Linear System Identification from Snapshot Data by Schrödinger Bridge

Kohei Morimoto KOHEI.MORIMOTO.73R@ST.KYOTO-U.AC.JP
Kenji Kashima KK@I.KYOTO-U.AC.JP
Graduate School of Informatics, Kyoto University, Kyoto, Japan

Editors: A. Abate, L. Balzano, N. Ozay, D. Panagou

Abstract

In many practical applications, such as single-cell RNA sequencing, individual entities cannot be tracked due to economic costs or experimental limitations, but population distributions can be observed at specific time points. Such snapshot measurements make traditional system identification challenging. This paper proposes a system identification method for linear Gaussian systems from snapshot measurement data using Schrödinger Bridges (SB). Our method employs an EM-like algorithm that alternates between trajectory estimation using SB and system parameter inference from the estimated trajectories. The linear Gaussian assumption for system states and noise allows us to exploit analytical solutions for the SB computation and parameter updates, enabling efficient computation. We also propose a data-driven method for estimating linear Gaussian SB, where parameters of marginal distributions are estimated incorporating dynamic constraints. In numerical simulations, we show our method achieves superior identification accuracy compared to moment-based baseline approaches and time efficiency compared to methods based on the nonlinear SB.

Keywords: Schrödinger Bridge; scRNA-seq; snapshot data; system identification

1. Introduction

In recent years, there has been a growing demand for analytical methods that can handle collective data where individual entities cannot be tracked, but population distributions can be observed at specific time points. This scenario appears in various fields such as wildlife migration and human flow analysis due to economic costs or experimental constraints. For instance, in wildlife migration studies, obtaining trajectory data for how each individual in a flock of sheep moves will be difficult, but obtaining the distribution of the flock at multiple time points will be relatively easier. Single-cell RNA sequencing data (scRNA-seq) represents another prominent case where such snapshot data analysis is essential. While scRNA-seq offers gene expression information at single-cell resolution, its fundamental limitation is that cells are destroyed during data collection, making it impossible to directly track gene expression trajectories over time. Instead, researchers must collect gene expression measurements from different groups of cells at various time points, with each cell being measured only once.

Recent studies have proposed using the Schrödinger Bridge (SB) problem for recovering the sample trajectories from snapshot data (Koshizuka and Sato (2023); Shen et al. (2024); Vargas et al. (2021b); Wang et al. (2024)). The SB estimates the most natural stochastic process that connects snapshot data observed at different time points given underlying system dynamics. Specifically, given marginal distributions, the SB computes the optimal transport path that minimizes the Kullback-Leibler (KL) divergence from the reference dynamics while satisfying the marginal constraints.

In this study, we propose a system identification method for linear Gaussian systems from snapshot data using SB. Our method implements an EM-like alternating optimization algorithm that repeats the trajectory estimation using SB and the estimation of the system parameters from the estimated trajectory. The key advantages of our approach are as follows. First, it can naturally handle snapshot data observed at irregular time intervals. This is particularly advantageous in biological experiments where regular sampling intervals are often difficult to maintain, and some time points may have missing observations. Second, our method achieves superior computational efficiency through analytical solutions. By assuming Gaussian distributions for system state and noise, we can use exact solutions for SB in linear Gaussian systems (Ito and Kashima (2024); Beghi (1996)). This allows us to obtain closed-form solutions for the trajectory estimation through SB and the system parameter updates in our EM-like algorithm, enabling computationally efficient estimation.

Although a similar approach for iteratively refining the reference dynamics has been proposed by Shen et al. (2024), their primary goal is to improve SB results rather than system identification. While they update system parameters during optimization, they neither focus on identification accuracy nor provide theoretical analysis for the estimation problem. Their method makes no assumptions about the system state distribution and relies on computationally intensive nonlinear fitting and gradient descent. In contrast, our method achieves efficient system identification by using analytical solutions for both SB problem and dynamics updates because of the linear Gaussian assumption.

We also propose a novel data-driven method for estimating SB for linear Gaussian systems. The original SB formulation for linear Gaussian systems requires precise knowledge of marginal distributions, which is often challenging to obtain from limited samples. To address this challenge, our method combines maximum likelihood estimation with SB costs to estimate marginal distributions. Our method can recover the most natural distributions and trajectories that are consistent with the underlying dynamics, even when individual time points lack sufficient samples for accurate statistical estimation.

Notation Let $\mathbb{S}_n^+$ and $\mathbb{S}_n^{++}$ denote the sets of n-dimensional positive semidefinite matrices and positive definite matrices, respectively. For matrices, $\|\cdot\|_F$ denotes the Frobenius norm. Let $\mathcal{P}(\mathcal{X})$ denote the set of all probability distributions over $\mathcal{X}$. Let $\mathcal{N}_n(\xi, \Sigma)$ denote the multivariate normal distribution with mean $\xi \in \mathbb{R}^n$ and covariance $\Sigma \in \mathbb{S}_n^+$. Let $\mathscr{N}_n$ denote the set of all n-dimensional multivariate normal distributions. Let $\mathrm{Cov}_P(X) = \mathbb{E}_P\left[(X - \mu_X)(X - \mu_X)^\top\right]$

2. Background: Discrete-time linear Gaussian Schrödinger Bridge

The Schrödinger Bridge (SB) determines the most natural stochastic process connecting two probability distributions. Specifically, it finds a stochastic process that minimizes the Kullback-Leibler (KL) divergence from a given reference process while satisfying the constraints on initial and terminal distributions. In this paper, we focus on discrete-time SB for linear systems, as they have analytical solutions (Ito and Kashima (2024)) and can be solved fast.

Let $\bar{\mu}_0, \bar{\mu}_N \in \mathbb{R}^n$ be the mean vectors and $\bar{\Sigma}_0, \bar{\Sigma}_N \in \mathbb{S}_n^{++}$ be the covariance matrices for the initial and terminal distributions, respectively. The SB problem minimizes the KL divergence of the probability measure $\hat{P}_{0:N} \in \mathcal{P}\left(\mathbb{R}^{n\times(N+1)}\right)$ with reference process $\tilde{P}_{0:N} \in \mathcal{P}\left(\mathbb{R}^{n\times(N+1)}\right)$ under fixed marginal distributions at initial and terminal times. In what follows, $\hat{P}_t$ denotes the marginal of the distribution $\hat{P}_{0:N}$ at time t, and so on.

Problem 1 (SB problem for linear dynamical systems) *Suppose that a reference process $\tilde{P}_{0:N}$ is given by the discrete-time linear systems*

$$\begin{aligned} \tilde{P}_{0:N}: \quad & \tilde{x}_{k+1} = A\tilde{x}_k + w_k \\ & w_k \sim \mathcal{N}(0, \Gamma),\ \tilde{x}_0 \sim \mathcal{N}(\tilde{\mu}_0, \tilde{\Sigma}_0) \end{aligned} \tag{1}$$

Here, $\tilde{x}_k \in \mathbb{R}^n$ is the state vector and $w_k \in \mathbb{R}^m$ is independent noise following the standard normal distribution. The matrix $A \in \mathbb{R}^{n\times n}$ is an invertible transition matrix, and $\Gamma \in \mathbb{S}_n^+$ is the noise covariance. The initial state has mean $\bar{\mu}_0$ and covariance $\bar{\Sigma}_0$. The time index is $k \in \mathcal{T} := [0, 1, \dots, N]$, where $N \in \mathbb{N}$ is a finite time horizon. Then, given $\bar{\mu}_0$, $\bar{\Sigma}_0$, $\bar{\mu}_N$, $\bar{\Sigma}_N$, find

$$\begin{aligned} \min_{\hat{P}_{0:N}\in\mathcal{P}\left(\mathbb{R}^{n\times(N+1)}\right)} \quad & \mathrm{KL}\left(\hat{P}_{0:N} \,\middle\|\, \tilde{P}_{0:N}\right) \\ \text{s.t.} \quad & \hat{P}_0 = \mathcal{N}(\bar{\mu}_0, \bar{\Sigma}_0),\ \hat{P}_N = \mathcal{N}(\bar{\mu}_N, \bar{\Sigma}_N). \end{aligned}$$

It is known that the optimal $\hat{P}_{0:N}$ for Problem 1 does not depend on $\tilde{\mu}_0, \tilde{\Sigma}_0$ and is given as the state process corresponding to the optimal policy of the following maximum entropy control problem.

Problem 2 (Maximum entropy optimal density control problem) *For A, Γ, $\bar{\mu}_0$, $\bar{\Sigma}_0$, $\bar{\mu}_N$, and $\bar{\Sigma}_N$, decompose Γ as $\Gamma = BB^\top$ and find*

$$\begin{aligned} \min_{\pi} \quad & \mathbb{E}\left[\sum_{k=0}^{N-1}\left(\frac{1}{2}\|u_k\|^2 - H(\pi_k(\cdot|x_k))\right)\right] \\ \text{s.t.} \quad & x_{k+1} = Ax_k + Bu_k, \quad u_k \sim \pi_k(\cdot|x) \\ & x_0 \sim \mathcal{N}(\bar{\mu}_0, \bar{\Sigma}_0), \quad x_N \sim \mathcal{N}(\bar{\mu}_N, \bar{\Sigma}_N). \end{aligned}$$

Here, $H(\pi_k(\cdot|x_k))$ represents the entropy of the control policy at time k.

The solution to this problem can be constructed by solving the following coupled Riccati equations for the matrix variables $\{P_k\}_{k=0}^N$ and $\{Q_k\}_{k=0}^N$:

$$\begin{aligned} P_{k+1} &= AP_kA^\top + BB^\top & \qquad Q_{k+1} &= AQ_kA^\top - BB^\top \\ \bar{\Sigma}_0^{-1} &= P_0^{-1} + Q_0^{-1} & \qquad \bar{\Sigma}_N^{-1} &= P_N^{-1} + Q_N^{-1}. \end{aligned}$$

An explicit formula for the solutions P_k and Q_k can be derived under certain conditions. Then, the optimal feedback control law is given by the following Gaussian distribution:

$$\pi_k^*(u \mid x) = \mathcal{N}(u \mid -(I + B^\top Q_{k+1}^{-1}B)^{-1}B^\top Q_{k+1}^{-1}A(x - \mu_k^*) + \bar{u}_k^*, (I + B^\top Q_{k+1}^{-1}B)^{-1})$$

where $\bar{u}_k^*$ is the optimal mean input, and μ_k^* is the optimal mean trajectory given by

$$\begin{aligned} \bar{u}_k^* &= B^\top \Phi(N, k+1)^\top G_{N,0}^{-1}(\bar{\mu}_N - \Phi(N, 0)\bar{\mu}_0) \\ \mu_k^* &= \Phi(k, 0)\bar{\mu}_0 + \sum_{s=0}^{k-1}\Phi(k, s+1)B\bar{u}_s^*. \end{aligned} \tag{2}$$

Here, $\Phi(k,l)$ represents the state transition matrix defined as A^{k-l}, and G_{k_1,k_0} is the controllability Gramian given by $\sum_{k=k_0}^{k_1-1}\Phi(k_1,k+1)BB^\top\Phi(k_1,k+1)^\top$. Finally, the state process generated by the optimal control law (2), which is also an optimal solution to Problem 1, is given by

$$\begin{aligned} &x^*_{k+1} = A_{Q,k}x^*_k + B_{Q,k}w_k,\ w_k \sim \mathcal{N}(0,I),\ x^*_0 \sim \mathcal{N}(\bar{\mu}_0, \bar{\Sigma}_0) \\ &A_{Q,k} := A - B(I + B^\top Q_{k+1}^{-1}B)^{-1}B^\top Q_{k+1}^{-1}A,\ B_{Q,k} := B(I + B^\top Q_{k+1}^{-1}B)^{-1/2}. \end{aligned} \tag{3}$$

It is worth mentioning that this procedure is applicable to the following multi-marginal problem:

Problem 3 (Multi-marginal Schrödinger Bridge problem) *Given $\tilde{P}_{0:N}$ in Problem 1, μ_{t_i}, Σ_{t_i} for $t_i \in \mathcal{T}_S \subset [0,1,\dots,N]$, find*

$$\min_{\hat{P}_{0:N}\in\mathcal{P}\left(\mathbb{R}^{n\times(N+1)}\right)} \quad \mathrm{KL}\left(\hat{P}_{0:N} \,\middle\|\, \tilde{P}_{0:N}\right) \quad \text{s.t. } \hat{P}_{t_i} = \mathcal{N}(\mu_{t_i}, \Sigma_{t_i}).$$

It is straightforward to solve this problem by dividing it into adjacent two time interval problems and solving Problem 1 in each interval, which is justified by the chain rule decomposition property of KL divergence.

3. Method

3.1. Linear system identification by Schrödinger bridge

In this section, we describe a SB-based linear system identification method (SBID) from snapshot data. Consider we have snapshot data consisting of n_t samples $\{x_t^i\}_{i=1}^{n_t}$ drawn from the marginal of $\hat{P}_{0:N}$ at measurement times $t \in \mathcal{T}_S := [t_0, t_1, ..., t_{K-1}]$. In this section, we investigate the case where the sample size is sufficiently large and the means and covariances at each time point can be estimated with high precision. We first calculate the means μ_{t_i} and covariances Σ_{t_i} of the marginals at each measurement time in advance. The SBID can be formulated as follows.

Problem 4 (SBID) *Given $\tilde{\mu}_0$, $\tilde{\Sigma}_0$, μ_{t_i}, Σ_{t_i} for $t_i \in \mathcal{T}_S$, find*

$$\min_{A,\Gamma} \min_{\hat{P}_{0:N}\in\mathcal{P}\left(\mathbb{R}^{n\times(N+1)}\right)} \mathrm{KL}\left(\hat{P}_{0:N} \,\middle\|\, \tilde{P}_{0:N}\right) \quad \text{s.t. } \hat{P}_{t_i} = \mathcal{N}\left(\mu_{t_i}, \Sigma_{t_i}\right) \quad (t_i \in \mathcal{T}_S)$$

where A and Γ determine $\tilde{P}_{0:N}$ by Eq. (1) with $\tilde{P}_0 = \hat{P}_0$.

The difference from Problem 3, multi-marginal SB, is that the parameters A and Γ of the reference distribution $\tilde{P}_{0:N}$ are estimated simultaneously. From the discussion below, we can confirm that the optimal solution does not depend on $\tilde{\mu}_0$ and $\tilde{\Sigma}_0$.

Our proposed algorithm for solving Problem 4 is in Alg. 1. In this algorithm, we use an EM (Expectation-Maximization) approach, alternating between optimization of $\hat{P}_{0:N}$ and (A,Γ). In the initialization step, we set initial values for the system parameters A and Γ. In the E-step, we estimate $\hat{P}_{0:N}$ given the parameters A and Γ. With fixed dynamics parameters, this problem reduces to Problem 3, which can be solved using the methods described in Sec. 2. The optimal process can

be represented as the time-variant linear systems given by Eq. (3). In the M-step, we optimize A and Γ by minimizing the KL divergence between $\tilde{P}_{0:N}$ and $\hat{P}_{0:N}$, which can be decomposed as

$$\begin{aligned}
&\mathrm{KL}\left(\hat{P}_{0:N} \,\middle\|\, \tilde{P}_{0:N}\right) = \mathbb{E}_{\hat{P}_{0:N}}\left[\ln \tilde{P}\left(x_1 \mid x_0\right) + \ln \tilde{P}\left(x_2 \mid x_1\right) \cdots + \ln \tilde{P}\left(x_N \mid x_{N-1}\right)\right] + \text{const.} \\
&= \frac{1}{2}\sum_{i=0}^{N-1} \mathbb{E}_{\hat{P}_{0:N}}\left[(x_{i+1} - Ax_i)^\top \Gamma^{-1}(x_{i+1} - Ax_i) - \ln\det\Gamma\right] + \text{const.} \\
&= \frac{1}{2}\sum_{i=0}^{N-1} \mathrm{tr}\left(\Gamma^{-1}\left(\mathbb{E}_{\hat{P}_{0:N}}\left[x_{i+1}x_{i+1}^\top - x_{i+1}x_i^\top A^\top - Ax_ix_{i+1}^\top - Ax_ix_i^\top A^\top\right]\right)\right) \\
&\qquad - \frac{N}{2}\ln\det\Gamma + \text{const.}
\end{aligned}$$

By minimizing this KL divergence for A, Γ, the update formula for A, Γ is derived as follows

$$A_{\text{new}} = \left(\sum_{n=0}^{N-1} \mathbb{E}_{\hat{P}_{0:N}}[x_{n+1}x_n^\top]\right)\left(\sum_{n=0}^{N-1} \mathbb{E}_{\hat{P}_{0:N}}[x_nx_n^\top]\right)^{-1} \tag{4}$$

$$\begin{aligned}
\Gamma_{\text{new}} = \frac{1}{N}\sum_{n=1}^{N}\Big(&\mathbb{E}_{\hat{P}_{0:N}}[x_nx_n^\top] - A_{\text{new}}\mathbb{E}_{\hat{P}_{0:N}}[x_{n-1}x_n^\top] \\
&- \mathbb{E}_{\hat{P}_{0:N}}[x_nx_{n-1}^\top]A_{\text{new}}^\top + A_{\text{new}}\mathbb{E}_{\hat{P}_{0:N}}[x_{n-1}x_{n-1}^\top]A_{\text{new}}^\top\Big)
\end{aligned} \tag{5}$$

Algorithm 1 SBID

Require:
Snapshot data $\{x_k^i\}_{i=1}^{n_k}$ at times $t_k \in \mathcal{T}_S$
Algorithm:
Calculate sample moments μ_{t_i}, Σ_{t_i}
Initialize A, Γ
Repeat until convergence:
E-step:
Solve Problem 3 with $\tilde{P}$ determined by current A and Γ
M-step: Update parameters
$A^{(n+1)} \leftarrow$ Eq. (4).
$\Gamma^{(n+1)} \leftarrow$ Eq. (5).
$n \leftarrow n+1$

The moments required in Eq. (4) and (5) can be analytically calculated using Eq. (3). Since Eq. (3) is linear Gaussian systems, the moments can be computed recursively: The first moment (mean) follows $\mu_{k+1} = A_{Q,k}\mu_k$ with $\mu_0 = \bar{\mu}_0$, and the second moment follows $\Sigma_{k+1} = A_{Q,k}\Sigma_k A_{Q,k}^\top + B_{Q,k}B_{Q,k}^\top$ with $\Sigma_0 = \bar{\Sigma}_0$. From these relations, we obtain $\mathbb{E}[x_nx_n^\top] = \Sigma_n + \mu_n\mu_n^\top$ and $\mathbb{E}[x_{n+1}x_n^\top] = A_{Q,n}(\Sigma_n + \mu_n\mu_n^\top)$. All these moments can be calculated analytically, enabling efficient computation of the parameter updates in Eq. (4) and (5).

By substituting A_{new}, Γ_{new} into the KL term, the minimum value is decomposed as

$$N\ln\det\left(\frac{1}{N}\sum_{i=0}^{N-1}\left(\mathbb{E}_{\hat{P}_{0:N}}\left[x_{i+1}x_{i+1}^\top\right] - A_{\text{new}}\mathbb{E}_{\hat{P}_{0:N}}\left[x_ix_i^\top\right](A_{\text{new}})^\top\right)\right) - H(\hat{P}_{0:N}). \tag{6}$$

Thus, minimizing the KL divergence with respect to $\hat{P}_{0:N}$ is equivalent to minimizing the moment consistency violation while maximizing the entropy $H(\hat{P}_{0:N})$.

3.2. Identification from small sample size data

In the previous section, we assumed that there were enough samples and that the marginal moments could be estimated with sufficient accuracy. This moment estimation becomes challenging when the sample size is small. To address this challenge, we propose a maximum likelihood SB problem (ML-SB) for linear Gaussian systems and its application to system identification (ML-SBID). Instead of imposing strict constraints on the marginal distribution as in Problem 1, we consider finding a stochastic process that both fits the observed snapshot data and the reference process. Precisely, we incorporate the likelihood of the snapshot data into the cost function in addition to the KL divergence cost of SB. The ML-SBID approach offers robustness when dealing with small sample sizes because it incorporates the likelihood of the observed data rather than relying on estimated moments.

Algorithm 2 ML-SB

Input:
Snapshot data $\{x_k^i\}_{i=1}^{n_k}$
Reference distribution parameters A, B
Initialization:
Initialize S, μ
Iterations: Repeat until convergence
1: Update non-diagonal blocks of S by solving SB problem (3) with marginals $\{S_{kk}\}_{k=1}^{K}$, μ_k
2: Update mean μ by Eq. (13)
3: **For** $k = 0$ **to** K:
4: Update marginal covariances S_{kk} by minimizing Eq. (12)

As in the previous section, we have snapshot data $\{x_t^i\}_{i=1}^{n_t}$ at measurement times $t \in \mathcal{T}_S := [t_0, t_1, ..., t_{K-1}]$. Then, ML-SBID minimizes the sum of the negative log-likelihoods of the samples and KL divergence.

Problem 5 (ML-SBID) *Given* $\{x_t^i\}_{i=1}^{n_t}$, *find*

$$\min_{\Gamma, A} \min_{\hat{P}_{0:N} \in \mathcal{N}_{n\times(N+1)}} -\sum_{t\in\mathcal{T}_S}\sum_{k=1}^{n_t} \ln \hat{P}_t(x_t^k) + \lambda \,\mathrm{KL}\left(\hat{P}_{0:N} \,\middle\|\, \tilde{P}_{0:N}\right)$$

where A *and* Γ *determine* $\tilde{P}_{0:N}$ *by Eq.* (1) *with* $\tilde{P}_0 = \hat{P}_0$.

Note that we set the initial mean $\bar{\mu}_0$ and covariance $\bar{\Sigma}_0$ of $\tilde{P}_0$ the same as that of $\hat{P}_0$. Similar to SBID, the ML-SBID problem can be solved using an EM approach by alternating between optimizing $\hat{P}_{0:N}$ and (A, Γ). For the M-step, i.e., the optimization of (A, Γ) with fixed $\hat{P}_{0:N}$, the update formulas for A and Γ are the same as in Eq. (4) and (5).

In what follows, we discuss the E-step, i.e., the optimization of $\hat{P}_{0:N}$ for a fixed (A, Γ):

Problem 6 (ML-SB) *Given* $\{x_t^i\}_{i=1}^{n_t}$ *for* $t \in \mathcal{T}_S$, *and system* (A, Γ), *find*

$$\min_{\hat{P}_{0:N} \in \mathcal{N}_{n\times(N+1)}} -\sum_{t\in\mathcal{T}_S}\sum_{k=1}^{n_t} \ln \hat{P}_t(x_t^k) + \lambda \,\mathrm{KL}\left(\hat{P}_{0:N} \,\middle\|\, \tilde{P}_{0:N}\right)$$

where A, Γ *and* $\hat{P}_{0:N}$ *determine* $\tilde{P}_{0:N}$ *by Eq.* (1) *with* $\tilde{P}_0 = \hat{P}_0$.

For solving Problem 6, we perform alternating optimization between the marginal distributions $\{\hat{P}_t : t \in \mathcal{T}_S\}$, and the path connecting them. When the marginal distributions are fixed, the optimization of the path is equivalent to Problem 3. In the following, we will discuss the optimization of the

marginal distributions when the path is fixed. The component of the KL divergence term that relates to the marginal distribution is

$$\mathrm{KL}\left(\hat{P}_{0:N} \,\middle\|\, \tilde{P}_{0:N}\right) = \mathrm{KL}\left(\hat{P}_{\mathcal{T}_S} \,\middle\|\, \tilde{P}_{\mathcal{T}_S}\right) + \text{const.} \tag{7}$$

where $\hat{P}_{\mathcal{T}_S}$ and $\tilde{P}_{\mathcal{T}_S}$ respectively represent the marginal of $\hat{P}_{0:N}$ and $\tilde{P}_{0:N}$ at time $\mathcal{T}_S$. The reference $\tilde{P}_{\mathcal{T}_S}$ can be written as $\mathcal{N}(\tilde{\mu}, \tilde{\Xi}^{-1})$ from Eq. (1), where $\tilde{\mu}$ is the mean vector

$$\tilde{\mu} := \begin{bmatrix} \bar{\mu}_0^\top & (A^{t_1-t_0}\bar{\mu}_0)^\top & \cdots & (A^{t_K-t_0}\bar{\mu}_0)^\top \end{bmatrix}^\top \tag{8}$$

and $\tilde{\Xi}$ is the block tridiagonal precision matrix

$$\tilde{\Xi} := \begin{bmatrix} \bar{\Sigma}_0^{-1} + (A^{t_1-t_0})' G_{t_1,t_0}^{-1} A^{t_1-t_0} & -(A^{t_1-t_0})' G_{t_1,t_0}^{-1} & 0 & \cdots & 0 \\ -G_{t_1,t_0}^{-1} A^{t_1-t_0} & P_1 & -(A^{t_2-t_1})' G_{t_2,t_1}^{-1} & \cdots & 0 \\ 0 & -G_{t_2,t_1}^{-1} A^{t_2-t_1} & P_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & -G_{t_K,t_{K-1}}^{-1} A^{t_K-t_{K-1}} & G_{t_K,t_{K-1}}^{-1} \end{bmatrix}, \tag{9}$$

where $P_s := (A^{t_{s+1}-t_s})^\top G_{t_{s+1},t_s}^{-1} A^{t_{s+1}-t_s} + G_{t_s,t_{s-1}}^{-1}$. Then, we parametrize $\hat{P}_{\mathcal{T}_S}$ as $\hat{P}_{\mathcal{T}_S} := \mathcal{N}(\mu, S)$, where $\mu \in \mathbb{R}^{nK}, S \in \mathbb{S}^+_{nK\times nK}$. Then, the marginal distributions $\hat{P}_{t_k}$ is written as $\mathcal{N}(\mu_k, S_{kk})$ and the optimization variable is μ and $\{S_{kk}\}_{k=0}^{K-1}$. Since the path is fixed, we can fill the non-diagonal blocks of S by computing the second moments obtained from Eq. (3). Under this parametrization, Eq. (7) becomes

$$\mathrm{KL}\left(\hat{P}_{\mathcal{T}_S} \,\middle\|\, \tilde{P}_{\mathcal{T}_S}\right) = \frac{1}{2}(\mu-\tilde{\mu})^\top \tilde{\Xi}(\mu-\tilde{\mu}) + \frac{1}{2}\operatorname{tr}(\tilde{\Xi}S) - \ln\det S + \text{const.} \tag{10}$$

For likelihood term of Problem 6, the marginal distribution at time t_k, denoted by $\hat{P}_{t_k}$, is given by $\mathcal{N}(\mu_k, S_{kk})$, where μ_k and S_{kk} are the corresponding block components of μ and S. Combining these results, the objective function in Problem 6 can be rewritten as

$$\begin{aligned} J\left(\mu, \{S_{kk}\}_{k=0}^{K-1}\right) := &\sum_{k=0}^{K}\sum_{i=1}^{n_k}(x_k^i-\mu_k)S_{kk}^{-1}(x_k^i-\mu_k) - n_k \ln\det S_{kk}^{-1} \\ &+ \lambda\left((\mu-\tilde{\mu})^\top\tilde{\Xi}(\mu-\tilde{\mu}) + \operatorname{tr}(\tilde{\Xi}S) - \ln\det S\right). \end{aligned} \tag{11}$$

The following theorem shows the convex property of the function (11).

Theorem 1 *The objective function (11) is quadratic and convex in terms of μ. Also, the function is convex in terms of the variable S_{kk} when $\lambda \le n_k$.*

Proof The first claim readily follows from the semi-positive definiteness of S_{kk} and $\tilde{\Xi}$. To prove the convexity with respect to S_{kk}, we analyze the structure of the determinant terms:

$$\begin{aligned} &-n_k \ln\det S_{kk}^{-1} - \lambda \ln\det S \\ &= -(n_k-\lambda)\ln\det S_{kk}^{-1} - \lambda\ln\det \begin{pmatrix} I & S_{00}^{-1/2}S_{01}S_{11}^{-1/2} & \cdots & S_{00}^{-1/2}S_{0K}S_{KK}^{-1/2} \\ S_{11}^{-1/2}S_{10}S_{00}^{-1/2} & I & \cdots & S_{11}^{-1/2}S_{1K}S_{KK}^{-1/2} \\ \vdots & \vdots & \ddots & \vdots \\ S_{KK}^{-1/2}S_{K0}S_{00}^{-1/2} & S_{KK}^{-1/2}S_{K1}S_{11}^{-1/2} & \cdots & I \end{pmatrix} \\ &= -(n_k-\lambda)\ln\det S_{kk}^{-1} - \lambda\ln\det\left(I - S_{kk}^{-1}D_k\right) + \text{const.} \end{aligned}$$

where D_k is the Schur complement with respect to k-th row and k-th column. Collecting the terms associated with S_{kk}, we obtain

$$\sum_{i=1}^{n_t}(x_k^i-\mu_t)S_{kk}^{-1}(x_k^i-\mu_t)-(n_k-\lambda)\ln\det S_{kk}^{-1}+\lambda\left(\mathrm{tr}(\tilde{\Xi}_{kk}S_{kk})-\ln\det\left(I-S_{kk}^{-1}D_k\right)\right). \quad (12)$$

Then, by performing the variable transformation $L_k = S_{kk}^{-1}$, you can see each term is convex. ■

Based on this theorem, we propose an iterative algorithm for ML-SB (Alg. 2). The algorithm alternates between two steps. First, with fixed marginal distributions, we solve the Problem 3 and compute the non-diagonal block of S by Eq. (3). Then, the marginal distributions are optimized with the non-diagonal blocks of S fixed, where the mean μ has an analytical update formula

$$\mu=\left(\begin{bmatrix}n_0S_{00}^{-1} & & & \\ & n_1S_{11}^{-1} & & \\ & & \ddots & \\ & & & n_KS_{KK}^{-1}\end{bmatrix}+\lambda T\tilde{\Xi}T^\top\right)^{-1}\begin{bmatrix}S_{00}^{-1}\sum_{i=1}^{n_0}x_0^i\\ S_{11}^{-1}\sum_{i=1}^{n_1}x_1^i\\ \vdots\\ S_{KK}^{-1}\sum_{i=1}^{n_K}x_K^i\end{bmatrix} \quad (13)$$

with

$$T=\begin{pmatrix}0 & 0 & \cdots & 0\\ -A^{t_1-t_0} & I & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ -A^{t_K-t_0} & 0 & \cdots & I\end{pmatrix}$$

and each diagonal block S_{kk} can be optimized through convex programming.

Remark 2 *As λ approaches zero, Problem 5 converges to Problem 4. This is because the likelihood term of Problem 5 becomes dominant, and the optimal marginal distributions converge to empirical distributions estimated directly from the samples. As a result, the optimization of the path becomes the same as the problem of minimizing the KL divergence with respect to the marginal of the sample.*

4. Simulations

For ML-SBID, we solve convex optimization problems by the MOSEK solver (MOSEK ApS (2023)) and the CVXPY modeler (Diamond and Boyd (2016)). All experiments for SBID and ML-SBID were conducted on a laptop PC with a Core i7-1185G7 CPU.

4.1. Sample Accuracy

In this section, we compare our SB-based methods (SBID and ML-SBID) with two baseline methods: mean-only identification (MeanOnlyID) estimates the transition matrix A by minimizing the prediction error of means $\sum_t\|\mu_{t+1}-A\mu_t\|^2$, while mean-covariance identification (MeanCovID) additionally considers the evolution of covariances

$$\min_{A,\Gamma}\sum_t\left(\|\mu_{t+1}-A\mu_t\|^2+\frac{1}{n}\|\Sigma_{t+1}-A\Sigma_tA^\top-\Gamma\|^2\right). \quad (14)$$

For MeanOnlyID, the optimal A can be obtained in closed form. For MeanCovID, we need to solve a nonlinear optimization problem due to the quadratic term $A\Sigma_t A^\top$ in the covariance error. We solved this problem using L-BFGS-B in *scipy.optimize* (Virtanen et al. (2020)). We randomly generate 10 dimensional systems as follows. The transition matrix A is generated as $A = 0.8I + 0.3R$, where R is a random matrix with elements uniformly drawn from $[-0.5, 0.5]$. The true noise covariance Γ is set to $0.01I$. The initial state covariance Σ_0 is randomly generated as a positive definite matrix with eigenvalues uniformly sampled from the interval $[0.3, 1.0]$. For each system, we collect sample trajectories with sample sizes varying from 2 to 256 samples and obtain snapshot data from the trajectories at time points $t \in [0, 1, \cdots, 20]$. All experiments are averaged over 10 random trials with different random seeds. For ML-SBID, we fix λ to 1.

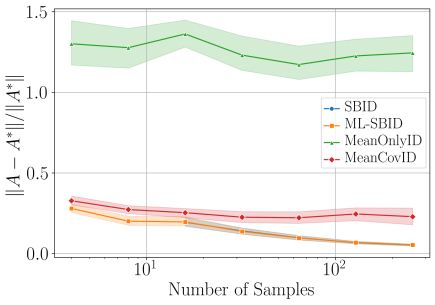

Figure 1: The vertical axis shows the relative error of estimated A, and the horizontal axis shows the number of trajectories. Error bars indicate the standard deviation over 10 random trials.

Figure 1 shows the parameter estimation error against varying sample sizes. Both SBID and ML-SBID achieve better accuracy compared to the baseline methods across all sample sizes. Although the accuracy of SBID and ML-SBID are almost the same, SBID can only be applied when the number of samples is larger than the dimension because the marginal distribution of the SB must be positive definite. This suggests that ML-SBID would be useful in scenarios with limited samples.

4.2. Time Efficiency

We evaluate our method's computational efficiency and prediction accuracy by comparing it with nonlinear SB methods, SBIRR (Shen et al. (2024)), DM-SB (Chen et al. (2023)), TrajectoryNet (Tong et al. (2020)), and Vanilla-SB (Vargas et al. (2021a)) using the Repressilator benchmark from Shen et al. (2024). The dynamics of the Repressilator model are

$$dX_1 = \frac{\beta}{1 + (X_3/k)^n} - \gamma X_1 + 0.1 dW_1 \tag{15}$$

$$dX_2 = \frac{\beta}{1 + (X_1/k)^n} - \gamma X_2 + 0.1 dW_2 \tag{16}$$

$$dX_3 = \frac{\beta}{1 + (X_2/k)^n} - \gamma X_3 + 0.1 dW_3 \tag{17}$$

where $\beta = 10$, $n = 3$, $k = 1$, and $\gamma = 1$. This model represents how three genes mutually suppress each other's expression.

The benchmark dataset is generated by the following procedure. The samples at time 0 are generated from the initial distribution $X_1, X_2 \sim \mathrm{Uniform}(1, 1.1), X_3 \sim \mathrm{Uniform}(2, 2.1)$ and samples from marginal distributions at each time point $t_k := 0.75(k-1)$ $(k = 1 \cdots 10)$ are obtained by the Euler-Maruyama method. For each time point, 50 independent samples is generated from the marginal distributions. The samples at times t_k $(k = 1, 3, \ldots, 9)$ are used as training data and those at times t_k $(k = 2, 4, 6, 8, 10)$ are used as test data. SBIRR has knowledge of the true dynamics except for its parameter and alternates between estimating the SB and optimizing β, n, γ. SBIRR has knowledge of the true functional form of the dynamics and alternates between estimating the

Method	t_2	t_4	t_6	t_8	t_{10}
SBIRR	$\mathbf{0.50 \pm 0.11}$	0.87 ± 0.10	$\mathbf{0.51 \pm 0.09}$	$\mathbf{0.61 \pm 0.10}$	1.49 ± 0.20
DM-SB	1.46 ± 0.08	1.08 ± 0.37	3.00 ± 0.54	2.18 ± 0.41	2.54 ± 1.21
TrajectoryNet	3.62 ± 0.05	2.86 ± 0.08	1.67 ± 0.07	3.45 ± 0.09	2.33 ± 0.08
Vanilla-SB	1.87 ± 0.05	1.23 ± 0.10	1.27 ± 0.15	1.19 ± 0.13	$\mathbf{1.16 \pm 0.14}$
SBID	0.63 ± 0.01	0.66 ± 0.04	0.58 ± 0.04	1.07 ± 0.07	1.56 ± 0.09
ML-SBID	0.65 ± 0.01	$\mathbf{0.65 \pm 0.04}$	$\mathbf{0.51 \pm 0.04}$	1.08 ± 0.07	1.66 ± 0.10

Table 1: Comparison of EMD values (mean ± standard deviation over 10 different random seeds). Results for DM-SB, TrajectoryNet, Vanilla-SB, and SBIRR are taken from Shen et al. (2024). The only source of randomness in SBID and ML-SBID is path sampling.

SB and optimizing β, n, and γ. For DM-SB, TrajectoryNet, and Vanilla-SB, the references are fixed to Brownian motion. On the other hand, our method assumes linear systems for dynamics and optimizes its system parameters. Since the noise term is fixed in SBIRR, we fix G to $0.1I$ and update only A in the optimization of SBID and ML-SBID. For evaluation of prediction error, we use the Earth Mover's Distance (EMD, Rubner et al. (1998)) between the marginal distributions of paths sampled from each model and the test data marginals. Table 1 and 2 show the EMD values at each test point and computation times. The results for SBIRR, DM-SB, TrajectoryNet, Vanilla-SB, taken from Shen et al. (2024), were obtained using a computing system equipped with an Intel Xeon Gold 6248 CPU and an NVIDIA Volta V100 GPU with 32 GB RAM. Our proposed methods (SBID and ML-SBID) achieve comparable or better accuracy than those in previous studies, even though we assume linear dynamics in our method. It is worth noting that our method significantly reduces computation time. SBID completes in about 0.2 seconds and ML-SBID in about 82 seconds, which is a substantial improvement compared to other methods' computation time of several hours.

5. Discussion and future work

In this paper, we proposed a system identification method from snapshot data for linear Gaussian systems using SB. The experimental results demonstrated that our method achieves better accuracy compared to direct moment-based approaches. We also showed that our method is efficient and comparable in performance to existing SB methods.

Method	Computation Time
SBIRR	2.43 ± 0.60 hour
DM-SB	15.63 ± 0.12 hour
TrajectoryNet	9.86 ± 0.43 hour
Vanilla-SB	0.23 ± 0.05 hour
SBID	0.20 ± 0.02 sec
ML-SBID	81.24 ± 2.17 sec

Table 2: Comparison of computation times for different methods. Results for DM-SB, TrajectoryNet, Vanilla-SB, and SBIRR are taken from Shen et al. (2024).

Several directions remain for future work. First, the current implementation for ML-SBID requires the inversion of $O(n_x N)$ dimensional matrices in computing the Schur complement when optimizing marginal covariances. There is potential for computational improvement, since the cost function for covariance at each time point should only depend on its neighboring marginals. Second, extending our method to SB for partially observed system (Chen et al. (2019, 2023)) would be a promising direction.

Acknowledgments

This work was supported in part by Moonshot R&D Grant Number JPMJMS2021 and by JSPS KAKENHI under Grant Number JP21H04875.

References

A. Beghi. On the relative entropy of discrete-time markov processes with given end-point densities. *IEEE Transactions on Information Theory*, 42(5):1529–1535, 1996. doi: 10.1109/18.532893.

Tianrong Chen, Guan-Horng Liu, Molei Tao, and Evangelos Theodorou. Deep momentum multi-marginal schrödinger bridge. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023. URL https://openreview.net/forum?id=ykvvv0gc4R.

Yongxin Chen, Giovanni Conforti, Tryphon T. Georgiou, and Luigia Ripani. Multi-marginal schrodinger bridges, 2019. URL https://arxiv.org/abs/1902.08319.

Steven Diamond and Stephen Boyd. CVXPY: A Python-embedded modeling language for convex optimization. *Journal of Machine Learning Research*, 17(83):1–5, 2016.

Kaito Ito and Kenji Kashima. Maximum entropy optimal density control of discrete-time linear systems and schrödinger bridges. *IEEE Transactions on Automatic Control*, 69(3):1536–1551, 2024. doi: 10.1109/TAC.2023.3305319.

Takeshi Koshizuka and Issei Sato. Neural lagrangian schrödinger bridge: Diffusion modeling for population dynamics. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=d3QNWD_pcFv.

MOSEK ApS. *MOSEK Optimizer API for Python. Version 10.1*, 2023.

Y. Rubner, C. Tomasi, and L.J. Guibas. A metric for distributions with applications to image databases. In *Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271)*, pages 59–66, 1998. doi: 10.1109/ICCV.1998.710701.

Yunyi Shen, Renato Berlinghieri, and Tamara Broderick. Multi-marginal schrödinger bridges with iterative reference refinement, 2024. URL https://arxiv.org/abs/2408.06277.

Alexander Tong, Jessie Huang, Guy Wolf, David Van Dijk, and Smita Krishnaswamy. TrajectoryNet: A dynamic optimal transport network for modeling cellular dynamics. In Hal Daumé III and Aarti Singh, editors, *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 9526–9536. PMLR, 13–18 Jul 2020. URL https://proceedings.mlr.press/v119/tong20a.html.

Francisco Vargas, Pierre Thodoroff, Austen Lamacraft, and Neil Lawrence. Solving schrödinger bridges via maximum likelihood. *Entropy*, 23(9), 2021a. ISSN 1099-4300. doi: 10.3390/e23091134. URL https://www.mdpi.com/1099-4300/23/9/1134.

Francisco Vargas, Pierre Thodoroff, Austen Lamacraft, and Neil Lawrence. Solving schrödinger bridges via maximum likelihood. *Entropy*, 23(9), 2021b. ISSN 1099-4300. doi: 10.3390/e23091134. URL https://www.mdpi.com/1099-4300/23/9/1134.

Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020. doi: 10.1038/s41592-019-0686-2.

Benjie Wang, Joel Jennings, and Wenbo Gong. Neural structure learning with stochastic differential equations. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=V1GM9xDvIY.