RAPID MIXING IS NOT ENOUGH: HARDNESS OF ACTION-CONDITIONED PREDICTIVE REPRESENTATION OPTIMIZATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Action-conditioned self-prediction learns shared state representations for reinforcement learning by predicting future representations under different actions. However, existing spectral solutions for the associated tabular objective rely on a common eigenbasis, leaving the complexity of global optimization beyond this setting unresolved. We study this question for centered and whitened representations in finite state spaces, removing the constant component and normalizing feature covariance. We prove worst-case computational hardness of approximating the optimal objective value to inverse polynomial additive accuracy, even for strictly positive, symmetric transition kernels that mix rapidly and retain fixed signal strength after centering. Our construction embeds a hard graph problem into valid transition kernels, and a further reduction extends the hardness to every fixed representation dimension. For symmetric kernels, we give exact solutions when they commute and efficient additive approximation for a single feature when the centered operators span a space of fixed dimension. The hard instances also admit exact realizations as finite transition datasets. For the scalar empirical cross-covariance objective, polynomially many independent stationary samples preserve the hardness gap with high probability, even when centering, whitening, and representation selection use the same data.

1 INTRODUCTION

Learning predictive state representations is a central problem in reinforcement learning because agents need to model how their actions change the environment (Schwarzer et al., 2021; Fujimoto et al., 2025). Action-conditioned self-prediction addresses this problem by predicting a future state representation from the current representation and the chosen action (Khetarpal et al., 2024). A shared encoder places the predictions for different actions in a common feature space, whose quality depends on the information it preserves across all actions. Understanding the cost of optimizing that shared space is therefore fundamental to this learning objective (Figure 1). However, the unresolved question is how efficiently one can optimize informative shared features when different actions favor different predictive directions.

Spectral methods construct features from diffusion operators (Wu et al., 2018; Gomez et al., 2024) or factor transition dynamics using separate state-action and next-state maps (Ren et al., 2023). The shared encoder objective uses the same state features on both sides of every action operator. For this objective, Khetarpal et al. (2024) derive a formulation with a separate linear predictor for each action. Under symmetric kernels with a common eigenbasis, they identify an optimal representation by selecting directions with the largest average squared action eigenvalues. However, their analysis leaves unresolved whether the globally optimal value can be approximated efficiently when the kernels lack a common eigenbasis. Centering excludes the constant component and whitening normalizes feature covariance, making the question whether polynomial time suffices for prescribed inverse polynomial additive accuracy. The shared feature space must serve all actions jointly, so individual spectral solutions provide no general guarantee for its global optimization. This gap makes it unclear whether strict positivity and rapid mixing of individual kernels suffice to make the shared objective tractable.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

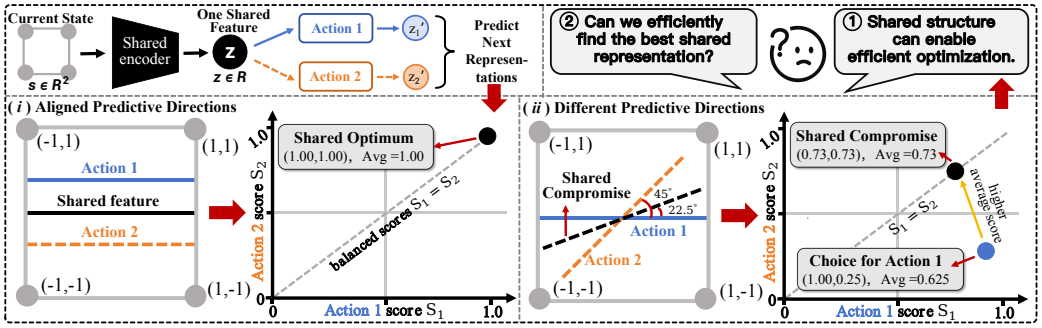

Figure 1: **Joint optimization of a shared predictive feature.** Two action predictors share a centered, whitened scalar feature, with scores $S_a = \cos^4 \theta_a$ normalized by their individual maxima. (i) Aligned directions admit a shared feature maximizing both scores. (ii) For directions $45°$ apart, the $22.5°$ feature scores $(0.73, 0.73)$, beating Action 1's preferred direction (0.625).

In this paper, we develop a graph construction to establish the computational limits of centered and whitened action-conditioned self-prediction. The construction assigns an action to each graph edge and combines a uniform transition rule with a term that encodes the connection between its endpoints. We spread this term across the state space so that all transition probabilities remain positive and each action retains fixed signal strength after centering. Centering removes the uniform background, allowing the shared prediction objective to recover the graph's connection structure. Its optimal value then determines the size of the largest set of mutually connected vertices, through a classical graph optimization identity (Motzkin & Straus, 1965). This correspondence yields worst-case hardness of approximating the optimal value to inverse polynomial additive accuracy for rapidly mixing symmetric kernels. We also identify shared structures that permit exact or approximate solutions and connect the hard instances to finite transition data. Our contributions are as follows.

(1) We construct strictly positive, symmetric transition kernels with rapid mixing and fixed centered operator norm. The hard family has a growing number of linearly independent centered action operators. Approximating its optimal value to inverse polynomial additive accuracy is computationally hard in the worst case. A further construction extends this hardness to every fixed representation dimension.

(2) For symmetric kernels, we give an exact value formula at every feasible representation dimension when they share a common eigenbasis. We also give an additive approximation scheme for a single feature, with exponential cost controlled by the number of linearly independent centered action operators.

(3) We realize the hard instances exactly as finite transition datasets. For the scalar empirical cross-covariance objective, polynomially many independent stationary samples per action preserve the hardness gap with high probability when centering, whitening, and representation selection use the same data.

2 Preliminaries

Markov dynamics. We consider a finite state space $\mathcal{S} = \{1, \ldots, n\}$, with $n \geq 2$, and $m \geq 1$ actions. Each action $a \in \{1, \ldots, m\}$ has a transition kernel $T_a \in \mathbb{R}^{n \times n}$, whose entry $T_a(s, t)$ is the probability of moving from s to t. The kernels are symmetric, have nonnegative entries, and satisfy $T_a \mathbf{1} = \mathbf{1}$, where $\mathbf{1}$ is the vector of ones. They are therefore doubly stochastic and share the uniform stationary distribution $\pi = \mathbf{1}/n$. For each action, expectations use $S \sim \pi$ followed by $S' \sim T_a(S, \cdot)$. Both source and target marginals are uniform. All actions receive equal weight in the objective. A kernel is strictly positive when all its entries are positive. Strict positivity makes the finite chain irreducible and aperiodic. Appendix A further discusses related work.

Centered and whitened prediction. Let $H \in \mathbb{R}^{n \times d}$ be a tabular representation shared by all actions, with row H_s encoding state s. Population centering and whitening require $\mathbb{E}[H_S] = 0$ and $\mathbb{E}[H_S^\top H_S] = I_d$. The rescaling $\Phi = H/\sqrt{n}$ gives the equivalent constraints $\Phi^\top \mathbf{1} = 0$ and $\Phi^\top \Phi = I_d$, which are feasible for $1 \leq d \leq n - 1$. Centering removes the constant direction

preserved by every kernel, and whitening fixes the feature covariance. The resulting features have zero mean and unit variance, with zero covariance between distinct coordinates.

For action a, the feature cross-covariance is $C_a = \mathbb{E}[H_S^\top H_{S'} \mid a] = \Phi^\top T_a \Phi$. We study the largest average squared cross-covariance achievable by one shared representation,

$$J_d(T_1, \dots, T_m) = \max_{\substack{\Phi \in \mathbb{R}^{n \times d} \\ \Phi^\top \mathbf{1} = 0,\ \Phi^\top \Phi = I_d}} \frac{1}{m} \sum_{a=1}^{m} \|\Phi^\top T_a \Phi\|_F^2. \tag{1}$$

Here $\|\cdot\|_F$ denotes the Frobenius norm. Symmetry gives $\|C_a\|_F^2 = \mathrm{Tr}(C_a^2)$, so this objective also has a trace form. The feasible set is compact, and the maximum is attained.

For a single feature, write $J = J_1$ and $\Phi = x$. The constraints reduce to $x^\top \mathbf{1} = 0$ and $\|x\|_2 = 1$, and the objective averages $(x^\top T_a x)^2$ over actions. The unit variance constraint makes each quadratic form the correlation between the current and next values of the same feature. Without centering, the normalized constant feature would attain value one for every action.

To connect this objective to prediction, assign each action a linear predictor $W_a \in \mathbb{R}^{d \times d}$ that maps H_S to $H_S W_a$. For any fixed feasible H, stationarity and whitening make the unique optimal predictor $W_a = C_a$, with

$$\min_{W_1, \dots, W_m} \frac{1}{m} \sum_{a=1}^{m} \mathbb{E}\big[\|H_{S'} - H_S W_a\|_2^2 \mid a\big] = d - \frac{1}{m} \sum_{a=1}^{m} \|C_a\|_F^2.$$

The identity follows by completing the square in each predictor. Stationarity makes the target feature covariance equal to the identity, so its total variance is fixed at d. Minimizing the population prediction loss jointly over the representation and predictors consequently has value $d - J_d$. For a scalar feature, the fitted coefficient can be negative, so both positive and negative correlations contribute to predictability. The full derivation is given in Appendix B.1.

3 Computational Hardness

In this section, we first construct Markov kernels that encode a graph (Section 3.1). Then we derive hardness results for global optimization and approximation (Section 3.2). Finally, we connect the construction to finite transition data (Section 3.3).

3.1 A Markov Embedding with Fixed Signal

Let $G = ([N], E)$ be a simple undirected graph with $N \geq 2$ vertices and $m = |E| \geq 1$ edges. Write $\omega(G)$ for its clique number, the largest number of mutually adjacent vertices. Each edge will index one action. Choose the state count $s = 2^{\lceil \log_2(N+1) \rceil}$, so $N + 1 \leq s < 2(N + 1)$. The Sylvester matrix of order s has orthogonal sign columns, including one constant column. Normalize any N nonconstant columns to obtain $\overline{Q} \in \mathbb{R}^{s \times N}$. Its columns are orthonormal and orthogonal to $\mathbf{1}$, and each coordinate equals $\pm 1/\sqrt{s}$.

For an edge $\{i, j\}$, let $A_{ij} = (e_i e_j^\top + e_j e_i^\top)/2$, where e_i is the ith coordinate vector in $\mathbb{R}^N$. We define the corresponding transition kernel by

$$\overline{T}_{ij} = \frac{\mathbf{1}\mathbf{1}^\top}{s} + \frac{1}{2} \overline{Q} A_{ij} \overline{Q}^\top. \tag{2}$$

The uniform component supplies a common stationary background. The sign columns distribute the edge perturbation across the state space, keeping every entry small enough to preserve positivity. Orthogonality preserves its action on the encoded graph coordinates.

Proposition 1 (Kernel validity and signal strength)**.** *Every kernel in equation 2 is rational, strictly positive, symmetric, and doubly stochastic. Its entries belong to* $\{1/(2s), 1/s, 3/(2s)\}$*. The centered operator* $B_{ij} = \overline{T}_{ij} - \mathbf{1}\mathbf{1}^\top/s$ *has rank two and operator norm* $1/4$*. Every chain is reversible with respect to the uniform distribution and has absolute spectral gap* $3/4$*.*

Proof. Write $\bar{q}_i$ for column i of $\overline{Q}$. Every entry of $\overline{Q}A_{ij}\overline{Q}^\top = (\bar{q}_i\bar{q}_j^\top + \bar{q}_j\bar{q}_i^\top)/2$ belongs to $\{-1/s, 0, 1/s\}$. Scaling by one half and adding the uniform entry gives the asserted positive rational values. The perturbation is symmetric and annihilates $\mathbf{1}$, which establishes the row and column sums. On the range of $\overline{Q}$, the centered operator is isometric to $A_{ij}/2$, with nonzero eigenvalues $\pm 1/4$. It vanishes on the orthogonal complement. The full kernel therefore has eigenvalues 1, $1/4$, $-1/4$, and zeros, proving the spectral claims. This separation lets the graph grow while all kernels remain strictly positive, with the same spectral gap and centered signal strength. □

3.2 Hardness of Approximating the Optimal Value

Theorem 2 (Exact value and approximation hardness)**.** *For the kernels in equation 2, the scalar optimum is*

$$\overline{J} := J((\overline{T}_{ij})_{\{i,j\}\in E}) = \frac{1}{8m}\left(1 - \frac{1}{\omega(G)}\right). \tag{3}$$

At a clique threshold $3 \le K \le N$, consecutive levels are separated by

$$\overline{\Delta}_{G,K} = \frac{1}{8mK(K-1)} > \frac{1}{4N^4}. \tag{4}$$

Under explicit rational matrix encoding, deciding whether $\overline{J} \ge \tau$ for a rational threshold τ is NP-hard. Returning an estimate of $\overline{J}$ with additive error at most $1/(16N^4)$ is also NP-hard.

Proof. For a centered unit feature $x \in \mathbb{R}^s$, write $z = \overline{Q}^\top x$ and $w = x - \overline{Q}z$. Orthogonality gives $\|z\|_2^2 + \|w\|_2^2 = 1$. Centering removes the uniform term in every action, leaving $x^\top \overline{T}_{ij} x = z_i z_j/2$. All predictions therefore depend only on the graph coordinates z, and their squared average is a nonnegative homogeneous polynomial of degree four.

The component w cannot occur in a maximizing feature. To see this, choose any edge and set $z = (e_i + e_j)/\sqrt{2}$, which yields the positive value $1/(16m)$. A maximizer must therefore have $z \neq 0$. If $c = \|z\|_2 < 1$, replacing x by $\overline{Q}z/c$ preserves feasibility and multiplies its positive objective by $c^{-4} > 1$. This contradicts maximality. Every maximizer consequently has $\|z\|_2 = 1$ and $w = 0$. Conversely, every unit graph vector lifts to a centered unit feature through $\overline{Q}$.

Set $y_i = z_i^2$. These squared coordinates range over the entire simplex $\Delta_N = \{y \in \mathbb{R}^N : y \ge 0,\ \mathbf{1}^\top y = 1\}$, since $z_i = \sqrt{y_i}$ supplies a preimage. The global optimization is therefore exactly

$$\overline{J} = \frac{1}{4m} \max_{y\in\Delta_N} \sum_{\{i,j\}\in E} y_i y_j.$$

The identity of Motzkin & Straus (1965) evaluates the simplex maximum as $(1 - 1/\omega(G))/2$, proving equation 3. The argument covers the full feature domain, including directions outside the encoded graph subspace.

For the reduction, CLIQUE asks whether $\omega(G) \ge K$ (Karp, 1972). Define $\tau^- = (1 - 1/(K-1))/(8m)$ and $\tau^+ = (1 - 1/K)/(8m)$. Formula equation 3 places every no instance at or below τ^- and every yes instance at or above τ^+. Subtraction gives equation 4, using $8mK(K-1) \le 4N^2(N-1)^2 < 4N^4$. Exact comparison with τ^+ decides CLIQUE. An estimate with the stated additive error also decides it by comparison with the midpoint of the two levels.

The Sylvester signs are recursively computable, and the entry alphabet in Proposition 1 gives an explicit rational construction. Its ms^2 entries require $O(ms^2 \log s)$ bits in total. The thresholds and requested accuracy have $O(\log N)$ bits. Since $s = O(N)$ and $m \le N(N-1)/2$, the reduction has polynomial size and runtime. Cases outside $m \ge 1$ and $3 \le K \le N$ are directly decidable and can be mapped to fixed yes or no instances. □

The hard family can also be required to lack a common eigenbasis. Adjoining a disjoint path on three vertices preserves every clique decision with $K \ge 3$ and supplies two adjacent edges. Their action kernels have a nonzero commutator, so they cannot share an eigenbasis. The centered edge operators are linearly independent because compression by $\overline{Q}$ recovers edge matrices with disjoint off-diagonal supports. Thus the number of independent action directions grows with the source graph.

Corollary 3 (Every fixed representation dimension)**.** *Fix a representation dimension $d \geq 1$. Exact thresholding of J_d is NP-hard for rational, strictly positive, symmetric stochastic kernels with centered operator norm $1/4$. Approximation at the inverse polynomial additive accuracy specified below is also NP-hard.*

Proof sketch. Use $s_d = 2^{\lceil \log_2(N+d) \rceil}$ states and add $d-1$ orthogonal Walsh columns u_j to the graph columns. Each added column defines an anchor action with centered operator $u_j u_j^\top/4$, repeated $R_d = (d+2)m$ times. A projection argument bounds any gain from allocating extra dimensions to the graph below the resulting loss in the repeated anchor contributions. Hence every maximizing subspace contains all anchor directions, leaving one dimension for the original scalar instance. With $M_d = m + R_d(d-1)$ total actions, the value is

$$\overline{J}_d = \frac{1}{M_d}\left[\frac{R_d(d-1)}{16} + \frac{1}{8}\left(1 - \frac{1}{\omega(G)}\right)\right].$$

Consecutive clique levels have gap $1/[8M_dK(K-1)]$. Approximating the value within one quarter of this gap decides the source instance. For fixed d, the state count, action count, and reciprocal gap are polynomial in the source size. Adding the uniform kernel preserves strict positivity for both the edge and anchor actions. A complete proof is given in Appendix B.3. □

3.3 Finite Transition Data

The scalar construction has an exact realization by observed transitions. For action e, include each ordered pair (u, v) with multiplicity $c_{uv}^{(e)} = 2s(\overline{T}_e)_{uv} \in \{1, 2, 3\}$. These integer counts total $2s^2$, with every row and column summing to $2s$. Both empirical marginals are therefore uniform. For $h = \sqrt{s}\,x$, the empirical average of $h_u h_v$ is exactly $x^\top \overline{T}_e x$. Empirical centering and whitening impose the same feasible set as the population constraints. The resulting dataset reproduces the global value and its gap exactly, and its total size is polynomial. The population predictor elimination identity also holds for these datasets because each action has matching source and target marginals.

For random samples, we analyze a scalar empirical cross-covariance objective with centering and whitening estimated from the data. Draw L stationary transitions $(S_{a\ell}, S'_{a\ell})$ for each action, independently across all actions and samples. Let $\widehat{F}_a = L^{-1}\sum_{\ell=1}^{L} e_{S_{a\ell}} e_{S'_{a\ell}}^\top$ be the empirical joint distribution and $\widehat{\pi} = m^{-1}\sum_a \widehat{F}_a \mathbf{1}$ the pooled source marginal. Here the coordinate vectors index the s states. Write $D_{\widehat{\pi}} = \operatorname{diag}(\widehat{\pi})$. When every entry of $\widehat{\pi}$ is positive, define

$$\widehat{J}_{\mathrm{ew}} = \max_{\substack{h \in \mathbb{R}^s,\ \widehat{\pi}^\top h = 0 \\ h^\top D_{\widehat{\pi}} h = 1}} \frac{1}{m}\sum_{a=1}^{m} (h^\top \widehat{F}_a h)^2. \tag{5}$$

The same transitions determine the feature constraints and the objective over which the encoder is selected.

Theorem 4 (Gap preservation under independent sampling)**.** *For the Walsh family and the empirical objective in equation 5, fix $0 < \delta < 1$ and $3 \leq K \leq N$. There is a universal constant C such that $L \geq CN^9 \log(8mN/\delta)$ independent stationary transitions per action suffice for the following event. With probability at least $1-\delta$, all entries of $\widehat{\pi}$ are positive. On the same event,*

$$\left|\widehat{J}_{\mathrm{ew}} - \overline{J}\right| \leq \frac{1}{16N^4} < \frac{\overline{\Delta}_{G,K}}{4}. \tag{6}$$

The comparison is uniform over the feasible features. Centering, whitening, and encoder selection may all use the same data.

Proof sketch. Whitening maps each feasible domain to a Euclidean unit sphere orthogonal to its corresponding mean direction. Projecting onto that centered subspace extends each objective to the full unit sphere without changing its maximum, by homogeneity. This puts the population and empirical problems on a common domain even though their centering directions and whitening metrics differ. Let B_a and $\widehat{B}_a$ denote the resulting symmetric operators obtained by whitening the flows and applying these projections. The empirical operators use the symmetric part of each whitened flow, which preserves its scalar quadratic form. Appendix B.4 gives the full proof.

Matrix Bernstein and marginal concentration control the flows, whitening metrics, and centering projectors simultaneously (Tropp, 2012). With probability at least $1-\delta$, their combined error $\varepsilon = \max_a \|\widehat{B}_a - B_a\|_2$ is $O(\sqrt{s\log(ms/\delta)/L} + s\log(ms/\delta)/L)$. The population operators have norm $1/4$, so uniformly over the common unit sphere, the squared quadratic predictions differ by at most $\varepsilon(1/2+\varepsilon)$. Averaging over actions and taking maxima preserves this bound. Since $s = O(N)$, the stated sample size makes $\varepsilon = O(N^{-4})$. Choosing the universal constant sufficiently large gives equation 6. The uniform comparison permits the representation to be chosen after observing the samples. The theorem gives a sufficient sample size for preserving the separation between the two clique levels. Thus the computational gap persists for both exact and sampled transition data. □

4 Optimization under Shared Structure

4.1 Exact Spectral Optimization

Suppose the kernels commute pairwise. Symmetry gives a common orthonormal eigenbasis $u_0, \ldots, u_{n-1}$. Choose $u_0 = \mathbf{1}/\sqrt{n}$ and write $T_a u_i = \lambda_{a,i} u_i$. Each nonconstant direction has score $\mu_i = m^{-1}\sum_a \lambda_{a,i}^2$. This shared spectral structure underlies the representation selection rule of Khetarpal et al. (2024). The centered objective admits the following formula.

Theorem 5 (Exact value under a common eigenbasis)**.** *Relabel the nonconstant common directions so that $\mu_1 \geq \cdots \geq \mu_{n-1}$. For every $1 \leq d \leq n-1$, $J_d(T_1, \ldots, T_m) = \sum_{i=1}^{d} \mu_i$. The representation $\Phi = [u_1 \ \cdots \ u_d]$ attains this value. Equivalently, the scores are the eigenvalues of $S = m^{-1}\sum_a T_a^2$ restricted to $\mathbf{1}^\perp$.*

Proof sketch. For any feasible Φ, orthogonal projection gives $\|\Phi^\top T_a \Phi\|_F^2 \leq \|T_a\Phi\|_F^2$. Averaging bounds the objective by $\mathrm{Tr}(\Phi^\top S \Phi)$, whose maximum is $\sum_{i=1}^{d}\mu_i$ by the Ky Fan variational principle. The selected common directions span a subspace invariant under every action. Projection therefore discards no component, so the upper bound is attained. This proves $J_d(T_1, \ldots, T_m) = \sum_{i=1}^{d}\mu_i$. Appendix B.5 gives the full proof. □

The compression bound still holds without commutativity, but equality can fail. For a single feature, it gives $J \leq U_{\mathrm{spec}} = \lambda_{\max}(S|_{\mathbf{1}^\perp})$. On the Walsh construction for the complete graph K_N, this bound is $U_{\mathrm{spec}} = 1/(8N)$, whereas $\overline{J} = 1/(4N^2)$. The ratio $U_{\mathrm{spec}}/\overline{J} = N/2$ shows how far this spectral relaxation can exceed the optimum, even with fixed centered signal.

4.2 Approximation through Effective Action Rank

We restrict the remaining analysis to a single feature. Define $P_0 = I - \mathbf{1}\mathbf{1}^\top/n$ and $B_a = (P_0 T_a P_0)|_{\mathbf{1}^\perp}$. These centered operators are symmetric contractions. Their effective action rank is $r = \dim \mathrm{span}\{B_1, \ldots, B_m\}$. It counts linearly independent operator directions, so repeated action labels do not increase it. Dependence can also occur among distinct, noncommuting actions. Repetition can change the objective's action weights, while leaving this rank unchanged. If $r = 0$, then $J = 0$; below assume $r \geq 1$.

Theorem 6 (Approximation through effective action rank)**.** *Suppose the kernels are rational, with total binary encoding length L. For every $0 < \varepsilon \leq 1$, a deterministic algorithm returns $\widehat{J}$ satisfying $|\widehat{J} - J| \leq \varepsilon$ using*

$$(1 + 1/\varepsilon)^{O(r)}\,\mathrm{poly}(n, m, L, \log(1/\varepsilon)) \tag{7}$$

bit operations.

Proof sketch. For a centered unit feature x, write $q(x) = (x^\top T_1 x, \ldots, x^\top T_m x)$. Euclidean norm duality and the Rayleigh variational principle give

$$\sqrt{mJ} = \max_{\substack{x \perp \mathbf{1} \\ \|x\|_2 = 1}} \|q(x)\|_2 = \max_{c \in \mathbb{S}^{m-1}} \lambda_{\max}\left(\sum_{a=1}^{m} c_a B_a\right).$$

Here $\mathbb{S}^{k-1}$ is the unit sphere in $\mathbb{R}^k$. The two maxima range over compact sets and may be interchanged. Each choice of action coefficients thus reduces the inner optimization to a symmetric eigenvalue problem. The coefficient maximum accounts for both signs of each action prediction before the result is squared.

Let $\mathcal{B}c = \sum_a c_a B_a$, and choose $U \in \mathbb{R}^{m\times r}$ whose orthonormal columns span $(\ker \mathcal{B})^\perp$. For any $c \in \ker \mathcal{B}$, we have $c^\top q(x) = x^\top(\mathcal{B}c)x = 0$, so $q(x)$ is orthogonal to that kernel. Norm duality therefore needs only coefficient vectors $c = Uu$ with $\|u\|_2 = 1$. Setting $C_j = \sum_a U_{aj} B_a$ gives

$$J = \frac{1}{m}\left[\max_{u\in\mathbb{S}^{r-1}} \lambda_{\max}\left(\sum_{j=1}^{r} u_j C_j\right)\right]^2. \tag{8}$$

This compression preserves the original action weights and removes coefficient directions that contribute zero to the combined operator. The orthonormal coordinates retain the Euclidean geometry needed to control approximation error after reducing the dimension.

Write $g(u) = \lambda_{\max}(\sum_j u_j C_j)$ and $F = \max_u g(u)$. The isometry U and the contraction bounds give $|g(u) - g(v)| \le \|\mathcal{B}U(u-v)\|_2 \le \sqrt{m}\|u-v\|_2$. Opposite coefficient directions ensure $F \ge 0$, and contractions give $F \le \sqrt{m}$. Choose a Euclidean $\varepsilon/4$-net $\mathcal{N}$ of the unit sphere, so every unit vector lies within distance $\varepsilon/4$ of a net point. Such a cover exists with at most $(1 + 8/\varepsilon)^r$ points. Evaluate g on the cover and set $F_{\mathcal{N}} = \max\{0, \max_{u\in\mathcal{N}} g(u)\}$.

The covering error satisfies $0 \le F - F_{\mathcal{N}} \le \sqrt{m}\varepsilon/4$. Applying equation 8 converts this spectral error into an objective error,

$$0 \le J - \frac{F_{\mathcal{N}}^2}{m} = \frac{(F - F_{\mathcal{N}})(F + F_{\mathcal{N}})}{m} \le \frac{\varepsilon}{2}.$$

Computing each largest eigenvalue to accuracy $\varepsilon\sqrt{m}/16$ uses less than half the remaining error budget. The remaining allowance accommodates rational approximations to the covering points and the coefficient basis.

For rational inputs, exact elimination on vectorized actions yields a rational basis and coefficient matrix. The coefficient Gram matrix is positive definite. Its square root converts the associated quadratic norm into Euclidean spherical coordinates. Rational determinant bounds place its smallest eigenvalue above $2^{-\operatorname{poly}(L)}$, so polynomially many bits suffice for this change of coordinates. Certified square roots and symmetric eigenvalue computations have polynomial cost at the required precision. A deterministic spherical grid has $(1 + 1/\varepsilon)^{O(r)}$ points, with its construction cost included in this asymptotic bound. Together these give equation 7. Appendix B.6 gives the complete proof. □

Fix r_0 and a polynomial p with $p(L) \ge 1$. On inputs with $r \le r_0$, running Theorem 6 at $\varepsilon = 1/[2p(L)]$ distinguishes $J \le \tau - 1/p(L)$ from $J \ge \tau + 1/p(L)$ in polynomial time. This guarantee includes noncommuting kernels. The Walsh edge operators have $r = m$, which grows with the input graph. Their hardness is consistent with the guarantee for fixed effective action rank. Under our kernel assumptions, exact probability CP rank R gives $r \le R$ (Wu & Li, 2026), and exact global transition factorization dimension k gives $r \le k(k-1)/2$ (Ren et al., 2023; Gao et al., 2026). These bounds place fixed R or k within the tractable regime of Theorem 6 (Appendix C.5). Appendix C provides stability bounds for supplied approximations with a common basis or a smaller action span.

5 Experiments

In this section, we conduct experiments to address the following questions: **RQ1:** Can features learned from finite trajectories distinguish adjacent clique levels (Section 5.2)? **RQ2:** How reliably do generic optimizers attain the certified optimum under a fixed evaluation budget (Section 5.3)? **RQ3:** How does exploiting a common eigenbasis compare with direct nonconvex optimization (Section 5.4)? **RQ4:** How does solution quality change when operators depart from a common eigenbasis (Section 5.5)?

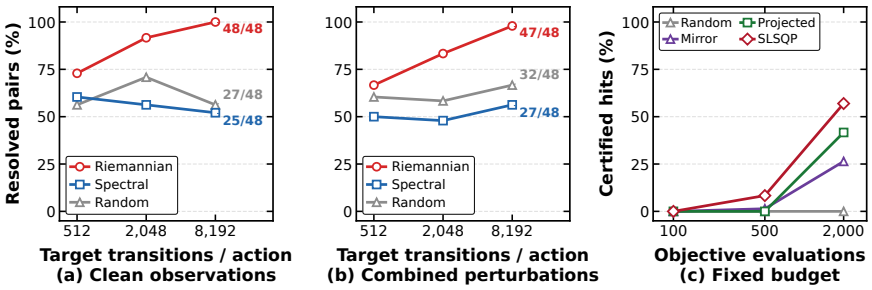

Figure 2: Finite data and optimization budgets. (a,b) Resolution rates on the same 48 graph pairs under clean observations and combined perturbations, using population values of the selected features. (c) Certified hit rates on 72 graphs across objective evaluation budgets, with absolute value error at most 10^{-7}.

5.1 Experimental Setup

Evaluation metrics. For graph pairs, we measure whether the population values of the selected features preserve the ordering of adjacent clique levels, together with the absolute error in their value difference. Exact clique numbers provide reference values through the identity of Motzkin & Straus (1965). An optimization hit has absolute value error at most 10^{-7} for graph objectives and 10^{-8} for exact shared structure. For perturbed operators, we report $U_{\mathrm{spec}} - f(x)$, the distance to the spectral upper bound, which bounds suboptimality.

Datasets. We evaluate scalar representations using Walsh graph instances and families of symmetric operators. The graph studies use 48 pairs at $N = 10, 14$ and 72 sparse or planted graphs at $N = 20, 36, 52$ (Gilbert, 1959). The structural study uses 96 operator families with exact or perturbed common eigenbases, without imposing stochasticity.

Baseline methods. Trajectory learners use Riemannian gradient ascent with multiple random starts (Absil et al., 2008), a spectral direction, or random directions. The spectral rule uses the average squared operator, extending selection under a common eigenbasis (Khetarpal et al., 2024). Budget comparisons include random search, projected ascent with simplex projection (Duchi et al., 2008), mirror ascent (Beck & Teboulle, 2003), and SLSQP (Kraft, 1988) through SciPy (Virtanen et al., 2020). Structural comparisons also use anchor diagonalization, which selects the best eigenvector of the first action under the shared objective. For more detailed descriptions of the experimental settings, please refer to Appendix F.1.

5.2 Learning from Finite Trajectories (RQ1)

We compare graph pairs with matched edge counts and clique numbers $K - 1$ and K, using clean observations or combined state replacement, missingness, and unequal action frequencies. Figure 2(a,b) shows that Riemannian feature learning resolves more pairs as the trajectory budget increases. At the largest target budget of 8,192 transitions per action, it resolves all 48 clean pairs and 47 perturbed pairs. Both baselines resolve at most two thirds of the pairs in either condition.

The ordering remains recoverable even as estimates of its magnitude deteriorate. At the same budget, the median relative gap error rises from 0.160 under clean observations to 0.436 under combined perturbations (Appendix F.2). These results support recovery of adjacent clique ordering from finite trajectories on the tested instances, while the recovered gap magnitude remains sensitive to observation and sampling perturbations.

5.3 Optimization under a Fixed Budget (RQ2)

We optimize the equivalent simplex objective to examine attainment of the certified value. Figure 2(c) reports the fraction attaining the certified value within 10^{-7}. The three optimization methods improve their hit rates with larger budgets, whereas random search records no hits. At 2,000 evaluations, SLSQP has the highest hit rate, 56.94%, followed by projected ascent at 41.67%. Even the best observed rate leaves many instances below the prescribed accuracy.

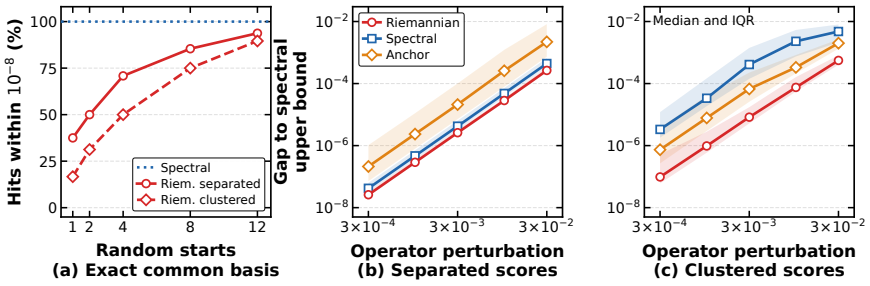

Figure 3: Exact and perturbed shared structure. (a) Riemannian hit rates using the best terminal value from the first k random starts, with tolerance 10^{-8}. (b,c) Median gaps to the spectral upper bound. Shaded bands indicate interquartile ranges.

At the same budget, SLSQP has a lower median error than projected ascent, 2.41×10^{-6} compared with 5.65×10^{-4} (Appendix F.3). Projected ascent has an interquartile range of 3.05×10^{-2}, approximately 8.4 times smaller than SLSQP's 2.56×10^{-1}. SLSQP thus more frequently attains the strict target and achieves a lower median error, while projected ascent concentrates the middle half of its errors within a smaller interval near zero.

5.4 Optimization with an Exact Common Basis (RQ3)

We compare separated and clustered spectral scores, with differences of 0.008 and 0.0002 between the two largest average squared eigenvalue scores. Figure 3(a) compares this result with Riemannian search. Spectral optimization attains the reference value to machine precision in both regimes. Riemannian search improves from a 27.08% hit rate with one initialization to 91.67% with twelve. This comparison illustrates the benefit of Theorem 5 for optimization. Exploiting the common basis gives direct access to the global value, while the tested local search remains sensitive to initialization even when that spectral solution exists.

5.5 Optimization under Structural Perturbations (RQ4)

We perturb each base family by symmetric operators of prescribed norm. Figure 3(b,c) separates the two score regimes and compares the achieved values with the spectral relaxation upper bound. Symmetric perturbations increase the median distances to the spectral upper bound, with the spectral direction showing greater sensitivity in the clustered regime. At perturbation 0.003, the spectral direction has median distance 4.23×10^{-6} in the separated regime and 4.11×10^{-4} in the clustered regime. Riemannian search reduces these distances to 2.61×10^{-6} and 8.26×10^{-6}, respectively. Joint optimization therefore yields a substantially larger reduction in the reported median distance when the leading scores are close.

The larger separation between methods in the clustered regime shows where direct optimization becomes particularly useful after this structure is perturbed. Riemannian search also exceeds anchor diagonalization on 91.67% of the perturbed instances, supporting the benefit of searching jointly across actions.

6 Conclusion

For centered and whitened action-conditioned self-prediction in finite state spaces, favorable properties of individual kernels are insufficient to guarantee efficient global optimization. We prove NP-hardness at prescribed inverse polynomial additive accuracy for strictly positive, symmetric kernels with rapid mixing and fixed centered signal strength. The hard family has growing effective action rank, and a further construction extends the hardness to every fixed representation dimension. For symmetric kernels, a common eigenbasis yields exact solutions at every feasible dimension. Fixed effective action rank permits efficient approximation at inverse polynomial additive accuracy for a single feature, including noncommuting families. Our experiments illustrate the benefit of exploiting an exact common basis and the sensitivity of numerical optimization to initialization and structural perturbations.

AI USE STATEMENT

In this work, we used generative AI tools for aiding and polishing the writing. We take responsibility for the final content of this work.

REFERENCES

P-A Absil, Robert Mahony, and Rodolphe Sepulchre. *Optimization algorithms on matrix manifolds*. Princeton University Press, 2008.

Alekh Agarwal, Sham Kakade, Akshay Krishnamurthy, and Wen Sun. Flambe: Structural complexity and representation learning of low rank mdps. *Advances in neural information processing systems*, 33:20095–20107, 2020.

Zhaojun Bai and Ding Lu. Variational characterization of monotone nonlinear eigenvector problems and geometry of self-consistent field iteration. *SIAM Journal on Matrix Analysis and Applications*, 45(1):84–111, 2024.

Amir Beck and Marc Teboulle. Mirror descent and nonlinear projected subgradient methods for convex optimization. *Operations Research Letters*, 31(3):167–175, 2003.

John Duchi, Shai Shalev-Shwartz, Yoram Singer, and Tushar Chandra. Efficient projections onto the l 1-ball for learning in high dimensions. In *Proceedings of the 25th international conference on Machine learning*, pp. 272–279, 2008.

Scott Fujimoto, Pierluca D'Oro, Amy Zhang, Yuandong Tian, and Michael Rabbat. Towards general-purpose model-free reinforcement learning. In *The Thirteenth International Conference on Learning Representations*, 2025.

Chenxiao Gao, Haotian Sun, Na Li, Dale Schuurmans, and Bo Dai. Spectral representation-based reinforcement learning, 2026. URL https://arxiv.org/abs/2512.15036.

Carles Gelada, Saurabh Kumar, Jacob Buckman, Ofir Nachum, and Marc G Bellemare. Deepmdp: Learning continuous latent space models for representation learning. In *International conference on machine learning*, pp. 2170–2179. PMLR, 2019.

Edgar N Gilbert. Random graphs. *The Annals of Mathematical Statistics*, 30(4):1141–1144, 1959.

Diego Gomez, Michael Bowling, and Marlos C Machado. Proper laplacian representation learning. In *International Conference on Learning Representations*, volume 2024, pp. 28911–28934, 2024.

Jean-Bastien Grill, Florian Strub, Florent Altché, Corentin Tallec, Pierre Richemond, Elena Buchatskaya, Carl Doersch, Bernardo Avila Pires, Zhaohan Guo, Mohammad Gheshlaghi Azar, et al. Bootstrap your own latent-a new approach to self-supervised learning. *Advances in neural information processing systems*, 33:21271–21284, 2020.

Christopher J Hillar and Lek-Heng Lim. Most tensor problems are np-hard. *Journal of the ACM (JACM)*, 60(6):1–39, 2013.

Richard M Karp. Reducibility among combinatorial problems. In *Complexity of Computer Computations: Proceedings of a symposium on the Complexity of Computer Computations, held March 20–22, 1972, at the IBM Thomas J. Watson Research Center, Yorktown Heights, New York, and sponsored by the Office of Naval Research, Mathematics Program, IBM World Trade Corporation, and the IBM Research Mathematical Sciences Department*, pp. 85–103. Springer, 1972.

Khimya Khetarpal, Zhaohan Daniel Guo, Bernardo Ávila Pires, Yunhao Tang, Clare Lyle, Mark Rowland, Nicolas Heess, Diana Borsa, Arthur Guez, and Will Dabney. A unifying framework for action-conditional self-predictive reinforcement learning. *CoRR*, abs/2406.02035, 2024.

Tamara G Kolda and Brett W Bader. Tensor decompositions and applications. *SIAM review*, 51(3): 455–500, 2009.

Dieter Kraft. A software package for sequential quadratic programming. *Forschungsbericht-Deutsche Forschungs- und Versuchsanstalt fur Luft- und Raumfahrt*, 1988.

Chen Ling, Jiawang Nie, Liqun Qi, and Yinyu Ye. Biquadratic optimization over unit spheres and semidefinite programming relaxations. *SIAM Journal on Optimization*, 20(3):1286–1310, 2010.

Xinwei Liu, Junyuan Liang, Jianting Zhang, and Wuhui Chen. Observation-grounded self-predictive reinforcement learning for visual continuous control. *arXiv preprint arXiv:2608.05989*, 2026.

Sridhar Mahadevan and Mauro Maggioni. Proto-value functions: A laplacian framework for learning representation and control in markov decision processes. *Journal of Machine Learning Research*, 8(10), 2007.

Theodore S Motzkin and Ernst G Straus. Maxima for graphs and a new proof of a theorem of turán. *Canadian Journal of Mathematics*, 17:533–540, 1965.

Chengzhuo Ni, Yaqi Duan, Munther Dahleh, Mengdi Wang, and Anru R Zhang. Learning good state and action representations for markov decision process via tensor decomposition. *Journal of Machine Learning Research*, 24(115):1–53, 2023.

Tianwei Ni, Benjamin Eysenbach, Erfan Seyedsalehi, Michel Ma, Clement Gehring, Aditya Mahajan, and Pierre-Luc Bacon. Bridging state and history representations: Understanding self-predictive rl. In *International Conference on Learning Representations*, volume 2024, pp. 23555–23569, 2024.

Corentin Pla, Hugo Richard, Marc Abeille, Nadav Merlis, and Vianney Perchet. On the hardness of reinforcement learning with transition lookahead. In *The 29th International Conference on Artificial Intelligence and Statistics*, 2026.

Tongzheng Ren, Tianjun Zhang, Lisa Lee, Joseph E. Gonzalez, Dale Schuurmans, and Bo Dai. Spectral decomposition representation for reinforcement learning. In *The Eleventh International Conference on Learning Representations*, 2023.

Max Schwarzer, Ankesh Anand, Rishab Goel, R Devon Hjelm, Aaron Courville, and Philip Bachman. Data-efficient reinforcement learning with self-predictive representations. In *International Conference on Learning Representations*, 2021.

Dmitry Shribak, Chen-Xiao Gao, Yitong Li, Chenjun Xiao, and Bo Dai. Diffusion spectral representation for reinforcement learning. *Advances in Neural Information Processing Systems*, 37: 110028–110056, 2024.

Yunhao Tang, Zhaohan Daniel Guo, Pierre Harvey Richemond, Bernardo Avila Pires, Yash Chandak, Rémi Munos, Mark Rowland, Mohammad Gheshlaghi Azar, Charline Le Lan, Clare Lyle, et al. Understanding self-predictive learning for reinforcement learning. In *International conference on machine learning*, pp. 33632–33656. PMLR, 2023.

Joel A Tropp. User-friendly tail bounds for sums of random matrices. *Foundations of computational mathematics*, 12(4):389–434, 2012.

Masatoshi Uehara, Xuezhou Zhang, and Wen Sun. Representation learning for online and offline rl in low-rank mdps. *CoRR*, abs/2110.04652, 2021.

Pauli Virtanen, Ralf Gommers, Travis E Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, et al. Scipy 1.0: fundamental algorithms for scientific computing in python. *Nature methods*, 17(3):261–272, 2020.

Junyi Wu and Dan Li. Factorized spectral representations for reinforcement learning. *ArXiv*, abs/2607.13498, 2026.

Yifan Wu, George Tucker, and Ofir Nachum. The laplacian in rl: Learning representations with efficient approximations. *CoRR*, abs/1810.04586, 2018.

David Zuckerman. Linear degree extractors and the inapproximability of max clique and chromatic number. In *Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*, pp. 681–690, 2006.

A Extended related work

Self-predictive representations. Predicting one representation from another gives a learning signal whose targets depend on the representation itself. BYOL uses an online network to predict a moving target network for image representation learning (Grill et al., 2020). SPR applies predictions of future latent states to data efficient reinforcement learning (Schwarzer et al., 2021). Analyses of self-predictive learning connect predictor and encoder dynamics to spectral structure and representation collapse (Tang et al., 2023), while Ni et al. (2024) relate state and history representations through predictive abstractions. These works study representation properties and learning dynamics. Our lower bound concerns the global value after centering and whitening have fixed the feasible representations, so guarantees for a particular update rule address a separate question.

Predictive objectives in control. DeepMDP learns latent models through reward and transition prediction (Gelada et al., 2019). MR.Q uses predictive representation learning in an agent evaluated across diverse control benchmarks (Fujimoto et al., 2025). OG-SPR combines latent prediction with observation and reward prediction for visual continuous control (Liu et al., 2026). These systems use objectives and model classes that extend beyond the tabular shared encoder objective in equation 1. Our theorem does not evaluate their training procedures or control performance. Its structural results identify sufficient conditions on explicit action operators. Applying them to learned nonlinear representations requires corresponding operator structure and approximation guarantees.

The shared spectral objective. The closest objective is developed by Khetarpal et al. (2024) for action-conditional self-prediction with separate linear predictors. Their Lemma 4 identifies the average trace objective under its predictor and dynamics assumptions, without requiring a common eigenbasis. Their Assumption 6 introduces shared eigenvectors, and Theorem 2 uses this structure to select a maximizing representation through average squared action eigenvalues. Our commuting result gives the corresponding centered value at every feasible representation dimension. The complexity result addresses the objective without that shared spectral assumption. Effective action rank supplies a different sufficient condition for scalar approximation, including noncommuting families. These results distinguish an optimal representation under shared spectral structure from global optimization over a larger input class.

Laplacian representations. Proto-value functions use spectral properties of a diffusion operator to construct basis functions for control (Mahadevan & Maggioni, 2007). Later work develops scalable Laplacian representation objectives (Wu et al., 2018) and methods for recovering ordered eigenvectors and their eigenvalues (Gomez et al., 2024). These constructions concern the spectrum of a specified graph or diffusion operator. Equation equation 1 couples several action operators through one feature space. Efficiently constructing spectral features for an individual operator therefore answers a different optimization question from maximizing this shared objective.

Spectral transition representations. SPEDER factors the policy independent transition law into state-action and next-state representations (Ren et al., 2023). It explicitly retains action dependence and uses separate maps for the two sides of the transition. Diff-SR extracts spectral representations through diffusion and energy models, with extensions to partial observability (Shribak et al., 2024). The framework of Gao et al. (2026) includes linear factorizations, latent variable models, and energy models. These constructions have their own representation objectives and learning assumptions. A finite linear factorization can restrict effective action rank under our symmetry assumptions, as shown in Appendix C.5. General latent or energy representations need not have a finite probability operator rank or a common eigenbasis. Their construction and inference guarantees consequently do not establish global tractability of equation 1.

Representation learning in low rank MDPs. FLAMBE gives sample efficient exploration and representation learning with maximum likelihood estimation and sampling oracles (Agarwal et al., 2020). Uehara et al. (2021) develop online and offline guarantees using learned representations and maximum likelihood estimation. Such results expose the statistical role of a low dimensional transition factorization while assigning representation estimation to an oracle. Our input consists of explicit rational transition matrices, and our computational target is the optimal shared prediction value. Fixed factorization dimension yields a conditional connection to our scalar algorithm. A

polynomial sample bound in a variable dimension does not supply a polynomial algorithm for this global objective.

Tensor representations and action rank. Ni et al. (2023) study representations of states and actions through a low Tucker rank transition tensor. FaStR assumes a CP factorization of transition probabilities and learns factorized spectral encoders (Wu & Li, 2026). CP rank and the ranks of tensor unfoldings are distinct quantities (Kolda & Bader, 2009). Our effective action rank is exactly the rank of the centered tensor unfolded along the action mode. A fixed probability CP rank therefore bounds this parameter and excludes the growing Walsh family, even though each individual Walsh operator has low matrix rank. Appendix C.5 derives this relationship and explains why factorization of a contrastive energy score alone is insufficient.

Quartic and spectral optimization. On the centered subspace, the scalar objective is a sum of squared quadratic forms on a sphere. Bai & Lu (2024) study this expression through nonlinear eigenvector problems, relate it to the joint numerical radius, and derive it from approximation of a tensor with symmetric matrix slices. This supplies a direct connection to our action sphere representation. General complexity results cover tensor optimization (Hillar & Lim, 2013) and biquadratic optimization over two unit spheres (Ling et al., 2010). Their coefficient models allow arrays without the positivity and normalization constraints of transition kernels. Our reduction realizes the graph polynomial inside rational stochastic matrices with fixed centered signal. The approximation algorithm then uses the action span dimension while preserving the original action weights.

Graph reductions and approximation precision. The simplex identity of Motzkin & Straus (1965), together with CLIQUE hardness (Karp, 1972), supplies the combinatorial source of our reduction. The Markov embedding determines the scaling that transfers consecutive clique levels to an additive gap. The raw optimum, its normalization, and its distance from saturation have different error criteria. Zuckerman (2006) establish polynomial factor inapproximability for maximum clique. Appendix D.1 transfers this result to the reciprocal clique quantity represented by the saturation deficit. The main theorem specifies inverse polynomial additive accuracy for the raw prediction objective. This also fixes the scale for finite data and perturbation comparisons.

Planning with transition look-ahead. Pla et al. (2026) prove hardness of planning with at least two steps of transition look-ahead, while one step admits polynomial time planning. Their reduction also encodes combinatorial structure through transitions. The computational targets differ. Their problem selects a policy using future transition information. Ours optimizes a shared representation for known action kernels. That planning result provides related complexity context, and the CLIQUE reduction above establishes the hardness of our representation objective.

B Proofs for the Main Results

We first derive the prediction objective (Appendix B.1) and complete the reduction and dimension extension (Appendices B.2 and B.3). We then prove the empirical comparison (Appendix B.4) and the two structural optimization results (Appendices B.5 and B.6). All references to main results use the numbering of the manuscript.

B.1 Predictor Elimination

For a feasible representation $H = \sqrt{n}\,\Phi$, both feature marginals have covariance I_d. Let $C_a = \mathbb{E}[H_S^\top H_{S'} \mid a]$. Expanding the squared prediction error gives the identity

$$\mathbb{E}[\|H_{S'} - H_S W_a\|_2^2 \mid a] = d + \|W_a\|_F^2 - 2\,\mathrm{Tr}(W_a^\top C_a) = d - \|C_a\|_F^2 + \|W_a - C_a\|_F^2.$$

The last term is nonnegative and vanishes uniquely at $W_a = C_a$. Averaging and optimizing over Φ proves that the joint minimum prediction loss is $d - J_d$. Symmetry is needed for the trace form $\|C_a\|_F^2 = \mathrm{Tr}(C_a^2)$, but completing the square only requires the two covariance identities.

The same calculation applies to empirical expectations whenever each action has the same source and target marginal used to center and whiten H. The balanced counts in Section 3.3 satisfy this condition exactly. Independently sampled transitions generally have different empirical marginals

across actions. Appendix B.4 gives the corresponding empirical comparison and records the distinction between squared cross-covariance and prediction loss after fitting the predictors.

B.2 Details of the Markov Reduction

The graph identity. Write $F_G(y) = \sum_{\{i,j\}\in E} y_i y_j$ for $y \in \Delta_N$. Among its maximizers, choose one with minimal support. If two supported vertices i, j are nonadjacent, moving mass between them while fixing their sum changes F_G affinely. An endpoint of this interval has at least the same value and smaller support, a contradiction. The support therefore forms a clique of size $k \le \omega(G)$. On that support, $F_G(y) = (1 - \sum_i y_i^2)/2 \le (1 - 1/k)/2 \le (1 - 1/\omega(G))/2$. Uniform mass on a maximum clique attains the last bound. This proves the identity used in Theorem 2 (Motzkin & Straus, 1965).

Encoding and boundary instances. The Sylvester recursion starts from the matrix (1) and replaces a sign matrix H by $\left(\begin{smallmatrix} H & H \\ H & -H \end{smallmatrix}\right)$. Although normalized columns contain $1/\sqrt{s}$, their products have rational entries with denominator s. Each kernel entry thus needs $O(\log s)$ bits, giving total matrix encoding length $O(ms^2 \log s)$. The thresholds in the main reduction also have logarithmic bit length. Empty graphs, thresholds at most two, and thresholds above N can be decided directly. They can then be sent to fixed yes or no instances, such as K_3 or a path on three vertices at threshold three. The reduction consequently covers general CLIQUE inputs with polynomial preprocessing.

Noncommutativity and action rank. Adjoining a disjoint path on three vertices preserves clique decisions with $K \ge 3$. For its adjacent edges $\{i, j\}$ and $\{j, k\}$, direct multiplication gives $[A_{ij}, A_{jk}] = (e_i e_k^\top - e_k e_i^\top)/4$. The Walsh kernels therefore have commutator $\overline{Q}(e_i e_k^\top - e_k e_i^\top)\overline{Q}^\top/16 \ne 0$. Real symmetric matrices with a common eigenbasis commute, so this pair excludes such a basis. Separately, compression by $\overline{Q}$ sends each centered kernel to $A_{ij}/2$. Their distinct off-diagonal supports prove linear independence and give effective action rank m.

B.3 Hardness at Every Fixed Representation Dimension

Proof of Corollary 3. The case $d = 1$ is Theorem 2. For $d \ge 2$, set $s_d = 2^{\lceil \log_2(N+d) \rceil}$ and choose $N + d - 1$ nonconstant normalized Walsh columns. Use the first N as graph columns and denote the remaining columns by $u_1, \dots, u_{d-1}$. Graph actions have centered operators $B_e = \overline{Q} A_e \overline{Q}^\top/2$. Each anchor action has centered operator $u_j u_j^\top/4$ and appears $R = (d + 2)m$ times. Adding $\mathbf{1}\mathbf{1}^\top/s_d$ gives positive rational stochastic kernels with centered norm $1/4$. There are $M = m + R(d - 1)$ actions, counted with repetition.

Let P be the orthogonal projection onto any feasible representation subspace, and let U be the anchor span. Put $p_j = u_j^\top P u_j$ and $t = d - 1 - \sum_j p_j$. Since $0 \le p_j \le 1$, the anchors contribute at most $R(d - 1 - t)/16$ to the unnormalized objective. Let $P_\perp$ project onto $U^\perp \cap \mathbf{1}^\perp$ and set $K = P_\perp P P_\perp$. This positive contraction has trace $1 + t$. Dimension counting gives a unit vector $v \in \mathrm{range}(P) \cap U^\perp$, so $Kv = v$. We can write $K = vv^\top + K'$ with $K' \succeq 0$, $K'v = 0$, and $\mathrm{Tr}\, K' = t$.

Every graph operator is supported on $U^\perp$, so its compressed contribution is $\mathrm{Tr}(B_e K B_e K)$. Expanding K bounds this by $(v^\top B_e v)^2 + 2\|B_e\|_2^2 t + \|B_e\|_2^2 t^2$. For the last term, diagonalize K' and bound each matrix entry of B_e by $\|B_e\|_2$. The scalar graph maximum is $H_* = (1 - 1/\omega(G))/8$. Using $\|B_e\|_2 = 1/4$ and $0 \le t \le d - 1$ yields the decisive comparison

$$\sum_{a=1}^{M} \|\Phi^\top B_a \Phi\|_F^2 \le \frac{R(d-1)}{16} + H_* - \frac{R - (d+1)m}{16}\, t.$$

The coefficient of t is strictly negative because $R - (d + 1)m = m > 0$.

Taking all anchor directions together with a scalar graph maximizer attains the right side at $t = 0$. Every maximizing subspace must therefore have $p_j = 1$ for every anchor, which forces it to contain U. Dividing by M gives the value stated in the main proof sketch. Adjacent clique levels have gap $1/[8MK(K - 1)]$, and error at most one quarter of this gap permits midpoint thresholding. For fixed d, all matrix counts, bit lengths, and reciprocal gaps are polynomial in the graph size. This proves both complexity assertions. □

The dimension extension also admits an exact transition dataset. For each action use counts $c_{uv} = 4s_d(T_a)_{uv}$, totaling $4s_d^2$ transitions. Graph counts lie in $\{2, 4, 6\}$ and anchor counts in $\{3, 4, 5\}$. Every row and column sums to $4s_d$, giving uniform empirical marginals. The resulting empirical objective agrees with J_d for every feasible representation. Repeated anchor actions remain separate equally weighted action labels, as required by the value formula.

B.4 Uniform Approximation with Empirical Whitening

We prove a scalar comparison for kernels sharing any strictly positive stationary distribution π. Symmetry is unnecessary in this subsection. Write $D = \mathrm{diag}(\pi)$, $F_a = DT_a$, $u = \sqrt{\pi}$, and $P = I - uu^\top$. The discriminant $S_a = D^{-1/2}F_aD^{-1/2}$ is a contraction. Indeed, Jensen's inequality gives $\|T_af\|_{L^2(\pi)} \le \|f\|_{L^2(\pi)}$ by stationarity. Set $B_a = P(S_a + S_a^\top)P/2$ and $\beta = \max_a \|B_a\|_2 \le 1$. Denote the population optimum under $\pi^\top h = 0$ and $h^\top Dh = 1$ by J_π.

Proposition 7 (Uniform empirical comparison)**.** *Draw L stationary transitions per action, independently across actions and samples. Define $\widehat{F}_a$ and $\widehat{\pi}$ as in equation 5, and let $\pi_{\min} = \min_i \pi_i$. For $0 < \delta < 1$, put $\ell = \log(4mn/\delta)$, $\ell_0 = \log(4n/\delta)$, $\xi = \sqrt{2\ell/(L\pi_{\min})} + 4\ell/(3L\pi_{\min})$, and $\eta = \sqrt{3\ell_0/(mL\pi_{\min})}$. If $\eta \le 1/2$, then with probability at least $1 - \delta$ the empirical marginal is positive and*

$$|\widehat{J}_{\mathrm{ew}} - J_\pi| \le e(2\beta + e), \qquad e = 2\xi + 5\eta. \tag{9}$$

The event controls the objective uniformly over features.

Proof. For one sampled pair define $X = D^{-1/2}e_Se_{S'}^\top D^{-1/2}$, so $\mathbb{E}X = S_a$. We have $\|X\|_2 \le 1/\pi_{\min}$ and $\|X - S_a\|_2 \le 2/\pi_{\min}$. The second moments satisfy $\mathbb{E}[XX^\top] \preceq I/\pi_{\min}$ and $\mathbb{E}[X^\top X] \preceq I/\pi_{\min}$. The second bound uses the target marginal π. Centering only decreases these second moments. Applying rectangular matrix Bernstein through symmetric dilation (Tropp, 2012), followed by a union bound over actions, gives $\max_a \|D^{-1/2}(\widehat{F}_a - F_a)D^{-1/2}\|_2 \le \xi$ with failure probability at most $\delta/2$.

The mL source states are independent with distribution π. A multiplicative Chernoff bound and a union bound over states give $\max_i |\widehat{\pi}_i/\pi_i - 1| \le \eta$ with failure probability at most $\delta/2$. Work on the intersection of these two events. Let $R = D^{-1/2}D_{\widehat{\pi}}D^{-1/2}$ and $A = R^{-1/2}$. Then $\|A\|_2 \le \sqrt{2}$ and $\|A - I\|_2 \le \eta$. The empirically whitened flow is $\widehat{S}_a = A(D^{-1/2}\widehat{F}_aD^{-1/2})A$, so $\|\widehat{S}_a - S_a\|_2 \le 2\xi + 3\eta$. Here the whitening contribution follows by expanding $AS_aA - S_a = (A - I)S_aA + S_a(A - I)$.

Let $\widehat{u} = \sqrt{\widehat{\pi}}$ and $\widehat{P} = I - \widehat{u}\widehat{u}^\top$. The relative marginal bound gives $\|\widehat{u} - u\|_2 \le \eta$ and $\|\widehat{P} - P\|_2 \le \eta$. Symmetrization cannot increase operator norm. Defining $\widehat{B}_a = \widehat{P}(\widehat{S}_a + \widehat{S}_a^\top)\widehat{P}/2$ therefore gives $\max_a \|\widehat{B}_a - B_a\|_2 \le 2\xi + 5\eta = e$. Expanding the two projectors accounts for the additional 2η, using the contraction bound for the population symmetric part.

Population whitening sends h to $x = D^{1/2}h$, a unit vector orthogonal to u. The projected quartic objective has the same maximum on the full unit sphere. For any unit x with $Px \ne 0$, normalizing Px multiplies its objective by $\|Px\|_2^{-4} \ge 1$. A vector with $Px = 0$ has value zero. The empirical problem has the same property with $\widehat{P}$. On this common full sphere, each pair of squared quadratic forms differs by at most $e(2\beta + e)$. Averaging and maximizing proves equation 9. The event is independent of which feasible feature is subsequently selected. □

For the Walsh kernels, $n = s = O(N)$, $\pi_{\min} = 1/s$, and $\beta = 1/4$. The preceding parameters give $e = O(\sqrt{s\log(ms/\delta)/L} + s\log(ms/\delta)/L)$. Taking $L \ge CN^9\log(8mN/\delta)$ makes $e \le cN^{-4}$, with any required universal $c > 0$ after increasing C. Equation equation 9 then gives error at most $1/(16N^4)$, and the main gap bound makes this strictly smaller than one quarter of the gap. The same choice ensures $\eta \le 1/2$. This completes the proof of Theorem 4.

For clarity, the empirical prediction loss has a different algebraic form under arbitrary samples. For a fixed scalar feature define $G_a = L^{-1}\sum_\ell h_{S_{a\ell}}^2$, $Q_a = L^{-1}\sum_\ell h_{S'_{a\ell}}^2$, and $C_a = L^{-1}\sum_\ell h_{S_{a\ell}}h_{S'_{a\ell}}$. When $G_a > 0$, minimizing over the scalar predictor gives $Q_a - C_a^2/G_a$. If $G_a = 0$, then $C_a = 0$ and the minimum is Q_a. Pooled whitening enforces the average source variance, but does not set

each G_a and Q_a to one. The sampled theorem concerns equation 5. The exact balanced datasets also satisfy the stronger empirical loss identity.

B.5 Tightness of the Spectral Relaxation

Proof of Theorem 5. For any feasible Φ, let $P_\Phi = \Phi\Phi^\top$ and $S = m^{-1}\sum_a T_a^2$. Orthogonal decomposition gives

$$\|T_a\Phi\|_F^2 - \|\Phi^\top T_a\Phi\|_F^2 = \|(I - P_\Phi)T_a\Phi\|_F^2 \geq 0.$$

After averaging, the objective is at most $\mathrm{Tr}(\Phi^\top S\Phi)$. The Ky Fan principle bounds this trace by the sum of the largest d eigenvalues of S on $\mathbf{1}^\perp$. Under a common eigenbasis, these eigenvalues are the scores μ_i. Selecting the corresponding common directions makes their span invariant under every T_a, so every discarded component above vanishes. The bound is attained and proves $J_d(T_1, \ldots, T_m) = \sum_{i=1}^d \mu_i$.. □

Selection must respect the common eigenvectors when scores coincide. An arbitrary vector in a repeated eigenspace of S can mix directions having different action eigenvalues, so the discarded components may be nonzero. For a concrete example, take orthogonal nonconstant normalized Walsh columns u, v and $T = \mathbf{1}\mathbf{1}^\top/n + \alpha(uu^\top - vv^\top)$, with $0 < \alpha < 1/2$. This is a positive symmetric stochastic kernel. Both u, v have squared score α^2, but $x = (u+v)/\sqrt{2}$ satisfies $x^\top Tx = 0$. The common directions still attain the correct value α^2.

For the complete graph, $\sum_{i<j} A_{ij}^2 = (N-1)I_N/4$ and $m = N(N-1)/2$. Squaring the Walsh centered operators gives $m^{-1}\sum_e B_e^2 = \overline{QQ}^\top/(8N)$ on the centered subspace. Its largest eigenvalue is $1/(8N)$. Theorem 2 gives $\overline{J} = 1/(4N^2)$, establishing the relaxation ratio $N/2$ in Section 4.1.

B.6 Bit Complexity of the Approximation Scheme

Proof of Theorem 6. We may regard each $B_a = P_0T_aP_0$ as a rational matrix on the full state space, with eigenvalue computations restricted to $\mathbf{1}^\perp$. For a centered unit x, write $q_a(x) = x^\top B_a x$. Norm duality gives $\sqrt{mJ} = \max_{\|c\|_2=1}\max_x \sum_a c_a q_a(x)$. The inner maximum is the largest restricted eigenvalue of $\sum_a c_a B_a$. If $\sum_a c_a B_a = 0$, then $c^\top q(x) = 0$ for every x. Removing this coefficient kernel proves the action sphere identity equation 8. Its maximum $F = \sqrt{mJ}$ belongs to $[0, \sqrt{m}]$, and its objective g is $\sqrt{m}$-Lipschitz by the contraction bounds.

To implement the coefficient reduction, select linearly independent input matrices $D_1, \ldots, D_r$ by rational elimination, and solve $B_a = \sum_j \Lambda_{aj} D_j$. The matrix $G = \Lambda^\top\Lambda$ is positive definite. With $U = \Lambda G^{-1/2}$, its columns form the required coefficient isometry. The action sphere matrix has the equivalent form

$$M(u) = \sum_{j=1}^r (G^{1/2}u)_j D_j, \qquad \|u\|_2 = 1. \tag{10}$$

Indeed, $\Lambda^\top U = G^{1/2}$. This representation keeps the original action multiplicities inside G. Each D_j is an input contraction, so coefficient errors v produce operator error at most $\sqrt{r}\,\|v\|_2$.

Here is an explicit grid whose generation fits the claimed runtime. Fix $\theta = \varepsilon/16$, let $k = \lceil\sqrt{r}\rceil$, and set the rational mesh $h = \theta/(4k)$. Enumerate all nonzero $z \in h\mathbb{Z}^r$ with $\|z\|_2 \leq 1 + \theta/4$, and use the directions $z/\|z\|_2$. Rounding any unit vector to this lattice changes it by at most $h\sqrt{r}/2 \leq \theta/8$. Normalization changes the distance by at most another factor two, so the directions form a θ-net. Disjoint cubes of side h around the enumerated points fit inside a ball of radius $1 + \theta/2$. The bound $\mathrm{vol}(B_2^r) \leq (C_0/\sqrt{r})^r$ then gives at most $(C_1/\theta)^r$ points. Recursive enumeration retains only prefixes whose squared norm satisfies the radius bound. Each retained prefix extends by zeros to a lattice point, so there are at most r times as many prefixes as full points. This also bounds generation cost.

All rational coefficients obtained by elimination have polynomial bit length. Clearing denominators and applying determinant bounds gives $2^{-p(n,m,L)}I \preceq G \preceq 2^{p(n,m,L)}I$ for some polynomial p. A rational approximation $\widetilde{H}$ to $G^{1/2}$ with operator error ρ can therefore be computed

in time polynomial in $n, m, L, \log(1/\rho)$. One implementation starts Newton iteration for the matrix square root above the spectral interval and rounds each matrix operation with guard bits. The scalar iteration reduces the initial spectral ratio in polynomially many steps and then converges quadratically. The displayed lower spectral bound controls inverses and roundoff with polynomially many guard bits. A positive definite output can be certified by its squared residual, using $\|\widetilde{H} - G^{1/2}\|_2 \le \|\widetilde{H}^2 - G\|_2/(\lambda_{\min}(\widetilde{H}) + \sqrt{\lambda_{\min}(G)})$. This inequality follows from the integral solution of the associated Sylvester equation.

For explicit error accounting, set $\tau = \varepsilon/128$. Approximate $G^{1/2}$ within $\tau/(2r)$, and normalize grid points to Euclidean error at most τ. The resulting matrix differs from $M(u)$ by at most $\tau + \sqrt{m}\tau \le 2\sqrt{m}\tau$, using equation 10 and the coefficient isometry. Compute each largest eigenvalue within τ. This requires only polynomial precision. In state coordinates one can use the rational basis $R_0 = [e_1 - e_n \ \cdots \ e_{n-1} - e_n]$ and the symmetric generalized eigenproblem with Gram matrix $R_0^\top R_0$. Its eigenvalues lie between one and n. Rational inertia tests and bisection consequently give certified eigenvalue intervals in polynomial bit time, including repeated eigenvalues.

Let $\widetilde{F}$ be the maximum computed eigenvalue, clipped below at zero. Covering and numerical errors together give $|\widetilde{F} - F| \le (\theta + 3\tau)\sqrt{m} = 11\varepsilon\sqrt{m}/128$. Squaring and dividing by m changes the value by at most $2(11\varepsilon/128) + (11\varepsilon/128)^2 < \varepsilon$. Each grid point has polynomial evaluation cost, and there are $(1 + 1/\varepsilon)^{O(r)}$ points. This proves the stated bit complexity. At $r = 0$ the value is exactly zero. For fixed r, choosing half the promised inverse polynomial threshold margin as the accuracy gives the decision consequence in Section 4.2. □

C Structural Extensions

This appendix extends the structural results to supplied approximations. A common stability bound (Appendix C.1) controls approximation by a shared basis (Appendix C.2) and a smaller action span (Appendix C.3). A separate block argument treats spectral clusters of a given reference operator (Appendix C.4). The first two results allow every feasible representation dimension. The action span and cluster algorithms concern a single feature. Appendix C.5 connects effective action rank to exact transition factorizations.

C.1 Stability of the Optimal Value

For this appendix, the objective extends to any family of symmetric contractions on the same centered subspace. Entrywise nonnegativity is unnecessary for the optimization and perturbation arguments. For such a family $T = (T_a)_{a=1}^m$, write $f_d^T(\Phi) = m^{-1}\sum_a \|\Phi^\top T_a \Phi\|_F^2$ on the centered Stiefel domain of equation 1. This convention lets a valid Markov family be compared with a structured operator approximation whose entries may have either sign.

Lemma 8 (Value stability)**.** *Let T_a and R_a be symmetric contractions on the same centered subspace. For every feasible Φ,*

$$|f_d^T(\Phi) - f_d^R(\Phi)| \le \frac{2d}{m}\sum_{a=1}^m \|T_a - R_a\|_2, \qquad |J_d(T) - J_d(R)| \le \frac{2d}{m}\sum_{a=1}^m \|T_a - R_a\|_2. \quad (11)$$

Proof. Put $X_a = \Phi^\top T_a \Phi$ and $Y_a = \Phi^\top R_a \Phi$. The difference of their squared Frobenius norms is bounded by $(\|X_a\|_F + \|Y_a\|_F)\|X_a - Y_a\|_F$. The contraction bounds give $\|X_a\|_F, \|Y_a\|_F \le \sqrt{d}$, while compression gives $\|X_a - Y_a\|_F \le \sqrt{d}\|T_a - R_a\|_2$. Averaging proves the pointwise assertion. Applying it at a maximizer of either family proves both directions of the value assertion. Compactness ensures that these maximizers exist. □

C.2 Approximation under a Supplied Common Basis

Suppose an orthogonal matrix $U = [\mathbf{1}/\sqrt{n} \ u_1 \ \cdots \ u_{n-1}]$ is supplied. Define $R_a = U\,\mathrm{diag}(U^\top T_a U)U^\top$, where diag keeps only the diagonal entries. The R_a are symmetric contractions sharing this basis. They fix the constant direction, and the remaining scores are $\nu_i =$

$m^{-1}\sum_a (u_i^\top T_a u_i)^2$. Write $V_d(U)$ for the sum of their largest d scores and $\delta_a = \|T_a - R_a\|_2$. Theorem 5 applies to the restricted operators and gives $J_d(R) = V_d(U)$.

Lemma 8 immediately yields $|J_d(T) - V_d(U)| \le 2d\, m^{-1}\sum_a \delta_a$. Computing the diagonal scores and selecting the largest ones therefore provides a value approximation once a suitable basis is available. The selected columns also give a feasible representation for the original family. Its loss relative to $J_d(T)$ is at most twice the displayed value error by applying the pointwise bound as well. Both guarantees use the measured discrepancy of the supplied basis. They do not require a procedure that finds the best approximate common basis, and they do not assert that a small discrepancy follows from noncommutativity alone.

C.3 Approximation through a Smaller Action Span

Let $R_1, \dots, R_m$ be supplied rational symmetric contractions on the centered subspace, with span dimension at most r. Define $\eta = m^{-1}\sum_a \|B_a - R_a\|_2$, where B_a are the centered input operators. The proof of Theorem 6 uses only symmetry, contraction, and rational encoding. It therefore applies to this surrogate family and returns a value $\widehat{J}_R$ with algorithmic error at most ε. The scalar case of Lemma 8 gives the combined guarantee

$$|\widehat{J}_R - J(B)| \le \varepsilon + 2\eta. \tag{12}$$

The runtime is $(1 + 1/\varepsilon)^{O(r)}$ times a polynomial in the state count, action count, input bit length, and $\log(1/\varepsilon)$. The encoding length here includes the supplied surrogate matrices. The term 2η records the error of replacing the action family. The term ε records the error of searching its coefficient sphere. These are separate sources of error and can be chosen separately when constructing an approximation. Repeated actions must retain their original weights in the Gram matrix from Appendix B.6, even if the surrogate span has much smaller dimension than the action list.

C.4 Approximation under Spectral Clusters

Let A be a supplied symmetric reference operator on $\mathbf{1}^\perp$. Partition its spectrum into clusters with spectral projectors $P_1, \dots, P_k$. Assume each cluster has dimension at most b, and eigenvalues from different clusters differ by at least $\gamma > 0$. For each action set $R_a = \sum_\ell P_\ell B_a P_\ell$. These block diagonal operators are contractions because each diagonal block is a compression of B_a. Write J_ℓ for the scalar optimum of the restricted family $(P_\ell B_a P_\ell)_a$ on the unit sphere in cluster ℓ.

In an eigenbasis of A, the (i, j) entry of $[A, B_a]$ is $(\lambda_i - \lambda_j)(B_a)_{ij}$. Every entry removed in R_a has $|\lambda_i - \lambda_j| \ge \gamma$. Squaring and summing these entries proves $\|B_a - R_a\|_2 \le \|B_a - R_a\|_F \le \|[A, B_a]\|_F/\gamma$. The separation assumption is essential to this estimate. A small commutator can coexist with large mixing within a cluster, which is why those coordinates remain in a joint block optimization.

For a unit vector x, let $t_\ell = \|P_\ell x\|_2^2$ and normalize each nonzero component to v_ℓ. Its vector of action predictions under R is $\sum_\ell t_\ell q(v_\ell)$, with $\sum_\ell t_\ell = 1$. Convexity of the squared Euclidean norm bounds its objective by $\sum_\ell t_\ell J_\ell \le \max_\ell J_\ell$. A vector supported on a maximizing block attains equality, so $J(R) = \max_\ell J_\ell$. Combining this identity with value stability gives

$$\left| J(B) - \max_\ell J_\ell \right| \le \frac{2}{m\gamma} \sum_{a=1}^{m} \|[A, B_a]\|_F. \tag{13}$$

Each block objective is 4-Lipschitz on its unit sphere, since the quadratic forms are bounded by one and are 2-Lipschitz. A sphere net at resolution $\varepsilon/4$ thus approximates each J_ℓ within ε, adding this amount to equation 13. The lattice construction in Appendix B.6 gives $(1 + 1/\varepsilon)^{O(b)}$ points per block. For rational input, a certified spectral partition and separation bound permit the projectors and block coordinates to be computed to the required precision. Their cost is polynomial in the input size, $\log(1/\varepsilon)$, and $\log(1 + 1/\gamma)$. Numerical error can be included by refining the net and arithmetic tolerances. The guarantee depends on the supplied reference operator, the separation, and the maximum cluster dimension.

C.5 Connections to Transition Factorizations

Action unfolding and CP rank. Write $\widetilde{B}_a = P_0 T_a P_0$ on the full state space and define $\mathcal{B}(s,a,t) = (\widetilde{B}_a)_{st}$. Each $\widetilde{B}_a$ vanishes on the constant direction, so restriction to $\mathbf{1}^\perp$ preserves all linear relations among the actions. If $M_{\mathrm{act}}(\mathcal{B})$ has row a equal to $\mathrm{vec}(\widetilde{B}_a)^\top$, then $r = \operatorname{rank} M_{\mathrm{act}}(\mathcal{B})$. This is the Tucker rank along the action mode (Kolda & Bader, 2009; Ni et al., 2023). Let $\mathcal{T}(s,a,t) = T_a(s,t)$ be the original probability tensor. Centering acts linearly on the other two modes, giving $r \le \operatorname{rank} M_{\mathrm{act}}(\mathcal{T})$.

Suppose the transition probability tensor has an exact CP representation with R components, as in the probability model of Wu & Li (2026). Centering gives

$$T_a(s,t) = \sum_{j=1}^{R} u_j(s) c_j(a) v_j(t), \qquad \widetilde{B}_a = \sum_{j=1}^{R} c_j(a)(P_0 u_j)(P_0 v_j)^\top.$$

The centered actions therefore lie in the span of R fixed matrices, proving $r \le R$. The Walsh family has $r = m$, so any exact probability CP representation of that family requires at least m components. With fixed R, symmetric stochastic kernels, and explicit rational input, Theorem 6 gives polynomial time scalar approximation at inverse polynomial accuracy.

A common next-state factor. Suppose $T_a(s,t) = \phi(s,a)^\top \mu(t)$ with k coordinates, as in the low rank model used by SPEDER (Ren et al., 2023). Write $T_a = F_a M^\top$ and $V = \mathrm{range}(M)$, so every row space lies in V. Symmetry places every column space in V as well. Stochasticity gives $T_a \mathbf{1} = \mathbf{1}$, hence $\mathbf{1} \in V$. All centered operators are consequently supported on $W = V \cap \mathbf{1}^\perp$, whose dimension q is at most $k-1$. For an orthonormal basis matrix U of W, write $\widetilde{B}_a = U C_a U^\top$ with symmetric $q \times q$ matrices C_a. The dimension of the symmetric matrix space gives

$$r \le \frac{q(q+1)}{2} \le \frac{k(k-1)}{2}.$$

This deduction connects a fixed global transition factorization dimension to Theorem 6. It also covers the finite linear and fixed finite latent variable models of Gao et al. (2026) when they satisfy our kernel assumptions. Bounds on individual matrix ranks do not give the same conclusion. Each centered Walsh operator has rank two and each full Walsh kernel has rank three, while their action span grows with m.

Factorization does not imply commutativity. Let $u = (1,-1,0)^\top$, $v = (0,1,-1)^\top$, and define

$$T_1 = \mathbf{1}\mathbf{1}^\top/3 + uu^\top/24, \qquad T_2 = \mathbf{1}\mathbf{1}^\top/3 + vv^\top/24.$$

Both kernels are rational, strictly positive, symmetric, and stochastic. Their joint probability tensor has a CP representation with three components, comprising the common uniform term and the two action specific outer products. It also has a common next-state space of dimension three. Since $u^\top v = -1$, their commutator is $[T_1, T_2] = (vu^\top - uv^\top)/576 \neq 0$. Thus neither factorization condition guarantees the common eigenbasis of Theorem 5.

Exact probabilities and learned scores. The bounds above apply to exact probability factorizations. FaStR also uses a factorized contrastive energy score (Wu & Li, 2026). Diffusion and energy representations can involve infinite dimensional feature spaces (Shribak et al., 2024; Gao et al., 2026). Exponentiation and normalization generally do not preserve a linear rank bound, so finite score dimension alone does not imply either probability factorization. Approximate models can be compared through Appendix C.3 when they supply rational symmetric contractions and control the average operator norm error η. The resulting scalar value error is at most $\varepsilon + 2\eta$. Prediction or contrastive fitting error must first be converted into this operator norm control. These are conditions on supplied transition operators, and learning suitable factors from observations remains a separate computational problem.

D Additional Hardness Results

The value identity yields consequences for several accuracy criteria (Appendix D.1), restricted action growth (Appendix D.2), and perturbations of the kernels (Appendix D.3). We also give a sepa-

rate construction with nonuniform stationarity and nonreversible dynamics (Appendix D.4). Unless specified otherwise, the objective in this appendix is scalar.

D.1 Hardness under Different Approximation Criteria

For the Walsh family define $V = 8m\overline{J} = 1 - 1/\omega(G)$. Consecutive normalized values at threshold K differ by $1/[K(K-1)] > 1/N^2$. An additive estimate of V with error at most $1/(4N^2)$ therefore decides CLIQUE by midpoint thresholding. An estimate of the raw value satisfying $|\widehat{J} - \overline{J}| \le \overline{J}/(4N^2)$ has additive error at most $1/(32mN^2)$. This is smaller than one quarter of the raw consecutive gap. Returning either estimate is NP-hard. The dependence on m cancels in the relative comparison because both the value and its gap carry the same action averaging factor.

Proposition 9 (Multiplicative hardness for the deficit)**.** *Set $D_{\mathrm{sat}} = 1 - 8m\overline{J} = 1/\omega(G)$. Fix $0 < \zeta < 1/2$. Returning an estimate $\widehat{D}$ satisfying*

$$N^{-(1/2-\zeta)} D_{\mathrm{sat}} \le \widehat{D} \le N^{1/2-\zeta} D_{\mathrm{sat}} \tag{14}$$

is NP-hard. This conclusion also holds for every fixed multiplicative factor greater than one.

Proof. The gap form of maximum clique inapproximability gives NP-hard promise instances whose yes and no clique levels have ratio at least $N^{1-\zeta}$ (Zuckerman, 2006). Put $f = N^{1/2-\zeta}$. On a yes instance, the largest permitted deficit estimate is f/ω_{yes}. On a no instance, the smallest is $1/(f\omega_{\mathrm{no}})$. These intervals are disjoint because $\omega_{\mathrm{yes}}/\omega_{\mathrm{no}} \ge N^{1-\zeta} > f^2$. A separator of polynomial bit length can be computed between them and decides the promise problem. Any fixed factor is at most f for all sufficiently large N, giving the last assertion. □

The multiplicative statement concerns distance from the normalized saturation value one. It does not give the same factor hardness for $\overline{J}$ itself. This distinction matters because V approaches one as the clique number increases, whereas D_{sat} retains its reciprocal dependence. The additive and relative statements above specify their own requested precision explicitly.

D.2 Hardness with Slowly Growing Action Counts

Proposition 10 (Polynomial action growth)**.** *For every fixed $\alpha > 0$, the Walsh threshold problem remains NP-hard on families satisfying $r_{\mathrm{eff}} = m \le n^{\alpha}$, where n is the number of Markov states. Approximation to a prescribed inverse polynomial additive accuracy remains NP-hard on these families.*

Proof. Start from a graph with N_0 vertices and m edges. Choose a fixed integer $c \ge \max\{1, 2/\alpha\}$ and add isolated vertices until the vertex count is $N' = N_0^c$. The Walsh lifting uses $n = 2^{\lceil \log_2(N'+1) \rceil}$ states, so $n \ge N'$. Since $m \le N_0^2$, we have $m \le n^{\alpha}$. Padding preserves the edges and clique number, hence also the exact Walsh value. The centered edge matrices remain independent, giving $r_{\mathrm{eff}} = m$. The explicit matrix output has polynomial size for fixed α, and the consecutive gap remains inverse polynomial in that output size. Every centered action still has norm $1/4$. □

The action rank in this construction is unbounded as the source instance grows. The proposition permits an arbitrarily small fixed polynomial exponent. It leaves the fixed rank algorithm of Theorem 6 intact. Padding changes the relation between state count and action count without removing the independent graph directions.

D.3 Perturbations that Preserve the Gap

Let $\widetilde{T}_e$ be symmetric stochastic kernels on the same states, and put $\kappa = m^{-1} \sum_e \|\widetilde{T}_e - \overline{T}_e\|_2$. Lemma 8 gives $|J(\widetilde{T}) - \overline{J}| \le 2\kappa$. If $\kappa \le \overline{\Delta}_{G,K}/8$, every no instance has perturbed value at most $\tau^- + \overline{\Delta}_{G,K}/4$, and every yes instance has value at least $\tau^+ - \overline{\Delta}_{G,K}/4$. The remaining separation is at least $\overline{\Delta}_{G,K}/2$. An additional optimization error at most $\overline{\Delta}_{G,K}/8$ still permits midpoint thresholding. This is a uniform robustness statement for the reduction's value gap. Any additional promised kernel properties, such as strict positivity, must also hold for the chosen perturbations.

D.4 Nonreversible Dynamics with Nonuniform Stationarity

We now use the weighted scalar objective $J_\pi = \max_{\pi^\top h=0,\ h^\top D_\pi h=1} m^{-1} \sum_e (h^\top F_e h)^2$, where $F_e = D_\pi T_e$ is a stationary flow. Set $t = 2^{\lceil \log_2(N+2) \rceil}$ and choose a nonconstant Walsh sign column $g \in \{-1, 1\}^t$. Let $a = (3\mathbf{1} + g)/2$, $Z = \sum_u a_u^2 = 5t/2$, and $\pi_u = a_u^2/Z$. The coordinates of a are one or two, giving stationary mass ratio four and $\pi_{\min} = 2/(5t)$. Choose N other normalized Walsh columns as Q, orthogonal to both $\mathbf{1}$ and g. They satisfy $Q^\top \sqrt{\pi} = 0$.

On any three distinct states, let C be the skew-symmetric matrix with entries $+1$ around an oriented cycle on three states, -1 on the reverse cycle, and zero elsewhere. It has zero row and column sums and norm $\sqrt{3}$. Writing $D = \mathrm{diag}(\pi)$, define

$$F_e = \pi\pi^\top + \frac{1}{4} D^{1/2} Q A_e Q^\top D^{1/2} + \frac{C}{100t^2}, \qquad T_e = D^{-1} F_e. \tag{15}$$

The symmetric term encodes the graph, and the circulation introduces nonreversibility without changing scalar quadratic forms.

All entries in this construction are rational and positive. In fact, $100t^2 (F_e)_{uv} = 16 a_u^2 a_v^2 + 10 a_u a_v r_{uv} + c_{uv}$ for integers $r_{uv}, c_{uv} \in \{-1, 0, 1\}$. Writing $p = a_u a_v \geq 1$, this count is at least $16p^2 - 10p - 1 \geq 5$. Orthogonality to $\sqrt{\pi}$ and the zero circulation sums give $F_e \mathbf{1} = F_e^\top \mathbf{1} = \pi$. Thus T_e is stochastic with stationary distribution π. The nonzero skew part of F_e violates detailed balance, proving nonreversibility. A positive symmetric stochastic matrix would have uniform stationary distribution, so T_e is also nonsymmetric.

For the discriminant $S_e = D^{-1/2} F_e D^{-1/2}$, the centered symmetric part is $Q A_e Q^\top / 4$, with norm $1/8$. The full centered discriminant has norm at most $1/8 + \sqrt{3}/(40t) < 1$, using $\|D^{-1/2}\|_2^2 = 5t/2$. Under $x = D^{1/2} h$, centering removes the stationary term and the skew part contributes zero. Writing $z = Q^\top x$ leaves $x^\top S_e x = z_i z_j / 4$. The positive witness and homogeneity argument from Theorem 2 again put every maximizer in the graph span. The graph identity yields

$$J_\pi = \frac{1}{32m} \left(1 - \frac{1}{\omega(G)} \right), \qquad \Delta_{G,K}^{\mathrm{nr}} = \frac{1}{32mK(K-1)} > \frac{1}{16N^4}. \tag{16}$$

These formulas give rational threshold hardness and additive hardness at accuracy $1/(64N^4)$. Exact transition counts are $100t^2 (F_e)_{uv}$, totaling $100t^2$ per action. Their row and column sums are $40t\, a_u^2$. Both empirical marginals equal π, so they reproduce the weighted centered objective and predictor elimination exactly. For independent stationary samples, Proposition 7 applies with $\pi_{\min} = 2/(5t)$ and $\beta = 1/8$. A sufficiently large universal constant in $L \geq C N^9 \log(8mN/\delta)$ then preserves the value within $1/(64N^4) < \Delta_{G,K}^{\mathrm{nr}}/4$. This sampling extension remains a scalar result.

E Geometry of the Prediction Objective

The reduction also gives explicit suboptimal local maxima (Appendix E.1) that persist under small perturbations (Appendix E.2). Simple boundary examples clarify what the value identity implies about maximizing features (Appendix E.3). These are properties of the objective itself.

E.1 Strict Suboptimal Local Maxima

Take $G = K_p \sqcup K_q$ with $p > q \geq 2$ and $m = \binom{p}{2} + \binom{q}{2}$. Let z_q put positive mass $1/\sqrt{q}$ on the smaller clique and zero elsewhere, and let $x_q = \overline{Q} z_q$. Its value is $(1 - 1/q)/(8m)$, below the global value $(1 - 1/p)/(8m)$ by $G_q = (p - q)/(8pqm)$. We show that x_q is nevertheless a strict local maximum on the centered unit sphere.

Write a nearby feature as $x = \overline{Q} z + w$ and define $\Psi(x) = (z_1^2, \ldots, z_N^2, \|w\|_2^2)$. Let t be its total mass outside the smaller clique, and write its coordinates on the smaller clique as $y_i = (1-t)/q + v_i$ with $\sum_i v_i = 0$. Set $A_q = (q-1)/(2q)$. The smaller clique contributes $A_q (1-t)^2 - \|v\|_2^2/2$, while all other graph edges contribute at most $t^2/2$. For $t \leq 1/2$, this gives

$$A_q - F_G(z_1^2, \ldots, z_N^2) \geq 2A_q t - (A_q + 1/2) t^2 + \frac{1}{2} \|v\|_2^2 \geq \frac{t}{8} + \frac{1}{2} \|v\|_2^2.$$

Equality with the local value requires $t = 0$ and $v = 0$. Keeping the coordinates on the smaller clique positive then forces $x = x_q$, proving strictness.

An explicit boundary margin will be useful below. Choose $0 < \rho \leq \min\{1, 1/(2q)\}$ and take the closed neighborhood $\|\Psi(x) - \Psi(x_q)\|_1 \leq \rho$ with those coordinates positive. They cannot reach zero inside this neighborhood because each initially has squared value $1/q$. On the boundary, $t \leq \rho/2$ and $\rho \leq 2t + \sqrt{q}\|v\|_2$. If $t \geq \rho/4$, the loss above is at least $\rho/32$. Otherwise $\|v\|_2 > \rho/(2\sqrt{q})$. In either case the Walsh objective loses at least $M_\rho = \rho^2/(128qm)$. The centered sphere ball of radius $\rho/4$ around x_q lies inside this neighborhood, since $\|\Psi(x) - \Psi(x_q)\|_1 \leq 2\|x - x_q\|_2$.

E.2 Persistence under Perturbations

Suppose symmetric stochastic kernels $\widetilde{T}_e$ satisfy $\kappa = m^{-1}\sum_e \|\widetilde{T}_e - \overline{T}_e\|_2 < \min\{M_\rho, G_q\}/4$. Pointwise stability changes every objective value by at most 2κ. The perturbed value at x_q therefore exceeds every boundary value of the closed neighborhood by at least $M_\rho - 4\kappa > 0$. A maximizer over that compact neighborhood lies in its interior and is a local maximum on the full centered sphere. Throughout the neighborhood, the original objective is at most its value at x_q. A feature on the larger clique thus has perturbed value exceeding every neighborhood value by at least $G_q - 4\kappa > 0$. The local maximum remains suboptimal. This argument asserts existence after perturbation. It makes no claim about an optimizer's probability of reaching that point.

E.3 Illustrative Graph Instances

Maximizing features need not identify a unique maximum clique. For a complete bipartite graph, placing total squared mass $1/2$ on each part gives $F_G = 1/4$, regardless of the distribution within each part. This attains the value associated with clique number two even when many vertices have positive coordinates. Sign changes of graph coordinates also preserve the objective. The reduction extracts the clique number from the optimal value and does not require every optimizer to have clique support.

The Walsh formula gives $\overline{J} = 1/16$ for a single edge, including any number of isolated vertices, $\overline{J} = 1/80$ for the cycle on five vertices, and $\overline{J} = 1/(4N^2)$ for K_N. The constant unit feature attains one before centering and is excluded from all these comparisons. These values provide direct implementation checks for normalization and action averaging.

An earlier sparse lifting uses $2N$ states, columns $q_i = (e_{i,+} - e_{i,-})/\sqrt{2}$, and kernels $T^{\mathrm{sp}}_{ij} = \mathbf{1}\mathbf{1}^\top/(2N) + QA_{ij}Q^\top/N$. Its centered action norm is $1/(2N)$ and its value is $(1 - 1/\omega(G))/(2mN^2)$. Thus the value for a single edge is $1/(4N^2)$ and the value for the complete graph is $1/N^4$. These belong to a different normalization. The active graph experiments in Appendix F use the Walsh value scale.

F Supplementary Experiments

This appendix provides the experimental settings and additional results supporting Section 5. We describe finite trajectory estimation (Appendix F.2), optimization under fixed budgets (Appendix F.3), and exact and perturbed shared structure (Appendix F.4). Additional robot prediction experiments examine sharing among action predictors (Appendix F.5).

F.1 Detailed Experimental Settings

Table 1 summarizes the experimental configurations. Each graph and operator configuration uses 12 random seeds. Graph pairs are reused across trajectory budgets and observation conditions. Optimization graphs are reused across methods and evaluation budgets. Each operator family is evaluated at all perturbation levels. We compute graph resolution rates with one outcome per pair and condition. Operator plots summarize instance distributions using medians and interquartile ranges. Robot results use 20 episode splits per dataset, with aggregation specified in Appendix F.5.

Table 1: Experimental configurations. Budgets, methods, and perturbation levels repeat the graph pairs, graphs, or base operator families listed below.

Study	Instances	Scale	Repeated conditions
Finite trajectories	48 graph pairs	$N = 10, 14$; $s = 16$	3 budgets; 2 observation conditions
Optimization budgets	72 graphs	$N = 20, 36, 52$	3 budgets; 4 methods
Shared structure	96 base families	$p = 8, 16$; $m = 4, 12$	6 perturbation levels, including zero
Robot prediction	3 datasets	$p = 2, 8, 4$; 12 action regions	20 splits; 5 training fractions

F.2 Finite Trajectory Experiments

Graph construction and sampling. The graph configurations are $(N, K-1, K, m) = (10, 3, 4, 12)$ and $(14, 4, 5, 22)$. Each pair has matched edge counts and clique numbers $K - 1$ and K. We plant the target clique and sample the remaining edges uniformly or with a preference for vertices in the same random block. The latter generator assigns vertices to three blocks and gives edges inside a block three times the sampling weight. Exact branch and bound rejects graphs with an incorrect clique number. Both configurations yield $s = 16$ states through the Walsh construction in equation 2. A trajectory discards its first $8s$ steps and retains a subsequent horizon of mL steps, where $L \in \{512, 2048, 8192\}$.

Observations and estimation. Clean observations use uniform actions. The perturbed condition combines 2% state replacement, 5% missingness, and action probabilities proportional to $a^{-1.2}$. Only transitions with both observations present enter estimation, so retained action counts vary. We add $1/2$ to each transition and marginal count and normalize transition rows. The pooled marginal defines empirical whitening. We symmetrize each whitened operator and project orthogonally to $\sqrt{\widehat{\pi}}$. Feature search uses unit directions orthogonal to $\mathbf{1}$. Population evaluation recenters and normalizes the selected direction under the uniform measure. This trajectory estimator has different sampling assumptions from Theorem 4.

Learners and evaluation. Riemannian ascent uses ten centered Gaussian initializations and at most 2,200 steps per start. It stops at tangent gradient norm 10^{-9} or after 25 unsuccessful backtracking trials. The initial step size is 0.25, the Armijo constant is 10^{-4}, and each rejection halves the step. An accepted step increases it by a factor of 1.25, capped at 0.8. The spectral method selects the leading eigenvector of the average squared estimated operator. Random search selects the best of 256 centered directions. A graph pair is resolved when the selected population value at clique level K exceeds that at level $K - 1$.

Figure 4(a) reports the absolute error in this value difference divided by the true adjacent gap. At $L = 8192$, the Riemannian median relative gap error is 0.160 under clean observations and 0.436 under combined perturbations. This distribution measures the accuracy of the recovered separation alongside the resolution rates in the main text.

F.3 Optimization under Fixed Budgets

Graphs and initialization. We generate graphs at $N \in \{20, 36, 52\}$. Sparse graphs sample each edge independently with probability 0.13. Planted graphs use probability 0.20 and add a clique of size $\max\{4, \lfloor\sqrt{N}\rfloor + 1\}$. Exact branch and bound supplies the reference clique number. The four optimizers maximize $F_G(y)/(4m)$, with starting points drawn uniformly from the simplex. Separate runs use budgets of 100, 500, and 2,000 objective evaluations.

Solver settings. Projected ascent uses step size 0.8 and Euclidean simplex projection. Mirror ascent multiplies coordinates by exponentiated gradient components with step size eight, clips the exponents to $[-20, 20]$, and renormalizes. Both methods restart when improvement is at most 10^{-14}. SLSQP uses a function tolerance of 10^{-11} and an iteration limit of 10,000 per start, subject to the shared evaluation cap. Its numerical derivative evaluations count toward this cap. Analytic gradients for projected and mirror ascent incur additional computation. Each method returns the largest objec-

tive value evaluated before exhausting its budget. Random search evaluates fresh simplex samples throughout.

A hit has absolute error at most 10^{-7} from the certified value. Figure 4(b) reports the shortfall from this value divided by the certified value, at 2,000 evaluations. Tiny negative errors from numerical roundoff are displayed as zero. Panel (c) shows certification costs, with median times of 0.0206, 0.0447, and 0.0807 ms for the three graph sizes. These costs describe exact reference computation on the sampled instances.

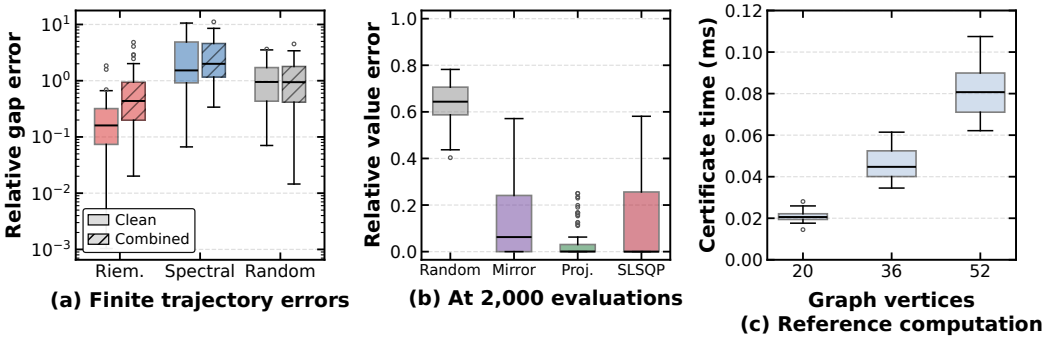

Figure 4: Additional error and cost distributions. (a) Relative adjacent gap error on 48 graph pairs per observation condition at $L = 8192$. (b) Relative value error on 72 graphs at 2,000 evaluations. (c) Certification time for 24 graphs per size. Boxes show quartiles and medians. Whiskers extend to observations within 1.5 interquartile ranges, with outliers shown individually.

F.4 Shared Structure Experiments

Exact common basis. We generate symmetric operators in dimensions $p = 8, 16$ with $m = 4, 12$ actions and optimize the scalar objective with $d = 1$. A Gaussian matrix followed by QR decomposition supplies a common orthogonal basis. Eigenvalues are initially uniform on $[-0.20, 0.20]$. We rescale the column with the largest average squared eigenvalue score so that its separation from the second largest score is 0.008 or 0.0002. These are score separations. The inputs are symmetric operators without an imposed stochasticity constraint.

Perturbations and solvers. Each base family receives symmetric Gaussian perturbations normalized to operator norm 0.0003, 0.001, 0.003, 0.01, or 0.03. Spectral optimization uses the leading eigenvector of the average squared operator. Anchor diagonalization evaluates every eigenvector of the first action and selects the best under the shared objective. Riemannian ascent uses 12 normalized Gaussian starts and at most 900 steps per start. Its gradient tolerance is 10^{-10}, initial step size is 0.20, and backtracking limit is 24 trials. The Armijo constant, rejection factor, growth factor, and maximum step size match the trajectory learner. Every method receives the operators. The generating basis supplies only the exact reference.

Reference values and aggregation. At zero perturbation, the largest generating score gives the exact optimum, with hit tolerance 10^{-8}. Comparisons with k starts use the largest terminal value among the first k initializations. For perturbed operators, the achieved objective is $f(x) = m^{-1} \sum_a (x^\top B_a x)^2$. We compute U_{spec} as the largest eigenvalue of the average squared operator and report $U_{\mathrm{spec}} - f(x)$. This quantity bounds suboptimality, and the true optimum is unknown. Each score regime contains 48 base families, reused across perturbations. Curves and bands summarize their median and interquartile range.

F.5 Additional Robot Prediction Experiments

Datasets and preprocessing. PushT, ALOHA simulated insertion, and xArm lift provide pairs of consecutive states within episodes. Each of 20 seeded splits assigns approximately 60%, 20%, and 20% of episodes to training, validation, and testing. Nested training subsets use 5%, 10%, 25%, 50%, and 100% of a shuffled episode order. Each subset refits state standardization, a principal

component projection with at most eight coordinates, and twelve action regions formed by quantiles along the leading action principal component. The resulting state dimensions are two, eight, and four. All preprocessing is fitted on training episodes and applied to the disjoint validation and test episodes.

Models and selection. The joint model uses $W_a = W_0 + \sum_{j=1}^{r} c_{aj} B_j$. Eight alternating ridge updates fit the bases and action coefficients, initialized by a singular value decomposition of centered independent heads. Controls use independent heads, complete pooling, equal training counts per action region, shuffled training action labels, or fixed random action coordinates. Each method selects its own validation optimum. Joint and random coordinate models search $1 \le r \le \min\{5, p^2\}$. All methods search ridge penalties in $\{10^{-5}, 10^{-3}, 0.03, 0.3\}$. This task measures regression sharing among predictors with fixed state preprocessing. The shared encoder objective and coefficient sphere algorithm remain separate theoretical objects.

Metrics and additional results. Gain is $(\mathrm{MSE}_{\mathrm{ind}} - \mathrm{MSE}_{\mathrm{method}})/\mathrm{MSE}_{\mathrm{ind}}$, using test mean squared error. We average the gains at the three smallest training fractions within each split, then compute means and paired differences across 20 splits. Table 2 shows positive joint gains over independent heads on all three datasets. Pooling, random coordinates, and shuffled labels outperform the joint model on ALOHA. The joint model exceeds these controls on PushT and xArm. Paired Wilcoxon tests assess differences in either direction, with Holm correction across nine dataset and control comparisons. The largest adjusted p-value is 0.00121. Repeated splits share source episodes, so uncertainty summaries describe this resampling protocol.

Table 2: Robot prediction at the three smallest training fractions. Joint gains are percentages relative to independent heads. Paired differences are percentage points. Positive differences favor the joint model. Entries give means with descriptive half widths $1.96s/\sqrt{20}$ across splits, where s is the sample standard deviation.

Dataset	Joint gain	Joint minus pooled	Joint minus random	Joint minus shuffled
PushT	2.72 ± 0.68	5.67 ± 0.71	3.91 ± 0.59	7.64 ± 0.90
ALOHA	83.33 ± 1.99	-1.99 ± 0.57	-1.64 ± 0.47	-1.09 ± 0.54
xArm	2.79 ± 0.20	1.72 ± 0.40	0.79 ± 0.33	6.43 ± 0.50