(\pi) = \lim_{T\to\infty} \frac{1}{T}\mathbb{E}\left[\int_0^T \ell(c(t), u(t))\, dt\right]$.

2.3 Policy Class

Definition 1 (Spectral Linear Policy). $u_k(t) = -\alpha_k c_k(t)$ with gain $\alpha_k > 0$.

The effective contraction rate is $\gamma_k = 1/\tau_k + \alpha_k$.

2.4 Regret Notion

Definition 2 (Mode-Decomposed Regret).

$$R_T = \sum_{k=1}^{K} R_{T,k} = \sum_{k=1}^{K} \left[\int_0^T \ell_k(c_k(t), u_k(t))\, dt - T \cdot J_k^*\right]$$

where J_k^* is the optimal ergodic cost for mode k.

Connection to existing kernel: the OU mode dynamics are the same as in the Harvestability paper (Harvestability/OUDynamics.lean), where τ_k governs the fin_harvestability. The control u_k accelerates harvesting.

3. Main Theorem

Theorem Candidate 1 (Ergodic Stability). Under the spectral linear policy with $\alpha_k > 0$ for all k:

$$J_{\text{erg}} = \sum_{k=1}^{K} \frac{q_k \sigma_k^2}{2\gamma_k} + r_k \alpha_k \cdot \frac{\sigma_k^2}{2\gamma_k}$$

The optimal gain is $\alpha_k^* = \sqrt{q_k/r_k} - 1/\tau_k$ (when positive), yielding:

$$J_{\text{erg}}^* = \sum_{k=1}^{K} \sigma_k^2 \sqrt{q_k r_k}$$

Theorem Candidate 2 (Finite-Time Regret). Under the same policy, the finite-time regret satisfies:

$$R_T \leq \sum_{k=1}^{K} \frac{C_k}{\gamma_k} (1 + \log T)$$

where $C_k = q_k \sigma_k^2/(2\gamma_k^2)$ depends on mode-specific parameters.

Corollary (Joint Guarantee). The spectral policy is simultaneously ergodically optimal (matching J^*) and logarithmic-regret, with the regret growing as $O(K \log T)$.

4. Proof Sketch

1. **Ergodic component.** Under OU dynamics with control, the stationary distribution of c_k is $\mathcal{N}(0, \sigma_k^2/(2\gamma_k))$. The ergodic cost follows by substitution. Optimal α_k is found by first-order condition.
2. **Regret decomposition.** The regret for mode k is the transient deviation from steady state. Using the OU Ornstein-Uhlenbeck concentration bounds, $\mathbb{E}[(c_k(t) - c_k^{\text{ss}})^2] \leq Ce^{-2\gamma_k t} +$ fluctuation.
3. **Integration.** Summing the transient over $[0, T]$: $R_{T,k} \leq C_k(1 + \log T)/\gamma_k$ by standard exponential integral bounds.
4. **Aggregation.** Sum over modes; the logarithmic dependence on T is inherited from each mode independently.

5. Empirics/Simulation

5.1 Synthetic Multi-Mode OU

- 10-mode OU system with $\tau_k \in [0.1, 10]$, $\sigma_k \in [0.5, 2.0]$.
- Compare spectral policy vs LQR vs model-free RL.
- Report: ergodic cost, regret trajectory, stability margin.

5.2 Portfolio Rebalancing

- Equity factor model with 5 PCA factors.
- Interpret rebalancing as spectral control.
- Measure: Sharpe ratio, max drawdown, and transaction cost.

5.3 Market Making

- Simplified Avellaneda-Stoikov model in spectral coordinates.
- Compare inventory control with and without mode-decomposed policy.

6. Limits

- **Weak mixing:** if $\gamma_k \to 0$ for some modes, regret bound degrades.
- **Model misspecification:** incorrect τ_k or σ_k estimates lead to suboptimal gains.
- **Partial observability:** if some modes are unobserved, the control is incomplete (connects to the Causal Identifiability paper).
- **Nonlinear dynamics:** the OU/linear-control setting does not extend trivially to nonlinear systems.

7. Related Work

- **Ergodic control:** Arapostathis et al. (2012), Borkar (1989) — long-run average cost optimization.
- **Online learning / regret:** Auer et al. (2002), Hazan (2016) — regret bounds for bandits and control.
- **Risk-sensitive control:** Whittle (1990), Borkar-Meyn (2002).
- **Spectral control:** our existing Harvestability paper provides the OU mode structure; the present paper adds active control and regret analysis.

8. Cross-Paper Connections

- **Universality (paper 1):** the mode contraction rates γ_k must respect the universality class structure. If the underlying dynamics belong to class $\mathcal{U}_s$, then $\tau_k \sim k^s$ and the control policy must work within that scaling law.
- **Causal Identifiability (paper 3):** modes that the controller acts on must be observable and causally identifiable. The observability index o_k from paper 3 constrains which modes can enter the control loop. Unidentifiable modes contribute uncontrollable noise.
- **Phase Transitions (paper 5):** near a generalization/control phase transition, the contraction rates for critical modes approach zero. The regret bound $R_T \leq \sum C_k/\gamma_k \cdot (1 + \log T)$ shows that the regret diverges exactly at the phase boundary.
- **Minimal Sufficient State (paper 4):** the controller only needs the first K^* modes. The spectral ergodic cost decomposes, and modes beyond K^* contribute negligible cost reduction relative to their estimation cost.

9. Adaptive Gain Estimation via Thompson Sampling

9.1 Problem

The optimal gain $\alpha_k^* = \sqrt{q_k/r_k} - 1/\tau_k$ requires known τ_k, σ_k. In practice, these are unknown and must be estimated online.

9.2 Mode-Wise Thompson Sampling

Maintain a posterior $p(\tau_k, \sigma_k \mid \text{data}_k)$ for each mode independently.

Algorithm (per mode k): 1. Sample $(\hat{\tau}_k, \hat{\sigma}_k) \sim p(\tau_k, \sigma_k \mid \text{data}_k)$. 2. Compute $\hat{\alpha}_k = \sqrt{q_k/r_k} - 1/\hat{\tau}_k$. Clip to $[\alpha_{\min}, \alpha_{\max}]$. 3. Apply control $u_k(t) = -\hat{\alpha}_k c_k(t)$ for the current epoch. 4. Observe trajectory $\{c_k(t)\}_{t \in \text{epoch}}$. 5. Update posterior: $p(\tau_k, \sigma_k) \leftarrow p(\tau_k, \sigma_k \mid \{c_k(t)\})$.

9.3 Regret of Adaptive Policy

Theorem Candidate 3 (Bayesian Regret of Adaptive Spectral Control). Under the Thompson sampling policy with conjugate priors, the Bayesian regret satisfies:

$$\mathbb{E}[R_T^{\text{adapt}}] \leq \sum_{k=1}^{K} \frac{C_k}{\gamma_k^*} \left(\sqrt{T \log T} + \log T \right)$$

The first term reflects the exploration cost; the second is the steady-state regret as in Theorem 2. The $\sqrt{T \log T}$ exploration cost is the price of not knowing τ_k, σ_k.

9.4 Posterior Convergence

For the OU process, the sufficient statistics for (τ_k, σ_k) are the lag-1 autocorrelation and the marginal variance of $c_k(t)$. Both converge at rate $O(1/\sqrt{T})$, so the posterior concentrates and the policy converges to the oracle gain α_k^*.

10. Nonlinear Extension Sketch

10.1 Beyond OU Dynamics

Replace the linear OU dynamics with a nonlinear mean-reverting SDE per mode:

$$dc_k = -g_k(c_k)\, dt + u_k\, dt + \sigma_k(c_k)\, dW_k$$

where g_k is a nonlinear restoring force (e.g., $g_k(c) = c/\tau_k + \beta c^3$ for a double-well potential).

10.2 Mode-Wise Lyapunov Stability

If each g_k admits a Lyapunov function $V_k(c_k)$ with $\mathcal{L}V_k \leq -\gamma_k V_k + D_k$ under the controlled dynamics, then:

- **Ergodic stability:** the process is positive recurrent with stationary measure.
- **Regret decomposition:** the mode-wise regret analysis extends with γ_k replaced by the effective contraction rate of the Lyapunov function.

10.3 Limitations

The mode-wise decoupling breaks down for nonlinear dynamics with significant cross-mode coupling. When $|g_{jk}(c_j, c_k)|$ is not small relative to $|g_k(c_k)|$, the regret analysis requires a joint multi-mode treatment. This is a genuine open problem connecting to nonlinear control theory and is deferred to future work.

11. Outlook

- **Adaptive control:** the Thompson sampling protocol (Section 9) provides a deployable algorithm with explicit Bayesian regret guarantees.
- **Multi-agent extension:** decentralized spectral control where agents manage different modes. Each agent runs its own Thompson sampler and communicates mode correlations.
- **Lean formalization:** the ergodic cost formula and optimal gain are algebraic and well-suited for LeanProofs/SpectralControl/ErgodicCost.lean. The Bayesian regret bound (Theorem 3) is a second formalization target.
- **Bridge to Spectral RL:** the regret bound here provides a foundation for the RL paper (paper 10), where the policy class is extended beyond linear to neural or kernel policies.
- **Nonlinear priority:** the Lyapunov extension (Section 10) identifies the precise boundary between tractable and open problems in spectral control.