Collinearity-Resilient Sparse Identification of Nonlinear Dynamics via Cluster-Stability and Representative Selection

Molin Ouyang[1], Fudong Ge[1,2,*], Zhiqiang Zuo[1,2] and Zhicheng Zhang[1,2]

[1]School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, PR China
[2]Tianjin Key Laboratory of Intelligent Unmanned Swarm Technology and System, Tianjin University, Tianjin 300072, PR China
*Corresponding author. E-mail address: gefd@tju.edu.cn

Abstract: Sparse identification of nonlinear dynamics (SINDy) exhibit structural fragilely when the candidate library contains highly correlated functions, as standard feature-wise selection tends to scatter selection frequencies across near-multicollinear terms and can therefore omit truly active variables. To overcome this limitation, in this paper, we propose a collinearity-aware SINDy framework that groups correlated candidates, evaluates cluster-level stability over repeated sparse-regression fits, and disambiguates each retained group by selecting a unique representative through independent-trajectory validation, followed by coefficient re-estimation on the fixed support. Experiments show that the proposed cluster-level approach achieves 85.0% exact recovery at a derivative-perturbation level of 0.30, compared with 67.7% for tuned STRidge and 35.3% for tuned feature-wise stability selection. Its applicability is further validated on a damped nonlinear pendulum, where 100.0% system-level exact recovery was achieved for the highly correlated terms θ and $\sin\theta$. Collectively, these results demonstrate that cluster-level stability and independent-trajectory disambiguation can improve the structural recovery in highly collinear candidate libraries.

Key Words: Sparse identification of nonlinear dynamics; Candidate-library collinearity; Stability selection; Representative selection; Numerical differentiation

1 Introduction

Sparse identification of nonlinear dynamics (SINDy) assumes that a system can be represented by a large candidate function library, yet only a small number of candidates are active [1]. For state snapshots $\mathbf{X}$ sampled along a trajectory and their derivatives $\dot{\mathbf{X}}$, the identification problem is formulated as

$$\dot{\mathbf{X}} = \Theta(\mathbf{X})\Xi + \mathbf{E}, \tag{1}$$

where $\Theta(\mathbf{X})$ denotes the candidate library, Ξ is the sparse coefficient matrix, and $\mathbf{E}$ represents noise and model mismatch. The nonzero entries of Ξ determine the active candidate functions and their coefficients [2]. A common approach for estimating Ξ is sequentially thresholded ridge regression (STRidge), which combines ridge regression with coefficient thresholding to identify the active candidate functions [3, 4]. The sparsity threshold λ determines which candidate terms are retained and therefore directly affects the identified support.

A support identified at a single sparsity threshold can be sensitive to the threshold choice and to perturbations in the data [5–8]. To reduce this dependence, feature-wise stability selection evaluates the recurrence of each candidate over repeated data subsampling and a spectrum of sparsity thresholds [9, 10]. Candidates with sufficiently high aggregated selection frequencies are then retained as stable terms.

Feature-wise selection frequencies are informative when candidate functions are sufficiently distinct. However, candidate libraries may contain highly correlated functions because of local functional equivalence over restricted state-space regions or redundant library construction. For example, x and $\sin(0.25x)/0.25$ are globally distinct but nearly identical over a limited interval. Under such conditions, STRidge may split selections among correlated candidates across subsamples, causing the selection frequency of a true term to fall below the stability cutoff even when the associated dynamical component is consistently selected [10, 11]. This behavior is referred to here as split voting.

Related methods addressed correlated predictors in different ways [12, 13]. Elastic Net promotes grouped selection and can be fitted efficiently along a regularization path [14]. Group Lasso and sparse-group Lasso imposed sparsity at the group level and at both the group and feature levels, respectively [15, 16]. Cluster stability selection aggregates selection evidence within correlated groups [17]. However, these methods are designed to retain correlated predictors or groups rather than identify the correct analytical term within a correlated group.

This motivates a two-stage strategy. Cluster-level stability selection is first used to determine whether a correlated candidate group represents an active dynamical component. Representatives of the retained multi-member clusters are then selected jointly using an independent calibration trajectory. The key idea is that candidate functions that are nearly indistinguishable over the training trajectory may become distinguishable when evaluated over a broader region of the state space. The main contributions of this work are as follows:

1) We identify the split-voting failure of feature-wise stability selection under severe candidate collinearity and introduce a diagnostic for this phenomenon.
2) A cluster-level stability framework is developed to aggregate selection evidence over highly correlated candidates and recover component-level support.
3) Retained correlated clusters are further resolved through an independent-trajectory disambiguation strategy, which selects representative terms jointly across clusters.

To evaluate the robustness and applicability of the proposed method, experiments are conducted under controlled derivative perturbations, varying collinearity strengths, correlated-variable baseline comparisons, numerical differentiation of noisy state measurements and a damped nonlinear pendulum with naturally occurring collinearity between θ and $\sin\theta$. Together, these experiments assess whether the proposed framework can reliably recover the correct equation structure under severe candidate collinearity and realis-

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

tic derivative-estimation errors.

The remainder of this paper is organized as follows. Section 2 presents the proposed collinearity-aware SINDy framework. In Section 3, we describe the benchmark systems, perturbation settings, numerical differentiation procedures, parameter choices and evaluation metrics. These are illustrated in Section 4, where the numerical results are reported.

2 Methods

2.1 Overview of the proposed framework

This work identifies a key failure mode of stability-based sparse equation discovery under severe candidate collinearity, where highly correlated functions in the candidate library induce a split-voting effect and cause the true functional term to fall below the feature-wise stability threshold. To address this, we extend the standard STRidge and feature-wise stability selection baselines with two successive mechanismsłcluster-level stability selection and within-cluster disambiguation.

As illustrated in Fig. 1, the framework first maps the collinearity structure of the candidate library into a correlation graph, from which highly correlated candidates are identified and grouped into clusters. Cluster-level stability selection is then performed on correlated clusters to determine whether the corresponding dynamical components should be retained. For each retained multi-member cluster, within-cluster disambiguation selects the appropriate representative combination using an independent calibration trajectory. The final support set combines these representatives with individually stable candidates outside the clusters, and the coefficients are re-estimated on the original training trajectory via ordinary least squares, yielding the identified governing equation.

The complete implementation procedure is summarized in Algorithm 1. The following subsections describe the key components of the proposed framework, including the baseline feature-wise stability selection, correlation-based cluster construction, cluster-level stability and split-vote diagnosis and within-cluster representative selection with final coefficient refitting.

2.2 STRidge and feature-wise stability selection

The proposed framework uses STRidge as the base sparse selector. For one state equation, the ridge estimate is

$$\hat{\xi} = \arg\min_{\xi} \left[\frac{1}{2m} \left\| \dot{\mathbf{x}}_c - \Theta^{\mathrm{std}} \xi \right\|_2^2 + \frac{\alpha_R}{2} \|\xi\|_2^2 \right], \tag{2}$$

where Θ^{std} is the standardized candidate library and $\dot{\mathbf{x}}_c$ is the centered target derivative. Candidates with $|\hat{\xi}_j| < \lambda$ are removed and ridge regression is repeated on the reduced library until the support no longer changes. Let $\hat{S}^{(b,\lambda)}$ denote the support obtained from the bth subsample at sparsity threshold λ. The feature-wise selection frequency is

$$\pi_j(\lambda) = \frac{1}{B} \sum_{b=1}^{B} \mathbb{I}\left[j \in \hat{S}^{(b,\lambda)} \right]. \tag{3}$$

Across a threshold spectrum Λ, the frequency is aggregated as

$$\pi_j^{\max} = \max_{\lambda \in \Lambda} \pi_j(\lambda). \tag{4}$$

Conventional feature-wise stability selection then retains candidate j when $\pi_j^{\max} \geq \pi_{\mathrm{thr}}$. Under severe candidate collinearity, selection evidence may be dispersed among correlated candidates through split voting. The proposed method therefore groups highly correlated candidates into clusters and evaluates their stability at the cluster level.

2.3 Detection of collinear candidate clusters

To begin with, we identify the collinearity structure of the candidate library from the training trajectory. Let θ_j denote candidate column j evaluated at the m samples along the training trajectory with sample mean

$$\bar{\theta}_j = \frac{1}{m} \mathbf{1}^{\mathsf{T}} \theta_j. \tag{5}$$

The set of candidates with nonzero centered variation is defined as

$$\mathscr{J}_{\mathrm{var}} = \left\{ j : \left\| \theta_j - \bar{\theta}_j \mathbf{1} \right\|_2 > \varepsilon_{\mathrm{var}} \right\}, \tag{6}$$

where $\varepsilon_{\mathrm{var}}$ is a small positive numerical tolerance. Constant or near-constant candidates are excluded from the correlation calculation because their centered norms are zero or numerically negligible, and are treated as singleton clusters.

For $j \in \mathscr{J}_{\mathrm{var}}$, the centered columns are normalized to unit ℓ_2 norm:

$$\tilde{\theta}_j = \frac{\theta_j - \bar{\theta}_j \mathbf{1}}{\left\| \theta_j - \bar{\theta}_j \mathbf{1} \right\|_2}. \tag{7}$$

The absolute correlation between candidates i and j is

$$C_{ij} = \left| \tilde{\theta}_i^{\mathsf{T}} \tilde{\theta}_j \right|, \qquad i, j \in \mathscr{J}_{\mathrm{var}}. \tag{8}$$

Thus, C_{ij} is the absolute Pearson correlation coefficient over the training trajectory. Two candidates are considered highly correlated when $C_{ij} \geq \tau_c$. An undirected correlation graph is constructed by connecting candidates i and j whenever this condition is satisfied. Each connected component defines a candidate cluster G_ℓ, while candidates not grouped with other candidates are treated as singleton clusters. The resulting L clusters form the partition

$$\mathscr{P} = \{G_1, G_2, \ldots, G_L\}. \tag{9}$$

The resulting partition describes the collinearity structure of the candidate library. Which clusters are retained as active dynamical components of the identified system is determined subsequently by cluster-level stability selection.

2.4 Cluster-level stability selection

A cluster G is counted as selected whenever at least one of its members appears in the fitted support. Its cluster-level selection frequency is

$$\Pi_G(\lambda) = \frac{1}{B} \sum_{b=1}^{B} \mathbb{I}\left[\hat{S}^{(b,\lambda)} \cap G \neq \varnothing \right]. \tag{10}$$

For a singleton cluster, $\Pi_G(\lambda)$ equals the corresponding feature-wise frequency. For a multi-member cluster, it aggregates selection evidence dispersed among correlated members. For final structural selection, the cluster-level frequency is aggregated over the threshold spectrum as

$$\Pi_G^{\max} = \max_{\lambda \in \Lambda} \Pi_G(\lambda), \tag{11}$$

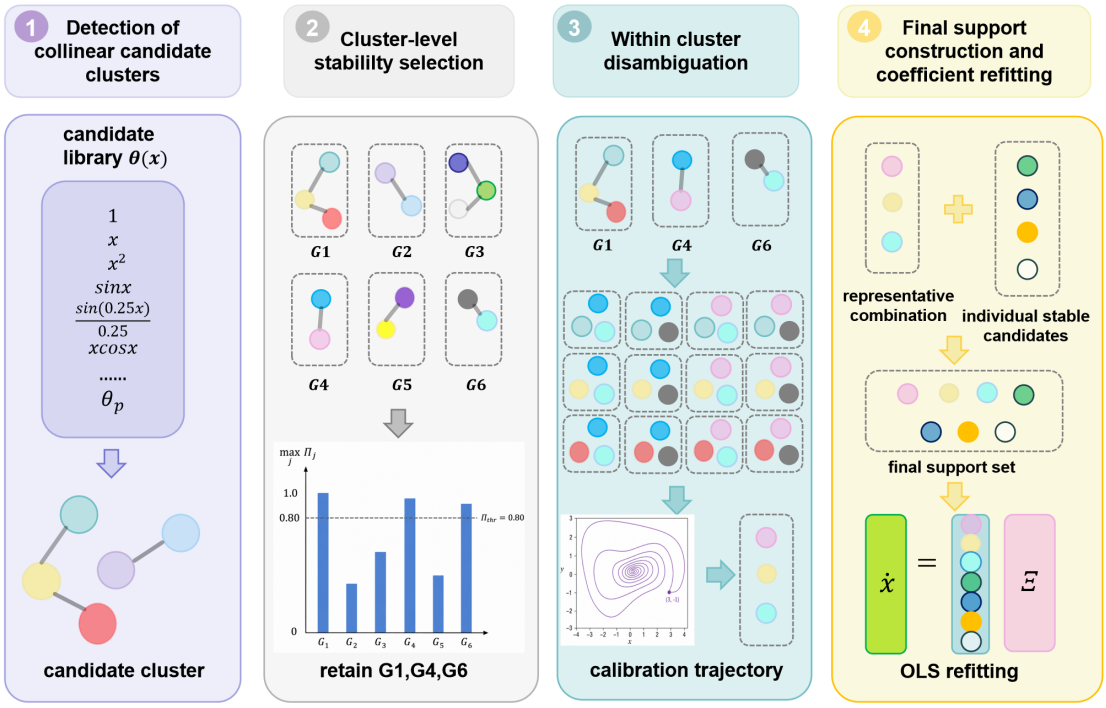


Fig. 1: Overview of the proposed collinearity-aware SINDy framework. The procedure consists of candidate clustering based on collinearity assessment, cluster-level stability selection, within-cluster representative disambiguation and final support construction with ordinary least-squares refitting.

Algorithm 1 Collinearity-aware SINDy with cluster stability and representative selection

Input:

Training data $(\mathbf{X}_{\text{tr}}, \dot{\mathbf{x}}_{\text{tr}})$, calibration data $(\mathbf{X}_{\text{cal}}, \dot{\mathbf{x}}_{\text{cal}})$ with N_{cal} samples and candidate library $\Theta(\cdot)$;
threshold set Λ, correlation threshold τ_c, stability cutoff π_{thr}, ridge parameter α_R and subsampling parameters (B, q).

1. Clustering

$\Theta_{\text{tr}} \leftarrow \Theta(\mathbf{X}_{\text{tr}})$, $\Theta_{\text{cal}} \leftarrow \Theta(\mathbf{X}_{\text{cal}})$.
Construct a correlation graph on nonconstant training library columns, connecting i and j when $C_{ij} \geq \tau_c$.
$\mathscr{P} \leftarrow$ connected components of the graph; assign constant or near-constant candidates to singleton clusters.

2. Cluster-level frequency selection

Standardize the training library and center the target derivative for STRidge fitting.
for all $(\lambda, b) \in \Lambda \times \{1, \ldots, B\}$ **do**
  $\hat{S}^{(b,\lambda)} \leftarrow$ support obtained by STRidge with parameters (λ, α_R) on a training subsample of size q.
end for
Compute cluster frequencies and retain stable clusters:

$$\Pi_G(\lambda) = \frac{1}{B}\sum_{b=1}^{B} \mathbf{1}\left\{\hat{S}^{(b,\lambda)} \cap G \neq \varnothing\right\}, \qquad \widehat{\mathscr{G}} = \left\{G \in \mathscr{P} : \max_{\lambda \in \Lambda} \Pi_G(\lambda) \geq \pi_{\text{thr}}\right\}.$$

3. Representative selection and final support construction

$\hat{S}_{\text{sing}} \leftarrow \{k : \{k\} \in \widehat{\mathscr{G}}\}$, $\widehat{\mathscr{G}}_{\text{multi}} \leftarrow \{G \in \widehat{\mathscr{G}} : |G| > 1\}$, $\mathscr{C} \leftarrow \prod_{G \in \widehat{\mathscr{G}}_{\text{multi}}} G$.
for all representative combinations $\mathbf{j} = (j_G)_{G \in \widehat{\mathscr{G}}_{\text{multi}}} \in \mathscr{C}$ **do**
  $\mathscr{M}(\mathbf{j}) \leftarrow \hat{S}_{\text{sing}} \cup \{j_G : G \in \widehat{\mathscr{G}}_{\text{multi}}\}$.
  $\hat{\xi}^{(\mathbf{j})} \leftarrow \text{OLS}\left(\Theta_{\mathscr{M}(\mathbf{j}),\text{tr}}, \dot{\mathbf{x}}_{\text{tr}}\right)$.
  $R_{\text{cal}}(\mathbf{j}) \leftarrow \left\|\dot{\mathbf{x}}_{\text{cal}} - \Theta_{\mathscr{M}(\mathbf{j}),\text{cal}}\hat{\xi}^{(\mathbf{j})}\right\|_2 / \sqrt{N_{\text{cal}}}$.
end for
$\mathbf{j}^* \leftarrow \arg\min_{\mathbf{j} \in \mathscr{C}} R_{\text{cal}}(\mathbf{j})$, $\hat{S}_{\text{final}} \leftarrow \mathscr{M}(\mathbf{j}^*)$.
$\hat{\xi}_{\text{final}} \leftarrow \text{OLS}\left(\Theta_{\hat{S}_{\text{final}},\text{tr}}, \dot{\mathbf{x}}_{\text{tr}}\right)$.
All OLS fits use the original, unstandardized libraries.

Output:

return $\hat{S}_{\text{final}}$, $\hat{\xi}_{\text{final}}$ and the identified state equation.

and the cluster is retained when

$$\Pi_G^{\max} \geq \pi_{\text{thr}}. \tag{12}$$

Thus, $\Pi_G^{\max}$ determines whether the dynamical component represented by cluster G is retained for subsequent representative selection.

2.5 Within-cluster disambiguation

Cluster-level selection determines whether a dynamical component is present, but does not identify its functional representative. Three representative-selection criteria are considered: stability-first, fit-first and simplicity-first. The proposed method adopts fit-first for final representative selection, while the three criteria are compared experimentally in Section 4.2.

Let $\hat{S}_{\text{sing}}$ denote the candidates belonging to retained singleton clusters. The set of retained multi-member clusters is defined as

$$\hat{\mathscr{G}}_{\text{multi}} = \left\{ G_\ell : \Pi_{G_\ell}^{\max} \geq \pi_{\text{thr}},\ |G_\ell| > 1 \right\}. \tag{13}$$

Thus, $\hat{\mathscr{G}}_{\text{multi}}$ contains all multi-member clusters that pass the cluster-level stability threshold. When several such clusters are retained, their representatives are selected jointly rather than sequentially. The possible representative combinations form the Cartesian product

$$\mathscr{C} = \prod_{G_\ell \in \hat{\mathscr{G}}_{\text{multi}}} G_\ell. \tag{14}$$

For a representative combination $\mathbf{j} = (j_\ell)_{G_\ell \in \hat{\mathscr{G}}_{\text{multi}}}$, the corresponding candidate support is

$$M(\mathbf{j}) = \hat{S}_{\text{sing}} \cup \left\{ j_\ell : G_\ell \in \hat{\mathscr{G}}_{\text{multi}} \right\}. \tag{15}$$

Under the fit-first criterion, coefficients are first estimated on the training trajectory:

$$\hat{\xi}(\mathbf{j}) = \arg\min_{\xi} \left\| \dot{\mathbf{x}}_{\text{tr}} - \Theta_{M(\mathbf{j}),\text{tr}} \xi \right\|_2^2. \tag{16}$$

The fitted model is then evaluated on an independent calibration trajectory:

$$R_{\text{cal}}(\mathbf{j}) = \sqrt{\frac{1}{N_{\text{cal}}} \left\| \dot{\mathbf{x}}_{\text{cal}} - \Theta_{M(\mathbf{j}),\text{cal}} \hat{\xi}(\mathbf{j}) \right\|_2^2}. \tag{17}$$

The representative combination with the minimum calibration error is selected:

$$\mathbf{j}^\star = \arg\min_{\mathbf{j} \in \mathscr{C}} R_{\text{cal}}(\mathbf{j}). \tag{18}$$

The resulting support defines $\hat{S}_{\text{final}}$. Once the support is fixed, the ridge-shrunk coefficients are discarded and OLS is applied to the original, unstandardized training library:

$$\hat{\xi}_{\text{final}} = \arg\min_{\xi} \left\| \dot{\mathbf{X}}_{\text{tr}} - \Theta_{\hat{S}_{\text{final}},\text{tr}} \xi \right\|_2^2. \tag{19}$$

Applying the same procedure independently to each state equation yields the full coefficient matrix $\hat{\Xi}_{\text{final}}$.

3 Experimental setup

3.1 Test systems and data generation

Two nonlinear systems are considered to evaluate the proposed method under controlled and naturally occurring collinearity. The van der Pol oscillator with $\mu = 1$ is given by

$$\dot{x} = y, \qquad \dot{y} = y - x^2 y - x \tag{20}$$

serves as the primary test system, where an artificial surrogate is added to the candidate library to induce controllable collinearity. A damped nonlinear pendulum as follows:

$$\Theta(\mathbf{X}) = [x_A,\ x_B,\ y,\ x^2,\ xy,\ y^2,\ x^2 y,\ \cos x], \tag{21}$$

with $x_A = x$, $x_B = \frac{\sin(0.25x)}{0.25}$ and true support $S^\star = \{x_A, y, x^2 y\}$ is used as a secondary system to assess the method under physically motivated, naturally occurring collinearity.

Consider that

$$\frac{\sin(0.25x)}{0.25} = x - \frac{0.25^2}{6} x^3 + \mathscr{O}(x^5), \tag{22}$$

x_A and x_B form a highly correlated but functionally distinct candidate pair. The collinearity strength is varied by replacing x_B with

$$x_B(a) = \frac{\sin(ax)}{a}, \qquad a \in \{0.25, 0.30, \ldots, 0.65\}. \tag{23}$$

The damped nonlinear pendulum is defined as

$$\dot{\theta} = \omega, \qquad \dot{\omega} = -0.2\omega - \sin\theta. \tag{24}$$

Both state equations are identified using

$$\Theta_{\text{pend}} = [1, \theta, \omega, \theta^2, \theta\omega, \omega^2, \theta^3, \sin\theta, \cos\theta]. \tag{25}$$

For small angular displacements, $\sin\theta \approx \theta$, making θ and $\sin\theta$ a naturally occurring correlated pair. For the van der Pol system, the training trajectory is generated from $[x(0), y(0)]^\mathsf{T} = [0.5, 0.5]^\mathsf{T}$ over $t \in [0, 25)$ and the calibration trajectory is generated from $[x(0), y(0)]^\mathsf{T} = [3, -1]^\mathsf{T}$ over $t \in [0, 15)$. The first and last five training samples are removed, leaving 2490 observations. Moreover, we require that the training and calibration trajectories for pendulum system are generated from $[\theta(0), \omega(0)]^\mathsf{T} = [0.4, 0]^\mathsf{T}$ and $[2, 0]^\mathsf{T}$, respectively. All training and calibration trajectories for both systems are sampled at $\Delta t = 0.01$. The calibration trajectories are used for parameter selection and within-cluster disambiguation.

3.2 Noise and numerical differentiation

The experiments include two complementary noise settings. In the controlled setting, Gaussian perturbations are added directly to the analytical derivatives to isolate structural-selection behavior from numerical differentiation. In the end-to-end setting, noise is added only to the state measurements and the derivatives are estimated numerically before equation identification [18–20].

For the end-to-end experiments, the state-noise levels are

$$\eta_x \in \{0.5\%, 1.0\%, 2.0\%, 3.0\%\}. \tag{26}$$

Centered finite differences and Savitzky–Golay differentiation are used to evaluate the sensitivity of end-to-end recovery to derivative-estimation quality [21]. For centered

finite differences, the candidate library is constructed from the noisy states. For Savitzky–Golay differentiation, the smoothed states are used to construct the library and the derivatives are obtained from the same local polynomial with window length 31 and polynomial degree 3. For the pendulum experiments, state-noise standard deviation is 4% of each clean-state standard deviation and derivatives are estimated using the same Savitzky–Golay procedure. All random processes are generated using NumPy with a common master seed

$$s_{\text{master}} = 20260625. \tag{27}$$

Within each paired experiment, the competing methods use identical noise realizations and, when subsampling is required, identical subsample indices.

3.3 Comparison methods and parameter settings

The proposed method is compared with two principal baselines: tuned STRidge and tuned feature-wise stability selection. Tuned STRidge evaluates the benefit over single-fit sparse regression, whereas feature-wise stability selection isolates the effect of replacing individual selection frequencies with cluster-level aggregation. Elastic Net and Group Lasso are included as additional baselines for correlated-variable selection.

Unless otherwise stated, the common parameters are

$$\alpha_R = 10^{-4}, \qquad \tau_c = 0.95, \qquad \pi_{\text{thr}} = 0.80, \tag{28}$$

where α_R is the ridge regularization coefficient, τ_c is the absolute-correlation threshold used to construct candidate clusters and π_{thr} is the cluster-level stability cutoff.

Each stability estimate is based on $B = 100$ subsamples of size $q = 100$, drawn without replacement from the training data. Within each trial, all stability-based methods use identical subsamples, which are also held fixed across sparsity thresholds. The STRidge threshold λ is evaluated over system-specific grids. For the van der Pol system, $\Lambda = \{0.6, 0.625, 0.65, 0.675, 0.7, 0.725, 0.75, 0.775, 0.8\}$, whereas the pendulum experiments use

$$\Lambda_{\text{pend}} = \{0.010, 0.015, \ldots, 0.090\}. \tag{29}$$

For tuned STRidge, λ is selected from the corresponding grid by minimizing the error on the independent calibration trajectory. For feature-wise stability selection, the cutoff π_{single} is selected by the same criterion from

$$\mathscr{P}_{\text{single}} = \{0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9\}. \tag{30}$$

The proposed method aggregates cluster-level selection frequencies over the full threshold grid, retains cluster G when $\Pi_G^{\max} \geq \pi_{\text{thr}}$ and applies fit-first disambiguation to each retained multi-member cluster.

Group Lasso uses the correlation-based clusters defined in Section 2.3 as predefined groups. For the disambiguated Elastic Net and Group Lasso variants, fit-first selection is applied to retained correlated candidates. All methods are evaluated using the same noisy data within each trial. After support selection, the coefficients are refitted by ordinary least squares on the original, unstandardized training library.

Table 1 summarizes the controlled van der Pol experiments. The derivative-noise level denotes the standard deviation of the Gaussian perturbation added to the analytical derivative.

3.4 Evaluation metrics and statistical analysis

The subsequent experiments evaluate support recovery primarily by exact recovery, with the true positive rate and false positive rate used to quantify missed and redundant selections. Let $\hat{S}$ and $S^\star$ denote the identified and true supports and let $\mathscr{J} = \{1, \ldots, p\}$ be the index set of all candidate functions. Exact recovery is achieved when $\hat{S} = S^\star$. The true positive rate and false positive rate are defined as

$$\text{TPR} = 1 - \frac{|S^\star \setminus \hat{S}|}{|S^\star|}, \qquad \text{FPR} = \frac{|\hat{S} \setminus S^\star|}{|\mathscr{J} \setminus S^\star|}. \tag{31}$$

Exact-recovery rates are summarized using Wilson 95% confidence intervals. Differences between methods are reported with paired-bootstrap 95% confidence intervals. Paired recovery outcomes are compared using two-sided exact McNemar tests, with Holm adjustment across the six noise levels.

Additional experiment-specific statistics are reported where appropriate. Representative-selection behavior is evaluated by comparing the outcomes of the fit-first and stability-first representative rules over valid trials. For comparisons involving correlated-variable baselines, cluster-level exact recovery and redundant selections are additionally reported.

4 Results

4.1 Repeatability of split voting

The split-vote experiment was designed as a mechanism test of feature-wise stability under severe candidate collinearity, rather than as a complete performance comparison. The derivative perturbation level was set to 0.20 and 300 independent outer trials were conducted. In each trial, the state trajectory and derivative perturbations were regenerated and 100 subsamples were used to compute stability frequencies. The mean absolute correlation between x_A and x_B was 0.999939, indicating that the two candidates were nearly indistinguishable over the training trajectory [22].

Fig. 2 shows a clear separation between feature-wise and cluster-level stability evidence. Across $\lambda \in [0.60, 0.80]$, the mean feature-wise frequencies of both x_A and x_B remain below the stability cutoff of 0.80, whereas the cluster frequency of $\{x_A, x_B\}$ remains above the cutoff, ranging from 0.928 to 1.000.

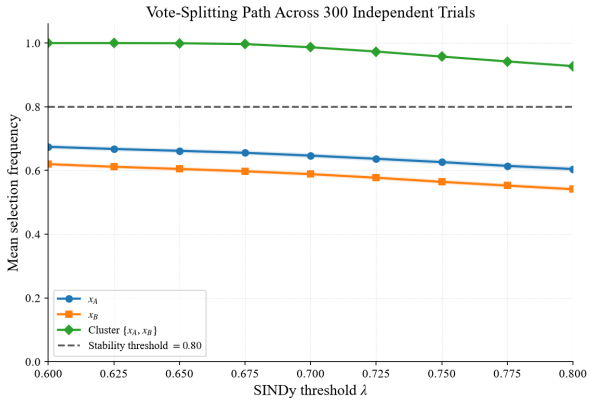


Fig. 2: Mean stability paths for the near-collinear pair over 300 independent trials.

A strict split-vote event occurs in 98.3% of trials, with a Wilson 95% confidence interval of [96.2%, 99.3%]. These

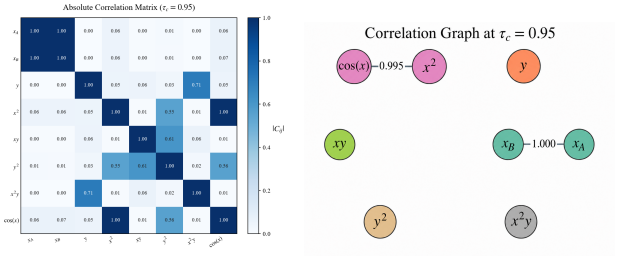
Table 1: Settings for the controlled van der Pol experiments

Task	Derivative-noise SD	Trials	Task-specific setting
Split-vote repeatability	0.20	300	$\lambda = 0.70$ for the representative result and $\lambda \in \Lambda$ for the stability path
Cluster detection and repair	0.20	1	Correlation graph constructed with $\tau_c = 0.95$
Disambiguation comparison	$\{0.05, 0.10, 0.20, 0.30, 0.40, 0.50\}$	300 per level	Stability-first, fit-first and simplicity-first
Noise robustness	$\{0.05, 0.10, 0.20, 0.30, 0.40, 0.50\}$	300 per level	Paired comparison of the principal methods
Collinearity strength	0.30	300 per value	$a \in \{0.25, 0.30, \ldots, 0.65\}$
Correlated-variable baselines	0.30	300	Elastic Net, Group Lasso and their disambiguated variants

results indicate that the underlying dynamical component is retained in almost every fit, but its specific functional representation changes across subsamples. Thus, a low feature-wise frequency can arise from competition within a highly correlated cluster rather than from the disappearance of the associated dynamical component.

4.2 Cluster detection and cluster-level recovery

At the correlation threshold $\tau_c = 0.95$, the absolute-correlation matrix and the corresponding graph identify two non-singleton correlated clusters: $G_1 = \{x_A, x_B\}$ and $G_2 = \{x^2, \cos x\}$ (see Fig. 3). The first cluster contains the true restoring term and its surrogate, whereas the second cluster is correlated but inactive in the governing equation. This contrast shows that correlation analysis alone identifies candidate redundancy, but not dynamical relevance.



(a) Absolute-correlation matrix (b) Correlation graph and detected clusters

Fig. 3: Correlation structure of the van der Pol candidate library.

The subsequent stability evaluation separates these two cases. Feature-wise stability selection retains y and x^2y, but neither x_A nor x_B exceeds the stability cutoff. In contrast, cluster-level stability retains G_1, y and x^2y, while rejecting the inactive cluster G_2. Thus, the cluster statistic recovers the omitted restoring component without automatically retaining every detected correlated cluster.

4.3 Within-cluster disambiguation and refitting

After cluster-level stability retained $G_1 = \{x_A, x_B\}$, the remaining task was to select a single representative from the retained correlated cluster. Three representative-selection rules were compared: stability priority, fit priority and simplicity priority. As shown in Fig. 4, fit priority yielded consistently higher point estimates of representative-selection accuracy than stability priority across the controlled derivative-noise levels.

Although simplicity priority achieves perfect accuracy in this benchmark, the result is predetermined by its construction, in which the true term $x_A = x$ is also the simpler candidate. It is therefore treated as a heuristic reference rather than

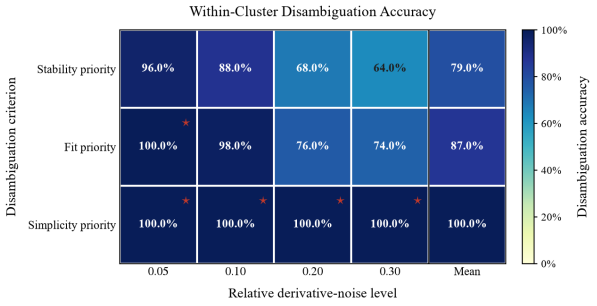


Fig. 4: Within-cluster disambiguation accuracy of different representative-selection criteria under derivative perturbations.

adopted as a generally applicable representative-selection rule. The paired outcomes in Fig. 5 further illustrate how the two data-driven criteria behave on the same valid trials, while Table 2 reports the corresponding statistical comparisons. For each derivative-noise level, 300 outer trials were conducted using the same noisy data and subsample indices for stability priority and fit priority. Only trials in which G_1 passed the cluster-level stability cutoff were included.

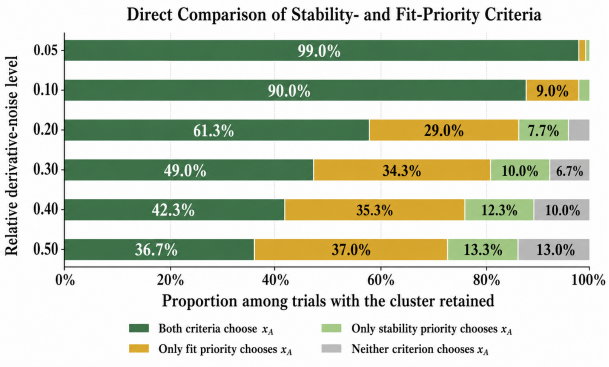


Fig. 5: Paired outcomes of stability-priority and fit-priority representative selection.

When derivative perturbations were weak, the two criteria usually agreed. Both selected the true representative x_A in 99.0% and 90.0% of valid trials at noise levels 0.05 and 0.10, respectively. As reported in Table 2, the paired difference at noise 0.05 was only 0.3 percentage points, and its confidence interval included zero. Accordingly, the exact McNemar test did not indicate a statistically supported difference at this level (Holm-adjusted $p = 1.000$).

From noise level 0.10 onward, trials in which only fit priority selected x_A consistently outnumbered those in which only stability priority did so. The paired improvements ranged from 8.0 to 24.3 percentage points, and all remained statistically significant after Holm adjustment (all adjusted

Table 2: Paired statistical comparison of fit-priority and stability-priority representative selection

Noise	Valid trials n_{valid}	Fit only b	Stability only c	Fit–stability 95% CI (pp)	Exact McNemar p	Holm-adjusted p
0.05	300	2	1	0.3 [−0.7, 1.7]	1.000	1.000
0.10	300	27	3	8.0 [4.7, 11.7]	8.430×10^{-6}	1.686×10^{-5}
0.20	300	87	23	21.3 [15.0, 27.7]	6.041×10^{-10}	3.021×10^{-9}
0.30	300	103	30	24.3 [17.3, 31.3]	1.452×10^{-10}	8.711×10^{-10}
0.40	300	106	37	23.0 [15.7, 30.3]	6.635×10^{-9}	2.589×10^{-8}
0.50	300	111	40	23.7 [16.0, 31.3]	6.474×10^{-9}	2.589×10^{-8}

$p \leq 1.686 \times 10^{-5}$). These results provide statistical evidence that fit priority is more reliable than stability priority under non-negligible derivative perturbations.

Moreover, Fig. 5 also shows that the proportion of trials in which neither criterion selected x_A increased from 2.0% at noise 0.20 to 13.0% at noise 0.50. Thus, independent-trajectory validation improves representative selection but does not eliminate its dependence on derivative-estimation quality.

This observed difference is consistent with the information used by the two rules. Stability priority relies on small differences between the repeated selection frequencies of two nearly interchangeable candidates. Fit priority instead tests the competing representations on an independent calibration trajectory, where the broader state-space coverage makes the surrogate x_B less interchangeable with the true term x_A. Fit priority is therefore used as the final representative-selection rule, while stability priority was retained as supporting evidence of structural recurrence. After x_A was selected, final OLS refitting on the fixed support recovered

$$\hat{y} = 0.9998\,y - 0.9997\,x^2y - 1.0001\,x_A, \tag{32}$$

in the representative detailed run, closely matching the true equation. This result illustrates the three decisions in the proposed workflow: candidate correlation defines the competing set, cluster-level stability determines component retention and within-cluster disambiguation selects the representative term before final coefficient refitting.

4.4 Robustness to controlled derivative perturbations

Derivative-noise level was varied over $\{0.05, 0.10, 0.20, 0.30, 0.40, 0.50\}$. At each level, 300 paired trials were conducted. Fig. 6 and Table 3 summarize exact recovery. At noise 0.05, all methods were close to the ceiling and the paired differences were not statistically supported. From 0.10 onward, the proposed method showed positive paired differences relative to both principal baselines. At noise 0.30, exact recovery was 85.0% for the proposed method, 67.7% for tuned STRidge and 35.3% for tuned feature-wise stability selection.

The feature-wise baseline did not fail only because of low sensitivity. At noise 0.30, its mean TPR was 96.8%, but its FPR was 12.9%. The proposed method achieved a TPR of 95.0% with an FPR of 3.0%, while tuned STRidge achieved 90.9% and 6.5%, respectively. Tuning a lower feature cutoff recovered dispersed true selections but admitted more false terms and redundant correlated members. Cluster-level aggregation produced a more favorable balance between retaining the true direction and suppressing false candidates.

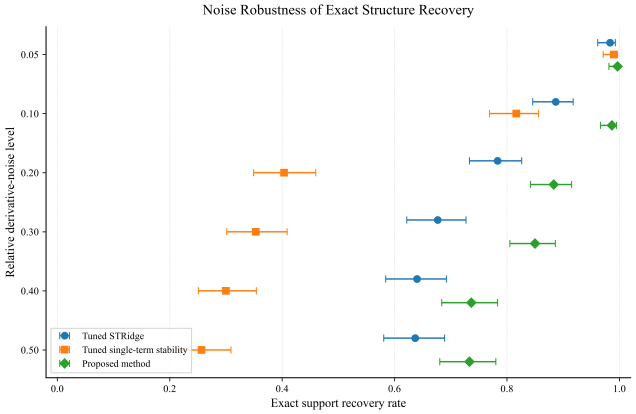


Fig. 6: Exact support-recovery rates under controlled derivative perturbations. Error bars indicate Wilson 95% confidence intervals.

4.5 Effect of collinearity strength

To determine where the method is useful, the surrogate was generalized to

$$x_B(a) = \frac{\sin(ax)}{a}, \qquad a \in \{0.25, 0.30, \cdots, 0.65\}, \tag{33}$$

while derivative noise was fixed at 0.30. Smaller a produced higher correlation with $x_A = x$. The same random perturbations and subsample indices were reused across values of a within each trial.

As is shown in Fig.7, at mean correlation 0.99994, the proposed method achieved 83.3% exact recovery, compared with 69.0% for tuned STRidge and 36.3% for feature-wise stability. The paired improvement over STRidge was 14.3 percentage points with a 95% interval of [7.7, 21.0] and $p = 6.445 \times 10^{-5}$. As correlation decreased, the improvement over STRidge contracted: 5.3 points at correlation 0.999581, 0.7 points at 0.998932 and zero at 0.998396 and below. All methods reached 100.0% exact recovery at correlation 0.996680.

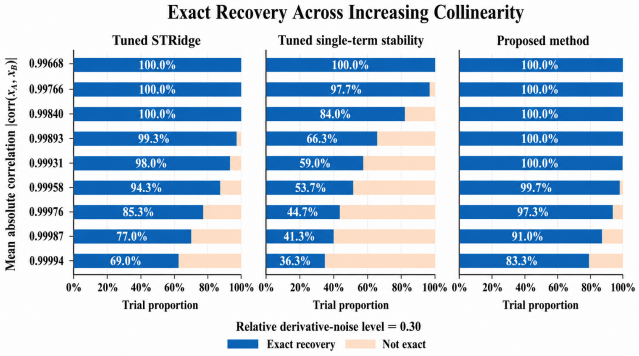


Fig. 7: Exact recovery as the mean absolute correlation between x_A and $x_B(a)$ varies.

Table 3: Exact recovery and paired improvements under controlled derivative perturbations

Noise	STRidge	Feature stability	Proposed	Proposed–STRidge 95% CI	p	Proposed–feature stability
0.05	98.3%	99.0%	99.7%	1.3 [0.0, 3.0] pp	0.2188	0.7 pp
0.10	88.7%	81.7%	98.7%	10.0 [6.3, 13.7] pp	2.272×10^{-7}	17.0 pp
0.20	78.3%	40.3%	88.3%	10.0 [5.0, 15.3] pp	2.871×10^{-4}	48.0 pp
0.30	67.7%	35.3%	85.0%	17.3 [11.0, 24.3] pp	9.323×10^{-7}	49.7 pp
0.40	64.0%	30.0%	73.7%	9.7 [3.0, 16.7] pp	0.008735	43.7 pp
0.50	63.7%	25.7%	73.3%	9.7 [2.3, 17.0] pp	0.01338	47.7 pp

The proposed procedure should therefore be interpreted as a correction for the severe-collinearity regime. It does not provide a universal gain when candidates are already distinguishable. This result also supports retaining the correlation detector rather than applying cluster-based post-processing indiscriminately.

4.6 Comparison with correlated-variable baselines

The derivative-noise level was fixed at 0.30 to compare tuned STRidge, feature-wise stability selection, Elastic Net, Group Lasso, their disambiguated variants and the proposed method (see Fig. 8 and Table 4). Both exact recovery of the analytical support, exact recovery at the cluster level and redundant retention of both x_A and x_B are evaluated.

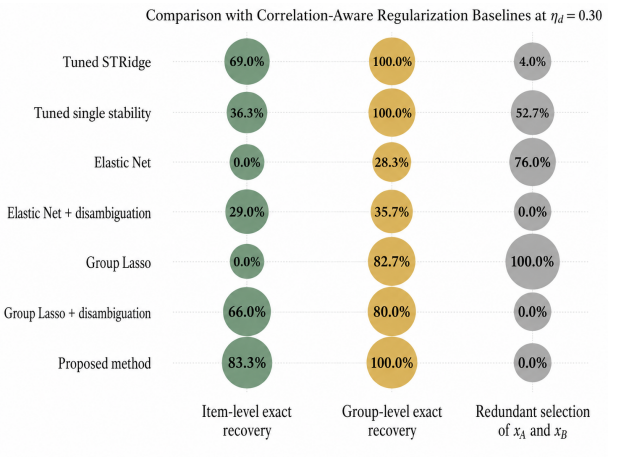


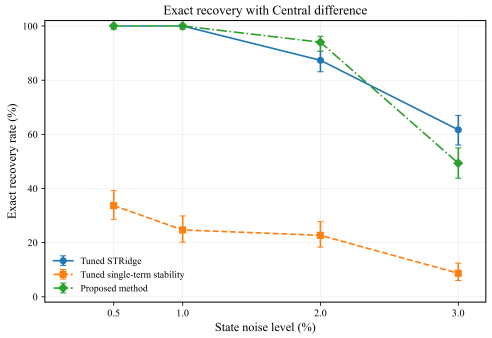
Fig. 8: Structural-recovery results for sparse-selection and correlated-regularization methods.

Original Group Lasso retained the true group in many trials but necessarily kept both members of the selected group, so feature-level exact recovery was zero and redundant selection was 100.0%. Applying representative disambiguation removed the redundancy and increased feature-level exact recovery to 66.0%, but it could not repair trials in which the front-end method failed to retain the correct cluster. Elastic Net similarly retained multiple correlated candidates in 76.0% of trials and post-processing could not compensate for its lower cluster-level recovery. And the proposed method combined 100.0% cluster-level recovery with zero redundant selection and the highest feature-level exact recovery. The comparison shows that cluster retention and within-cluster representative recovery are separate requirements.

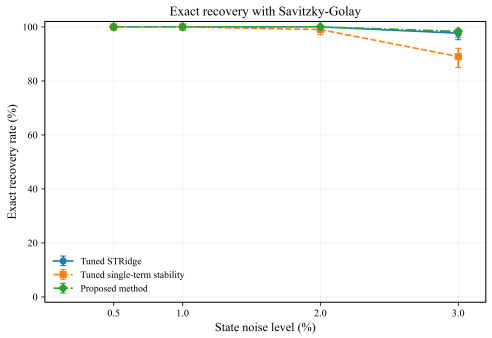
4.7 End-to-end numerical differentiation

Controlled derivative perturbations allow the structural-selection mechanism to be studied independently. However, practical equation discovery requires estimating derivatives from noisy measurements. Therefore, additional experiments were performed by adding noise only to the state variables with noise levels of $0.5\%, 1.0\%, 2.0\%, 3.0\%$. The derivatives were obtained using either centered finite differences or Savitzky–Golay smoothing and differentiation with a window length of 31 and a polynomial order of 3. Analytical derivatives were not used during model selection.

The results in Fig. 9 and Table 5 show that derivative-estimation quality strongly affects sparse structure recovery. With centered finite differences, the proposed method achieves 100.0%, 100.0% and 94.0% exact recovery at 0.5%, 1.0% and 2.0% state noise, respectively. At 2.0%, it outperforms STRidge by 6.67 percentage points and feature-wise stability by 71.33 percentage points. However, at 3.0% state noise, the proposed method decreases to 49.3%, while STRidge reaches 61.7%. This indicates that centered finite differences can amplify state noise enough to make the stability frequencies used for cluster selection unreliable. Therefore, the proposed method should not be viewed as a substitute for accurate derivative estimation.



(a) Centered finite differences



(b) Savitzky–Golay differentiation

Fig. 9: Exact recovery under numerical differentiation of noisy state measurements.

Savitzky–Golay preprocessing gives more stable results. At 3.0% state noise, the proposed method achieves 98.3% exact recovery, compared with 97.7% for STRidge and 89.0% for feature-wise stability. The difference between the proposed method and STRidge is not statistically sup-

Table 4: Comparison with correlated-variable baselines at derivative noise 0.30

Method	Feature-level exact recovery	Cluster-level exact recovery	Redundant x_A–x_B selection
Tuned STRidge	69.0%	100.0%	4.0%
Tuned feature-wise stability	36.3%	100.0%	52.7%
Elastic Net	0.0%	28.3%	76.0%
Elastic Net + disambiguation	29.0%	35.7%	0.0%
Group Lasso	0.0%	82.7%	100.0%
Group Lasso + disambiguation	66.0%	80.0%	0.0%
Proposed method	83.3%	100.0%	0.0%

Table 5: Exact recovery rates under numerical differentiation

Differentiation	State noise	STRidge	Feature-wise stability	Proposed
Centered difference	0.5%	100.0%	33.7%	100.0%
Centered difference	1.0%	100.0%	24.7%	100.0%
Centered difference	2.0%	87.3%	22.7%	94.0%
Centered difference	3.0%	61.7%	8.7%	49.3%
Savitzky–Golay	0.5%	100.0%	100.0%	100.0%
Savitzky–Golay	1.0%	100.0%	100.0%	100.0%
Savitzky–Golay	2.0%	100.0%	99.0%	100.0%
Savitzky–Golay	3.0%	97.7%	89.0%	98.3%

ported, whereas the improvement over feature-wise stability remains significant. Overall, these results suggest that the proposed framework mainly improves feature-wise stability under collinearity, while its advantage over a tuned STRidge baseline depends on the quality of numerical differentiation.

4.8 Natural collinearity in a damped pendulum

The damped-pendulum experiment examines whether the proposed method can resolve naturally occurring collinearity between the true nonlinear term $\sin\theta$ and its small-angle approximation θ. Unlike the constructed surrogate in the van der Pol experiment, this collinearity arises from a standard mixed candidate library and a restricted-angle training trajectory.

The restricted-angle training data produced a highly correlated pair $\{\theta, \sin\theta\}$, and the corresponding natural cluster was detected consistently across trials. This setting is challenging because θ can approximate $\sin\theta$ well over the training range, even though it is not the correct analytical term in the governing equation. The task therefore tests whether a method can recover the exact support $\{\omega, \sin\theta\}$ in the $\dot{\omega}$ equation, rather than merely fit the local training trajectory. To separate threshold-selection effects from structural ambiguity, an Oracle STRidge baseline was also included. This baseline uses knowledge of the true support to select the best attainable threshold along the same STRidge threshold path.

As shown in Table 6, all methods except tuned feature-wise stability recovered the kinematic equation $\dot{\theta} = \omega$ in every trial. The main difficulty lay in the second equation, where the exact nonlinear restoring term had to be distinguished from its small-angle surrogate. Tuned STRidge achieved only 2.7% exact recovery for $\dot{\omega}$ and Oracle STRidge achieved the same rate. This agreement indicates that the low STRidge recovery was not primarily caused by validation-based threshold selection. Rather, in most trials, the exact support did not appear along the prescribed STRidge threshold path under near-collinearity.

Tuned feature-wise stability improved the recovery of $\dot{\omega}$ to 34.7%, but its system-level exact recovery remained 32.0%. This result shows that repeated-selection evidence can partially mitigate threshold sensitivity, but feature-wise aggregation still suffers when evidence is distributed between θ and $\sin\theta$. In contrast, the proposed method achieved 100.0% exact recovery for both equations and therefore for the full system.

The improvement comes from separating two decisions that are coupled in feature-wise selection. Cluster-level stability first retains the correlated direction represented by $\{\theta, \sin\theta\}$ and independent-trajectory disambiguation then selects $\sin\theta$ as the representative term. This experiment therefore extends the evidence beyond the deliberately constructed van der Pol surrogate and shows that the same mechanism appears in a standard physical system with a nonpolynomial true function, noisy state measurements and numerical differentiation. It also illustrates why a simplicity-first rule is not generally reliable: the simpler candidate θ is locally accurate in the restricted-angle regime, but the correct restoring term is $\sin\theta$.

5 Conclusion

A collinearity-aware SINDy framework was developed by combining correlation-based candidate clustering, cluster-level stability selection, strict split-vote diagnostics, independent-trajectory representative selection and final OLS refitting. The method repaired repeatable vote splitting in the van der Pol system, improved exact recovery under moderate derivative perturbations, reduced redundant selection relative to correlated-variable baselines and recovered the correct damped-pendulum equations under natural small-angle collinearity. Future work should evaluate real measurements, incomplete libraries, irregular and non-Gaussian data, clusters containing multiple true terms, alternative clustering and threshold-aggregation rules and scalable representative-search algorithms.

References

[1] Brunton, S. L., Proctor, J. L., & Kutz, J. N. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, 113(15), 3932–3937, 2016.

[2] Tibshirani, R. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B*, 58(1), 267–288, 1996.

Table 6: Single-equation and system-level exact recovery for the damped nonlinear pendulum

Method	Exact recovery of $\dot{\theta}$	Exact recovery of $\dot{\omega}$	System-level exact recovery
Tuned STRidge	100.0%	2.7%	2.7%
Oracle STRidge	100.0%	2.7%	2.7%
Tuned feature-wise stability	90.7%	34.7%	32.0%
Proposed method	100.0%	100.0%	100.0%

[3] Rudy, S. H., Brunton, S. L., Proctor, J. L., & Kutz, J. N. Data-driven discovery of partial differential equations. *Science Advances*, 3(4), e1602614, 2017.

[4] Zhang, L., & Schaeffer, H. On the convergence of the SINDy algorithm. *Multiscale Modeling & Simulation*, 17(3), 948–972, 2019.

[5] Fasel, U., Kutz, J. N., Brunton, B. W., & Brunton, S. L. Ensemble-SINDy: Robust sparse model discovery in the low-data, high-noise limit, with active learning and control. *Proceedings of the Royal Society A*, 478(2260), 20210904, 2022.

[6] Jiang, Y., Du, J., Yang, Y., & Deng, Z. Regularized least absolute deviation-based sparse identification of dynamical systems. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 33(1), 013103, 2023.

[7] Hao, Z., Yang, H., & Huang, Q. A robust sparse identification method for nonlinear dynamic systems affected by non-stationary noise. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 33(8), 083119, 2023.

[8] Zhang, S., Liu, Y., Wang, J., & Lu, J. Adaptive integral alternating minimization method for robust learning of nonlinear dynamical systems from highly corrupted data. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 33(12), 123112, 2023.

[9] Meinshausen, N., & Bühlmann, P. Stability selection. *Journal of the Royal Statistical Society: Series B*, 72(4), 417–473, 2010.

[10] Maddu, S., Cheeseman, B. L., Sbalzarini, I. F., & Müller, C. L. Stability selection enables robust learning of differential equations from limited noisy data. *Proceedings of the Royal Society A*, 478(2262), 20210916, 2022.

[11] Nouraie, M., Smith, C., & Muller, S. Stability selection via variable decorrelation. *Statistics and Computing*, 36, 160, 2026.

[12] Bühlmann, P., Rütimann, P., van de Geer, S., & Zhang, C.-H. Correlated variables in regression: Clustering and sparse estimation. *Journal of Statistical Planning and Inference*, 143(11), 1835–1858, 2013.

[13] Liu, R., & Sopasakis, A. Enhancing sparse identification of nonlinear dynamics with Earth-Mover distance and group similarity. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 35(3), 033139, 2025.

[14] Zou, H., & Hastie, T. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B*, 67(2), 301–320, 2005.

[15] Yuan, M., & Lin, Y. Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B*, 68(1), 49–67, 2006.

[16] Simon, N., Friedman, J., Hastie, T., & Tibshirani, R. A sparse-group lasso. *Journal of Computational and Graphical Statistics*, 22(2), 231–245, 2013.

[17] Alexander, D. H., & Lange, K. Stability selection for genome-wide association. *Genetic Epidemiology*, 35(7), 722–728, 2011.

[18] Tang, Y., Liao, W., Kuske, R., & Kang, S.-H. WeakIdent: Weak formulation for identifying differential equations using narrow-fit and trimming. *Journal of Computational Physics*, 483, 112069, 2023.

[19] Messenger, D. A., & Bortz, D. M. Asymptotic consistency of the WSINDy algorithm in the limit of continuum data. *IMA Journal of Numerical Analysis*, 45(6), 3264–3312, 2025.

[20] Forootani, A., Goyal, P., & Benner, P. A robust sparse identification of nonlinear dynamics approach by combining neural networks and an integral form. *Engineering Applications of Artificial Intelligence*, 149, 110360, 2025.

[21] Savitzky, A., & Golay, M. J. E. Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, 36(8), 1627–1639, 1964.

[22] Prokop, A., Frolov, N., & Gelens, L. Enhancing model identification with SINDy via nullcline reconstruction. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 34(6), 063135, 2024.

[23] Bertsimas, D., & Gurnee, W. Learning sparse nonlinear dynamics via mixed-integer optimization. *Nonlinear Dynamics*, 111(7), 6585–6604, 2023.

[24] Goyal, P., Pontes Duff, I., & Benner, P. Guaranteed stable quadratic models and their applications in SINDy and operator inference. *Physica D: Nonlinear Phenomena*, 483, 134893, 2025.

[25] Fung, S. W., Fasel, U., & Juniper, M. P. Rapid Bayesian identification of sparse nonlinear dynamics from scarce and noisy data. *Proceedings of the Royal Society A*, 481(2307), 20240200, 2025.

[26] Zheng, H., & Lin, G. LES-SINDy: Laplace-enhanced sparse identification of nonlinear dynamical systems. *Journal of Computational Physics*, 546, 114443, 2026.

[27] López, C., Naranjo, Á., Salazar, D., & Moore, K. J. Weak-form modified sparse identification of nonlinear dynamics. *Journal of Computational Physics*, 544, 114410, 2026.

[28] Champion, K., Lusch, B., Kutz, J. N., & Brunton, S. L. Data-driven discovery of coordinates and governing equations. *Proceedings of the National Academy of Sciences*, 116(45), 22445–22451, 2019.

[29] Kaheman, K., Kutz, J. N., & Brunton, S. L. SINDy-PI: A robust algorithm for parallel implicit sparse identification of nonlinear dynamics. *Proceedings of the Royal Society A*, 476(2242), 20200279, 2020.

[30] Lathourakis, C., & Cicirello, A. Physics enhanced sparse identification of dynamical systems with discontinuous nonlinearities. *Nonlinear Dynamics*, 112, 11237–11264, 2024.

[31] Yonezawa, A., Yonezawa, H., Yahagi, S., Kajiwara, I., Kijimoto, S., Taniuchi, H., & Murakami, K. Sparse identification of nonlinear dynamics with library optimization mechanism: Recursive long-term prediction perspective. *IEEE Transactions on Cybernetics*, 56(5), 2475–2488, 2026.