PATH-GUIDED PARTICLE-BASED SAMPLING

Mingzhou Fan, Ruida Zhou, Chao Tian, Xiaoning Qian
Presenter: Mingzhou Fan

July, 2024

Background

- **Goal:** Sample a target distribution with **unnormalized** distribution $\hat{p}$;
- **Particle-based sampling**: Construct a vector field $\phi_t : \mathbb{R}^n \to \mathbb{R}^n$ such that the process

$$\frac{\mathrm{d}x_t}{\mathrm{d}t} = \phi_t(x_t), \quad x_0 \sim p_0$$

  would drive the distribution of particles to a target distribution;
- **Langevin Dynamics (LD):**

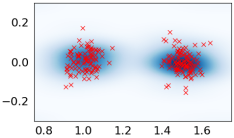
$$dx_t = \nabla \ln \hat{p}(x_t)dt + \sqrt{2}d\mathbf{B}_t,$$

  Discretized to:

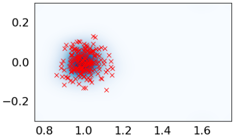
$$x_{t+\Delta t} = x_t + \Delta t \nabla \ln \hat{p}(x_t) + \sqrt{2\Delta t}\epsilon;$$

Motivation

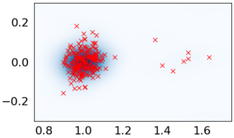
Existing particle-based methods suffer from slow mixing for some target distributions, taking LD as an example:



(a) Target



(b) LD 100 iters



(c) LD 4,000 iters

Annealing-based methods can utilize relatively smooth transitions between **intermediate distributions** to achieve better performance.
Is it possible to combine annealing-based methods with particle-based sampling?

Methodology

PGPS: Path-Guided Particle-based Sampling

- Given a partition-free density process $\{\hat{p}_t\}_{t\in[0,1]}$ and its normalized densities $\{p_t\}_{t\in[0,1]}$, with $\hat{p}_0 = p_0$ being the initial distribution and p_1 being the target with existed derivatives.
- We wish to construct a vector field $\phi_t^\theta : \mathbb{R}^n \to \mathbb{R}^n$ such that the process
$$\frac{\mathrm{d}x_t}{\mathrm{d}t} = \phi_t^\theta(x_t), \quad x_0 \sim p_0$$
would evolve the distribution of particles through the path;
- The time evolution of the density should satisfy the continuity equation:
$$\frac{\partial}{\partial t} p_t(x) = -\nabla \cdot (p_t(x)\phi_t^\theta(x))$$
- **Goal:** Find ϕ_t^θ to guide the samples to follow a normalized density path $\{p_t\}_{t\in[0,1]}$.

Methodology

Using $\hat{p}_t$ to guide x_t

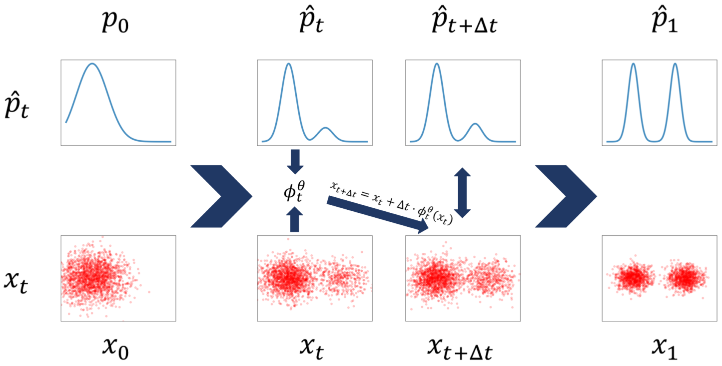


Figure: Schematic of PGPS

Algorithm

Minimizing the infinite sample loss to find a solution to the continuity function:

$$L_t^{inf}(\theta) = \mathbb{E}_{x\sim p_t} \left| \frac{\partial \ln \hat{p}_t(x)}{\partial t} + (\nabla \ln \hat{p}_t(x) + \nabla) \cdot \phi_t^\theta(x) - \mathbb{E}_{x\sim p_t} \frac{\partial \ln \hat{p}_t(x)}{\partial t} \right|^2;$$

Algorithm can be summarized into:

1. Sample initial distribution and get $\{x^i\}$;
2. Learn ϕ_t^θ, by minimizing

$$L_t(\theta) = \frac{1}{N} \sum_i \left| \frac{\partial \ln \hat{p}_t(x^i)}{\partial t} + (\nabla \ln \hat{p}_t(x^i) + \nabla) \cdot \phi_t^\theta(x^i) - \frac{1}{N} \sum_i \frac{\partial \ln \hat{p}_t(x^j)}{\partial t} \right|^2;$$

3. Update particles $x_t \leftarrow x_t + \phi_t^\theta(x_t)\Delta t$;
4. Update time $t \leftarrow t + \Delta t$;
5. if $t \leq 1$, go to step 2;

Theoretical Analysis

Approximation error $\delta := \min_{\phi_t \in \Phi_t} \sqrt{\int_0^1 \mathbb{E}_{x \sim p_t}[\|\phi_t^\theta(x) - \phi_t(x)\|^2] dt}$

Wasserstein error

The Wasserstein distance of the resulting distribution by the algorithm and the target distribution can be bounded by

$$W_2(p_{\hat{x}_1^\theta}, p_{x_1}) \leq \delta \sqrt{\exp(1 + 2K_1)} + \sqrt{\Delta t} \sqrt{\frac{C(\exp(1 + K_1^2) - 1)}{1 + K_1^2}},$$

where C and K_1 are Lipschitz constants for the vector fields.

Loss $\leftrightarrow$ Approximation error

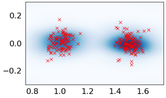
For any ϕ_t^θ, the approximation error

$$\delta \leq K \cdot L_t^{inf}(\theta),$$

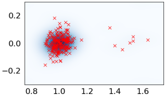
where $K > 0$ is a universal constant factor.

Empirical Results

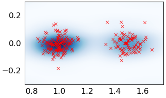
PGPS on the motivate example:



(a) Target



(b) LD 4,000 iters



(c) PGPS 650 iters

Inference performance on benchmark datasets with benchmark methods

Accuracy (ACC) $\uparrow$				
	PGPS	SVGD	SGLD	PFG
SONAR	**0.7981** $\pm$ 0.023	0.7962 $\pm$ 0.016	0.7942 $\pm$ 0.024	0.7673 $\pm$ 0.033
WINEWHITE	0.4520 $\pm$ 0.010	0.4520 $\pm$ 0.010	**0.4831** $\pm$ 0.049	0.4520 $\pm$ 0.010
WINERED	**0.5938** $\pm$ 0.018	0.5770 $\pm$ 0.018	0.5107 $\pm$ 0.096	0.5723 $\pm$ 0.019
AUSTRALIAN	0.8620 $\pm$ 0.009	0.8626 $\pm$ 0.006	0.7362 $\pm$ 0.157	**0.8643** $\pm$ 0.006
HEART	**0.2556** $\pm$ 0.142	0.1801 $\pm$ 0.042	0.2384 $\pm$ 0.135	0.1762 $\pm$ 0.033
GLASS	**0.5850** $\pm$ 0.080	0.5383 $\pm$ 0.076	0.4561 $\pm$ 0.152	0.4505 $\pm$ 0.071
COVERTYPE	**0.5899** $\pm$ 0.095	0.4867 $\pm$ 0.006	0.5221 $\pm$ 0.084	0.5088 $\pm$ 0.053

Expected Calibration Error (ECE) $\downarrow$				
	PGPS	SVGD	SGLD	PFG
SONAR	0.2517 $\pm$ 0.057	0.1712 $\pm$ 0.020	0.3394 $\pm$ 0.049	**0.1678** $\pm$ 0.050
WINEWHITE	**0.0750** $\pm$ 0.011	0.0988 $\pm$ 0.012	0.0935 $\pm$ 0.024	0.0876 $\pm$ 0.018
WINERED	**0.0366** $\pm$ 0.005	0.0402 $\pm$ 0.004	0.0868 $\pm$ 0.029	0.0449 $\pm$ 0.005
AUSTRALIAN	0.1703 $\pm$ 0.066	0.1713 $\pm$ 0.064	0.3517 $\pm$ 0.078	**0.1457** $\pm$ 0.047
HEART	**0.4579** $\pm$ 0.071	0.5117 $\pm$ 0.064	0.5110 $\pm$ 0.114	0.4887 $\pm$ 0.089
GLASS	**0.1142** $\pm$ 0.008	0.1155 $\pm$ 0.006	0.2157 $\pm$ 0.025	0.1289 $\pm$ 0.021
COVERTYPE	**0.0743** $\pm$ 0.016	0.0950 $\pm$ 0.012	0.1301 $\pm$ 0.038	0.0926 $\pm$ 0.078

Thanks!