Dimension-free Statistical and Computational Guarantee for Optimal Transport

Triad: Tianyi Lin (darren_lin@berkeley.edu), Michael Jordan (jordan@cs.berkeley.edu), Marco Cuturi (cuturi@google.com)

Context. Recent years have witnessed an explosion of ML papers inspired by optimal transport (OT) theory [Villani, 2008]. In particular, using the minimal distance estimation (MDE) [Basu et al., 2011] framework with OT distances has resulted in several works, from training of generative models [Arjovsky et al., 2017] and auto-encoders [Tolstikhin et al., 2018], to clustering [Cuturi and Doucet, 2014], trajectory inference [Hashimoto et al., 2016], and non-parametric testing [Ramdas et al., 2017]. Several fields of science have also been influenced by OT, notably single cell genomics [Schiebinger et al., 2019] or neuroimaging [Janati et al., 2020].

Why this project? For OT ideas to continue to bear fruit in ML, it will be necessary to tackle longstanding challenges, from both statistical and computational points of view: the computation is notoriously expensive, and the plug-in estimation suffers from the curse of dimensionality [Fournier and Guillin, 2015, Weed and Bach, 2019]. Therefore, the aforementioned applications should, in principle, somewhat fail in high-dimensional setting despite being theoretically motivated [Bernton et al., 2019]. Although assuming smoothness of the underlying densities may help [Weed and Berthet, 2019, Hütter and Rigollet, 2021], the computational complexities of these estimators degrade exponentially with the dimension. In practice, these issues do play a role since they entail a lack of robustness and instability with respect to inputs [Ling and Okada, 2007, Pele and Werman, 2009, Paty and Cuturi, 2019].

Plan. Our goal in this project is to bridge the statistical-computational gap along two directions: (1) compute a plug-in estimator based on low-dimensional projections; (2) compute a dimension-free statistical estimator based on kernel mean embeddings and positive definite operator characterization.

1. Project and Estimate. Paty and Cuturi [2019] proposed to seek the k-dimensional subspace ($k > 1$) that maximizes the OT distance between two measures after projection, resulting in a *projection robust* Wasserstein (PRW) distance, providing a generalization of the popular sliced Wasserstein distance [Rabin et al., 2011]. During the past year, we have focused on statistical and computational properties of the plug-in estimation of PRW and derived a bunch of results, including dimension-free sample complexity, consistency and central-limit theorem under model misspecification, nonconvex max-min computational model and provably efficient algorithms based on Riemannian optimization (NeurIPS'20 + AISTATS'21 papers). We will extend these works by studying their differentiability, in order for them to fit naturally in the `OTT` toolbox[1].

2. RKHS Potentials. Vacher et al. [2021] proposed an alternative approach a few weeks ago, with follow-up potential: their dimension-free computational upper bound was established using an interior-point method (IPM) to solve a large-scale conic problem. We find great opportunity in replacing IPMs with new algorithms, able to exploit the problem structure and bypass the computation, storage, and factorization of a large-scale Schur complement matrix. Our plan includes: (1) Find the dimension-free statistical estimators for multimarginal OT (MOT) [Pass, 2015] and Wasserstein barycenter problem (WBP) [Cuturi and Doucet, 2014] (both problems have been recognized as the backbone of numerous applications). (2) Develop efficient algorithms for new estimators for OT, MOT and WBP and demonstrate their practical implications in ML.

Potential for collaboration. These projects build on several previous projects on algorithmic OT in collaboration with Google researcher M. Cuturi, developing provably efficient algorithms for OT, MOT and WBP and providing novel computational hardness results for fixed-support WBP (3 published, 1 submitted works together with T. Lin and M.I. Jordan). An important expected outcome will be to improve the `OTT` toolbox with new algorithms.

[1]https://github.com/google-research/ott

References

M. Arjovsky, S. Chintala, and L. Bottou. Wasserstein generative adversarial networks. In *ICML*, pages 214–223, 2017.

A. Basu, H. Shioya, and C. Park. *Statistical Inference: The Minimum Distance Approach*. CRC Press, 2011.

E. Bernton, P. E. Jacob, M. Gerber, and C. P. Robert. On parameter estimation with the wasserstein distance. *Information and Inference: A Journal of the IMA*, 8(4):657–676, 2019.

M. Cuturi and A. Doucet. Fast computation of Wasserstein barycenters. In *ICML*, pages 685–693, 2014.

N. Fournier and A. Guillin. On the rate of convergence in Wasserstein distance of the empirical measure. *Probability Theory and Related Fields*, 162(3-4):707–738, 2015.

T. Hashimoto, D. Gifford, and T. Jaakkola. Learning population-level diffusions with generative RNNs. In *ICML*, pages 2417–2426, 2016.

J-C. Hütter and P. Rigollet. Minimax estimation of smooth optimal transport maps. *The Annals of Statistics*, 49(2):1166–1194, 2021.

H. Janati, T. Bazeille, B. Thirion, M. Cuturi, and A. Gramfort. Multi-subject MEG/EEG source imaging with sparse multi-task regression. *NeuroImage*, page 116847, 2020.

H. Ling and K. Okada. An efficient earth mover's distance algorithm for robust histogram comparison. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(5):840–853, 2007.

B. Pass. Multimarginal optimal transport: theory and applications. *ESAIM: Mathematical Modelling and Numerical Analysis*, 49(6):1771–1790, 2015.

F-P. Paty and M. Cuturi. Subspace robust Wasserstein distances. In *ICML*, pages 5072–5081, 2019.

O. Pele and M. Werman. Fast and robust earth mover's distances. In *ICCV*, pages 460–467. IEEE, 2009.

J. Rabin, G. Peyré, J. Delon, and M. Bernot. Wasserstein barycenter and its application to texture mixing. In *International Conference on Scale Space and Variational Methods in Computer Vision*, pages 435–446, 2011.

A. Ramdas, N. G. Trillos, and M. Cuturi. On Wasserstein two-sample testing and related families of nonparametric tests. *Entropy*, 19(2):47, 2017.

G. Schiebinger, J. Shu, M. Tabaka, B. Cleary, V. Subramanian, A. Solomon, J. Gould, S. Liu, S. Lin, and P. Berube. Optimal-transport analysis of single-cell gene expression identifies developmental trajectories in reprogramming. *Cell*, 176(4):928–943, 2019.

I. Tolstikhin, O. Bousquet, S. Gelly, and B. Schoelkopf. Wasserstein auto-encoders. In *ICLR*, 2018.

A. Vacher, B. Muzellec, A. Rudi, F. Bach, and F-X. Vialard. A dimension-free computational upper-bound for smooth optimal transport estimation. *ArXiv Preprint: 2101.05380*, 2021.

C. Villani. *Optimal Transport: Old and New*, volume 338. Springer Science & Business Media, 2008.

J. Weed and F. Bach. Sharp asymptotic and finite-sample rates of convergence of empirical measures in Wasserstein distance. *Bernoulli*, 25(4A):2620–2648, 2019.

J. Weed and Q. Berthet. Estimation of smooth densities in Wasserstein distance. In *COLT*, pages 3118–3119. PMLR, 2019.