Geometric framework for robust order detection in delay-coordinates Dynamic Mode Decomposition

Yoav Harris

Geometric framework for robust order detection in delay-coordinates Dynamic Mode Decomposition

Research Thesis

In Partial Fulfillment of the Requirements for
the Degree of Master of Science in Electrical Engineering

Submitted to the Senate of the Technion—Israel Institute of Technology

Shvat 5786 Haifa February 2026

Acknowledgment

The research thesis was done in the Viterbi Faculty of Electrical and Computer Engineering under the joint supervision of Prof. Ronen Talmon and Dr. Hadas Benisty.

The generous financial help of the Technion and the Leonid and Diane Sherman Interdisciplinary Graduate School Fellowship is gratefully acknowledged.

I hereby declare that the research presented in this thesis, including the collection, processing, and presentation of data, as well as references to and comparisons with prior work, was carried out with full academic integrity and in accordance with the ethical standards of the scientific community. The reporting of the research and its results has been conducted honestly and completely, in accordance with these standards.

List of Publications

- Yoav Harris, Hadas Benisty, and Ronen Talmon. "Geometric framework for robust order detection in delay-coordinates Dynamic Mode Decomposition." Submitted 2026.

  This publication covers the same material as the thesis and relates to Chapters 2–6.

Contents

List of Figures

List of Tables

Abstract

Delay-coordinates dynamic mode decomposition (DC–DMD) is widely used to extract coherent spatiotemporal components from high-dimensional time series. A central practical difficulty is choosing the model order, or equivalently deciding which extracted components represent the underlying dynamics in the presence of noise. This work adopts a geometric perspective on order detection in DC–DMD. We introduce a new definition of true and spurious modes based on their relationship to the signal subspace, and use it to analyze the geometry and structure of recovered DMD modes. Building on this framework, we propose two complementary, fully data-driven methods for robust component selection in DC–DMD. The first is based on a geometric criterion derived from subspace-residual analysis, which exploits the fact that true modes concentrate their energy in signal-dominated directions, while spurious modes necessarily retain significant energy outside any moderate subspace overestimate. The second is based on a structure-based criterion grounded in a new operator-theoretic analysis of DC–DMD; using a block-companion formulation, we provide a new derivation of the Kronecker–Vandermonde (KV) structure of the modes and show that spurious modes deviate from this KV structure according to their geometric residuals. The analysis further yields principled insight into commonly used energy- and norm-based heuristics, clarifying when and why they can fail under delay-coordinates. Extensive numerical experiments confirm the theoretical predictions and demonstrate that the proposed geometric and structure-based methods achieve consistently better order detection than existing baselines across a wide range of noise levels, spectral separations, damping regimes, and embedding lengths.

Notations

D	Snapshot dimension
L	Embedding length (delay-coordinates)
N	Number of snapshots
M	Working rank / number of DMD modes
m	True signal order
$\boldsymbol{x}_k$	k-th snapshot $\in \mathbb{C}^D$
$\boldsymbol{s}_k$	Clean signal snapshot
$\boldsymbol{n}_k$	Measurement noise
$\widetilde{\boldsymbol{x}}_k$	Delay vector $\in \mathbb{C}^{DL}$
$\boldsymbol{X}_0, \boldsymbol{X}_1$	Paired snapshot matrices $\in \mathbb{C}^{DL\times(N-L)}$
$\boldsymbol{S}_0, \boldsymbol{S}_1$	Clean signal matrices
$\boldsymbol{E}_0, \boldsymbol{E}_1$	Noise matrices
$\boldsymbol{A}$	Propagator operator
$\boldsymbol{A}_M$	Reduced propagator $\in \mathbb{C}^{M\times M}$
$\boldsymbol{U}_M$	Left singular vectors of $\boldsymbol{X}_0$
$\boldsymbol{\Sigma}_M$	Singular values
$\boldsymbol{V}_M$	Right singular vectors of $\boldsymbol{X}_0$
λ_j	j-th eigenvalue
ρ_j	Eigenvalue magnitude
θ_j	Eigenvalue phase
$\boldsymbol{\phi}_j$	j-th mode (signal)
$\widehat{\lambda}_j$	j-th estimated eigenvalue
$\widehat{\boldsymbol{\phi}}_j$	j-th DMD mode (generic)
$\widehat{\boldsymbol{\phi}}_j^{\mathrm{p}}$	j-th projected DMD mode

$\widehat{\boldsymbol{\phi}}_j^{\mathrm{e}}$	j-th exact DMD mode
$\mathcal{S}$	Signal subspace
$\mathcal{U}_M$	Truncated subspace col($\boldsymbol{U}_M$)
$\mathcal{U}_m$	Best m-dimensional proxy to $\mathcal{S}$
$\boldsymbol{P}_{\mathcal{U}_M}$	Orthogonal projector onto $\mathcal{U}_M$
$\boldsymbol{C}_L$	Block companion matrix
$\|\cdot\|_2$	Spectral norm / Euclidean norm
$\|\cdot\|_F$	Frobenius norm

Abbreviations

DMD	Dynamic Mode Decomposition
DC–DMD	Delay-Coordinates Dynamic Mode Decomposition
SVD	Singular Value Decomposition
SNR	Signal-to-Noise Ratio
KV	Kronecker–Vandermonde
ESR	Estimated-Subspace Residual
SSR	Signal-Subspace Residual
BIC	Bayesian Information Criterion
MDL	Minimum Description Length
GAP	Spectral Gap
STC	Spatiotemporal Consistency

1. Introduction

High-dimensional time series arise in a wide range of physical and biological systems, including fluid flows [35, 4], neural recordings [11, 8], climate dynamics [15, 43], and mechanical systems [20, 33]. Such measurements are inevitably contaminated by noise and unmodeled effects. Although the data may involve many coupled observables, the underlying dynamics are often governed by a comparatively small number of coherent components. A central challenge in data-driven dynamical analysis is therefore to extract these latent components and characterize their temporal behavior directly from measured trajectories [3].

Dynamic Mode Decomposition (DMD) is a widely used framework for this purpose. Originally introduced in the fluid dynamics community, DMD approximates the observed evolution by a best-fit linear operator learned from data snapshots and interprets its eigenpairs as spatiotemporal building blocks of the dynamics [35, 44, 22]. The associated eigenvalues encode oscillatory and growth or decay behavior, while the corresponding modes encode spatial structure. Variants of DMD and related operator-theoretic methods have since been applied across many scientific and engineering domains [32, 6, 9].

From an operator-theoretic perspective, DMD can be interpreted as a finite-dimensional, data-driven approximation of the Koopman operator, which provides a linear representation of nonlinear dynamics through its action on observables [25, 26, 47, 46]. This connection becomes particularly natural under lifting strategies that enrich the observable space and expose additional spectral content.

A common enhancement of DMD is the use of *delay coordinates*. By augmenting each snapshot with a window of consecutive measurements, delay embedding provides a systematic means of expanding the space of observables, improving spectral resolution, and strengthening the link between data-driven linear models and Koopman operator theory [12, 28, 13]. Delay coordinates originate in nonlinear dynamics and attractor reconstruction [42], and are closely related to classical subspace identification, Hankel methods, and matrix-pencil techniques [19, 17]. In the context of DMD, delay-coordinate formulations, often referred to as Hankel-DMD or Delay-Coordinates Dynamic Mode Decomposition (DC–DMD), have been shown to enhance mode separation and recover more accurate spectral information from short or noisy time series [44, 1, 23, 27, 36]. One concrete illustration is the standing-wave case, where a single real oscillation is generated by a pair of complex-conjugate components. In this setting, standard ($L = 1$) DMD often splits the oscillation into two separate modes whose individual spatial patterns are not physically meaningful, while delay coordinates recover the correct oscillatory structure; see [5].

In the presence of measurement noise, the snapshot matrices are effectively full rank, so practical implementations typically work with a truncation rank or model order that may

exceed the unknown true dynamical order. When the working order is overestimated, the computed decomposition necessarily includes additional components that have no counterpart in the underlying dynamics. In delay-coordinate settings, the lifting induces spatiotemporal organization in the embedded space, so these additional components can appear coherent across delays and can be more difficult to separate from true modes than in standard (non-delayed) DMD; as a result, common heuristics such as singular-value gap rules, eigenvalue screening, or mode-norm rankings can become unreliable.

The problem of identifying which DMD components reflect true dynamics, and which are spurious artifacts of noise and order overestimation, has therefore received considerable attention. Existing approaches include spectral heuristics [44, 22], likelihood- and information-theoretic criteria [31, 37, 45], robustness-based methods [34, 14], and structure-aware post hoc tests that exploit delay-induced spatiotemporal structure [5]. Despite this extensive body of work, there is currently no definitive solution to the problem of mode selection in DC–DMD. In particular, spurious modes are typically treated as residual artifacts, and their geometric relationship to the signal-dominated subspaces induced by the data remains largely implicit.

In this thesis, we adopt a geometric perspective on mode selection in DC–DMD. A central contribution of our work is a precise, perturbation-robust definition of *true* and *spurious* modes based on their geometric relationship to the underlying signal subspace. Rather than defining spurious modes implicitly through energy thresholds or stability heuristics, we formalize the distinction using subspace geometry and residual structure. This definition serves as a foundation for both analysis and algorithm design, and provides a unifying framework for interpreting existing selection heuristics.

Building on this geometric framework, we develop two complementary mode-selection algorithms derived from a geometric separation principle between signal-consistent and noise-consistent modes. Under standard assumptions, true DMD modes align with a low-dimensional signal subspace and concentrate their energy within it. In contrast, spurious modes arising from isotropic or weakly structured noise cannot concentrate within any moderate overestimate of this subspace and necessarily retain a non-negligible component outside it.

The first algorithm operationalizes this principle by constructing a data-driven proxy of the signal subspace and quantifying each mode's geometric residual relative to it. The resulting residual-based score follows directly from the geometry of finite-rank least-squares fits under noise.

The second algorithm builds on a structural analysis of delay-coordinate modes. Indeed, Kronecker–Vandermonde (KV) structure in true DC–DMD modes was previously identified in [5] under an explicit exponential signal model. Here, we provide a fundamentally different derivation based on the block-companion structure of least-squares propagators. This operator-based viewpoint yields two key theoretical insights. First, it shows that KV structure is induced by the delay coordinates themselves rather than by a specific signal model. Second, and critically, it reveals that a mode's adherence to the KV structure is strictly determined by its geometric residual relative to the data-driven subspace. Thus, true modes exhibit strong KV structure, while spurious modes deviate from it. This observation motivates continuous KV-deviation metrics that quantify the

degree of structural conformity, rather than treating KV structure as a binary property exclusive to true modes.

Finally, our analysis provides a principled explanation for the empirical behavior of commonly used energy- and norm-based heuristics. By decomposing the exact-mode norm into eigenvalue-dependent and residual-dependent contributions, we show that delay coordinates fundamentally alter the spectral statistics of spurious modes. In particular, as the embedding length grows, spurious eigenvalues concentrate near the unit circle, rendering magnitude- and norm-based rankings unreliable or even misleading in weakly damped regimes. This clarifies when such heuristics can be expected to fail, and why residual- and structure-based criteria provide more robust alternatives in delay-coordinate settings.

We validate the proposed theory and algorithms through extensive numerical experiments covering a wide range of operating regimes, including varying noise levels, spectral separation, damping, amplitude heterogeneity, and embedding length. Across all tested scenarios, the proposed geometric and structure-based criteria yield significantly more reliable order detection than commonly used baselines, including spectral gap rules, information-theoretic criteria, robustness-based methods, and mode-norm heuristics. In particular, the experiments confirm the theoretical predictions that delay coordinates fundamentally alter the behavior of spurious components: as the embedding length increases, spurious eigenvalues concentrate near the unit circle and spurious modes increasingly exhibit structured, KV-like organization, causing magnitude- and energy-based rankings to degrade. In contrast, the proposed residual- and structure-based scores remain stable and discriminative, accurately separating true and spurious modes even in regimes where traditional heuristics fail.

2. Theoretical Background

This chapter formalizes the delay-coordinates DMD setting and introduces the notation and terminology used throughout the thesis. The objective is to precisely define the objects produced by the method, and to distinguish between the components that correspond to the underlying signal and those that arise spuriously due to measurement noise and order overestimation.

2.1. System and data model

We observe a discrete signal given by:

$$\boldsymbol{x}_k = \boldsymbol{s}_k + \boldsymbol{n}_k \in \mathbb{C}^D, \tag{2.1}$$

where $\boldsymbol{n}$ represents a stochastic measurement noise and $\boldsymbol{s}$ follows a deterministic time-evolution rule given by:

$$\boldsymbol{s}_{k+1} = f(\boldsymbol{s}_k), \qquad \boldsymbol{s}_k \in \mathbb{C}^D. \tag{2.2}$$

Given the measurement model (2.1) and the latent evolution (2.2), our objective is to recover the dynamics of $\{\boldsymbol{s}_k\}$ from a finite, noisy observation window of N consecutive samples $\{\boldsymbol{x}_k\}_{k=0}^{N-1}$, with no access to data outside this interval.

We pursue two complementary outcomes: (i) recover the map f governing $\boldsymbol{s}_{k+1} = f(\boldsymbol{s}_k)$ when it is identifiable from the data; or (ii) when full identification is not possible, obtain a faithful characterization of the dynamics, such as its dominant time scales and invariant patterns. Such representations support tasks including denoising, by separating $\boldsymbol{s}_k$ from $\boldsymbol{n}_k$, and prediction, by forecasting $\boldsymbol{s}_{k+h}$ beyond the observation window.

2.2. Delay-coordinates DMD

We refer to each measurement vector $\boldsymbol{x}_k$ as a *snapshot*. DMD and related methods commonly use delay-coordinates where each delay vector stacks L consecutive snapshots:

$$\widetilde{\boldsymbol{x}}_k \coloneqq \begin{bmatrix} \boldsymbol{x}_k \\ \boldsymbol{x}_{k+1} \\ \vdots \\ \boldsymbol{x}_{k+L-1} \end{bmatrix} \in \mathbb{C}^{DL}, \qquad k = 0, \ldots, N-L. \tag{2.3}$$

This embedding produces an augmented trajectory matrix, from which we form the paired snapshot matrices used in the DMD least-squares fit.

Since the embedding is linear, we define $\widetilde{\boldsymbol{s}}_k$ and $\widetilde{\boldsymbol{n}}_k$ in accordance with (2.1). Collecting the delay vectors yields

$$\boldsymbol{X}^{(L)} \coloneqq \begin{bmatrix} \widetilde{\boldsymbol{x}}_0 & \cdots & \widetilde{\boldsymbol{x}}_{N-L} \end{bmatrix} = \boldsymbol{S}^{(L)} + \boldsymbol{E}^{(L)} \in \mathbb{C}^{DL\times(N-L+1)}. \tag{2.4}$$

From this point on, fix $L \geq 1$ and, for brevity, write $\boldsymbol{X}, \boldsymbol{S}, \boldsymbol{E}$ for $\boldsymbol{X}^{(L)}, \boldsymbol{S}^{(L)}, \boldsymbol{E}^{(L)}$.

Classical delay–embedding theory shows that, under generic conditions, a delay map embeds the underlying attractor [42], providing a standard basis for using delay-coordinates to enrich observed dynamics. In practice, delay-coordinates are widely used to improve modal separation and spectral resolution in data–driven analysis of dynamical systems [19, 1, 44, 23].

To estimate a one-step propagator on delay vectors, we pair each embedded snapshot with its immediate successor and form the standard matrices $\boldsymbol{X}_0$ and $\boldsymbol{X}_1$.

DMD approximates the temporal evolution map f from (2.2) by a linear propagator over the available data window:

$$\widetilde{\boldsymbol{x}}_{k+1} \approx \boldsymbol{A}\,\widetilde{\boldsymbol{x}}_k. \tag{2.5}$$

which motivates pairing each delay vector with its immediate successor in time. The paired snapshot matrices are defined as

$$\boldsymbol{X}_0 = \begin{bmatrix} \widetilde{\boldsymbol{x}}_0 & \widetilde{\boldsymbol{x}}_1 & \cdots & \widetilde{\boldsymbol{x}}_{N-L-1} \end{bmatrix}, \quad \boldsymbol{X}_1 = \begin{bmatrix} \widetilde{\boldsymbol{x}}_1 & \widetilde{\boldsymbol{x}}_2 & \cdots & \widetilde{\boldsymbol{x}}_{N-L} \end{bmatrix}, \tag{2.6}$$

both of size $DL \times (N-L)$. An analogous construction defines $\boldsymbol{S}_0, \boldsymbol{S}_1, \boldsymbol{E}_0, \boldsymbol{E}_1$, following (2.4).

Linear evolution (i.e. $\boldsymbol{s}_{k+1} = \boldsymbol{A}\boldsymbol{s}_k$), along with mild, well-justified assumptions regarding the nature of the matrix $\boldsymbol{A}$ (see details in Appendix), imply that any trajectory is of the form

$$\boldsymbol{s}_k \;=\; \sum_{j=1}^{m} b_j\,\boldsymbol{\phi}_j\,\lambda_j^k \;=\; \boldsymbol{\Phi}\,\boldsymbol{\Lambda}^k\boldsymbol{b}, \tag{2.7}$$

where $(\lambda_j, \boldsymbol{\phi}_j)$ are eigenpairs of $\boldsymbol{A}$ and the amplitudes b_j are set by the initial condition. Throughout we parameterize the eigenvalues as

$$\lambda_j \;=\; \rho_j\,\mathrm{e}^{\mathrm{i}\theta_j}, \qquad \rho_j \in [0,1], \tag{2.8}$$

where ρ_j are magnitudes, θ_j are phases, and $\rho_j \leq 1$ ensures bounded temporal behavior. The triples $(\rho_j, \theta_j, \boldsymbol{\phi}_j)$ together with b_j are the signal parameters; throughout we refer to them as the true underlying signal components. We define the order m as the number of components in the trajectory. This modal form holds exactly when $\boldsymbol{A}$ is diagonalizable, and in the non-diagonalizable case polynomial–exponential dynamics are typically represented in finite-data DMD by clusters of nearby simple eigenvalues; see [44, 22, 1].

Equivalently, the paired snapshot matrices satisfy $\operatorname{rank}(\boldsymbol{S}_0) = \operatorname{rank}(\boldsymbol{S}_1) = m$, since all columns lie in $\operatorname{span}\{\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_m\}$. We therefore use "order" and "signal rank" interchangeably. As is typical in a wide range of systems, the number of underlying dynamical

components is substantially smaller than the number of observables, so we adopt the standard assumption that

$$m \ll \min(N-L,\, DL).$$

The maximal algebraic rank of the signal matrices is $\min(N-L,\, DL)$; throughout we focus on the common regime $N-L < DL$ (the complementary case $DL < N-L$ is uncommon and discussed in Appendix A).

Although the linear evolution is postulated for the latent signal, $\boldsymbol{S}_1 = \boldsymbol{A}_s\boldsymbol{S}_0$, in practice we estimate from the measurements by solving $\boldsymbol{X}_1 \approx \boldsymbol{A}\boldsymbol{X}_0$. This implicitly assumes only that the noise is additive and non-dynamical (i.e., not strongly correlated with the signal), so that the measurement relation is a perturbation of the signal relation.

Effect of noise and order ambiguity

In the measurements $\boldsymbol{X}_\bullet = \boldsymbol{S}_\bullet + \boldsymbol{E}_\bullet$, additive noise obscures the low-rank structure of the signal, rendering $\boldsymbol{X}_0$ and $\boldsymbol{X}_1$ effectively full rank. Consequently, the true order m cannot be inferred from ranks alone. Noise affects the estimation process in two coupled ways:

- It perturbs the spatial structure of the signal subspace, with a magnitude that depends on the noise level.
- It introduces additional components that have no signal counterpart. These *spurious* modes arise even under infinitesimal noise.

Because the true order m is unknown, we proceed with a working rank $M \geq m$. Whether M coincides with m, and how deviations between them manifest, lies at the core of this work and is formalized as an *order detection* problem in Section 2.3.

2.2.1. DMD algorithm

Given the paired snapshot matrices $(\boldsymbol{X}_0, \boldsymbol{X}_1)$, a natural unconstrained least–squares solution, via the Moore–Penrose pseudoinverse [29, 2], is

$$\boldsymbol{A}_{\mathrm{MP}} = \boldsymbol{X}_1\boldsymbol{X}_0^{\dagger} \tag{2.9}$$

However, under measurement noise, this produces a (numerically) full-rank operator that absorbs high-variance directions and thereby *models noise as dynamics*. This undermines the identifiability of the true components. The rank-constrained formulation in (2.10) instead enforces the low-rank signal model, restricting the fit to an M-dimensional subspace so the ensuing eigenvalue-mode pairs reflect coherent structure rather than noise.

The low-rank assumption motivates solving a rank-constrained least squares:

$$\boldsymbol{A} \in \arg\min_{\substack{\boldsymbol{A}\in\mathbb{C}^{DL\times DL} \\ \mathrm{rank}(\boldsymbol{A})=M}} \|\boldsymbol{X}_1 - \boldsymbol{A}\boldsymbol{X}_0\|_F^2 \tag{2.10}$$

DMD extracts the eigenpairs of the propagator $\boldsymbol{A}$. These encode the spatial and temporal structure of the data and correspond to the underlying signal parameters defined in (2.7), with their practical estimation illustrated in Section 2.2.2. This yields an explicit spatiotemporal representation: the eigenvalues capture temporal evolution while the eigenvectors capture spatial structure. In this form the signal can be *modeled and reconstructed* directly, rather than merely advanced one step by $\boldsymbol{A}$.

The practical solution to the rank-constrained least-squares problem (2.10) uses the truncated Singular Value Decomposition (SVD) of $\boldsymbol{X}_0$:

$$\begin{gathered}\boldsymbol{X}_0 \;\approx\; \boldsymbol{U}_M\,\boldsymbol{\Sigma}_M\,\boldsymbol{V}_M^{\mathrm{H}},\\ \boldsymbol{U}_M \in \mathbb{C}^{DL\times M},\;\; \boldsymbol{\Sigma}_M \in \mathbb{R}^{M\times M},\;\; \boldsymbol{V}_M \in \mathbb{C}^{(N-L)\times M}.\end{gathered} \tag{2.11}$$

These matrices are used to construct the truncated propagator:

$$\boldsymbol{A}_M \;\coloneqq\; \boldsymbol{U}_M^{\mathrm{H}}\,\boldsymbol{X}_1\,\boldsymbol{V}_M\,\boldsymbol{\Sigma}_M^{-1} \;\in\; \mathbb{C}^{M\times M}. \tag{2.12}$$

$$\begin{aligned}\boldsymbol{A}_M &= \boldsymbol{W}\,\widehat{\boldsymbol{\Lambda}}\,\boldsymbol{W}^{-1},\\ \widehat{\boldsymbol{\Lambda}} &= \operatorname{diag}(\widehat{\lambda}_1,\dots,\widehat{\lambda}_M) \in \mathbb{C}^{M\times M},\\ \boldsymbol{W} &= [\,\boldsymbol{w}_1\;\cdots\;\boldsymbol{w}_M\,] \in \mathbb{C}^{M\times M}.\end{aligned} \tag{2.13}$$

Projected DMD uses the **same truncated basis** that defined $\boldsymbol{A}_M$,

$$\widehat{\boldsymbol{\phi}}_j^{\mathrm{p}} \coloneqq \boldsymbol{U}_M\boldsymbol{w}_j, \tag{2.14}$$

so $\widehat{\boldsymbol{\phi}}_j^{\mathrm{p}} \in \operatorname{col}(\boldsymbol{U}_M)$. We denote the column space of $\boldsymbol{U}_M$ by

$$\mathcal{U}_M \coloneqq \operatorname{col}(\boldsymbol{U}_M). \tag{2.15}$$

Exact DMD uses the (untruncated) data that appears on the right-hand side of the construction of $\boldsymbol{A}_M$,

$$\widehat{\boldsymbol{\phi}}_j^{\mathrm{e}} \coloneqq \boldsymbol{X}_1\boldsymbol{V}_M\boldsymbol{\Sigma}_M^{-1}\boldsymbol{w}_j, \tag{2.16}$$

so $\widehat{\boldsymbol{\phi}}_j^{\mathrm{e}} \in \operatorname{col}(\boldsymbol{X}_1)$ and is not restricted to any subspace defined by the singular vectors of $\boldsymbol{X}_0$.

Notation. In what follows, a hatted mode without a superscript, $\widehat{\boldsymbol{\phi}}_j$, denotes a generic representative that may be either the projected DMD mode $\widehat{\boldsymbol{\phi}}_j^{p}$ or the exact DMD mode $\widehat{\boldsymbol{\phi}}_j^{e}$. Statements written with $\widehat{\boldsymbol{\phi}}_j$ hold for both choices unless p or e is shown explicitly.

Since $\|\boldsymbol{w}_j\|_2 = 1$ and $\boldsymbol{U}_M$ has orthonormal columns, by their construction, projected modes are unit vectors:

$$\|\boldsymbol{\phi}_j^{\mathrm{p}}\|_2 = \|\boldsymbol{U}_M\boldsymbol{w}_j\|_2 = 1.$$

Since the columns of $\boldsymbol{U}_M$ are orthonormal, the orthogonal projector onto $\mathcal{U}_M$ is

$$\boldsymbol{P}_{\mathcal{U}_M} \coloneqq \boldsymbol{U}_M\boldsymbol{U}_M^{\mathrm{H}}. \tag{2.17}$$

The relationship between the exact modes and their projected counterparts can be explicitly written:

$$\boldsymbol{P}_{\mathcal{U}_M}\,\widehat{\boldsymbol{\phi}}_j^{\mathrm{e}} \;=\; \widehat{\lambda}_j\,\widehat{\boldsymbol{\phi}}_j^{\mathrm{p}}, \tag{2.18}$$

which shows that they differ only by projection and scaling (this follows from the definition of $\boldsymbol{A}_M$ in (2.12)).

In the ideal noiseless case with correct order ($M = m$), the projected and exact DMD constructions coincide: they yield the same eigenvalues and the same modes in the delay-coordinates space. These modes live in the delay-coordinates space $\mathbb{C}^{DL}$ (which reduces to $\mathbb{C}^D$ when $L = 1$), rather than in the original snapshot space, and should be read as delay-coordinates realizations of the underlying dynamics.

At this stage, we emphasize the distinction between the *underlying signal eigenvalues* $\{\lambda_j\}_{j=1}^m$ and *underlying signal modes* $\{\boldsymbol{\phi}_j\}_{j=1}^m$ in (2.7), and the eigenvalues and modes computed by DMD. The latter are estimates of the former. These estimates become useful once we fix amplitudes from the initial embedded snapshot, yielding a finite-rank reconstruction and a direct prediction rule.

2.2.2. Finite-rank reconstruction

Collecting the modes (from either lifting scheme) into

$$\widehat{\boldsymbol{\Phi}} = [\,\widehat{\boldsymbol{\phi}}_1\;\cdots\;\widehat{\boldsymbol{\phi}}_M\,] \in \mathbb{C}^{DL\times M},$$

Amplitudes $\widehat{\boldsymbol{b}}$ are determined by enforcing the initial condition, i.e., by solving $\widehat{\boldsymbol{\Phi}}\widehat{\boldsymbol{b}} \approx \widetilde{\boldsymbol{x}}_0$, which fixes the overall scaling.

The M–term approximation of the embedded snapshot is then

$$\widehat{\boldsymbol{x}}_k \;=\; \widehat{\boldsymbol{\Phi}}\,\widehat{\boldsymbol{\Lambda}}^{\,k}\,\widehat{\boldsymbol{b}}, \qquad k = 0,\ldots,N-L. \tag{2.19}$$

Equation (2.19) extends naturally to prediction: for any $h \geq 1$,

$$\widehat{\boldsymbol{x}}_{k+h} \;=\; \widehat{\boldsymbol{\Phi}}\,\widehat{\boldsymbol{\Lambda}}^{\,k+h}\,\widehat{\boldsymbol{b}}, \qquad k > N-L.$$

Here $\widehat{\boldsymbol{x}}_{k+h}$ is a prediction in the augmented (delay-coordinates) space. The corresponding prediction in the original data space is obtained by extracting its first D entries, as dictated by the delay-coordinates embedding in (2.3).

2.3. Signal space, order detection, and component selection

The core problem addressed in this thesis is that the true order m is unknown. Common strategies range from heuristic thresholds to information criteria and robustness tests (see Chapter 3). We adopt a post hoc approach: choose an upper bound $M \geq m$, compute DMD with order M to obtain $\{(\widehat{\lambda}_j, \widehat{\boldsymbol{\phi}}_j)\}_{j=1}^M$, and then retain the m modes that

correspond to the underlying dynamics (labeled *true*), labeling the remaining $M - m$ as *spurious*. This formulation serves as the working setup for the component–selection methods developed later in the thesis.

We define the *signal subspace* of the underlying dynamics as

$$\mathcal{S} \coloneqq \operatorname{span}\{\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_m\}, \tag{2.20}$$

where $\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_m$ are the true underlying signal modes from equation (2.7). By construction, the clean snapshot matrices satisfy

$$\operatorname{col}(\boldsymbol{S}_0) = \operatorname{col}(\boldsymbol{S}_1) = \mathcal{S}. \tag{2.21}$$

We assume that the noise level is sufficiently low such that the signal subspace can still be identified from the principal components of the data.

In practice, the leading left singular vectors of the data matrix $\boldsymbol{X}_0$ provide a data-driven basis for representing components of the signal subspace $\mathcal{S}$. Fix a working rank $M \geq m$, and let

$$\boldsymbol{U}_M \coloneqq [\boldsymbol{u}_1, \ldots, \boldsymbol{u}_M], \qquad \mathcal{U}_M \coloneqq \operatorname{col}(\boldsymbol{U}_M) \tag{2.22}$$

denote the corresponding M-dimensional truncation subspace.

We quantify proximity between subspaces using the spectral norm of the difference between their orthogonal projectors. For subspaces $\mathcal{S}$ and $\mathcal{U}$ with projectors $\boldsymbol{P}_{\mathcal{S}}$ and $\boldsymbol{P}_{\mathcal{U}}$, we use the projector distance

$$\left\|\boldsymbol{P}_{\mathcal{S}} - \boldsymbol{P}_{\mathcal{U}}\right\|_2 \in [0, 1]. \tag{2.23}$$

When $\dim \mathcal{S} = \dim \mathcal{U}$, this quantity equals $\sin \theta_{\max}(\mathcal{S}, \mathcal{U})$, where $\theta_{\max}$ is the largest principal angle.

If the true order m were known and $\mathcal{S}$ were observable, the best m-dimensional proxy to $\mathcal{S}$ *within* $\mathcal{U}_M$ need not be spanned by the first m singular vectors. Accordingly, we define $\mathcal{U}_m$ by selecting the subset of m singular directions (among the M available) that minimizes the projector distance (2.23) to $\mathcal{S}$. Specifically, for any index set $\mathcal{I} \subset \{1, \ldots, M\}$, write $\boldsymbol{U}_{\mathcal{I}}$ for the submatrix formed by the columns $\{\boldsymbol{u}_i\}_{i \in \mathcal{I}}$, and define

$$\mathcal{I}_m \in \arg \min_{\mathcal{I} \subset \{1,\ldots,M\},\, |\mathcal{I}| = m} \left\|\boldsymbol{P}_{\mathcal{S}} - \boldsymbol{P}_{\operatorname{col}(\boldsymbol{U}_{\mathcal{I}})}\right\|_2, \tag{2.24}$$

$$\mathcal{U}_m \coloneqq \operatorname{col}(\boldsymbol{U}_{\mathcal{I}_m}). \tag{2.25}$$

In particular, $\mathcal{U}_m$ does not necessarily coincide with the span of the leading m singular vectors.

Assumption 2.1 (Signal-subspace proximity)**.** *Let $\mathcal{S}$ be the signal subspace and let $\mathcal{U}_m \subseteq \mathcal{U}_M$ be the m-dimensional proxy subspace defined by the subset-selection rule* (2.24)-(2.25)*. We assume*

$$\left\|\boldsymbol{P}_{\mathcal{S}} - \boldsymbol{P}_{\mathcal{U}_m}\right\|_2 \leq \eta, \qquad \eta \in [0, 1), \tag{2.26}$$

Since $\dim \mathcal{S} = \dim \mathcal{U}_m = m$*,* (2.26) *is equivalent to*

$$\theta_{\max}(\mathcal{S}, \mathcal{U}_m) \leq \arcsin \eta.$$

Under Assumption 2.1,

$$\mathcal{S} \approx \mathcal{U}_m \subseteq \mathcal{U}_M, \tag{2.27}$$

where "≈" refers to the projector-distance bound (2.26), and the inclusion holds since $\mathcal{U}_m$ is spanned by a subset of the columns of $\boldsymbol{U}_M$.

Heuristically, η improves with higher Signal-to-Noise Ratio (SNR), better spectral separation, better spatial conditioning, and larger embedding dimension or window length.

Choosing a working rank $M \geq m$ yields M candidate DMD components even though the underlying dynamics are m-dimensional, so component selection is unavoidable. While many practical rules exist, they often leave implicit what it means for a recovered component to be *true* under perturbations. In this work, we adopt an explicit perturbation-robust definition of true components and use it to derive principled consequences for selection.

3. Literature Review

This chapter reviews existing approaches for order detection and mode selection in DMD, which can be grouped into four broad families. The following discussion assumes the mathematical framework established in Chapter 2. Comprehensive overviews of DMD variants, selection heuristics, and algorithmic foundations are provided by [36, 9].

3.1. Spectral and likelihood-based mode selection

Spectral heuristics are widely used for order detection in DMD and choose the truncation rank using gap- or threshold-based rules applied to the singular values of the snapshot matrix [44, 22]. In the context of delay-coordinates, analogous rules are applied to the singular values of the corresponding block-Hankel snapshot matrices [23]. Eigenvalue screening rules based on magnitude or phase are also commonly employed in practice [22]. Closely related approaches formulate order detection as penalized likelihood model selection, trading off fit quality with a complexity penalty, for example via Minimum Description Length (MDL) or Bayesian Information Criterion (BIC) [31, 37, 41, 45]. Within the DMD framework, likelihood-based selection has been pursued through penalized optimization formulations [18, 30] or Bayesian inference approaches [28].

3.2. Robustness-based mode selection

Robustness-based strategies retain components that persist under data resampling, held-out refits, or other perturbations of the snapshots. A prominent example is bagging optimized DMD (BOP-DMD), which fits ensembles across randomized data subsets and selects modes based on their empirical consistency [34]. Other stability-based methods assess mode validity using diagnostics such as time-reversal consistency [14].

3.3. Post hoc and structure-aware mode selection

These methods compute DMD at an overestimated truncation rank and then perform mode selection in a post-processing step. In contrast to rank estimation or eigenvalue-centric selection methods, some approaches are mode-centric and exploit structural properties of the computed modes. In settings involving delay-coordinates, representative examples test for KV structure across delays in the resulting delay-coordinates modes [5]. Related post-processing strategies have also been developed in closely connected subspace and matrix-pencil settings [39, 38]. A complementary post hoc direction uses

residual-based verification to filter or rank eigenpairs by numerical consistency, aiming to mitigate spectral pollution; while this provides an eigenpair-level diagnostic, it does not consider the organization of modes induced by delay-coordinates [10].

3.4. Energy- and amplitude-based heuristics

These rank modes by simple magnitude proxies such as Euclidean norms or reconstruction amplitudes [35, 44]. These should be understood as practical magnitude proxies, rather than as likelihood- or energy-based models. These heuristics are appealing due to their simplicity and low computational cost. However, norms and amplitudes depend on algorithmic choices and can be sensitive to noise.

3.5. Summary of Gaps

Taken together, the above approaches address order detection and mode selection from complementary perspectives, but important gaps remain. Spectral heuristics rely on global summaries and degrade under noise, spectral crowding, or damping. Information-criterion methods require refitting models across hypothesized orders, making them computationally expensive and sensitive to noise-model mismatch. Robustness-based approaches emphasize empirical stability but do not provide a structural interpretation of the recovered components or a principled definition of true versus spurious modes. Norm- and amplitude-based heuristics are simple to apply but can become unreliable in delay-coordinates settings and lack a clear mechanistic justification.

Structure-aware post hoc methods explicitly formalize the expected organization of *true* modes arising from delay-coordinates, but the corresponding characterization of *spurious* modes remains largely implicit. Spurious components are typically treated as unstructured residual artifacts rather than as objects with their own systematic structure. As a result, existing methods do not exploit spurious-mode geometry in selection rules, leaving their behavior under delay-coordinates, and their interaction with true modes, insufficiently understood.

4. Mode geometry, signal subspace residual, and true and spurious modes

This chapter presents a geometric analysis of the DMD modes and, based on it, derives a practical method for mode selection.

By definition, the signal subspace $\mathcal{S}$ spans the m underlying signal modes $\{\boldsymbol{\phi}_j\}_{j=1}^m$ (2.20). However, in practice, m is unknown and the underlying signal modes are accessed only through the $M > m$ computed DMD modes $\{\widehat{\boldsymbol{\phi}}_j\}_{j=1}^M$. We now define a partition of the computed DMD modes into *true* and *spurious* modes, where the true modes can be viewed as noisy estimators of the underlying signal modes.

Definition 4.1 (Signal-subspace residual vector)**.** *For any vector $\boldsymbol{v} \in \mathbb{C}^{DL}$, the Signal-Subspace Residual (SSR) vector is defined by*

$$\boldsymbol{r}_{\mathcal{S}}(\boldsymbol{v}) \;\coloneqq\; (\boldsymbol{I} - \boldsymbol{P}_{\mathcal{S}})\,\boldsymbol{v}, \tag{4.1}$$

where $\boldsymbol{P}_{\mathcal{S}}$ is the orthogonal projector onto the signal subspace $\mathcal{S}$.

We define the *true* modes to be the m modes with the smallest residual norms, and the remaining $M - m$ modes to be *spurious*.

Definition 4.2 (True and spurious modes)**.** *Given M computed modes $\{\widehat{\boldsymbol{\phi}}_j\}_{j=1}^M$ (either exact or projected), let $\mathcal{J}_{\text{true}} \subset \{1, \dots, M\}$ be a set of m indices attaining the m smallest values of*

$$\left\|\boldsymbol{r}_{\mathcal{S}}(\widehat{\boldsymbol{\phi}}_j)\right\|_2^2, \qquad j = 1, \dots, M,$$

and let $\mathcal{J}_{\text{spur}} \coloneqq \{1, \dots, M\} \setminus \mathcal{J}_{\text{true}}$. For each $j \in \{1, \dots, M\}$, $\widehat{\boldsymbol{\phi}}_j$ is a true mode *if $j \in \mathcal{J}_{\text{true}}$, and a* spurious mode *otherwise.*

Definition 4.2 suggests that the modes can be classified as true or spurious by evaluating their residual energy outside the signal subspace $\mathcal{S}$. However, since $\mathcal{S}$ cannot be directly observed, this definition serves only as an ideal reference for distinguishing between true and spurious modes. To bridge this gap, we introduce a practical criterion for differentiating true and spurious DMD modes.

By Assumption 2.1, the m-dimensional subspace $\mathcal{U}_m$ is close to $\mathcal{S}$. However, the true order m, and therefore $\mathcal{U}_m$, are unknown. We postulate that $\mathcal{U}_M$, where $M > m$ is an arbitrary overestimate of the true order, can serve as a practical proxy for $\mathcal{S}$. We therefore propose to measure residual energy with respect to $\mathcal{U}_M$ rather than $\mathcal{S}$, using the orthogonal projector $\boldsymbol{P}_{\mathcal{U}_M}$.

(a) Signal Subspace Residual (SSR) (b) Estimated Subspace Residual (ESR)

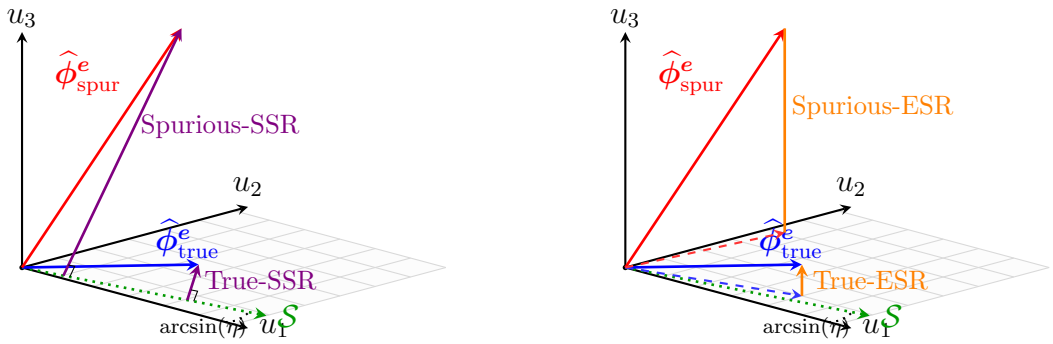

Figure 4.1.: Geometric illustration of signal-subspace residual (SSR) and estimated-subspace residual (ESR) in the case $m = 1$ and $M = 2$, where the true signal subspace $\mathcal{S}$ (green) lies inside the truncation subspace $\mathcal{U}_M = \text{span}\{u_1, u_2\}$. True and spurious modes are drawn in blue and red, respectively, together with their projections onto $\mathcal{U}_M$ (dashed). (a) SSR corresponds to the in-plane deviation of a mode from $\mathcal{S}$, shown as violet arrows. (b) ESR corresponds to the out-of-plane component of a mode, shown as orange arrows. In this constructed example, the true mode has small SSR and ESR, whereas the spurious mode has larger residual in both senses.

Definition 4.3 (Estimated-subspace residual vector)**.** *For an exact DMD mode $\widehat{\boldsymbol{\phi}}^e_j$, define the Estimated-Subspace Residual (ESR) vector by*

$$\boldsymbol{r}_{\mathcal{U}_M}\big(\widehat{\boldsymbol{\phi}}^e_j\big) \;\coloneqq\; (\boldsymbol{I} - \boldsymbol{P}_{\mathcal{U}_M})\,\widehat{\boldsymbol{\phi}}^e_j, \qquad j = 1, \ldots, M. \tag{4.2}$$

We note that Definition 4.3 is only informative for exact DMD modes because the projected modes lie in $\mathcal{U}_M$ by construction, and so their residual is $\boldsymbol{r}_{\mathcal{U}_M}\big(\widehat{\boldsymbol{\phi}}^p_j\big) \equiv 0$.

Figure 4.1 summarizes the geometry of the SSR and the ESR and highlights that ESR captures the $\mathcal{U}_M^{\perp}$ component of SSR.

We propose to estimate the partition of modes into true and spurious based on the squared norm of the ESR vectors:

$$\mathcal{R}_j \;\coloneqq\; \big\|\boldsymbol{r}_{\mathcal{U}_M}(\widehat{\boldsymbol{\phi}}^e_j)\big\|_2^2. \tag{4.3}$$

Specifically, the scores in logarithmic scale $\zeta_j \coloneqq \log(\mathcal{R}_j + \varepsilon)$ for some small $\varepsilon > 0$, are clustered into two groups, and the cluster with the smaller values ζ_j is labeled *true*. Algorithm 4.1 summarizes this procedure.

In Section 6, we empirically show that Algorithm 4.1 reliably separates true and spurious modes and outperforms standard baselines. The intuition behind the separating power based on the norm of the ESR vectors can be described as follows. If $\widehat{\boldsymbol{\phi}}^e_j$ is a true mode, it is dominated by the signal component, so its energy concentrates on the signal subspace $\mathcal{S}$ and therefore, by Assumption 2.1, on $\mathcal{U}_m$. Since $\mathcal{U}_m \subset \mathcal{U}_M$, a true

mode typically places very little energy outside $\mathcal{U}_M$ (and hence in $\mathcal{U}_M^\perp$), so $\left\|\boldsymbol{r}_{\mathcal{U}_M}(\widehat{\boldsymbol{\phi}}_j^e)\right\|_2^2$ is small.

Conversely, if $\widehat{\boldsymbol{\phi}}_j^e$ is a spurious mode, it is a noise-driven artifacts of the finite-rank least-squares fit rather than of coherent low-dimensional structure in the data. Such modes are not associated with the signal subspace and therefore are not confined to any low-dimensional subspace, typically placing non-negligible energy outside $\mathcal{U}_M$. This behavior is consistent with well-documented observations of spectral pollution and noise-induced modes in practical DMD computations [36, 10]. Consequently, a spurious mode generally has a larger component in $\mathcal{U}_M^\perp$, and thus $\|\boldsymbol{r}_{\mathcal{U}_M}(\widehat{\boldsymbol{\phi}}_j^e)\|_2^2$ is large relative to that of a true mode.

Require: Eigenpairs $\{(\widehat{\lambda}_j, \widehat{\boldsymbol{\phi}}_j^e)\}_{j=1}^M$, small $\varepsilon > 0$; clustering routine and aggregate as in Algorithm 4.2
Ensure: Order estimate $\widehat{m}$ and labels $\{\ell_j\}$
1: **for** $j = 1, \ldots, M$ **do**
2: $\quad \mathcal{R}_j \leftarrow \|\widehat{\boldsymbol{\phi}}_j^e\|_2^2 - |\widehat{\lambda}_j|^2$ ▷ by (4.6)
3: $\quad \zeta_j \leftarrow \log(\mathcal{R}_j + \varepsilon)$
4: **end for**
5: Form score vector $\boldsymbol{f} = (\zeta_1, \ldots, \zeta_M)^\top$
6: Apply binary mode selection framework (Algorithm 4.2) to $\boldsymbol{f}$ to obtain labels $\{\ell_j\}$ and $\widehat{m}$

Algorithm 4.1.: Mode selection based on the norms of the residual vectors.

To compute the residual norm efficiently, following (2.18), we decompose the exact mode as

$$\widehat{\boldsymbol{\phi}}_j^e = \widehat{\lambda}_j \, \widehat{\boldsymbol{\phi}}_j^p + \boldsymbol{r}_{\mathcal{U}_M}(\widehat{\boldsymbol{\phi}}_j^e). \tag{4.4}$$

Since $\|\widehat{\boldsymbol{\phi}}_j^p\|_2 = 1$ and $\boldsymbol{r}_{\mathcal{U}_M}(\widehat{\boldsymbol{\phi}}_j^e) \in \mathcal{U}_M^\perp$, the two terms in (4.4) are orthogonal, and hence

$$\left\|\widehat{\boldsymbol{\phi}}_j^e\right\|_2^2 = |\widehat{\lambda}_j|^2 + \left\|\boldsymbol{r}_{\mathcal{U}_M}(\widehat{\boldsymbol{\phi}}_j^e)\right\|_2^2. \tag{4.5}$$

Therefore,

$$\left\|\boldsymbol{r}_{\mathcal{U}_M}(\widehat{\boldsymbol{\phi}}_j^e)\right\|_2^2 = \|\widehat{\boldsymbol{\phi}}_j^e\|_2^2 - |\widehat{\lambda}_j|^2. \tag{4.6}$$

The expression in (4.6) implies that the residual norms can be computed directly from the norms of the exact modes and the magnitudes of the eigenvalues, without forming $\boldsymbol{P}_{\mathcal{U}_M}$. This adds only a negligible cost beyond the cost of SVD used in the standard delay-coordinates DMD. For a complete complexity analysis, see Appendix F.

We end this chapter with two remarks. First, we note that the clustering step in Algorithm 4.1 is generic and can be used with any per-mode scalar score. In particular, one may pass $f_j = \log(\mathcal{R}_j + \varepsilon)$ (or any other appropriate scaling) into the framework in Algorithm 4.2.

Second, since the scores $\mathcal{R}_j$ can be computed directly from the rank-M DMD eigenpairs via (4.6), the mode selection in Algorithm 4.1 applies to the classical case, where $L = 1$ without delay-coordinates.

Require: Score vector $\boldsymbol{f} = (f_1, \ldots, f_M)^\top \in \mathbb{R}^M$ (smaller is better), clustering routine with $k = 2$ (e.g., K-Means, GMM), aggregate agg (default: mean)
Ensure: Labels $\ell_j \in \{\text{true}, \text{spurious}\}$ and order estimate $\widehat{m}$
1: (optional) Normalize $\boldsymbol{f}$ (e.g., affine rescaling)
2: $\{\mathcal{C}_1, \mathcal{C}_2\} \leftarrow \textsc{Cluster}(\boldsymbol{f}, k = 2)$
3: $\mathcal{C}_{\text{true}} \leftarrow \arg\min_{C \in \{\mathcal{C}_1, \mathcal{C}_2\}} \text{agg}(\{f_j\}_{j \in C})$
4: **for** $j = 1, \ldots, M$ **do**
5: $\quad \ell_j \leftarrow$ true if $j \in \mathcal{C}_{\text{true}}$ else $\ell_j \leftarrow$ spurious
6: **end for**
7: $\widehat{m} \leftarrow |\mathcal{C}_{\text{true}}|$

Algorithm 4.2.: Binary mode selection framework

5. Kronecker–Vandermonde structure in modes

Assuming an exponential signal model as in (2.7), Bronstein *et al.*[5] showed that, when using delay-coordinates, each true mode exhibits a KV structure of the form

$$\boldsymbol{v}_L(\lambda_j) \otimes \boldsymbol{\phi}_j \;=\; \begin{bmatrix} \boldsymbol{\phi}_j \\ \lambda_j\, \boldsymbol{\phi}_j \\ \vdots \\ \lambda_j^{L-1}\, \boldsymbol{\phi}_j \end{bmatrix} \in \mathbb{C}^{DL}, \tag{5.1}$$

where $\boldsymbol{v}_L(\lambda) \in \mathbb{C}^L$ is the Vandermonde vector:

$$\boldsymbol{v}_L(\lambda) \;=\; [\,1,\; \lambda,\; \ldots,\; \lambda^{L-1}\,]^\top, \tag{5.2}$$

and $\otimes$ is the Kronecker product, defined by

$$\boldsymbol{a} \otimes \boldsymbol{b} \;\coloneqq\; \begin{bmatrix} a_1\, \boldsymbol{b} \\ a_2\, \boldsymbol{b} \\ \vdots \\ a_L\, \boldsymbol{b} \end{bmatrix} \in \mathbb{C}^{DL}. \tag{5.3}$$

for any $\boldsymbol{a} \in \mathbb{C}^L$ and $\boldsymbol{b} \in \mathbb{C}^D$. See App. C for a compact derivation of this result. Here, we extend this analysis in three ways. First, we derive the same KV structure without assuming a signal model, showing that this structure is induced by the delay-coordinates rather than by the signal model. Second, our new derivation shows that KV structure can arise in both *true* and *spurious* components, often with different strengths (with true modes adhering significantly more strongly). This motivates continuous KV-deviation metrics that quantify the extent to which each mode conforms to the KV template, while avoiding undue sensitivity to small, localized perturbations. Third, we link the new derivation of the KV structure to the ESR (Definition 4.3) and use their relationship to devise a new mode-selection algorithm, complementary to Algorithm 4.1, which is based on the ESR, and to the algorithm proposed in Ref. [5].

We note that this KV structure implies that, in delay-coordinates, a mode no longer represents a purely spatial pattern in $\mathbb{C}^D$, but rather a coupled spatiotemporal pattern in $\mathbb{C}^{DL}$: its D-dimensional lagged segments are tied across lags by a geometric progression parameterized by λ_j.

5.1. Block-companion minimizer and compression

To derive the KV structure without assuming a particular signal model, we examine the following unconstrained least-squares problem:

$$\min_{\boldsymbol{A}\in\mathbb{C}^{DL\times DL}} \|\boldsymbol{X}_1 - \boldsymbol{A}\boldsymbol{X}_0\|_F^2, \tag{5.4}$$

where $\boldsymbol{X}_0, \boldsymbol{X}_1 \in \mathbb{C}^{DL\times N}$ are the snapshot matrices defined in (2.6). Although the minimizer of (5.4) is not necessarily unique, there exists a minimizer whose structure is dictated by the delay-coordinates.

Lemma 5.1 (Block-companion minimizer)**.** *The optimization problem* (5.4) *admits the minimizer* $\boldsymbol{C}_L$ *with the following block-companion form:*

$$\boldsymbol{C}_L = \begin{bmatrix} \boldsymbol{0} & \boldsymbol{I}_D & & & \boldsymbol{0} \\ & \ddots & \ddots & & \\ & & \boldsymbol{0} & \boldsymbol{I}_D & \boldsymbol{0} \\ \boldsymbol{B}_1 & \boldsymbol{B}_2 & \cdots & \boldsymbol{B}_{L-1} & \boldsymbol{B}_L \end{bmatrix} \in \mathbb{C}^{DL\times DL}, \tag{5.5}$$

where $\boldsymbol{B}_\ell \in \mathbb{C}^{D\times D}$.

Proof See Appendix D.1.

Lemma 5.1 shows that the delay-coordinates are explicitly encoded in a minimizer: the top $L-1$ block rows implement the one-step shift between consecutive delay blocks, while the last block row defines a linear predictor of the next sample from the stacked history $\widetilde{\boldsymbol{x}}_k = [\,\boldsymbol{x}_k^\top, \ldots, \boldsymbol{x}_{k+L-1}^\top\,]^\top$, namely $\boldsymbol{x}_{k+L} \approx \boldsymbol{B}\,\widetilde{\boldsymbol{x}}_k$, where $\boldsymbol{B} = [\,\boldsymbol{B}_1 \ \cdots \ \boldsymbol{B}_L\,] \in \mathbb{C}^{D\times DL}$. Throughout, we fix $\boldsymbol{C}_L$ by taking $\boldsymbol{B}$ to be the Moore-Penrose (pseudoinverse) least-squares predictor; Appendix D.1 makes this choice explicit.

Lemma 5.2 (Eigenvectors of block-companion matrices (see, e.g., Ref. [24]))**.** *Any eigenvector of* $\boldsymbol{C}_L$ *associated with eigenvalue* $\mu \in \mathbb{C}$ *has the following Kronecker–Vandermonde form:*

$$\boldsymbol{v}_L(\mu) \otimes \boldsymbol{\varphi}, \tag{5.6}$$

where $\boldsymbol{v}_L(\mu) = [1, \mu, \ldots, \mu^{L-1}]^\top$, *and* $\boldsymbol{\varphi} \in \mathbb{C}^D \setminus \{\boldsymbol{0}\}$.

Proof. See Appendix D.2.

We now connect the block-companion solution $\boldsymbol{C}_L \in \mathbb{C}^{DL\times DL}$ of (5.5) to the truncated DMD propagator $\boldsymbol{A}_M \in \mathbb{C}^{M\times M}$ defined in (2.12).

Proposition 5.3. *The truncated DMD propagator can be recast as*

$$\boldsymbol{A}_M \;=\; \boldsymbol{U}_M^H\, \boldsymbol{C}_L\, \boldsymbol{U}_M, \tag{5.7}$$

where $\boldsymbol{U}_M$ *is defined in* (2.11).

Proof. See Appendix D.4.

Based on Proposition 5.3, we connect the eigenvectors of $\boldsymbol{C}_L$, which have a KV structure (Lemma 5.2), to the projected-DMD modes.

Theorem 5.4. *Let $(\widehat{\lambda}_j, \boldsymbol{w}_j)$ be an eigenpair of the reduced propagator $\boldsymbol{A}_M$ defined in* (2.12)*, and let $\widehat{\boldsymbol{\phi}}_j^p$ and $\widehat{\boldsymbol{\phi}}_j^e$ be the associated projected and exact DMD modes defined in* (2.14) *and* (2.16)*, respectively. Then, $(\widehat{\lambda}_j, \widehat{\boldsymbol{\phi}}_j^p)$ satisfies*

$$(\boldsymbol{C}_L - \widehat{\lambda}_j \boldsymbol{I})\, \widehat{\boldsymbol{\phi}}_j^p \;=\; \boldsymbol{r}_{\mathcal{U}_M}(\widehat{\boldsymbol{\phi}}_j^e), \tag{5.8}$$

where $\boldsymbol{r}_{\mathcal{U}_M}(\widehat{\boldsymbol{\phi}}_j^e)$ is the estimated-subspace residual vector (Definition 4.3).

Proof. See Appendix D.5. Theorem 5.4 shows that the estimated subspace residual vector $\boldsymbol{r}_{\mathcal{U}_M}(\widehat{\boldsymbol{\phi}}_j^e)$ governs (through the spectrum of $\boldsymbol{C}_L$) the extent to which the projected mode $\widehat{\boldsymbol{\phi}}_j^p$ conforms to the KV structure exhibited by the eigenvectors of the block-companion minimizer $\boldsymbol{C}_L$. In particular, $\boldsymbol{r}_{\mathcal{U}_M}(\widehat{\boldsymbol{\phi}}_j^e) = \boldsymbol{0}$ implies that $\widehat{\boldsymbol{\phi}}_j^p$ is an eigenvector of $\boldsymbol{C}_L$, and hence has exact Kronecker–Vandermonde structure. Specifically, $\boldsymbol{r}_{\mathcal{U}_M}(\widehat{\boldsymbol{\phi}}_j^e)$ is exactly the eigenrelation residual of the projected eigenpair $(\widehat{\lambda}_j, \widehat{\boldsymbol{\phi}}_j^p)$ with respect to $\boldsymbol{C}_L$.

Theorem 5.4 therefore provides a structural link that motivates KV-deviation metrics as a complementary criterion: the ESR norm $\left\|\boldsymbol{r}_{\mathcal{U}_M}(\widehat{\boldsymbol{\phi}}_j^e)\right\|_2$ directly measures the extent to which the projected DMD mode violates the companion eigenrelation. Since the eigenvectors of $\boldsymbol{C}_L$ have KV structure, smaller ESR suggests closer KV consistency, while larger ESR indicates a stronger departure from the ideal KV form. In Sec. 4, the ESR norm was used to separate true and spurious components; the link established here motivates KV deviation as an additional structural score for mode selection.

5.2. Nested DMD

Above, we showed that delay-embedded DMD induces KV structure in the computed modes. We postulate that true modes adhere more closely to the KV template than spurious modes. We therefore introduce direct KV-deviation scores computed from the mode entries and $\widehat{\lambda}_j$, and use them as structural features for separating true and spurious modes.

Here, we propose a different, direct quantification of the deviation of each mode from the KV structure, and then, use this deviation as a feature for separating true from spurious modes.

To motivate the proposed quantification, consider the *ideal KV template* for a projected DMD mode:

$$\widehat{\boldsymbol{\phi}}_j^p \;=\; \boldsymbol{v}_L(\widehat{\lambda}_j) \otimes \widehat{\boldsymbol{\phi}}_j^{(0)}, \qquad \widehat{\boldsymbol{\phi}}_j^{(0)} \in \mathbb{C}^D, \tag{5.9}$$

where $\boldsymbol{v}_L(\lambda) \coloneqq [1, \lambda, \ldots, \lambda^{L-1}]^\top$. Reshaping this mode, of length DL, as a $D \times L$ matrix by organizing blocks of D coordinates as L columns, yields

$$\left[\, \widehat{\boldsymbol{\phi}}_j^{(0)},\; \widehat{\lambda}_j\, \widehat{\boldsymbol{\phi}}_j^{(0)},\; \ldots,\; \widehat{\lambda}_j^{L-1} \widehat{\boldsymbol{\phi}}_j^{(0)} \,\right]. \tag{5.10}$$

Viewing this matrix as a sequence of columns, the ℓth column can be recast as

$$\widetilde{b}_j \, \widetilde{\boldsymbol{\phi}}_j \, \widehat{\lambda}_j^{\ell}, \tag{5.11}$$

where $\widetilde{b}_j \coloneqq \left\|\widehat{\boldsymbol{\phi}}_j^{(0)}\right\|_2$ and $\widetilde{\boldsymbol{\phi}}_j \coloneqq \widehat{\boldsymbol{\phi}}_j^{(0)} / \left\|\widehat{\boldsymbol{\phi}}_j^{(0)}\right\|_2$. This expression coincides with the one-mode specialization of (2.7), with the lag index ℓ replacing the time index k. Therefore, this reshaped form is consistent with the signal model *assumed* by an order-1 DMD.

Consequently, we reshape each computed projected mode $\widehat{\boldsymbol{\phi}}_j^{p} = [\,\widehat{\boldsymbol{\phi}}_j^{(0)}; \ldots ; \widehat{\boldsymbol{\phi}}_j^{(L-1)}\,]$ as above, obtaining the following mode matrix,

$$\widehat{\boldsymbol{\Phi}}_j \;\coloneqq\; \left[\, \widehat{\boldsymbol{\phi}}_j^{(0)}, \ldots, \widehat{\boldsymbol{\phi}}_j^{(L-1)} \,\right] \in \mathbb{C}^{D \times L}. \tag{5.12}$$

Then, we apply an order-1 DMD to the column sequence of $\widehat{\boldsymbol{\Phi}}_j$, along the lag axis, yielding a single DMD eigenvalue $\widetilde{\lambda}_j \in \mathbb{C}$ and a corresponding DMD mode $\widetilde{\boldsymbol{\phi}}_j^{(0)} \in \mathbb{C}^D$.

Following the standard DMD reconstruction convention (Sec. 2.2.2), we reconstruct the rank-1 mode matrix by:

$$\widehat{\boldsymbol{\Phi}}_j^{(\mathrm{KV})} \;\coloneqq\; \left(\widetilde{\boldsymbol{\phi}}_j^{(0)} \, \widetilde{\boldsymbol{\phi}}_j^{(0)H} \, \widehat{\boldsymbol{\phi}}_j^{(0)}\right) \boldsymbol{v}_L^{\top}(\widetilde{\lambda}_j), \tag{5.13}$$

where $\widehat{\boldsymbol{\phi}}_j^{(0)}$ is the first column of $\widehat{\boldsymbol{\Phi}}_j$.

Based on the reconstructed matrix $\widehat{\boldsymbol{\Phi}}_j^{(\mathrm{KV})}$, we define a single KV-deviation score as the mean squared reconstruction error:

$$\mathcal{R}_j^{(\mathrm{KV})} \;\coloneqq\; \frac{\left\|\widehat{\boldsymbol{\Phi}}_j - \widehat{\boldsymbol{\Phi}}_j^{(\mathrm{KV})}\right\|_F^2}{DL}. \tag{5.14}$$

The score $\mathcal{R}_j^{(\mathrm{KV})}$ is a direct KV-deviation metric: it vanishes when the reshaped mode follows an exact rank-1 KV template along the lag axis and increases as the mode departs from KV structure.

This is exactly the rank-1 spatiotemporal template *assumed* by an order-1 DMD model along the lag axis (treating lag index as the evolution variable).

From this point onward we follow Algorithm 4.1: we apply the same score scaling and then the binary mode selection framework (Algorithm 4.2). Here the per-mode feature is scalar and derived from $\mathcal{R}_j^{(\mathrm{KV})}$. Algorithm 5.1 summarizes the procedure.

The nested rank-1 construction replaces one delay-coordinates DMD fit of unknown order with M independent order-1 fits along the lag axis, one per mode.

The added cost scales linearly with D, L, and M and is negligible compared with the truncated SVD used to form the projected modes; see Appendix F for detailed asymptotic costs.

As a simpler variant, we consider a *fixed-eigenvalue KV fit* (FEKVF). Here "fixed eigenvalue" means that we do not re-estimate a lag-axis eigenvalue: we set the lag multiplier to the original DMD eigenvalue $\widehat{\lambda}_j$ and fit only the spatial coefficient in the KV template. For each mode this amounts to the least-squares problem

$$\min_{\boldsymbol{\varphi} \in \mathbb{C}^D} \left\|\widehat{\boldsymbol{\Phi}}_j - \boldsymbol{\varphi} \, \boldsymbol{v}_L(\widehat{\lambda}_j)^{\top}\right\|_F^2,$$

Require: Projected modes $\{\widehat{\boldsymbol{\phi}}_j^{p}\}_{j=1}^{M} \subset \mathbb{C}^{DL}$, embedding length L, spatial dimension D, small $\varepsilon > 0$; clustering routine and aggregation rule as in Algorithm 4.2
Ensure: Order estimate $\widehat{m}$ and labels $\{\ell_j\}_{j=1}^{M}$
1: **for** $j = 1, \ldots, M$ **do**
2: $\quad\widehat{\boldsymbol{\Phi}}_j \leftarrow \mathrm{reshape}(\widehat{\boldsymbol{\phi}}_j^{p},\, D \times L)$
3: $\quad$Run rank-1 DMD on $\widehat{\boldsymbol{\Phi}}_j$ to obtain $(\widetilde{\boldsymbol{\phi}}_j^{(0)}, \widetilde{\lambda}_j)$
4: $\quad\boldsymbol{y}_0 \leftarrow$ first column of $\widehat{\boldsymbol{\Phi}}_j$
5: $\quad\widehat{\boldsymbol{\Phi}}_j^{(\mathrm{KV})} \leftarrow \big(\widetilde{\boldsymbol{\phi}}_j^{(0)}\, \widetilde{\boldsymbol{\phi}}_j^{(0)H}\, \boldsymbol{y}_0\big)\, \boldsymbol{v}_L(\widetilde{\lambda}_j)^{\top}$
6: $\quad\mathcal{R}_j^{(\mathrm{KV})} \leftarrow \|\widehat{\boldsymbol{\Phi}}_j - \widehat{\boldsymbol{\Phi}}_j^{(\mathrm{KV})}\|_F^2/(DL)$
7: $\quad\zeta_j \leftarrow \log(\mathcal{R}_j^{(\mathrm{KV})} + \varepsilon)$
8: **end for**
9: Stack features $\boldsymbol{\zeta} \leftarrow [\,\zeta_1, \ldots, \zeta_M\,]$
10: Apply binary mode selection framework (Algorithm 4.2) to $\boldsymbol{\zeta}$ to obtain labels $\{\ell_j\}$ and $\widehat{m}$

Algorithm 5.1.: KV deviation via nested rank-1 DMD (binary mode selection framework)

yielding a single scalar deviation score in closed form (Appendix E). Compared with the nested rank-1 method, FEKVF removes the inner order-1 DMD step and is therefore computationally cheaper per mode.

5.3. Connection to existing order-detection methods

5.3.1. Spatiotemporal coupling ratio-based method

Both the approach of Bronstein *et al.* (Ref. [5]) and our approach use KV structure to form per-mode scores that are small when the structure is present. Bronstein *et al.*[5] do so via entrywise quotient checks across consecutive lag blocks, testing whether $(\widehat{\boldsymbol{\phi}}_j^{(\ell+1)})_d/(\widehat{\boldsymbol{\phi}}_j^{(\ell)})_d \approx \widehat{\lambda}_j$ and aggregating the results. In delay-coordinates, however, ratio tests can be numerically fragile: whenever $(\widehat{\boldsymbol{\phi}}_j^{(\ell)})_d$ is small, the quotient is ill-conditioned and can take arbitrarily large values, so the aggregate can be dominated by a few unstable coordinates. Our analysis suggests that KV adherence is a continuous property shared by true and spurious modes to different degrees, so quotient checks often behave like a near-binary test (pass or fail). In contrast, our KV-fit score measures a graded distance from the KV template through a global reconstruction error. We compare these criteria numerically in Sec. 6.

5.3.2. Mode-norm

A common order-detection method in standard DMD ranks modes by the Euclidean norm of the exact-DMD mode, $\|\widehat{\boldsymbol{\phi}}_j^{e}\|_2$, motivated by the intuition that true modes carry

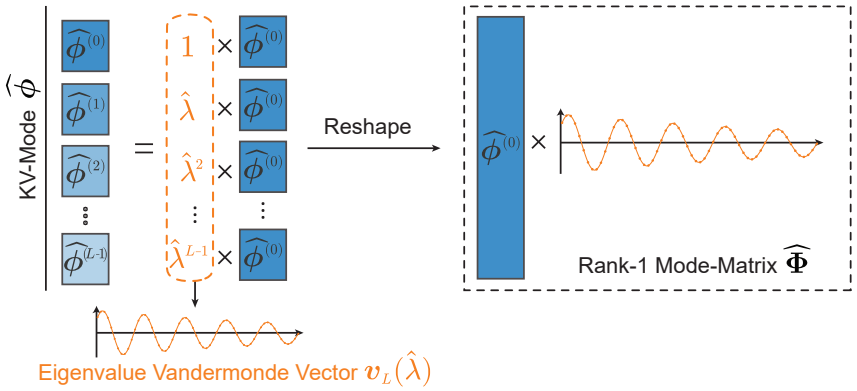

Figure 5.1.: Kronecker–Vandermonde structure of a DMD mode. A KV mode is formed by stacking L lag segments of length D, where each segment is a scaled and rotated copy of a base spatial vector $\widehat{\boldsymbol{\phi}}^{(0)}$. The multipliers $\widehat{\lambda}^{\ell}$ form the Vandermonde vector $\boldsymbol{v}_L(\widehat{\lambda}) = [1, \widehat{\lambda}, \ldots, \widehat{\lambda}^{L-1}]^\top$, illustrated at the bottom. Reshaping the KV mode yields the rank-1 outer product $\widehat{\boldsymbol{\phi}}^{(0)}\, \boldsymbol{v}_L(\widehat{\lambda})^\top$, mirroring the rank-1 structure underlying order-1 DMD.

larger energy than spurious modes. Our analysis in Sec. 4 sheds new light on this heuristic. In (4.5), the squared-bnorm of exact DMD modes $\|\widehat{\boldsymbol{\phi}}_j^e\|_2^2$ is expressed as the sum of the squared magnitude of the eigenvalue $|\widehat{\lambda}_j|^2$ and the ESR squared-norm $\|\boldsymbol{r}_{\mathcal{U}_M}(\widehat{\boldsymbol{\phi}}_j^e)\|_2^2$.

When using delay-coordinates, we empirically observe that the magnitude of the eigenvalue loses its discriminative power. Specifically, as the embedding length L grows, the bulk of spurious eigenvalue magnitudes concentrates in an annular band near the unit circle, as illustrated in Fig. 6.1 (left panel). This behavior is qualitatively consistent with annulus concentration phenomena studied for random polynomial roots; see Soundararajan [40] for a detailed analysis. We first consider the case where L is sufficiently large, so that spurious modes satisfy $|\widehat{\lambda}_j| \approx 1$ for $j \in \mathcal{J}_{\text{spur}}$.

Consider first a weakly damped signal, where the underlying eigenvalues satisfy $|\widehat{\lambda}_j| \approx 1$, for $j \in [m]$, where we assume that the associated DMD eigenvalues satisfy $|\widehat{\lambda}_j| \approx |\lambda_j|$ for $j \in \mathcal{J}_{\text{true}}$ to simplify the discussion.

In the weakly damped case with L sufficiently large, the eigenvalue magnitudes $|\widehat{\lambda}_j|^2$ are essentially the same for $j \in \mathcal{J}_{\text{true}}$ and $j \in \mathcal{J}_{\text{spur}}$, and thus carry little to no information about mode identity. By the decomposition of the exact DMD mode squared-norm in (4.5), the only component with discriminative power is the ESR squared-norm $|\boldsymbol{r}_{\mathcal{U}_M}(\widehat{\boldsymbol{\phi}}_j^e)|_2^2$. Using the mode norm $\|\widehat{\boldsymbol{\phi}}_j^e\|_2^2$ therefore confounds this informative ESR norm with the non-informative eigenvalue magnitude $|\widehat{\lambda}_j|^2$. Since the eigenvalue magnitudes may vary independently of mode identity, they mask or dilute the separation between true and spurious modes. For this reason, it is preferable to work directly with the ESR

norm, which isolates the meaningful geometric information without contamination from eigenvalue scaling.

Next, consider a strongly damped signal, where the underlying eigenvalues satisfy $|\lambda_j| < 1$ for $j \in [m]$. When L is sufficiently large, spurious modes typically have $|\widehat{\lambda}_j| \approx 1$, while true DMD modes have smaller $|\widehat{\lambda}_j|$. In addition, true modes typically have smaller ESR energies than spurious modes. Together, these two facts imply that true modes tend to have *smaller* exact-mode norms than spurious modes. This is the opposite of how the heuristic is commonly applied, which treats larger norms as more likely to be true.

For completeness, we note that for small-L embeddings (and in the standard DMD case $L = 1$), spurious eigenvalues are often observed well inside the unit disk. In this regime, true modes typically have significantly larger $|\widehat{\lambda}_j|^2$ than spurious modes, so mode-norm ranking *is* an effective separator. However, the ESR norm contributes in the opposite direction, being larger for spurious modes, and therefore it weakens the separation provided by $|\widehat{\lambda}_j|^2$ alone. The numerical evaluation in Sec. 6 compares mode-norm ranking with a detector based purely on $|\widehat{\lambda}_j|$ for $L = 1$.

6. Numerical Experiments

This chapter presents results of numerical experiments. We first evaluate the performance of our proposed order-detection methods and compare them with existing baselines. We then report experiments that illustrate the spurious-eigenvalue magnitude behavior discussed in Sec. 5.3.2.

6.1. Experimental setup

We simulate signals according to eq. (1) using the linear m-mode signal described in (7), with additive Gaussian noise. (We note that we observed qualitatively similar performance under other noise distributions, including bi-Gaussian, heavy-tailed Student's t, uniform, and heteroscedastic noise, but report only Gaussian results here for brevity.) The ground-truth spatial modes $\boldsymbol{\phi}_j$ are drawn independently and normalized to unit norm.

Across experiments we control the following generation parameters: the SNR, the number of true modes m, the eigenvalue magnitudes (we use a common value by setting $\rho_j = \rho$ for all j so that damping is controlled by the single scalar ρ), the minimal phase gap $\Delta\theta$, and the amplitude heterogeneity κ_b. The minimal phase gap is defined as

$$\Delta\theta = \min_{j \neq k} \min\big(|\theta_j - \theta_k|,\, 2\pi - |\theta_j - \theta_k|\big). \tag{6.1}$$

Amplitude heterogeneity is quantified by

$$\kappa_b = \frac{\max_j |b_j|}{\min_j |b_j|}. \tag{6.2}$$

6.2. Order Detection Evaluations

We compare our methods (ESR-energy, KV deviation via nested rank-1 DMD, and fixed-eigenvalue KV fit) with the baselines BIC, Spectral Gap (GAP), and Spatiotemporal Consistency (STC) from Chapter 3. Performance is measured by the order-hit probability, defined as the fraction of trials in which the estimated order equals the actual order m, across 500 trials.

In each experiment, we vary a single parameter (SNR, spectral separation, damping, amplitude heterogeneity) while keeping all other quantities at fixed working-point values.

Fig. 6.2, shows the order-hit probability as a function of SNR under Gaussian noise. Our methods outperform all three baselines across the entire test range. As SNR increases, all methods exhibit a transition region where hit probability rises from near

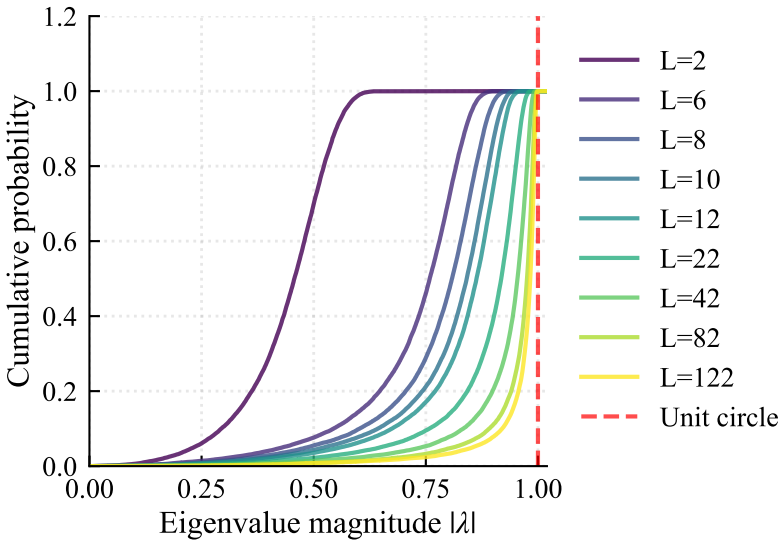

Figure 6.1.: Spurious eigenvalue magnitude statistics versus embedding length L. For each L, we pool the magnitudes $|\widehat{\lambda}|$ of all spurious eigenvalues from all Monte–Carlo trials and plot the resulting empirical CDF (one curve per L). Overall, spurious eigenvalue magnitudes shift upward with increasing L, but the pooled distributions retain a non-negligible lower tail.

zero to near one. The KV-deviation methods reach this transition at the lowest SNR, followed closely by ESR-energy. BIC shows a similar pattern but with a later transition. STC and GAP increase more gradually and do not approach perfect accuracy. Increasing the number of modes, m, shifts all transitions toward higher SNR values with the same ordering among methods.

Fig. 6.3 presents the order-hit probability as a function of the minimal phase separation $\Delta\theta$ under Gaussian noise. Across the tested range, our methods consistently achieve higher hit probabilities than all three baselines. As $\Delta\theta$ increases, all methods move from near-zero to near-one hit probability, with the KV-deviation criteria transitioning at the smallest gaps and ESR-energy close behind. BIC follows the same pattern but becomes reliable only at larger separations, GAP requires substantially wider gaps before saturating, and STC increases more slowly and reaches only moderate accuracy. Again, increasing m shifts the transition regions toward larger phase gaps while leaving the relative ordering of methods unchanged.

Fig. 6.4 shows the order-hit probability as a function of the common magnitude ρ under Gaussian noise. Throughout the tested range, our methods consistently outperform the baselines, achieving higher hit probabilities for all values of ρ. As ρ increases toward one, all methods improve, but the same performance hierarchy persists: the KV-deviation scores become reliable at the smallest damping, ESR-energy follows closely, and BIC only catches up at larger values of ρ. GAP eventually attains high accuracy but only when ρ is very close to one. In contrast to the phase-separation experiment, STC does

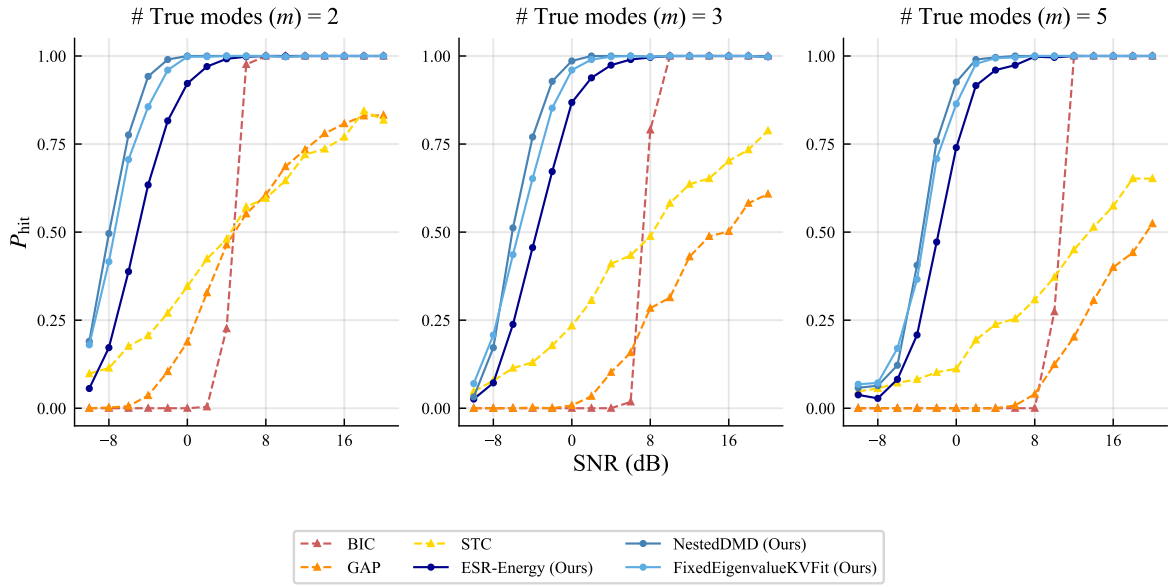

Figure 6.2.: Order-hit probability vs. SNR under Gaussian noise. Panels show $m = 2$ (left), $m = 3$ (center), and $m = 5$ (right). Working point: $N = 200$, $D = 45$, $L = 64$, $M = 15$; eigenvalues satisfy $\rho_j = 0.98$ with equally spaced phases and minimal separation $\Delta\theta = 0.01$; all amplitudes equal to one. Only the SNR varies.

not saturate within the tested range and instead increases slowly and approximately monotonically. As before, increasing m shifts the transition regions toward larger ρ while leaving the ordering among methods unchanged.

Figure 6.5 shows the order-hit probability as a function of amplitude heterogeneity κ_b under Gaussian noise. Across the tested range, our methods consistently outperform the baselines, maintaining substantially higher hit probabilities as the amplitudes become more uneven. As κ_b increases, the baseline methods (STC, BIC, GAP) deteriorate sharply and their hit probabilities collapse over most of the range. In contrast, the KV-deviation methods and ESR-energy remain stable and accurate, with only a modest decline in the most heterogeneous regime and only for $m = 5$.

To summarize the one-dimensional sweeps in Figs. 6.2–6.5, Table 6.1 reports normalized AUC values aggregated over each sweep.

Figure 6.6 shows the order-hit probability as a function of the number of true modes m. Our methods remain detailed and robust until a point of collapse determined by the signal parameters (dimensions, SNR, etc.), here observed at $m = 15$. In contrast, the performance of the baseline methods collapses much earlier. BIC is particularly sensitive to the model order, failing almost completely for $m > 3$. STC and GAP also degrade significantly, while ESR-energy and the KV-deviation methods remains robust for a wider range of m.

Fig. 6.7 presents the hit probability in the absence of delay-coordinates, i.e., $L = 1$. To satisfy the inequality $DL < N - L$, we set $D = 220$ while keeping the other parameters

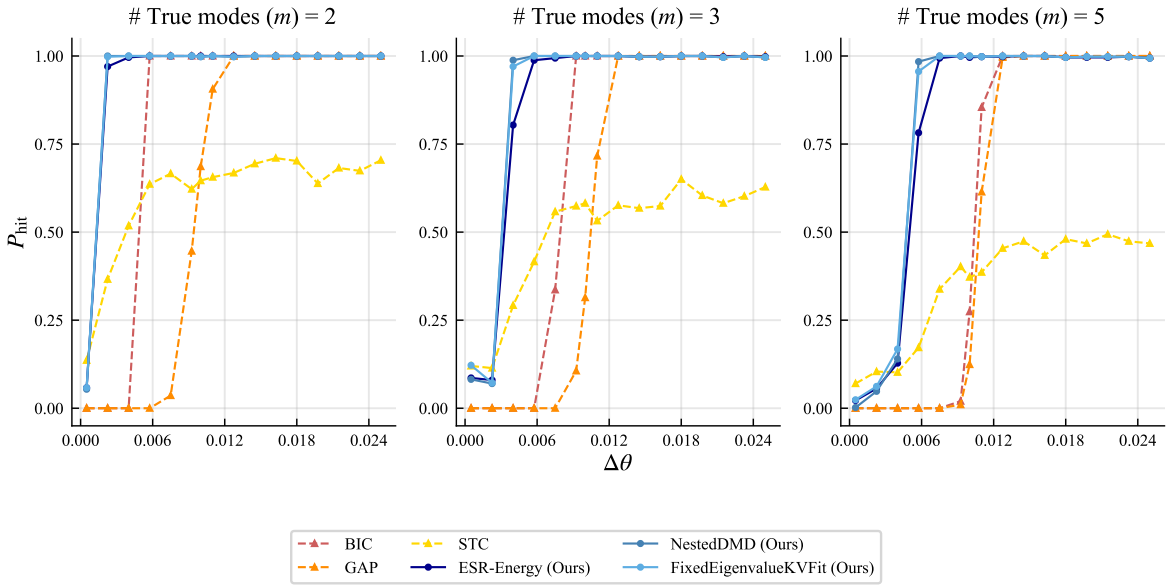

Figure 6.3.: Order-hit probability vs. phase separation $\Delta\theta$ (defined in (6.1)) under Gaussian noise. Panels show $m = 2$ (left), $m = 3$ (center), and $m = 5$ (right). Working point matches Fig. 6.2, with SNR fixed at 10 dB; only $\Delta\theta$ varies.

fixed. We compare the performance of our ESR-energy to BIC and GAP, and also include two $L = 1$ baselines that use eigenvalue magnitude and exact-mode norm as the per-mode scores along with the binary mode selection framework (Algorithm 4.2).

Even without delay-coordinates, ESR-energy remains interpretable and reliable across all tested ranges of SNR, $\Delta\theta$, and damping parameter ρ. In this setting, eigenvalue magnitude and mode-norm baselines behave similarly but become reliable sooner, with eigenvalue magnitude consistently outperforming mode norm. This supports the analysis in Sec. 5.3.2, and also shows that ESR-energy retains practical value when delay-coordinates are infeasible or undesirable.

Table 6.2 reports the corresponding normalized AUC values for the no-delay setting.

6.3. Spurious eigenvalue behavior under delay-coordinates

To test the effect of the embedding length L on spurious eigenvalues, we fix $m = 5$ and $M = 15$ and run 500 trials. From the ten spurious eigenvalues in each trial we record

$$\rho_{\min}(L) \;=\; \min_{j \notin \mathcal{J}_{\text{true}}} |\widehat{\lambda}_j|.$$

Figure 6.1 shows that spurious eigenvalue magnitudes shift upward with L and increasingly concentrate near the unit circle. At the same time, a non-negligible lower tail

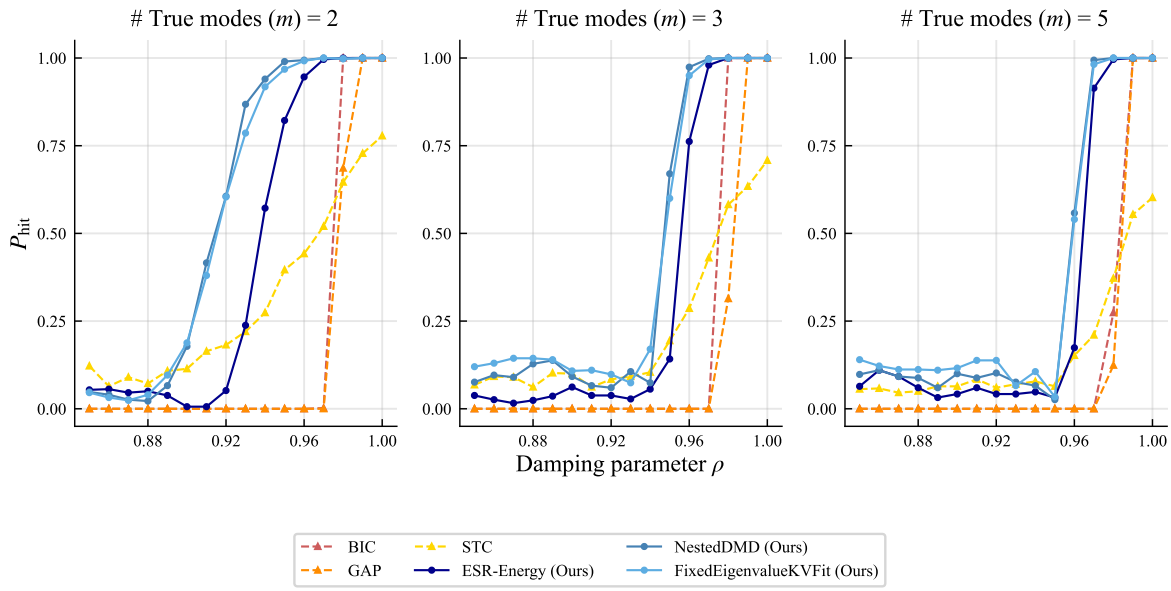

Figure 6.4.: Order-hit probability vs. damping factor ρ under Gaussian noise. Panels show $m = 2$ (left), $m = 3$ (center), and $m = 5$ (right). Working point matches Figs. 6.2–6.3; only ρ varies.

persists for all tested L, indicating substantial variability across trials even when the typical magnitudes are large.

This behavior supports the empirical observation highlighted in Sec. 5.3.2: as L increases, the bulk of spurious eigenvalue magnitudes shifts upward toward the unit circle, while a substantial lower tail persists. Consequently, even at large L, eigenvalue magnitude alone cannot reliably separate true and spurious components.

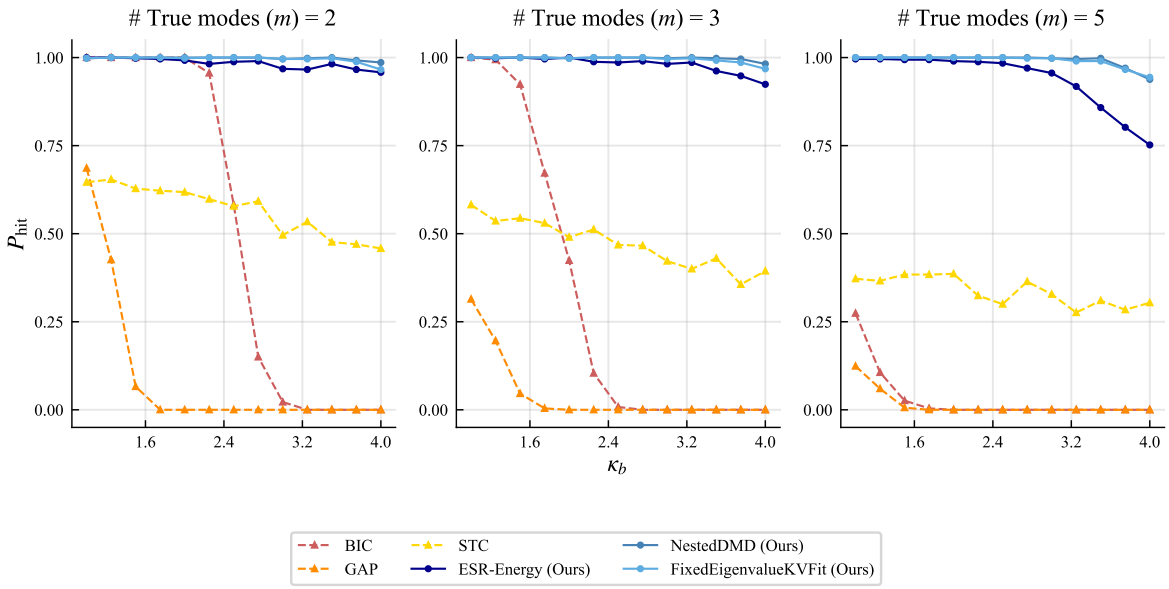

Figure 6.5.: Order-hit probability vs. amplitude heterogeneity κ_b (defined in (6.2)) under Gaussian noise. Panels show $m = 2$ (left), $m = 3$ (center), and $m = 5$ (right). As κ_b increases, the baseline methods (STC, BIC, GAP) degrade, while the KV-deviation methods and ESR-energy remain accurate across most of the tested range. Working point matches Figs. 6.2–6.4.

	$m = 2$					$m = 3$					$m = 5$				
Method	SNR	$\Delta\theta$	r	κ_b	M	SNR	$\Delta\theta$	r	κ_b	M	SNR	$\Delta\theta$	r	κ_b	M
BIC	0.497	0.811	0.154	0.507	**0.994**	0.400	0.690	0.149	0.290	**0.992**	0.320	0.594	0.122	0.025	0.265
GAP	0.424	0.629	0.137	0.064	0.680	0.212	0.590	0.120	0.034	0.317	0.122	0.574	0.111	0.013	0.107
STC	0.477	0.610	0.287	0.559	0.251	0.392	0.494	0.208	0.461	0.212	0.282	0.356	0.145	0.332	0.153
ESR-Energy	0.813	0.955	0.413	0.973	0.646	0.764	0.876	0.303	0.971	0.639	0.707	0.807	0.264	0.930	0.605
NestedDMD	**0.913**	**0.962**	**0.575**	**0.994**	0.949	**0.849**	**0.894**	0.397	**0.992**	0.931	**0.774**	0.826	0.320	**0.985**	0.868
FixedEigenvalueKVFit	0.895	**0.962**	0.567	0.992	0.966	0.833	**0.894**	**0.410**	0.990	0.947	0.767	**0.828**	**0.336**	0.984	**0.884**

Table 6.1.: Normalized AUC for DC experiments (delay-embedded) across one-dimensional parameter sweeps. Columns grouped by number of true modes (m) contain AUC values for: SNR, phase separation ($\Delta\theta$), damping coefficient (r), amplitude ratio ($\kappa_b = b_{\max}/b_{\min}$), and order-overestimation (M).

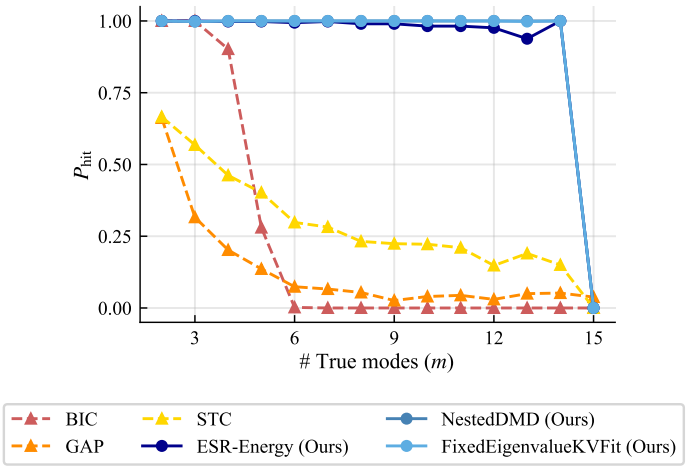

Figure 6.6.: Order-hit probability vs. number of true modes m under Gaussian noise. Working point as in Fig. 6.2, except $M = 3m$. Our methods remain effective as the model complexity grows, while baselines deteriorate rapidly.

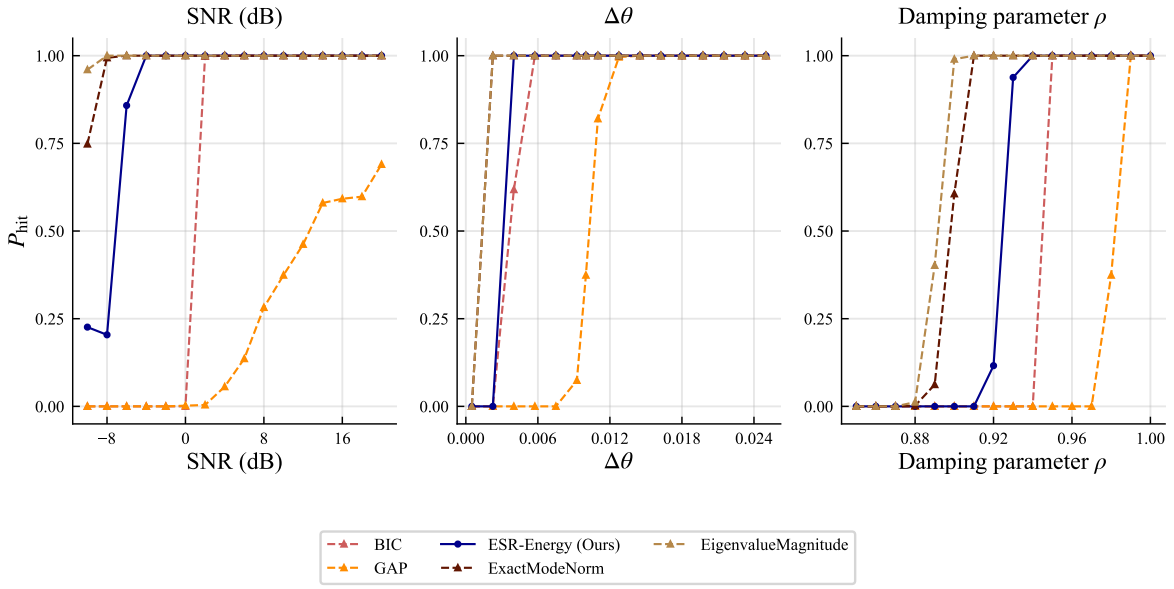

Figure 6.7.: Order detection performance without delay-coordinates ($L = 1$). Representative parameter variations under Gaussian noise for $m = 3$ over SNR (left), phase separation $\Delta\theta$ (middle), and damping ρ (right). For SNR, ESR-energy becomes reliable at lower levels than BIC, while mode-norm and eigenvalue-magnitude detectors become reliable sooner still, with eigenvalue magnitude performing best. For $\Delta\theta$, the methods remain close, with ESR-energy only slightly earlier at the transition, and again mode norm and eigenvalue magnitude shifting the transition earlier (eigenvalue magnitude best). GAP is unreliable across the tested ranges. Working point matches Figs. 6.2–6.5, except that $L = 1$ and $D = 220$.

	$m = 2$				$m = 3$				$m = 5$			
Method	SNR	$\Delta\theta$	r	κ_b	SNR	$\Delta\theta$	r	κ_b	SNR	$\Delta\theta$	r	κ_b
BIC	0.759	0.961	0.910	0.995	0.619	0.856	0.356	0.987	0.476	0.735	0.214	0.819
GAP	0.453	0.627	0.132	0.069	0.222	0.592	0.119	0.033	0.135	0.575	0.113	0.015
ESR-Energy	0.988	0.963	0.996	0.998	0.899	0.886	0.498	0.991	0.793	0.812	0.291	0.987
ExactModeNorm	**1.000**	0.996	**1.000**	**1.000**	0.984	**0.961**	0.677	**0.996**	0.867	**0.885**	0.363	**0.990**
EigenvalueMagnitude	**1.000**	**1.000**	**1.000**	**1.000**	**0.992**	**0.961**	**0.726**	**0.996**	**0.894**	**0.885**	**0.378**	**0.990**

Table 6.2.: Normalized AUC for No-DC experiments ($L = 1$, no delays) across one-dimensional parameter sweeps. Columns grouped by number of true modes (m) contain AUC values for: SNR, phase separation ($\Delta\theta$), damping coefficient (r), and amplitude ratio ($\kappa_b = b_{\max}/b_{\min}$).

7. Conclusions and Future Work

Our results indicate that, in delay-coordinates DMD, the most informative signals for mode selection are carried by the induced *mode organization* and mode geometry providing the most informative selection signals. Consistent with this viewpoint, the KV-deviation criteria, which explicitly test delay-induced structure, are the most reliable across the tested regimes and outperform both classical baselines and the ESR-based selector. Moreover, the spurious spectrum is not invariant under delay-coordinates: spurious eigenvalue magnitudes grow systematically with the embedding length L. This behavior helps explain why the mode-norm method, which is often effective in standard ($L = 1$) DMD, can lose discriminative power in delay-coordinates.

A practical consideration is sensitivity to the design parameters L and M. The embedding length L should be selected for modeling and spectral estimation, not for detection. Empirically, our selectors are effective from modest L values and do not deteriorate as L grows. The truncation rank M is best interpreted through the overestimation gap $M - m$: while M must exceed m to allow post hoc mode selection, overly large M shrinks residual energy for all modes and can reduce the discriminative power of ESR. This sensitivity suggests that the *full* per-mode energy allocation across singular directions may contain more stable information than an aggregated residual summary. In contrast, the KV-deviation criteria largely mitigate the dependence on M, remaining reliable even under aggressive overestimation.

Future Work

Several directions appear promising. First, it would be valuable to characterize how true and spurious modes distribute energy across singular directions and to develop selection rules that exploit distributional differences directly. Second, we observe that spurious spectral statistics vary systematically with the embedding length L; explaining the mechanism behind this dependence would deepen our understanding of how delay-coordinates reshape the spurious spectrum. Finally, although ESR and KV deviation are linked conceptually through the companion eigenrelation, a systematic characterization of how residual geometry relates to KV departure remains open.

Appendices

A. Relations between D, L and N

This appendix revisits several identities from the main text in the *wide* delay-embedded setting, where the snapshot matrix $\boldsymbol{X}_0 \in \mathbb{C}^{DL\times(N-L)}$ satisfies

$$DL < N - L. \tag{.1}$$

In this regime, $\boldsymbol{X}_0$ has more columns than rows. Throughout we assume that $\boldsymbol{X}_0$ has full row rank,

$$\operatorname{rank}(\boldsymbol{X}_0) = DL, \tag{.2}$$

which is a generic outcome under sufficiently rich excitation or additive noise.

A.1. Least-squares propagator and the projector onto $\operatorname{col}(\boldsymbol{X}_0)$

Full row rank implies that $\boldsymbol{X}_0\boldsymbol{X}_0^*$ is invertible and that the Moore–Penrose pseudoinverse admits the form

$$\boldsymbol{X}_0^\dagger \;=\; \boldsymbol{X}_0^{\mathrm{H}}\,(\boldsymbol{X}_0\boldsymbol{X}_0^{\mathrm{H}})^{-1}. \tag{.3}$$

Let $\mathcal{U} = \operatorname{col}(\boldsymbol{X}_0)$ and denote by $\boldsymbol{P}_{\mathcal{U}}$ the orthogonal projector onto $\mathcal{U}$. Since $\operatorname{rank}(\boldsymbol{X}_0) = DL$, we have $\mathcal{U} = \mathbb{C}^{DL}$ and therefore

$$\boldsymbol{P}_{\mathcal{U}} = \boldsymbol{I}_{DL}. \tag{.4}$$

Consequently, the least-squares propagator

$$\boldsymbol{A}_{\mathrm{MP}} \;=\; \boldsymbol{X}_1\boldsymbol{X}_0^\dagger \tag{.5}$$

is uniquely defined.

A.2. Implications for residual-based diagnostics

In the wide full-row-rank regime, $\mathcal{U} = \mathbb{C}^{DL}$ eliminates any component outside $\mathcal{U}$. In particular, for an exact DMD mode $\widehat{\boldsymbol{\phi}}^{\,e}$ the decomposition

$$\widehat{\boldsymbol{\phi}}^{\,e} \;=\; \boldsymbol{P}_{\mathcal{U}_M}\widehat{\boldsymbol{\phi}}^{\,e} + (\boldsymbol{P}_{\mathcal{U}} - \boldsymbol{P}_{\mathcal{U}_M})\widehat{\boldsymbol{\phi}}^{\,e} + (\boldsymbol{I} - \boldsymbol{P}_{\mathcal{U}})\widehat{\boldsymbol{\phi}}^{\,e} \tag{.6}$$

reduces to

$$\widehat{\boldsymbol{\phi}}^{\,e} \;=\; \boldsymbol{P}_{\mathcal{U}_M}\widehat{\boldsymbol{\phi}}^{\,e} + (\boldsymbol{I} - \boldsymbol{P}_{\mathcal{U}_M})\widehat{\boldsymbol{\phi}}^{\,e}, \tag{.7}$$

because $(\boldsymbol{I} - \boldsymbol{P}_{\mathcal{U}})\widehat{\boldsymbol{\phi}}^{e} = \boldsymbol{0}$. Thus, residual-type scores relative to $\mathcal{U}_M$ quantify only the truncation effect $(\boldsymbol{I} - \boldsymbol{P}_{\mathcal{U}_M})\widehat{\boldsymbol{\phi}}^{e}$, which shrinks monotonically as M increases toward $\min(DL, N - L) = DL$.

Accordingly, the wide full-row-rank regime is an extreme case in which $\mathcal{U} = \mathrm{col}(\boldsymbol{X}_0) = \mathbb{C}^{DL}$ and thus $\boldsymbol{P}_{\mathcal{U}} = \boldsymbol{I}_{DL}$. In this setting, residual-type scores relative to $\mathcal{U}_M$ reflect only the truncation term $(\boldsymbol{I} - \boldsymbol{P}_{\mathcal{U}_M})\widehat{\boldsymbol{\phi}}^{e}$, which shrinks as $M \uparrow DL$. We therefore focus on the more common tall regime $N - L < DL$, which is typical in high-dimensional measurements (and even more so under delay coordinates) and is also more challenging from an estimation standpoint since fewer snapshots are available relative to the number of unknown parameters.

A.3. Block-companion structure

In the delay-embedded construction, the shift relations in the first $L - 1$ block rows hold exactly by definition. Consequently, the block-companion matrix $\boldsymbol{C}_L$ is feasible for the constrained least-squares problem that enforces these shift equations. In the wide full-row-rank regime, the identity $\boldsymbol{A}_{\mathrm{MP}} = \boldsymbol{C}_L \boldsymbol{P}_{\mathcal{U}}$ simplifies (since $\boldsymbol{P}_{\mathcal{U}} = \boldsymbol{I}_{DL}$) to

$$\boldsymbol{A}_{\mathrm{MP}} = \boldsymbol{C}_L. \tag{.8}$$

Hence the Moore–Penrose solution already acts as an exact shift on all DL coordinates, and no additional projection through $\boldsymbol{P}_{\mathcal{U}}$ takes place.

B. Bound on signal–subspace deviation under noise

This appendix specializes classical subspace perturbation theory to the delay–embedded trajectory matrix and quantifies the deviation between the true signal subspace

$$\mathcal{S} = \mathrm{col}(\boldsymbol{S}_0)$$

and its empirical estimate $\mathrm{col}(\boldsymbol{U}_m)$, where $\boldsymbol{U}_m$ contains the leading m left singular vectors of

$$\boldsymbol{X}_0 = \boldsymbol{S}_0 + \boldsymbol{E}.$$

Model and notation

The clean signal follows the exponential model

$$\boldsymbol{s}_k = \sum_{j=1}^{m} b_j \, \boldsymbol{\phi}_j \, \lambda_j^k, \qquad \lambda_j = \rho_j \, \mathrm{e}^{\mathrm{i}\theta_j},$$

with spatial vectors $\boldsymbol{\phi}_j \in \mathbb{C}^D$. We assume a bounded-radius regime $0 < \rho_{\min} \leq \rho_j \leq \rho_{\max} \leq 1$.

Let N denote the number of snapshots $\{\boldsymbol{s}_k\}_{k=0}^{N-1}$. After delay embedding with L delays, the number of usable delay vectors is

$$N_{\text{eff}} := N - L + 1,$$

and the standard DC–DMD snapshot matrices satisfy

$$\boldsymbol{X}_0, \boldsymbol{X}_1 \in \mathbb{C}^{DL\times(N-L)}.$$

In particular, the noiseless part of $\boldsymbol{X}_0$ admits the factorization

$$\boldsymbol{S}_0 = \widetilde{\boldsymbol{\Phi}}_L \, \boldsymbol{B} \, \boldsymbol{V}_{N-L}^{\top},$$

where the lifted spatial factor is

$$\widetilde{\boldsymbol{\Phi}}_L := \boldsymbol{V}_L \odot \boldsymbol{\Phi} = \big[\boldsymbol{v}_L(\lambda_1) \otimes \boldsymbol{\phi}_1 \;\; \cdots \;\; \boldsymbol{v}_L(\lambda_m) \otimes \boldsymbol{\phi}_m\big] \in \mathbb{C}^{DL\times m},$$

with base spatial factor $\boldsymbol{\Phi} = [\boldsymbol{\phi}_1 \cdots \boldsymbol{\phi}_m]$. The amplitudes are given by $\boldsymbol{B} = \operatorname{diag}(b_1, \ldots, b_m)$, and the Vandermonde matrix is defined generally as

$$\boldsymbol{V}_n = \big[\boldsymbol{v}_n(\lambda_1) \;\; \cdots \;\; \boldsymbol{v}_n(\lambda_m)\big] \in \mathbb{C}^{n\times m}, \qquad \boldsymbol{v}_n(\lambda) = [\,1, \lambda, \ldots, \lambda^{n-1}\,]^{\top}.$$

Let $\theta_{\max}$ denote the largest principal angle between $\operatorname{col}(\boldsymbol{U}_m)$ and $\mathcal{S}$.

Assumptions

We assume:

(A1) *Phase separation*:

$$\Delta_\theta := \min_{j\neq j'} \min\big(|\theta_j - \theta_{j'}|,\, 2\pi - |\theta_j - \theta_{j'}|\big) > 0.$$

(A2) *Nondegeneracy*: $\sigma_m(\boldsymbol{\Phi}) > 0$ and $\min_j |b_j| > 0$.

(A3) L and $N - L$ exceed a constant multiple of $1/\Delta_\theta$.

Signal–subspace deviation

Lemma .1 (Signal–subspace deviation)**.** *If $\|\boldsymbol{E}\|_2$ is sufficiently small, then*

$$\|\boldsymbol{P}_{\operatorname{col}(\boldsymbol{U}_m)} - \boldsymbol{P}_{\mathcal{S}}\|_2 = \sin\theta_{\max} \le \eta,$$

where

$$\eta = \frac{\|\boldsymbol{E}\|_2}{\sigma_m(\boldsymbol{\Phi})\, \min_j |b_j|\, \alpha_L\, \alpha_{N-L} - \|\boldsymbol{E}\|_2}.$$

Here α_n denotes any lower bound on $\sigma_m(\boldsymbol{V}_n)$ valid under assumptions (A1)–(A3).

Sketch. Wedin's $\sin\Theta$ theorem gives

$$\sin\theta_{\max} \le \frac{\|\boldsymbol{E}\|_2}{\sigma_m(\boldsymbol{S}_0) - \|\boldsymbol{E}\|_2}.$$

Using the factorization of $\boldsymbol{S}_0$,

$$\sigma_m(\boldsymbol{S}_0) \ge \sigma_m(\widetilde{\boldsymbol{\Phi}}_L)\, \min_j |b_j|\, \sigma_m(\boldsymbol{V}_{N-L}).$$

Since $\widetilde{\boldsymbol{\Phi}}_L = \boldsymbol{V}_L \odot \boldsymbol{\Phi}$,

$$\sigma_m(\widetilde{\boldsymbol{\Phi}}_L) \ge \sigma_m(\boldsymbol{\Phi})\, \sigma_m(\boldsymbol{V}_L).$$

Combining the inequalities and using $\sigma_m(\boldsymbol{V}_n) \ge \alpha_n$ yields the result. □

Vandermonde conditioning

Under (A1), classical results on Vandermonde conditioning imply that

$$\sigma_m(\boldsymbol{V}_n) \ge \alpha_n,$$

for some $\alpha_n > 0$ depending on the minimal phase gap Δ_θ, the order m, and the bounded-radius regime $[\rho_{\min}, \rho_{\max}]$. Smaller phase gaps worsen the conditioning and reduce α_n. When $\rho_{\max} = 1$ (unit-modulus nodes), one may take a bound of the form

$$\alpha_n \ge c_{\mathrm{sep}}(\Delta_\theta, m)\sqrt{n},$$

whereas for $\rho_{\max} < 1$ the dependence on n need not exhibit the same $\sqrt{n}$ growth and may level off, with constants controlled additionally by $[\rho_{\min}, \rho_{\max}]$. See, e.g., [7, 21].

Discussion

The deviation $\sin\theta_{\max}$ scales linearly with $\|\boldsymbol{E}\|_2$ and is controlled by the conditioning of the lifted signal factors. The dependence on L and $N - L$ enters through the Vandermonde conditioning constants α_L and α_{N-L}; when $\rho_{\max} = 1$ this recovers the familiar $\sqrt{L\,(N-L)}$ scaling, while for $\rho_{\max} < 1$ the improvement with n may saturate. The result provides a quantitative justification for Assumption 2.1 in the main text.

C. Delay-coordinates yield Kronecker–Vandermonde spatiotemporal modes

This appendix gives a short data-side derivation of the structural result identified in [5]: a signal that is a sum of spatially modulated exponentials in time becomes, after delay-coordinates, a sum of spatiotemporally coupled KV templates, each still modulating a temporal exponential across the embedded columns. The argument requires no operator assumptions and applies only to the true components, providing a concise complement to the operator-side derivation in Chapter 5, which extends the KV characterization to *all* modes (true and spurious).

Signal model

We adopt the exponential signal model from (2.7). For an embedding length L, collect L successive samples into

$$\widetilde{\boldsymbol{s}}_k = \begin{bmatrix} \boldsymbol{s}_k \\ \boldsymbol{s}_{k+1} \\ \vdots \\ \boldsymbol{s}_{k+L-1} \end{bmatrix} \in \mathbb{C}^{DL}.$$

Embedding of a single exponential component

For the jth component $b_j\,\boldsymbol{\phi}_j\,\lambda_j^k$,

$$\widetilde{\boldsymbol{s}}_k^{(j)} = b_j \begin{bmatrix} \boldsymbol{\phi}_j\,\lambda_j^k \\ \boldsymbol{\phi}_j\,\lambda_j^{k+1} \\ \vdots \\ \boldsymbol{\phi}_j\,\lambda_j^{k+L-1} \end{bmatrix} = b_j\,\big(\boldsymbol{\psi}_L(\lambda_j) \otimes \boldsymbol{\phi}_j\big)\,\lambda_j^k,$$

where $\boldsymbol{\psi}_L(\lambda_j) = [1, \lambda_j, \ldots, \lambda_j^{L-1}]^\top$. Delay-coordinates therefore fuse the spatial vector $\boldsymbol{\phi}_j$ and temporal exponential λ_j^k into the KV template $\boldsymbol{\psi}_L(\lambda_j) \otimes \boldsymbol{\phi}_j$.

KV decomposition of the embedded trajectory matrix

Let $N_{\text{eff}} = T - L + 1$ be the number of embedded columns. The trajectory matrix becomes

$$\widetilde{\boldsymbol{S}} = \sum_{j=1}^{m} b_j\,\big(\boldsymbol{\psi}_L(\lambda_j) \otimes \boldsymbol{\phi}_j\big)\big[1, \lambda_j, \ldots, \lambda_j^{N_{\text{eff}}-1}\big].$$

Writing $\boldsymbol{\psi}_{N_{\text{eff}}}(\lambda_j)$ for the temporal Vandermonde vector, this is

$$\widetilde{\boldsymbol{S}} = \sum_{j=1}^{m} b_j\,\big(\boldsymbol{\psi}_L(\lambda_j) \otimes \boldsymbol{\phi}_j\big)\boldsymbol{\psi}_{N_{\text{eff}}}(\lambda_j)^\top.$$

Conclusion

Under delay-coordinates, every true component becomes a rank-1 KV term whose spatial and temporal factors are inseparably coupled. These data-side KV modes coincide with the ideal KV templates assumed in Bronstein *et al.*[5]. The operator-side analysis of Chapter 5 generalizes this perspective by showing that *all* projected-DMD modes, including spurious ones, satisfy an approximate KV relation whose deviation is measured by their estimated-subspace residual.

D. Block-Companion Structure and Least-Squares Propagators

We consider the least-squares problem

$$\min_{\boldsymbol{A}\in\mathbb{C}^{DL\times DL}} \|\boldsymbol{A}\boldsymbol{X}_0 - \boldsymbol{X}_1\|_F^2,$$

with delay-embedded data matrices $\boldsymbol{X}_0, \boldsymbol{X}_1 \in \mathbb{C}^{DL\times T_{\text{eff}}}$. Among all minimizers we single out the Moore–Penrose (minimum-norm) minimizer

$$\boldsymbol{A}_{\text{MP}} := \boldsymbol{X}_1\boldsymbol{X}_0^{\dagger},$$

which will be related to the reduced propagator via an orthogonal projection.

D.1. Existence of a block-companion minimizer

Partition $\boldsymbol{A}$ into $D \times D$ block rows

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}^{(0)} \\ \vdots \\ \boldsymbol{A}^{(L-1)} \end{bmatrix}, \qquad \boldsymbol{A}^{(\ell)} \in \mathbb{C}^{D\times DL}.$$

The objective splits as

$$\|\boldsymbol{A}\boldsymbol{X}_0 - \boldsymbol{X}_1\|_F^2 = \sum_{\ell=0}^{L-1} \|\boldsymbol{A}^{(\ell)}\boldsymbol{X}_0 - \boldsymbol{X}_1^{(\ell)}\|_F^2.$$

In DC–DMD,

$$\boldsymbol{X}_1^{(\ell)} = \boldsymbol{S}_L^{(\ell)}\boldsymbol{X}_0, \qquad \ell = 0, \dots, L-2,$$

for the block-shift matrix $\boldsymbol{S}_L$. Thus setting $\boldsymbol{A}^{(\ell)} = \boldsymbol{S}_L^{(\ell)}$ for $\ell < L-1$ is already optimal. Minimizing only the last block row yields some $\boldsymbol{B}$, giving the block–companion minimizer

$$\boldsymbol{C}_L = \boldsymbol{S}_L + (\boldsymbol{e}_L \otimes \boldsymbol{I}_D)\boldsymbol{B}.$$

Minimizing only the last block row amounts to

$$\min_{\boldsymbol{B}\in\mathbb{C}^{D\times DL}} \|\boldsymbol{B}\boldsymbol{X}_0 - \boldsymbol{X}_1^{(L-1)}\|_F^2.$$

Throughout, we fix

$$\boldsymbol{B} := \boldsymbol{X}_1^{(L-1)}\boldsymbol{X}_0^{\dagger},$$

the Moore–Penrose least-squares predictor.

D.2. Eigenvectors of the block-companion matrix are KV

Let

$$\boldsymbol{C}_L = \begin{bmatrix} \boldsymbol{0} & \boldsymbol{I}_D & & \\ & \ddots & \ddots & \\ & & \boldsymbol{0} & \boldsymbol{I}_D \\ \boldsymbol{B}_1 & \boldsymbol{B}_2 & \cdots & \boldsymbol{B}_L \end{bmatrix}.$$

Write an eigenvector as

$$\boldsymbol{v} = [\, \boldsymbol{v}^{(0)}; \ldots; \boldsymbol{v}^{(L-1)} \,].$$

For $\ell = 0, \ldots, L-2$,

$$\boldsymbol{v}^{(\ell+1)} = \lambda\, \boldsymbol{v}^{(\ell)},$$

so $\boldsymbol{v}^{(\ell)} = \lambda^\ell \boldsymbol{v}^{(0)}$. The final block row imposes

$$(\boldsymbol{B}_1 + \lambda \boldsymbol{B}_2 + \cdots + \lambda^{L-1} \boldsymbol{B}_L)\boldsymbol{v}^{(0)} = \lambda^L \boldsymbol{v}^{(0)}.$$

Hence every eigenvector has KV form

$$\boldsymbol{v} = \boldsymbol{v}_L(\lambda) \otimes \boldsymbol{v}^{(0)}, \qquad \boldsymbol{v}_L(\lambda) = [\, 1, \lambda, \ldots, \lambda^{L-1} \,]^\top.$$

D.3. The Moore–Penrose minimizer satisfies $\boldsymbol{X}_1 \boldsymbol{X}_0^\dagger = \boldsymbol{C}_L \boldsymbol{P}_{\mathcal{U}}$

Let $\mathcal{U} = \mathrm{col}(\boldsymbol{X}_0)$ and $\boldsymbol{P}_{\mathcal{U}}$ its projector.

Lemma .2 (Moore–Penrose projection identity)**.** *Let $\mathcal{U} = \mathrm{col}(\boldsymbol{X}_0)$ and let*

$$\boldsymbol{P}_{\mathcal{U}} := \boldsymbol{X}_0 \boldsymbol{X}_0^\dagger$$

be the orthogonal projector onto $\mathcal{U}$. Consider the least-squares problem

$$\min_{\boldsymbol{A} \in \mathbb{C}^{DL \times DL}} \|\boldsymbol{A}\boldsymbol{X}_0 - \boldsymbol{X}_1\|_F^2,$$

and define the Moore–Penrose (minimum-norm) minimizer

$$\boldsymbol{A}_{\mathrm{MP}} := \boldsymbol{X}_1 \boldsymbol{X}_0^\dagger.$$

Then, for any minimizer $\boldsymbol{A}_\star$,

$$\boldsymbol{A}_{\mathrm{MP}} = \boldsymbol{A}_\star\, \boldsymbol{P}_{\mathcal{U}}.$$

Proof. Any minimizer $\boldsymbol{A}_\star$ satisfies the normal equations

$$(\boldsymbol{A}_\star \boldsymbol{X}_0 - \boldsymbol{X}_1)\boldsymbol{X}_0^H = \boldsymbol{0}, \qquad \text{equivalently} \qquad \boldsymbol{A}_\star \boldsymbol{X}_0 \boldsymbol{X}_0^H = \boldsymbol{X}_1 \boldsymbol{X}_0^H.$$

A standard characterization of the solution set of this Frobenius least-squares problem is

$$\boldsymbol{A}_\star = \boldsymbol{X}_1 \boldsymbol{X}_0^\dagger + \boldsymbol{Z}(\boldsymbol{I} - \boldsymbol{P}_{\mathcal{U}}), \qquad \boldsymbol{Z} \in \mathbb{C}^{DL \times DL}.$$

Right-multiplying by $\boldsymbol{P}_{\mathcal{U}}$ and using $\boldsymbol{P}_{\mathcal{U}}(\boldsymbol{I} - \boldsymbol{P}_{\mathcal{U}}) = \boldsymbol{0}$ yields

$$\boldsymbol{A}_\star \boldsymbol{P}_{\mathcal{U}} = (\boldsymbol{X}_1 \boldsymbol{X}_0^\dagger + \boldsymbol{Z}(\boldsymbol{I} - \boldsymbol{P}_{\mathcal{U}}))\boldsymbol{P}_{\mathcal{U}} = \boldsymbol{X}_1 \boldsymbol{X}_0^\dagger = \boldsymbol{A}_{\mathrm{MP}}.$$

□

D.4. The reduced propagator is the orthogonal compression of C_L

Let $\boldsymbol{X}_0 = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^H$ be the SVD and let $\boldsymbol{U}_M$ be its first M columns. Define the reduced propagator

$$\boldsymbol{A}_M := \boldsymbol{U}_M^H \, \boldsymbol{A}_{\mathrm{MP}} \, \boldsymbol{U}_M.$$

Lemma .3. *With $\boldsymbol{C}_L$ from Subsection D.1,*

$$\boldsymbol{A}_M = \boldsymbol{U}_M^H \boldsymbol{C}_L \boldsymbol{U}_M.$$

Proof. From Lemma .2, $\boldsymbol{A}_{\mathrm{MP}} = \boldsymbol{C}_L \boldsymbol{P}_{\mathcal{U}}$. Since $\mathcal{U}_M \subset \mathcal{U}$, we have $\boldsymbol{P}_{\mathcal{U}} \boldsymbol{U}_M = \boldsymbol{U}_M$, hence

$$\boldsymbol{A}_M = \boldsymbol{U}_M^H \boldsymbol{A}_{\mathrm{MP}} \boldsymbol{U}_M = \boldsymbol{U}_M^H \boldsymbol{C}_L \boldsymbol{U}_M.$$

□

D.5. Proof of Theorem 5.4

Proof. Fix an eigenpair $(\widehat{\lambda}_j, \boldsymbol{w}_j)$ of $\boldsymbol{A}_M$ from (2.12), and let $\widehat{\boldsymbol{\phi}}_j^p$ and $\widehat{\boldsymbol{\phi}}_j^e$ be the associated projected and exact DMD modes from (2.14) and (2.16). Recall that $\widehat{\boldsymbol{\phi}}_j^p \in \mathcal{U}_M$, so $\boldsymbol{P}_{\mathcal{U}_M} \widehat{\boldsymbol{\phi}}_j^p = \widehat{\boldsymbol{\phi}}_j^p$.

We begin by projecting the exact mode onto $\mathcal{U}_M$ and identifying the reduced propagator within the expression. Using (2.16),

$$\boldsymbol{P}_{\mathcal{U}_M} \widehat{\boldsymbol{\phi}}_j^e = \boldsymbol{U}_M \underbrace{\left(\boldsymbol{U}_M^H \boldsymbol{X}_1 \boldsymbol{V}_M \boldsymbol{\Sigma}_M^{-1} \right)}_{\boldsymbol{A}_M} \boldsymbol{w}_j = \boldsymbol{U}_M \, \boldsymbol{A}_M \, \boldsymbol{w}_j, \tag{.9}$$

We now invoke the eigen-relation $\boldsymbol{A}_M \boldsymbol{w}_j = \widehat{\lambda}_j \boldsymbol{w}_j$ and the definition of projected DMD modes (2.14) to obtain

$$\boldsymbol{P}_{\mathcal{U}_M} \widehat{\boldsymbol{\phi}}_j^e = \boldsymbol{U}_M \boldsymbol{A}_M \boldsymbol{w}_j = \widehat{\lambda}_j \, \boldsymbol{U}_M \boldsymbol{w}_j = \widehat{\lambda}_j \, \widehat{\boldsymbol{\phi}}_j^p. \tag{.10}$$

Hence,

$$\boldsymbol{r}_{\mathcal{U}_M}\!\left(\widehat{\boldsymbol{\phi}}_j^e\right) = (\boldsymbol{I} - \boldsymbol{P}_{\mathcal{U}_M}) \widehat{\boldsymbol{\phi}}_j^e = \widehat{\boldsymbol{\phi}}_j^e - \hat{\lambda}_j \, \widehat{\boldsymbol{\phi}}_j^p. \tag{.11}$$

Next, by Lemma .2 and the existence of the block-companion minimizer $\boldsymbol{C}_L$ from Subsection D.1, we have

$$\boldsymbol{A}_{\mathrm{MP}} = \boldsymbol{X}_1 \boldsymbol{X}_0^{\dagger} = \boldsymbol{C}_L \, \boldsymbol{P}_{\mathcal{U}}, \qquad \mathcal{U} = \mathrm{col}(\boldsymbol{X}_0).$$

Since $\widehat{\boldsymbol{\phi}}_j^p = \boldsymbol{U}_M \boldsymbol{w}_j \in \mathcal{U}$, it follows that $\boldsymbol{P}_{\mathcal{U}} \widehat{\boldsymbol{\phi}}_j^p = \widehat{\boldsymbol{\phi}}_j^p$ and thus

$$\boldsymbol{C}_L \widehat{\boldsymbol{\phi}}_j^p = \boldsymbol{A}_{\mathrm{MP}} \widehat{\boldsymbol{\phi}}_j^p = \boldsymbol{X}_1 \boldsymbol{X}_0^{\dagger} \boldsymbol{U}_M \boldsymbol{w}_j.$$

Using the (untruncated) SVD $\boldsymbol{X}_0 = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^H$ and the fact that $\boldsymbol{U}_M$ consists of the first M columns of $\boldsymbol{U}$, we obtain

$$\boldsymbol{X}_0^{\dagger} \boldsymbol{U}_M = \boldsymbol{V} \boldsymbol{\Sigma}^{-1} \boldsymbol{U}^H \boldsymbol{U}_M = \boldsymbol{V}_M \boldsymbol{\Sigma}_M^{-1},$$

and therefore

$$\boldsymbol{C}_L \widehat{\boldsymbol{\phi}}_j^p = \boldsymbol{X}_1 \boldsymbol{V}_M \boldsymbol{\Sigma}_M^{-1} \boldsymbol{w}_j = \widehat{\boldsymbol{\phi}}_j^e$$

by (2.16). Combining with (.11) yields

$$(\boldsymbol{C}_L - \hat{\lambda}_j \boldsymbol{I}) \widehat{\boldsymbol{\phi}}_j^p = \widehat{\boldsymbol{\phi}}_j^e - \hat{\lambda}_j \, \widehat{\boldsymbol{\phi}}_j^p = \boldsymbol{r}_{\mathcal{U}_M}\big(\widehat{\boldsymbol{\phi}}_j^e\big),$$

which is (5.8). □

E. Fixed-Eigenvalue KV Fit (FEKVF)

The fixed-eigenvalue KV fit (FEKVF) provides a lightweight alternative to the nested rank-1 DMD of Sec. 5.2. After reshaping each projected mode into a per-mode lag matrix

$$\widehat{\boldsymbol{\phi}}_j^p \;\mapsto\; \widehat{\boldsymbol{\Phi}}_j \;\coloneqq\; \mathrm{reshape}\big(\widehat{\boldsymbol{\phi}}_j^p,\, D \times L\big) \in \mathbb{C}^{D \times L},$$

FEKVF evaluates how well $\widehat{\boldsymbol{\Phi}}_j$ conforms to a KV evolution that is consistent with its existing eigenvalue $\widehat{\lambda}_j$. Unlike the nested fit, no inner DMD is performed; the eigenvalue is held fixed.

Closed-form residual

For the eigenvalue $\widehat{\lambda}_j$, define the Vandermonde vector

$$\boldsymbol{a}_j \;\coloneqq\; \boldsymbol{v}_L(\widehat{\lambda}_j) = [\,1, \widehat{\lambda}_j, \ldots, \widehat{\lambda}_j^{L-1}\,]^\top.$$

Projecting $\widehat{\boldsymbol{\Phi}}_j$ onto the fixed-eigenvalue KV cone amounts to the rank-1 least-squares problem

$$\min_{\boldsymbol{u} \in \mathbb{C}^D} \big\| \widehat{\boldsymbol{\Phi}}_j - \boldsymbol{u}\, \boldsymbol{a}_j^\top \big\|_F^2,$$

whose minimizer is

$$\boldsymbol{u}_j^\star \;=\; \frac{\widehat{\boldsymbol{\Phi}}_j \, \boldsymbol{a}_j^*}{\|\boldsymbol{a}_j\|_2^2}, \qquad \widehat{\boldsymbol{\Phi}}_j^{(\mathrm{KV})} \;\coloneqq\; \boldsymbol{u}_j^\star \, \boldsymbol{a}_j^\top.$$

The corresponding normalized residual is

$$\mathcal{R}_j^{(\mathrm{KV})} \;\coloneqq\; \frac{\|\widehat{\boldsymbol{\Phi}}_j - \widehat{\boldsymbol{\Phi}}_j^{(\mathrm{KV})}\|_F^2}{\|\widehat{\boldsymbol{\Phi}}_j\|_F^2} \;=\; 1 - \frac{\|\widehat{\boldsymbol{\Phi}}_j \, \boldsymbol{a}_j^*\|_2^2}{\|\widehat{\boldsymbol{\Phi}}_j\|_F^2 \, \|\boldsymbol{a}_j\|_2^2}.$$

We form a scalar KV-conformity score

$$\zeta_j := \log\big(\mathcal{R}_j^{(\mathrm{KV})} + \varepsilon\big),$$

where $\varepsilon > 0$ prevents singularities. Smaller values indicate closer agreement with the KV pattern implied by $\widehat{\lambda}_j$.

Computation

All operations in FEKVF reduce to one $D \times L$ matrix-vector product and a small number of scalar reductions per mode. In particular, no inner DMD problem is solved and the eigenvalue is reused rather than re-estimated. As a result, the per-mode effort is substantially smaller than for the nested rank-1 fit, and the procedure applies independently to each mode and is easily implemented in a batched fashion. A formal comparison of the asymptotic costs of FEKVF and nested DMD is given in Appendix F.

Require: Mode matrices $\{\widehat{\boldsymbol{\Phi}}_j \in \mathbb{C}^{D\times L}\}_{j=1}^{M}$, eigenvalues $\{\widehat{\lambda}_j\}$, small $\varepsilon > 0$; clustering routine and aggregation rule as in Algorithm 4.2
Ensure: Order estimate $\widehat{m}$ and labels $\{\ell_j\}$
for $j = 1, \ldots, M$ **do**
  $\boldsymbol{a} \leftarrow [\,1, \widehat{\lambda}_j, \ldots, \widehat{\lambda}_j^{L-1}\,]^\top$
  $\boldsymbol{w} \leftarrow \boldsymbol{a}^*$
  $\boldsymbol{u}^\star \leftarrow \widehat{\boldsymbol{\Phi}}_j\, \boldsymbol{w} / \|\boldsymbol{a}\|_2^2$
  $\widehat{\boldsymbol{\Phi}}_j^{(\mathrm{KV})} \leftarrow \boldsymbol{u}^\star\, \boldsymbol{a}^\top$
  $\mathcal{R}_j^{(\mathrm{KV})} \leftarrow \|\widehat{\boldsymbol{\Phi}}_j - \widehat{\boldsymbol{\Phi}}_j^{(\mathrm{KV})}\|_F^2 / \|\widehat{\boldsymbol{\Phi}}_j\|_F^2$
  $\zeta_j \leftarrow \log(\mathcal{R}_j^{(\mathrm{KV})} + \varepsilon)$
end for
Stack $\boldsymbol{\zeta} = [\zeta_1, \ldots, \zeta_M]$
Apply binary mode selection framework (Algorithm 4.2) to obtain $\{\ell_j\}$ and $\widehat{m}$

Algorithm 1.: Fixed-Eigenvalue KV Fit (FEKVF)

Remarks. FEKVF shares the same conceptual interpretation as the nested rank-1 DMD: both measure global KV conformity of the reconstructed mode trajectories. The key difference is that FEKVF holds the eigenvalue fixed and performs a single rank-1 projection, producing a scalar KV-conformity score ζ_j that feeds directly into the binary mode selection framework (Algorithm 4.2).

F. Computational Complexity of the Procedures

This appendix summarizes the computational costs associated with the procedures used throughout the paper. Throughout, D denotes the spatial dimension, L the embedding length, DL the lifted dimension, and $M > m$ the truncation rank.

F.1. DMD construction

The delay-embedded snapshot matrices $\boldsymbol{X}_0, \boldsymbol{X}_1 \in \mathbb{C}^{DL\times T}$ are processed by a rank-M SVD,

$$\boldsymbol{X}_0 \approx \boldsymbol{U}_M \boldsymbol{\Sigma}_M \boldsymbol{V}_M^\top.$$

Computing this truncated SVD has cost

$$O(DL\,T\,M),$$

which is the dominant cost in the full pipeline when T is not extremely small. The projected modes $\widehat{\boldsymbol{\phi}}_j^p$ require a single matrix–vector multiplication per mode and cost $O(DL\,M)$ in total. Exact modes $\widehat{\boldsymbol{\phi}}_j^e = \boldsymbol{X}_1 \boldsymbol{V}_M \boldsymbol{\Sigma}_M^{-1} e_j$ require an additional multiplication by $\boldsymbol{X}_1$, giving the same overall $O(DL\,M)$ cost.

F.2. Nested DMD and block-companion construction

The nested-DMD operator (Section 5.2) is obtained from the rank-M SVD factors via a least-squares fit of a block-companion structure. The fit is linear in M blocks of spatial dimension D, yielding total cost

$$O(DL\,M)$$

for computing the block coefficients and the nested eigenpairs. This cost is negligible compared with the truncated SVD of $\boldsymbol{X}_0$.

F.3. KV evaluation and fixed-eigenvalue KV fit

Given eigenvalues $\{\widehat{\lambda}_j\}$, the KV structure induces L shifted blocks of spatial vectors. Evaluating the full KV matrix requires $O(DLM)$ operations.

The fixed-eigenvalue KV fit (Appendix E) solves a least-squares problem

$$\min_{\boldsymbol{A}} \|\boldsymbol{S}_0 - \boldsymbol{A}\boldsymbol{V}\|_F^2$$

where the block shifts are determined by the $\widehat{\lambda}_j$. This regression has cost

$$O(DLM)$$

when solved via QR or normal equations, and is again dominated by the rank-M SVD of $\boldsymbol{X}_0$.

F.4. ESR-score computation

Each exact mode is a vector in $\mathbb{C}^{DL}$. Using the orthogonal decomposition (Sec. 4)

$$\widehat{\boldsymbol{\phi}}_j^e \;=\; \widehat{\lambda}_j\,\widehat{\boldsymbol{\phi}}_j^p \;+\; \boldsymbol{r}_{\mathcal{U}_M}(\widehat{\boldsymbol{\phi}}_j^e), \qquad \|\widehat{\boldsymbol{\phi}}_j^p\|_2 = 1,$$

the ESR score (residual energy) satisfies

$$\mathcal{R}_j \;=\; \big\|\boldsymbol{r}_{\mathcal{U}_M}(\widehat{\boldsymbol{\phi}}_j^e)\big\|_2^2 \;=\; \|\widehat{\boldsymbol{\phi}}_j^e\|_2^2 - |\widehat{\lambda}_j|^2.$$

Thus no explicit projector applications are required, and evaluating all $\mathcal{R}_j$ (and derived scalar scores such as ζ_j) over M modes costs

$$O(DL\,M),$$

with $O(M)$ additional memory.

F.5. Binary mode selection

The binary mode selection rule of Algorithm 4.2 operates on a per-mode scalar score $f_j \in \mathbb{R}$.

Clustering the M scalar scores into $k = 2$ groups (e.g., K-Means or a two-component GMM) costs

$$O(M)$$

per iteration for K-Means, and similarly

$$O(M)$$

per EM iteration for a diagonal-covariance GMM. In all cases, with $k = 2$, this clustering cost is negligible compared with the other steps of the pipeline.

Code and Data Availability

The data that support the findings of this study are available within this thesis. The data were generated via numerical simulations, and the Python code used to generate the data and reproduce the figures is available on GitHub at `https://github.com/YoavHarris/geometric-dc-dmd-order-detection`. A citable archival release is available via Zenodo [16].

All algorithms presented in this work have been implemented and are included in the repository for reproducibility and future research.

Bibliography

[1] H. Arbabi and I. Mezić. 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.

[2] J. C. A. Barata and M. S. Hussein. The moore–penrose pseudoinverse: A tutorial review of the theory. *Brazilian Journal of Physics*, 42(1-2):146–165, 2012.

[3] P. Benner, S. Gugercin, and K. Willcox. A survey of projection-based model reduction methods for parametric dynamical systems. *SIAM Review*, 57(4):483–531, 2015.

[4] G. Berkooz, P. Holmes, and J. L. Lumley. The proper orthogonal decomposition in the analysis of turbulent flows. *Annual Review of Fluid Mechanics*, 25:539–575, 1993.

[5] E. Bronstein, A. Wiegner, D. Shilo, and R. Talmon. The spatiotemporal coupling in delay-coordinates dynamic mode decomposition. *Chaos*, 32(12):123127, 2022.

[6] M. Budišić, R. Mohr, and I. Mezić. Applied koopmanism. *Chaos*, 22(4):047510, 2012.

[7] E. J. Candès and C. Fernandez-Granda. Towards a mathematical theory of super-resolution. *Communications on Pure and Applied Mathematics*, 67(6):906–956, 2014.

[8] M. M. Churchland, J. P. Cunningham, M. T. Kaufman, J. D. Foster, P. Nuyujukian, S. I. Ryu, and K. V. Shenoy. Neural population dynamics during reaching. *Nature*, 487(7405):51–56, 2012.

[9] M. J. Colbrook. The multiverse of dynamic mode decomposition algorithms. *SIAM Review*, 66(3):483–530, 2024.

[10] M. J. Colbrook, L. J. Ayton, and M. Szőke. Residual dynamic mode decomposition: Robust and verified koopmanism. *Journal of Fluid Mechanics*, 955:A21, 2023.

[11] J. P. Cunningham and B. M. Yu. Dimensionality reduction for large-scale neural recordings. *Nature Neuroscience*, 17(11):1500–1509, 2014.

[12] S. Das and D. Giannakis. Delay-coordinate maps and the spectra of koopman operators. *Journal of Statistical Physics*, 175:1107–1145, 2019.

[13] S. Das and D. Giannakis. Koopman spectra in reproducing kernel hilbert spaces. *Applied and Computational Harmonic Analysis*, 49:573–607, 2020.

[14] S. T. M. Dawson, M. S. Hemati, M. O. Williams, and C. W. Rowley. Characterizing and correcting for the effect of sensor noise in the dynamic mode decomposition. *Experiments in Fluids*, 57(3):42, 2016.

[15] A. Hannachi, I. T. Jolliffe, and D. B. Stephenson. Empirical orthogonal functions and related techniques in atmospheric science: A review. *International Journal of Climatology*, 27(9):1119–1152, 2007.

[16] Y. Harris. Geometric-spectral framework for order detection in delay-embedded dmd, 2026.

[17] Y. H. Hua and T. K. Sarkar. Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 38(5):814–824, 1990.

[18] M. R. Jovanović, P. J. Schmid, and J. W. Nichols. Sparsity-promoting dynamic mode decomposition. *Physics of Fluids*, 26(2):024103, 2014.

[19] J.-N. Juang and R. S. Pappa. An eigensystem realization algorithm for modal parameter identification and model reduction. *Journal of Guidance, Control, and Dynamics*, 8(5):620–627, 1985.

[20] G. Kerschen, J. C. Golinval, A. F. Vakakis, and L. A. Bergman. The method of proper orthogonal decomposition for dynamical characterization and order reduction of mechanical systems: An overview. *Nonlinear Dynamics*, 41(1–3):147–169, 2005.

[21] S. Kunis, D. Potts, and M. Tasche. On the condition number of vandermonde matrices. *Linear Algebra and its Applications*, 500:283–299, 2016.

[22] J. N. Kutz, S. L. Brunton, B. W. Brunton, and J. L. Proctor. *Dynamic Mode Decomposition: Data-Driven Modeling of Complex Systems*. SIAM, 2016.

[23] S. Le Clainche and J. M. Vega. Higher order dynamic mode decomposition and its applications. *SIAM Journal on Applied Dynamical Systems*, 16(2):882–925, 2017.

[24] D. S. Mackey, N. Mackey, C. Mehl, and V. Mehrmann. Vector spaces of linearizations for matrix polynomials. *SIAM Journal on Matrix Analysis and Applications*, 28(4):971–1004, 2006.

[25] I. Mezić. Spectral properties of dynamical systems, model reduction and decompositions. *Nonlinear Dynamics*, 41(1-3):309–325, 2005.

[26] I. Mezić. Analysis of fluid flows via spectral properties of the koopman operator. *Annual Review of Fluid Mechanics*, 45:357–378, 2013.

[27] S. Pan, N. Arnold-Medabalimi, and K. Duraisamy. Sparsity-promoting algorithms for the discovery of informative koopman-invariant subspaces. *Journal of Fluid Mechanics*, 917:A18, 2021.

[28] S. Pan and K. Duraisamy. Physics-informed probabilistic learning of linear embeddings of nonlinear dynamics with guaranteed stability. *SIAM Journal on Applied Dynamical Systems*, 19(1):480–509, 2020.

[29] R. Penrose. On best approximate solutions of linear matrix equations. *Mathematical Proceedings of the Cambridge Philosophical Society*, 52(1):17–19, 1956.

[30] J. L. Proctor, S. L. Brunton, and J. N. Kutz. Dynamic mode decomposition with control. *SIAM Journal on Applied Dynamical Systems*, 15(1):142–161, 2016.

[31] J. Rissanen. Modeling by shortest data description. *Automatica*, 14(5):465–471, 1978.

[32] C. W. Rowley, I. Mezić, S. Bagheri, P. Schlatter, and D. S. Henningson. Spectral analysis of nonlinear flows. *Journal of Fluid Mechanics*, 641:115–127, 2009.

[33] A. Saito and T. Kuno. Data-driven experimental modal analysis by dynamic mode decomposition. *Journal of Sound and Vibration*, 481:115434, 2020.

[34] D. Sashidhar and J. N. Kutz. Bagging, optimized dynamic mode decomposition (bop-dmd) for robust, stable forecasting with spatial and temporal uncertainty-quantification. *Philosophical Transactions of the Royal Society A*, 380(2222):20210195, 2022.

[35] P. J. Schmid. Dynamic mode decomposition of numerical and experimental data. *Journal of Fluid Mechanics*, 656:5–28, 2010.

[36] P. J. Schmid. Dynamic mode decomposition and its variants. *Annual Review of Fluid Mechanics*, 54:225–254, 2022.

[37] G. Schwarz. Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464, 1978.

[38] Y.-I. Segman, A. Amar, and R. Talmon. Samp++: Robust structure-aware matrix pencil. In *Proceedings of EUSIPCO 2025*, pages –, 2025.

[39] Y.-I. Segman, A. Amar, and R. Talmon. Structure-aware matrix pencil method. *arXiv preprint arXiv:2502.17047*, 2025.

[40] K. Soundararajan. Equidistribution of zeros of polynomials. *American Mathematical Monthly*, 126(5):409–412, 2019.

[41] P. Stoica and Y. Selén. Model-order selection: A review of information criterion rules. *IEEE Signal Processing Magazine*, 21(4):36–47, 2004.

[42] F. Takens. Detecting strange attractors in turbulence. In D. A. Rand and L.-S. Young, editors, *Dynamical Systems and Turbulence, Warwick 1980*, volume 898 of *Lecture Notes in Mathematics*, pages 366–381. Springer, 1981.

[43] D. W. J. Thompson and J. M. Wallace. Annular modes in the extratropical circulation. part i: Month-to-month variability. *Journal of Climate*, 13(5):1000–1016, 2000.

[44] J. H. Tu, C. W. Rowley, D. M. Luchtenburg, S. L. Brunton, and J. N. Kutz. On dynamic mode decomposition: Theory and applications. *Journal of Computational Dynamics*, 1(2):391–421, 2014.

[45] M. Wax and T. Kailath. Detection of signals by information theoretic criteria. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 33(2):387–392, 1985.

[46] M. O. Williams, M. S. Hemati, S. T. M. Dawson, I. G. Kevrekidis, and C. W. Rowley. Extending data-driven koopman analysis to actuated systems. *IFAC-PapersOnLine*, 49(18):704–709, 2016.

[47] M. O. Williams, I. G. Kevrekidis, and C. W. Rowley. A data–driven approximation of the koopman operator: Extending dynamic mode decomposition. *Journal of Nonlinear Science*, 25:1307–1346, 2015.

פרק בעברית

מסגרת גיאומטרית לגילוי סדר מודל חסין בפירוק לרכיבים דינמיים

חיבור על מחקר

לשם מילוי חלקי של הדרישות לקבלת התואר
מגיסטר למדעים בהנדסת חשמל

יואב האריס

הוגש לסנאט הטכניון – מכון טכנולוגי לישראל

פברואר 2026 חיפה שבט התשפ"ו

תודות

המחקר נערך בפקולטה להנדסת חשמל ומחשבים על שם ויטרבי תחת הנחייתם המשותפת של פרופ׳ רונן טלמון וד״ר הדס בניסטי.

אני מודה לטכניון ולמלגת ליאוניד ודיאן שרמן לתואר שני במחקר רב תחומי על התמיכה הכספית הנדיבה בהשתלמותי.

מחבר חיבור זה מצהיר כי המחקר, כולל איסוף הנתונים, עיבודם והצגתם, התייחסות והשוואה למחקרים קודמים וכו׳, נעשה כולו בצורה ישרה, כמצופה ממחקר מדעי המבוצע לפי אמות המידה האתיות של העולם האקדמי. כמו כן, הדיווח על המחקר ותוצאותיו בחיבור זה נעשה בצורה ישרה ומלאה, לפי אותן אמות מידה.

תקציר

סדרות עיתיות מרובות־מימד מופיעות בתחומים רבים, כגון מכניקת נוזלית, מדידות עצביות, נתונים פיננסים, נתוני אקלים ומערכות מכניות. לעיתים קרובות הן נשלטות על ידי מספר קטן יחסית של רכיבים דינמיים. ניתוח סדרות אלה מתבסס לעיתים קרובות על ההנחה שהדינמיקה נשלטת על ידי מספר קטן של רכיבים קוהרנטיים. במסגרת זו, noitisopmoceD edoM cimanyD מפרש את ההתפתחות בזמן דרך אופנים מרחביים וערכים עצמיים של אופרטור לינארי הנלמד מהדאטה. הרחבה באמצעות קואורדינטות השהייה מאפשרת לשפר את הרזולוציה הספקטרלית ולהעשיר את המידע הזמין לפירוק, אך כאשר סדר העבודה נבחר גדול מן הסדר האמיתי, מתקבלים אופנים נוספים שיש לסווג בדיעבד כחלק מתהליך מציאת הסדר.

פירוק אופנים דינמי (DMD) עם קואורדינטות השהייה (CD) נמצא בשימוש רחב על מנת לחלץ רכיבים מרחביים וזמניים מסדרות עיתיות במימד גבוה בפועל, רעש מדידה הופך את מטריצות הנתונים למלאות־דרגה, ולכן החישוב מבוצע לרוב בסדר עבודה Mהעולה על סדר האות m. מצב זה מוביל להופעת אופנים נוספים שאינם מייצגים דינמיקה אמיתית ומציב את בעיית מציאת הסדר כבעיית סיווג אופנים.

קיימות גישות רבות למציאת סדר ב DMD. ביניהן שערוך דרגת מטריצת הדאטה, שיטות קריטריונים אינפורמציוניים הנגזרים מהנחות סבירות, ושיטות הבוחנות את עמידות ברכיבים להפרעות. גישות אלו נבדלות זו מזו בהנחות המודל ובאופן שבו הן מפרשות אופנים סרק, אך לרוב אינן מספקות הגדרה מפורשת המבחינה בין רכיבים אמיתיים לרכיבים הנובעים מהערכת יתר של הסדר. עבודה זו ממקמת את עצמה בתוך מסגרת זו ומבקשת לנסח הבחנה כזו באופן ישיר.

עבודה זו מאמצת נקודת מבט גאומטרית למציאת הסדר ומנסחת הבחנה פורמלית בין אופנים אמיתיים ואופני סרק דרך הקשר שלהם לתת־מרחב האות. אנו מגדירים את זהות הרכיב כאמיתי או סרק על בסיס עוצמת ההכלה שלו בתת המרחב שפורס האות הנקי. הגדרה פורמלית (אם כי קונספטואלית בלבד) זו מאפשרת לנו לגזור קריטריון החלטה פרקטי, המאפשר לנתח ולסווג את הרכיבים לפי הגאומטריה שלהם.

במסגרת זו, אנו מציעים שתי גישות נפרדות (אך קשורות) לבחירת רכיבים רובוסטית ב־DMD־CD. הראשונה הינה הקריטריון הגאומטרי הנגזר מריכוך ההגדרה הפורמלית שמוזכרת לעיל. זו מנצלת את העובדה שאופנים אמיתיים מרכזים את האנרגיה שלהם בתת מרחב האות, ואילו אופני סרק, המייצגים רעש איזוטרופי, מפזרים את האנרגיה שלהם באופן אחיד יותר, ושומרים רכיב משמעותי מחוץ לכל תת מרחב. רכיב זה הינו שארית גאומטרית הנמדדת ביחס לתת־המרחב המשוער, והוא מאפשר להבדיל בין אופני אמת לאופני סרק.

השיטה השנייה מבוססת על קריטריון מבני המעוגן בניתוח אופרטורי חדש של CD־DMD. הניתוח קושר את האופנים למטריצת מלווה־בלוקים. קישור זה גורר מסקנות אודות מבנה קרונקר־ואנדרמונדה של האופנים במסגרת ניסוח מלווה־בלוקים, ומראה שחריגות של אופנים ממבנה זה נשלטות על ידי השאריות הגאומטריות הנזכרות לעיל. תוצאה זו מראה כי מבנה קרונקר־ואנדרמונדה אינו מאפיין בלעדי של אופנים אמיתיים, אלא תכונה הנובעת ממבנה ההשהיות עצמו. בהתאם לכך, מידת ההתאמה למבנה זה נתפסת כתכונה רציפה, ולא בינארית. והקשר בינה לבין השאריות הגאומטריות מאפשר

לקשור בין הניתוח המבני לבין הקריטריון הגאומטרי במסגרת תאורטית אחת.

הניתוחים הנכללים בעבודה, הן הגאומטרי והן האלגברי, מציגים באור חדש שיטה קיימת לברירת אופנים, מהפופולאריות ביותר שבשימוש כיום: סיווג על פי נורמה של אופן. אנו מציגים ניתוח מעמיק שמבהיר מדוע וכיצד שיטה זו נכשלת כאשר משתמשים בקואורדינטות השהייה.

סימולציות וניסויים אמפיריים מקיפים מאשרים את התחזיות התאורטיות, ומדגימים ששיטות הסיווג מבוססות הגאומטריה והמבנה משיגות תוצאות זיהוי־סדר עדיפות בעקביות, בהשוואה לשיטות קיימות. היתרון של השיטות החדשות נשמר ואף מודגש בניסויים לאורך טווח רחב של עוצמות רעש, הפרדה ספקטרלית, משטרי דעיכה, ומספר ההשהיות.