Zeroth-Order Non-Log-Concave Sampling with Variance Reduction and Applications to Inverse Problems

M. Berk Sahin, Behzad Sharif, Abolfazl Hashemi

Introduction

- We consider a **fundamental** problem of sampling from a **non-log-concave** distribution $\pi \propto \exp(-f(x))$ *via Langevin Monte Carlo (LMC)* algorithm:

$$x_{(k+1)\gamma} = x_{k\gamma} - \gamma \nabla f(x_{k\gamma}) + \sqrt{2}(B_{(k+1)\gamma} - B_{k\gamma}),$$

where $(B_{\mathrm{t}})_{\mathrm{t}\geq 0}$ denotes Brownian motion, and γ is step size.

- Many applications involve **unavailable** or prohibitively **expensive** gradients, motivating **zeroth-order (ZO) LMC** methods[1,2,3]

$$x_{(k+1)\gamma} = x_{k\gamma} - \frac{\gamma}{b}\sum_{i=1}^{b} \tilde{\nabla} f_\mu(x_{k\gamma}, u_k^i) + \sqrt{2}(B_{(k+1)\gamma} - B_{k\gamma}),$$

where

$$\tilde{\nabla} f_\mu(x_{k\gamma}, u_k^i) = \frac{f(x_{k\gamma}+\mu u_k^i) - f(x_{k\gamma})}{\mu} u_k^i, \qquad \text{where} \quad u_k^i \sim \mathrm{N}(0, I).$$

i: Gaussian vector index $\qquad k$: iteration index

1. Liu et al: OPT2020 Workshop 2020. 2. Roy et al: Bernoulli 2022. 3. He et al: NeurIPS 2024.

Motivation

Challenge 1: Non-asymptotic convergence guarantees for ZO-LMC sampling is limited by **strongly log-concave** target distributions π, which does not hold in general.

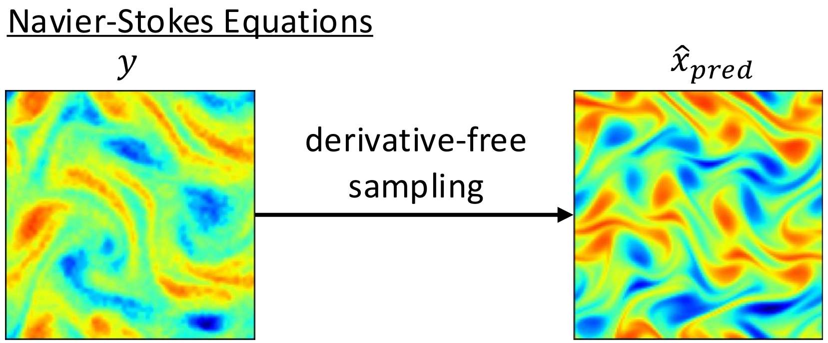
Challenge 2: For accurate ZO estimation of the gradient, batch size should be $b = \mathcal{O}(d)$, which is **impractical** in high-dimensional settings due to memory budgets or computation costs per iteration.

- **Derivative-free** methods[1,2,3] gained increasing attention to solve black-box inverse problems

$$y = A(x) + n, \qquad n \sim \mathrm{N}(0, \sigma_{noise}^2 I),$$

where we have input-output (black-box) access to the forward model A.

Challenge 3: Existing derivative-free methods rely on heuristic approximations, preventing convergence guarantees in sampling.



1. Huang et al: ICML 2024. 2. Tang et al: AISTATS 2024. 3. Zheng et al: TMLR 2025.

Methods

- We propose a **variance-reduced ZO estimator** to reduce per-iteration cost $\mathrm{b} = \mathcal{O}(d) \rightarrow \mathcal{O}(1)$:

$$g_k = \begin{cases} \frac{1}{b}\sum_{i=1}^{b} \tilde{\nabla} f_\mu(x_{k\gamma}, u_k^i) & w.p. \quad p \\ g_{k-1} + \frac{1}{b'}\sum_{i=1}^{b'} \tilde{\nabla} f_\mu(x_{k\gamma}, u_k^i) - \tilde{\nabla} f_\mu(x_{(k-1)\gamma}, u_k^i) & w.p. \;\; 1-p, \end{cases}$$

where $b' \ll b$ and p and is probability of using a large batch size.

- This estimate is accurate and reduces variance given a Lipschitz continuous potential function.

Assumption 1: *The potential f is L_1-Lipschitz continuous: $\|f(x_1) - f(x_2)\|_2 \leq L_1\|x_1 - x_2\|_2$ for all $x_1, x_2 \in \mathbb{R}^d$ and for some $L_1 > 0$.*

- Using this estimator, we propose **variance-reduced ZO LMC sampling** algorithm

$$x_{(k+1)\gamma} = x_{k\gamma} - \gamma \boldsymbol{g_k} + \sqrt{2}(B_{(k+1)\gamma} - B_{k\gamma}).$$

- We additionally require the following standard assumption for controlling the discretization error of LMC.

Assumption 2: *The gradient of the potential f is L_2-Lipschitz continuous: $\|\nabla f(x_1) - \nabla f(x_2)\|_2 \leq L_2\|x_1 - x_2\|_2$ for all $x_1, x_2 \in \mathbb{R}^d$ and for some $L_2 > 0$*

Theoretical Results

- In our theoretical analysis, we consider the time interpolation of the proposed algorithm:

$$x_t = x_{k\gamma} - (t - k\gamma)\boldsymbol{g_k} + \sqrt{2}(B_t - B_{k\gamma}) \qquad \text{for } t \in [k\gamma, (k+1)\gamma].$$

Theorem 1 (Informal): *Let $\pi \propto exp(-f(x))$ be the target distribution with f satisfying Assumption 1 and 2. Then, for some $\gamma > 0$, $p \in (0,1]$, $\mu > 0$, and $b > 0$, the distribution $\bar{\nu}_{N\gamma}$ of random sample x_t for $t \sim U[0, N\gamma]$ satisfies $\mathrm{FI}(\bar{\nu}_{N\gamma}||\pi) \leq \varepsilon$ after $\mathcal{O}(\mathrm{d}^7/\varepsilon^4)$ iterations with $\mathcal{O}(1)$ function evaluations per iteration.*

Theorem 2 (Informal): *Let $\pi \propto exp(-f(x))$ be the target distribution with f satisfying Assumption 1 and 2. Then, for some $\gamma_k > 0$, $p_k \in (0,1]$, $\mu_k > 0$, and $b_k > 0$, the distribution $\bar{\nu}_{N\gamma}$ of random sample x_t for $t \sim U[0, N\gamma]$ converges weakly to π with $\mathcal{O}(1)$ function evaluations per iteration.*

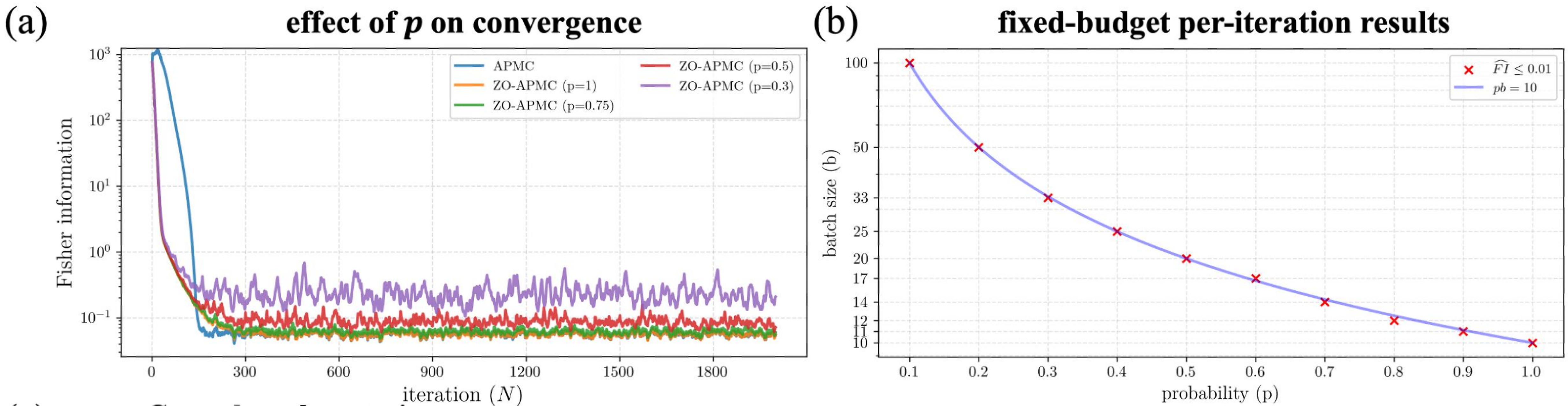
- π satisfies **Poincaré inequality** → $\|\bar{\nu}_{N\gamma} - \pi\|_{\mathrm{TV}}^2 \leq \varepsilon$ *after* $\mathcal{O}(\mathrm{d}^7/\varepsilon^4)$ *iterations with* $\mathcal{O}(1)$ *function evaluations per iteration.*
- Using Bayes' rule, and score network prior, we propose and analyze **ZO-APMC** for black-box inverse problems:

$$x_{(k+1)\gamma} = x_{k\gamma} - (t - k\gamma)(g_k - \alpha_k \mathcal{S}_\theta(x_{k\gamma}, \sigma_k)) + \sqrt{2}(B_t - B_{k\gamma})$$

- **Controlled perturbation** + **bounded error** → **same theoretical guarantees!**

Empirical Results: Toy Experiments

- 2D experiments with random A; results report the mean FI of ZO-APMC (proposed method) over 20 seeds.
- ZO-APMC **converges in FI** for different p values.
- ZO-APMC converges to $\varepsilon = 0.01$ with the same per-iteration cost pb + (1−p)b' = $\mathcal{O}(1)$ and different p and b values.



Empirical Results: MRI and Black-Hole Imaging

- Across MRI reconstruction and black-hole imaging, ZO-APMC **consistently outperforms** black-box sampling baselines on almost all metrics.

MRI Reconstruction

	PSNR (dB)↑	SSIM↑	NRMSE↓	SD↓	MSE↓
PnPDM	30.81	0.946	3.76e-2	2.16e-2	8.46e-4
DPS	34.38	0.965	2.54e-2	2.06e-2	4.07e-4
APMC	<u>36.55</u>	<u>0.973</u>	<u>1.99e-2</u>	<u>2.0e-2</u>	<u>2.55e-4</u>
Forward-GSG	27.8	0.918	5.42e-2	3.26e-2	19.1e-4
Central-GSG	27.78	0.917	5.43e-2	3.27e-2	19.2e-4
SCG	7.1	0.711	7.67	1.38	0.21
DPG	32.17	0.953	5.4e-2	**2.69e-2**	6.5e-4
EnKG	31.32	0.934	5.72e-2	2.92e-2	6.72e-4
ZO-APMC (ours)	**35.29**	**0.966**	**2.28e-2**	2.99e-2	**3.29e-4**

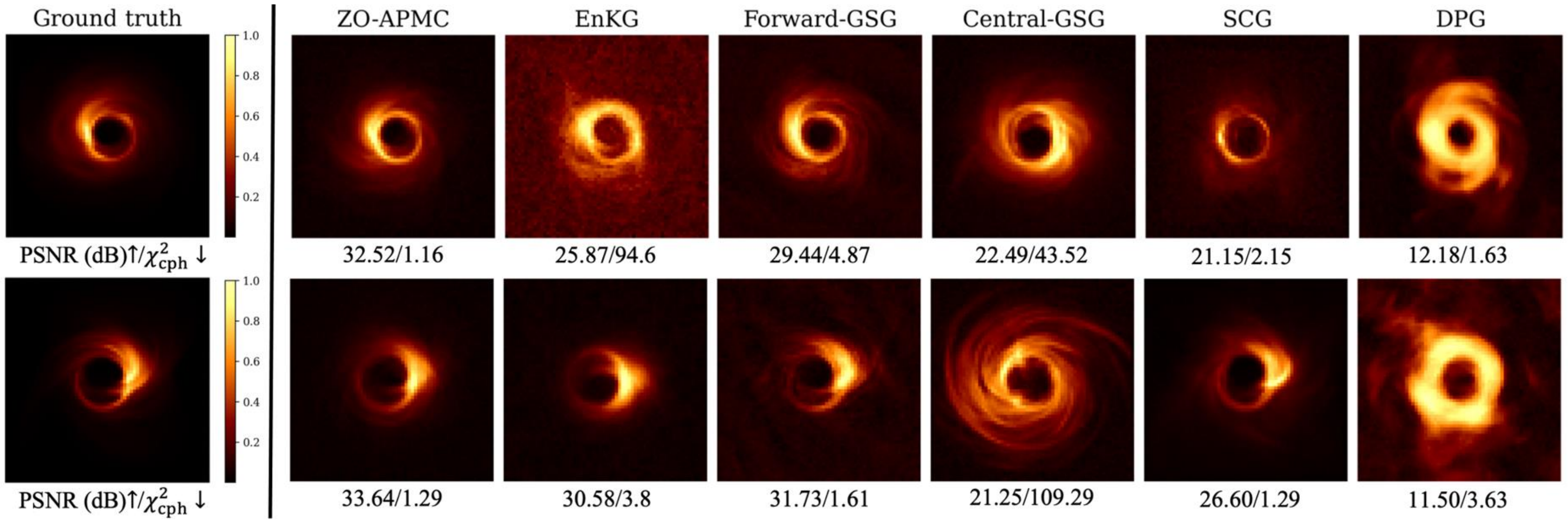
Black-Hole Imaging

	PSNR↑	Blurred PSNR↑	$\chi^2_{\mathrm{cph}}\downarrow$	$\chi^2_{\mathrm{camp}}\downarrow$	SD↓
PnPDM	<u>26.48</u>	32.31	<u>11.48</u>	23.54	4.5e-2
DPS	25.61	<u>30.84</u>	12.39	17.72	<u>4.32e-2</u>
APMC	26.23	31.32	11.78	19.23	4.34e-2
Forward-GSG	26.21	31.47	6.77	14.06	2.99e-2
Central-GSG	21.63	23.73	80.31	78.5	4.5e-2
SCG	22.21	25.51	23.72	14.23	1.7e-2
DPG	12.33	14.02	8.17	30.44	**1.6e-2**
EnKG	22.86	27.69	64.37	33.44	0.925
ZO-APMC (Ours)	**26.71**	**32.86**	**5.42**	**11.23**	3.02e-2

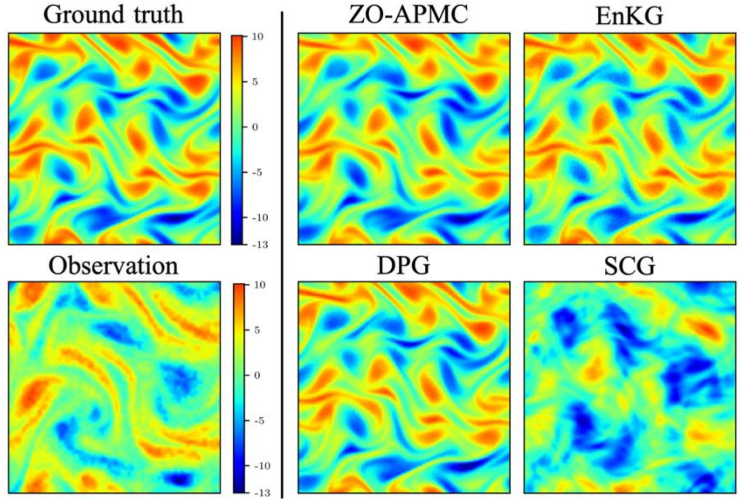
Empirical Results: Representative Cases

- For black-hole imaging, ZO-APMC reconstructs most closely match the ground truth compared to other baselines.
- For Navier-Stokes problem, ZO-APMC generates samples comparable to EnKG and DPG, while providing theoretical guarantees.

Black-Hole Imaging



Navier-Stokes



Conclusion

- We proposed **variance-reduced ZO** sampling algorithm with **first non-asymptotic** convergence guarantees for non-log-concave sampling.
- We introduce ZO-APMC, the **first derivative-free** black-box inverse problem solver with non-asymptotic convergence guarantees.
- We **validated** our theory experimentally and demonstrated **strong empirical performance** of ZO-APMC in **high-dimensional** settings.
- Future work includes extension of our method to **underdamped Langevin dynamics** and **latent diffusion models** for **improved** sampling efficiency.