Learning Stochastic Dynamics from Data

Ziheng Guo

Department of Mathematics
University of Houston

April 28, 2025

Joint work with Igor Cialenco, Ming Zhong and James Greene

Content

- Problem Setup
- Learning Framework and Performance Measure
- Examples

Problem Setup

What are we doing?

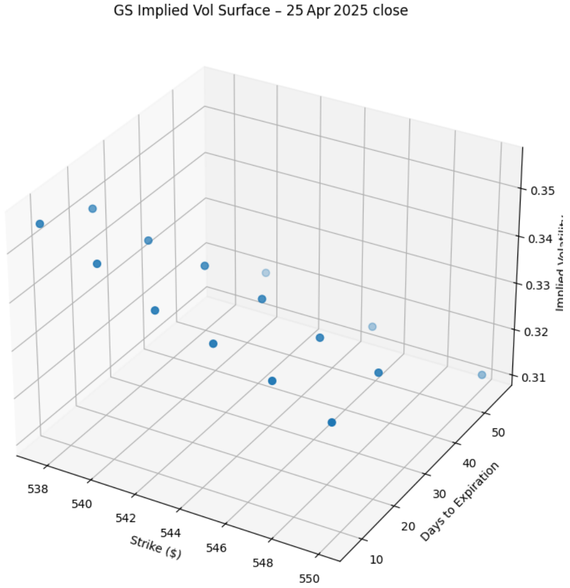
Given observation of trajectories, estimate SDE parameters.

- We consider SDE

$$\mathrm{d}\mathbf{x}_t = \boldsymbol{f}(\mathbf{x}_t)\,\mathrm{d}t + \sigma(\mathbf{x}_t)\,\mathrm{d}\mathbf{w}_t, \quad \mathbf{x}_t, \mathbf{w}_t \in \mathbb{R}^d, \tag{1}$$

  with some given initial condition $\mathbf{x}_0 \sim \mu_0$, and where $\boldsymbol{f} : \mathbb{R}^d \to \mathbb{R}^d$ is the drift term, $\sigma : \mathbb{R}^d \to \mathbb{R}^{d\times d}$ is the diffusion coefficient, and $\mathbf{w}$ represents a vector of independent standard Brownian Motions.
- $\Sigma = \Sigma(\boldsymbol{x}) : \mathbb{R}^d \to \mathbb{R}^{d\times d}$ where $\Sigma = \sigma\sigma^\intercal$.
- Given observation data in the form of $\{\mathbf{x}_t, \mathrm{d}\mathbf{x}_t\}_{t\in[0,T]}$ for $\mathbf{x}_0 \sim \mu_0$.
- **Question**: How to estimate f and Σ given the observation of X_t?

Problem Setup



Learning Framework and Performance Measure

The estimation of drift function f will be the minimizer to the following loss function

$$\mathcal{E}_{\mathcal{H}}(\tilde{\boldsymbol{f}}) = \mathbb{E}\left[\frac{1}{2}\int_{t=0}^{T}\Big(\langle\tilde{\boldsymbol{f}}(\mathbf{x}_t), \Sigma^{\dagger}(\mathbf{x}_t)\tilde{\boldsymbol{f}}(\mathbf{x}_t)\rangle \, \mathrm{d}t - 2\langle\tilde{\boldsymbol{f}}(\mathbf{x}_t), \Sigma^{\dagger}(\mathbf{x}_t)\, \mathrm{d}\mathbf{x}_t\rangle\Big)\right], \tag{2}$$

- $\mathcal{H}$ as function space of $\tilde{\boldsymbol{f}}$. Designed to be convex and compact and it is also data-driven.
- $\Sigma^{\dagger}$ is the pseudo-inverse of Σ (when σ is assumed to be SPD, $\Sigma^{\dagger} = \Sigma^{-1}$).
- This loss function is derived from Girsanov theorem and the corresponding Randon-Nykodim derivative or likelihood ratio for stochastic processes.

Learning Framework and Performance Measure

The estimation of Σ is the minimizer of the following loss function

$$\mathcal{E}(\tilde{\Sigma}) = \mathbb{E}\Big[[\mathbf{x}, \mathbf{x}]_T - \int_{t=0}^{T} \tilde{\Sigma}(\mathbf{x}_t)\, \mathrm{d}t)\Big]^2. \tag{3}$$

where $[\mathbf{x}, \mathbf{x}]_T$ is the quadratic variation of the stochastic process $\mathbf{x}_t$ over time interval $[0, T]$.

$$[\mathbf{x}_i, \mathbf{x}_j]_t = \lim_{|\Delta t_k| \to 0} \sum_k (\mathbf{x}_i(t_{k+1}) - \mathbf{x}_i(t_k))(\mathbf{x}_j(t_{k+1}) - \mathbf{x}_j(t_k)),$$

- In particular, if Σ is constant, then the estimation can be simplified to $\tilde{\Sigma} = \mathbb{E}\frac{[\mathbf{x},\mathbf{x}]_T}{T}$.
- In case where both $\boldsymbol{f}$ and Σ are unknown, Σ can be estimated first, allowing the estimated covariance matrix to be used to implement

- If we have access to original drift function $\boldsymbol{f}$, then we will use the following error to compute the difference between $\hat{\boldsymbol{f}}$ (our estimator) to $\boldsymbol{f}$ with the following norm

$$||\boldsymbol{f} - \hat{\boldsymbol{f}}||_{L^2(\rho)}^2 = \int_{\mathbb{R}^d} ||\boldsymbol{f}(\mathbf{x}) - \hat{\boldsymbol{f}}(\mathbf{x})||_{\ell^2(\mathbb{R}^d)}^2 \, \mathrm{d}\rho(\mathbf{x}). \tag{4}$$

- The weighted measure ρ, defined on $\mathbb{R}^d$, is given as follows

$$\rho(\mathbf{x}) = \mathbb{E}\Big[\frac{1}{T}\int_{t=0}^{T} \delta_{\mathbf{x}_t}(\mathbf{x})\Big], \quad \text{where } \mathbf{x}_t \text{ evolves from } \mathbf{x}_0 \text{ by equation 1.} \tag{5}$$

- Under normal circumstances, $\boldsymbol{f}$ is most likely non-accessible. Thus we look at a performance measure that compares the difference between $\mathbf{X}(\boldsymbol{f}, \mathbf{x}_0, T) = \{\mathbf{x}_t\}_{t \in [0, T]}$ (the observed trajectory that evolves from $\mathbf{x}_0 \sim \mu_0$ with the unknown $\boldsymbol{f}$) and $\hat{\mathbf{X}}(\hat{\boldsymbol{f}}, \mathbf{x}_0, T) = \{\hat{\mathbf{x}}_t\}_{t \in [0, T]}$.
- Difference between the two trajectories is measured as

$$||\mathbf{X} - \hat{\mathbf{X}}|| = \mathbb{E}\Big[\frac{1}{T}\int_{t=0}^{T} ||\mathbf{x}_t - \hat{\mathbf{x}}_t||^2_{\ell^2(\mathbb{R}^d)}\, \mathrm{d}t\Big]. \tag{6}$$

- In most general case, we compare the distribution of the trajectories over different initial conditions and all possible noise at some chosen time snapshots using the Wasserstein distance at any given time $t \in [0, T]$.
- The Wasserstein distance of order two between μ_t^M and $\hat{\mu}_t^M$ is calculated as

$$\mathcal{W}_2(\mu_t^M, \hat{\mu}_t^M \mid \mu_0) = \left(\inf_{\pi \in \Pi(\mu_t^M, \hat{\mu}_t^M \mid \mu_0)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \|x - y\|^2 \, d\pi(x, y) \right)^{1/2}. \tag{7}$$

where

$$\mu_t^M = \frac{1}{M} \sum_{i=1}^{M} \delta_{\mathbf{x}^{(i)}(t)}, \quad \hat{\mu}_t^M = \frac{1}{M} \sum_{i=1}^{M} \delta_{\hat{\mathbf{x}}^{(i)}(t)} \tag{8}$$

Examples
1D Drift

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

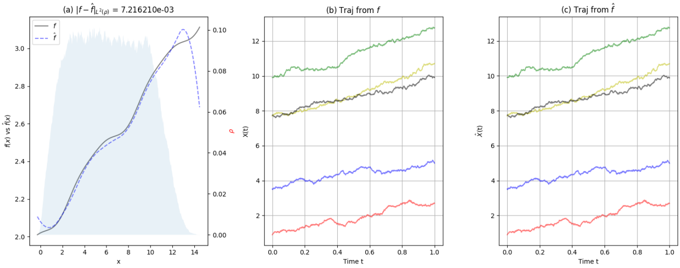


Table: One-dimensional Drift Function Estimation Summary

		Wasserstein Distance	
Number of Basis	8	Wasserstein Distance	
Maximum Degree	2	$t = 0.25$	0.0291
Relative $L^2(\rho)$ Error	0.007935	$t = 0.50$	0.0319
Relative Trajectory Error	0.0020239 ± 0.002046	$t = 1.00$	0.0403

- We also tested our estimation method to van der Pol oscillator, which is a classical example of a self-sustained oscillator with nonlinear damping, which has many applications in biology and physics and set

$$\begin{cases} \mathrm{d}x_t &= y_t \, \mathrm{d}t, \\ \mathrm{d}y_t &= \Big(\mu\big(1 - x_t^2\big)\, y_t - x_t\Big)\, \mathrm{d}t + \sigma \, \mathrm{d}w_t^y \end{cases} \tag{9}$$

- We set parameters as $\mu = 1$ and $\sigma = 0.1$

Examples
2D Drift

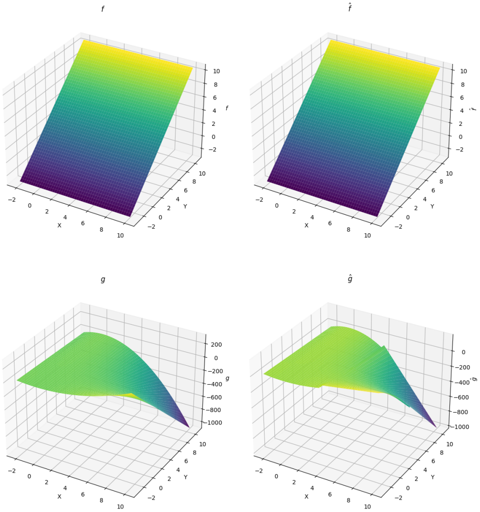


Figure: Comparison of true and estimated surfaces in each dimension

Examples
2D Drift

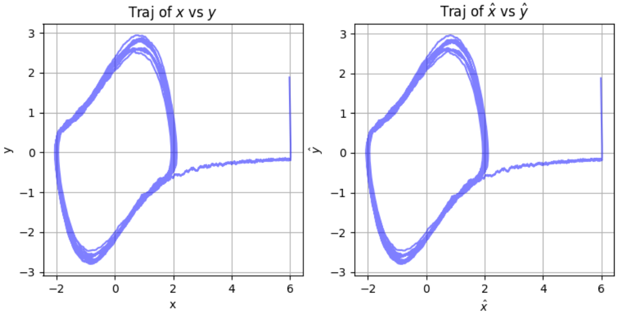


Figure: Trajectory controlled by estimated drift keeps the oscillating property

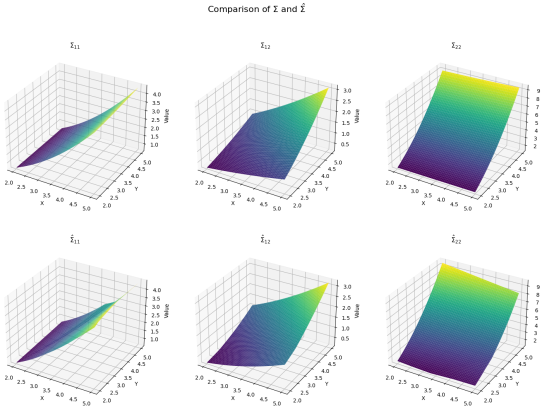
Table: Van der Pol oscillator drift estimation summary

Relative $L^2(\rho)$ Error	0.0297
Relative Trajectory Error	0.019 ± 0.071
Wasserstein Distance at $t = 25$	0.0521
Wasserstein Distance at $t = 50$	0.0548
Wasserstein Distance at $t = 100$	0.0539

Examples

2D Variance

- We estimate the covariance matrix Σ for the two-dimensional ($d = 2$) case. We assume that both $\boldsymbol{f}$ and Σ are unknown. We set σ as: $\begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}$ where the components $\sigma_{11} = 0.4\mathbf{x}_1$, $\sigma_{12} = \sigma_{21} = 0.025\mathbf{x}_1\mathbf{x}_2$, $\sigma_{22} = 0.6\mathbf{x}_2$ are all state dependent.



Consider the stochastic heat equation driven by an additive noise

$$\mathrm{d}\mathbf{u}(t,\mathbf{x}) - \theta\Delta\mathbf{u}(t,\mathbf{x})\,\mathrm{d}t = \sigma\,\mathrm{d}\mathbf{w}(t,\mathbf{x}) \tag{10}$$

on a smooth bounded domain $\mathbf{x} \in G \subset \mathbb{R}^d$, with initial condition $\mathbf{u}(0,\mathbf{x}) = 0$, zero boundary condition, and Δ being the Laplace operator with zero boundary conditions in a suitable underlying Hilbert space H.

- We are interested in the estimation of θ.
- Consider orthonormal basis $\{h_k\}_{k\in\mathbb{N}} \subset H$.
- By spectral approach, we have the projection operator $P : H \to H^N$, where $H^N = \mathrm{span}\{h_1, \ldots, h_N\}$. Then $\mathbf{u}^N = P^N\mathbf{u} = \sum_{k=1}^N \mathbf{u}_k(t)h_k(\mathbf{x})$ is the Fourier approximation of the solution $\mathbf{u}$ by the first N eigenmodes

$$\mathrm{d}\sum_{k=1}^{N}\mathbf{u}_k(t)h_k(\mathbf{x}) + \theta\sum_{k=1}^{N}\mathbf{u}_k(t)\lambda_k h_k(\mathbf{x})\,\mathrm{d}t = \sigma\,\mathrm{d}\sum_{k=1}^{N} q_k h_k(\mathbf{x})\mathbf{w}_k(t).$$

Since $\{h_k(\mathbf{x})\}_{k=1}^{N}$ are orthogonal to each other, we get that

$$\mathrm{d}\mathbf{u}_k(t) + \theta\lambda_k\mathbf{u}_k(t)\,\mathrm{d}t = \sigma q_k\,\mathrm{d}\mathbf{w}_k(t), \quad k = 1,\ldots,N.$$

Then θ can be estimated by 2 and we obtain the loss function

$$\mathcal{E}(\tilde{\theta}) = \mathbb{E}\left[\frac{1}{2}\sum_{k=1}^{N}\frac{\tilde{\theta}^2\lambda_k^2}{\sigma^2 q_k^2}\int_0^T \mathbf{u}_k^2\,\mathrm{d}t + \sum_{k=1}^{N}\frac{\tilde{\theta}\lambda_k}{\sigma^2 q_k^2}\int_0^T \mathbf{u}_k\,\mathrm{d}\mathbf{u}_k\right]. \tag{11}$$

We focus on one dimensional stochastic heat equation, $d = 1$, and take the domain $G = [0, \pi]$. In this case $h_k(x) = \sin(kx)$ and $\lambda_k = k^2$. The parameters are set as follows: $T = 1$, $\Delta t = 0.01$, $\sigma = 0.1$, and $q_k = 1$, that corresponds to space-time white noise. We set the parameter of interest $\theta = 2$.

	N = 1	N = 2	N = 5	N = 10	N = 20
M = 1	0.5230	3.2796	2.3315	1.9893	2.0000
M = 10	1.7456	2.2964	2.0765	2.0036	2.0000
M = 50	2.3217	1.8248	2.0433	2.0009	2.0000
M = 100	1.7596	2.0183	2.0082	2.0004	2.0000

Table: SPDE θ estimation under different number of modes N and trajectory number M

Consider the stochastic heat equation driven by an additive noise

$$\mathrm{d}\mathbf{u}(t,\mathbf{x}) - \theta(\mathbf{x})\Delta\mathbf{u}(t,\mathbf{x})\,\mathrm{d}t = \sigma\,\mathrm{d}\mathbf{w}(t,\mathbf{x}), \quad \mathbf{x} \in [0, 2\pi], \tag{12}$$

with initial condition $\mathbf{u}(0,\mathbf{x}) = 0$, zero boundary condition. Δ being the Laplace operator with zero boundary conditions in a suitable underlying Hilbert space H.

$$\theta(x) = \begin{cases} \theta_1 & \text{if } 0 \leq \mathbf{x} < \pi, \\ \theta_2 & \text{if } \pi \leq \mathbf{x} \leq 2\pi \end{cases},$$

where we assume θ_1 and θ_2 are unknown.
Then we obtain that

$$\mathrm{d}\mathbf{u}_j(t) + \sum_{k=1}^{\infty} \langle \theta(\mathbf{x}) h_k(\mathbf{x}), h_j(\mathbf{x})\rangle \lambda_k \mathbf{u}_k(t)\,\mathrm{d}t = \sigma q_j\,\mathrm{d}\mathbf{w}_j(t), \quad j = 1, \ldots, N. \tag{13}$$

We set the parameters as follows: $T = 1$, $\delta t = 0.01$, $\sigma = 0.5$, $q_k = 1$, $\lambda_k = \frac{k^2}{4}$ and $M = 1$.

	$(\hat{\theta}_1, \hat{\theta}_2)$	N	L^2 Error
$(\theta_1 = 2, \theta_2 = 4)$	$(2.053, 3.993)$	10	0.0535
$(\theta_1 = 2, \theta_2 = 4)$	$(1.999, 4.000)$	20	0.0010
$(\theta_1 = 1, \theta_2 = 5)$	$(1.044, 5.086)$	10	0.0966
$(\theta_1 = 1, \theta_2 = 5)$	$(0.999, 5.000)$	20	0.0010

Table: SPDE θ estimation

We also consider an interacting agent system with correlated stochastic noise. For a system of N agents, where each agent is associated with a state vector $\mathbf{x}_i \in \mathbb{R}^{d'}$. The agents' states are governed by the following SDEs

$$\mathrm{d}\mathbf{x}_i(t) = \frac{1}{N} \sum_{j=1, j\neq i}^{N} \phi(||\mathbf{x}_j(t) - \mathbf{x}_i(t)||)(\mathbf{x}_j(t) - \mathbf{x}_i(t))\,\mathrm{d}t + \sigma(\mathbf{x}_i(t))\,\mathrm{d}\mathbf{w}(t), \quad i = 1, \cdots, N.$$

Here $\phi : \mathbb{R}^+ \to \mathbb{R}$ is an interaction kernel that governs how agent j influences the behavior of agent i, and $\sigma : \mathbb{R}^{d'} \to \mathbb{R}^{d'}$ is a symmetric positive definite matrix that represents the noise.

If we define the vectorized notations,

$$\boldsymbol{f}_\phi(\mathbf{x}) = \frac{1}{N}\begin{bmatrix} \sum_{j=2}^{N} \phi(||\mathbf{x}_j - \mathbf{x}_1||)(\mathbf{x}_j - \mathbf{x}_1) \\ \vdots \\ \sum_{j=1}^{N-1} \phi(||\mathbf{x}_j - \mathbf{x}_N||)(\mathbf{x}_j - \mathbf{x}_N) \end{bmatrix} \quad \text{and} \quad \tilde{\sigma} = \begin{bmatrix} \sigma(\mathbf{x}_1) & 0 & \cdots & \mathbf{0} \\ 0 & \sigma(\mathbf{x}_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma(\mathbf{x}_N) \end{bmatrix}.$$

Here $\boldsymbol{f} : \mathbb{R}^d \to \mathbb{R}^d$ and $\tilde{\sigma} : \mathbb{R}^d \to \mathbb{R}^{d\times d}$.

We test our learning with the following parameters $N = 20$, $d' = 2$ (hence $d = Nd' = 40$), $\phi(r) = r - 1$, $T = 1$

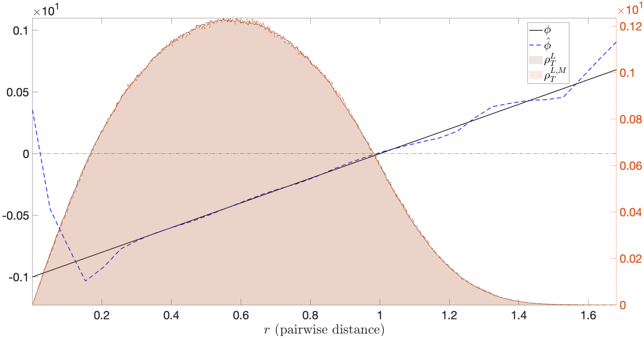


Figure: Comparison of true ϕ vs learned $\hat{\phi}$.