THE CONNECTIVITY COST OF SHARED COORDINATES IN WORLD MODELS

Anonymous authors
Paper under double-blind review

ABSTRACT

In world models, each action can be implemented with a concise connectivity pattern under its own optimal coordinate system, but the optimal coordinate systems for different actions are generally inconsistent. World models require all actions to share the same coordinate system in order to read and update a consistent representation, yet this causes an increase in the number of connections required for the actions. Existing work has primarily focused on coordinate alignment methods that reduce the number of connections, but a specific characterization of how the connectivity scale changes when actions move from their respective coordinates to shared coordinates is still lacking. This paper investigates this connectivity expansion induced by shared state coordinates. In the symmetric linear model analyzed in this paper, we prove that for families of actions lacking common structure, each action requires only a linear number of connections under its own optimal coordinates, whereas the minimum number of connections typically grows to quadratic order under shared coordinates. The reason is that the connectivity structures required by different actions generally cannot be sufficiently aligned and reused within a single coordinate system, and this gap persists even when approximate compression is allowed. We further show that if different actions share a fixed sparse connectivity pattern under the same coordinate system, the number of connections required for the shared state can still remain at a linear scale, and we present a structure that avoids this increase in asymptotic order. Numerical experiments reproduce the exact connection counts, the three budget regimes, and the same effect in learned world models on a physical system with a known interaction graph.

1 INTRODUCTION

World models learn environment states and how they change under actions, thereby providing predictions of future states for tasks such as robotic manipulation and interactive prediction (Ha & Schmidhuber, 2018; Hafner et al., 2019; 2020; Finn et al., 2016; Hafner et al., 2021; Hansen et al., 2024; Hafner et al., 2025). During continuous interaction, the model needs to maintain an internal state that evolves over time, for different actions to read and update (Ha & Schmidhuber, 2018; Hafner et al., 2020; Singh et al., 2021). To leverage the structure of objects and relations in the environment, many models organize the internal state into multiple units that can be read and updated separately, with the hope that each state update involves only a small number of relevant units (Battaglia et al., 2016; Sanchez-Gonzalez et al., 2018; Kipf et al., 2018; 2020; Locatello et al., 2020; Kipf et al., 2022; Sanchez-Gonzalez et al., 2020; Veerapaneni et al., 2019; Yu et al., 2024). Such local connectivity provides both a structured state representation and a reduction in the number of state relations that need to be processed per update.

The connectivity pattern changes with the choice of internal coordinates, and related questions have long been studied in state-space representations and matrix decompositions (De Schutter, 2000; Bunse-Gerstner et al., 1993; Cardoso & Souloumiac, 1996; Pham, 2001). One line of work studies how state coordinate transformations can yield sparser or simpler realizations of the dynamics, that is, how to find the optimal coordinates for an individual action. However, a persistent state requires different actions to repeatedly operate on the same coordinate system, which has given rise to another line of work more closely related to this paper. Joint diagonalization and joint block diagonalization seek a transformation that simultaneously simplifies multiple update matrices in the same coordinate

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.

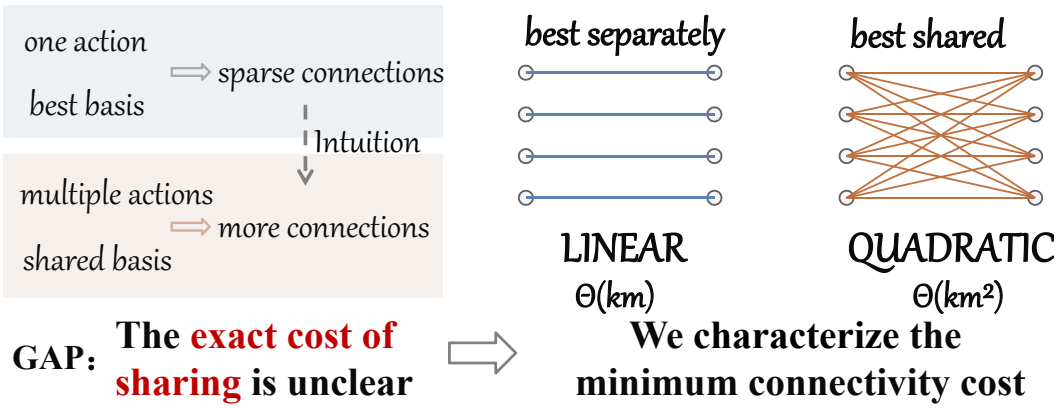

Figure 1: One action is sparse in its own best basis, while several actions constrained to a single shared basis need more connections. This paper compares the separate and shared optima, $\Theta(km)$ against $\Theta(km^2)$ for generic symmetric families, and characterizes the minimum connectivity cost of sharing.

system (Abed-Meraim & Belouchrani, 2004; Maehara & Murota, 2011; Cai et al., 2015). These works mainly address whether different actions admit shared coordinates in which their updates can be simplified and how to find or approximate such coordinates (Chabriel et al., 2014; Bouchard et al., 2020; Cai & Liu, 2017). To our knowledge, these studies do not typically analyze the difference in connectivity scale before and after imposing shared coordinates, nor do they characterize whether this difference changes the asymptotic order of the connection count. Thus, prior work has not yet answered the following question: when each action can be implemented with only a small number of connections on its own, can sharing a single state coordinate system still maintain the same low connection scale?

We prove that, in the symmetric linear model studied in this paper, sharing state coordinates increases the minimum number of required connections from linear to quadratic order. Even if each action can restrict state updates to a linear number of unit connections under its own optimal coordinates, there is no guarantee that only the same order of connections is needed once all actions share the coordinates. This is because the optimal coordinates that maintain concise updates for each action are inconsistent, so the small number of connections originally required by different actions cannot be sufficiently aligned and reused in the shared coordinates. The model must retain a large number of additional cross-unit connections to implement all actions simultaneously. This phenomenon is not merely a fragile consequence of strictly requiring connections to be zero, and it persists even when some connections are discarded to obtain approximate representations. The above mechanism also implies that if all actions follow a common sparse connectivity pattern under the same state coordinate system, with different actions only varying the specific updates on these connections, the required connections still grow linearly with the number of state units. Based on this, we present a class of structures that can avoid the aforementioned connectivity expansion.

The main contributions of this paper are as follows.

- The scaling of the minimum number of connections when different actions use their respective optimal coordinates versus when all actions share the same coordinates has not been characterized, and this paper proves that in a symmetric linear model, generic action families exhibit growth from linear to quadratic order.
- We further show that this growth is not a consequence of exact sparsity, and when the shared connection budget falls below a certain level, approximate representations still incur non-negligible update errors.
- We characterize a class of common sparse connectivity structures that can avoid this growth, and use numerical experiments to verify the roles of connectivity expansion, approximation error, and common structure.

2 BACKGROUND & MOTIVATION

World models maintain an internal state that evolves over time and update this state according to the current action. To exploit objects and local relations in the environment, many structured models divide the state into multiple units that can be accessed separately, such as graph nodes, entities, or object slots, and encourage each update to involve only a small number of relevant units (Watters et al., 2017; Greff et al., 2019; Jiang et al., 2020; Pfaff et al., 2021; Kossen et al., 2020; Wu et al., 2023). In such representations, the pairs of units that exchange information under an action naturally define its connectivity structure. An update is local when only a small number of these relations are required. An important but easily overlooked fact is that these units are not uniquely determined by the dynamics. The same state can be reorganized through a change of coordinates, so whether an update appears local can also depend on the choice of internal coordinates.

This dependence on coordinates is not specific to world models. In linear systems and matrix analysis, the same dynamics can have different representations under different state coordinates, and coordinate transformations have long been used to expose simpler matrix structures such as diagonal or block structures (De Schutter, 2000; Bunse-Gerstner et al., 1993; Cardoso & Souloumiac, 1996). A simple example illustrates the issue. Suppose two actions each operate independently on state units in their own preferred coordinates, so that each action requires only a linear number of connections when considered separately. If the two coordinate systems differ by a rotation that mixes the units, then fixing one coordinate system can cause the update of the other action to spread across many pairs of units. Therefore, the fact that every action is individually local does not imply that all actions remain local in the same state representation. This distinction is especially important for persistent world models, where different actions repeatedly operate on the same state.

The closest line of work is joint diagonalization and joint block diagonalization. Rather than choosing coordinates separately for each matrix, these methods seek a common transformation under which multiple matrices simultaneously exhibit simpler structure. Prior work has studied when such common transformations exist, how they can be identified and computed, and how approximate common structure can be found (Abed-Meraim & Belouchrani, 2004; Chabriel et al., 2014; Cai et al., 2015; Bouchard et al., 2020), their perturbation sensitivity (Afsari, 2008; Cai & Li, 2019), robust variants for nearly commuting matrices (He & Kressner, 2024), and applications (Eynard et al., 2015; Nazerian et al., 2023). These studies naturally address whether several operators can be simplified in shared coordinates and how such coordinates can be obtained. The connectivity question in persistent world models is different. Even when no ideal common diagonal or block-diagonal structure exists, all actions must still operate on the same state coordinates. The question that matters is how many additional connections shared coordinates require, compared with letting each action use its own optimal coordinates. It is also important to know whether this difference only adds a finite number of connections at a fixed dimension or changes the asymptotic order of the connection count as the number of state units grows. To our knowledge, existing work on joint matrix decompositions has mainly focused on common structure itself and on methods for finding it, rather than on an asymptotic characterization of this shared-versus-separate connectivity gap.

The central gap lies elsewhere. It is the lack of a direct comparison between two representation requirements that naturally arise in persistent state models. One allows each action to choose the coordinates that make it as local as possible. The other requires all actions to repeatedly read and update the same state coordinates. The first determines the minimum connectivity that each action can achieve individually, while the second determines the connectivity required by a persistent shared state. Comparing these two optima directly measures the structural cost of sharing coordinates across actions. Exact support alone is also too sensitive to arbitrarily small perturbations, so the same question must be studied under a finite connection budget by measuring the amount of update information that cannot be retained. Finally, if some action families can reuse the same sparse relations in shared coordinates, it is necessary to identify what cross-action structure allows shared locality to be preserved. This paper studies these questions by examining the connectivity gap created by shared coordinates, whether the gap persists under approximation, and what forms of cross-action structure keep shared connectivity small.

3 SETUP

We now fix the state units, the count of active connections, and the two optima that the rest of the paper compares. Let the state space be $V = \mathbb{R}^d$ with $d = mr$. We view it as m state units of equal size r,

$$V = V_1 \oplus \cdots \oplus V_m, \qquad \dim V_i = r. \tag{1}$$

A block therefore represents one directed relation from one state unit to another. We study a family of k action-conditioned linear updates

$$\mathcal{A} = (A_1, \ldots, A_k), \qquad A_a \in \mathrm{Sym}_d(\mathbb{R}). \tag{2}$$

Here $\mathrm{Sym}_d(\mathbb{R})$ and $O(d)$ denote real symmetric matrices and the orthogonal group. The symmetric class is our reference model because orthogonal coordinate changes preserve the class and every individual action admits an eigenbasis. An orthogonal basis $Q \in O(d)$ represents the same action as $Q^\top A_a Q$.

For a matrix M partitioned into $m \times m$ blocks of size $r \times r$, define

$$\mathrm{bnnz}_r(M) = \#\{(i,j) : \|M_{ij}\|_F > 0\}. \tag{3}$$

Here M_{ij} is block (i,j), $\|\cdot\|_F$ is the Frobenius norm, and the count measures active directed relations between units. For symmetric matrices, an off-diagonal relation appears in both directions and contributes two blocks. We compare

$$C_{\mathrm{sep}}(\mathcal{A}) = \sum_{a=1}^{k} \min_{Q_a \in O(d)} \mathrm{bnnz}_r(Q_a^\top A_a Q_a), \tag{4}$$

$$C_{\mathrm{sh}}(\mathcal{A}) = \min_{Q \in O(d)} \sum_{a=1}^{k} \mathrm{bnnz}_r(Q^\top A_a Q), \tag{5}$$

$$\mathfrak{F}(\mathcal{A}) = C_{\mathrm{sh}}(\mathcal{A}) - C_{\mathrm{sep}}(\mathcal{A}). \tag{6}$$

C_{sep} asks how local the actions can be when each chooses its own coordinates, C_{sh} asks the same when one coordinate system serves every action, and $\mathfrak{F}$ is their difference. Throughout, "generic" means outside a measure-zero set, and asymptotics take $m \to \infty$ with k, r fixed.

So far the comparison rests on exact support, which treats every nonzero block equally, including arbitrarily small ones. We also ask how much update energy must be discarded under a finite shared budget B. Define

$$\mathcal{E}_B(\mathcal{A}) = \inf_{Q \in O(d)} \inf_{\substack{\widetilde{A}_a \in \mathrm{Sym}_d(\mathbb{R}) \\ \sum_a \mathrm{bnnz}_r(\widetilde{A}_a) \le B}} \frac{\sum_{a=1}^{k} \|Q^\top A_a Q - \widetilde{A}_a\|_F^2}{\sum_{a=1}^{k} \|A_a\|_F^2}. \tag{7}$$

The numerator is the energy omitted by the best low-budget representation and the denominator makes it comparable across scales, so $\mathcal{E}_B = 0$ means the budget represents the family exactly. We write quadratic-scale budgets as

$$B_m = \lfloor \rho k m^2 \rfloor, \tag{8}$$

where ρ is the retained directed-block density across the k actions.

4 OBSTRUCTION

We first quantify the cost of forcing individually simple actions into one basis, then test whether a finite approximation budget removes that cost.

4.1 CONNECTION GROWTH

Theorem 1 (Connection growth under a shared basis)**.** *Fix $k \ge 2$, $m \ge 1$, $r \ge 1$, and $d = mr$. For a generic family $\mathcal{A} \in \mathrm{Sym}_d(\mathbb{R})^k$,*

$$C_{\mathrm{sep}}(\mathcal{A}) = km, \qquad C_{\mathrm{sh}}(\mathcal{A}) = (k-1)m^2 + m. \tag{9}$$

Consequently $C_{\mathrm{sep}} = \Theta(km)$ while $C_{\mathrm{sh}} = \Theta(km^2)$ for fixed k, r.

Proof sketch. Each symmetric action can choose its own eigenbasis, placing its update in m diagonal blocks, and generic full rank rules out fewer. The upper bound uses an eigenbasis of the first action, spending m blocks on it and at most m^2 on each of the other $k-1$. For the matching lower bound, fix any joint support of total budget $B < (k-1)m^2 + m$: its free entries, together with the shared basis directions *minus the within-block rotations that act trivially on the support*, span strictly fewer dimensions than the ambient space. Removing that redundancy closes the former $m-1$ gap. The representable families form a measure-zero set, as does the finite union over supports; see Appendix A. □

One common basis can absorb the orientation of roughly one action; each further generic action then contributes another quadratic-scale set of relations in the same coordinates, which is where the coefficient $k-1$ comes from.

General real operators. The symmetric class is the reference model because an orthogonal change of coordinates fully diagonalizes a single action. Learned Jacobians are not symmetric, and an orthogonal basis brings a single real operator only to real Schur form, which already occupies $m(m+1)/2$ blocks. For even block width r and generic $\mathcal{A} \in (\mathbb{R}^{d\times d})^k$,

$$C_{\mathrm{sep}}(\mathcal{A}) = k\frac{m(m+1)}{2}, \qquad C_{\mathrm{sh}}(\mathcal{A}) = (k-1)m^2 + \frac{m(m+1)}{2}. \tag{10}$$

In the symmetric regime the separate cost is $\Theta(m)$ and the shared cost $\Theta(m^2)$, so the ratio diverges with m. In the general real regime the separate cost is already $\Theta(m^2)$, sharing still adds a quadratic term, and the ratio stays bounded, tending to $2 - 1/k$. A large relative expansion is a property of individually diagonalizable families. Appendix A treats odd block widths and the asymptotics in r.

4.2 Approximation at a Finite Budget

Define the critical shared density

$$\rho_c(k) = 1 - \frac{1}{k}. \tag{11}$$

We draw each A_a independently from a Frobenius-isotropic Gaussian on $\mathrm{Sym}_d(\mathbb{R})$, meaning that its coefficients in any Frobenius-orthonormal basis of the symmetric matrix space are independent and identically distributed (i.i.d.) standard Gaussian.

Theorem 2 (Finite-budget obstruction)**.** *Let $B_m = \lfloor \rho k m^2 \rfloor$. If $\rho < \rho_c(k)$, there are constants $c, \gamma > 0$, depending only on k, r, ρ, such that for all sufficiently large m,*

$$\Pr[\mathcal{E}_{B_m}(\mathcal{A}) \geq c] \geq 1 - e^{-\gamma m^2}. \tag{12}$$

If $\rho > \rho_c(k)$, then for all sufficiently large m, every symmetric family satisfies $\mathcal{E}_{B_m}(\mathcal{A}) = 0$.

Proof sketch. Low-budget shared representations form a closed cone and $\mathcal{E}_B$ is the squared normalized distance to it. For each support, nets for the free matrix parameters and for the shared basis cover the normalized representations, and conjugation is Lipschitz in the basis, so the nets compose. A spherical-cap bound then puts a random ambient direction near any low-budget representation with probability at most $e^{-\Omega(m^2)}$, the exponent being the codimension left when $\rho < 1 - 1/k$. Above the threshold one action's eigenbasis is an exact representation. Appendix B gives the net, the finite-dimensional constant and the failure probability. □

Below the critical density, discarded relations carry a non-vanishing fraction of the operator energy with high probability, so approximate sparsity does not remove the obstruction. Two refinements make the threshold usable at measurable sizes. The construction of Theorem 1 needs

$$B_* = (k-1)m^2 + m, \qquad \text{i.e.} \qquad \rho_*(m,k) = \frac{B_*}{km^2} = \rho_c(k) + \frac{1}{km}, \tag{13}$$

so the usable threshold at finite m sits above ρ_c by exactly $1/(km)$. At the critical density every finite generic family still fails to be represented exactly, yet the attainable normalized error tends to 0 with m. Subcritical, critical and supercritical are three distinct regimes, shown in Section 6.2.

If instead the state is converted whenever the action changes, a generic relative basis uses all m^2 block relations, so over H steps with switch rate q_H the average interaction count is $m + q_H m^2$; a constant switch rate moves the quadratic term to coordinate conversion, where it survives (Appendix D).

The obstruction results raise the next question. Which structural families can share coordinates while keeping low connectivity? Section 5 answers this through common sparse compatibility.

5 COMPATIBILITY

The previous section describes generic incompatible families. We now identify a shared structural envelope that keeps connectivity linear and extend it to approximate structure.

5.1 SHARED SPARSE STRUCTURE

Let $G = ([m], E)$ be an undirected graph with state units $[m] = \{1, \dots, m\}$ and edge set E. Its allowed block envelope contains all m diagonal blocks and the two directed blocks associated with each edge, so one action has envelope size

$$s_G = m + 2|E|. \tag{14}$$

Proposition 1 (Common sparse envelope)**.** *Suppose there is a basis Q_* such that every $Q_*^\top A_a Q_*$ has support inside the same graph envelope G. Then*

$$C_{\mathrm{sh}}(\mathcal{A}) \le k(m + 2|E|) = k s_G. \tag{15}$$

If G has bounded degree, then $C_{\mathrm{sh}} = O(km)$.

Proof. Use Q_*. Each action has at most m diagonal blocks and two directed blocks per graph edge. Summing over actions gives the displayed bound, and bounded degree implies $|E| = O(m)$. □

The actions may place different numerical updates on the allowed relations; shared locality asks only that they agree on where interactions live. The claim is about scale: a common sparse envelope keeps shared connectivity linear in m and so avoids the quadratic expansion of Theorem 1. Neither this proposition nor Theorem 3 uses symmetry, since both need only one common support and one common basis, so both hold for general real families with a directed envelope.

Exact support becomes dense under arbitrarily small perturbations, so we next retain a sparse component and measure the residual energy.

Theorem 3 (Approximate common structure)**.** *Suppose a basis Q_* and matrices S_a, R_a satisfy*

$$Q_*^\top A_a Q_* = S_a + R_a, \tag{16}$$

where every S_a lies in the same graph envelope G. Then

$$\mathcal{E}_{k s_G}(\mathcal{A}) \le \frac{\sum_a \|R_a\|_F^2}{\sum_a \|A_a\|_F^2}. \tag{17}$$

Proof sketch. Choose the feasible shared basis $Q = Q_*$ and approximate each action by S_a. The representation uses at most $k s_G$ blocks, and its discarded energy is exactly $\sum_a \|R_a\|_F^2$. The optimum in the definition of $\mathcal{E}_B$ can only reduce this value. A formal version appears in Appendix C. □

The bound makes the positive result continuous: a small perturbation may activate every block in an exact support count while its energy stays concentrated near the common envelope. Theorem 2 places generic families a constant distance from low-budget representations, and Theorem 3 gives small low-budget error whenever a family lies close to a shared sparse structure.

5.2 MEASURING COMPATIBILITY

To compare these regimes on one axis, let Ω be a symmetric directed block envelope with $|\Omega| \le s$, shared by all actions. Define

$$\Gamma_s(\mathcal{A}) = \inf_{Q \in O(d)} \inf_{\substack{|\Omega| \le s \\ \mathrm{supp}(S_a) \subseteq \Omega \ \forall a}} \frac{\sum_a \|Q^\top A_a Q - S_a\|_F^2}{\sum_a \|A_a\|_F^2}. \tag{18}$$

Here supp denotes active block positions. Γ_s is the normalized distance of the entire action family to its best common sparse envelope of size s. Small Γ_s means the actions can reuse one set of relation locations after a shared coordinate change.

Theorem 4 (Budget–compatibility sandwich)**.** *For every action family and every s,*

$$\mathcal{E}_{ks}(\mathcal{A}) \le \Gamma_s(\mathcal{A}) \le \mathcal{E}_s(\mathcal{A}). \tag{19}$$

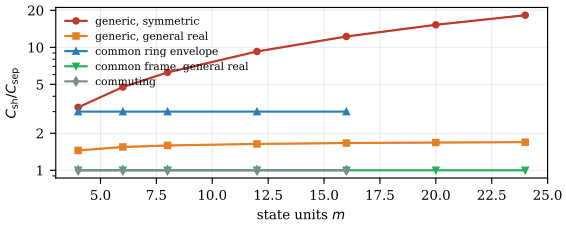

Figure 2: Exact connectivity, $k = 4$, $r = 2$. Markers are measured ratios of block counts, dashed lines the closed forms of the two generic families. Only the symmetric generic family diverges.

Proof sketch. A common envelope of s positions lets each of the k actions use at most s blocks, so it is feasible for total budget ks and gives the left inequality. For the right inequality, take any total-budget-s solution. The union of all active positions across actions has size at most s and is a valid common envelope. Taking infima gives the result. Appendix C records the full argument. □

A low common-envelope error therefore guarantees a low total-budget error, while a family that resists a total budget of ks must be far from every common envelope of size s.

Corollary 1 (Generic common-envelope incompatibility)**.** *Let $s_m = \lfloor \rho m^2 \rfloor$ with $\rho < 1 - 1/k$. For the isotropic random family in Theorem 2, there are $c, \gamma > 0$ such that, for all sufficiently large m,*

$$\Pr[\Gamma_{s_m}(\mathcal{A}) \geq c] \geq 1 - e^{-\gamma m^2}. \tag{20}$$

Reason. Since $ks_m \leq \lfloor \rho k m^2 \rfloor$, budget monotonicity and Theorem 2 give $\mathcal{E}_{ks_m} \geq c$ with the stated probability. Theorem 4 then gives $\Gamma_{s_m} \geq c$. □

Generic families are therefore far from every low-density common envelope, while families near one such envelope admit accurate low-budget shared representations. Structural compatibility is the condition connecting individual locality to locality of a persistent shared state.

Latent gauge. A learned latent state is defined only up to an invertible reparameterization, so before measuring any of these quantities on learned operators we fix the gauge. We estimate the latent covariance on a held-out calibration split and apply exact whitening. In the whitened coordinates any remaining $GL(d)$ reparameterization reduces to a common orthogonal transform, and C_{sep}, C_{sh}, $\mathcal{E}_B$ and Γ_s are all invariant under common orthogonal conjugation. The measured shared-coordinate cost then reflects the learned dynamics itself, independently of the chosen latent chart. Appendix E.1 states the transport identity and the numerical checks.

6 Experiments

The experiments follow the order of the theory: the connectivity cost of shared coordinates in the exact and then the approximate setting, followed by common sparse structure and its approximate form. Synthetic experiments use the explicit constructions of the theory, and learned-model analyses use the one-step residual Jacobian $J - I$. Two scores recur below. Under a per-action budget b, E_b^{sep} is the discarded energy when every action keeps its own basis and E_b^{sh} the same quantity under one shared basis, so their difference $F_b = E_b^{\mathrm{sh}} - E_b^{\mathrm{sep}}$ prices the shared coordinate system; the subscript is dropped when the budget is fixed. Appendix E gives full protocols.

6.1 Connectivity Scaling

We first test Theorem 1. Fixing $k = 4$, $r = 2$ and increasing m, we compare a commuting family sharing an eigenbasis, a compatible family with a common sparse ring envelope, and a generic family with independently oriented actions. The shared/separate ratio stays at 1 for the commuting family and exactly 3 for the ring family, while for the generic family it grows from 3.25 at $m = 4$ to 18.25 at $m = 24$, matching the exact formula. At $m = 16$, for instance, that family occupies 64 blocks under action-specific bases and 784 under one shared basis. When the preferred coordinates of different actions are incompatible, a shared system costs quadratically more connections.

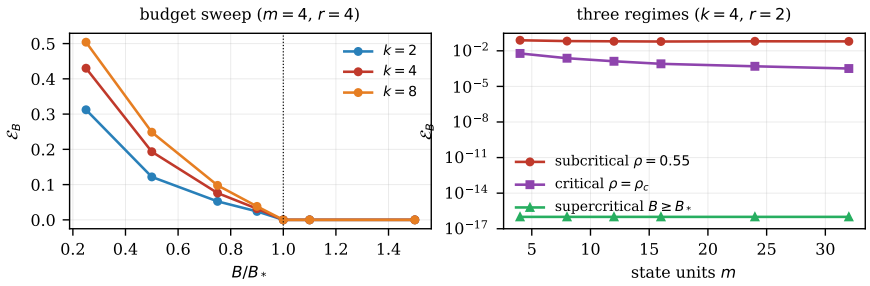

Figure 3: Finite budgets. Left, omitted energy against B/B_*, reaching zero at B_*. Right, the three regimes as m grows: subcritical $\rho = 0.55$, critical $\rho = \rho_c$, supercritical $B \geq B_*$.

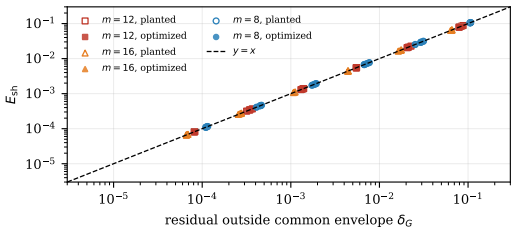

Figure 4: Approximate common structure, $k = 4$, $r = 4$. The axis is the measured energy δ_G leaving the common envelope. The planted basis lands on the diagonal and the optimized basis on or below it, at all 84 settings.

The same grid for general real matrices behaves differently. There the action-specific connectivity is already quadratic and the predicted ratio approaches $2 - 1/k$: with the same $k = 4$ the measured ratio grows slowly to 1.69 at $m = 24$, matching the closed form to 0.2%. Since the Jacobians learned by the flat world model are predominantly nonsymmetric, the shared-coordinate cost in learned models is read through this regime rather than the symmetric scaling law.

6.2 Obstruction under a Finite Budget

Exact support is strict, whereas practical representations discard weak connections, which is what $\mathcal{E}_B$ measures. We sample Frobenius-isotropic Gaussian symmetric families over $k \in \{2, 4, 8\}$, $r \in \{2, 4\}$ and several m, and express the budget as a fraction of $B_* = m + (k-1)m^2$, keeping the highest-energy blocks it allows.

The error decreases continuously with the budget and reaches exactly zero at $B = B_*$, staying zero beyond: all 300 instances at or above B_* have $\mathcal{E}_B < 10^{-12}$. Well below the threshold it remains clearly nonzero: with $k = 4$, $r = 2$ at density $\rho = 0.55$ it stays near 0.065 across an eightfold increase in m, from 4 to 32. At $\rho = \rho_c$ the error is positive at every finite m but decays steadily, from 6.0×10^{-3} at $m = 4$ to 3.2×10^{-4} at $m = 32$. Subcritical budgets retain a dimension-independent cost, the critical density marks a boundary case, and B_* suffices for exact representation. The finite-size correction $\rho_*(m, k) - \rho_c = 1/(km)$ and the full parameter sweep are reported in Appendix E.2.

6.3 Shared Structure

We next test shared structure, constructing families with a known common sparse basis and perturbing them away from full compatibility, with $m \in \{8, 12, 16\}$, $k = 4$, $r = 4$ and a budget large enough for the planted envelope. Instead of treating the perturbation magnitude as a measure of incompatibility, we measure the fraction δ_G of operator energy that leaves the common envelope.

At the fully compatible endpoint $\delta_G = 0$ and the shared error is numerically zero. After perturbation the planted basis gives an error equal to the residual energy, matching it to 2×10^{-16} across all

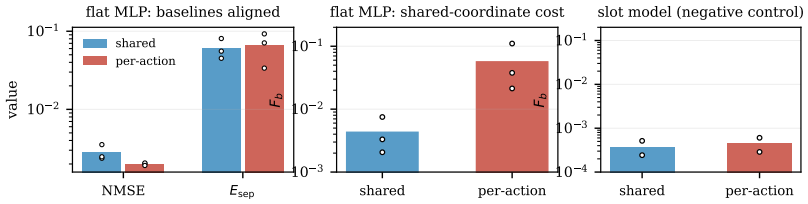

Figure 5: Learned models on the Neural Relational Inference spring system, $m = 16$, $k = 4$, interaction density 0.15, per-action budget $b = 0.5m$. Prediction error and E_{sep} are aligned across the two graph conditions while F_b separates; the object-wise slot model keeps F_b near zero in both. Circles are independently trained seeds.

84 settings, and optimizing the shared basis stays at or below δ_G. This illustrates Proposition 1 and Theorem 3: common sparse structure is graded, and if a fraction δ_G of the dynamics leaves the envelope then some shared representation has error at most δ_G. The ring family gives the exact-support counterpart, where shared connectivity stays linear.

6.4 Learned World Models

Finally we test the same mechanism in learned nonlinear world models, using the spring system of the Neural Relational Inference (NRI) benchmark (Kipf et al., 2018), with $m = 16$ particles and $k = 4$ actions and varying only the ground-truth interaction support. In the shared support condition all four actions use the same edge locations with independent strengths; in the per-action support condition each action samples its own. Both use the same model class, data scale and evaluation protocol. At held-out states we compute the one-step residual Jacobian $J - I$ and compare action-specific and shared bases under the same budget. Jacobian-based local linearization has been used to probe nonlinear recurrent dynamics (Sussillo & Barak, 2013; Smith et al., 2021); the analysis here is state-local, and Appendix E.4 reports the matching comparison with one basis fixed across states. We also report prediction error and E_{sep} to check that the two conditions are not separated by fit quality or by the compressibility of individual actions alone.

For the flat MLP the two graph conditions have prediction errors of the same order and very similar E_{sep}. With those baselines aligned, the shared support condition gives $F_b \approx 0.004$ against $F_b \approx 0.056$ for per-action support, about an order of magnitude, with nonoverlapping ranges across the three seeds of the longer training schedule, 0.002 to 0.008 against 0.021 to 0.110. As a negative control, the slot model already organizes its representation by objects, and there the penalty is close to zero under both graph conditions. This supports the structural variable of Proposition 1: with single-action difficulty and prediction quality largely unchanged, reusing the same interaction locations across actions substantially reduces the cost of shared coordinates. The symmetric component of these Jacobians carries about 29% of their energy, so the learned results are read through the general real regime of Eq. (10).

7 Conclusion

This paper compares action-wise locality with locality of a persistent representation shared across actions. In the symmetric reference model generic families move from linear separate connectivity to quadratic shared connectivity and a subcritical budget retains non-vanishing approximation error, while for general real operators the same mechanism leaves a bounded relative cost. A common sparse envelope restores linear shared connectivity, approximate common structure controls low-budget error by the residual energy leaving the envelope, and Γ_s places both regimes on one axis. Experiments reproduce all three budget regimes, and in learned world models changing only whether actions reuse the same interaction locations moves the penalty by about an order of magnitude, a factor of 13 in the aggregate, at matched prediction quality. Cross-action structural compatibility is therefore the condition linking local action dynamics to a local shared world state.

AI USE STATEMENT

AI tools were used solely to assist with language polishing and code organization. We developed the research content and wrote the manuscript, and we take full responsibility for its content and accuracy.

REPRODUCIBILITY STATEMENT

All theoretical claims are stated with their assumptions in Sections 4 and 5 and proved in Appendices A–D, including the dimension count behind the exact block equality, every constant of the covering argument in closed form, and the critical density. The experimental protocol, scoring definitions, gauge fixing and replication unit are given in Appendix E.1, the synthetic grids in Appendices E.2 and E.3, and the environment, architectures, optimizer and per-seed results in Appendix E.4. Synthetic results use explicit constructions wherever the theory supplies one, and search-based scores are calibrated against an instance with a closed-form solution. Code, seeds and configurations are released with the paper.

REFERENCES

Karim Abed-Meraim and Adel Belouchrani. Algorithms for joint block diagonalization. In *2004 12th European Signal Processing Conference, Vienna, Austria, September 6-10, 2004*, pp. 209–212. IEEE, 2004. URL https://ieeexplore.ieee.org/document/7080203/.

Bijan Afsari. Sensitivity analysis for the problem of matrix joint diagonalization. *SIAM J. Matrix Anal. Appl.*, 30(3):1148–1171, 2008. doi: 10.1137/060655997. URL https://doi.org/10.1137/060655997.

Peter W. Battaglia, Razvan Pascanu, Matthew Lai, Danilo Jimenez Rezende, and Koray Kavukcuoglu. Interaction networks for learning about objects, relations and physics. In Daniel D. Lee, Masashi Sugiyama, Ulrike von Luxburg, Isabelle Guyon, and Roman Garnett (eds.), *Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016, December 5-10, 2016, Barcelona, Spain*, pp. 4502–4510, 2016. URL https://proceedings.neurips.cc/paper/2016/hash/3147da8ab4a0437c15ef51a5cc7f2dc4-Abstract.html.

Florent Bouchard, Bijan Afsari, Jérôme Malick, and Marco Congedo. Approximate joint diagonalization with riemannian optimization on the general linear group. *SIAM J. Matrix Anal. Appl.*, 41(1):152–170, 2020. doi: 10.1137/18M1232838. URL https://doi.org/10.1137/18M1232838.

Angelika Bunse-Gerstner, Ralph Byers, and Volker Mehrmann. Numerical methods for simultaneous diagonalization. *SIAM J. Matrix Anal. Appl.*, 14(4):927–949, 1993. doi: 10.1137/0614062. URL https://doi.org/10.1137/0614062.

Yunfeng Cai and Ren-Cang Li. Perturbation analysis for matrix joint block diagonalization. *Linear Algebra and its Applications*, 581:163–197, November 2019. ISSN 0024-3795. doi: 10.1016/j.laa.2019.07.012. URL https://doi.org/10.1016/j.laa.2019.07.012.

Yunfeng Cai and Chengyu Liu. An algebraic approach to nonorthogonal general joint block diagonalization. *SIAM J. Matrix Anal. Appl.*, 38(1):50–71, 2017. doi: 10.1137/16M1080756. URL https://doi.org/10.1137/16M1080756.

Yunfeng Cai, Decai Shi, and Shufang Xu. A matrix polynomial spectral approach for general joint block diagonalization. *SIAM J. Matrix Anal. Appl.*, 36(2):839–863, 2015. doi: 10.1137/130935264. URL https://doi.org/10.1137/130935264.

Jean-François Cardoso and Antoine Souloumiac. Jacobi angles for simultaneous diagonalization. *SIAM J. Matrix Anal. Appl.*, 17(1):161–164, 1996. doi: 10.1137/S0895479893259546. URL https://doi.org/10.1137/S0895479893259546.

Gilles Chabriel, Martin Kleinsteuber, Eric Moreau, Hao Shen, Petr Tichavský, and Arie Yeredor. Joint matrices decompositions and blind source separation: A survey of methods, identification, and applications. *IEEE Signal Process. Mag.*, 31(3):34–43, 2014. doi: 10.1109/MSP.2014.2298045. URL https://doi.org/10.1109/MSP.2014.2298045.

Bart De Schutter. Minimal state-space realization in linear system theory: an overview. *Journal of Computational and Applied Mathematics*, 121(1-2):331–354, September 2000. ISSN 0377-0427. doi: 10.1016/S0377-0427(00)00341-1. URL https://doi.org/10.1016/S0377-0427(00)00341-1.

Alan Edelman, Tomás A. Arias, and Steven Thomas Smith. The geometry of algorithms with orthogonality constraints. *SIAM J. Matrix Anal. Appl.*, 20(2):303–353, 1998. doi: 10.1137/S0895479895290954. URL https://doi.org/10.1137/S0895479895290954.

Davide Eynard, Artiom Kovnatsky, Michael M. Bronstein, Klaus Glashoff, and Alexander M. Bronstein. Multimodal manifold analysis by simultaneous diagonalization of laplacians. *IEEE Trans. Pattern Anal. Mach. Intell.*, 37(12):2505–2517, 2015. doi: 10.1109/TPAMI.2015.2408348. URL https://doi.org/10.1109/TPAMI.2015.2408348.

Chelsea Finn, Ian J. Goodfellow, and Sergey Levine. Unsupervised learning for physical interaction through video prediction. In Daniel D. Lee, Masashi Sugiyama, Ulrike von Luxburg, Isabelle Guyon, and Roman Garnett (eds.), *Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016, December 5-10, 2016, Barcelona, Spain*, pp. 64–72, 2016. URL https://proceedings.neurips.cc/paper/2016/hash/d9d4f495e875a2e075a1a4a6e1b9770f-Abstract.html.

Thomas Frerix and Joan Bruna. Approximating orthogonal matrices with effective givens factorization. In Kamalika Chaudhuri and Ruslan Salakhutdinov (eds.), *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019, Long Beach, California, USA*, volume 97 of *Proceedings of Machine Learning Research*, pp. 1993–2001. PMLR, 2019. URL http://proceedings.mlr.press/v97/frerix19a.html.

Klaus Greff, Raphaël Lopez Kaufman, Rishabh Kabra, Nick Watters, Chris Burgess, Daniel Zoran, Loic Matthey, Matthew M. Botvinick, and Alexander Lerchner. Multi-object representation learning with iterative variational inference. In Kamalika Chaudhuri and Ruslan Salakhutdinov (eds.), *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019, Long Beach, California, USA*, volume 97 of *Proceedings of Machine Learning Research*, pp. 2424–2433. PMLR, 2019. URL http://proceedings.mlr.press/v97/greff19a.html.

David Ha and Jürgen Schmidhuber. Recurrent world models facilitate policy evolution. In Samy Bengio, Hanna M. Wallach, Hugo Larochelle, Kristen Grauman, Nicolò Cesa-Bianchi, and Roman Garnett (eds.), *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018, December 3-8, 2018, Montréal, Canada*, pp. 2455–2467, 2018. URL https://proceedings.neurips.cc/paper/2018/hash/2de5d16682c3c35007e4e92982f1a2ba-Abstract.html.

Danijar Hafner, Timothy P. Lillicrap, Ian Fischer, Ruben Villegas, David Ha, Honglak Lee, and James Davidson. Learning latent dynamics for planning from pixels. In Kamalika Chaudhuri and Ruslan Salakhutdinov (eds.), *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019, Long Beach, California, USA*, volume 97 of *Proceedings of Machine Learning Research*, pp. 2555–2565. PMLR, 2019. URL http://proceedings.mlr.press/v97/hafner19a.html.

Danijar Hafner, Timothy P. Lillicrap, Jimmy Ba, and Mohammad Norouzi. Dream to control: Learning behaviors by latent imagination. In *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*. OpenReview.net, 2020. URL https://openreview.net/forum?id=S1lOTC4tDS.

Danijar Hafner, Timothy P. Lillicrap, Mohammad Norouzi, and Jimmy Ba. Mastering atari with discrete world models. In *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*. OpenReview.net, 2021. URL https://openreview.net/forum?id=0oabwyZbOu.

Danijar Hafner, Jurgis Pasukonis, Jimmy Ba, and Timothy P. Lillicrap. Mastering diverse control tasks through world models. *Nat.*, 640(8059):647–653, 2025. doi: 10.1038/S41586-025-08744-2. URL https://doi.org/10.1038/s41586-025-08744-2.

Nicklas Hansen, Hao Su, and Xiaolong Wang. TD-MPC2: scalable, robust world models for continuous control. In *The Twelfth International Conference on Learning Representations, ICLR 2024, Vienna, Austria, May 7-11, 2024*. OpenReview.net, 2024. URL https://openreview.net/forum?id=Oxh5CstDJU.

Haoze He and Daniel Kressner. Randomized joint diagonalization of symmetric matrices. *SIAM J. Matrix Anal. Appl.*, 45(1):661–684, 2024. doi: 10.1137/22M1541265. URL https://doi.org/10.1137/22m1541265.

Jindong Jiang, Sepehr Janghorbani, Gerard de Melo, and Sungjin Ahn. SCALOR: generative world models with scalable object representations. In *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*. OpenReview.net, 2020. URL https://openreview.net/forum?id=SJxrKgStDH.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In Yoshua Bengio and Yann LeCun (eds.), *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015. URL http://arxiv.org/abs/1412.6980.

Thomas Kipf, Gamaleldin Fathy Elsayed, Aravindh Mahendran, Austin Stone, Sara Sabour, Georg Heigold, Rico Jonschkowski, Alexey Dosovitskiy, and Klaus Greff. Conditional object-centric learning from video. In *The Tenth International Conference on Learning Representations, ICLR 2022, Virtual Event, April 25-29, 2022*. OpenReview.net, 2022. URL https://openreview.net/forum?id=aD7uesX1GF_.

Thomas N. Kipf, Ethan Fetaya, Kuan-Chieh Wang, Max Welling, and Richard S. Zemel. Neural relational inference for interacting systems. In Jennifer G. Dy and Andreas Krause (eds.), *Proceedings of the 35th International Conference on Machine Learning, ICML 2018, Stockholmsmässan, Stockholm, Sweden, July 10-15, 2018*, volume 80 of *Proceedings of Machine Learning Research*, pp. 2693–2702. PMLR, 2018. URL http://proceedings.mlr.press/v80/kipf18a.html.

Thomas N. Kipf, Elise van der Pol, and Max Welling. Contrastive learning of structured world models. In *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*. OpenReview.net, 2020. URL https://openreview.net/forum?id=H1gax6VtDB.

Jannik Kossen, Karl Stelzner, Marcel Hussing, Claas Voelcker, and Kristian Kersting. Structured object-aware physics prediction for video modeling and planning. In *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*. OpenReview.net, 2020. URL https://openreview.net/forum?id=B1e-kxSKDH.

Francesco Locatello, Dirk Weissenborn, Thomas Unterthiner, Aravindh Mahendran, Georg Heigold, Jakob Uszkoreit, Alexey Dosovitskiy, and Thomas Kipf. Object-centric learning with slot attention. In Hugo Larochelle, Marc'Aurelio Ranzato, Raia Hadsell, Maria-Florina Balcan, and Hsuan-Tien Lin (eds.), *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*, 2020. URL https://proceedings.neurips.cc/paper/2020/hash/8511df98c02ab60aea1b2356c013bc0f-Abstract.html.

Takanori Maehara and Kazuo Murota. Algorithm for error-controlled simultaneous block-diagonalization of matrices. *SIAM J. Matrix Anal. Appl.*, 32(2):605–620, 2011. doi: 10.1137/090779966. URL https://doi.org/10.1137/090779966.

Amirhossein Nazerian, Kshitij Bhatta, and Francesco Sorrentino. Exact decomposition of optimal control problems via simultaneous block diagonalization of matrices. *IEEE Open Journal of Control Systems*, 2:24–35, 2023. ISSN 2694-085X. doi: 10.1109/OJCSYS.2022.3231553. URL https://doi.org/10.1109/OJCSYS.2022.3231553.

Tobias Pfaff, Meire Fortunato, Alvaro Sanchez-Gonzalez, and Peter W. Battaglia. Learning mesh-based simulation with graph networks. In *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*. OpenReview.net, 2021. URL https://openreview.net/forum?id=roNqYL0_XP.

Dinh Tuan Pham. Joint approximate diagonalization of positive definite hermitian matrices. *SIAM J. Matrix Anal. Appl.*, 22(4):1136–1152, 2001. doi: 10.1137/S089547980035689X. URL https://doi.org/10.1137/S089547980035689X.

Alvaro Sanchez-Gonzalez, Nicolas Heess, Jost Tobias Springenberg, Josh Merel, Martin A. Riedmiller, Raia Hadsell, and Peter W. Battaglia. Graph networks as learnable physics engines for inference and control. In Jennifer G. Dy and Andreas Krause (eds.), *Proceedings of the 35th International Conference on Machine Learning, ICML 2018, Stockholmsmässan, Stockholm, Sweden, July 10-15, 2018*, volume 80 of *Proceedings of Machine Learning Research*, pp. 4467–4476. PMLR, 2018. URL http://proceedings.mlr.press/v80/sanchez-gonzalez18a.html.

Alvaro Sanchez-Gonzalez, Jonathan Godwin, Tobias Pfaff, Rex Ying, Jure Leskovec, and Peter W. Battaglia. Learning to simulate complex physics with graph networks. In *Proceedings of the 37th International Conference on Machine Learning, ICML 2020, 13-18 July 2020, Virtual Event*, volume 119 of *Proceedings of Machine Learning Research*, pp. 8459–8468. PMLR, 2020. URL http://proceedings.mlr.press/v119/sanchez-gonzalez20a.html.

Uri Shalit and Gal Chechik. Coordinate-descent for learning orthogonal matrices through givens rotations. In *Proceedings of the 31th International Conference on Machine Learning, ICML 2014, Beijing, China, 21-26 June 2014*, volume 32 of *JMLR Workshop and Conference Proceedings*, pp. 548–556. JMLR.org, 2014. URL http://proceedings.mlr.press/v32/shalit14.html.

Gautam Singh, Skand Vishwanath Peri, Junghyun Kim, Hyunseok Kim, and Sungjin Ahn. Structured world belief for reinforcement learning in POMDP. In Marina Meila and Tong Zhang (eds.), *Proceedings of the 38th International Conference on Machine Learning, ICML 2021, 18-24 July 2021, Virtual Event*, volume 139 of *Proceedings of Machine Learning Research*, pp. 9744–9755. PMLR, 2021. URL http://proceedings.mlr.press/v139/singh21a.html.

Jimmy T. H. Smith, Scott W. Linderman, and David Sussillo. Reverse engineering recurrent neural networks with jacobian switching linear dynamical systems. In Marc'Aurelio Ranzato, Alina Beygelzimer, Yann N. Dauphin, Percy Liang, and Jennifer Wortman Vaughan (eds.), *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6-14, 2021, virtual*, pp. 16700–16713, 2021. URL https://proceedings.neurips.cc/paper/2021/hash/8b77b4b5156dc11dec152c6c71481565-Abstract.html.

David Sussillo and Omri Barak. Opening the black box: Low-dimensional dynamics in high-dimensional recurrent neural networks. *Neural Comput.*, 25(3):626–649, 2013. doi: 10.1162/NECO_A_00409. URL https://doi.org/10.1162/NECO_a_00409.

Stanisław J. Szarek. Metric entropy of homogeneous spaces. *Banach Center Publications*, 43(1): 395–410, 1998. ISSN 1730-6299. doi: 10.4064/-43-1-395-410. URL https://doi.org/10.4064/-43-1-395-410.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, pp. 5998–6008, 2017. URL https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html.

Rishi Veerapaneni, John D. Co-Reyes, Michael Chang, Michael Janner, Chelsea Finn, Jiajun Wu, Joshua B. Tenenbaum, and Sergey Levine. Entity abstraction in visual model-based reinforcement learning. In Leslie Pack Kaelbling, Danica Kragic, and Komei Sugiura (eds.), *3rd Annual Conference on Robot Learning, CoRL 2019, Osaka, Japan, October 30 - November 1, 2019, Proceedings*,

volume 100 of *Proceedings of Machine Learning Research*, pp. 1439–1456. PMLR, 2019. URL http://proceedings.mlr.press/v100/veerapaneni20a.html.

Nicholas Watters, Daniel Zoran, Theophane Weber, Peter W. Battaglia, Razvan Pascanu, and Andrea Tacchetti. Visual interaction networks: Learning a physics simulator from video. In Isabelle Guyon, Ulrike von Luxburg, Samy Bengio, Hanna M. Wallach, Rob Fergus, S. V. N. Vishwanathan, and Roman Garnett (eds.), *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, pp. 4539–4547, 2017. URL https://proceedings.neurips.cc/paper/2017/hash/8cbd005a556ccd4211ce43f309bc0eac-Abstract.html.

Ziyi Wu, Nikita Dvornik, Klaus Greff, Thomas Kipf, and Animesh Garg. Slotformer: Unsupervised visual dynamics simulation with object-centric models. In *The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1-5, 2023*. OpenReview.net, 2023. URL https://openreview.net/forum?id=TFbwV6I0VLg.

Zhongwei Yu, Jingqing Ruan, and Dengpeng Xing. Learning causal dynamics models in object-oriented environments. In Ruslan Salakhutdinov, Zico Kolter, Katherine A. Heller, Adrian Weller, Nuria Oliver, Jonathan Scarlett, and Felix Berkenkamp (eds.), *Forty-first International Conference on Machine Learning, ICML 2024, Vienna, Austria, July 21-27, 2024*, volume 235 of *Proceedings of Machine Learning Research*, pp. 57597–57638. PMLR / OpenReview.net, 2024. URL https://proceedings.mlr.press/v235/yu24j.html.

A EXACT OBSTRUCTION

This appendix gives the full proof of Theorem 1. We first record the dimension of a fixed symmetric block support.

A.1 DIMENSION OF A FIXED SUPPORT

Consider one symmetric $d \times d$ matrix with $d = mr$. Suppose its block support contains q diagonal blocks and p symmetric off-diagonal block pairs. The directed active-block count is $b = q + 2p$.

Lemma 1 (Fixed-support dimension)**.** *The linear space of symmetric matrices supported on this pattern has dimension*

$$q\frac{r(r+1)}{2} + pr^2 = \frac{br^2 + qr}{2}. \tag{21}$$

Proof. A diagonal $r \times r$ block is symmetric, so it has $r(r+1)/2$ free entries. An off-diagonal pair is determined by one arbitrary $r \times r$ block B, with the transposed block $B^\top$ forced on the opposite side, so the pair has r^2 free entries. Summing yields the first expression. Substituting $p = (b-q)/2$ gives the second. □

For a joint support pattern $\mathcal{S}$ across k actions, let B be the sum of their directed active-block counts. Let $L_{\mathcal{S}} \subset \mathrm{Sym}_d(\mathbb{R})^k$ be the corresponding product support space.

Corollary 2 (Joint-support parameter bound)**.**

$$\dim L_{\mathcal{S}} \le \frac{r^2 B + rkm}{2}. \tag{22}$$

Proof. Apply Lemma 1 to each action. The total number of active diagonal blocks across all actions is at most km, hence

$$\dim L_{\mathcal{S}} = \sum_a \frac{b_a r^2 + q_a r}{2} \le \frac{r^2 B + rkm}{2}. \tag{23}$$

□

A.2 GENERIC LOWER BOUND AND DETERMINISTIC UPPER BOUND

The ambient space $\mathrm{Sym}_d(\mathbb{R})^k$ has dimension

$$N_m = k\frac{d(d+1)}{2}, \tag{24}$$

and the orthogonal group has dimension $d(d-1)/2$.

Full proof of Theorem 1. We prove the separate cost first. Every symmetric A_a has an orthogonal eigenbasis. In that basis it is block diagonal with at most m active diagonal blocks, so its minimum block count is at most m. A full-rank block matrix must have at least one active block in each of its m block rows. Therefore any full-rank matrix has block count at least m. A generic symmetric matrix is full rank because the zero-determinant set is an algebraic hypersurface of measure zero. Thus every action in a generic family has individual optimum m, giving $C_{\mathrm{sep}} = km$.

For the shared lower bound, fix a joint support pattern $\mathcal{S}$ with total directed budget B. Define the smooth map

$$\Phi_{\mathcal{S}} : L_{\mathcal{S}} \times O(d) \to \mathrm{Sym}_d(\mathbb{R})^k, \tag{25}$$

$$\Phi_{\mathcal{S}}((M_a)_{a=1}^k, Q) = (QM_1Q^\top, \dots, QM_kQ^\top). \tag{26}$$

Its domain dimension is at most

$$t_B = \frac{r^2 B + rkm}{2} + \frac{d(d-1)}{2}. \tag{27}$$

Using $d = mr$ and Corollary 2,

$$N_m - t_B \geq \frac{r^2((k-1)m^2 - B) + rm}{2}. \tag{28}$$

If $B \leq (k-1)m^2$, the right-hand side is strictly positive. Hence the differential of $\Phi_{\mathcal{S}}$ has rank strictly below the ambient dimension everywhere. Sard's theorem implies that the image of $\Phi_{\mathcal{S}}$ has Lebesgue measure zero in $\mathrm{Sym}_d(\mathbb{R})^k$. There are finitely many symmetric block-support patterns for fixed m and k, so the union of all images with $B \leq (k-1)m^2$ also has measure zero. Every family outside this exceptional union therefore satisfies

$$C_{\mathrm{sh}} \geq (k-1)m^2 + 1. \tag{29}$$

For the deterministic upper bound, choose an eigenbasis of A_1 as the shared basis. The first action then uses at most m blocks. Every other transformed action uses at most all m^2 directed blocks. Consequently

$$C_{\mathrm{sh}} \leq m + (k-1)m^2. \tag{30}$$

Combining the separate equality and the two shared bounds proves the theorem.

The dimension calculation also explains why the threshold depends on $k-1$. One shared orthogonal basis can absorb the orientation of roughly one action. Each additional generically oriented action contributes another quadratic-scale collection of relations that must be expressed in the same coordinates.

A.3 Redundancy within a block and the exact shared count

The bound of Theorem 1 is an equality because the orthogonal group contains directions that do not move the block support at all. Let $\mathcal{S}$ be a joint block support and let $L_{\mathcal{S}}$ be the space of symmetric families supported on $\mathcal{S}$. If $\mathcal{S}$ has n_{d} diagonal and n_{p} off-diagonal block pairs summed over actions, then

$$\dim L_{\mathcal{S}} = n_{\mathrm{d}} \frac{r(r+1)}{2} + n_{\mathrm{p}} r^2. \tag{31}$$

Conjugation by $Q \in O(d)$ generates at most $\dim O(d) = d(d-1)/2$ further directions, but the block-diagonal subgroup $O(r)^m$ maps every block to itself, adding no new direction to the support pattern.

Lemma 2 (Inert directions). *Let $R \in O(r)^m$ act block-diagonally. Then $R^\top \cdot$ preserves $L_{\mathcal{S}}$ for every block support $\mathcal{S}$, so the orbit map $O(d) \times L_{\mathcal{S}} \to \mathrm{Sym}_d^k$ has differential of rank at most*

$$\dim L_{\mathcal{S}} + \frac{d(d-1)}{2} - m\frac{r(r-1)}{2}. \tag{32}$$

Proof. Block (i,j) of $R^\top M R$ equals $R_i^\top M_{ij} R_j$, which vanishes exactly when M_{ij} does. Hence the $O(r)^m$ directions lie in the kernel of the orbit differential modulo $L_{\mathcal{S}}$, and the rank drops by $\dim O(r)^m = m\, r(r-1)/2$. □

Comparing this rank with the ambient dimension $k\, d(d+1)/2$ and applying Sard's theorem to the orbit map gives a measure-zero image whenever the rank is strictly smaller. Carrying out the count for a total budget B shows that the image is measure zero for every $B < (k-1)m^2 + m$, while the eigenbasis of one action realizes $B = (k-1)m^2 + m$. Both bounds therefore meet, which removes the $m-1$ slack of the earlier two-sided statement.

A.4 General real families

For $\mathcal{A} \in (\mathbb{R}^{d\times d})^k$ an orthogonal change of coordinates can reach real Schur form but not diagonal form. A real Schur form is block upper triangular at the scalar level, so at block granularity it occupies the $m(m+1)/2$ positions on and above the diagonal, and this is generically exact when the block width r is even: the 2×2 blocks carrying complex eigenvalue pairs can then be aligned so that each lies inside a single $r \times r$ block. Repeating the dimension count of Lemma 2 with $\dim L_{\mathcal{S}} = n\, r^2$ and ambient dimension kd^2 gives Eq. (10). The ratio is

$$\frac{C_{\mathrm{sh}}}{C_{\mathrm{sep}}} = \frac{(k-1)m^2 + m(m+1)/2}{k\, m(m+1)/2} \xrightarrow[m\to\infty]{} 2 - \frac{1}{k}, \tag{33}$$

bounded for every k, in contrast with the symmetric ratio $((k-1)m+1)/k$, which diverges. For odd r a complex pair can be forced to straddle two block boundaries, so the separate count can exceed $k\,m(m+1)/2$ by a term of order m; the regime conclusion is unaffected, but the exact equality should be stated for even r only.

B Proof of the Finite-Budget Obstruction

This appendix proves Theorem 2. We work in the Euclidean space $\mathcal{H} = \mathrm{Sym}_d(\mathbb{R})^k$ with the product Frobenius norm $\|\mathcal{A}\|_F^2 = \sum_a \|A_a\|_F^2$ and dimension $N_m = kd(d+1)/2$.

B.1 The exact set of low budget

For a fixed joint support pattern $\mathcal{S}$, define

$$\mathcal{C}_{\mathcal{S}} = \{(QM_1Q^\top, \dots, QM_kQ^\top) : Q \in O(d), (M_a)_a \in L_{\mathcal{S}}\}. \tag{34}$$

Writing $b(\mathcal{S})$ for the total directed block count of a joint support pattern, the set of families exactly representable with total budget B is

$$\mathcal{M}_B = \bigcup_{\mathcal{S}: b(\mathcal{S}) \le B} \mathcal{C}_{\mathcal{S}}. \tag{35}$$

Lemma 3 (Closed-cone property)**.** *For fixed m, k, r, B, the set $\mathcal{M}_B$ is a closed cone in $\mathcal{H}$.*

Proof. Fix $\mathcal{S}$. Closure under real scaling is immediate. Let $\mathcal{X}_n = (Q_nM_{1,n}Q_n^\top, \dots, Q_nM_{k,n}Q_n^\top) \in \mathcal{C}_{\mathcal{S}}$ converge to $\mathcal{X}$. Compactness of $O(d)$ gives a subsequence $Q_n \to Q$. Orthogonal invariance yields $\|(M_{a,n})_a\|_F = \|\mathcal{X}_n\|_F$, so the support-space coefficients are bounded. In fact

$$(M_{a,n})_a = (Q_n^\top X_{a,n} Q_n)_a \to (Q^\top X_a Q)_a =: \mathcal{M}. \tag{36}$$

The finite-dimensional linear space $L_{\mathcal{S}}$ is closed, hence $\mathcal{M} \in L_{\mathcal{S}}$ and $\mathcal{X} \in \mathcal{C}_{\mathcal{S}}$. Thus every $\mathcal{C}_{\mathcal{S}}$ is closed. Only finitely many support patterns exist, so their finite union $\mathcal{M}_B$ is closed. The cone property passes to the union. □

Lemma 4 (Distance form)**.** *For every nonzero family $\mathcal{A}$, with* dist *denoting distance in the product Frobenius norm,*

$$\mathcal{E}_B(\mathcal{A}) = \frac{\mathrm{dist}(\mathcal{A}, \mathcal{M}_B)^2}{\|\mathcal{A}\|_F^2}. \tag{37}$$

Proof. For a feasible pair $(Q, (\widetilde{A}_a)_a)$ in the definition of $\mathcal{E}_B$, set $X_a = Q\widetilde{A}_aQ^\top$. Then $(X_a)_a \in \mathcal{M}_B$ and $\|A_a - X_a\|_F = \|Q^\top A_aQ - \widetilde{A}_a\|_F$. Conversely every element of $\mathcal{M}_B$ has such a representation. Taking the infimum gives the distance formula. □

B.2 Covering ingredients

We use three elementary covering facts. Constants below are uniform in m and may depend on fixed k and r.

Lemma 5 (Euclidean sphere net)**.** *The unit sphere in an s-dimensional Euclidean space has an α-net of size at most $(3/\alpha)^s$ for $0 < \alpha < 1$.*

Proof. Take a maximal α-separated set. The radius-$\alpha/2$ balls around its points are disjoint and lie in the ball of radius $1 + \alpha/2$. Comparing s-dimensional volumes gives at most $(1 + 2/\alpha)^s \le (3/\alpha)^s$ points. Maximality makes the set an α-net. □

Lemma 6 (Orthogonal-group net)**.** *There is a universal constant C_O such that $O(d)$ has an operator-norm η-net of size at most*

$$(C_O/\eta)^{d(d-1)/2} \tag{38}$$

for $0 < \eta < 1$, where distance is measured by the operator norm $\|\cdot\|_{\mathrm{op}}$.

The statement is a standard metric-entropy bound for orthogonal groups under operator-norm-type metrics; see, e.g., Szarek (1998). It can be obtained by covering local exponential-coordinate charts of dimension $d(d-1)/2$ and then using finitely many charts. We record it explicitly because its exponent is the part used below.

Lemma 7 (Conjugation is Lipschitz)**.** *For $Q, R \in O(d)$ and any matrix M,*

$$\|Q^\top MQ - R^\top MR\|_F \le 2\|Q-R\|_{\mathrm{op}}\|M\|_F. \tag{39}$$

Proof. Insert and subtract $Q^\top MR$:

$$Q^\top MQ - R^\top MR = Q^\top M(Q-R) + (Q-R)^\top MR. \tag{40}$$

Apply $\|AXB\|_F \le \|A\|_{\mathrm{op}}\|X\|_F\|B\|_{\mathrm{op}}$ and $\|Q\|_{\mathrm{op}} = \|R\|_{\mathrm{op}} = 1$. □

Fix a support pattern $\mathcal{S}$ of product-space dimension $s_{\mathcal{S}} = \dim L_{\mathcal{S}}$ and let $\mathbb{S}^{N_m-1}$ denote the unit sphere in $\mathcal{H}$ and define

$$\mathcal{T}_{\mathcal{S}} = \mathcal{C}_{\mathcal{S}} \cap \mathbb{S}^{N_m-1}. \tag{41}$$

Combining Lemmas 5–7, first take an $\alpha/4$-net of the unit sphere of $L_{\mathcal{S}}$ and an $\alpha/8$ operator-norm net of $O(d)$. Their conjugated products form an α-net of $\mathcal{T}_{\mathcal{S}}$ of size at most the following bound on the covering number $N(\mathcal{T}_{\mathcal{S}}, \alpha)$,

$$N(\mathcal{T}_{\mathcal{S}}, \alpha) \le (C/\alpha)^{s_{\mathcal{S}}+d(d-1)/2} \le (C/\alpha)^{t_B}. \tag{42}$$

The number of joint directed support patterns of total budget at most B is at most

$$\sum_{j=0}^{B} \binom{km^2}{j} \le \exp\{C_1 m^2\} \tag{43}$$

for fixed budget density, after absorbing constants depending on k and the density.

B.3 Subcritical regime

Let U be uniform on the unit sphere $\mathbb{S}^{N_m-1}$. For a fixed unit vector v and $0 < \alpha < 1/4$, a spherical-cap estimate gives

$$\Pr(\|U - v\|_F \le 2\alpha) \le (C_2\alpha)^{N_m-1}. \tag{44}$$

Use an α-net for every normalized support image and union bound over the net points and support patterns. With t_B as above,

$$\Pr(\mathrm{dist}(U, \mathcal{M}_B) \le \alpha) \le \exp\{C_3 m^2\}(C_4/\alpha)^{t_B}(C_5\alpha)^{N_m-1}. \tag{45}$$

Equivalently,

$$\log \Pr(\mathrm{dist}(U, \mathcal{M}_B) \le \alpha) \le Am^2 + (N_m - 1 - t_B)\log\alpha \tag{46}$$

for a constant A.

Now set $B = B_m = \lfloor \rho k m^2 \rfloor$ and let

$$\delta = 1 - \frac{1}{k} - \rho > 0. \tag{47}$$

The same dimension algebra as in Appendix A yields

$$N_m - t_B \ge \frac{r^2((k-1)m^2 - B_m) + rm}{2} \ge \frac{r^2 k\delta}{2}m^2 + \frac{rm}{2}. \tag{48}$$

Hence $N_m - 1 - t_B \ge am^2$ for some $a > 0$ and all sufficiently large m. Choose a fixed α small enough that $A + a\log\alpha < -\gamma$ for some $\gamma > 0$. Then

$$\Pr(\mathrm{dist}(U, \mathcal{M}_{B_m}) \le \alpha) \le e^{-\gamma m^2}. \tag{49}$$

A Frobenius-isotropic Gaussian tuple $\mathcal{A}$ has a rotationally invariant Gaussian law in the Euclidean space $\mathcal{H}$. Its normalized direction $U = \mathcal{A}/\|\mathcal{A}\|_F$ is then uniform on the unit sphere and independent of the radius. Since $\mathcal{M}_B$ is a cone,

$$\frac{\mathrm{dist}(\mathcal{A}, \mathcal{M}_B)}{\|\mathcal{A}\|_F} = \mathrm{dist}(U, \mathcal{M}_B). \tag{50}$$

Lemma 4 then gives $\mathcal{E}_{B_m}(\mathcal{A}) \ge \alpha^2$ except on an event of probability at most $e^{-\gamma m^2}$. Taking $c = \alpha^2$ proves the claim.

B.4 SUPERCRITICAL REGIME

The eigenbasis construction from Theorem 1 represents every symmetric family exactly with

$$B_{\text{eig}} = m + (k-1)m^2 \tag{51}$$

blocks. If $\rho > 1 - 1/k$, then $\rho k - (k-1) > 0$. Hence

$$B_m - B_{\text{eig}} \geq [\rho k - (k-1)]m^2 - m - 1 > 0 \tag{52}$$

for all sufficiently large m. The exact eigenbasis representation is then feasible for budget B_m, and $\mathcal{E}_{B_m} = 0$.

B.5 EXPLICIT COVERING CONSTANTS

The covering argument above uses a metric-entropy bound for $O(d)$. We record a self-contained version so that no constant is left implicit.

Lemma 8 (Orthogonal-group net)**.** *Let $h = d(d-1)/2$. For $0 < \eta < 1$ the group $O(d)$ admits an operator-norm η-net of cardinality at most $2(1 + 2\pi/\eta)^h$.*

Proof. Every $Q \in SO(d)$ is a direct sum of planar rotations and an identity block, with eigenvalue -1 occurring with even multiplicity and absorbed into rotations by π; hence $Q = e^X$ for skew-symmetric X with $\|X\|_{\text{op}} \leq \pi$. Take a maximal η-separated set in the operator-norm ball of radius π inside the h-dimensional space of skew-symmetric matrices. Disjoint balls of radius $\eta/2$ lie in the ball of radius $\pi + \eta/2$, so volume comparison bounds the cardinality by $(1 + 2\pi/\eta)^h$, and maximality gives covering. The Duhamel identity $e^X - e^Y = \int_0^1 e^{(1-t)X}(X - Y)e^{tY}\,dt$ has orthogonal factors on both sides, so $\|e^X - e^Y\|_{\text{op}} \leq \|X - Y\|_{\text{op}}$ and the exponential transports the net. The remaining component is a fixed reflection times $SO(d)$, giving the factor 2. □

With $\|QMQ^\top - RMR^\top\|_F \leq 2\|Q - R\|_{\text{op}}\|M\|_F$, a support-space net of radius $\varepsilon/2$ and a group net of radius $\varepsilon/4$ compose into an ε-net of the conjugated image, of cardinality at most $2(5/\varepsilon)^s((1 + 8\pi)/\varepsilon)^h$ when the support space has dimension at most s. Combining with the cap bound $\Pr(\|U - v\| \leq R) \leq (2R)^{N-1}$ for $R \leq 1/2$ gives the following computable statement.

Theorem 5 (Finite-dimensional probabilistic bound)**.** *In the symmetric case set*

$$\begin{gathered} N = k\tfrac{d(d+1)}{2}, \qquad h = \tfrac{d(d-1)}{2}, \qquad s = \min\Big\{N, \lfloor \tfrac{r^2 B + r\min(B, km)}{2} \rfloor\Big\}, \qquad M = k\tfrac{m(m+1)}{2}, \\ \Delta = N - 1 - s - h, \qquad K = (M{+}1)\log 2 + s\log 5 + h\log(1{+}8\pi) + (N{-}1)\log 6. \end{gathered} \tag{53}$$

If $\Delta > 0$ then for every $u > 0$, with $\varepsilon = \min\{1/12, e^{-(K+u)/\Delta}\}$ and $c = \varepsilon^2$,

$$\Pr\big[\mathcal{E}_B(\mathcal{A}) \geq c\big] \geq 1 - e^{-u}. \tag{54}$$

The condition $\Delta > 0$ is exactly subcriticality: $h/N \to 1/k$ and $s/N \to \rho$, so $\Delta > 0$ iff $\rho < 1 - 1/k$. Relaxing $s \leq N$, $h \leq r^2m^2/2$, $N \leq k(r^2 + r)m^2/2$ and $M + 1 \leq (k + 1)m^2$ yields $K \leq A_{\mathsf{S}}m^2$ with $A_{\mathsf{S}} = (k + 1)\log 2 + \frac{k(r^2+r)}{2}\log 30 + \frac{r^2}{2}\log(1 + 8\pi)$, and $\Delta \geq am^2$ with $a = kr^2\delta/2$ and $\delta = 1 - 1/k - \rho$. Taking $u = m^2$ gives the dimension-free constant

$$c_{\mathsf{S}}(k, r, \rho) = \min\Big\{\tfrac{1}{144},\ \exp\big(-\tfrac{2(A_{\mathsf{S}}+1)}{a}\big)\Big\}, \qquad \gamma = 1, \tag{55}$$

which depends on k, r, ρ but not on m, as Theorem 2 requires. Two caveats should be read with it. The constant degrades as $\log(1/c) = O(1/\delta)$ near the threshold, so any report of c must state δ. And it is conservative: at $m = 16$, $r = 2$, $k = 2$, $\rho = 0.25$ one gets $\Delta = 271$ and $c \approx 6.9 \times 10^{-15}$ at failure probability e^{-256}. This certifies that the error cannot be driven to zero uniformly in m; it is not a quantitative account of the errors of order 10^{-1} measured in Section 6.2.

B.6 THE CRITICAL DENSITY

At $\rho = \rho_c$ exactly, the budget is $B_0 = (k - 1)m^2$, which is m blocks short of B_*.

Theorem 6 (Critical behavior)**.** *Let $B_0 = (k - 1)m^2$ with $k \geq 2$, $m \geq 2$. Every generic symmetric family satisfies $\mathcal{E}_{B_0}(\mathcal{A}) > 0$, while*

$$\inf_{\mathcal{A} \neq 0} \textit{(attainable normalized error at } B_0) \longrightarrow 0 \qquad (m \to \infty). \tag{56}$$

Proof sketch. Strict positivity is Theorem 1 together with the closed-cone distance form: $B_0 < B_*$, so no exact representation exists and the distance to a closed cone not containing $\mathcal{A}$ is positive. For the vanishing upper envelope, take the eigenbasis of one action and drop the m smallest-energy diagonal blocks of a second action; the discarded energy is at most an $O(1/m)$ fraction of the pooled energy after normalization. Hence the critical point is a boundary case: positive at every finite m, but with no dimension-free lower bound. □

This is the third regime shown in Figure 3. It also explains why the finite-dimensional threshold $\rho_*(m,k) = \rho_c + 1/(km)$ of Eq. (13), and not ρ_c itself, is the density at which the measured error becomes exactly zero.

C Compatibility Results

This appendix collects the formal positive results and the common-envelope comparison used in Section 5.

C.1 Exact and approximate graph envelopes

For an undirected graph $G = ([m], E)$, let $L_G \subset \mathrm{Sym}_d(\mathbb{R})$ be the linear space whose active blocks are restricted to the m diagonal positions and the two directed positions corresponding to each edge. Its directed envelope size is

$$s_G = m + 2|E|. \tag{57}$$

Formal proof of Proposition 1. By assumption, $Q_*^\top A_a Q_* \in L_G$ for every action. Choosing the feasible shared basis Q_* gives at most s_G active blocks per action, hence $C_{\mathrm{sh}} \le k s_G = k(m+2|E|)$. If the maximum degree of G is bounded by Δ, then $2|E| \le \Delta m$, so $s_G \le (1+\Delta)m$ and the cost is $O(km)$. □

Formal proof of Theorem 3. Assume $Q_*^\top A_a Q_* = S_a + R_a$ with $S_a \in L_G$. Set the shared basis to Q_* and the approximant of action a to $\widetilde{A}_a = S_a$. The total block count is at most $k s_G$. The objective value of this feasible point is

$$\frac{\sum_a \|Q_*^\top A_a Q_* - S_a\|_F^2}{\sum_a \|A_a\|_F^2} = \frac{\sum_a \|R_a\|_F^2}{\sum_a \|A_a\|_F^2}. \tag{58}$$

Taking the infimum over all feasible shared representations can only reduce this value. □

C.2 Compatibility with a common envelope

A symmetric directed block envelope Ω is a set of block positions closed under transposition. For size s, the compatibility quantity can equivalently be written as

$$\Gamma_s(\mathcal{A}) = \inf_{Q,\Omega} \frac{\sum_a \|Q^\top A_a Q - P_\Omega(Q^\top A_a Q)\|_F^2}{\sum_a \|A_a\|_F^2}, \tag{59}$$

where $|\Omega| \le s$ and P_Ω keeps exactly the blocks in Ω. This form makes clear that Γ_s is non-increasing in s.

Formal proof of Theorem 4. Take any feasible common-envelope representation in the definition of Γ_s. Each action uses at most the s positions in Ω, so the total directed block count across k actions is at most ks. The same basis and approximants are feasible for $\mathcal{E}_{ks}$, which proves $\mathcal{E}_{ks} \le \Gamma_s$ after taking infima. For the other direction, take any feasible representation for $\mathcal{E}_s$. Across all actions it uses at most s directed positions in total. Let Ω be the union of those positions. Then $|\Omega| \le s$, the same basis and approximants are feasible for Γ_s, and their error is unchanged. Taking infima gives $\Gamma_s \le \mathcal{E}_s$. □

Formal proof of Corollary 1. Let $s_m = \lfloor \rho m^2 \rfloor$ and $B_m = \lfloor \rho k m^2 \rfloor$. Since $k\lfloor x \rfloor \le \lfloor kx \rfloor$ for $k \in \mathbb{N}$,

$$k s_m \le B_m. \tag{60}$$

The budgeted error is non-increasing in its budget, so $\mathcal{E}_{ks_m} \geq \mathcal{E}_{B_m}$. The subcritical part of Theorem 2 gives $\mathcal{E}_{B_m} \geq c$ with probability at least $1 - e^{-\gamma m^2}$. The left inequality in Theorem 4 then implies $\Gamma_{s_m} \geq \mathcal{E}_{ks_m} \geq c$ on the same event. □

C.3 Residual at a fixed graph

For a fixed graph G and a fixed basis Q, write

$$\delta_G(\mathcal{A}; Q) = \inf_{S_a \in L_G} \frac{\sum_a \|Q^\top A_a Q - S_a\|_F^2}{\sum_a \|A_a\|_F^2}, \qquad \delta_G^\star(\mathcal{A}) = \inf_{Q \in O(d)} \delta_G(\mathcal{A}; Q). \tag{61}$$

The graph fixes the permitted locations and lets the numerical values vary by action. Throughout the experiments δ_G denotes the residual at the planted basis, which is the quantity on the horizontal axis of Figure 4.

Proposition 2 (Graph residual controls budget error)**.**

$$\mathcal{E}_{ks_G}(\mathcal{A}) \leq \Gamma_{s_G}(\mathcal{A}) \leq \delta_G^\star(\mathcal{A}) \leq \delta_G(\mathcal{A}). \tag{62}$$

Proof. The first inequality is Theorem 4. For the second, the graph envelope itself is one admissible common envelope of size s_G in the definition of Γ_{s_G}, and optimizing over all envelopes can only improve on the fixed graph. The third holds because the planted basis is one admissible choice of Q. □

The planted-basis residual is thus a directly computable upper bound on every quantity in the chain, which is what makes it a convenient horizontal axis.

C.4 Isotropic residual geometry

The symmetric matrix space has dimension

$$D = \frac{d(d+1)}{2}. \tag{63}$$

The graph support space has dimension

$$D_G = m\frac{r(r+1)}{2} + |E|r^2. \tag{64}$$

For a Frobenius-isotropic random direction N in $\mathrm{Sym}_d(\mathbb{R})$, let $P_{L_G^\perp}$ project onto the orthogonal complement of L_G. Rotational symmetry gives

$$\mathbb{E}\|P_{L_G^\perp} N\|_F^2 = q_G \mathbb{E}\|N\|_F^2, \qquad q_G = 1 - \frac{D_G}{D}. \tag{65}$$

The experiment normalizes S and N separately and uses $S + \sigma N$. In high dimension, $\langle S, N\rangle_F$ concentrates near zero, while the off-envelope fraction of N concentrates near q_G. Thus the planted-basis residual is predicted by

$$\delta_G(\sigma) \approx \frac{\sigma^2 q_G}{1 + \sigma^2}. \tag{66}$$

This is a planted-basis geometry prediction, and the optimized quantities $\delta_G^\star$, Γ_{s_G}, and $\mathcal{E}_{ks_G}$ sit at or below it.

D Coordinates Chosen per Action

This appendix quantifies the trajectory cost of storing the state in action-specific coordinates. Let

$$A_a = Q_a D_a Q_a^\top \tag{67}$$

be an eigen-decomposition for action a, with D_a using m diagonal blocks.

Proposition 3 (Trajectory cost of action-specific coordinates)**.** *Consider a trajectory of H action updates with N_{sw} action switches. For a generic set of relative bases $Q_b^\top Q_a$, each switch uses all m^2 block relations. The total block-interaction count is*

$$C_{\mathrm{switch}} = Hm + N_{\mathrm{sw}} m^2, \tag{68}$$

and the average per step is

$$\frac{C_{\mathrm{switch}}}{H} = m + q_H m^2, \qquad q_H = \frac{N_{\mathrm{sw}}}{H}. \tag{69}$$

Thus a constant-order switch rate yields $\Theta(m^2)$ average interaction count, while $q_H = O(1/m)$ keeps the average at $O(m)$.

Proof. An update performed in the current action's own eigenbasis uses the m diagonal blocks of D_a. When the next action changes from a to b, the same state must be converted by the relative orthogonal transform $Q_b^\top Q_a$. A fixed block of a generic orthogonal matrix is nonzero outside the zero set of a nontrivial polynomial in the matrix entries. The finite union over block positions has Haar measure zero, so a generic relative basis has all m^2 blocks active. Summing m for each of the H updates and m^2 for each of the N_{sw} switches gives the formula. Dividing by H gives the average and the two asymptotic regimes. □

This calculation places the two representation choices on the same scale. A persistent shared basis pays for incompatibility inside action updates, while action-specific coordinates can concentrate the same quadratic-scale interaction at basis changes.

E Experimental Details

E.1 Protocol and Calibration

The synthetic and learned-model experiments use the same basic definitions. Exact block counts use an $m \times m$ partition into $r \times r$ blocks. Finite-budget error is the discarded Frobenius energy divided by the pooled Frobenius energy of the full action family. The total budget B may be redistributed across actions, and learned-model comparisons give every action the same per-action budget b; each result states which convention it uses. In the main text,

$$F_b = E_b^{\mathrm{sh}} - E_b^{\mathrm{sep}} \tag{70}$$

is computed instance by instance, and both endpoint errors are reported alongside it. Reporting the absolute difference together with its two endpoints keeps the comparison well defined at every value of E_{sep}.

Learned-model experiments use the one-step residual Jacobian $A_a = J_a - I$. Using $H = 1$ makes the Jacobian correspond directly to the action-conditioned update that the theory describes. To control the latent gauge, we estimate the covariance on a fixed calibration split and apply exact whitening. After whitening, an invertible linear reparameterization reduces to a common orthogonal transform, and the block error used here is invariant to common orthogonal conjugation. All shared and per-action comparisons use the same whitening transform, the same budget, and the same pooled denominator.

Experiments that require basis search optimize over the orthogonal group and initialize from the identity, action-derived bases, and multiple random starts, following orthogonality-preserving coordinate updates (Edelman et al., 1998; Shalit & Chechik, 2014; Frerix & Bruna, 2019) and established approximate joint-diagonalization constructions (Cardoso & Souloumiac, 1996; Pham, 2001; Abed-Meraim & Belouchrani, 2004). The numerical implementation is first calibrated on a two-dimensional analytic instance, where both grid evaluation and optimization recover the closed-form solution. On learned Jacobians the manifold solver improves consistently on Givens and variable-projection searches, and every reported comparison pairs shared and per-action values produced within one solver batch.

The flat model is a multilayer perceptron over the concatenated state and action embedding. The slot model encodes each particle into one latent unit, uses action-conditioned multi-head attention (Vaswani et al., 2017) for cross-unit interaction, and applies feed-forward residual updates per unit. Both are trained with Adam (Kingma & Ba, 2015) on one-step prediction, with observations

standardized on the training split, and the potential model of the regime comparison replaces the transition by one gradient step on a learned per-action potential.

The replication unit follows the question being tested. Independently generated synthetic families are independent replicates. For learned models, independently trained seeds are the primary replicate, while multiple held-out states and multiple budgets are nested within each seed. Full random seeds, training settings, and state-level results are stored with the code.

E.2 Connectivity and Budgets

Table 1: Measured block counts against the closed forms, at $k = 4$ and $r = 2$, averaged over independent families at the strict numerical-zero threshold. The separate count matches km in the symmetric regime and $km(m+1)/2$ in the general real regime at every m. The shared count matches its closed form exactly for the symmetric families, and to within 0.6% for the general real families.

family	m	C_{sep}	closed form	C_{sh}	closed form
commuting	4	16	16	16	16
commuting	8	32	32	32	32
commuting	16	64	64	64	64
ring envelope	4	16	16	48	48
ring envelope	8	32	32	96	96
ring envelope	16	64	64	192	192
generic, symmetric	4	16	16	52	52
generic, symmetric	8	32	32	200	200
generic, symmetric	16	64	64	784	784
generic, symmetric	24	96	96	1752	1752
generic, general real	4	40	40	58	58
generic, general real	8	144	144	229	228
generic, general real	16	544	544	906	904
generic, general real	24	1200	1200	2032	2028

Exact counts. Table 1 compares the measured counts with the closed forms. The three symmetric families separate cleanly: the commuting family pays nothing for sharing, the ring envelope pays a constant factor, and the generic family grows quadratically. The general real families sit in the bounded regime, where the separate count is already quadratic. The general real shared count sits within 0.6% of its closed form, with the relative excess shrinking from 0.5% at $m = 8$ to 0.2% at $m = 24$.

Threshold and block ordering. A block is counted as present when it carries more than a set fraction of the family energy. Table 1 uses a strict numerical zero, the threshold that matches the exact-support definition of bnnz_r. For the general real families we sort the complex eigenvalue pairs to the front of the Schur form, so that each 2×2 block lies inside a single $r \times r$ block of the imposed partition. All counts reported here use this ordering, and they match the closed form at every m and r tested.

Full budget sweep. The Theorem 2 experiments cover $k = 2, 4, 8$, $r = 2, 4$, and several values of m. Each family is evaluated from clearly subcritical budgets to budgets above B_*. All 300 instances with $B \geq B_*$ satisfy $\mathcal{E}_B < 10^{-12}$, with the largest residual at the level of numerical precision. This zero error is obtained by the explicit basis construction alone.

The three regimes. Figure 6 shows the full (k, r) grid, and Figure 7 separates the regimes by holding the density fixed while m grows. Subcritical curves are flat over an eightfold increase in m, which is what a dimension-free constant requires, and the same holds when the budget is set as a fixed fraction of B_* as well as an absolute density. At the critical density the error is positive at every m but decays steadily, consistent with Theorem 6. Repeating the subcritical sweep with a basis search on the same families reproduces the same flat profile up to a bounded factor that stays constant in m.

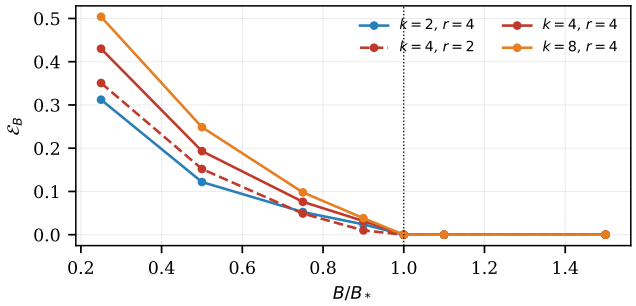

Figure 6: Finite-budget error over the (k, r) grid. Every curve falls to exactly zero at $B = B_*$ and stays there. Larger k and larger r raise the subcritical error at the same fraction of the threshold.

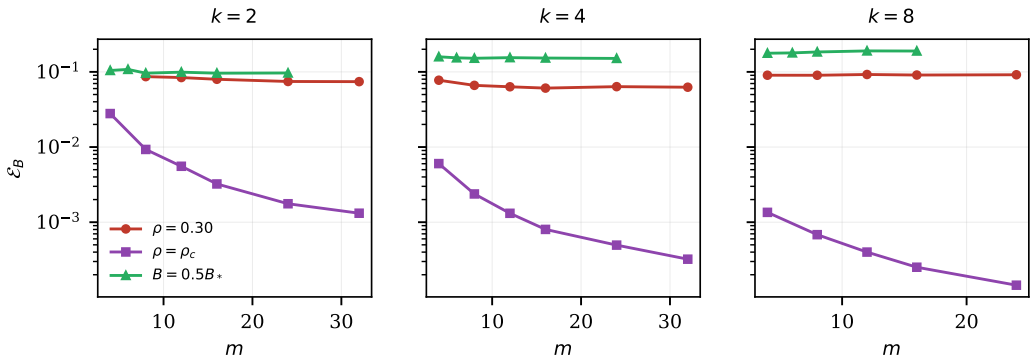

Figure 7: The three regimes as m grows, at $r = 2$. A fixed subcritical density ($\rho = 0.30$) and a fixed fraction of B_* ($B = 0.5B_*$) both give an error that is flat in m; the critical density gives an error that is positive at every m but decays.

Correction at finite size. The exact budget density is $\rho_*(m, k) = B_*/(km^2) = 1 - 1/k + 1/(km)$, hence $\rho_*(m, k) - \rho_c = 1/(km)$ is the finite-size correction implied directly by the exact count in Theorem 1. The numerical results match this identity for all tested values of k: the measured offsets are 0.125, 0.0625 and 0.03125 for $k = 2, 4, 8$ at $m = 4$, equal to $1/(km)$ to the digits shown. As a coordinate-choice control, the action-eigenbasis construction reaches zero at B_* while the identity basis retains nonzero error at the same budget, so the transition tracks the shared basis together with the budget.

E.3 Common Envelope Experiments

Exact common envelope. The Theorem 3 construction starts from a known common sparse basis in which all actions share the same envelope. When the budget can contain the full envelope, the unperturbed endpoint has δ_G, E_{sep} and E_{sh} all at numerical zero. This is the finite-budget realization of Proposition 1; the ring family in Theorem 1 gives the corresponding exact-support realization in which shared connectivity remains linear.

Approximate common structure. After continuously perturbing the planted family, we measure the residual energy δ_G that leaves the common envelope in the planted basis, which puts the structural distance on the same scale as the error it controls. Across 84 valid settings with $m = 8, 12, 16$, using the planted basis and discarding the off-envelope component gives $E_{\text{sh}}^{\text{true}} = \delta_G$ with a maximum absolute discrepancy of 1.8×10^{-16}. Further optimization of the shared basis stays at or below this residual bound to numerical precision. For $m = 16$, δ_G increases smoothly from 0 to about 0.065, while the optimized shared error remains at or below the same bound. The main text uses δ_G on the horizontal axis, and the sequence of perturbation levels is monotone in δ_G.

Sharing of support. A complementary experiment varies the imposed fraction of shared support q_{shared}, which sets how many support positions the actions are required to share. Reducing that fraction raises F_b overall, the direction Proposition 1 predicts, and this sweep serves as a mechanism check alongside the planted-envelope results.

Sandwich between budget and compatibility. We additionally evaluate $\mathcal{E}_{ks} \leq \Gamma_s \leq \mathcal{E}_s$ on 168 parameter settings with $m = 8$, $r = 4$ and $k = 4$, and observe no violations. The relation follows directly from containment of the feasible sets, and the sweep confirms the implementation across the whole grid.

E.4 Learned Models

Table 2: Per-seed results at $b = 0.5m$, averaged over held-out states. The two conditions of the flat model are aligned in prediction error and in E_{sep}, and separate in F_b with no overlap between the two ranges. The slot model keeps F_b close to zero in both conditions.

model	support	seed	prediction error	E_{sep}	E_{sh}	F_b
flat	shared	0	0.0024	0.0807	0.0882	0.0075
flat	shared	1	0.0035	0.0451	0.0471	0.0021
flat	shared	2	0.0025	0.0556	0.0589	0.0033
flat	per-action	0	0.0020	0.0337	0.0550	0.0213
flat	per-action	1	0.0021	0.0925	0.2029	0.1104
flat	per-action	2	0.0019	0.0709	0.1087	0.0377
slot	shared	0	0.0003	0.0738	0.0743	0.0005
slot	shared	1	0.0002	0.0498	0.0500	0.0002
slot	per-action	0	0.0008	0.0798	0.0804	0.0006
slot	per-action	1	0.0009	0.0588	0.0591	0.0003

Results per seed. Table 2 gives every run. For the flat model the two conditions are matched on the two quantities that could otherwise explain the effect: prediction error and the per-action-basis error E_{sep} are of the same order across conditions, so the separation in F_b is carried by the support structure alone. The two F_b ranges are disjoint. We report the aggregate together with the full per-seed range. A separate batch trained on a shorter schedule, with four seeds per condition, reproduces the same direction with a tighter spread.

Control with the slot model. The slot model already organizes the latent state by objects. Under this architectural prior, both the shared and the per-action support environments have F_b close to zero, with only a small difference between them. This serves as a negative control: when the representation already fixes a strong object-wise decomposition, changing support sharing in the environment leaves the shared-coordinate cost close to zero.

Learned Jacobian regime. In the flat MLP, the symmetric component of the one-step residual Jacobian accounts for about 29% of the total energy, placing the learned operators in the general real-matrix regime of Eq. (10). As an auxiliary control, symmetrizing the same Jacobians substantially lowers the separate error and increases the shared penalty. In a separately trained potential model, whose transition is one gradient step on a learned per-action potential, the Jacobians are symmetric to numerical precision and prediction quality is comparable to that of the flat model, while E_{sep} falls to about 6×10^{-4} compared with about 10^{-1} for the flat model. This supports the distinction between regimes according to whether a single action can become highly local under an orthogonal change of coordinates, and it places each learned model in the regime that matches its operator class.

Gauge calibration. Applying an invertible linear reparameterization to the flat latent state leaves model predictions unchanged. After exact whitening, Jacobians from different gauges differ only by a common orthogonal transform. Algebraic checks of prediction invariance, Jacobian conjugacy, covariance conjugacy and whitening transport are all at numerical precision. Exact whitening is the

transport that keeps the scoring equivariant, and the main text relies on this exact invariance together with orthogonally invariant scoring.

Local and global bases. The learned-Jacobian analysis in the main text lets each held-out state choose its own optimized basis, which makes it a state-local linear diagnostic matched to the one-step theory. We also evaluate a global version that fixes one basis across states, and the two formulations give shared-coordinate costs of the same size in these models. The state-local form is the one reported throughout.