

Physics-informed discovery of non-smooth dynamical systems from trajectory data

Davide Murari

dm2011@cam.ac.uk

DAMTP, University of Cambridge
Phi-ML meets Engineering seminar series
The Alan Turing Institute

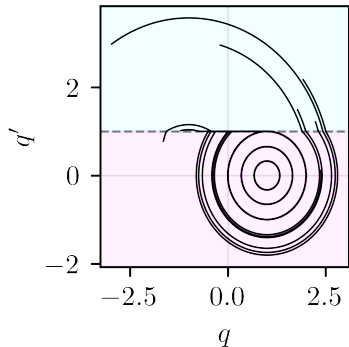
with Erik Jansson, Chris Budd, and Carola-Bibiane Schönlieb

24 September 2026

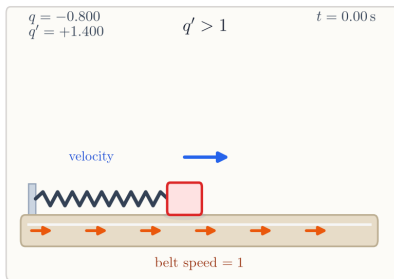


A motivating example

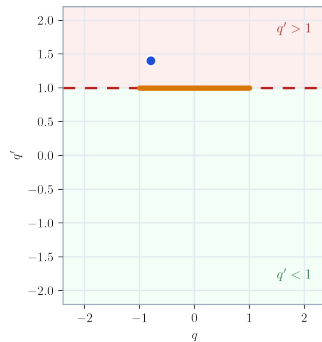
$$q''(t) + q(t) = \text{sign}(1 - q'(t)).$$



Block-spring on a moving belt



Phase portrait



Two coupled learning problems

1. Geometry

Infer the switching hyperplanes from noisy trajectory observations.

$$\text{Example: } \Sigma := \left\{ (q, p) \in \mathbb{R}^2 \mid p = 1 \right\}.$$

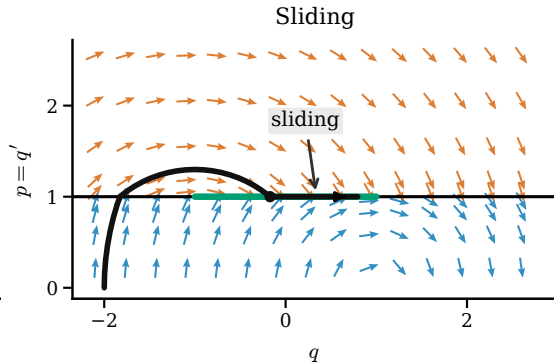
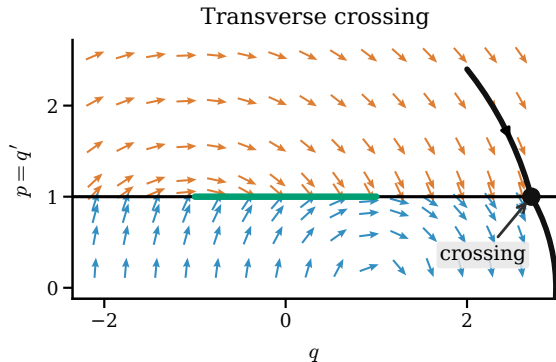
2. Dynamics

Learn the smooth vector field within the regions separated by the recovered geometry.

$$\text{Example: } q' = p, \quad p' = -q + 1 \quad (p < 1), \quad q' = p, \quad p' = -q - 1 \quad (p > 1)$$

Learn the geometry first. Preserve it while learning the dynamics.

Behaviour at the switching hyperplane



Piecewise-smooth vector fields

$$\dot{x} = f(x, t) = f_k(x, t), \quad x \in \mathcal{P}_k,$$

where the polyhedral regions $\mathcal{P}_k$ are induced by affine switching hyperplanes.

Inside each region

f_k is continuously differentiable.

Across an interface

The limiting vector fields may disagree.

We interpret solutions in the Filippov sense.

The Filippov sliding vector field

For a switch $\Sigma = \{x : a^\top x + b = 0\}$, let

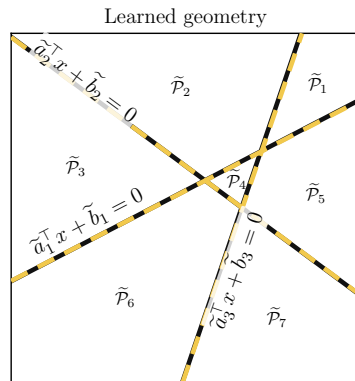
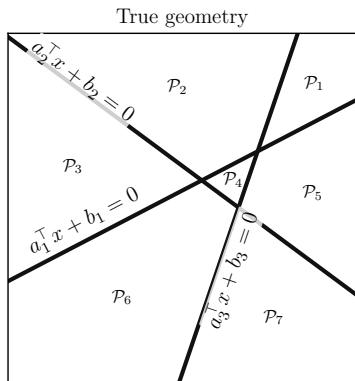
$$\nu^+ = a^\top f^+(x, t), \quad \nu^- = a^\top f^-(x, t).$$

On an attracting sliding region, $\nu^+ < 0 < \nu^-$ and

$$f_\Sigma = \alpha f^+ + (1 - \alpha) f^-, \quad \alpha = \frac{\nu^-}{\nu^- - \nu^+}, \quad a^\top f_\Sigma = 0.$$

The learned one-sided dynamics determine motion on the switch as well as away from it.

The geometric setup



— True lines

- - Learned lines

Learning the switching geometry

Derivative jumps reveal the switch

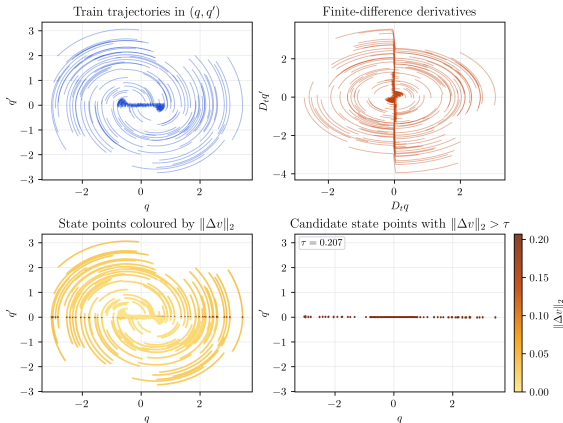
From noisy states $y_j = x(t_j) + \eta_j$,
estimate velocities

$$\hat{v}_j = \sum_{k=-r}^r w_k y_{j+k}.$$

Score consecutive changes:

$$J(j) = \|\hat{v}_{j+1} - \hat{v}_j\|_2.$$

**RANSAC: Fit the
highest-jump points.**



From jump scores to a candidate cloud

1. Rank all derivative-jump intervals globally.
2. Retain the largest fraction ρ .
3. Add neighbouring intervals to accommodate the differentiation stencil.
4. Use both interval endpoints as candidate points.

The cloud is overcomplete: it contains switch points and outliers.

Why robust consensus fitting?

Candidate outliers arise from

- ▶ smooth high-curvature motion,
- ▶ observation noise, and
- ▶ derivative-estimation artefacts.

RANSAC

Find the affine proposal with the largest coherent support.

Total least squares

Refine the winning hyperplane using only its inliers.

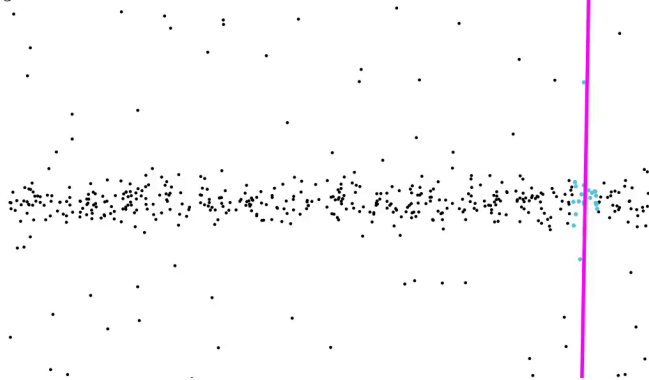
RANSAC searches for the largest coherent support

iteration 1/20

$$\ell : y = +47.5x - 208.3$$

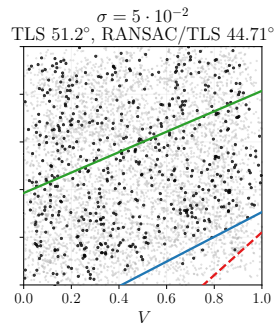
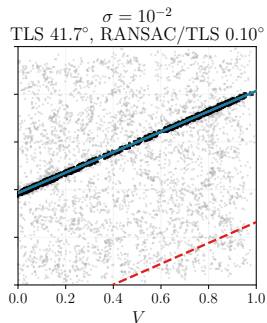
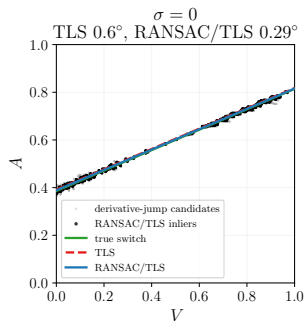
Score: 24

Best score: 24



1. **Sample** a minimal subset of candidates (two points in $\mathbb{R}^2$).
2. Fit a proposal ℓ .
3. **Count** all points within the distance threshold.
4. Keep the best support and refit its inliers by TLS.

Three geometry-recovery regimes



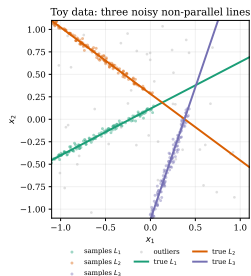
Geometry learning returns a fixed arrangement

$$\underbrace{A_P \in \mathbb{R}^{d_P \times d}}_{\text{row } i = a_i^\top},$$

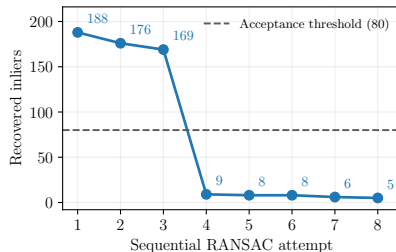
$$\underbrace{b_P \in \mathbb{R}^{d_P}}_{\text{offsets}}$$

$$\Sigma_i = \{x : a_i^\top x = b_i\}.$$

A three-line arrangement



If d_P is unknown: sequential RANSAC



Fit $\rightarrow$ remove inliers $\rightarrow$ repeat.

Three supports exceed the threshold, so $\hat{d}_P = 3$.

Encode the partition through its signs

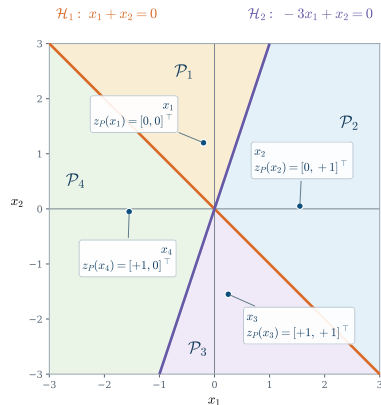
Recovered hyperplanes define

$$z_P(x) = \mathcal{H}(b_P - A_P x) \in \{0, 1\}^{d_P}.$$

The switching set is

$$\Gamma_P := \bigcup_{i=1}^{d_P} \{x \in \mathbb{R}^d : (b_P - A_P x)_i = 0\}.$$

- ▶ z_P is constant within each cell of $\mathbb{R}^d \setminus \Gamma_P$.
- ▶ One shared model can distinguish all cells.



Learning geometry-preserving dynamics

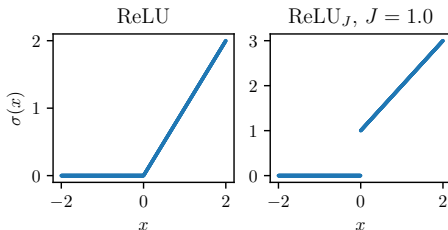
Two network types with distinct roles

We use the ReLU and the jump-ReLU activation¹, applied componentwise to vector inputs:

$$\text{ReLU}(x) = \max\{0, x\}, \quad \text{ReLU}_J(x) = \text{ReLU}(x) + J \odot \mathcal{H}(x), \quad \mathcal{H}(x) = \mathbf{1}_{\{x \geq 0\}}.$$

Φ : jump-ReLU network

Ψ : ordinary ReLU network



[1] F. Della Santa and S. Pieraccini, “Discontinuous neural networks and discontinuity learning,” *J. Comput. Appl. Math.* 419 (2023), 114678.

A geometry-preserving discontinuous network

$$\mathcal{N}(x, t) = \Psi(x, t, \Phi(z_P(x))) .$$

Geometry

$z_P(x)$ supplies the fixed regime code.

Discontinuous features

Φ maps the code to region-dependent features.

Within-region dynamics

Ψ learns the within-region vector fields.

The only discontinuous input is the prescribed switching code.

Guarantee 1: no new jumps

If x_0 is away from the prescribed hyperplanes, z_P is locally constant near x_0 . Therefore

$$\mathcal{N}(x, t) = \Psi(x, t, \Phi(z_P(x)))$$

is continuous near (x_0, t) .

$\text{disc}(\mathcal{N})$ is contained in the prescribed hyperplane arrangement Γ_P .

Guarantee 2: sufficient expressivity

Let F be smooth within every region of a fixed polyhedral partition. On a compact state domain Ω and a time interval $[0, T]$, for every $\varepsilon > 0$ there is a geometry-preserving network such that

$$\sup_{x \in \Omega \setminus \Gamma_P, t \in [0, T]} \|F(x, t) - \mathcal{N}(x, t)\|_\infty < \varepsilon.$$

Interpretation

Constraining where jumps may occur does not sacrifice universal approximation of the smooth regional dynamics.

Time-dependent forcing

For periodically forced systems, use harmonic features

$$\tau(t) = (\cos(\omega t), \sin(\omega t)), \quad \mathcal{N}(x, t) = \Psi(x, \tau(t), \Phi(z_P(x))).$$

- ▶ Known forcing frequency can be incorporated directly.
- ▶ Other continuous time features do not introduce new state-space jumps.
- ▶ Random Fourier features are a viable approach when ω is unknown.

Training by trajectory matching

$$\mathcal{L}_{\text{traj}}(\mathcal{N}) = \frac{1}{NJd} \sum_{n=1}^N \sum_{j=1}^J \|y_j^n - \phi_{\mathcal{N}}(t_j^n; t_0^n, y_0^n)\|_2^2.$$

$\phi_{\mathcal{N}}(t; t_0, y_0)$ is a numerical approximation to the two-time solution map of $\dot{x} = \mathcal{N}(x, t)$, with $x(t_0) = y_0$.

- ▶ Training: realise $\phi_{\mathcal{N}}$ with RK4 and gradually increase J .
- ▶ Inference: simulate the learned ODE with a Filippov-aware solver.

Experiments

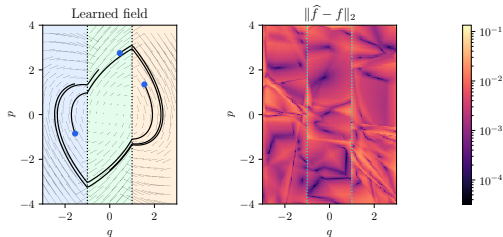
Providing the geometry vs enforcing it

$$\dot{q} = p, \quad \dot{p} = -q - \alpha(q)p, \quad \Sigma = \{q = \pm 1\}.$$

Shared information: exact switching code $z_P(x)$ and the same noisy vector-field samples.

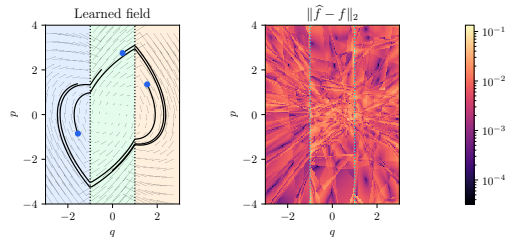
Geometry enforced

$$\mathcal{N}(x) = \Psi(x, \Phi(z_P(x)))$$

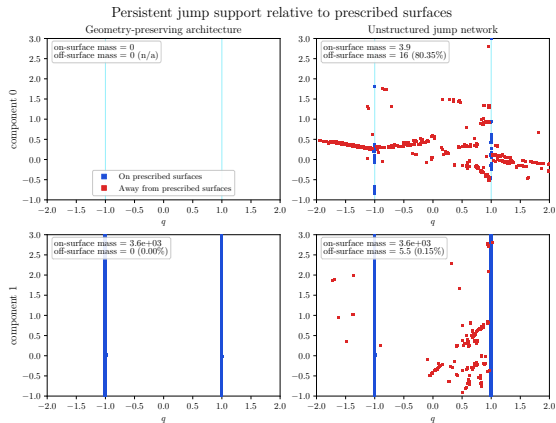


Geometry provided as input

$$\tilde{\mathcal{N}}(x) = G(x, z_P(x))$$



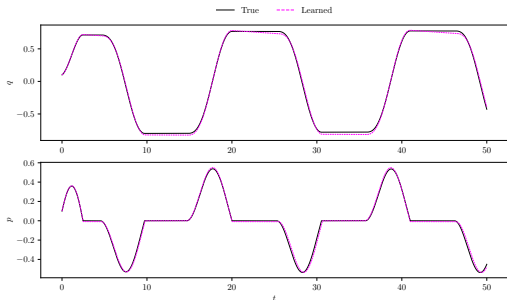
Where the unstructured jump network fails



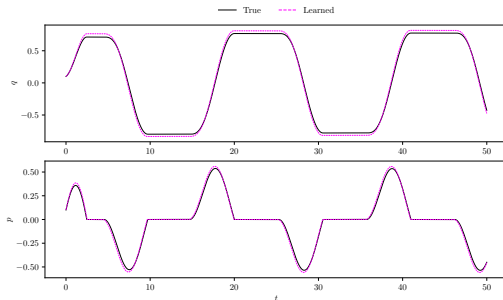
Correct geometry must be enforced by the architecture, not just supplied as an input.

End-to-end learning of dry friction

$$\dot{q} = p, \quad \dot{p} = -q - cp + A \cos(\omega t) - \mu \operatorname{sign}(p), \quad \Sigma = \{p = 0\}.$$



Recovered switching line



True switching line supplied

Both use the second-order model and known ω . Noise magnitude: $\sigma = 10^{-2}$.

Physical structure explains the accurate rollouts

How much do these two physical ingredients help?

Enforce the second-order structure

Recovered switch, $\sigma = 10^{-2}$

Model	Rollout RMSE	Final RMSE
Learn both components	$8.82 \cdot 10^{-2}$	$1.56 \cdot 10^{-1}$
Enforce $\dot{q} = p$	$1.50 \cdot 10^{-2}$	$2.35 \cdot 10^{-2}$

5.9× lower rollout error

Encode the forcing frequency

True switch, second-order model, $\sigma = 10^{-2}$

Time features	Rollout RMSE	Final RMSE
Random Fourier	$3.47 \cdot 10^{-2}$	$5.22 \cdot 10^{-2}$
Known ω	$2.19 \cdot 10^{-2}$	$3.51 \cdot 10^{-2}$

1.6× lower rollout error

Physics enters both the architecture and the choice of time features.

The PP04 system

V = global ice volume

A = Antarctic ice cover

C = atmospheric CO_2

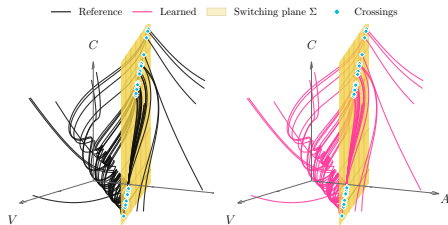
Dynamics

Periodic forcing: $l_{65}(t) = \mu \sin(\omega t)$

$$\dot{V} = \frac{-xC - y l_{65}(t) + z - V}{\tau_V},$$

$$\dot{A} = \frac{V - A}{\tau_A},$$

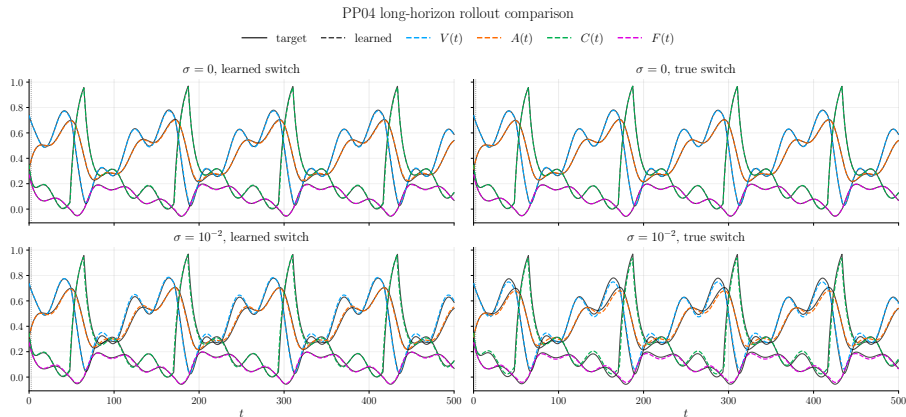
$$\dot{C} = \frac{\alpha l_{65}(t) - \beta V + \gamma \mathcal{H}(-s(V, A, C)) + \delta - C}{\tau_C}.$$



Reference and **learned** trajectories cross Σ .

Switch: $s(V, A, C) = 0.3V - 0.7A + 0.27$.

PP04 long-horizon rollouts



Clean/noisy data and recovered/known switching geometry.

Structured models are smaller and more accurate

Representative mild-noise results, $\sigma = 10^{-2}$:

System	Model	Parameters	Rollout RMSE	VF RMSE	VF-band RMSE
Dry friction	Disc	521	$1.50 \cdot 10^{-2}$	$2.02 \cdot 10^{-2}$	$2.31 \cdot 10^{-2}$
	NODE	8,705	$4.88 \cdot 10^{-2}$	$1.03 \cdot 10^{-1}$	$4.34 \cdot 10^{-1}$
PP04	Disc	448	$1.11 \cdot 10^{-2}$	$7.90 \cdot 10^{-3}$	$2.44 \cdot 10^{-2}$
	NODE	8,899	$2.61 \cdot 10^{-2}$	$1.05 \cdot 10^{-2}$	$3.11 \cdot 10^{-2}$

The discontinuous models use 17–20× fewer parameters and achieve lower errors on both benchmarks.

Disc: learned-geometry discontinuous model. NODE: smooth baseline.

Conclusions

Summary and extensions

1. Geometry and dynamics benefit from different inductive biases.
2. Interpretability benefits from separating the two learning problems.
3. The geometry-preserving architecture cannot invent off-switch discontinuities.
4. Physical structure improves accuracy, interpretability, and parameter efficiency.

Next directions

Experimental data, non-affine interfaces, interacting switches, and discontinuous latent dynamics.

Thank you

Questions?

dm2011@cam.ac.uk

Learning piecewise-smooth dynamical systems

Davide Murari, Erik Jansson, Chris Budd,
and Carola-Bibiane Schönlieb

arXiv:2608.19785 (2026)



Scan to read
the paper



Single-hyperplane recovery

1. Savitzky–Golay velocity estimation.
2. Global ranking of derivative-jump magnitudes.
3. Candidate expansion to neighbouring intervals.
4. RANSAC affine proposals and support counting.
5. TLS refinement of the winning inliers.
6. Normalisation and deterministic sign convention.

When does a derivative jump stand out?

At a transversal crossing (p, τ) , write the vector-field jump as

$$f^+(p, \tau) - f^-(p, \tau) = \kappa e, \quad \kappa > 0, \quad \|e\|_2 = 1.$$

For the difference Δ between consecutive linear velocity estimates, the local analysis gives

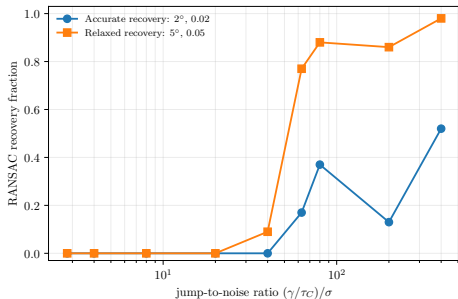
$$\mathbb{E}\|\Delta_{\text{cross}}\|_2 \geq \kappa|R| - \beta - \alpha, \quad \mathbb{E}\|\Delta_{\text{smooth}}\|_2 \leq \beta + \sigma\|c\|_2\sqrt{d}.$$

Hence a sufficient separation condition is

$$\underbrace{\kappa|R|}_{\text{jump magnitude} \times \text{stencil response}} > \underbrace{2\beta + \alpha}_{\text{local approximation errors}} + \underbrace{\sigma\|c\|_2\sqrt{d}}_{\text{differentiated observation noise}}.$$

R : response of the differenced velocity stencil to the crossing; c : its coefficients; α, β : bounded Taylor and smooth-motion remainders; d : state dimension.

PP04: an empirical recovery transition



Jump-to-noise ratio

$$\frac{\gamma/\tau_C}{\sigma}$$

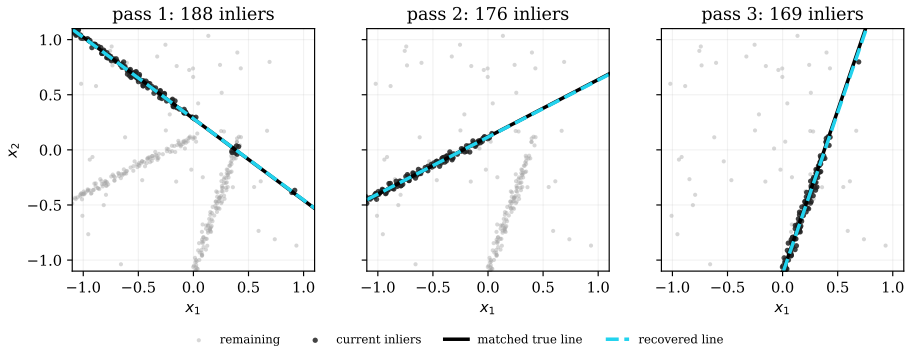
compares the PP04 jump rate with observation noise.

- Fixed $\sigma = 0.05$.
- 100 RANSAC seeds per γ .
- **Accurate:** angle $< 2^\circ$, offset < 0.02 .
- **Relaxed:** angle $< 5^\circ$, offset < 0.05 .

Recovery becomes reliable once the jump is strong enough relative to noise.

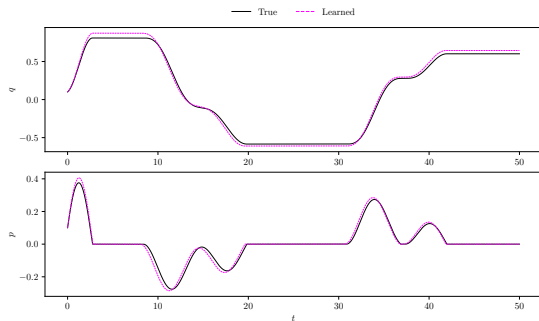
Caveat: changing γ also changes the trajectories and candidate cloud, so the transition need not be monotone.

Sequential recovery of multiple lines

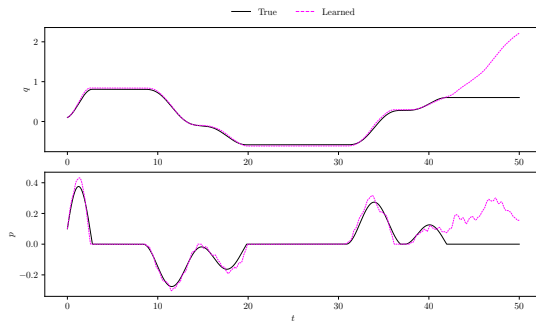


Greedy removal works for well-separated interfaces; interacting switches require joint model selection.

Known versus unknown forcing frequency



Known harmonic feature



Random Fourier features

Evaluation metrics

$$\text{RMSE}_{\text{rollout}}^2 = \frac{1}{N(K+1)} \sum_{n=1}^N \sum_{k=0}^K \|\hat{x}_n(t_k) - x_n(t_k)\|_2^2,$$

$$\text{RMSE}_f^2 = \frac{1}{M} \sum_{m=1}^M \|\hat{f}(x_m, t_m) - f(x_m, t_m)\|_2^2.$$

- ▶ Final-time RMSE probes accumulated trajectory error.
- ▶ Switching-band RMSE isolates local interface accuracy.

Event-aware rollout evaluation

1. Detect a change in the switching sign during a macro-step.
2. Localise the event by bisection.
3. Project the event point onto the hyperplane.
4. Cross the interface or enter the attracting sliding regime.
5. Propagate sliding motion with the Filippov convex combination.