Score Based Diffusion in Infinite Dimensions

Nikola B. Kovachki

NVIDIA

ICIAM 2023
Theoretical Foundations and Algorithmic Innovation in Operator Learning
August 22nd

Outline

Generative Modeling

Generative Models in the Wild

Figure: DALL-E 2

Figure: Imagen

Generative Models in Scientific Computing

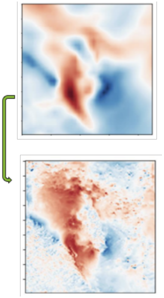


Figure: Inverse Problems

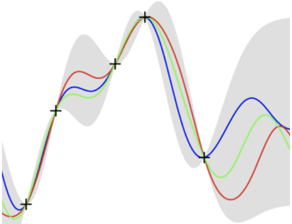


Figure: UQ

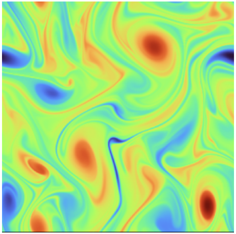


Figure: Chaos

Problem Formulation

Unconditional

Goal: sample measure μ supported on $\mathcal{U}$

Given: data samples $\{u_j\}_{j=1}^N \overset{i.i.d.}{\sim} \mu$

Map: $\Psi : \mathcal{X} \to \mathcal{U}, \ \Psi_\sharp \eta = \mu, \ \eta$ measure on $\mathcal{X}$

Conditional

Goal: sample measure $\mu(\cdot|y)$ for every $y \in \mathcal{Y}$

Given: data samples $\{(u_j, y_j)\}_{j=1}^N \overset{i.i.d.}{\sim} \mu$

Map: $\Psi : \mathcal{X} \times \mathcal{Y} \to \mathcal{U}, \ \Psi(\cdot, y)_\sharp \eta = \mu(\cdot|y)$

Current Approaches

Deterministic

- Generative Adversarial Networks
- Normalizing Flows
- Triangular Maps
- Variational Autoencoders
- Diffusion Models

Stochastic

- MCMC methods
- Stochastic Interpolants
- Diffusion Models

Score Matching in Finite Dimensions

(Song, Y., Ermon, S., 2019)

Langevin Dynamics

Assumptions

$$\mathcal{X} = \mathcal{U} = \mathbb{R}^d, \ \mu \text{ has density } p \in C^1(\mathbb{R}^d)$$

Langevin Equation

$$dx_t = \nabla_x \log p(x_t)\, dt + \sqrt{2} dz_t$$
$$x_0 \sim N(0, I) := \eta$$
$$z_t \text{ is a standard Wiener process}$$

Map

$$\Psi_T(x_0) := x_T, \quad T \gg 0$$
$$(\Psi_T)_\sharp \eta \to \mu \text{ as } T \to \infty$$

The Score and Denoising

Perturbation of the Score

$$\nu_\sigma = \mu * N(0, \sigma^2 I), \;\; \sigma > 0$$
$$\nu_\sigma \text{ has density } p_\sigma \in C^\infty(\mathbb{R}^d)$$

Denoising Score Matching

$$\sigma^2 \nabla \log p_\sigma = \arg\min_{s_\sigma} \mathbb{E}_{\xi \sim N(0,\sigma^2 I)} \mathbb{E}_{u \sim \mu} \left| \xi + s_\sigma(u + \xi) \right|^2$$

Map

$$\Psi_{T,\sigma}(x_0) := x_T^{(\sigma)}, \;\; T \gg 0, \;\; \sigma \ll 1$$
$$(\Psi_{T,\sigma})_\sharp \eta \to \mu \text{ as } T \to \infty, \text{ and } \sigma \to 0$$

Multiple Noise Scales

Langevin Equations

$$\text{Pick}: 0 < \sigma_1 < \sigma_2 < \cdots < \sigma_J$$

$$dx_t^{(\sigma_j)} = \sigma_j^2 \nabla_x \log p_{\sigma_j}\left(x_t^{(\sigma_j)}\right) dt + \sqrt{2} dz_t^{(\sigma_j)}$$

$$x_0^{(\sigma_j)} \sim \mathcal{L}\left(x_{T_{j+1}}^{(\sigma_{j+1})}\right),\ j = J-1, \ldots, 1,\ x_0^{\sigma_J} \sim N(0, \sigma_J^2 I)$$

$$z_t^{(\sigma_j)} \text{ is a } \sigma_j^2 - \text{Wiener process}$$

Map

$$\Psi := \Psi_{T_1,\sigma_1} \circ \cdots \circ \Psi_{T_J,\sigma_J}$$

Forward and Reverse SDEs

Key Idea

Allow σ to vary continuously

Forward Process

$$du_t = \sqrt{t} dz_t, \ \ u_0 \sim \mu$$

Reverse Process

$$\text{SDE: } dx_t = -t\nabla_x p_t(x_t)\, dt + \sqrt{t} d\bar{z}_t, \ \ x(T) \sim \mathcal{L}(u_T)$$

$$\text{ODE: } dx_t = -t\nabla_x p_t(x_t)\, dt, \ \ x(T) \sim \mathcal{L}(u_T)$$

Score Matching in Infinite Dimensions

(Lim, J., Kovachki, N.B., Baptista, R., et. al., 2023)

Operator Learning

Key Idea

Treat Ψ as a map between function spaces

Find generalization yielding tractable approximation

Benefits for Generative Modeling

Mathematical understanding

Scale to larger resolutions

Consistent error at any resolution

Consistent sampling cost at any resolution

Gaussian Densities

Assumptions

$\mathcal{X} = \mathcal{U} = H$ infinite-dimensional separable Hilbert space

$\mu_\sigma = N(0, \sigma^2 C)$ centered Gaussian measure on H

$\mu(H_0) = 1$, with H_0 the Cameron-Martin space of μ_σ

Perturbation

$\nu_\sigma = \mu * \mu_\sigma$

$\nu_\sigma \sim \mu_\sigma$ equivalent in the sense of measures

Convergence

$$W_p(\nu_\sigma, \mu) \leq K(p, C)\sigma$$

The Score Operator

Density

$$\frac{d\nu_\sigma}{d\mu_\sigma}(w) = \exp(\Phi_\sigma(w))$$

The Score

$$D_{H_0}\Phi_\sigma = D_{H_0} \log \frac{d\nu_\sigma}{d\mu_\sigma}$$

$$D_{H_0}\Phi_\sigma : H \to H_0^*, \text{ assume Fréchet differentiability}$$

Denoising Score Matching

$$\sigma^2 C D_{H_0}\Phi_\sigma = \arg\min_{s_\sigma} \mathbb{E}_{\xi\sim\mu_\sigma}\mathbb{E}_{u\sim\mu}\left\|\sigma^{-1}C^{-1/2}(u - s_\sigma(u+\xi))\right\|_H^2$$

Preconditioned Langevin Dynamics

Preconditioned Langevin Equation

$$dx_t = -x_t + \sigma^2 C D_{H_0} \Phi_\sigma(x_t)\, dt + \sqrt{2} dz_t^{(\sigma)}$$

$z_t^{(\sigma)}$ is a $\sigma^2 C$-Wiener process

x_t has invariant measure ν_σ

Reparametrization

Optimize: $\arg\min_\theta \mathbb{E}_{\xi \sim \mu_\sigma} \mathbb{E}_{u \sim \mu} \| \xi + F_\theta(u + \xi) \|_H^2$

Discretize: $dx_t = F_\theta(x_t)\, dt + \sqrt{2} dz_t^{(\sigma)}$

Regularity Gap

Assumptions

$$H = \dot{L}^2(\mathbb{T}^d), \ \ \mu = N(0, C_1), \ \ \mu_\sigma = N(0, \sigma^2 C_2)$$
$$C_1 = (-\Delta)^{-\alpha_1}, \ \ C_2 = (-\Delta)^{-\alpha_2}, \ \ \alpha_1, \alpha_2 > d/2$$

Regularity Gap

$\mu(H_0) = 1$ is satisfied iff $\alpha_1 - \alpha_2 > d/2$

Dissipative Dynamics

$\mu(H_0) = 1$ satisfied if μ is a pushforward under a smoothing map

Smoothing Operators

Perturbation

$$\nu_\sigma = (A_\sigma)_\sharp \mu * \mu_\sigma, \quad A_\sigma : H \to H_0$$

$\nu_\sigma \sim \mu_\sigma$ equivalent with no assumptions on μ

Idea: apply smoothing to data $A_\sigma \to I$ as $\sigma \to 0$

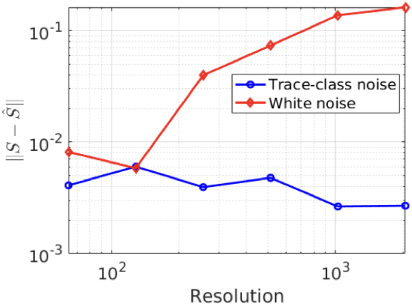
Example

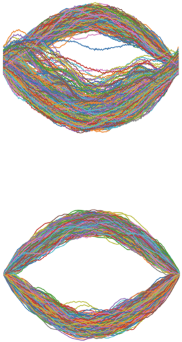
$A_\sigma = e^{\sigma\Delta}$ solution operator for heat equation

Numerical Examples

Mixture of Gaussian Fields

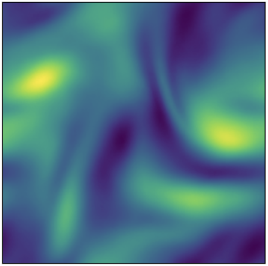
Using trace-class noise yields a resolution invariant map

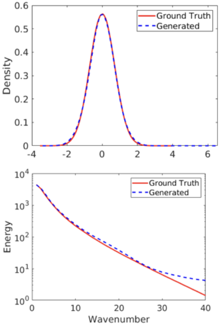




Super Resolution for Navier-Stokes

Super-resolution for NS preserves data statistics





Super Resolution for MNIST

Figure: 64×64

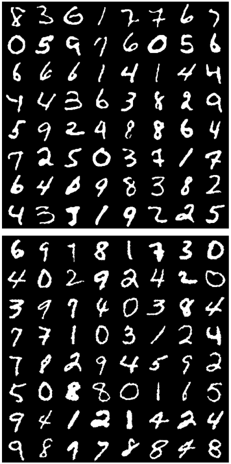


Figure: 128×128

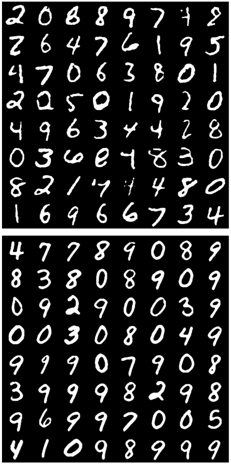
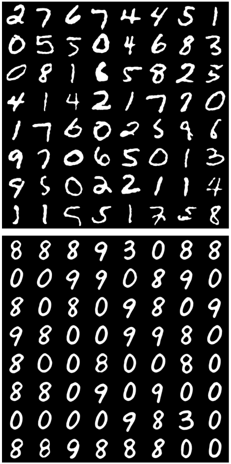


Figure: 256×256



Conclusion

This Talk

- Infinite dimensional score matching
- Operator learning + trace-class covariance
- Learn consistently across resolutions

Future Directions

- SDE formulation in infinite dimensions
- Flow ODE in infinite dimensions
- Bayesian inverse problems
- Rates for convergence of approximation