Proceedings of Machine Learning Research vol 196:1–2, 2024 37th Annual Conference on Learning Theory

Algorithms for mean-field variational inference via polyhedral optimization in the Wasserstein space

Yiheng Jiang YJ2070@NYU.EDU
Courant Institute of Mathematical Sciences, New York University

Sinho Chewi SCHEWI@IAS.EDU
School of Mathematics, Institute for Advanced Study

Aram-Alexandre Pooladian ARAM-ALEXANDRE.POOLADIAN@NYU.EDU
Center for Data Science, New York University

Editors: Shipra Agrawal and Aaron Roth

Abstract

We develop a theory of finite-dimensional polyhedral subsets over the Wasserstein space and optimization of functionals over them via first-order methods. These sets are defined with respect to a fixed reference measure ρ, and a collection of *compatible* transport maps $\mathcal{M}$ (Panaretos and Zemel, 2020), written

$$\mathcal{P}_\diamond := \left\{ (\textstyle\sum_{T\in\mathcal{M}} \lambda_T T)_\sharp \rho \;\middle|\; \lambda \in \mathbb{R}_+^{|\mathcal{M}|} \right\},$$

where $\mathbb{R}_+^{|\mathcal{M}|}$ is the non-negative orthant. The assumption of compatibility entails strong consequences. Letting W_2 denote the 2-Wasserstein metric over probability distributions, we show that $(\mathcal{P}_\diamond, W_2)$ is *isometric* to $(\mathbb{R}_+^{|\mathcal{M}|}, \|\cdot\|_Q)$, where $\|\cdot\|_Q$ is a twisted *Euclidean* norm. This isometry allows us to optimize functionals over $\mathcal{P}_\diamond$ via lightweight first-order algorithms, and their stochastic variants. Moreover, this isometry property also holds when $\lambda \in \mathcal{K} \subseteq \mathbb{R}_+^{|\mathcal{M}|}$ for any convex set $\mathcal{K}$, which permits us to analyze algorithms such as Frank–Wolfe.

The second part of our paper concerns our main application: the problem of mean-field variational inference, which seeks to approximate a distribution $\pi \propto \exp(-V)$ over $\mathbb{R}^d$ by a product measure with respect to the KL divergence, where $0 \prec \ell_V I \preceq \nabla^2 V \preceq L_V I$, with condition number $\kappa = L_V/\ell_V$. The premise of this application is the observation that the space of product measures $\mathcal{P}(\mathbb{R})^{\otimes d}$ can be approximated by $\mathcal{P}_\diamond$ for suitably chosen $\mathcal{M}$ when π has suitable structure, i.e.,

$$\operatorname*{arg\,min}_{\mu\in\mathcal{P}(\mathbb{R})^{\otimes d}} \mathrm{KL}(\mu \,\|\, \pi) =: \pi^\star \approx \pi^\star_\diamond := \operatorname*{arg\,min}_{\mu\in\mathcal{P}_\diamond} \mathrm{KL}(\mu \,\|\, \pi)\,.$$

Concretely, we show that $\sqrt{\ell_V}\, W_2(\pi^\star_\diamond, \pi^\star) \le \varepsilon$ when ρ is the standard Gaussian, and

- $\mathcal{M}$ follows a piecewise linear construction, with $|\mathcal{M}| \le \tilde{O}(\kappa^2 d^{3/2}/\varepsilon)$,
- $\mathcal{M}$ follows a higher-order construction, with $|\mathcal{M}| \le \tilde{O}(\kappa^{3/2} d^{5/4}/\varepsilon^{1/2})$.

We accompany the above results with stochastic optimization guarantees, and have implemented the piecewise linear construction here. Our proofs hinge on regularity theory for the optimal transport map from ρ to $\pi^\star$, as well as novel smoothness results for entropy over the family $\mathcal{P}_\diamond$.[1]
Keywords: mean-field variational inference, optimization, Wasserstein gradient flows

1. Extended abstract. Full version appears as [arXiv:2312.02849].

References

Victor M Panaretos and Yoav Zemel. *An invitation to statistics in Wasserstein space*. Springer Nature, 2020.