Commun. Comput. Phys.
doi: 10.4208/cicp.OA-2025-0042

Vol. **39**, No. 3, pp. 941-968
March 2026

Reconstruction of Dynamical Systems Without Time Label

Zhijun Zeng[1], Chenglong Bao[2,3], Pipi Hu[4], Yi Zhu[2,3] and Zuoqiang Shi[2,3,*]

[1] *Department of Mathematical Sciences, Tsinghua University, China.*
[2] *Yau Mathematical Sciences Center, Tsinghua University, China.*
[3] *Yanqi Lake Beijing Institute of Mathematical Sciences and Applications, China.*
[4] *Microsoft Research AI4Science, China.*

Received 26 February 2025; Accepted (in revised version) 16 July 2025

Abstract. In this paper, we study the method to reconstruct dynamical systems from data without time labels. Data without time labels appear in many applications, such as molecular dynamics, single-cell RNA sequencing, etc. Reconstruction of dynamical system from time sequence data has been studied extensively. However, these methods do not apply if time labels are unknown. Without time labels, sequence data become distribution data. Based on this observation, we propose to treat the data as samples from a probability distribution and try to reconstruct the underlying dynamical system by minimizing the distribution loss, sliced Wasserstein distance more specifically. Extensive experiment results demonstrate the effectiveness of the proposed method.

AMS subject classifications: 65L09, 34A55, 93B30

Key words: Dynamical system recovery, data without time label, Wasserstein distance.

1 Introduction

Dynamical models are crucial for enhancing our comprehension of the natural world. By harnessing massive datasets to reveal the underlying governing equations that describe the behavior of complex physical systems, we can significantly advance our ability to model, simulate, and understand these systems across diverse scientific disciplines. It is a common situation that high-dimensional observations are generated by hidden dynamical systems operating within a low-dimensional space, a concept related to the manifold

*Corresponding author. *Email addresses:* zqshi@tsinghua.edu.cn (Z. Shi), zengzj22@mails.tsinghua.edu.cn (Z. Zeng)

hypothesis. [12]. Accordingly, reconstructing the underlying dynamical system is crucial for simulating real-world scientific systems and educing its mechanisms.

For an evolutionary system, the observations consist of trajectory data with time labels $\{(t_i, \boldsymbol{x}_i)\}_{i=1}^n$, typically modeled using a dynamic form $\boldsymbol{x}_{\boldsymbol{\theta}}(t)$ or its differential equation representation $\frac{d\boldsymbol{x}}{dt} = \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{\theta})$. However, in certain contexts—such as microscopy—technical limitations preclude the acquisition of precise time labels, and the observed data are instead obtained as unlabeled high-dimensional point clouds $\{\boldsymbol{x}_i\}_{i=1}^n$. Although reconstructing the hidden dynamics from such unlabeled data remains challenging, one can still postulate reasonable assumptions regarding the observation times and attempt to reconstruct the dynamics.

System identification

The conventional system identification task involves determining the form of the underlying system and estimating its parameters from trajectory data with time labels $\{(t_i, \boldsymbol{x}_i)\}_{i=1}^n$. Traditionally, the forward solver-based nonlinear least squares (FSNLS) method is a standard approach for system identification in models governed by differential equations, and its procedure can be summarized in four steps: (1) proposing an initial set of parameters, (2) solving the forward process on the collocation points using a numerical solver, (3) comparing the generated solution with observational samples and updating the parameters, and (4) iterating steps (2) and (3) until the convergence criteria are satisfied. As a well-established topic, [1] and [22] provide a comprehensive overview of extant results. As models become larger and demand greater realism in FSNLS, two main challenges persist—solver accuracy and the selection of the initial candidate parameter—which significantly influence the final estimate, yet fully satisfactory solutions remain elusive. Recently, the Neural Ordinary Differential Equation (ODE) approach and its derivatives [7, 17, 26, 40] have combined FSNLS with a neural network ansatz for the forcing term and provided a powerful tool for inferring the complex dynamics of physical systems. Subsequent efforts embed numerical integrators and semigroup constraints within neural networks, thereby markedly improving reconstruction accuracy [5, 30]. However, both FSNLS and Neural ODE update the model by computing pointwise residuals between simulated data and trajectory observations, which cannot be applied to unlabeled data.

To address these issue, an alternative approach, known as the sparse identification of nonlinear dynamics (SINDy) [3, 4], avoids forward computation by evaluating the nonlinear candidate basis functions on the given dataset and estimates the system parameters using sparse regression methods. Numerous subsequent studies have augmented the performance of this foundational concept by refining the estimation of differentiation and least square [11, 18, 24]. Both the strong-form and weak-form variants of SINDy rely on time label information to compute constraint equations, thereby precluding their direct application to unlabeled data.