Beyond DSA: Conjugacy-based Comparison of Dynamical Systems

Prakhar Godara
Department of Psychology
New York University, NY 10003, USA
prakhargodara@gmail.com

Pang Shiang Tay
Department of Mathematics
Göttingen University, 37073, Germany
pangshiang.tay@stud.uni-goettingen.de

Marcelo G Mattar
Department of Psychology
New York University, NY 10003, USA
mm13100@nyu.edu

Abstract

Comparing whether two dynamical systems implement the same computation despite differences in coordinates or measurements is a central problem in neuroscience and machine learning. Dynamical Similarity Analysis [DSA; Ostrow et al., 2023] addresses this problem by aligning finite-dimensional Koopman approximations of the two systems through an orthogonal similarity transformation. Here we show that orthogonal alignment is neither necessary nor sufficient for topological conjugacy: genuinely conjugate systems may be related by a non-orthogonal basis-transfer matrix that DSA cannot capture, while non-conjugate systems may have orthogonally equivalent Koopman operators that DSA will fail to distinguish. We then use this observation to formulate *Conjugacy-based Similarity Analysis* (CSA), which restricts alignments to those induced by candidate state-space homeomorphisms rather than arbitrary orthogonal matrices. We prove that CSA's fitted alignment is the finite-data projection of the composition operator associated with the candidate homeomorphism, and use controlled examples to show why this distinction matters when observable dictionaries are chosen explicitly or implicitly from data. Together, these results clarify what Koopman-based similarity measures must ensure to support claims of identifying conjugacies between computational systems.

1 Introduction

Comparing the computations performed by two dynamical systems is a central methodological question across neuroscience and machine learning. In neuroscience, such comparisons are used to validate computational models against neural recordings [Schrimpf et al., 2021, Pagan et al., 2022], to align brain–computer interfaces across recording sessions [Degenhart et al., 2020, Karpowicz et al., 2025], and to ask whether different circuits implement equivalent computations [Chaudhuri et al., 2019]. In machine learning, similarity measures between learned representations are used to characterize training dynamics, architectural invariances, and model families [Kriegeskorte et al., 2008, Raghu et al., 2017, Williams et al., 2021]. Across both fields, the underlying question is the same: when do two systems compute in the same way?

A natural way to make the question of computational equivalence precise is to ask whether one system can be obtained from the other by a relabelling of states. Concretely, suppose a homeomorphism h matches each state x of one system to a state $h(x)$ of the other. If running the first system

for one step and then relabelling produces the same result as relabelling first and then running the second system, then the two systems generate the same trajectories up to a renaming of their state variables. In dynamical systems theory this property is called *topological conjugacy*. Conjugacy thus formalizes the intuition behind "same computation" by ignoring the particular coordinates in which states are described and asking only whether the underlying transition rule is the same. Geometric similarity, by contrast, fails to capture this invariance: it compares the shapes traced out by trajectories rather than the rules that generate them [Kriegeskorte et al., 2008, Raghu et al., 2017, Williams et al., 2021].

Testing conjugacy directly is difficult because the relabelling function h is unknown and lives in an infinite-dimensional space of candidate maps. Koopman operator theory offers a route around this. The Koopman operator represents a possibly nonlinear dynamical system through its linear action on observables, providing a principled language for learning and comparing dynamical structure from data [Koopman, 1931]. Building on this framework, Dynamical Similarity Analysis (DSA) was recently introduced to compare dynamical systems in terms of their finite-dimensional Koopman approximations rather than raw state-space geometry [Ostrow et al., 2023]. DSA has since been adopted in systems neuroscience and deep learning as the operator-level alternative to geometric similarity.

DSA, however, looks for the wrong kind of match between Koopman operators. Conjugacy implies that a relabelling induces a specific transformation of measurements, and it is this transformation that should align the two Koopman operators. DSA ignores this, searching instead for any orthogonal alignment between Koopman matrices, whether or not it corresponds to a relabelling of states. This mismatch creates two failure modes: DSA can make genuinely different systems look similar when an arbitrary orthogonal alignment hides their difference, and can make genuinely equivalent systems look different when the correct state-by-state correspondence is outside its alignment class. In other words, DSA compares learned operator representations without requiring the alignment to be a valid change of coordinates in the underlying dynamical system. Consequently, neither a small nor a large DSA distance supports a clean conclusion about the dynamics: a small distance need not imply the same computation, and a large distance need not imply different computations.

To address this problem, we introduce *conjugacy-based similarity analysis* (CSA), a template for comparing finite-dimensional Koopman approximations through candidate conjugacies rather than arbitrary orthogonal alignments. Given a parameterized family of candidate relabellings $\{h_\theta\}_{\theta\in\Theta}$ (eg: monotone polynomials of bounded degree, etc.), CSA constructs from each h_θ the matrix P_θ that records how measurements transform under h_θ, and then asks how well this P_θ aligns the two Koopman matrices. The score is the smallest alignment error obtained as θ varies over the family. Because every P_θ comes, by construction, from an actual relabeling of states, any alignment found by CSA could have been produced by a conjugacy, a guarantee that does not hold for alignments found by DSA. Empirically, CSA correctly identifies conjugate systems in the very regimes where DSA's failure modes predict that it cannot, both when the measurement dictionary is chosen explicitly and implicitly from data.

Our contributions are threefold. **First**, we show that orthogonal alignment of Koopman matrices is not, in general, a valid criterion for conjugacy, even before approximation error is considered (Section 3). **Second**, we introduce CSA (Section 4) as a template for conjugacy-aware Koopman comparison, and prove that the matrix it fits is the correct finite-dimensional shadow of the composition operator induced by a candidate relabelling, so that CSA estimates a well-defined operator-theoretic object rather than a heuristic alignment. **Third**, we instantiate this template in controlled settings (Section 5) and show empirically that the distinction between orthogonal alignment and relabelling-induced alignment changes the conclusions one draws about conjugacy, both under state-space transformations and under changes of measurement basis.

2 Theory

2.1 Dynamical systems and conjugacy

We consider discrete-time dynamical systems $f : X \to X$ and $g : Y \to Y$, where X and Y are state spaces. Two such systems are *topologically conjugate* if there exists a homeomorphism $h : X \to Y$ such that

$$h \circ f = g \circ h, \tag{1}$$

in which case h provides a state-by-state correspondence under which the two systems generate the same trajectories. Testing conjugacy directly is difficult because the relabelling h is unknown and, in general, must be searched for over a large function space of candidate maps. The remainder of this section introduces the linear operators that allow us to work with conjugacy through measurements rather than through h directly.

2.2 Koopman operators

The Koopman operator lifts a possibly nonlinear dynamical system to a linear operator on observables [Koopman, 1931]. Let $\mathcal{H}_X$ and $\mathcal{H}_Y$ be Hilbert spaces of scalar-valued observables on X and Y. The Koopman operators associated with f and g are

$$(\mathcal{K}_f u)(x) = u(f(x)), \qquad (\mathcal{K}_g v)(y) = v(g(y)),$$

for observables $u \in \mathcal{H}_X$ and $v \in \mathcal{H}_Y$. Thus, although f and g may be nonlinear maps on state space, $\mathcal{K}_f$ and $\mathcal{K}_g$ are linear operators on observable spaces. This replaces a nonlinear state-space comparison problem with a linear operator comparison problem, at the cost of working in infinite-dimensional function spaces.

2.3 Composition operators induced by homeomorphisms

Any homeomorphism between state spaces induces a corresponding linear map between observable spaces. Specifically, a homeomorphism $h : X \to Y$ induces the *composition operator*

$$\mathcal{C}_h : \mathcal{H}_Y \to \mathcal{H}_X, \qquad (\mathcal{C}_h v)(x) = v(h(x)). \tag{2}$$

which pulls back observables on Y to observables on X along h. The composition operator is the natural counterpart, at the level of measurements, of the relabelling h at the level of states. When f and g are conjugate, this specific operator—not an arbitrary linear map between observable spaces—connects $\mathcal{K}_f$ and $\mathcal{K}_g$. Indeed, the conjugacy relation in Eq. (1) implies that this composition operator intertwines the two Koopman operators. Thus, for any $v \in \mathcal{H}_Y$,

$$(\mathcal{C}_h \mathcal{K}_g v)(x) = (\mathcal{K}_g v)(h(x)) = v(g(h(x))) = v(h(f(x))) = (\mathcal{C}_h v)(f(x)) = (\mathcal{K}_f \mathcal{C}_h v)(x). \tag{3}$$

Therefore

$$\mathcal{C}_h \mathcal{K}_g = \mathcal{K}_f \mathcal{C}_h. \tag{4}$$

This is the operator-theoretic consequence of topological conjugacy: the intertwiner is not an arbitrary operator on observables, but specifically the composition operator induced by the same state-space map h that conjugates the dynamics.

2.4 Matrix representation of the intertwining relation

To express the operator intertwining (4) in coordinates, let $\{\phi_i\}_{i\geq 1}$ and $\{\psi_j\}_{j\geq 1}$ be orthonormal bases of $\mathcal{H}_X$ and $\mathcal{H}_Y$, respectively, and write

$$\Phi(x) = (\phi_1(x), \phi_2(x), \ldots)^\top, \qquad \Psi(y) = (\psi_1(y), \psi_2(y), \ldots)^\top.$$

The matrix representations F and G of the Koopman operators $\mathcal{K}_f$ and $\mathcal{K}_g$ in these bases are defined by

$$\Phi(f(x)) = F\Phi(x), \qquad \Psi(g(y)) = G\Psi(y), \tag{5}$$

and the matrix representation of the composition operator $\mathcal{C}_h$ is the matrix P satisfying

$$\Psi(h(x)) = P\Phi(x). \tag{6}$$

The jth row of P gives the coefficients, in the Φ-basis, of the pulled-back observable $\mathcal{C}_h\psi_j = \psi_j \circ h$; thus P is not an arbitrary change of coordinates but the coordinate representation of a pullback. Combining Eqs. (1), (5) and (6) yields the matrix-level counterpart of the operator intertwining (4):

$$GP\Phi(x) = G\Psi(h(x)) = \Psi(g(h(x))) = \Psi(h(f(x))) = P\Phi(f(x)) = PF\Phi(x). \tag{7}$$

Since this holds for all x, the matrices satisfy

$$PF = GP, \tag{8}$$

or equivalently, when P is invertible,

$$G = PFP^{-1}. \tag{9}$$

This is the matrix-level signature of conjugacy: the two Koopman representations are related not by an arbitrary alignment, but by the matrix representation of the composition operator induced by the state-space relabelling h.

The next section asks whether the class of alignments used by DSA matches the class of alignments implied by conjugacy.

3 Why DSA fails as a test of conjugacy

DSA attacks the conjugacy problem by directly aiming to approximate Eq. (9). However, as aforementioned, the matrix relation $PF = GP$ is not an arbitrary similarity: P is the coordinate representation of a pullback induced by a candidate state-space map h, and must therefore satisfy Eq. (6) for some h—i.e., it cannot be chosen freely. Any criterion intended to detect conjugacy must therefore search over alignments that could plausibly arise from such pullbacks. The following two subsections show that DSA's orthogonal search fails this requirement, both in the alignment class it considers (§3.1) and in the information it discards when computing distances (§3.2).

3.1 DSA searches over the wrong alignment class

Given finite-dimensional Koopman approximations F and G of the same size, DSA computes

$$d_{\mathrm{DSA}}(F, G) = \min_{Q \in O(m)} \|G - QFQ^{-1}\|_F. \tag{10}$$

That is, DSA asks whether the two Koopman matrices can be made similar by an orthogonal change of coordinates, whereas $PF = GP$ asks for a composition-induced P. These two classes are not in containment in either direction.

On one hand, DSA is too permissive because not every orthogonal alignment is composition-induced. A defining property of a composition operator is that it is distributive over multiplication of observable functions, but a generic orthonormal operator does not have this property (Appendix A provides a simple example). Crucially, this is not an artifact coming from the finite-dimensional approximations: classical results in ergodic theory exhibit measure-preserving systems whose true (infinite-dimensional) Koopman operators are unitarily equivalent without being topologically conjugate [Rédei and Werndl, 2012], and Appendix B works through one such case for Bernoulli shifts.

On the other hand, DSA is also too restrictive because not every composition-induced alignment is orthogonal. The composition operator $\mathcal{C}_h$ is unitary only under additional assumptions, most notably when h is measure preserving [Budišić et al., 2012] (see Appendix C). Outside that special case, the correct intertwiner falls outside the orthogonal class searched by DSA, and DSA can report a large discrepancy even between genuinely conjugate systems.

3.2 The mismatch is representational, not merely numerical

Beyond optimizing over an inappropriate domain, DSA ignores too much information when computing distances. For instance, when solving the optimization problem in Eq. (10), DSA ignores which subspaces are F, G operators in and how those spaces are relatively oriented. Rather it only views them as numerical arrays and looks for similarity. The form this issue takes depends on how the Koopman approximation is learned. In EDMD, the observable dictionary is chosen explicitly, so the basis-transfer problem is directly visible. In Hankel-DMD one identifies the right subspace by taking the SVD of the Hankel matrix [Arbabi and Mezic, 2017]. When comparing the corresponding Koopman matrices, DSA ignores the information of *which* lags were chosen by the SVD procedure, which is arguably relevant for evaluating conjugacy.

These observations identify the design requirement for a conjugacy-aware comparison criterion. Rather than comparing Koopman matrices as bare numerical arrays, one should compare them through alignments that can be interpreted as projected pullbacks of candidate state-space maps. The next section introduces *conjugacy-based similarity analysis* (CSA), which implements this idea: CSA first constructs the finite-dimensional transfer induced by a candidate relabelling, and only then asks whether that transfer intertwines the two Koopman approximations.

4 Conjugacy-based Similarity Analysis (CSA)

Let $f : X \to X$ and $g : Y \to Y$ be discrete-time dynamical systems as above. We now assume finite observable dictionaries

$$\Phi = (\phi_1, \dots, \phi_m)^\top : X \to \mathbb{R}^m, \qquad \Psi = (\psi_1, \dots, \psi_n)^\top : Y \to \mathbb{R}^n,$$

and finite-dimensional Koopman approximations $F \in \mathbb{R}^{m\times m}$ and $G \in \mathbb{R}^{n\times n}$ (where m need not be equal to n), for instance fitted by EDMD [Williams et al., 2015] or Hankel-DMD [Arbabi and Mezic, 2017], satisfying

$$\Phi(f(x)) \approx F\Phi(x), \qquad \Psi(g(y)) \approx G\Psi(y).$$

CSA restricts the admissible alignments to those that arise from candidate state-space maps. Let $\{h_\theta : X \to Y : \theta \in \Theta\}$ be a parameterized family of candidate homeomorphisms. This family specifies the class of conjugacies over which we search. In low-dimensional examples, it can be chosen explicitly; in higher-dimensional settings, one could use flexible parameterizations of diffeomorphisms, such as invertible neural architectures or residual networks with invertibility constraints [Chen et al., 2024, Huang et al., 2021]. For each h_θ, we approximate the induced pullback composition operator

$$\mathcal{C}_{h_\theta} v = v \circ h_\theta$$

as a transfer from the Ψ-dictionary to the Φ-dictionary. Concretely, given sample points $x_1, \dots, x_N \in X$, we solve

$$P_\theta = \arg\min_{P\in\mathbb{R}^{n\times m}} \sum_{i=1}^{N} \big\|\Psi(h_\theta(x_i)) - P\Phi(x_i)\big\|_2^2. \tag{11}$$

Here $P_\theta \in \mathbb{R}^{n\times m}$. It need not be square when the two dictionaries have different dimensions. The purpose of this inner problem is to ensure that the alignment matrix is not arbitrary: P_θ is the finite-data approximation to the projected pullback induced by h_θ. In the infinite-data limit, and with complete bases of the Hilbert spaces, Equation (11) recovers the coordinate representation of the corresponding composition operator; see Appendix H for proofs.

Once we have restricted P_θ to the class of projected composition operators arising from Θ, we can optimize over the space of homeomorphisms

$$d_{\mathrm{CSA}}(F, G) := \inf_{\theta\in\Theta} \frac{\big\|P_\theta F - GP_\theta\big\|_F}{\|P_\theta\|_F}. \tag{12}$$

Thus Eq. (12) measures the residual of the finite-dimensional intertwining relation

$$P_\theta F \approx GP_\theta.$$

When $m = n$ and P_θ is invertible, this is equivalent to the similarity relation

$$G \approx P_\theta F P_\theta^{-1}.$$

However, CSA does not require P_θ to be square or invertible. The residual form is therefore more general than direct matrix similarity. The normalisation by $\|P_\theta\|_F$ makes the intertwining residual relative to the size of the projected pullback. Equation (12) is the finite-dimensional form of the CSA score. Since the Frobenius norm does not generally admit a finite infinite-dimensional continuation for composition operators, Appendix M introduces a weighted Hilbert-Schmidt norm and proves convergence of the finite-sample score to a well-defined infinite-dimensional objective d^∞ (Propositions 6 and 7). Moreover, Proposition 8 shows that $d^\infty = 0$ if and only if the searched family contains a conjugacy. Pseudocode and optimisation details are given in Appendix I.

5 Example comparisons between CSA and DSA

We now test the implications of Sections 3 and 4 in controlled settings where the relevant state-space transformations and observable bases are known. The controlled setup is essential: by knowing the true homeomorphism h in advance, we can construct pairs of systems that are exactly conjugate by construction and ask whether each method recognizes them as such. As DSA is proposed as an

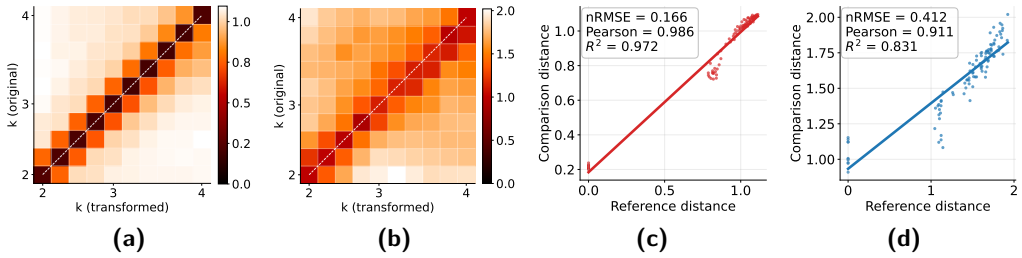


Figure 1: **State-space transformations with fixed dictionary.** Pairwise (comparison) distance matrices for the two methods - CSA (a) and DSA (b), when one copy of the logistic-map family is transformed by $h_\theta(x) = x + \theta \sin(\pi x)$ and both Koopman operators are represented in the same Fourier dictionary for both the methods. Scatter plots for reference and comparison matrix entries for CSA (c) and DSA (d).

operator-level comparison that should not depend on how the Koopman matrices are learned [Ostrow et al., 2023], we instantiate the same test in two pipelines: EDMD with explicit finite dictionaries (§5.1), and a Hankel-DMD/Krylov pipeline closer to the original paper's native setting (§5.2). In each pipeline we run a parameter-indexed family of systems through a known representational change on one side of the comparison — a state-space conjugacy or a different observable basis — and ask how faithfully a method preserves the intended pairwise geometry. In every experiment, CSA is given a parametric family that contains the true homeomorphism; this isolates the question of whether the method's construction is well-aligned with the conjugacy criterion, deferring questions of family-misspecification to the discussion (see Appendix K).

5.1 EDMD-based implementation

We instantiate the EDMD pipeline on the logistic-map family

$$f_k(x) = k\,x(1-x), \qquad x \in [0,1], \qquad k \in [2,4],$$

sweeping k across 10 equally spaced values from $k = 2$ (single stable fixed point) to $k = 4$ (chaotic). The logistic map is used only as a controlled scalar-parameter family with non-trivial dynamics; nothing in what follows depends on its specific form. For a chosen observable dictionary $\Phi(x) = (\phi_1(x), \ldots, \phi_m(x))^\top$, we obtain the Koopman matrix F_k representing $\mathcal{K}_{f_k}$ via EDMD, with samples drawn from the uniform (Lebesgue) measure on $[0,1]$. The dictionary is either Fourier (the constant plus 12 paired sine–cosine harmonics, $m = 25$) or Legendre (the first $m = 25$ shifted polynomials), depending on the experiment.

Each logistic parameter k produces one EDMD Koopman matrix in a chosen dictionary and state-space coordinate system. Computing $M \in \{\mathrm{DSA}, \mathrm{CSA}\}$ on every pair across the 10 logistic values yields a 10×10 distance matrix. We build two such matrices: a reference B, in which both sides of every pair share the same configuration, and a comparison D, in which one side has been moved to a different configuration — a different dictionary, or a state-space conjugacy applied to the data — while the other has not. Because every transformed system is, by construction, conjugate to its untransformed counterpart, a conjugacy-invariant method should preserve the reference structure in the comparison: $D \approx B$ throughout. We summarize agreement by the normalized affine RMSE and the Pearson correlation between the entries of D and B; precise definitions for each experiment are in Appendix D and the reference matrices are in Appendix F.

5.1.1 CSA, but not DSA, recognizes systems related by a known state-space conjugacy

We first test whether each method correctly identifies systems that have been put through a known smooth conjugacy. We hold the dictionary fixed at the Fourier dictionary and vary only the state-space coordinates. From each f_k we form the conjugate twin $\tilde{f}_{k,\theta} = h_\theta \circ f_k \circ h_\theta^{-1}$ via the smooth homeomorphism $h_\theta(x) = x + \theta \sin(\pi x)$ with $|\theta| < 1/\pi$, and fit its EDMD Koopman matrix on the conjugated regression pairs $\{(h_\theta(x_i), h_\theta(f_k(x_i)))\}$ in the same Fourier dictionary. The two sides of the comparison therefore produce Koopman matrices for systems that are exactly conjugate by the known map h_θ, but represented in identical observable coordinates — so any departure of D from

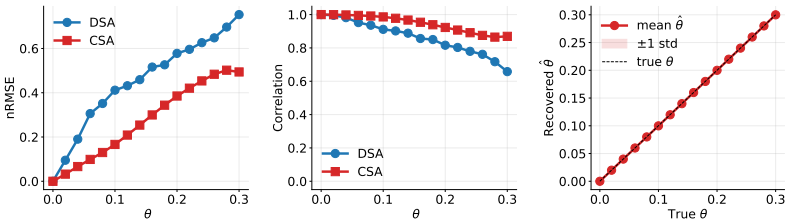


Figure 2: **Robustness to state-space transformations.** (a) nRMSE as a function of θ for the two methods - CSA (red), DSA (blue). (b) Pearson correlation between the entries of the comparison and reference matrices as a function of θ. (c) CSA's recovered conjugacy parameter $\hat{\theta}$ on the ten conjugate pairs (mean $\pm$ s.d.) as a function of θ.

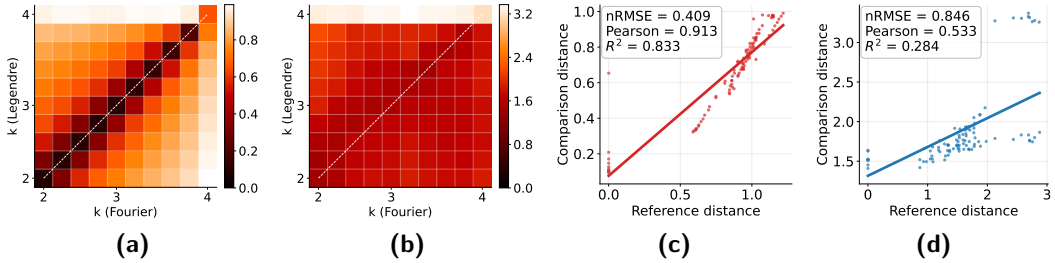


Figure 3: **Cross-dictionary comparison.** Pairwise (comparison) distance matrices for the two methods - CSA (a) and DSA (b), when the same logistic-map family is represented in different dictionaries (Fourier (horizontal axis) vs. Legendre (vertical axis)), $\theta = 0$. Scatter plots of reference and comparison matrix entries for CSA (c) and DSA (d).

B measures how the comparison method handles the homeomorphism. CSA is given the parametric family $\{h_\theta\}_{\theta \in (\frac{-1}{\pi}, \frac{1}{\pi})}$ and searches over θ; DSA searches over orthogonal $Q \in O(m)$.

At a representative $\theta = 0.1$, the comparison matrices D for the two methods, together with their entrywise scatter against B, separate the two methods cleanly (Figure 1). CSA preserves the diagonal banding of the reference and lies tightly along an affine line in the scatter (Fig. 1c), while DSA loses the banding and the scatter is visibly broadened (Fig. 1d). Across the full sweep $\theta \in [0, 0.30]$, DSA's nRMSE rises rapidly and its Pearson correlation falls to about 0.66 by the end of the sweep, whereas CSA's nRMSE grows much more slowly and its correlation stays above 0.86 throughout (Fig. 2a,b). Beyond reporting a small distance, CSA also identifies the homeomorphism responsible for it: the optimized CSA parameter $\hat{\theta}$ on the conjugate pairs $(f_k, \tilde{f}_{k,\theta})$ tracks the diagonal $\hat{\theta} = \theta$ across the sweep (Fig. 2c). Together, these results show that, in this controlled setting, DSA fails as a conjugacy invariant, given that the orthogonal alignment class cannot bridge the non-orthogonal pullback induced by h_θ. CSA, on the other hand, both identifies the conjugacy and quantifies its parameter.

5.1.2 CSA, but not DSA, recovers equivalence when the same dynamics are written in different bases

The first experiment held the dictionary fixed and varied the state-space coordinates. We now do the reverse: we fix the dynamics — the same logistic system f_k on each side — and represent its Koopman operator in two distinct dictionaries. The state-space homeomorphism is the identity, $h = \mathrm{id}$, so a conjugacy-invariant method should recognize the two representations as describing the same operator. The cross-basis comparison matrices D between the Legendre and Fourier dictionaries at matched dimension $m = 25$ separate the two methods (Fig. 3a,b): CSA recovers the reference geometry; DSA does not. The corresponding scatter plots against the reference B (Fig. 3c,d) make the gap quantitative: CSA's nRMSE is 0.409 and its Pearson correlation is 0.913, while DSA's nRMSE is 0.846 and its correlation drops to 0.533.

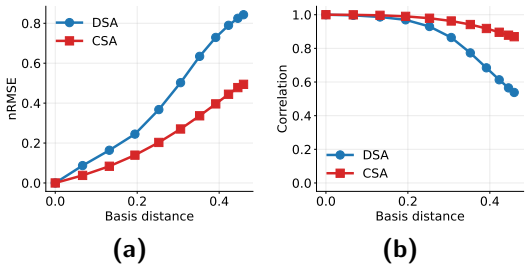


Figure 4: **Robustness to basis mismatch.** (a) nRMSE as a function of basis distance (Appendix E) for the two methods - CSA (red), DSA (blue). (b) Pearson correlation between the entries of the comparison and reference matrices as a function of distance between the bases.

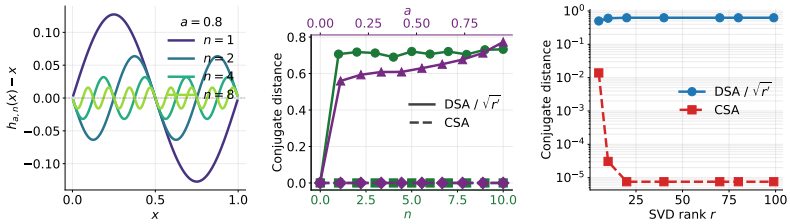


Figure 5: **Hankel-DMD/Krylov comparison under smooth conjugacies.** (a) Deviations $h_{a,n}(x) - x$ for $a = 0.8$. (b) Conjugate-pair distances under sweeps over n at fixed $a = 0.8$ and over a at fixed $n = 4$, at rank $r = 20$. (c) Conjugate-pair distances as a function of rank at $(a, n) = (0.5, 4)$.

We hypothesized that DSA's failure should depend on the degree of misalignment between the two dictionaries' subspaces. To test this, we constructed a parametric family of subspaces interpolating between the Fourier and Legendre endpoints by sampling uniformly along the Grassmannian geodesic between the two subspaces of $L^2([0, 1])$ (construction in Appendix E). For each intermediate subspace, we compute the Koopman operators (for all considered values of k) in that subspace and the corresponding nRMSE and correlation against the Fourier reference. As the projector distance grows, DSA's nRMSE rises nearly linearly and its Pearson correlation drops sharply, whereas CSA remains close to the reference geometry until the projector distance is large (Fig. 4). This pattern is exactly what Section 3 predicts: the two dictionaries describe identical operators, but their subspaces in $L^2([0, 1])$ are not orthogonally related, so no $Q \in O(m)$ recovers the matched-basis distance. The non-orthogonal P found by CSA does.

5.2 Hankel-DMD/Krylov-subspace implementation

We next move to the setting in which DSA is natively defined: Koopman operators learned from time-delay embeddings. From a scalar trajectory we form a Hankel matrix of delay dimension d, truncate its SVD at rank r, and fit the Koopman matrix in the truncated subspace by least squares (Appendix G). Crucially, in this pipeline the observable basis is not specified in advance — it is selected implicitly by the SVD of each system's own Hankel matrix. Two conjugate systems can therefore pick different SVD-truncated bases even when their underlying dynamics are equivalent, putting Hankel-DMD squarely in the regime where Section 3 predicts DSA's orthogonal alignment to be too weak. We refer to standard DSA in this setting as *DSA-Krylov* and to its conjugacy-constrained counterpart as *CSA-Krylov*.

Our reference dynamics are the contracting linear map $f_\rho(y) = \frac{1}{2} + \rho(y - \frac{1}{2})$ with $\rho \in (0, 1)$, which we conjugate by the two-parameter C^∞ family of homeomorphisms

$$h_{a,n}(x) = x - \frac{a}{2\pi n} \sin\big(2\pi n\,(x - \tfrac{1}{2})\big), \qquad n \in \mathbb{Z}_{\geq 1}, \quad a \in [0, 1).$$

Since $h'_{a,n}(x) = 1 - a\cos\big(2\pi n(x - \frac{1}{2})\big) \geq 1 - a > 0$, each $h_{a,n}$ is a diffeomorphism of $[0, 1]$ fixing both endpoints; the amplitude a controls how strongly $h_{a,n}$ departs from the identity, while n sets the number of oscillations of the deviation $h_{a,n}(x) - x$ over the interval (Figure 5(a)). The

conjugate twin of f_ρ is $g_{a,n}^{(\rho)} = h_{a,n}^{-1} \circ f_\rho \circ h_{a,n}$, and we report the average over the two conjugate pairs $(g_{a,n}^{(\rho)}, f_\rho)$ at $\rho \in \{0.5, 0.8\}$. For each system we generate 50 trajectories of length $T = 150$, embed at delay dimension $d = 100$ and fit the Koopman matrix with ridge $\lambda = 10^{-6}$ (Appendix G); convergence under this truncation is verified in Appendix J. CSA-Krylov searches over the family used to build the twin, with $n \in \{0, \ldots, 10\}$ ($n = 0$ denoting the identity) and $a \in [0, 0.95]$. For visual comparison on a common scale, we plot $d_{\mathrm{DSA}}/\sqrt{r'}$ alongside the CSA-Krylov distance, where r' is the dimension of the two matrices compared by DSA (Appendix G); the $1/\sqrt{r'}$ factor is the per-dimension normalization of a Frobenius residual and is irrelevant to the conjugacy-invariance question.

We first ask whether each method's distance is flat under varying conjugacies, as a conjugacy invariant should be. Two sweeps at rank $r = 20$ — the number of oscillations $n \in \{0, \ldots, 10\}$ at fixed amplitude $a = 0.8$, and the amplitude $a \in [0, 0.95]$ at fixed $n = 4$ — make the contrast clear (Fig. 5b). In both sweeps the underlying dynamics is the same contraction f_ρ — only the coordinate chart in which it is represented changes — so any conjugacy invariant should be flat. CSA-Krylov is approximately flat: its distance stays below 10^{-4} over either sweep, and its minimiser recovers the true (a, n) to three decimals at every point where $h_{a,n}$ is not the identity (at the identity the twin coincides with f_ρ, whose Krylov dictionary spans only $\{1, x\}$, and every member of the family attains zero residual). DSA-Krylov, by contrast, rises from nearly zero (at $a = 0$ or $n = 0$, where $h_{a,n}$ is the identity and the two systems coincide) to 0.56–0.77 once the homeomorphism departs from the identity. DSA-Krylov's distance therefore tracks features of the relabelling which have no bearing on conjugacy.

We next confirm that this gap is structural rather than a truncation artefact. The rank sweep at $(a, n) = (0.5, 4)$ on a log scale (Fig. 5c) shows that CSA-Krylov stays at $\sim 10^{-5}$ for $r \geq 10$, while $d_{\mathrm{DSA}}/\sqrt{r'}$ stays between 0.5 and 0.63 at every rank. The mechanism is the one anticipated by §3: DSA fits each system's SVD independently, so the two truncations pick mismatched observable directions that no orthogonal $Q \in O(r)$ can bridge. CSA-Krylov constructs the Krylov subspace using the truncated SVD basis and solves the appropriate basis-transfer problem.

6 Discussion

In this paper we showed that Koopman-based similarity analysis depends critically on the structure of the alignment being approximated. If the target notion is conjugacy, then the relevant alignment between observable spaces is not an arbitrary orthogonal transformation, but the pullback induced by a state-space homeomorphism. This distinction matters both theoretically and empirically. Orthogonal alignment can identify similarities that do not correspond to any valid relabelling of states, and it can miss genuine conjugacies when the induced pullback is non-orthogonal or when the observable bases differ. Thus, a small distance between finite Koopman matrices is not, by itself, evidence of shared dynamics unless the admissible alignment class reflects the intended equivalence relation.

CSA instantiates this principle by restricting alignments to projected composition operators induced by candidate homeomorphisms. In controlled EDMD and Hankel-DMD settings, this was sufficient to recover conjugacy structure in regimes where DSA was sensitive to basis choice, state-space transformations, or SVD-selected observable subspaces. These experiments were intentionally controlled: CSA was given the relevant basis information and a homeomorphism family containing the true map. While this makes the theoretical point explicit, in practice, the space of homeomorphisms might be misspecified, potentially limiting the use of CSA. This problem is considered in Appendix K where we bound the effect of misspecifying the family. To further gain traction over the practical viability of CSA, in Appendix L we test the robustness of the score to noise in the measurement of the observables.

Nonetheless, the goal for this paper was not to present a fully automated comparison pipeline, but to isolate the mathematical issue that any conjugacy-aware pipeline must confront. Holding the learned operators fixed, the difference between CSA and DSA comes from the distance being optimized, and therefore from the structural assumptions each distance makes about equivalence.

This also clarifies what Koopman methods can and cannot buy us. Passing to linear operators is powerful, but it does not eliminate the need to reason about the state-space correspondence. The conjugacy reappears as a composition operator on observables, and its finite-dimensional projection

is the object that must align the learned Koopman representations. In this sense, CSA lies between direct state-space matching and unconstrained operator matching: it uses the operator representation, but preserves the state-space map that gives the comparison its dynamical meaning.

Conjugacy is a strong notion of equivalence, and other scientific goals may call for weaker notions. For example, two systems may not be conjugate but may nevertheless share predictive structure under a fixed set of delays, or implement similar input–output maps in a task-relevant subspace. Such similarities may be meaningful, but they should not be confused with conjugacy. Likewise, spectral similarity or orthogonal matrix similarity may reveal useful structure without licensing claims about shared dynamics. The central question is therefore not simply how to design more flexible similarity metrics, but how to match the metric to the equivalence relation that the scientific claim requires.

CSA is not a metric in the formal sense, and its conclusions depend on the candidate homeomorphism family and observable spaces supplied to it. These are limitations of the present instantiation, but they are also part of the point: any method that aims to detect conjugacy must make assumptions about admissible relabellings and about how observables transform across systems. Making those assumptions explicit is preferable to hiding them inside an unconstrained alignment problem.

The main contribution of this work is therefore a clarification with practical consequences. Koopman-based comparisons should not treat finite operator matrices as basis-free objects whose similarity can be interpreted independently of the observable spaces and transformations that produced them; identifying the pullback structure required by conjugacy shows why DSA-style orthogonal alignment can be misleading and points to a principled alternative. This matters wherever conjugacy is the claim being made — across recording sessions for BCI alignment, across animals for cross-subject comparison, between models and data for model validation, or across architectures and training regimes in machine learning. In each case, before asking whether two systems are similar, we must ask in what sense they are meant to be the same.

Acknowledgements

The authors acknowledge Alex Williams for useful discussions on the subject. The authors further acknowledge the help of anonymous reviewers that aided in making the argument more robust. P.G. further acknowledges Paul Liebenow for introducing him to the Koopman operator formalism and Devarshi Mukherjee for helping navigate some basic literature on operator theory during the initial stages of this work. P.G. was funded by the Walter Benjamin Programme of the Deutsche Forschungsgemeinschaft (DFG), project number GO 4136/1-1, which supported this work.

References

Mitchell Ostrow, Adam Eisen, Leo Kozachkov, and Ila Fiete. Beyond geometry: Comparing the temporal structure of computation in neural circuits with dynamical similarity analysis. *Advances in Neural Information Processing Systems*, 36:33824–33837, 2023.

Martin Schrimpf, Idan Asher Blank, Greta Tuckute, Carina Kauf, Eghbal A. Hosseini, Nancy Kanwisher, Joshua B. Tenenbaum, and Evelina Fedorenko. The neural architecture of language: Integrative modeling converges on predictive processing. *Proceedings of the National Academy of Sciences*, 118(45):e2105646118, 2021. doi: 10.1073/pnas.2105646118. URL `https://www.pnas.org/doi/abs/10.1073/pnas.2105646118`.

Marino Pagan, Vincent D Tang, Mikio C Aoi, Jonathan W Pillow, Valerio Mante, David Sussillo, and Carlos D Brody. A new theoretical framework jointly explains behavioral and neural variability across subjects performing flexible decision-making. *bioRxiv*, pages 2022–11, 2022.

Alan D Degenhart, William E Bishop, Emily R Oby, Elizabeth C Tyler-Kabara, Steven M Chase, Aaron P Batista, and Byron M Yu. Stabilization of a brain–computer interface via the alignment of low-dimensional spaces of neural activity. *Nature biomedical engineering*, 4(7):672–685, 2020.

Brianna M Karpowicz, Yahia H Ali, Lahiru N Wimalasena, Andrew R Sedler, Mohammad Reza Keshtkaran, Kevin Bodkin, Xuan Ma, Daniel B Rubin, Ziv M Williams, Sydney S Cash, et al. Stabilizing brain-computer interfaces through alignment of latent dynamics. *Nature Communications*, 16(1):4662, 2025.

Rishidev Chaudhuri, Berk Gerçek, Biraj Pandey, Adrien Peyrache, and Ila Fiete. The intrinsic attractor manifold and population dynamics of a canonical cognitive circuit across waking and sleep. *Nature neuroscience*, 22(9):1512–1520, 2019.

Nikolaus Kriegeskorte, Marieke Mur, and Peter A Bandettini. Representational similarity analysis-connecting the branches of systems neuroscience. *Frontiers in systems neuroscience*, 2:249, 2008.

Maithra Raghu, Justin Gilmer, Jason Yosinski, and Jascha Sohl-Dickstein. Svcca: Singular vector canonical correlation analysis for deep learning dynamics and interpretability. *Advances in neural information processing systems*, 30, 2017.

Alex H Williams, Erin Kunz, Simon Kornblith, and Scott Linderman. Generalized shape metrics on neural representations. *Advances in neural information processing systems*, 34:4738–4750, 2021.

Bernard O Koopman. Hamiltonian systems and transformation in hilbert space. *Proceedings of the National Academy of Sciences*, 17(5):315–318, 1931.

Miklós Rédei and Charlotte Werndl. On the history of the isomorphism problem of dynamical systems with special regard to von neumann's contribution. *Archive for history of exact sciences*, 66(1):71–93, 2012.

Marko Budišić, Ryan Mohr, and Igor Mezić. Applied koopmanism. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 22(4), 2012.

Hassan Arbabi and Igor Mezic. Ergodic theory, dynamic mode decomposition, and computation of spectral properties of the koopman operator. *SIAM Journal on Applied Dynamical Systems*, 16 (4):2096–2126, 2017.

Matthew O. Williams, Ioannis G. Kevrekidis, and Clarence W. Rowley. A data–driven approximation of the koopman operator: Extending dynamic mode decomposition. *Journal of Nonlinear Science*, 25(6):1307–1346, 2015.

Ruiqi Chen, Giacomo Vedovati, Todd Braver, and ShiNung Ching. Dform: Diffeomorphic vector field alignment for assessing dynamics across learned models. *arXiv preprint arXiv:2402.09735*, 2024.

Hao Huang, Boulbaba Ben Amor, Xichan Lin, Fan Zhu, and Yi Fang. Residual networks as flows of velocity fields for diffeomorphic time series alignment. *arXiv preprint arXiv:2106.11911*, 2021.

William T Redman, Juan Bello-Rivas, Maria Fonoberova, Ryan Mohr, Yannis G Kevrekidis, and Igor Mezić. Identifying equivalent training dynamics. *Advances in Neural Information Processing Systems*, 37:23603–23629, 2024.

Igor Mezic. Spectrum of the koopman operator, spectral expansions in functional spaces, and state-space geometry. *J. Nonlinear Sci.*, 30(5):2091–2145, 2020.

Milan Korda and Igor Mezić. On convergence of extended dynamic mode decomposition to the koopman operator. *Journal of Nonlinear Science*, 28(2):687–710, 2018.

A Unitary operators need not be composition operators

A composition operator $\mathcal{C}_h$ induced by a homeomorphism h acts on observables by

$$(\mathcal{C}_h f)(x) = f(h(x)).$$

Such operators preserve pointwise products:

$$\mathcal{C}_h(fg)(x) = (fg)(h(x)) = f(h(x))\, g(h(x)) = (\mathcal{C}_h f)(x)\, (\mathcal{C}_h g)(x). \tag{13}$$

Multiplicativity is a necessary property of composition operators[1].

Now consider the two-dimensional subspace

$$V = \mathrm{span}\{1, x\} \subset L^2([-1, 1]),$$

with orthonormal basis

$$e_1(x) = \frac{1}{\sqrt{2}}, \qquad e_2(x) = \sqrt{\frac{3}{2}}\, x.$$

Define $U : V \to V$ by

$$Ue_1 = e_2, \qquad Ue_2 = e_1.$$

Its matrix in the basis $\{e_1, e_2\}$ is

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

so U is unitary. However, U is not a composition operator, because it does not preserve products. Indeed,

$$e_1 e_2 = \frac{1}{\sqrt{2}} e_2,$$

so

$$U(e_1 e_2) = \frac{1}{\sqrt{2}} e_1,$$

whereas

$$(Ue_1)(Ue_2) = e_2 e_1 = e_1 e_2 = \frac{1}{\sqrt{2}} e_2.$$

Hence

$$U(e_1 e_2) \neq (Ue_1)(Ue_2).$$

Therefore U is unitary but not a composition operator.

B Spectral isomorphism doesn't imply dynamical conjugacy

Proposition 1. Let

$$X := \{0, 1\}^{\mathbb{Z}} \times \{0, 1\}, \qquad Y := \{0, 1, 2\}^{\mathbb{Z}} \times \{0, 1\},$$

equipped with the product probability measures

$$\mu := \mu_2 \times \eta, \qquad \nu := \mu_3 \times \eta,$$

where μ_2 and μ_3 are the Bernoulli product measures on $\{0, 1\}^{\mathbb{Z}}$ and $\{0, 1, 2\}^{\mathbb{Z}}$, respectively, and η is the uniform measure on $\{0, 1\}$. Define

$$T := \sigma_2 \times \mathrm{id}, \qquad S := \sigma_3 \times \mathrm{id},$$

where σ_2 and σ_3 are the two-sided Bernoulli shifts. Then the Koopman operators $\mathcal{K}_T$ and $\mathcal{K}_S$ are unitarily equivalent, but T and S are not conjugate.

[1] This is precisely the reason why the Eigenfunctions of the Koopman operator form a multiplicative lattice [Budišić et al., 2012].

Proof. We prove the two claims separately.

First, $\mathcal{K}_T$ and $\mathcal{K}_S$ are unitarily equivalent. For $k = 2, 3$, let

$$X_k := \{0, 1, \dots, k-1\}^{\mathbb{Z}}, \qquad (\sigma_k x)_n = x_{n+1},$$

with μ_k the corresponding Bernoulli product measure. It is a standard fact that

$$L^2(X_k, \mu_k) = \mathbb{C}\mathbf{1} \oplus \bigoplus_{j=1}^{\infty} \mathcal{H}_j^{(k)},$$

where on each $\mathcal{H}_j^{(k)}$, the Koopman operator $\mathcal{K}_{\sigma_k}$ acts as a copy of the bilateral shift on $\ell^2(\mathbb{Z})$. Since this decomposition is the same for $k = 2$ and $k = 3$, there exists a unitary

$$W : L^2(X_2, \mu_2) \to L^2(X_3, \mu_3)$$

such that

$$W\mathcal{K}_{\sigma_2} = \mathcal{K}_{\sigma_3} W.$$

Now

$$L^2(X, \mu) \cong L^2(X_2, \mu_2) \oplus L^2(X_2, \mu_2), \qquad L^2(Y, \nu) \cong L^2(X_3, \mu_3) \oplus L^2(X_3, \mu_3),$$

and under these identifications,

$$\mathcal{K}_T = \mathcal{K}_{\sigma_2} \oplus \mathcal{K}_{\sigma_2}, \qquad \mathcal{K}_S = \mathcal{K}_{\sigma_3} \oplus \mathcal{K}_{\sigma_3}.$$

Hence $\widetilde{W} := W \oplus W$ is unitary and satisfies

$$\widetilde{W}\mathcal{K}_T = \mathcal{K}_S \widetilde{W}.$$

Thus $\mathcal{K}_T$ and $\mathcal{K}_S$ are unitarily equivalent.

Next, T and S are not conjugate. If there were a homeomorphism $h : X \to Y$ with

$$h \circ T = S \circ h,$$

then T and S would have the same number of periodic points of each period. But

$$\#\mathrm{Fix}(\sigma_k^n) = k^n,$$

since an n-periodic bi-infinite sequence is determined by a word of length n. Therefore

$$\#\mathrm{Fix}(T^n) = 2 \cdot 2^n, \qquad \#\mathrm{Fix}(S^n) = 2 \cdot 3^n.$$

These are unequal for every $n \geq 1$, so no such h can exist. Hence T and S are not conjugate. □

B.1 Why the Bernoulli shifts are spectrally isomorphic

For completeness, we briefly indicate the spectral decomposition used above. For $k = 2, 3$, write

$$X_k = \{0, 1, \dots, k-1\}^{\mathbb{Z}}, \qquad \mathcal{H}_k = L^2(X_k, \mu_k).$$

Let

$$L^2(\{0, 1, \dots, k-1\}, p_k) = \mathbb{C}\mathbf{1} \oplus \mathcal{H}_k^0,$$

where p_k is the uniform measure on the alphabet, and choose an orthonormal basis

$$e_1^{(k)}, \dots, e_{k-1}^{(k)}$$

for $\mathcal{H}_k^0$.

For $n \in \mathbb{Z}$, define the coordinate observables

$$e_{n,\alpha}^{(k)}(x) := e_\alpha^{(k)}(x_n).$$

More generally, for a finite set $J = \{n_1 < \cdots < n_r\} \subset \mathbb{Z}$ and multi-index $\alpha = (\alpha_1, \dots, \alpha_r)$, define

$$\psi_{J,\alpha}^{(k)}(x) := \prod_{j=1}^{r} e_{\alpha_j}^{(k)}(x_{n_j}).$$

Together with the constant function 1, these form an orthonormal basis of $\mathcal{H}_k$.

The Koopman operator acts by translation of the index set:

$$\mathcal{K}_{\sigma_k}\psi^{(k)}_{J,\alpha} = \psi^{(k)}_{J+1,\alpha}, \qquad J+1 := \{n+1 : n \in J\}.$$

Hence each translation orbit of a finite pattern generates an invariant subspace on which $\mathcal{K}_{\sigma_k}$ acts as the bilateral shift on $\ell^2(\mathbb{Z})$. Since there are countably many such orbits for both $k = 2$ and $k = 3$, both operators decompose as

$$\mathbb{C}\mathbf{1} \oplus \bigoplus_{j=1}^{\infty} (\text{copy of bilateral shift}).$$

Thus $\mathcal{K}_{\sigma_2}$ and $\mathcal{K}_{\sigma_3}$ are unitarily equivalent.

The important point is that this intertwiner is an abstract unitary operator on L^2. It is not induced by a homeomorphism of state spaces. This is precisely why spectral equivalence does not imply conjugacy. It is straightforward to see that CSA doesn't mistakenly qualify them as conjugate as it searches over the space of valid homeomorphisms.

Remark 1. There is some existing literature proposing that spectral conjugacy can, in some cases, imply topological conjugacy. In Section 3.3 of [Redman et al., 2024], it is claimed that spectral conjugacy within basins of attraction is sufficient for topological conjugacy. This is based on the proof in [Mezic, 2020] that within a basin of attraction, the Koopman operator is topologically conjugate to a Koopman operator of a linear dynamical system, i.e. one given by exponential decay. Paraphrased, the claim is that within basins of attraction $\mathcal{B}_1$ and $\mathcal{B}_2$, and coordinate charts $h_i : \mathcal{B}_i \to \mathbb{R}^n$ respectively such that $U^t_i(h_i(x)) = e^{A_i t}h_i(x)$, where A_i is some matrix. Then there exists some linear map $M : \mathbb{R}^n \to \mathbb{R}^n$ such that $h_1^{-1} \circ M \circ h_2$ is a topological conjugacy between the basins of attraction. In both cases, the components of the map h_i give eigenfunctions of their respective Koopman operators.

Let us grant that "interesting" neural networks converge to some solution, so the relevant dynamics occur in some basin of attraction. From a modelling perspective, it is not a priori clear that the eigenfunctions as generated would span the Krylov subspace of the Koopman operator, even if the Krylov subspace is generated from within the basin of attraction.

B.2 Trivial extension to non-ergodic and non-measure-preserving systems

The counter-example above can be trivially generalized to non-measure-preserving systems as well. This is done by extending the state-space to include a non-measure-preserving map that is identical for both the systems. For instance,

$$y_{t+1} = C(y_t) := \frac{y_t}{2}.$$

Define

$$\widehat{T} := \sigma_2 \times C, \qquad \widehat{S} := \sigma_3 \times C.$$

The lack of conjugacy still follows from the Bernoulli component, since the periodic point counts in the shift coordinate remain different. The point of this variant is simply that the obstruction is not specific to the measure-preserving setting.

C When is P unitary?

Since P is the matrix representation of $\mathcal{C}_h$ in orthonormal bases, P is unitary if and only if $\mathcal{C}_h$ is unitary. Equivalently,

$$P^*P = I \qquad \iff \qquad \langle \mathcal{C}_h u, \mathcal{C}_h v\rangle_X = \langle u, v\rangle_Y \quad \text{for all } u, v. \tag{14}$$

If $\mathcal{H}_X = L^2(X, \mu_X)$ and $\mathcal{H}_Y = L^2(Y, \mu_Y)$, this is equivalent to

$$\int_X |u(h(x))|^2 \, d\mu_X(x) = \int_Y |u(y)|^2 \, d\mu_Y(y) \qquad \text{for all } u, \tag{15}$$

or, equivalently,

$$h_{\#}\mu_X = \mu_Y. \tag{16}$$

Thus P is unitary precisely when h preserves the relevant measure. In the special case where $X, Y \subset \mathbb{R}^d$ and both observable spaces use Lebesgue measure, a smooth invertible map h is unitary only if

$$|\det Dh(x)| = 1 \qquad \text{a.e. on } X. \tag{17}$$

Thus many perfectly valid conjugacies do not induce unitary operators. For example, the affine map

$$h(x) = ax + b \tag{18}$$

is unitary on L^2 with Lebesgue measure only when $|a| = 1$. Likewise, the map

$$h_\theta(x) = x + \theta \sin(\pi x), \qquad x \in [0, 1], \tag{19}$$

has derivative

$$h'_\theta(x) = 1 + \theta\pi \cos(\pi x), \tag{20}$$

which is not identically of unit modulus unless $\theta = 0$. Therefore, although h_θ may be invertible for sufficiently small θ, the induced operator $\mathcal{C}_{h_\theta}$ is generally not unitary, and hence neither is its matrix representation P_θ.

It is therefore important to distinguish invertibility of the conjugacy from unitarity of the induced composition operator. Conjugacy only requires h to be bijective and to satisfy $h \circ f = g \circ h$. Unitarity is an additional and much stronger requirement.

As a simple example where P is unitary, consider a mirror reflection. Let $X = Y = [-1, 1]$, let $h(x) = -x$, and suppose g is the reflected version of f, i.e.

$$g = h \circ f \circ h^{-1}. \tag{21}$$

Then

$$(\mathcal{C}_h u)(x) = u(-x). \tag{22}$$

Since the reflection $x \mapsto -x$ preserves Lebesgue measure,

$$\|\mathcal{C}_h u\|^2_{L^2([-1,1])} = \int_{-1}^{1} |u(-x)|^2 \, dx = \int_{-1}^{1} |u(y)|^2 \, dy = \|u\|^2_{L^2([-1,1])}, \tag{23}$$

where we used the change of variables $y = -x$. Hence $\mathcal{C}_h$ is unitary, and therefore P is unitary in any pair of orthonormal bases.

If, moreover, the basis on X is chosen as the reflected image of the basis on Y, i.e.

$$\phi_i = \mathcal{C}_h \psi_i = \psi_i(-x), \tag{24}$$

then the matrix representation of $\mathcal{C}_h$ is simply

$$P = I. \tag{25}$$

If instead the same basis is used on both sides, then P need not be the identity, but it will still be unitary. For example, in a polynomial or Fourier basis, reflection acts by parity, so P becomes diagonal with entries ± 1.

D Construction of comparison and reference distance matrices

Let $\{k_1, \ldots, k_N\}$ be the set of logistic-map parameters under consideration. For each k_i, let

$$K_i^{(\Phi,\theta)}$$

denote the finite-dimensional Koopman matrix obtained from the system at parameter k_i, represented in dictionary Φ, with state-space transformation parameter θ. Here $\theta = 0$ denotes the untransformed system.

For a comparison method $M \in \{\mathrm{DSA}, \mathrm{CSA}\}$, the associated pairwise distance matrix is

$$D^M\big[(\Phi_1, \theta_1), (\Phi_2, \theta_2)\big] \in \mathbb{R}^{N \times N},$$

with entries

$$D^M\big[(\Phi_1, \theta_1), (\Phi_2, \theta_2)\big]_{ij} = d_M\Big(K_i^{(\Phi_1,\theta_1)}, K_j^{(\Phi_2,\theta_2)}\Big).$$

The matrix denoted by D in the main text is the comparison matrix relevant to the experiment under consideration, while B is the corresponding matched reference matrix.

State-space transformation experiment. When the dictionary is fixed to Φ and one side is transformed by h_θ, the comparison matrix is

$$D = D^M\big[(\Phi, 0), (\Phi, \theta)\big],$$

and the matched reference matrix is the symmetrized within-transformation baseline

$$B = \frac{1}{2}\left(D^M\big[(\Phi, 0), (\Phi, 0)\big] + D^M\big[(\Phi, \theta), (\Phi, \theta)\big]\right).$$

Cross-dictionary experiment. When the underlying dynamics are unchanged but the dictionaries differ, the comparison matrix is

$$D = D^M\big[(\Phi, 0), (\Psi, 0)\big].$$

and the matched reference matrix is the symmetrized within-dictionary baseline

$$B = \frac{1}{2}\left(D^M\big[(\Phi, 0), (\Phi, 0)\big] + D^M\big[(\Psi, 0), (\Psi, 0)\big]\right).$$

For the basis-mismatch sweep (Figure 4), where $\Psi = \Phi^{(t)}$ (Appendix E), the reference is instead held fixed at the Fourier endpoint, $B = D^M\big[(\Phi, 0), (\Phi, 0)\big]$.

Matrix-level summary statistics. Given a comparison matrix D and reference matrix B, we use all N^2 entries of both, including the diagonal, and fit the best affine map from B to D. The reported normalized affine RMSE measures the residual error after this affine alignment, and the Pearson correlation quantifies the extent to which the pairwise geometry is preserved up to affine rescaling.

E Distance between observable subspaces

This appendix details two ingredients of the basis-change experiment of Section 5.1.2: the construction of a one-parameter family of m-dimensional observable subspaces interpolating between the Fourier and Legendre endpoint dictionaries, and the projector distance used to label each member of that family on the horizontal axis of Figure 4.

Subspace family between Fourier and Legendre. The two endpoint dictionaries are taken at matched dimension $m = 25$: the Fourier dictionary contains the constant together with the first twelve cosine/sine harmonics, and the Legendre dictionary contains the first m Legendre polynomials on $[0, 1]$. We embed both into a common ambient dictionary by simple concatenation, of dimension $p = 2m$, and evaluate it on a dense uniform grid $\{x_\ell\}_{\ell=1}^N$ of $N = 4000$ points in $[0, 1]$ to obtain the ambient feature matrix $A \in \mathbb{R}^{N\times p}$. The empirical Gram matrix

$$\hat{G} \;=\; \frac{1}{N}\, A^\top A \;\in\; \mathbb{R}^{p\times p} \tag{26}$$

is the discrete approximation of the $L^2([0, 1])$ inner product $\langle \phi_i, \phi_j\rangle = \int_0^1 \phi_i(x)\phi_j(x)\,\mathrm{d}x$ on the ambient features under Lebesgue measure on $[0, 1]$. All inner products and orthogonality statements below are with respect to $\hat{G}$; the dense uniform grid sets the discretisation, and refining it recovers the continuum L^2 pairing.

Let $S = \hat{G}^{1/2}$. Pre-multiplication by S is the change of variables that turns $\hat{G}$-orthonormality into Euclidean orthonormality: a coefficient matrix $C \in \mathbb{R}^{p\times m}$ satisfies $C^\top \hat{G} C = I_m$ iff $U = SC$ satisfies $U^\top U = I_m$. Let $C_F, C_L \in \mathbb{R}^{p\times m}$ be the selector matrices that pick the Fourier and Legendre blocks of the ambient dictionary, re-orthonormalised in the $\hat{G}$ inner product. The two endpoint subspaces are then represented by the Euclidean-orthonormal frames

$$U_F = SC_F, \qquad U_L = SC_L \;\in\; \mathbb{R}^{p\times m}. \tag{27}$$

For $t \in [0, 1]$ the intermediate frame is the principal geodesic from U_F to U_L on the Grassmannian $\mathrm{Gr}(m, p)$,

$$U(t) \;=\; U_F E_1 \cos(t\Omega) + W \sin(t\Omega), \tag{28}$$

where $U_F^\top U_L = E_1 \cos\Omega\, E_2^\top$ is the thin SVD, $\Omega = \mathrm{diag}(\omega_1, \ldots, \omega_m)$ collects the principal angles between U_F and U_L, and $W = (U_L E_2 - U_F E_1 \cos\Omega)\sin^{-1}\Omega$ is the orthonormal completion. By

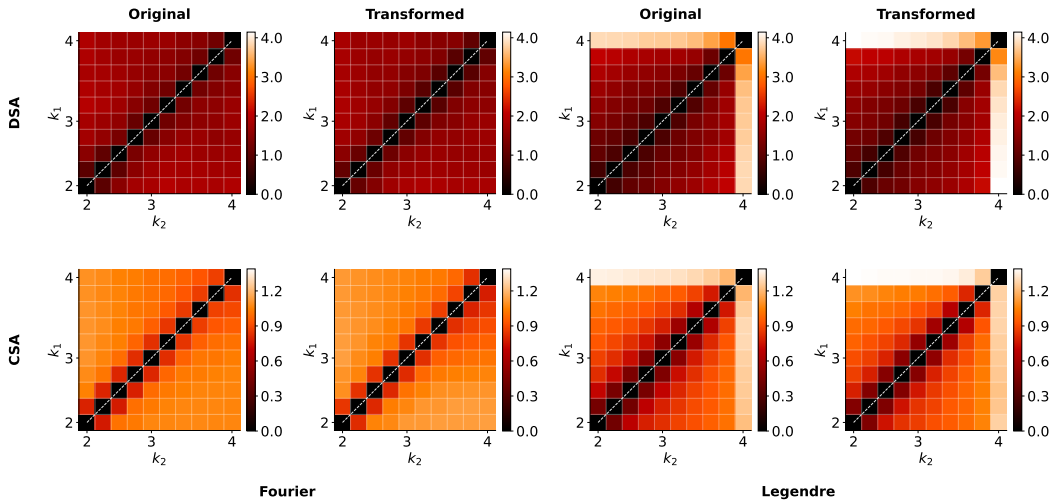


Figure 6: Pairwise distance matrices of within transformation pairs for both DSA (top) and CSA (bottom). k_i represents the k parameter of the Logistic map for the ith element in the pair as in the main text.

construction $U(0) = U_F E_1$ and $U(1) = U_L E_2$, and $U(t)$ remains Euclidean-orthonormal in the ambient space throughout. The corresponding intermediate dictionary, in L^2-coefficient form, is recovered by inverting the change of variables, $C(t) = S^{-1}U(t)$, and re-orthonormalising in the $\hat{G}$ inner product; on any evaluation grid the dictionary itself is $\Phi^{(t)}(x) = A_{\text{ambient}}(x)\, C(t)$. The Koopman operators reported in Figure 4 are fitted by the same EDMD procedure of Section 5.1 using $\Phi^{(t)}$ as the dictionary.

Projector distance between subspaces. Given two Euclidean-orthonormal frames $U_1, U_2 \in \mathbb{R}^{p\times m}$ in the ambient feature space, the basis distance we report is the normalised Frobenius distance between the corresponding orthogonal projectors,

$$d_{\text{basis}}(U_1, U_2) \;=\; \frac{\|U_1 U_1^\top - U_2 U_2^\top\|_F}{\sqrt{2m}}. \tag{29}$$

It vanishes when the two m-dimensional subspaces coincide and equals 1 when they are mutually orthogonal. In terms of the principal angles $\omega_1, \ldots, \omega_m$ between the subspaces it admits the closed form $d^2_{\text{basis}} = \frac{1}{m}\sum_{i=1}^m \sin^2\omega_i$, so it coincides, up to the $1/\sqrt{m}$ normalisation, with the chordal Grassmann distance on $\mathrm{Gr}(m,p)$. Because the endpoint frames (27) carry the empirical Gram square root S, the Euclidean projector distance (29) computed on $U_F, U(t), U_L$ is exactly the projector distance between the underlying function subspaces in $L^2([0,1])$ under the empirical inner product (26). The horizontal axis of Figure 4 is $d_{\text{basis}}(U_F, U(t))$, which increases monotonically with t from zero at $t = 0$ to the endpoint distance $d_{\text{basis}}(U_F, U_L)$ at $t = 1$.

F Intra-transformation distance matrices

In Figure 6 we show that both DSA and CSA are able to identify conjugacy across two Logistic systems when both of those systems are represented in the same basis Φ (either Fourier or Legendre) and they share the same state-space (either x (called original) or $\hat{x} = h_\theta(x)$ (called transformed)). These matrices presented in this figure are used to generate the reference matrices in Appendix D. In particular we have plotted

$$D^M\big[(\Phi,\theta),(\Phi,\theta)\big] \in \mathbb{R}^{N\times N},$$

for $M \in \{\text{DSA}, \text{CSA}\}$.

G Krylov subspace construction via delay-embedding SVD

This appendix details how the Koopman operators compared by DSA-Krylov and CSA-Krylov in Section 5.2 are constructed from scalar time-series data. The procedure follows the Hankel-DMD framework [Arbabi and Mezic, 2017, Ostrow et al., 2023], which we summarise here for completeness and to fix notation.

G.1 Delay-embedding and the Hankel matrix

Let $f\colon [0,1] \to [0,1]$ be a one-dimensional map with Koopman operator $\mathcal{U}$, and let $g(x) = x$ be the identity observable. From a scalar time series $\{x_0, x_1, \ldots, x_{T-1}\}$ with $x_{t+1} = f(x_t)$, we form the *delay-embedding* (Hankel) matrix with embedding dimension d:

$$H = \begin{pmatrix} x_0 & x_1 & \cdots & x_{d-1} \\ x_1 & x_2 & \cdots & x_d \\ \vdots & \vdots & \ddots & \vdots \\ x_{T-d} & x_{T-d+1} & \cdots & x_{T-1} \end{pmatrix} \in \mathbb{R}^{N \times d}, \quad N = T - d + 1. \tag{30}$$

Each row is a delay-embedded state vector $z_t = (x_t, x_{t+1}, \ldots, x_{t+d-1})^\top \in \mathbb{R}^d$, and each column j contains the Krylov observable $\mathcal{U}^j g$ evaluated along the orbit: $H_{\cdot,j} = (g(x_j), g(x_{j+1}), \ldots, g(x_{j+N-1}))^\top$. The column space of H therefore spans the data-driven Krylov subspace $\mathcal{K}_d(\mathcal{U}, g) = \operatorname{span}\{g, \mathcal{U}g, \ldots, \mathcal{U}^{d-1}g\}$.

When multiple independent trajectories are available, each is embedded separately (so that no delay window crosses a trajectory boundary) and the resulting matrices are stacked vertically, increasing N while keeping d fixed.

G.2 SVD truncation

We compute the singular value decomposition

$$H = U \, \Sigma \, V^\top, \tag{31}$$

with $U \in \mathbb{R}^{N \times d}$, $\Sigma = \operatorname{diag}(\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_d)$, and $V \in \mathbb{R}^{d \times d}$ orthogonal. Truncating at rank $r \leq d$ defines the r-dimensional subspace

$$V_r = V_{:,\,1:r} \in \mathbb{R}^{d \times r}, \tag{32}$$

whose columns form an orthonormal basis for the leading r right-singular directions. The fraction of total energy captured is $\sum_{i=1}^{r} \sigma_i^2 \big/ \sum_{i=1}^{d} \sigma_i^2$.

Both DSA-Krylov and CSA-Krylov use the leading $r' = \min(r, \#\{k : \sigma_k > 10^{-8}\sigma_1\})$ directions, normalised by the singular values. The projected Krylov observables at a delay-embedded point z are

$$\phi(z) = \Sigma_{r'}^{-1} V_{r'}^\top z \in \mathbb{R}^{r'}, \tag{33}$$

which are orthonormal over the training data and coincide with the eigen-time-delay coordinates of the DSA implementation [Ostrow et al., 2023]. The dimension r' may differ between systems; we write r'_i for system i.

G.3 Koopman operator in the truncated subspace

Partition the delay vectors into current and next-step pairs:

$$Z = (z_0, z_1, \ldots, z_{N-2})^\top \in \mathbb{R}^{(N-1)\times d}, \quad Z' = (z_1, z_2, \ldots, z_{N-1})^\top \in \mathbb{R}^{(N-1)\times d}, \tag{34}$$

and project both onto the truncated subspace, $A = Z V_{r'} \Sigma_{r'}^{-1}$ and $B = Z' V_{r'} \Sigma_{r'}^{-1}$ in $\mathbb{R}^{(N-1)\times r'}$, so that the rows of A and B are $\phi(z_t)^\top$ and $\phi(z_{t+1})^\top$ respectively. Writing the Koopman matrix $K \in \mathbb{R}^{r' \times r'}$ in the column-vector convention $\phi(z_{t+1}) \approx K\phi(z_t)$, the regularised least-squares fit minimises $\|AK^\top - B\|_F^2 + \lambda \|K^\top\|_F^2$,

$$K^\top = \left(A^\top A + \lambda I_{r'}\right)^{-1} A^\top B. \tag{35}$$

G.4 DSA-Krylov: orthogonal alignment

With K_1, K_2 fitted via (35), the smaller one zero-padded to the size of the larger (so that $r = \max(r_1', r_2')$ in (36)), DSA-Krylov seeks the orthogonal matrix that best aligns them:

$$d_{\text{DSA-Krylov}}(K_1, K_2) = \min_{Q \in O(r)} \|K_1 - Q\, K_2\, Q^\top\|_F, \tag{36}$$

solved by Riemannian optimisation on $O(r)$ with spectral initialisation [Ostrow et al., 2023].

Since the two systems may have been fitted in different SVD bases $V_{1,r}$ and $V_{2,r}$, the true basis change need not be orthogonal even when the systems are dynamically equivalent. Restricting to $O(r)$ therefore conflates coordinate artefacts with dynamical differences.

G.5 CSA-Krylov: conjugacy-constrained basis change

CSA-Krylov replaces the orthogonal search with a basis change *derived* from a candidate conjugacy h_θ, in two steps.

Step 1: regression for P_θ. Fix a dense grid $\{x_i\}_{i=1}^M \subset (0,1)$ and evaluate the projected Krylov observables for both systems:

$$\Phi_1 = \big[\phi_1(x_1), \ldots, \phi_1(x_M)\big]^\top \in \mathbb{R}^{M \times r_1'}, \tag{37}$$

$$\Phi_2 = \big[\phi_2(h_\theta(x_1)), \ldots, \phi_2(h_\theta(x_M))\big]^\top \in \mathbb{R}^{M \times r_2'}, \tag{38}$$

where $\phi_i(x) = \Sigma_{i,r'}^{-1} V_{i,r'}^\top\, (x, f_i(x), \ldots, f_i^{d-1}(x))$ uses the known dynamics to build Krylov observables at arbitrary evaluation points. The basis-change matrix is then

$$P_\theta = \big[(\Phi_1^\top \Phi_1 + \lambda I)^{-1}\, \Phi_1^\top\, \Phi_2\big]^\top \in \mathbb{R}^{r_2' \times r_1'}. \tag{39}$$

Step 2: intertwining residual. The CSA-Krylov distance is the normalised intertwining error minimised over the conjugacy family:

$$d_{\text{CSA-Krylov}}(K_1, K_2) = \min_\theta\ \frac{\|P_\theta\, K_1 - K_2\, P_\theta\|_F}{\|P_\theta\|_F}. \tag{40}$$

For each integer n, the minimum over a is found by a coarse grid search followed by bounded Brent refinement around the best coarse candidates (Algorithm 1); the reported distance is the minimum over n.

H The CSA transfer matrix as a projected composition operator

We now make precise the sense in which the CSA transfer matrix is a finite-dimensional projection of the composition operator induced by a candidate state-space map. This is the direct analogue, for the basis-transfer operator, of the standard EDMD fact that the fitted Koopman matrix is an empirical projection of the true Koopman operator and converges to the corresponding population projection as the number of samples grows (see [Korda and Mezić, 2018]).

Let (X, μ_X) and (Y, μ_Y) be probability spaces, and let

$$\mathcal{H}_X = L^2(X, \mu_X), \qquad \mathcal{H}_Y = L^2(Y, \mu_Y).$$

Let $h_\theta : X \to Y$ be measurable, and define the pullback, or composition, operator

$$\mathcal{C}_{h_\theta} : \mathcal{H}_Y \to \mathcal{H}_X, \qquad (\mathcal{C}_{h_\theta} u)(x) = u(h_\theta(x)).$$

Throughout this subsection, we assume that $\mathcal{C}_{h_\theta} \psi_j \in L^2(X, \mu_X)$ for the finite collection of observables used below. For the complete-basis convergence statement at the end, we assume the stronger condition that $\mathcal{C}_{h_\theta}$ is a bounded operator from $\mathcal{H}_Y$ to $\mathcal{H}_X$.

Let

$$V_\Phi = \operatorname{span}\{\phi_1, \ldots, \phi_m\} \subset \mathcal{H}_X, \qquad V_\Psi = \operatorname{span}\{\psi_1, \ldots, \psi_n\} \subset \mathcal{H}_Y,$$

and write

$$\Phi(x) = \begin{bmatrix} \phi_1(x) \\ \vdots \\ \phi_m(x) \end{bmatrix}, \qquad \Psi(y) = \begin{bmatrix} \psi_1(y) \\ \vdots \\ \psi_n(y) \end{bmatrix}.$$

We assume for simplicity that the observables are real-valued. The complex-valued case is identical after replacing transposes by Hermitian transposes.

Given sample points $x_1, \dots, x_M \in X$, define the empirical measure

$$\widehat{\mu}_M = \frac{1}{M} \sum_{\ell=1}^{M} \delta_{x_\ell}.$$

The CSA basis-transfer matrix for a fixed candidate map h_θ is defined by

$$P_{\theta,M} = \arg\min_{P \in \mathbb{R}^{n \times m}} \frac{1}{M} \sum_{\ell=1}^{M} \|\Psi(h_\theta(x_\ell)) - P\Phi(x_\ell)\|_2^2 . \tag{41}$$

Here we use the row convention: the jth row of $P_{\theta,M}$ contains the coefficients, in the Φ-dictionary, of the approximation to the pulled-back observable $\mathcal{C}_{h_\theta}\psi_j = \psi_j \circ h_\theta$. Thus $P_{\theta,M}$ is a row-coefficient matrix. If one instead represents linear maps by column coefficients, the corresponding operator matrix is $P_{\theta,M}^\top$.

Proposition 2 (Finite-data transfer as an empirical projection)**.** Assume that the empirical Gram matrix

$$G_{\Phi,M} = \frac{1}{M} \sum_{\ell=1}^{M} \Phi(x_\ell)\Phi(x_\ell)^\top$$

is invertible. Let $\Pi_\Phi^{\widehat{\mu}_M}$ denote the $L^2(\widehat{\mu}_M)$-orthogonal projection onto V_Φ. Then, for each $j = 1, \dots, n$,

$$\sum_{i=1}^{m} (P_{\theta,M})_{ji}\phi_i = \Pi_\Phi^{\widehat{\mu}_M} \mathcal{C}_{h_\theta}\psi_j.$$

Equivalently, $P_{\theta,M}$ is the row-coefficient matrix of the empirical projected composition operator

$$\Pi_\Phi^{\widehat{\mu}_M} \mathcal{C}_{h_\theta}\big|_{V_\Psi} : V_\Psi \to V_\Phi.$$

Proof. The least-squares problem in (41) separates over the rows of $P_{\theta,M}$. Let $p_{j,M}^\top$ denote the jth row. Then $p_{j,M}$ solves

$$\min_{p \in \mathbb{R}^m} \frac{1}{M} \sum_{\ell=1}^{M} \big(\psi_j(h_\theta(x_\ell)) - p^\top \Phi(x_\ell)\big)^2 .$$

Since $G_{\Phi,M}$ is invertible, the minimizer is unique and satisfies

$$p_{j,M} = G_{\Phi,M}^{-1} \frac{1}{M} \sum_{\ell=1}^{M} \Phi(x_\ell)\psi_j^\top(h_\theta(x_\ell)).$$

On the other hand, the empirical projection of $\mathcal{C}_{h_\theta}\psi_j$ onto V_Φ is, by definition,

$$\Pi_\Phi^{\widehat{\mu}_M} \mathcal{C}_{h_\theta}\psi_j = \arg\min_{f \in V_\Phi} \|f - \mathcal{C}_{h_\theta}\psi_j\|_{L^2(\widehat{\mu}_M)}.$$

Writing $f = p^\top \Phi$, this becomes exactly

$$\arg\min_{p \in \mathbb{R}^m} \frac{1}{M} \sum_{\ell=1}^{M} \big(p^\top \Phi(x_\ell) - \psi_j(h_\theta(x_\ell))\big)^2 .$$

Thus

$$p_{j,M}^\top \Phi = \Pi_\Phi^{\widehat{\mu}_M} \mathcal{C}_{h_\theta}\psi_j.$$

Applying this argument for each $j = 1, \dots, n$ proves the claim. □

We next pass from the empirical projection to the population projection.

Proposition 3 (Infinite-data limit for fixed dictionaries)**.** Assume that $x_1, \dots, x_M$ are i.i.d. samples from μ_X and that

$$G_\Phi = \int_X \Phi(x)\Phi(x)^\top \, d\mu_X(x)$$

is invertible. Assume also that the entries of $\Phi(x)\Phi(x)^\top$ and $\Phi(x)\psi_j(h_\theta(x))$ are integrable for each $j = 1, \dots, n$.

Define the population transfer matrix

$$P_\theta = \arg\min_{P \in \mathbb{R}^{n \times m}} \int_X \|\Psi(h_\theta(x)) - P\Phi(x)\|_2^2 \, d\mu_X(x). \tag{42}$$

Then

$$P_{\theta,M} \to P_\theta$$

almost surely in any finite-dimensional matrix norm. Moreover, for each $j = 1, \dots, n$,

$$\sum_{i=1}^{m} (P_\theta)_{ji}\phi_i = \Pi_\Phi^{\mu_X} \mathcal{C}_{h_\theta}\psi_j,$$

where $\Pi_\Phi^{\mu_X}$ is the $L^2(\mu_X)$-orthogonal projection onto V_Φ. Equivalently, P_θ is the row-coefficient matrix of the population projected composition operator

$$\Pi_\Phi^{\mu_X} \mathcal{C}_{h_\theta}\big|_{V_\Psi} : V_\Psi \to V_\Phi.$$

Proof. The population least-squares problem also separates row by row. The jth row of P_θ, denoted $p_j^\top$, is the unique minimizer of

$$\min_{p \in \mathbb{R}^m} \int_X \big(\psi_j(h_\theta(x)) - p^\top \Phi(x)\big)^2 \, d\mu_X(x).$$

Its normal equations give

$$p_j = G_\Phi^{-1} \int_X \Phi(x)\psi_j^\top(h_\theta(x)) \, d\mu_X(x).$$

Similarly, the jth row of $P_{\theta,M}$ satisfies

$$p_{j,M} = G_{\Phi,M}^{-1} \frac{1}{M} \sum_{\ell=1}^{M} \Phi(x_\ell)\psi_j^\top(h_\theta(x_\ell)).$$

By the strong law of large numbers,

$$G_{\Phi,M} \to G_\Phi$$

and

$$\frac{1}{M} \sum_{\ell=1}^{M} \Phi(x_\ell)\psi_j(h_\theta(x_\ell)) \to \int_X \Phi(x)\psi_j^\top(h_\theta(x)) \, d\mu_X(x)$$

almost surely. Since matrix inversion is continuous on the set of invertible matrices, $p_{j,M} \to p_j$ almost surely for each j. There are finitely many rows, so $P_{\theta,M} \to P_\theta$ almost surely in any matrix norm.

Finally, the identity

$$\sum_{i=1}^{m} (P_\theta)_{ji}\phi_i = \Pi_\Phi^{\mu_X} \mathcal{C}_{h_\theta}\psi_j$$

follows because the population projection

$$\Pi_\Phi^{\mu_X} \mathcal{C}_{h_\theta}\psi_j = \arg\min_{f \in V_\Phi} \|f - \mathcal{C}_{h_\theta}\psi_j\|_{L^2(\mu_X)}$$

is obtained by the same normal equations. □

Thus, for fixed dictionaries, CSA estimates the finite-dimensional operator

$$\Pi_{\Phi}^{\mu_X} \mathcal{C}_{h_\theta}\big|_{V_\Psi} : V_\Psi \to V_\Phi.$$

This is the precise sense in which the fitted transfer matrix is a projected composition operator. It is not an arbitrary linear map between coefficient spaces, and it is not merely an orthogonal change of basis. It is the matrix representation of the pullback by h_θ, restricted to the source dictionary and projected into the target dictionary.

We finally record the complete-basis limit. This statement explains how the projected composition operators recover the true composition operator as the dictionaries become complete.

Proposition 4 (Complete-basis limit)**.** Suppose that $\{\phi_i\}_{i\geq 1}$ and $\{\psi_j\}_{j\geq 1}$ are orthonormal bases of $\mathcal{H}_X$ and $\mathcal{H}_Y$, respectively. Let

$$V_m^X = \operatorname{span}\{\phi_1, \dots, \phi_m\}, \qquad V_n^Y = \operatorname{span}\{\psi_1, \dots, \psi_n\},$$

and let $\Pi_m^X : \mathcal{H}_X \to V_m^X$ and $\Pi_n^Y : \mathcal{H}_Y \to V_n^Y$ denote the corresponding orthogonal projections. If

$$\mathcal{C}_{h_\theta} : \mathcal{H}_Y \to \mathcal{H}_X$$

is bounded, then

$$\Pi_m^X \mathcal{C}_{h_\theta} \Pi_n^Y \to \mathcal{C}_{h_\theta}$$

in the strong operator topology as $m, n \to \infty$. That is, for every $u \in \mathcal{H}_Y$,

$$\left\|\Pi_m^X \mathcal{C}_{h_\theta} \Pi_n^Y u - \mathcal{C}_{h_\theta} u\right\|_{\mathcal{H}_X} \to 0.$$

Proof. Fix $u \in \mathcal{H}_Y$. Then

$$\begin{aligned}\left\|\Pi_m^X \mathcal{C}_{h_\theta} \Pi_n^Y u - \mathcal{C}_{h_\theta} u\right\|_{\mathcal{H}_X} &\leq \left\|\Pi_m^X \mathcal{C}_{h_\theta}(\Pi_n^Y u - u)\right\|_{\mathcal{H}_X} + \left\|(\Pi_m^X - I)\mathcal{C}_{h_\theta} u\right\|_{\mathcal{H}_X} \\ &\leq \|\mathcal{C}_{h_\theta}\| \, \|\Pi_n^Y u - u\|_{\mathcal{H}_Y} + \left\|(\Pi_m^X - I)\mathcal{C}_{h_\theta} u\right\|_{\mathcal{H}_X}.\end{aligned}$$

Because $\{\psi_j\}_{j\geq 1}$ is an orthonormal basis of $\mathcal{H}_Y$, $\Pi_n^Y u \to u$ in $\mathcal{H}_Y$. Because $\{\phi_i\}_{i\geq 1}$ is an orthonormal basis of $\mathcal{H}_X$, $\Pi_m^X v \to v$ in $\mathcal{H}_X$ for every $v \in \mathcal{H}_X$, and in particular for $v = \mathcal{C}_{h_\theta} u$. Hence both terms on the right-hand side converge to zero. □

Combining Propositions 2, 3, and 4, we obtain the following interpretation: for fixed dictionaries, the CSA least-squares transfer is the empirical projection of the composition operator induced by h_θ; as the number of samples grows, it converges to the population projected composition operator; and, under complete orthonormal dictionaries and boundedness of the pullback operator, these projected operators converge strongly to the true composition operator.

I Pseudocode and optimisation details

We describe the computational procedures for CSA (Section 4) and CSA-Krylov (Section 5.2). Both share a two-level structure: an *inner* regression that fits the basis-transfer matrix P_θ for each candidate θ, and an *outer* search over θ that minimises the intertwining residual. The code necessary to reproduce our results is publicly available at `https://github.com/prakhargodara/Koopman-operators---CSA`.

I.1 CSA

In the EDMD setting the observable dictionaries Φ and Ψ are known explicitly, so the regression for P_θ reduces to a standard least-squares problem.

Algorithm 1 Conjugacy-based Similarity Analysis (CSA)

Require: Koopman matrices $F \in \mathbb{R}^{m\times m}$, $G \in \mathbb{R}^{n\times n}$; dictionaries Φ, Ψ; conjugacy family $\{h_\theta\}_{\theta\in\Theta}$; evaluation grid $\{x_i\}_{i=1}^M$; regularisation $\lambda > 0$; coarse grid $\theta_1, \dots, \theta_L \in \Theta$; number of refinements n_{ref}
Ensure: d_{CSA}, θ^*

$\mathbf{\Phi}_1 \leftarrow [\Phi(x_1), \dots, \Phi(x_M)]^\top \in \mathbb{R}^{M\times m}$ ▷ Pre-compute dictionary on grid
$G_\Phi \leftarrow \mathbf{\Phi}_1^\top \mathbf{\Phi}_1 + \lambda I_m$ ▷ Regularised Gram matrix
$R \leftarrow G_\Phi^{-1} \mathbf{\Phi}_1^\top$ ▷ Regression operator, $m \times M$

function OBJECTIVE(θ)
  $\mathbf{\Phi}_2 \leftarrow [\Psi(h_\theta(x_1)), \dots, \Psi(h_\theta(x_M))]^\top$ ▷ Dictionary of system 2 at $h_\theta(x)$
  $P_\theta \leftarrow (R\,\mathbf{\Phi}_2)^\top$ ▷ Least-squares transfer matrix, $n \times m$
  return $\|P_\theta F - G\,P_\theta\|_F^2 \;/\; \|P_\theta\|_F^2$
end function

Coarse search: evaluate OBJECTIVE(θ_j) for $j = 1, \dots, L$
$\theta^* \leftarrow \arg\min_j$ OBJECTIVE(θ_j)

Local refinement: for the top-n_{ref} coarse candidates, run Brent's method (bounded scalar optimisation) on OBJECTIVE($\cdot$) in a small window $[\theta_j - \delta, \theta_j + \delta]$ around each candidate; update θ^* if any refinement improves the objective

$d_{\mathrm{CSA}} \leftarrow \sqrt{\text{OBJECTIVE}(\theta^*)}$
return d_{CSA}, θ^*

I.2 CSA-Krylov

In the delay-embedding setting the dictionaries are determined by the data through the SVD of the Hankel matrix (Appendix G). Given the known dynamics f_i and the SVD basis $V_{i,r'}$, the projected Krylov observables can be evaluated at any point x:

$$\phi_i(x) \;=\; \Sigma_{i,r'}^{-1} V_{i,r'}^\top \begin{pmatrix} x \\ f_i(x) \\ f_i^2(x) \\ \vdots \\ f_i^{d-1}(x) \end{pmatrix} \;\in\; \mathbb{R}^{r'}.$$

This is the analogue of the EDMD dictionary $\Phi(x)$: a scalar state is mapped to an r'-dimensional feature vector whose entries are the leading normalised SVD components of the delay-embedding (Appendix G). CSA-Krylov then constructs P_θ and optimises over θ exactly as CSA (Algorithm 1), with ϕ_1, ϕ_2 replacing Φ, Ψ.

J Convergence diagnostics

We test that the Koopman operators we fit converge as the representation truncation refines.

For the EDMD panels, we sweep $k \in \{2.0, 2.2, \dots, 4.0\}$ and fit the Koopman matrix F of $f_k(x) = k\,x(1-x)$ on 10^5 uniform samples of $[0,1]$ (ridge 10^{-6}), matching the EDMD experiments of Section 5.1; both EDMD-Fourier and EDMD-Legendre are swept across $n_b \in \{11, 21, 41, 81, 121\}$ basis functions. For the Hankel-DMD panel we use the Koopman matrices of Figure 5c: linear contractions f_ρ at $\rho \in \{0.5, 0.8\}$ and their conjugate twins $g^{(\rho)} = h_{a^*,n^*}^{-1} \circ f_\rho \circ h_{a^*,n^*}$ at the operating point $(a^*, n^*) = (0.5, 4)$, fitted from the same 50 training trajectories of length $T = 150$ per system, at delay dimension $d = 100$, in the normalised coordinates $\phi = \Sigma_{r'}^{-1} V_{r'}^\top z$ truncated at $r' = \min(r, \#\{k : \sigma_k > 10^{-8}\sigma_1\})$ (Appendix G), with ridge $\lambda = 10^{-6}$; the rank r is swept across $\{2, 3, 4, 5, 6, 8, 10, 15, 20, 40, 70, 80, 99\}$ and ρ_x is evaluated on 100 held-out trajectories per system.

Figure 7 reports the held-out state-prediction residual

$$\rho_x(K, f) \;=\; \frac{\sum_t \big(x_{t+1} - c^\top K\,\Phi(x_t)\big)^2}{\sum_t x_{t+1}^2}, \qquad c \;=\; \arg\min_c \sum_t \big(x_t - c^\top \Phi(x_t)\big)^2,$$

where the readout c is fit on training data so that $c^\top \Phi(x)$ recovers the state. Predicting the 1-D state with a fixed readout makes ρ_x comparable across truncation levels. Fourier-EDMD converges algebraically in n_b with a peak at $k = 4$ where the state is most non-periodic. On the other hand, Legendre-EDMD on f_k has converged at every n_b. For the Hankel-DMD panel the linear contractions saturate at the floor already at rank $r = 2$ — their Koopman spectrum is essentially two-dimensional — while the smooth-homeomorphism conjugates require a larger truncation to be resolved: at $\rho = 0.8$ the residual drops by about seven orders of magnitude between $r = 2$ and $r = 8$. All four systems then plateau at $\rho_x \approx 10^{-12}$, the bias floor of the ridge ($\rho_x \approx \lambda^2$ in these normalised coordinates), and for $r \geq 15$ the cutoff fixes r' (at 2, 8 and 12 for f_ρ, $g^{(0.5)}$ and $g^{(0.8)}$), so the fitted matrices no longer change with r. All three representations exhibit monotone refinement.

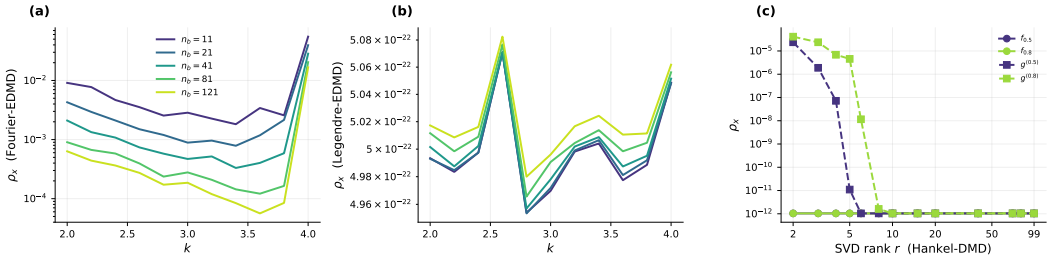


Figure 7: State-prediction residual ρ_x versus truncation, by representation. **(a)** Fourier-EDMD on f_k; lighter shades denote larger n_b. **(b)** Legendre-EDMD on f_k. **(c)** Hankel-DMD with the Koopman matrices of Figure 5c (linear contractions f_ρ and their smooth-conjugate twins $g^{(\rho)}$ at $(a^*, n^*) = (0.5, 4)$); the nonlinear conjugates need a larger truncation than the linear maps before ρ_x saturates.

K Robustness to misspecification of the homeomorphism class

In Section 5 we give CSA a family $\{h_\theta\}_{\theta\in\Theta}$ containing the true homeomorphism, however, in practice this may not be guaranteed. Therefore, here we bound the changes in the CSA score as a function of homeomorphism-family misspecification.

With F, G, Φ, Ψ as in Section 4, $\{x_i\}_{i=1}^N \subset X$ the evaluation grid and $R = (\mathbf{\Phi}^\top\mathbf{\Phi})^{-1}\mathbf{\Phi}^\top$ the regression operator of Algorithm 1, define for a bounded measurable $h : X \to Y$

$$P_h = (R\,\mathbf{\Psi}_h)^\top, \quad \mathbf{\Psi}_h = [\Psi(h(x_1)), \ldots, \Psi(h(x_N))]^\top, \qquad J(h) = \frac{\|P_h F - G P_h\|_F}{\|P_h\|_F},$$

so that P_{h_θ} is the regularised solution of Eq. (11) and $d_{\mathrm{CSA}}(F, G) = \inf_{\theta\in\Theta} J(h_\theta)$ by Eq. (12).

Proposition 5. Assume $\psi_1 \equiv 1$, each ψ_a is Lipschitz on Y, and $p_{\min} := \|R\mathbf{1}_N\|_2 > 0$ (automatic when the constant function also belongs to the Φ-dictionary). Then J is Lipschitz over bounded measurable maps $X \to Y$:

$$|J(h) - J(\tilde h)| \leq L_J\,\|h - \tilde h\|_\infty, \qquad L_J := \frac{2\big(\|F\|_2 + \|G\|_2\big)}{p_{\min}}\,\|R\|_2\,\sqrt{N}\,S_\Psi, \qquad S_\Psi^2 := \sum_{a=1}^{n} \mathrm{Lip}(\psi_a)^2,$$

where $\|h - \tilde h\|_\infty := \sup_{x\in X} \|h(x) - \tilde h(x)\|_2$. Consequently, for every bounded measurable reference map $h^* : X \to Y$,

$$d_{\mathrm{CSA}}(F, G) \;\leq\; J(h^*) + L_J \inf_{\theta\in\Theta} \|h_\theta - h^*\|_\infty.$$

In particular, $J(h^*) = 0$ gives $d_{\mathrm{CSA}}(F, G) \leq L_J \inf_{\theta\in\Theta} \|h_\theta - h^*\|_\infty$.

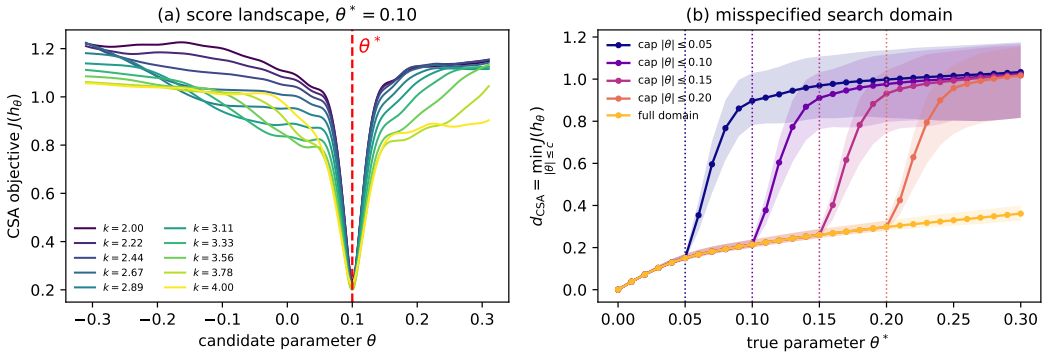


Figure 8: **Score landscape and continuous degradation under misspecification** (Fourier dictionary, $m = 25$, conjugate twins generated at $\theta^* = 0.10$). (a) $J(h_\theta)$ versus θ for the ten logistic parameters k. (b) Returned score $\min_{|\theta| \le c} J(h_\theta)$ versus the true parameter θ^* for search domains capped at $c \in \{0.05, 0.10, 0.15, 0.20\}$ (dotted verticals) and for the full domain; lines: mean over k, bands: min–max.

Proof. For any bounded measurable $h, \tilde{h}$,

$$\|P_h - P_{\tilde{h}}\|_F = \|R(\mathbf{\Psi}_h - \mathbf{\Psi}_{\tilde{h}})\|_F \le \|R\|_2 \, \|\mathbf{\Psi}_h - \mathbf{\Psi}_{\tilde{h}}\|_F \le \|R\|_2 \sqrt{N} \, S_\Psi \, \|h - \tilde{h}\|_\infty,$$

since $|\psi_a(h(x_i)) - \psi_a(\tilde{h}(x_i))| \le \mathrm{Lip}(\psi_a) \, \|h - \tilde{h}\|_\infty$ at each of the N grid points. Because $\psi_1 \equiv 1$, the first row of every P_h equals $(R\mathbf{1}_N)^\top$, so $\|P_h\|_F \ge p_{\min} > 0$. Writing $\nu(P) := \|PF - GP\|_F$ and $\kappa := \|F\|_2 + \|G\|_2$, we have $|\nu(P) - \nu(\widetilde{P})| \le \kappa \|P - \widetilde{P}\|_F$ and $\nu(P) \le \kappa \|P\|_F$. Hence

$$|J(h) - J(\tilde{h})| \le \frac{|\nu(P_h) - \nu(P_{\tilde{h}})|}{\|P_h\|_F} + \frac{\nu(P_{\tilde{h}}) \, \big| \, \|P_{\tilde{h}}\|_F - \|P_h\|_F \big|}{\|P_h\|_F \, \|P_{\tilde{h}}\|_F} \le \frac{2\kappa}{p_{\min}} \, \|P_h - P_{\tilde{h}}\|_F \le L_J \, \|h - \tilde{h}\|_\infty.$$

The score bound follows from $J(h_\theta) \le J(h^*) + L_J \|h_\theta - h^*\|_\infty$ by taking the infimum over $\theta \in \Theta$. □

Figure 8 checks both aspects of the proposition numerically in the setting of Section 5.1.1: the landscape $\theta \mapsto J(h_\theta)$ is smooth with a single minimum at the true parameter (recovered to within 10^{-3} for every k), and when the search domain is capped at $|\theta| \le c$ — for this family $\|h_\theta - h_{\theta^*}\|_\infty = |\theta - \theta^*|$, so the overshoot $\theta^* - c$ is exactly the misspecification distance — the returned score leaves the full-domain curve continuously, with initial slope of the order of the landscape's empirical Lipschitz constant, as the linear bound predicts.

L Robustness to measurement noise

We repeat the conjugacy experiment of Section 5.1.1 ($\theta^* = 0.10$, Fourier dictionary, $m = 25$) with i.i.d. Gaussian noise of standard deviation σ (relative to the unit state interval) added to every observed state — inputs and outputs of both systems — before EDMD, over the factorial sweep of $N \in \{10^3, 10^4, 10^5\}$ samples against $\sigma \in \{0, 0.01, 0.02, 0.05\}$, with five independent replicates per cell (Figure 9). Everything CSA *infers* is stable up to 5% noise: the recovered parameter $\hat{\theta}$ stays within 2×10^{-3} of θ^* for $\sigma \le 0.02$ and within 8×10^{-3} at $\sigma = 0.05$ (panel a); the comparison–reference agreement stays at Pearson 0.98–0.99 (panel b); and each system's conjugate twin is identified by row-wise argmin in $50/50$ replicates — all with the sole exception of the jointly extreme cell $(N, \sigma) = (10^3, 0.05)$, which degrades gracefully (Pearson 0.90, twin identification $45/50$) rather than catastrophically. Absolute scores, by contrast, shrink with σ (panel c): input noise attenuates the EDMD regression (errors-in-variables), mollifying the high-mode content of *both* operators consistently and lowering the finite-dictionary floor. CSA's relative statements are therefore noise-robust, but raw scores should not be compared across noise levels. Finally, we only consider measurement noise, whereas under *process* noise the same construction applies with $\mathcal{K}_f$, $\mathcal{K}_g$ replaced by the stochastic Koopman operators.

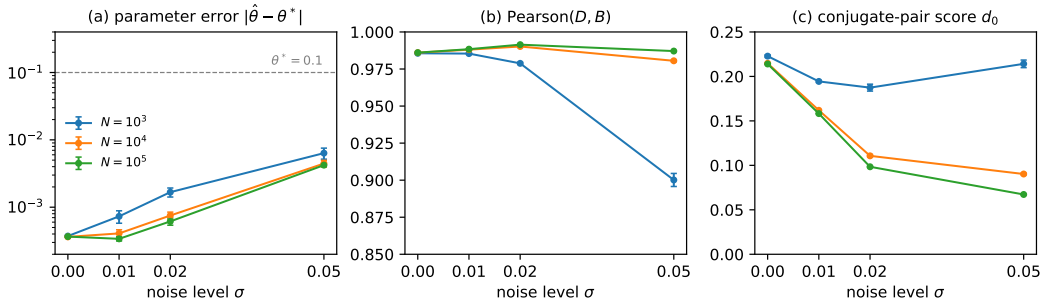


Figure 9: **Robustness to measurement noise** (mean $\pm$ std over five replicates). (a) Recovered-parameter error, averaged over the ten conjugate pairs. (b) Pearson correlation between comparison and reference distance matrices. (c) Mean CSA score of the exactly conjugate pairs: absolute scores attenuate with noise even as the inference remains stable.

M Convergence of the CSA score

Appendix H establishes convergence of the fitted transfer matrix; here we establish convergence of the CSA *score*. The natural infinite-dimensional continuation of the Frobenius norm in Eq. (12) — the Hilbert–Schmidt norm — is generically infinite for composition operators, so we work with a weighted norm that can be equivalent to the Frobenius norm at any basis size used in practical applications, but is bounded in the infinite limit. We show that the weighted score converges, as the sample sizes and the dictionary sizes grow, to an infinite-dimensional objective d^{∞}; that d^{∞} vanishes precisely when the searched family contains an almost-everywhere conjugacy (Proposition 8). We also show the CSA score's relationship with the direct state-space conjugacy residual by the exact formula Eq. (49).

Notation. Eq. (12) of the main text is the finite-sample, finite-dictionary case, with the sample and dictionary sizes left implicit. In this appendix the notation is fully explicit: hats with the subscripts N, N' mark finite-sample quantities, the superscripts m, n mark the dictionary sizes, and symbols carrying neither refer to the complete infinite limit. Thus $\widehat{P}^{m,n}_{\theta,N}$, $\widehat{F}^m_N$, $\widehat{G}^n_{N'}$ denote the computed matrices entering Eq. (12); $P^{m,n}_{\theta}$, F^m, G^n their infinite-sample counterparts; and $\mathcal{K}_f$, $\mathcal{K}_g$, $\mathcal{C}_{h_\theta}$ the operators of Section 2. For the scores, the superscript ∞ marks the complete limit.

Assumptions. As in Appendix H, $\mathcal{H}_X = L^2(X, \mu_X)$ and $\mathcal{H}_Y = L^2(Y, \mu_Y)$ for probability measures μ_X, μ_Y. We assume:

- **(A1)** each $h_\theta : X \to Y$ is a homeomorphism and Θ is finite (the grid searched in Algorithm 1); and $\mathcal{K}_f$, $\mathcal{K}_g$, $\mathcal{C}_{h_\theta}$ are well defined and bounded;
- **(A2)** $\{\phi_i\}_{i\geq 1}$, $\{\psi_j\}_{j\geq 1}$ are orthonormal bases of $\mathcal{H}_X$, $\mathcal{H}_Y$ with $\phi_1 \equiv \psi_1 \equiv 1$; the orthogonal projections Π^X_m, Π^Y_n onto their initial segments are as in Proposition 4;
- **(A3)** $n_0 \geq 1$ and strictly positive $\{w_j\}_{j\geq 1}$ are fixed with $w_j = 1$ for $j \leq n_0$ and $\sum_j w_j < \infty$; we write $W_n = \mathrm{diag}(w_1, \ldots, w_n)$ and, for a bounded operator $A : \mathcal{H}_Y \to \mathcal{H}_X$, $\|A\|^2_W := \sum_{j\geq 1} w_j \|A\psi_j\|^2_{\mathcal{H}_X} \leq \|A\|^2 \sum_j w_j < \infty$;
- **(A4)** $x_1, \ldots, x_N$ are i.i.d. from μ_X and $y_1, \ldots, y_{N'}$ are i.i.d. from μ_Y; all empirical moments appearing below are integrable.

For $\theta \in \Theta$, let $\widehat{P}^{m,n}_{\theta,N}$ be the minimizer of Eq. (11) (the matrix $P_{\theta,M}$ of Appendix H with $M = N$), and let $\widehat{F}^m_N$, $\widehat{G}^n_{N'}$ be the EDMD matrices from the analogous least squares for $\Phi \circ f$ and $\Psi \circ g$. The weighted finite-sample score is

$$\widehat{d}^{m,n}_{N,N'} := \min_{\theta\in\Theta} \widehat{J}^{m,n}_{N,N'}(\theta), \qquad \widehat{J}^{m,n}_{N,N'}(\theta) := \frac{\left\| W_n^{1/2} \left(\widehat{P}^{m,n}_{\theta,N} \widehat{F}^m_N - \widehat{G}^n_{N'} \widehat{P}^{m,n}_{\theta,N} \right) \right\|_F}{\left\| W_n^{1/2} \widehat{P}^{m,n}_{\theta,N} \right\|_F}. \tag{43}$$

If $n \le n_0$ then $W_n = I_n$ and (43) is *verbatim* Eq. (12); choosing n_0 no smaller than the largest dictionary dimension used in this paper makes every reported score a member of the weighted family (see Remark 2).

The population counterparts at dictionary sizes (m, n) are built from the projected operators

$$\mathcal{K}_f^m := \Pi_m^X \mathcal{K}_f \Pi_m^X, \qquad \mathcal{K}_g^n := \Pi_n^Y \mathcal{K}_g \Pi_n^Y, \qquad \mathcal{C}_\theta^{m,n} := \Pi_m^X \mathcal{C}_{h_\theta} \Pi_n^Y, \tag{44}$$

through the intertwining defects $\mathcal{R}_\theta^{m,n} := \mathcal{K}_f^m \mathcal{C}_\theta^{m,n} - \mathcal{C}_\theta^{m,n} \mathcal{K}_g^n$ and, in the complete limit, $\mathcal{R}_\theta := \mathcal{K}_f \mathcal{C}_{h_\theta} - \mathcal{C}_{h_\theta} \mathcal{K}_g$, whose vanishing is the intertwining relation of Eq. (4):

$$d^{\,m,n} := \min_{\theta \in \Theta} J^{m,n}(\theta), \quad J^{m,n}(\theta) := \frac{\|\mathcal{R}_\theta^{m,n}\|_W}{\|\mathcal{C}_\theta^{m,n}\|_W}, \qquad d^{\,\infty} := \min_{\theta \in \Theta} J^{\infty}(\theta), \quad J^{\infty}(\theta) := \frac{\|\mathcal{R}_\theta\|_W}{\|\mathcal{C}_{h_\theta}\|_W}. \tag{45}$$

All W-norms here are finite by (A3), and $\|\mathcal{C}_{h_\theta}\|_W \ge \sqrt{w_1} = 1$ since $\mathcal{C}_{h_\theta} \psi_1 = 1$.

Matrix–operator correspondence. Let

$$(F^m)_{ik} = \langle \mathcal{K}_f \phi_i, \phi_k \rangle, \qquad (G^n)_{jl} = \langle \mathcal{K}_g \psi_j, \psi_l \rangle, \qquad (P_\theta^{m,n})_{ji} = \langle \mathcal{C}_{h_\theta} \psi_j, \phi_i \rangle,$$

for $i, k \le m$ and $j, l \le n$; these are the population versions of $\widehat{F}_N^m$, $\widehat{G}_{N'}^n$, $\widehat{P}_{\theta,N}^{m,n}$ (Proposition 3 and its analogues for $\mathcal{K}_f$, $\mathcal{K}_g$). Expanding in the orthonormal bases, for $j \le n$,

$$\mathcal{R}_\theta^{m,n} \psi_j = \sum_{k=1}^{m} (P_\theta^{m,n} F^m - G^n P_\theta^{m,n})_{jk} \phi_k,$$

while $\mathcal{R}_\theta^{m,n} \psi_j = \mathcal{C}_\theta^{m,n} \psi_j = 0$ for $j > n$. Hence, by orthonormality,

$$\|\mathcal{R}_\theta^{m,n}\|_W = \|W_n^{1/2} (P_\theta^{m,n} F^m - G^n P_\theta^{m,n})\|_F, \qquad \|\mathcal{C}_\theta^{m,n}\|_W = \|W_n^{1/2} P_\theta^{m,n}\|_F.$$

Since $\psi_1 = \phi_1 = 1$, the first row of $P_\theta^{m,n}$ is $(1, 0, \dots, 0)$, so $\|W_n^{1/2} P_\theta^{m,n}\|_F \ge \sqrt{w_1} = 1$. Thus $J^{m,n}(\theta)$ is exactly the population counterpart of the finite-dimensional weighted CSA objective.

Proposition 6. Under (A1)–(A4), for fixed (m, n),

$$\widehat{d}_{N,N'}^{\,m,n} \longrightarrow d^{\,m,n} \qquad \text{almost surely as } N, N' \to \infty.$$

Proof. For every $\theta \in \Theta$,

$$\widehat{P}_{\theta,N}^{m,n} \to P_\theta^{m,n}, \qquad \widehat{F}_N^m \to F^m, \qquad \widehat{G}_{N'}^n \to G^n \quad \text{a.s.},$$

by Proposition 3 and, for the EDMD matrices, Theorem 2 of Korda and Mezić [2018]. Since $\|W_n^{1/2} P_\theta^{m,n}\|_F \ge 1$, continuity of

$$(P, F, G) \mapsto \frac{\|W_n^{1/2}(PF - GP)\|_F}{\|W_n^{1/2} P\|_F}$$

gives $\widehat{J}_{N,N'}^{m,n}(\theta) \to J^{m,n}(\theta)$ a.s. for every θ. Because Θ is finite, this convergence is uniform over Θ, and therefore

$$\widehat{d}_{N,N'}^{\,m,n} = \min_{\theta \in \Theta} \widehat{J}_{N,N'}^{m,n}(\theta) \longrightarrow \min_{\theta \in \Theta} J^{m,n}(\theta) = d^{\,m,n}.$$

□

Proposition 7. Under (A1)–(A3), as $\min(m, n) \to \infty$, $\sup_{\theta \in \Theta} |J^{m,n}(\theta) - J^{\infty}(\theta)| \to 0$, and hence $d^{\,m,n} \to d^{\,\infty}$.

Proof. Set $B_C := \max_{\theta \in \Theta} \|\mathcal{C}_{h_\theta}\|$, finite since Θ is. Because orthogonal projections have operator norm at most one, $\|\mathcal{C}_\theta^{m,n}\| \le B_C$, $\|\mathcal{K}_f^m\| \le \|\mathcal{K}_f\|$, $\|\mathcal{K}_g^n\| \le \|\mathcal{K}_g\|$, and hence $\|\mathcal{R}_\theta^{m,n}\|, \|\mathcal{R}_\theta\| \le B_R := B_C(\|\mathcal{K}_f\| + \|\mathcal{K}_g\|)$ for all m, n, θ.

Fix θ and j. For $n \ge j$, $\Pi_n^Y \psi_j = \psi_j$, hence from Proposition 4,

$$\mathcal{C}_\theta^{m,n} \psi_j = \Pi_m^X \mathcal{C}_{h_\theta} \psi_j \longrightarrow \mathcal{C}_{h_\theta} \psi_j,$$

and similarly, using additionally that $\mathcal{K}_f^m \to \mathcal{K}_f$ strongly (Theorem 3 of Korda and Mezić [2018]),

$$\mathcal{R}_\theta^{m,n}\psi_j \longrightarrow \mathcal{R}_\theta\psi_j.$$

Moreover, since $\|\psi_j\| = 1$, $w_j\|(\mathcal{R}_\theta^{m,n} - \mathcal{R}_\theta)\psi_j\|^2 \le 4B_R^2 w_j$; because $\sum_j w_j < \infty$, dominated convergence gives

$$\|\mathcal{R}_\theta^{m,n} - \mathcal{R}_\theta\|_W \to 0, \qquad \|\mathcal{C}_\theta^{m,n} - \mathcal{C}_{h_\theta}\|_W \to 0.$$

Since Θ is finite, both convergences are uniform in θ. Furthermore, $\|\mathcal{C}_\theta^{m,n}\|_W, \|\mathcal{C}_{h_\theta}\|_W \ge 1$ because $\psi_1 = \phi_1 = 1$. Hence $\sup_{\theta\in\Theta} |J^{m,n}(\theta) - J^\infty(\theta)| \to 0$, and taking minima over θ yields $d^{m,n} \to d^\infty$. □

Combining the two propositions,

$$\underbrace{\widehat{d}_{N,N'}^{m,n}}_{\text{computed; = Eq. (12) for } n\le n_0} \xrightarrow[\text{a.s.}]{N,N'\to\infty} d^{m,n} \xrightarrow{m,n\to\infty} d^\infty.$$

Remark 2. For $w \equiv 1$ the norm $\|\cdot\|_W$ is the Hilbert–Schmidt norm, which boundedness of $\mathcal{C}_{h_\theta}$ does not make finite: if $(h_\theta)_*\mu_X$ has density bounded below with respect to μ_Y, then $\|\mathcal{C}_{h_\theta}\psi_j\|$ is bounded away from zero and the unweighted sum diverges. No convergence is therefore claimed for the unweighted continuation of Eq. (12); the summable tail is what makes the limit finite. Only the nonzero values of d^∞ depend on the tail (and on the Y-side basis); its zero set does not (Proposition 8).

It is straightforward to relate the Koopman formulation to the state-space conjugacy residual if we assume additionally **(A5)**: $Y \subseteq \mathbb{R}^d$ and the coordinate observables $\iota^{(a)}(y) = y_a$, $a = 1, \dots, d$, lie in $L^2(\mu_Y)$. The direct residual of Eq. (1), $r(\theta) := \int_X \|h_\theta(f(x)) - g(h_\theta(x))\|_2^2 \, d\mu_X(x)$, involves no Koopman operators and is reached from $\mathcal{R}_\theta$ through the coordinates:

$$\mathcal{R}_\theta\, \iota^{(a)} = (h_\theta \circ f)_a - (g \circ h_\theta)_a \quad (a = 1, \dots, d), \qquad \text{so} \qquad r(\theta) = \sum_{a=1}^{d} \|\mathcal{R}_\theta\, \iota^{(a)}\|_{\mathcal{H}_X}^2. \tag{46}$$

Proposition 8 (Zero set: the weighted score vanishes iff the family contains a conjugacy)**.** Under (A1)–(A3) and (A5),

$$d^\infty = 0 \iff \min_{\theta\in\Theta} r(\theta) = 0 \iff \exists\, \theta \in \Theta :\ h_\theta \circ f = g \circ h_\theta \ \ \mu_X\text{-a.e.}$$

The equivalence holds for every orthonormal basis and every weight sequence admissible under (A3). If moreover f and g are continuous and $\operatorname{supp}\mu_X = X$, the conjugacy identity holds everywhere on X.

Proof. Since Θ is finite, $d^\infty = 0$ means $\|\mathcal{R}_\theta\|_W = 0$ for some θ; as all $w_j > 0$ and $\{\psi_j\}$ is complete, $\mathcal{R}_\theta = 0$ on $\mathcal{H}_Y$, and evaluating on the coordinates via (46) gives $h_\theta \circ f = g \circ h_\theta$ μ_X-a.e. Conversely, if $h_\theta \circ f = g \circ h_\theta$ on a μ_X-full set, then $u \circ h_\theta \circ f = u \circ g \circ h_\theta$ μ_X-a.e. for every $u \in \mathcal{H}_Y$ (the compositions are well defined by (A1)), i.e. $\mathcal{R}_\theta = 0$ and $d^\infty = 0$; the equivalence with $r(\theta) = 0$ is immediate from (46). Only positivity of the weights and completeness of the basis were used, whence the independence claim. Finally, if f and g are continuous then $h_\theta \circ f - g \circ h_\theta$ is continuous and vanishes μ_X-a.e., hence on $\operatorname{supp}\mu_X = X$. □

Beyond sharing a zero set, the identity-observable direction gives an explicit relation between the CSA score and the direct state-space residual. For clarity, let $d = 1$ and define

$$\bar{\iota} := \iota - \int_Y y \, d\mu_Y(y), \qquad \sigma_\iota := \|\bar{\iota}\|_{L^2(\mu_Y)} > 0, \qquad \delta_\theta := h_\theta \circ f - g \circ h_\theta,$$

so that $r(\theta) = \|\delta_\theta\|_{\mathcal{H}_X}^2$, and suppose

$$\psi_{j_\iota} = \frac{\bar{\iota}}{\sigma_\iota} \tag{47}$$

for some j_ι (e.g. - for the shifted Legendre dictionary of Section 5.1, $j_\iota = 2$); the vector case can be computed similarly. Since $\mathcal{R}_\theta 1 = 0$, Eq. (46) gives

$$\mathcal{R}_\theta\psi_{j_\iota} = \frac{\delta_\theta}{\sigma_\iota}, \tag{48}$$

and splitting the j_ι-term of $\|\mathcal{R}_\theta\|_W^2$ yields,

$$J^\infty(\theta)^2 = \frac{w_{j_\iota}\, r(\theta)/\sigma_\iota^2 + \sum_{j \neq j_\iota} w_j \|\mathcal{R}_\theta \psi_j\|_{\mathcal{H}_X}^2}{\|\mathcal{C}_{h_\theta}\|_W^2}. \tag{49}$$

At finite dictionary sizes the same direction carries the residual together with explicit projection errors: for $n \geq j_\iota$,

$$\mathcal{R}_\theta^{m,n} \psi_{j_\iota} = \frac{\delta_\theta + E_\theta^{m,n}}{\sigma_\iota}, \qquad E_\theta^{m,n} := (\Pi_m^X - I)\delta_\theta + \Pi_m^X \mathcal{K}_f (\Pi_m^X - I) h_\theta + \Pi_m^X \mathcal{C}_{h_\theta} (I - \Pi_n^Y) g \tag{50}$$

(insert $\Pi_m^X = I + (\Pi_m^X - I)$ and $\Pi_n^Y = I - (I - \Pi_n^Y)$ into $\mathcal{K}_f^m \mathcal{C}_\theta^{m,n} \psi_{j_\iota} - \mathcal{C}_\theta^{m,n} \mathcal{K}_g^n \psi_{j_\iota}$, and use that constants lie in both dictionaries). Consequently,

$$J^{m,n}(\theta)^2 = \frac{\frac{w_{j_\iota}}{\sigma_\iota^2} \left\| \delta_\theta + E_\theta^{m,n} \right\|_{\mathcal{H}_X}^2 + \sum_{\substack{1 \leq j \leq n \\ j \neq j_\iota}} w_j \|\mathcal{R}_\theta^{m,n} \psi_j\|^2}{\|\mathcal{C}_\theta^{m,n}\|_W^2}. \tag{51}$$

Thus the finite-dimensional score contains the direct state-space residual through the identity-observable direction, together with its projection error and the defects along the remaining observables. By strong convergence of the projections and boundedness of $\mathcal{K}_f$ and $\mathcal{C}_{h_\theta}$, $E_\theta^{m,n} \to 0$ in $\mathcal{H}_X$ as $\min(m, n) \to \infty$.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: [N/A]

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: [N/A]

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

Justification: [N/A]

Guidelines:

- The answer [N/A] means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: [N/A]

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [Yes]

Justification: [N/A]

Guidelines:

- The answer [N/A] means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: [N/A]

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: [N/A]

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).
   - It should be clear whether the error bar is the standard deviation or the standard error of the mean.

- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer:

   Justification: In this paper, we have run all the codes on the personal laptop. The total time it takes to run them is somewhere between 2 hours and 16 hours.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: [N/A]

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [No]

    Justification: This paper is predominantly about the use of operator theory in evaluating dynamical conjugacy. It is unclear what the societal impact would be.

    Guidelines:

    - The answer [N/A] means that there is no societal impact of the work performed.
    - If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.

- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: We have no datasets. The uploaded code can be executed as described in the readme.

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: [N/A]

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.

- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [Yes]

    Justification: [N/A]

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: [N/A]

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: [N/A]

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

Answer: [N/A]

Justification: [N/A]

Guidelines:

- The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.