MITSUBISHI ELECTRIC RESEARCH LABORATORIES
https://www.merl.com

Quantifying Similarity Between Nonlinear Systems From Data Using Matrix Profiles

Chakrabarty, Ankush; Nikovski, Daniel N.

TR2026-108 July 14, 2026

Abstract

Determining which dynamical systems exhibit behavior most similar to a target system is critical for transfer learning and data-efficient system identification. However, quantifying similarity between nonlinear systems from input-output data is challenging: trajectories differ in length, initial conditions, and input sequences, making direct comparison unreliable. Classical system identification depends on correct model structure selection, while frequency-domain methods often miss transient dynamics. This paper proposes a data-driven methodology based on matrix profiles: a time-series tool for efficient subsequence matching. The approach operates directly on input-output trajectories without state estimation or parameter identification. By embedding trajectories in an interleaved inputoutput space and computing cross-matrix profiles, the method captures dynamical features including transient behavior under varying conditions. We derive analytical bounds on matrix profile distance as functions of parameter and input discrepancies, establishing that systems closer in parameter space yield smaller matrix profile distances when control trajectories are similar. Numerical experiments demonstrate improved similarity assessment compared to Euclidean distance and dynamic time warping baselines, achieving substantial reductions in trajectory error and parameter-space deviation across both linear and nonlinear systems.

International Conference on Control, Decision, and Information (CoDIT) 2026

Mitsubishi Electric Research Laboratories, Inc.
201 Broadway, Cambridge, Massachusetts 02139

Quantifying Similarity Between Nonlinear Systems From Data Using Matrix Profiles

Ankush Chakrabarty and Daniel Nikovski

Abstract—**Determining which dynamical systems exhibit behavior most similar to a target system is critical for transfer learning and data-efficient system identification. However, quantifying similarity between nonlinear systems from input-output data is challenging: trajectories differ in length, initial conditions, and input sequences, making direct comparison unreliable. Classical system identification depends on correct model structure selection, while frequency-domain methods often miss transient dynamics. This paper proposes a data-driven methodology based on matrix profiles: a time-series tool for efficient subsequence matching. The approach operates directly on input-output trajectories without state estimation or parameter identification. By embedding trajectories in an interleaved input-output space and computing cross-matrix profiles, the method captures dynamical features including transient behavior under varying conditions. We derive analytical bounds on matrix profile distance as functions of parameter and input discrepancies, establishing that systems closer in parameter space yield smaller matrix profile distances when control trajectories are similar. Numerical experiments demonstrate improved similarity assessment compared to Euclidean distance and dynamic time warping baselines, achieving substantial reductions in trajectory error and parameter-space deviation across both linear and nonlinear systems.**

I. Introduction

Determining which dynamical systems within a large collection exhibit behavior most similar to a target system is critical in control engineering, robotics, and applied machine learning. Consider a fleet of industrial robots with variations in actuator dynamics, friction, or load distributions. Identifying which previously characterized systems are most similar enables effective warm-starting of optimization routines [1], principled reuse of identified models [2], and data-efficient adaptation (i.e., transfer learning) across domains including building energy management, medical devices, and aerospace applications. Despite its importance, quantifying similarity between nonlinear dynamical systems from input–output data is challenging. Directly matching full trajectories is difficult because they often differ in length, require alignment, start from different initial states, and are driven by different input sequences—all of which can make output comparisons unreliable, even for the same system.

Classical system identification estimates parametric models for comparison but depends on selecting the correct model structure [3], and proximity in parameter space may not reflect actual behavioral similarity under varying conditions. Moreover, the identification of even a linear model is possible only up to a linear transform of the constructed state space [4], so comparing the identified system matrices is misleading, as two data sets from the same underlying dynamical system can lead to very different system matrices. This problem becomes only worse for non-linear system models whose parameters are often identified by means of non-linear optimization that is prone to getting trapped in local minima and cannot be expected to converge to the same parameter estimates even when executed again on the same data. Model-free frequency-domain fingerprinting avoids structural assumptions but often misses transient or input-driven dynamics [5]. The aforementioned difficulties of classical system identification have recently led to the emergence of data-driven control methods [6], but their ideas are not easy to apply to reasoning about similarities between dynamical systems, because the sampled data sets from such systems are generally very different from one another.

Recent advances in transfer learning and meta-learning demonstrate that knowledge from source systems can accelerate learning on target systems when appropriate similarity guides source selection [2], [7]–[11] . However, these methods often require explicit system models or computationally expensive retraining to assess transferability [12]. A scalable, data-driven similarity metric operating directly on observed trajectories (without parameter identification or model retraining) would significantly enhance practical deployment.

A similarity metric between two systems can be defined by the principle that similar systems produce the same outputs when given the same inputs in the same states. However, identical states and inputs are rarely achievable in practice and we instead would like to compare trajectories from normal operation. Although system states are typically unobservable, Takens' theorem [13] guarantees that a delay embedding of sufficient length uniquely represents the true state for autonomous systems; including past control inputs extends this to controllable systems [14]. Thus, when two trajectories have identical past observations and controls, they can be considered in the same state and expected to yield identical next outputs if their dynamics match.

If the trajectories are long enough to cover neighborhoods of all states and control inputs, we should find closely matching subsequences across systems. This motivates a similarity metric based on how well each subsequence of observations and controls in one trajectory best matches at least one subsequence in the other. Efficient computation can be achieved using time-series algorithms for computing matrix profiles (e.g., STOMP [15]), which rapidly find nearest-neighbor subsequence matches without requiring global alignment—ideal for comparing systems with transient or input-driven dynamics. Thus, we propose a matrix-profile-based methodology for quantifying similarity between nonlinear discrete-time dynamical systems directly from input-output data. By embedding trajectories in an interleaved input-output space and computing cross-matrix profiles, our approach captures salient dynamical features (including transient behavior and input-output coupling) without explicit model identification.

Ankush Chakrabarty is with Insulet Corporation, Acton, MA 01720 and Daniel Nikovski is with Mitsubishi Electric Research Laboratories (MERL), Cambridge, MA 02139, USA. achakrabarty@ieee.org, nikovski@merl.com.

We establish formal connections between matrix profile distances and parameter discrepancies, providing theoretical guarantees on metric behavior and demonstrating utility for knowledge transfer.

Our similarity metric addresses source selection in data-driven control. For meta-learning [8], [10], it enables automatic identification of relevant source systems, improving efficiency and reducing negative transfer risk. For transfer learning [2], [16], ranking systems by matrix profile distance provides principled source selection without model retraining. Our approach complements classical system identification [3], [17] by offering behavioral similarity measures for model validation and prescreening. Unlike methods relying on linearization or frequency-domain representations, matrix profile algorithms operate on raw input–output sequences and are robust to sampling variations, noise, and trajectory lengths [18]–[20].

Our major *contributions* include: (i) a novel data-driven similarity metric for nonlinear systems based on matrix profiles, operating on input-output trajectories without state estimation or parameter identification; (ii) theoretical bounds on matrix profile distance as functions of parameter and input discrepancies; (iii) empirical validation on linear and nonlinear oscillators, demonstrating improved similarity quantification versus baselines including dynamic time warping.

II. Problem Statement

We consider a family of nonlinear discrete-time dynamical systems

$$x^+ = f(x,u,\theta), \quad y = h(x) + \eta, \tag{1}$$

with state $x \in \mathbb{R}^{n_x}$, control input $u \in \mathcal{U} \subset \mathbb{R}^{n_u}$, and output $y \in \mathbb{R}^{n_y}$ corrupted by isotropic Gaussian noise $\eta \sim \mathcal{N}(0, \sigma I)$. The functions $f : \mathbb{R}^{n_x} \times \mathbb{R}^{n_u} \times \Theta \to \mathbb{R}^{n_x}$ and $h : \mathbb{R}^{n_x} \to \mathbb{R}^{n_y}$ are the state-transition and output functions, respectively, and their analytical forms are unknown; e.g., they may be high-fidelity simulators where we do not have access to the internal equations or state. The systems (1) are parameterized by $\theta \in \Theta$, where Θ is a product of finite intervals in $\mathbb{R}^{n_\theta}$. We make the following assumptions on the class of systems considered herein.

Assumption 1. The dynamics $f : \mathbb{R}^n \times \mathcal{U} \times \Theta \to \mathbb{R}^n$ and output function $h : \mathbb{R}^n \to \mathbb{R}^p$ satisfy the following properties:

(A1) (*Contractivity*) There exists a contraction constant $\rho \in (0,1)$ such that, for all $\theta \in \Theta$, $u \in \mathcal{U}$, and $x, \tilde{x} \in \mathbb{R}^n$,

$$\|f(x,u,\theta) - f(\tilde{x},u,\theta)\| \le \rho \, \|x - \tilde{x}\|. \tag{2a}$$

(A2) (*Continuity w.r.t. parameters*) There exists a Lipschitz constant $L_\theta > 0$ such that, for all $x \in \mathbb{R}^n$, $u \in \mathcal{U}$, and $\theta_1, \theta_2 \in \Theta$,

$$\|f(x,u,\theta_1) - f(x,u,\theta_2)\| \le L_\theta \, \|\theta_1 - \theta_2\|. \tag{2b}$$

(A3) (*Continuity w.r.t. inputs*) There exists a Lipschitz constant $L_u > 0$ such that, for all $x \in \mathbb{R}^n$, $\theta \in \Theta$, and $u_1, u_2 \in \mathcal{U}$,

$$\|f(x,u_1,\theta) - f(x,u_2,\theta)\| \le L_u \, \|u_1 - u_2\|. \tag{2c}$$

(A4) (*Continuity of output*) There exists a Lipschitz constant $L_h > 0$ such that, for all $x_1, x_2 \in \mathbb{R}^n$,

$$\|h(x_1) - h(x_2)\| \le L_h \, \|x_1 - x_2\|. \tag{2d}$$

These assumptions are broad enough to encompass many nonlinear systems in identification. The last two impose standard Lipschitz continuity, ensuring smooth responses of the dynamics and output to input and state changes. Assumption (A1) makes the dynamics contractive in state, keeping trajectories stable and excluding marginally unstable cases. Assumption (A2) limits sensitivity to parameter variations, so small parameter changes cause only small deviations in state evolution.

We consider a family of dynamical systems parameterized by $\theta \in \Theta$. Each system evolves according to unknown state and output maps f and h, but only input-output data trajectories are observed. Specifically, for each source system $s = 1, \dots, N_{\rm src}$, we have access to sequences $\{u_t^{(s)}, y_t^{(s)}\}_{t=1}^{T_s}$ generated under a distinct parameter vector $\theta^{(s)} \in \Theta$. These trajectories collectively form the *source dataset* $\mathcal{D}_{\rm src}$.

Now consider a *target system* governed by an unknown parameter $\theta^\star \in \Theta$. We only observe input-output measurements $\{u_t^\star, y_t^\star\}_{t=1}^{T^\star}$ from the system, which define the *target dataset* $\mathcal{D}_{\rm tgt}$. Our primary objective is to determine, directly from the observed trajectories, without access to the maps f or h, and without explicit state information, which source system (or subset of source systems) exhibit dynamical behavior most similar to the target. The output of the procedure is the observed trajectory of the most similar source system.

The main challenge lies in defining a similarity measure that is both informative and tractable. Such a measure should: (i) operate directly on observable data trajectories, without requiring model access, state estimation, or parameter identification; (ii) remain sensitive to the underlying dynamical structure even under differences in inputs, initial conditions, or time horizons; and (iii) scale efficiently to large libraries of source systems.

III. Proposed Solution

We propose a matrix-profile (MP) based similarity operator that scores, ranks, and retrieves source systems most similar to a target system directly from input-output trajectories; based on the concept of matrix profiles proposed in [15], [21]. The approach comprises an interleaved representation that preserves instantaneous input-output coupling, a pan-library construction with guard padding to prevent cross-system matches, a cross-matrix-profile AB-join to obtain per-window nearest-neighbor distances; and an aggregation rule that yields a system-level similarity score and ranking. Each step is discussed in more detail hereafter.

A. Interleaved Representation

For any system with output $y = \{y_t\}_{t=1}^T$ and input $u = \{u_t\}_{t=1}^T$, define the interleaving operator

$$\mathcal{I}(y,u) \triangleq [\, y_1, \, u_1, \, y_2, \, u_2, \, \dots, \, y_T, \, u_T \,]^\top \in \mathbb{R}^{2T}. \tag{3}$$

We denote by $Z^{(s)} \in \mathbb{R}^{2T_s}$ the interleaved sequence for source system $s \in \{1, \dots, N_{\rm src}\}$, and by $Z^\star \in \mathbb{R}^{2T^\star}$ the interleaved target sequence. We choose a subsequence length $m = wn_u + (w+1)n_y$, where w is the length of the time-lag window, so that any length-m window that starts on an output element (i.e., an index aligned with a y-entry) also ends on an output element.

Since $\mathcal{I}(y,u)$ begins with y_1 at index 1, as shown in (3), the output entries occur at all odd indices $\{1, 3, 5, \dots\}$ while input entries occur at all even indices $\{2, 4, 6, \dots\}$. Therefore, to ensure each window begins and ends on an output, we restrict window starting positions to the set of odd indices:

$$\mathcal{K} \triangleq \{\, i \in \{1, \dots, 2T_t - m + 1\} \mid i \text{ is odd} \,\}. \tag{4}$$

B. Pan-Library Construction with Guard Padding

To enable a *single* AB-join[1] against the entire source library, we concatenate all interleaved source sequences with guard segments of NaNs:

$$S_{\text{lib}} \triangleq Z^{(1)} \oplus \texttt{NaN}_m \oplus Z^{(2)} \oplus \texttt{NaN}_m \oplus \cdots \oplus Z^{(N_{\text{src}})} \in \mathbb{R}^{L_{\text{lib}}}, \quad (5)$$

where $\oplus$ denotes concatenation, $\texttt{NaN}_m$ is a length-m vector of undefined values, and $L_{\text{lib}} \triangleq \sum_{s=1}^{N_{\text{src}}}(2T_s + m) - m$. We create an index map

$$q : \{1, \dots, L_{\text{lib}}\} \to \{1, \dots, N_{\text{src}}\} \cup \{-1\}, \quad (6)$$

that records the source-system identity of each position in S_{lib} ($q(k) = -1$ for NaN guard regions). The NaN padding ensures that no admissible window in the AB-join can straddle two systems.

C. Cross-Matrix Profile AB-Join

Given the target $Z^\star$ and the library S_{lib}, the cross-matrix profile AB-join with window length m returns, for each valid starting index i in the target, the minimum Euclidean distance to any length-m window in the library and the corresponding index:

$$\text{MP}^\star[i] \triangleq \min_{j \in \mathcal{J}} \left\| Z^\star_{i:i+m-1} - S_{\text{lib},\, j:j+m-1} \right\|_2, \quad (7a)$$

$$\mathcal{I}^\star[i] \triangleq \arg\min_{j \in \mathcal{J}} \left\| Z^\star_{i:i+m-1} - S_{\text{lib},\, j:j+m-1} \right\|_2, \quad (7b)$$

where $i \in \{1, \dots, 2T^\star - m + 1\}$ and $\mathcal{J} \triangleq \{\, j : q(j) \neq -1 \ \wedge\ j,\, j+m-1$ both lie within the same $Z^{(s)}\,\}$. We then enforce the alignment set $\mathcal{K}$ from (4):

$$\text{MP}^\star_{\text{filt}} \triangleq \{\text{MP}^\star[i] : i \in \mathcal{K}\}, \quad (8a)$$

$$\mathcal{I}^\star_{\text{filt}} \triangleq \{\mathcal{I}^\star[i] : i \in \mathcal{K}\}. \quad (8b)$$

Distances may be computed with or without z-normalization. Using raw (non-normalized) distances preserves amplitude information (often diagnostic of parameter differences); z-normalized distances emphasize shape while discounting scale. The choice is a design knob and can be ablated; although we recommend switching off z-normalization.

D. Per-System Aggregation and Ranking

For source system s, collect the distances of all $\mathcal{K}$-aligned target windows whose best match lies within $Z^{(s)}$, that is

$$\mathcal{D}^{(s)}_{\text{MP}} \triangleq \{\, \text{MP}^\star_{\text{filt}}[k] \ : \ q(\mathcal{I}^\star_{\text{filt}}[k]) = s \,\}. \quad (9)$$

Define the *similarity score* via an aggregate statistic

$$\boxed{S^{(s)} \triangleq \begin{cases} \text{mean}(\mathcal{D}^{(s)}_{\text{MP}}), & \text{if } |\mathcal{D}^{(s)}_{\text{MP}}| > 0, \\ +\infty, & \text{otherwise.} \end{cases}} \quad (10)$$

Lower $S^{(s)}$ indicates higher similarity. Alternative aggregation operators include median, trimmed-mean, or min, trading off robustness vs. sensitivity. The *ranking of similarity* is obtained by sorting $S^{(s)}$ (or $\tilde{S}^{(s)}$) in ascending order,

$$(s_1, \dots, s_{N_{\text{src}}}) \ \text{ such that } \ S^{(s_1)} \leq \cdots \leq S^{(s_{N_{\text{src}}})}, \quad (11)$$

and the top-k set $\{s_1, \dots, s_k\}$ denotes the k most similar systems estimated by the proposed approach.

[1]An *AB-join* is the cross-matrix-profile operation between two time series S_1 and S_2, returning for every subsequence of S_1 its nearest-neighbor distance within S_2. Efficient implementations such as STOMP and SCAMP exploit incremental dot-product updates to compute the AB-join in near-linear time with respect to the total sequence length [22].

E. Computational Complexity

Let $L_q \triangleq 2T_t$ and L_{lib} be the lengths of the target and library sequences. The AB-join with window m has time complexity $\mathcal{O}(L_q\, L_{\text{lib}})$ for STOMP-style exact MP (with favorable constants due to incremental dot-product updates and FFT amortization) and memory $\mathcal{O}(L_{\text{lib}})$. The stride filter and aggregation steps are linear in L_q and the number of matches, respectively. Thus the overall complexity is dominated by the single AB-join and scales linearly in the size of the concatenated library, rather than requiring $\mathcal{O}(N_{\text{src}})$ separate joins.

A full pseudocode is provided in Algorithm 1.

Algorithm 1 MATRIX-PROFILE RETRIEVAL

Require: Source datasets $\{(u^{(s)}, y^{(s)})\}_{s=1}^{N_{\text{src}}}$, target dataset $(u^\star, y^\star)$; window size m (must be <u>odd</u>); normalization flag NORM $\in \{\text{raw}, z\}$; coverage weight $\lambda \geq 0$.

$Z^{(s)} \leftarrow \mathcal{I}(y^{(s)}, u^{(s)})$ for all s; $Z^\star \leftarrow \mathcal{I}(y^\star, u^\star)$.
$S_{\text{lib}} \leftarrow Z^{(1)} \oplus \texttt{NaN}_m \oplus \cdots \oplus Z^{(N_{\text{src}})}$; build index map $q(\cdot)$ as in (6).
$\text{MP}^\star, \mathcal{I}^\star \leftarrow \mathsf{ABJoin}(Z^\star, S_{\text{lib}}, m, \text{NORM})$, excluding windows intersecting NaNs.
Construct $\mathcal{K}$ as in (4); set $\text{MP}^\star_{\text{filt}}, \mathcal{I}^\star_{\text{filt}}$ as in (8).
for $s = 1$ **to** N_{src} **do**
    $\mathcal{D}^{(s)} \leftarrow \{\, \text{MP}^\star_{\text{filt}}[k] \mid q(\mathcal{I}^\star_{\text{filt}}[k]) = s \,\}$
    if $|\mathcal{D}^{(s)}| = 0$ **then**
        $S^{(s)} \leftarrow +\infty$
    else
        $S^{(s)} \leftarrow \text{mean}(\mathcal{D}^{(s)})$
    end if
end for
Sort systems by ascending $S^{(s)}$ and **return** top-k and full ranking

---

IV. THEORETICAL ANALYSIS

We make the following additional assumptions in this section to bound the performance of our proposed approach.

Assumption 2. All systems start from the same initial state: $x_0(\theta_1) = x_0(\theta_2) = x_0$ for all $\theta_1, \theta_2 \in \Theta$, and all data trajectories are of length T.

Assumption 3. All admissible control sequences are bounded in energy and amplitude:

$$\sup_{t \in \{0, \dots, T-1\}} \|u_t\| < \infty, \qquad \frac{1}{T} \sum_{t=0}^{T-1} \|u_t\|^2 < \infty.$$

Assumption 2 is not necessary, just made for ease of exposition. Assumption 3 is also not unreasonable, since we expect real-world actuation to be bounded. For a given parameter $\theta_i \in \Theta$ and control sequence $U^{(i)} = \{u_t^{(i)}\}_{t=0}^{T-1}$, we denote the state trajectory as $X_t^{(i)} = x_t(\theta_i, U^{(i)})$, and the output trajectory as $Y_t^{(i)} = y_t(\theta_i, U^{(i)}) = h(X_t^{(i)})$.

Definition 1. For two control sequences $U^{(1)} = \{u_t^{(1)}\}_{t=0}^{T-1}$ and $U^{(2)} = \{u_t^{(2)}\}_{t=0}^{T-1}$, we define the average control discrepancy

$$\Delta_U = \left(\frac{1}{T} \sum_{t=0}^{T-1} \left\| u_t^{(1)} - u_t^{(2)} \right\|^2 \right)^{1/2}.$$

Note we slightly abuse notation here for compactness by dropping the dependence on the superscripts. By Assumption 3, the set of admissible control trajectories $\mathcal{U} = \{\{u_t\}_{t=0}^{T-1}\}$ is bounded in both ℓ_∞ and ℓ_2 norms.

Theorem 1. *Consider the noise-free case: $\eta \equiv 0$. Under Assumptions 1–3, there exist constants $C_1, C_2 > 0$ depending only on $\rho, L_\theta, L_u, L_h, m$, and T, such that for any $\theta_1, \theta_2 \in \Theta$ and control sequences $U^{(1)}, U^{(2)}$:*

$$\mathcal{D}_{MP}(\theta_1, \theta_2) \le C_1 \|\theta_1 - \theta_2\| + C_2 \Delta_U. \tag{12}$$

Furthermore, suppose that for parameters $\theta_1, \theta_2, \theta_3 \in \Theta$ and their respective controls, we have $\Delta_U^{(i,j)} \le \epsilon$ for all $i, j \in \{1,2,3\}$, and that

$$\|\theta_1 - \theta_2\| + \frac{C_2}{C_1}\epsilon < \|\theta_1 - \theta_3\|.$$

Then $\mathcal{D}_{MP}(\theta_1, \theta_2) < \mathcal{D}_{MP}(\theta_1, \theta_3)$.

Proof. For any $t \ge 0$, we have:

$$\begin{aligned}
\|x_{t+1}^{(1)} - x_{t+1}^{(2)}\| &= \|f(x_t^{(1)}, u_t^{(1)}, \theta_1) - f(x_t^{(2)}, u_t^{(2)}, \theta_2)\| \\
&\le \|f(x_t^{(1)}, u_t^{(1)}, \theta_1) - f(x_t^{(2)}, u_t^{(1)}, \theta_1)\| \\
&\quad + \|f(x_t^{(2)}, u_t^{(1)}, \theta_1) - f(x_t^{(2)}, u_t^{(2)}, \theta_1)\| \\
&\quad + \|f(x_t^{(2)}, u_t^{(2)}, \theta_1) - f(x_t^{(2)}, u_t^{(2)}, \theta_2)\|.
\end{aligned}$$

Let $\delta_t = \|x_t^{(1)} - x_t^{(2)}\|$. By assumption, we get

$$\delta_{t+1} \le \rho\delta_t + L_u\|u_t^{(1)} - u_t^{(2)}\| + L_\theta\|\theta_1 - \theta_2\|.$$

Since $\delta_0 = 0$ by Assumption 2, iterating this recursion yields:

$$\delta_t \le \sum_{k=0}^{t-1} \rho^{t-1-k}[L_u\|u_k^{(1)} - u_k^{(2)}\| + L_\theta\|\theta_1 - \theta_2\|].$$

For the control term, by the Cauchy-Schwarz inequality:

$$\begin{aligned}
&\sum_{k=0}^{t-1} \rho^{t-1-k}\|u_k^{(1)} - u_k^{(2)}\| \\
&\le \left(\sum_{k=0}^{t-1} \rho^{2(t-1-k)}\right)^{1/2} \left(\sum_{k=0}^{t-1} \|u_k^{(1)} - u_k^{(2)}\|^2\right)^{1/2} \\
&\le \left(\sum_{j=0}^{\infty} \rho^{2j}\right)^{1/2} \sqrt{t \cdot T} \cdot \Delta_U \\
&= \frac{\sqrt{tT}}{\sqrt{1-\rho^2}}\Delta_U.
\end{aligned}$$

Since $\sum_{k=0}^{t-1} \rho^k \le 1/(1-\rho)$, and $t \le T$, we obtain

$$\delta_t \le \frac{L_u T}{\sqrt{1-\rho^2}}\Delta_U + \frac{L_\theta}{1-\rho}\|\theta_1 - \theta_2\|.$$

Since the output map is Lipschitz continuous, we know that

$$\|y_t^{(1)} - y_t^{(2)}\| = \|h(x_t^{(1)}) - h(x_t^{(2)})\| \le L_h\delta_t.$$

For brevity of notation, we define

$$\alpha = \frac{L_h L_\theta}{1-\rho}, \quad \beta = \frac{L_h L_u T}{\sqrt{1-\rho^2}},$$

which leads to the rewritten inequality

$$\|y_t^{(1)} - y_t^{(2)}\| \le \alpha\|\theta_1 - \theta_2\| + \beta\Delta_U.$$

Now, from the matrix profile, we can bound the interleaved embedding difference by

$$\begin{aligned}
\Delta Z_t^{(1,2)} &=: \|Z_t(\theta_1, U^{(1)}) - Z_t(\theta_2, U^{(2)})\|^2 \\
&= \sum_{j=0}^{m-1} \left[\|y_{t+j\tau}^{(1)} - y_{t+j\tau}^{(2)}\|^2 + \|u_{t+j\tau}^{(1)} - u_{t+j\tau}^{(2)}\|^2\right].
\end{aligned}$$

For the output component, we get

$$\sum_{j=0}^{m-1} \|y_{t+j\tau}^{(1)} - y_{t+j\tau}^{(2)}\|^2 \le m[\alpha\|\theta_1 - \theta_2\| + \beta\Delta_U]^2.$$

For the control terms, using Definition 1, we get

$$\sum_{j=0}^{m-1} \|u_{t+j\tau}^{(1)} - u_{t+j\tau}^{(2)}\|^2 \le mT\Delta_U^2.$$

Therefore,

$$\begin{aligned}
&\|\Delta Z_t^{(1,2)}\|^2 \le m[(\alpha\|\theta_1 - \theta_2\| + \beta\Delta_u)^2 + T\Delta_U^2] \\
&\le m[\alpha^2\|\theta_1 - \theta_2\|^2 + 2\alpha\beta\|\theta_1 - \theta_2\|\Delta_U + (\beta^2 + T)\big(\Delta_U\big)^2].
\end{aligned}$$

Using $\sqrt{\xi_1^2 + 2\xi_1\xi_2 + \xi_2^2} = \xi_1 + \xi_2$ for $\xi_1, \xi_2 \ge 0$ and the inequality $\sqrt{\xi_1^2 + \xi_2^2} \le |\xi_1| + |\xi_2|$, we get

$$\|\Delta Z_t^{(1,2)}\| \le \sqrt{m}[\alpha\|\theta_1 - \theta_2\| + (\beta + \sqrt{T})\Delta_U].$$

By definition of the matrix profile distance and the fact that the minimum is at most the same-time distance:

$$\begin{aligned}
\mathcal{D}_{\text{MP}}(\theta_1, \theta_2) &\le \frac{1}{N}\sum_{t=0}^{N-1} \|Z_t(\theta_1, U^{(1)}) - Z_t(\theta_2, U^{(2)})\| \\
&\le \sqrt{m}[\alpha\|\theta_1 - \theta_2\| + (\beta + \sqrt{T})\Delta_U].
\end{aligned}$$

Setting $C_1 = \sqrt{m}\alpha$ and $C_2 = \sqrt{m}(\beta + \sqrt{T})$ yields the constants

$$\begin{aligned}
C_1 &= \sqrt{m} \cdot \frac{L_h L_\theta}{1-\rho}, \\
C_2 &= \sqrt{m}\left(\frac{L_h L_u T}{\sqrt{1-\rho^2}} + \sqrt{T}\right),
\end{aligned}$$

and proves (12).

Now suppose $\Delta_U^{(i,j)} \le \epsilon_{ij} \le \epsilon$ for all $i, j \in \{1,2,3\}$. Then:

$$\begin{aligned}
\mathcal{D}_{\text{MP}}(\theta_1, \theta_2) &\le C_1\|\theta_1 - \theta_2\| + C_2\epsilon_{12}, \\
\mathcal{D}_{\text{MP}}(\theta_1, \theta_3) &\le C_1\|\theta_1 - \theta_3\| + C_2\epsilon_{13}.
\end{aligned}$$

Therefore, $\mathcal{D}_{\text{MP}}(\theta_1, \theta_2) < \mathcal{D}_{\text{MP}}(\theta_1, \theta_3)$ holds if:

$$C_1\|\theta_1 - \theta_2\| + C_2\epsilon_{12} < C_1\|\theta_1 - \theta_3\| + C_2\epsilon_{13},$$

or

$$\|\theta_1 - \theta_2\| + \frac{C_2}{C_1}(\epsilon_{12} - \epsilon_{13}) < \|\theta_1 - \theta_3\|,$$

which implies $\|\theta_1 - \theta_2\| + \frac{C_2}{C_1}\epsilon < \|\theta_1 - \theta_3\|$. □

We can also relax the noise-free assumption using straightforward arguments. The following proposition hints that the method degrades gracefully with noise, with the noise contribution scaling with $\sqrt{m}$.

Lemma 1. *Let $\eta^{(1)}, \eta^{(2)} \in \mathbb{R}^d$ be independent Gaussian random vectors with $\eta^{(i)} \sim \mathcal{N}(0, \sigma^2 I_d)$. Then, particularly for large d,*

$$\mathbb{E}\left[\left\|\eta^{(1)} - \eta^{(2)}\right\|\right] \approx \sqrt{2d}\,\sigma.$$

More precisely, the norm concentrates sharply around this value.

Proof. Since $\eta^{(1)}$ and $\eta^{(2)}$ are independent, their difference has distribution $\chi = \eta^{(1)} - \eta^{(2)} \sim \mathcal{N}(0, 2\sigma^2 I_d)$. For a Gaussian vector $\chi \sim \mathcal{N}(0, c^2 I_d)$, the Euclidean norm $\|\chi\|$ concentrates sharply around $c\sqrt{d}$ for large d [23, Theorem 3.1.1]. This concentration implies that the expectation is well-approximated by this value, i.e., $\mathbb{E}[\|\chi\|] \approx c\sqrt{d}$. Applying this approximation with $c = \sqrt{2}\sigma$ gives $\mathbb{E}[\|\chi\|] \approx (\sqrt{2}\sigma)\sqrt{d} = \sqrt{2d}\,\sigma$[2]. $\square$

Proposition 1 (Robustness to Measurement Noise)**.** *Under the assumptions of Theorem 1 with noisy measurements, we have:*

$$\mathbb{E}[\mathcal{D}_{MP}(\theta_1, \theta_2)] \leq C_1\|\theta_1 - \theta_2\| + C_2\Delta_U + C_3\sigma,$$

where $C_3 = 2\sqrt{mn_y}$ and the expectation is over the measurement noise.

Proof. Let $\tilde{Z}_t(\theta_i, U^{(i)})$ denote the noisy interleaved embedding constructed from $\tilde{y}_t^{(i)}$ and $u_t^{(i)}$, and let $Z_t(\theta_i, U^{(i)})$ denote the corresponding noise-free embedding to maintain parity with the proof of Theorem 1. Since noise enters additively in the outputs, we have

$$\|\Delta\tilde{Z}_t^{(1,2)}\| \leq \|\Delta Z_t^{(1,2)}\| + \|\eta_t^{(1)} - \eta_t^{(2)}\|.$$

Taking expectations and using the independence of noise from the trajectories,

$$\mathbb{E}\left[\left\|\Delta\tilde{Z}_t^{(1,2)}\right\|\right] \leq \left\|\Delta Z_t^{(1,2)}\right\| + \mathbb{E}\left[\left\|\eta_t^{(1)} - \eta_t^{(2)}\right\|\right].$$

For independent Gaussian noise vectors $\eta_t^{(i)} \in \mathbb{R}^{mn_y}$ with variance σ^2,

$$\mathbb{E}\left[\left\|\eta_t^{(1)} - \eta_t^{(2)}\right\|\right] \approx \sqrt{2mn_y}\,\sigma,$$

by invoking Lemma 1. Applying the deterministic bound from Theorem 1 and summing over all subsequences gives

$$\mathbb{E}[\mathcal{D}_{\text{MP}}(\theta_1, \theta_2)] \leq C_1\|\theta_1 - \theta_2\| + C_2\Delta_U + 2\sqrt{mn_y}\,\sigma,$$

which proves the claim with $C_3 = 2\sqrt{mn_y}$. $\square$

V. Results

In this section, we demonstrate the potential of our proposed approach on two examples: a linear harmonic oscillator that conforms to the assumptions (A1)–(A4), and a nonlinear hysteretic oscillator system.

Example 1. Linear oscillator with scalar output

We validate the proposed matrix profile-based approach on a family of forced damped harmonic oscillators described by:

$$\dot{x_1} = x_2,\ \dot{x_2} = \frac{1}{m_p}\left(u(t) - c_p x_2 - k_p x_1\right),\ y = x_1.$$

where x_1 denotes displacement, x_2 denotes velocity, and the parameter vector $\theta = [m_p, c_p, k_p]^\top$ consists of mass, damping coefficient, and spring stiffness, respectively. The control input $u(t) = A\sin(\omega t)$ represents sinusoidal forcing with amplitude A and frequency ω.

We generated $N_{\text{src}} = 32$ source systems with parameters sampled from $m_p \in [0.5, 1.0]$, $c_p \in [0.1, 0.2]$, and $k_p \in [1.0, 2.0]$, and $N_{\text{qry}} = 8$ test systems sampled from the same distributions. Each system was simulated for $T = 40$ seconds with time step $\Delta t = 0.01$ using forward Euler integration, resulting in trajectories of length 4000. Initial conditions x_0 and forcing parameters A, ω were randomly sampled for each system to ensure diversity in operating conditions.

For each query system, we extracted a subsequence of length $T_t = 500$ time steps from a random starting point, yielding an interleaved trajectory of dimension $2T_t = 1000$. We set the subsequence length to $m = 201$ (odd, to begin and end with output samples) and stride $\sigma = 2$ to maintain alignment consistency. As a baseline, we compared against dynamic time warping (DTW) applied to output trajectories only, using the FastDTW algorithm with default Euclidean distance [24]. DTW was chosen instead of only Euclidean distance comparison because it explicitly accounts for temporal misalignment and phase variability between trajectories. Unlike Euclidean distance, which assumes perfect pointwise alignment (which is intractable to compute when trajectory lengths differ), DTW nonlinearly warps the time axis to achieve optimal alignment between sequences, making it robust to local timing differences that frequently arise in dynamical systems.

Figure 1 illustrates the results for a representative query system. The top two panels show the state trajectories for the query system (black dashed), the closest system identified by the matrix profile method (blue, system index 117), and the closest system identified by DTW (red, system index 267). The proposed method successfully identifies a source system whose dynamics closely track the query system across both displacement and velocity states, particularly capturing the phase and amplitude characteristics during the initial transient period ($t \in [0, 500]$). In contrast, the DTW-selected system exhibits significant phase misalignment and amplitude discrepancies, especially in the velocity state (x_2). This divergence is quantified in the bottom panel, which displays the mean absolute error (MAE) between the query and source trajectories, respectively. The matrix profile method achieves a final cumulative MAE of approximately 0.14, compared to 0.33 for DTW, which is a reduction of over 57%.

To verify that similarity in the data space corresponds to similarity in the parameter space, we computed the Euclidean distance between the query parameters $\theta^{(q)}$ and the identified source parameters. For the representative query system shown in Figure 1, the matrix profile method identified a source with parameter distance $\|\theta^{(q)} - \theta^{(117)}\|_2 = 0.0824$, while the DTW method identified a source with distance $\|\theta^{(q)} - \theta^{(267)}\|_2 = 0.1953$. This confirms that the proposed method captures behavioral similarity in trajectories but also correlates with proximity in the underlying parameter manifold.

Across 100 query systems that are randomly (and uniquely) sampled, the proposed pan-MP approach achieved a mean parameter-space deviation[3] of 0.383 ± 0.230, compared to 0.544 ± 0.254 for DTW, while reducing runtime from approximately 67 s to 3 s per query. While this is encouraging, it is important to report that in 20% instances, the MP approach is outperformed by using DTW distances. These are usually when the query system parameters are quite close to one of the source systems, and therefore, the DTW distance is very small. For most cases, however, MP outperforms DTW by a significant margin.

[2]The exact value is $\mathbb{E}[\|\chi\|] = 2\sigma\frac{\Gamma((d+1)/2)}{\Gamma(d/2)}$, which is asymptotically equivalent to $\sqrt{2d}\sigma$ as $d \to \infty$.

[3]This is calculated by taking the mean absolute error pointwise in time between the true query system parameters and the source system parameters identified to be the most similar.

TABLE I

MEAN $\pm$ STANDARD DEVIATION OF PARAMETER-SPACE DISTANCE $\|\theta_{\text{BEST}} - \theta_q\|$ FOR DIFFERENT QUERY AND SUBSEQUENCE LENGTHS.

QUERY LENGTH	$m = 51$	$m = 101$	$m = 201$	$m = 301$	$m = 401$
250	0.502 $\pm$ 0.238	0.321 $\pm$ 0.167	0.316 $\pm$ 0.177	–	–
500	0.389 $\pm$ 0.228	0.459 $\pm$ 0.249	0.345 $\pm$ 0.187	0.366 $\pm$ 0.228	0.367 $\pm$ 0.223
1000	0.363 $\pm$ 0.186	0.347 $\pm$ 0.182	0.344 $\pm$ 0.180	0.289 $\pm$ 0.171	0.268 $\pm$ 0.163
2000	0.410 $\pm$ 0.201	0.416 $\pm$ 0.240	0.273 $\pm$ 0.174	0.291 $\pm$ 0.140	0.362 $\pm$ 0.197
4000	0.346 $\pm$ 0.158	0.296 $\pm$ 0.229	0.344 $\pm$ 0.205	0.321 $\pm$ 0.195	0.381 $\pm$ 0.201

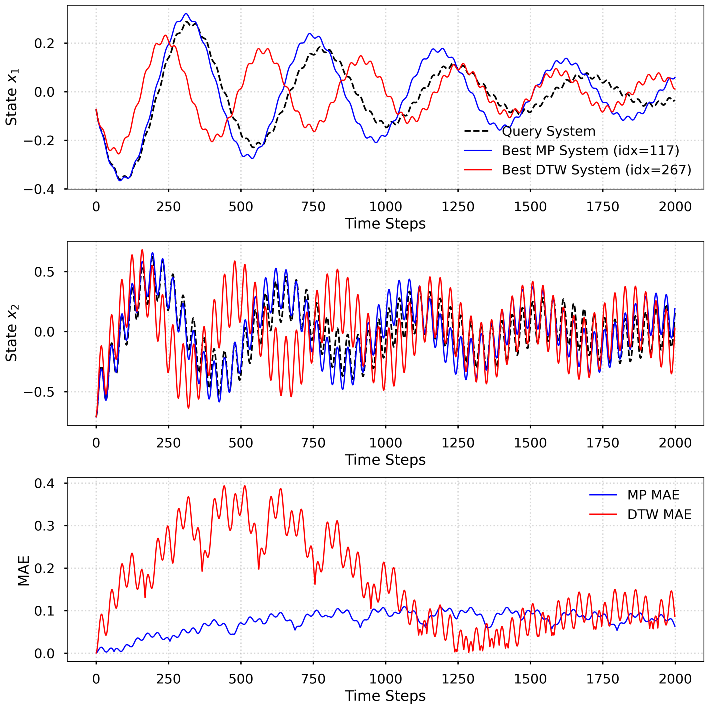

Fig. 1. Comparison of matrix profile (MP) and dynamic time warping (DTW) similarity methods on a forced damped harmonic oscillator. Top two panels show state trajectories for displacement (x_1) and velocity (x_2). Bottom panel shows the mean absolute error (MAE) over time. The MP method achieves superior trajectory matching and lower cumulative error.

From Table I, we observe that the mean parameter-space deviation decreases with increasing subsequence length m up to approximately $m = 200$–300, beyond which improvements saturate. Longer query lengths generally yield more stable estimates (lower standard deviations), indicating that the method benefits from richer temporal context. The smallest errors are consistently observed for pairs around $(1000, 301)$ and $(2000, 201)$, suggesting an optimal balance between subsequence size and query duration. Notably, short queries (≤ 500) exhibit higher variability, reflecting greater sensitivity to initialization and local dynamics, which is reasonable as we assume no model information; thus, providing very limited data will raise uncertainty of the MP estimate.

Example 2. Nonlinear system with vector output

We further evaluate our matrix profile approach on a family of forced Bouc-Wen oscillators, a canonical nonlinear system for modeling hysteretic restoring forces:

$$\dot{x_1} = x_2$$
$$\dot{x_2} = \frac{1}{m_b}\Big(u(t) - c_b\, x_2 - k_b\, x_1 - x_3\Big)$$
$$\dot{x_3} = \alpha_b x_2 - \beta_b |x_2|\, |x_3|^{\nu_b - 1} x_3 - \gamma_b x_2 |x_3|^{\nu_b} + \delta_b x_2 |x_3|^{\nu_b}$$

where x_1 is displacement, x_2 is velocity, x_3 encodes the internal hysteretic variable, $y = x$, and the parameter vector $\theta = [m_b, c_b, k_b, \alpha_b, \beta_b, \gamma_b, \delta_b]^\top$; $\nu_b = 1$. The input $u(t)$ is a parameterized sinusoidal signal, and initial conditions are randomly sampled for each run.

A set of $N_{\text{src}} = 400$ source systems was created, each with random parameters from compact intervals: $m \in [0.5, 1.0]$, $c \in [0.1, 0.2]$, $k \in [1.0, 2.0]$, and $\alpha, \beta, \gamma, \delta \in [0.5, 1.5]$. Each source and each of $N_{\text{qry}} = 100$ query systems was simulated for $T = 40$s at $\Delta t = 0.01$, producing trajectories of length $4,000$. For similarity search, query subsequences of length $T_t = 500$ were randomly extracted and interleaved with their respective inputs before running the matrix profile computation with stride $\sigma = 2$ and window length $m = 201$.

The matrix profile (MP) ranking was benchmarked against direct Euclidean distance (ED) computed between output trajectory segments (state x_1 only). The cumulative mean absolute error (MAE) across all three states was computed to quantify trajectory matching quality.

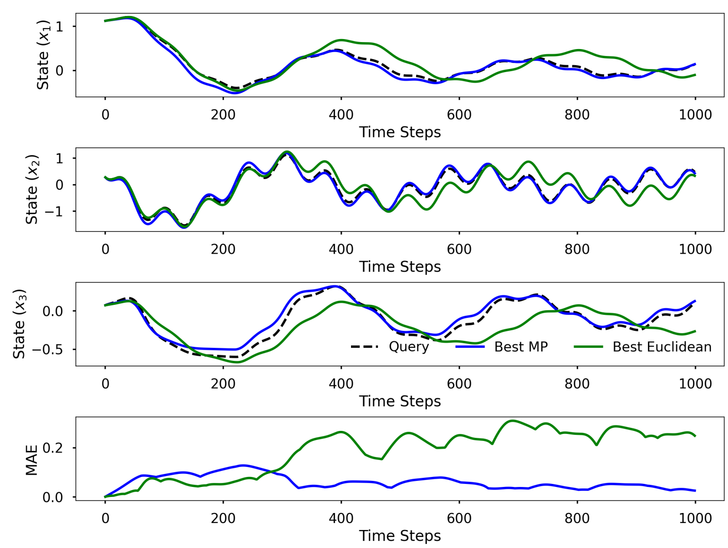

Fig. 2. Comparison of matrix profile (MP) and Euclidean distance (ED) similarity methods on a Bouc-Wen hysteretic oscillator. Top three panels show state trajectories for displacement (x_1), velocity (x_2), and hysteretic variable (x_3).

Figure 2 presents a representative comparison. The three upper panels show state trajectories for displacement (x_1), velocity (x_2), and the hysteretic variable (x_3). The query system (black dashed line) is compared against the best-matching source identified by the matrix profile method (blue, system index not shown) and by the Euclidean distance baseline (green).

Both methods successfully identify source systems that track the query trajectory reasonably well across all three states. However, notable differences emerge in trajectory fidelity. The MP-selected system exhibits closer alignment with the query, particularly in the velocity state (x_2) where it captures both phase and amplitude characteristics more accurately during the entire time-span shown. The hysteretic variable (x_3) also shows tighter tracking for the MP selection. The bottom panel quantifies this difference via MAE computed across all three outputs. While the Euclidean MAE grows monotonically and plateaus around 0.30, the MP method maintains low error consistently along the trajectory.

TABLE II
COMPARISON OF MATRIX PROFILE (MP) AND EUCLIDEAN DISTANCE METHODS

METHOD	θ ERROR (Mean $\pm$ Std)	y ERROR (Mean $\pm$ Std)	CPU TIME (seconds)
MP	0.920 ± 0.174	14.35 ± 9.34	$\sim$39.5
Euclid	1.034 ± 0.181	21.77 ± 17.89	$\sim$0.003
DTW	1.012 ± 0.159	18.43 ± 11.44	$\sim$275.8

From Table II, the MP method achieves noticeably lower parameter and trajectory reconstruction errors compared to the direct Euclidean baseline, indicating its superior ability to capture temporally aligned structure. Although MP incurs higher computational cost, the substantial gain in accuracy justifies the additional runtime for most practical settings. We do not compare in detail with DTW in this example because it is an order of magnitude slower than MP, so Euclidean is the better baseline.

VI. CONCLUSIONS

We proposed a matrix-profile–based framework for quantifying similarity between nonlinear dynamical systems directly from input–output data, without requiring state estimation or model-based parameter identification. Theoretical investigations established performance bounds and robustness to noise, while experiments on linear and nonlinear oscillators demonstrated significant improvements over Euclidean and DTW baselines in both parameter-space and trajectory accuracy. Future work will integrate the method into adaptive and meta-learning pipelines for data-efficient control transfer, and extend it to multivariate, stochastic, and high-dimensional systems.

REFERENCES

[1] A. Shahzad, E. C. Kerrigan, and G. A. Constantinides, "A warm-start interior-point method for predictive control," in *UKACC International Conference on Control*. IEEE, 2010, pp. 1–6.

[2] K. Niu, M. Zhou, C. T. Abdallah, and M. Hayajneh, "Deep transfer learning for system identification using long short-term memory neural networks," *IEEE Transactions on Automatic Control*, 2022, arXiv:2204.03125.

[3] L. Ljung, *System Identification: Theory for the User*, 2nd ed. Upper Saddle River, NJ: Prentice Hall PTR, 1999.

[4] W. S. Levine, *The Control Handbook (three volume set)*. CRC press, 2018.

[5] R. Pintelon and J. Schoukens, *System Identification: A Frequency Domain Approach*, 2nd ed. Hoboken, NJ: John Wiley & Sons, 2012.

[6] C. D. Persis and P. Tesi, "Formulas for data-driven control: Stabilization, optimality, and robustness," *IEEE Transactions on Automatic Control*, vol. 65, no. 3, pp. 909–924, 2020.

[7] S. M. Richards, N. Azizan, J.-J. Slotine, and M. Pavone, "Adaptive-control-oriented meta-learning for nonlinear systems," in *Robotics: Science and Systems*, 2021.

[8] D. Muthirayan, D. Kalathil, and P. P. Khargonekar, "Meta-learning online control for linear dynamical systems," *IEEE Transactions on Automatic Control*, vol. 70, no. 6, pp. 4163–4169, 2025.

[9] C. Finn, P. Abbeel, and S. Levine, "Model-agnostic meta-learning for fast adaptation of deep networks," *International Conference on Machine Learning*, pp. 1126–1135, 2017.

[10] A. Chakrabarty, V. M. Deshpande, G. Wichern, and K. Berntorp, "Physics-constrained meta-learning for online adaptation and estimation in positioning applications," in *IEEE Conference on Decision and Control (CDC)*, 2024.

[11] J. Yan, A. Chakrabarty, A. Rupenyan, and J. Lygeros, "MPC of uncertain nonlinear systems with meta-learning for fast adaptation of neural predictive models," in *2024 IEEE 20th International Conference on Automation Science and Engineering (CASE)*. IEEE, 2024, pp. 1910–1915.

[12] L. Xin, L. Ye, G. Chiu, and S. Sundaram, "Learning dynamical systems by leveraging data from similar systems," *IEEE Transactions on Automatic Control*, 2025.

[13] F. Takens, "Detecting strange attractors in turbulence," in *Dynamical Systems and Turbulence, Warwick 1980: proceedings of a symposium held at the University of Warwick 1979/80*. Springer, 1981, pp. 366–381.

[14] J. Stark, "Delay embeddings for forced systems. I. Deterministic forcing," *Journal of Nonlinear Science*, vol. 9, no. 3, pp. 255–332, 1999.

[15] Y. Zhu, Z. Zimmerman, N. S. Senobari, C.-C. M. Yeh, G. Funning, A. Mueen, P. Brisk, and E. Keogh, "Matrix profile II: Exploiting a novel algorithm and GPUs to break the one hundred million barrier for time series motifs and joins," in *Proc. IEEE Int. Conf. on Data Mining (ICDM)*. IEEE, 2016, pp. 739–748.

[16] M. E. Taylor and P. Stone, "Transfer learning for reinforcement learning domains: A survey," *Journal of Machine Learning Research*, vol. 10, no. 7, pp. 1633–1685, 2009.

[17] L. Ljung, "Perspectives on system identification," *Annual Reviews in Control*, vol. 34, no. 1, pp. 1–12, 2010.

[18] N. Williams, I. Walters, J. Li, G. M. Matrone, S. Leng, G. Leen, S. Lowry, and M. Charalambides, "A general metric for the similarity of both stochastic and deterministic system dynamics," *Entropy*, vol. 23, no. 9, p. 1191, 2021.

[19] M. Ostrow, A. Eisen, L. Polk, D. Owens, E. Musselman, M. Musselman, I. R. Bar, J. Snider, and G. S. Escola, "Beyond geometry: Comparing the temporal structure of computation in neural circuits with dynamical similarity analysis," *Advances in Neural Information Processing Systems*, vol. 36, 2023.

[20] S. Zhang, Z. Ye, Y. Yan, Z. Song, Y. Wu, and J. Wu, "KoopSTD: Reliable similarity analysis between dynamical systems via approximating Koopman spectrum with timescale decoupling," in *International Conference on Machine Learning (ICML)*, 2025.

[21] C.-C. M. Yeh, Y. Zhu, L. Ulanova, N. Begum, Y. Ding, H. A. Dau, D. F. Silva, A. Mueen, and E. Keogh, "Matrix profile I: All pairs similarity joins for time series: A unifying view that includes motifs, discords and shapelets," in *Proc. IEEE Int. Conf. on Data Mining (ICDM)*, 2016, pp. 1317–1322.

[22] S. M. Law, "STUMPY: A Powerful and Scalable Python Library for Time Series Data Mining," *The Journal of Open Source Software*, vol. 4, no. 39, p. 1504, 2019.

[23] R. Vershynin, *High-Dimensional Probability: An Introduction with Applications in Data Science*, ser. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2018.

[24] S. Salvador and P. Chan, "Toward accurate dynamic time warping in linear time and space," *Intelligent data analysis*, vol. 11, no. 5, pp. 561–580, 2007.