When to Learn: Conformal Scores as Online Update Criteria

Ben Gao[1,2] Jordan Patracone[1,2] Stéphane Chrétien[3] Olivier Alata[1]

[1]Université Jean Monnet, Laboratoire Hubert Curien [2]Inria [3]Université Lyon 2, Laboratoire ERIC

Key Elements

We introduce a lightweight and novel framework designed for adaptive model updates in streaming environments: **Conformal Online Learning (COL)**, inspired by conformal prediction but repurposed for learning. We demonstrate the effectiveness of COL through its application to online learning of Koopman linear embeddings for non-linear dynamical systems

Koopman Operator [1]

Instead of directly modeling the nonlinear dynamics $x_{t+1} = Tx_t$, Koopman operator theory considers the evolution of lifted states through a linear operator:

$$\Phi_\theta(x_{t+1}) \approx K\Phi_\theta(x_t).$$

This approach enables the use of linear spectral analysis tools.

Conformal Online Learning of Koopman embeddings

The goal is to incrementally Φ_θ and K from sequentially observed dynamics. At each time step t, the parameters (θ_t, K_t) are updated by minimizing the multi-step prediction loss within $\mathcal{D}_t = \{x_{t-w}, \dots, x_t\}$. The score function at time step t is defined as

$$s_t(x_t, (\theta_t, K_t)) := \sum_{\tau=1}^{w} \|\Phi_{\theta_t}(x_t) - K_t^\tau \Phi_{\theta_t}(x_{t-\tau})\|^2$$

to evaluate the consistency of (θ_t, K_t).

Prediction Score Set

Given a newly observed state x_t and a conformity threshold $q_t > 0$, the *prediction score set* at time t is defined as

$$S_t = s_t(x_t, \mathrm{Param}_t),$$

where $\mathrm{Param}_t = \{(\theta, K) \text{ such that } s = s_t(x_t, (\theta, K)) \le q_t\}$. This set contains all prediction scores attainable at x_t by Koopman models that satisfy the current calibration constraint.

Dynamic threshold by Conformal PI Control [2]

Rather than fixing the quantile threshold q_t in advance, the method updates it online in response to conformity violations. After observing whether $s_t \in S_t$, a binary error signal $e_t = \mathbf{1}\{s_t \notin S_t\}$ is computed, and the threshold q_t is adjusted via:

$$q_{t+1} = q_t + \underbrace{\gamma(e_t - \alpha)}_{\text{Proportional term (P)}} + \underbrace{r_t\left(\sum_{i=1}^{t}(e_i - \alpha)\right)}_{\text{Integral term (I)}}$$

Dynamic Regret Theorem

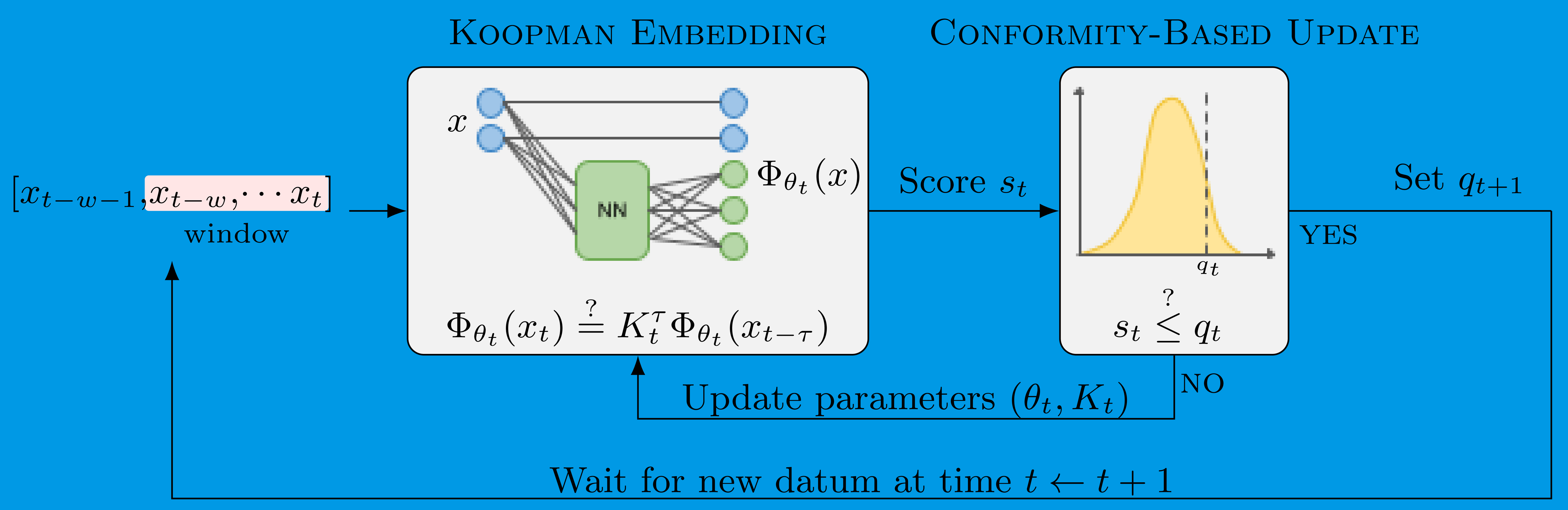
Let (θ_t, K_t) be the parameters produced by COL and let $(\theta_t^*, K_t^*) \in \operatorname{argmin}_{(\theta,K)} \mathcal{L}_t(\theta, K)$ denote any time-dependent optimal model minimizing the loss at step t. Further assume:

- (A1) Each $\mathcal{L}_t$ is L-smooth with $\|\nabla \mathcal{L}_t(\theta, K)\| \le B$;
- (A2) Bounded total variation and squared variation for oracle path: $V_T := \sum_{t=1}^{T} \|(\theta_{t+1}^*, K_{t+1}^*) - (\theta_t^*, K_t^*)\| < \infty$, $S_T := \sum_{t=1}^{T} \|(\theta_{t+1}^*, K_{t+1}^*) - (\theta_t^*, K_t^*)\|^2 < \infty$;
- (A3) $\sum_{t=1}^{T} q_t \le \mathcal{O}(\alpha h(T))$ for some sublinear, nonnegative, nondecreasing function h;

Then the dynamic regret satisfies:

$$\sum_{t=1}^{T} [\mathcal{L}_t(\theta_t, K_t) - \mathcal{L}_t(\theta_t^*, K_t^*)] \le \mathcal{O}(\alpha h(T) + V_T + S_T).$$

Shifting from producing conformal prediction sets to **assessing the model's consistency** on new observations, model updates are triggered only when its conformal score $s_t(\text{model}_t)$ is greater than a dynamic threshold q_t, and continue until $s_t(\text{model}_t) \le q_t$.



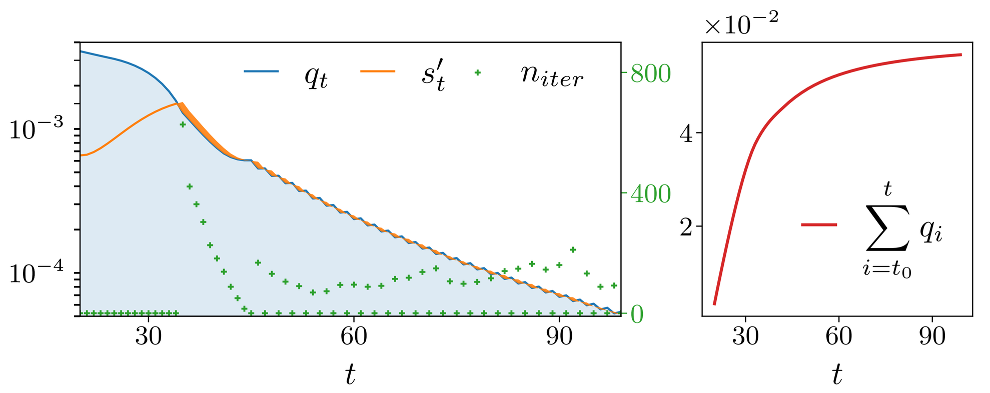
Take a picture to download the full paper

Compare to baselines

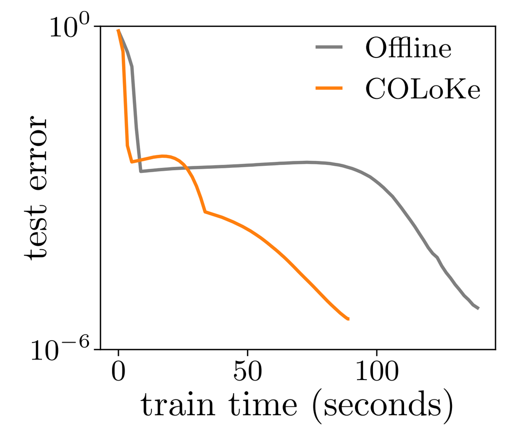
Models are trained using the training trajectories while computing the online prediction error. After the training, models are evaluated on the held-out trajectories to compute generalization error. We report averages across five splits, along with the standard deviation of the averages.

	ODMD	OEDMD	OnlineAE	OLoKe	COLoKe
Single attractor	$1.1 \cdot 10^{-3}$ $(\pm 3.6 \cdot 10^{-5})$ $4.6 \cdot 10^{-5}$ $(\pm 7.3 \cdot 10^{-7})$	$2.5 \cdot 10^{-2}$ $(\pm 2.8 \cdot 10^{-4})$ $1.5 \cdot 10^{-2}$ $(\pm 4.7 \cdot 10^{-4})$	$1.0 \cdot 10^{-2}$ $(\pm 7.7 \cdot 10^{-4})$ $7.4 \cdot 10^{-5}$ $(\pm 2.8 \cdot 10^{-5})$	$2.1 \cdot 10^{-6}$ $(\pm 6.6 \cdot 10^{-7})$ $7.5 \cdot 10^{-6}$ $(\pm 2.5 \cdot 10^{-6})$	$\mathbf{2.4 \cdot 10^{-7}}$ $(\pm \mathbf{3.6 \cdot 10^{-8}})$ $\mathbf{7.6 \cdot 10^{-7}}$ $(\pm \mathbf{9.6 \cdot 10^{-8}})$
Duffing oscillator	$2.5 \cdot 10^{-4}$ $(\pm 7.8 \cdot 10^{-6})$ $1.9 \cdot 10^{-4}$ $(\pm 1.5 \cdot 10^{-6})$	$6.8 \cdot 10^{-3}$ $(\pm 4.5 \cdot 10^{-3})$ $3.8 \cdot 10^{-3}$ $(\pm 3.3 \cdot 10^{-4})$	$8.7 \cdot 10^{-3}$ $(\pm 2.5 \cdot 10^{-3})$ $2.0 \cdot 10^{-3}$ $(\pm 6.2 \cdot 10^{-4})$	$5.5 \cdot 10^{-5}$ $(\pm 1.0 \cdot 10^{-5})$ $2.3 \cdot 10^{-4}$ $(\pm 4.0 \cdot 10^{-5})$	$\mathbf{3.1 \cdot 10^{-6}}$ $(\pm \mathbf{2.3 \cdot 10^{-7}})$ $\mathbf{7.3 \cdot 10^{-5}}$ $(\pm \mathbf{1.9 \cdot 10^{-5}})$
VdP oscillator	$2.1 \cdot 10^{-3}$ $(\pm 3.6 \cdot 10^{-5})$ $1.1 \cdot 10^{-3}$ $(\pm 4.7 \cdot 10^{-6})$	$2.1 \cdot 10^{-3}$ $(\pm 3.2 \cdot 10^{-5})$ $1.1 \cdot 10^{-3}$ $(\pm 7.8 \cdot 10^{-6})$	$1.7 \cdot 10^{-2}$ $(\pm 3.0 \cdot 10^{-3})$ $3.8 \cdot 10^{-3}$ $(\pm 1.0 \cdot 10^{-3})$	$6.6 \cdot 10^{-4}$ $(\pm 1.5 \cdot 10^{-4})$ $9.2 \cdot 10^{-4}$ $(\pm 3.0 \cdot 10^{-4})$	$\mathbf{3.8 \cdot 10^{-4}}$ $(\pm \mathbf{1.2 \cdot 10^{-5}})$ $\mathbf{6.0 \cdot 10^{-4}}$ $(\pm \mathbf{1.4 \cdot 10^{-4}})$
Lorenz system	$2.7 \cdot 10^{-1}$ $(\pm 1.3 \cdot 10^{-3})$ $1.0 \cdot 10^{-1}$ $(\pm 5.8 \cdot 10^{-4})$	$5.5 \cdot 10^{-1}$ $(\pm 2.2 \cdot 10^{-2})$ $2.7 \cdot 10^{-1}$ $(\pm 3.1 \cdot 10^{-2})$	$5.9 \cdot 10^{-1}$ $(\pm 8.4 \cdot 10^{-2})$ $3.8 \cdot 10^{-2}$ $(\pm 2.6 \cdot 10^{-3})$	$7.6 \cdot 10^{-3}$ $(\pm 1.8 \cdot 10^{-4})$ $4.7 \cdot 10^{-3}$ $(\pm 3.0 \cdot 10^{-4})$	$\mathbf{6.5 \cdot 10^{-3}}$ $(\pm \mathbf{1.0 \cdot 10^{-4}})$ $\mathbf{3.3 \cdot 10^{-3}}$ $(\pm \mathbf{1.1 \cdot 10^{-4}})$
ETD $\star$	$\mathbf{1.4 \cdot 10^{-1}}$ $(\pm \mathbf{2.9 \cdot 10^{-1}})$ $1.2 \cdot 10^{-1}$ $(\pm 2.9 \cdot 10^{-1})$	$2.7 \cdot 10^{-1}$ $(\pm 3.0 \cdot 10^{-1})$ $1.5 \cdot 10^{-1}$ $(\pm 2.6 \cdot 10^{-1})$	$2.1 \cdot 10^{-1}$ $(\pm 3.4 \cdot 10^{-1})$ $7.9 \cdot 10^{-2}$ $(\pm 7.3 \cdot 10^{-2})$	$2.1 \cdot 10^{-1}$ $(\pm 3.4 \cdot 10^{-2})$ $9.7 \cdot 10^{-2}$ $(\pm 8.5 \cdot 10^{-2})$	$2.1 \cdot 10^{-1}$ $(\pm 8.6 \cdot 10{-}2)$ $\mathbf{7.3 \cdot 10^{-2}}$ $(\pm \mathbf{6.3 \cdot 10^{-2}})$

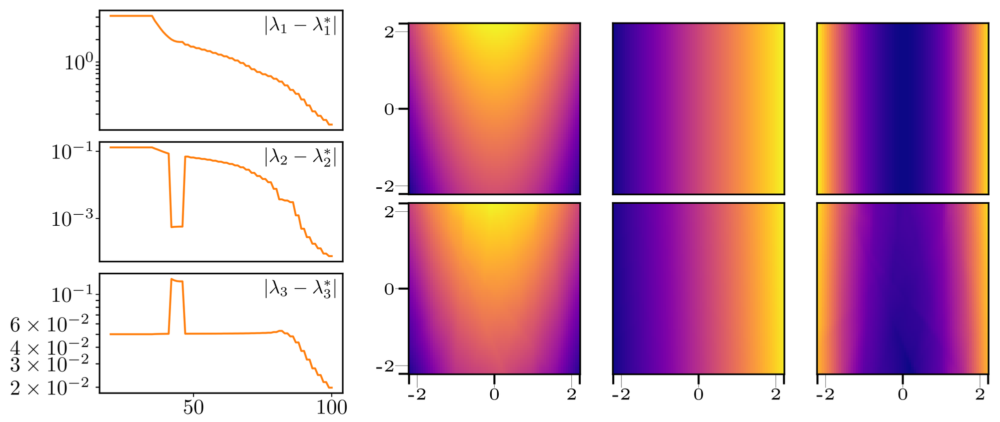
Adaptive update behavior



Learning efficiency



Convergence of eigen-properties



Acknowledgements

This work was sponsored by a public grant overseen by Auvergne-Rhône-Alpes region, Grenoble Alpes Metropole and BPIFrance, as part of project I-Démo Région "Green AI".

References

1. S. L. Brunton, M. Budišić, E. Kaiser, and J. N. Kutz, "Modern Koopman Theory for Dynamical Systems," *SIAM Review*, vol. 64, no. 2, pp. 229–340, 2022.
2. A. Angelopoulos, E. Candès, and R. J. Tibshirani, "Conformal PID Control for Time Series Prediction," *Advances in Neural Information Processing Systems*, vol. 36, pp. 23047–23074, 2023.