ILLINOIS TECH
College of Computing

Learning Stochastic Dynamics from Data

Ziheng Guo [a] & Igor Cialenco & Ming Zhong

[a]All authors are from Dept. of Applied Math, Illinois Tech. Email: zguo22@hawk.iit.edu

Introduction

- Stochastic Differential Equation (SDE) is a fundamental modeling tool in various science and engineering fields.
- Compared with traditional deterministic models which often fall short in capturing the stochasticity in the nature, the significance of SDE lies in the ability to model complex systems influenced by random perturbations.
- SDE models are widely used in (but not limited to):
  1. Quantum Mechanics in Physics
  2. Molecular Dynamics in Chemistry
  3. Material Sciences
- The calibration of these models is crucial for their effective application.
- We present a noise guided trajectory based system identification method for inferring the dynamical structure from observation generated by SDEs.

Learning Framework

- We consider the following SDE
$$dX_t = f(X_t)dt + dW_t, \quad X_t, W_t \in R^d, \qquad (1)$$
where $f: R^d \to R^d$ is a drift term, and W represents the Brownian noise with a symmetric positive definite covariance matrix $D = D(x): R^d \to R^{d\times d}$.
- We consider the scenario when we are given continuous observation data in the form of $\{X_t, dX_t\}_{t\in[0,T]}$ for $X_0 \sim \mu_0$.
- The estimation of drift funtion f will be the minimizer to the following loss function
$$\mathcal{E}_H(\tilde{f}) = E_{X_0 \in \mu_0}\Big[\frac{1}{2T}\int_{t=0}^{T} <\tilde{f}(X_t), D^{-1}(X_t)\tilde{f}(X_t)> \ dt$$
$$- 2\int_{t=0}^{T} <\tilde{f}(X_t), D^{-1}(X_t)dX_t> \Big],$$
where $\tilde{f} \in H$; the function space H is designed to be convex and compact and it is also data-driven.
- In the case of uncorrelated noise, i.e. $D(X) = \sigma^2(X)I$, where I is the $d \times d$ identity matrix and $\sigma: R^d \to R^+$ is a scalar function depends on the state, representing the noise level. The above equation can be simplified to
$$\mathcal{E}_H^{\text{Sim}}(\tilde{f}) = E_{X_0 \in \mu_0}\Big[\frac{1}{2T}\int_{t=0}^{T} \frac{||\tilde{f}(X_t)||^2}{\sigma^2(X_t)} dt - 2\int_{t=0}^{T} \frac{<\tilde{f}(X_t), dX_t>}{\sigma^2(X_t)}\Big].$$
- The uniqueness of our method is that we incorporate the covariance matrix into the learning and hence improving the estimation especially when the noise is correlated.

Performance Measures

- If we have access to original drift function f, then we will use the following error to compute the difference between $\hat{f}$ (our estimator) to f with the following norm
$$||f - \hat{f}||^2_{L^2(\rho)} = \frac{1}{T}\int_{X\in\Omega} ||f(X) - \hat{f}(X)||^2_{\ell^2(R^d)}\, d\rho(X).$$
Here the weighted measure ρ is defined on Ω, where it defines the region of X explored due to the dynamics defined by (1).
- In real life situation, f is most likely non-accessible. Hence we will look at a performance measure that compare the difference between $X(f, X_0, T) = \{X_t\}_{t\in[0,T]}$ and $\hat{X}(\hat{f}, X_0, T) = \{\hat{X}_t\}_{t\in[0,T]}$. Then, the difference between the two trajectories is measured as follows
$$||X - \hat{X}|| = E_{X_0\sim\mu_0}\Big[\frac{1}{T}\int_{t=0}^{T} ||X_t - \hat{X}_t||^2_{\ell^2(R^d)}\, dt\Big].$$
- We also compare the distribution of the trajectories over different initial conditions and all possible noise at some chosen time snapshots using the Wasserstein distance.

Examples I

- We initiate our numerical study with a one-dimensional ($d = 1$) drift function that incorporates both polynomial and trigonometric components, given by
$f = 2 + 0.08X - 0.05\sin(X) + 0.02\cos^2(X)$.

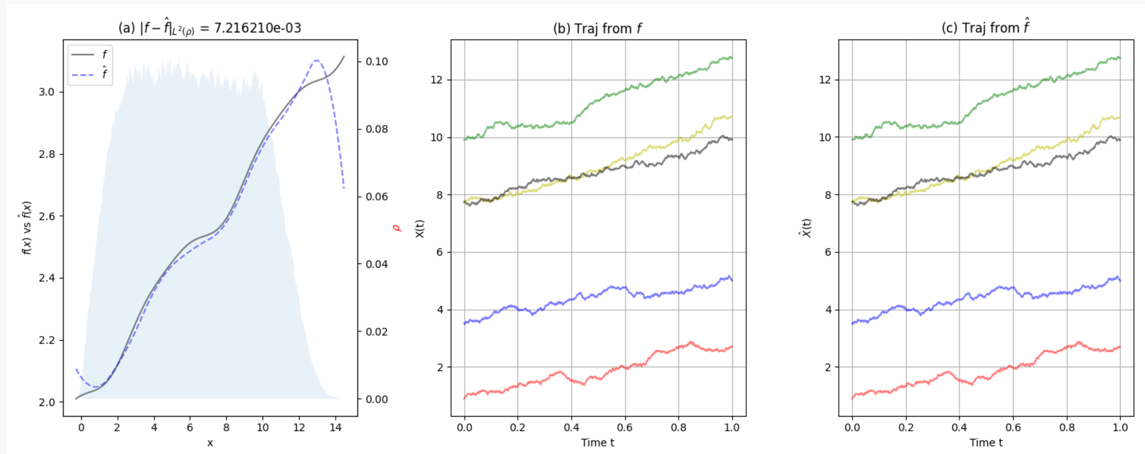

Figure: Estimation Summary of 1D Example

- Left: Comparison of f and $\hat{f}$. Middle: 5 trajectories generated by f. Right: 5 trajectories generated by $\hat{f}$ with same noise.

Table: One-dimensional Drift Function Estimation Summary

Number of Basis	8
Maximum Degree	2
Relative $L^2(\rho)$ Error	0.007935
Relative Trajectory Error	0.0020239 ± 0.002046
Wasserstein Distance (t=1.00)	0.0403

Examples II

- We extend our numerical test to two-dimensional case ($d = 2$) and set
$$f_1 = 2\sin(0.2X_1) + 1.5\cos(0.1X_2)$$
$$f_2 = 3\sin(0.3X_1)\cos(0.1X_2).$$

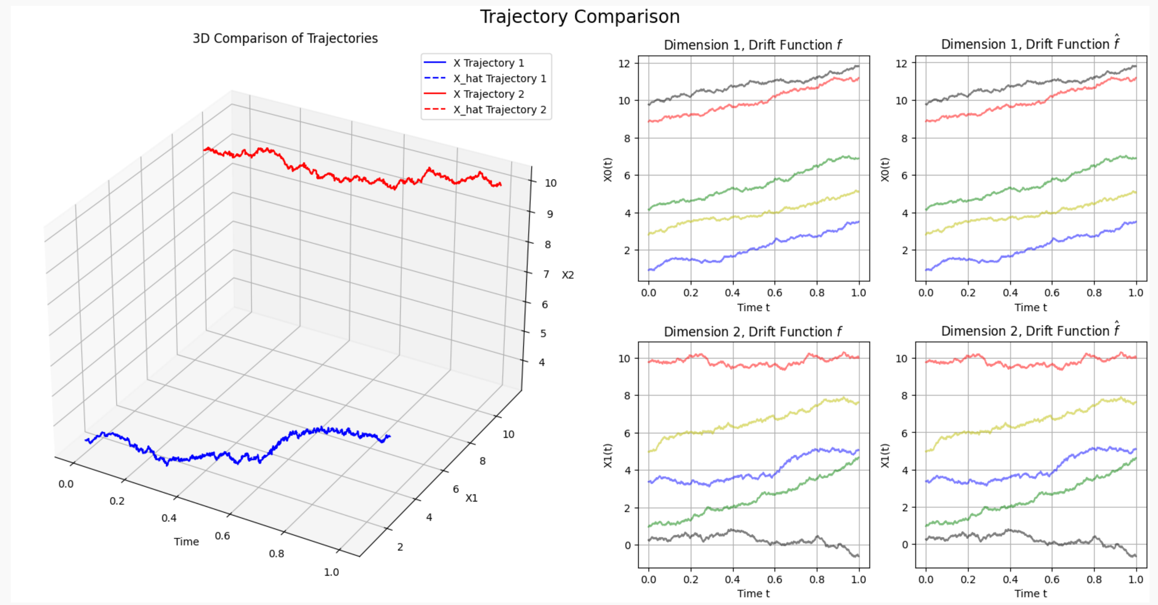

Figure: Two-dimensional Trigonometric Trajectory Comparison

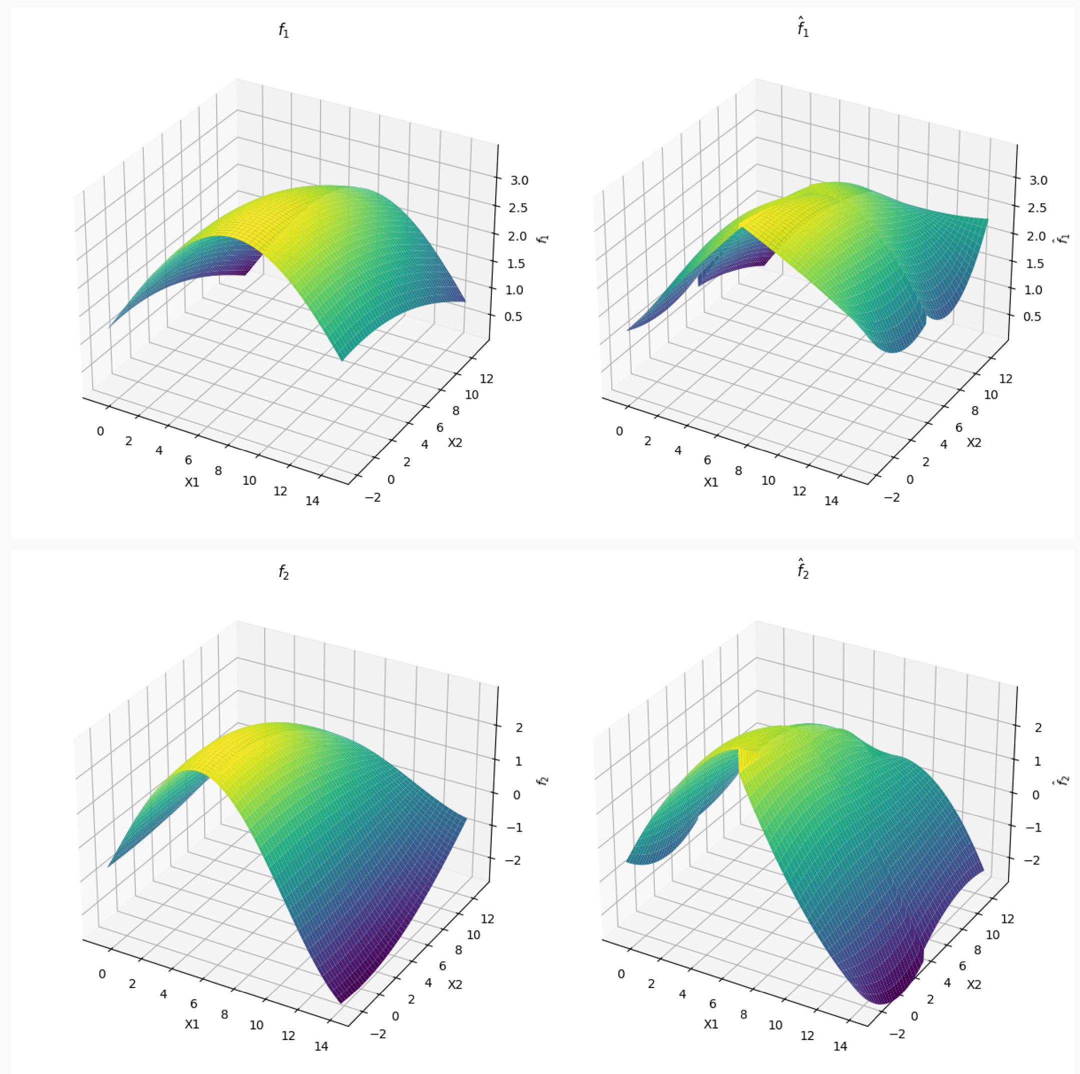

Figure: Comparison of f and $\hat{f}$ in 2D. (a) Surface of Dimension 1 (b) Surface of Dimension 2

Summary

- We have presented a novel approach of learning the drift and diffusion from a noise-guide likelihood function. Such learning can handle general noise structure as long as the covariance information of the noise is known.
- Our experiment results confirms that our method is capable in functional estimation of the drift term for SDE.