Improved Complexity Bounds in Wasserstein Barycenter Problem

Darina Dvinskikh[1]; Daniil Tiapkin[2]

[1]WIAS, Berlin; [2]HSE University, Moscow

Contribution.

We propose two algorithms to compute Wasserstein barycenters which are build upon the saddle-point problem reformulation.

- The first algorithm, Mirror Prox for WB, based on mirror prox with specific prox-function, has no limitations in contrast to regularized-based methods, which are numerically unstable under a small value of the regularization parameter.
- The second algorithm, Dual Extrapolation for WB, based on dual extrapolation scheme under the weaker convergence requirements of area-convexity, improves the complexity for the Wasserstein barycenter problem.

Approach	Paper	Complexity
IBP	[KTD+19]	$\widetilde{O}\left(\frac{mn^2\|C\|_\infty^2}{\varepsilon^2}\right)$
Accelerated IBP	[GDG19]	$\widetilde{O}\left(\frac{mn^2\sqrt{n}}{\varepsilon}\right)$
FastIBP	[LHC+20]	$\widetilde{O}\left(\frac{mn^2\sqrt[3]{n}}{\varepsilon\sqrt[3]{\varepsilon}}\right)$
Mirror prox with specific norm	This work	$\widetilde{O}\left(\frac{mn^2\sqrt{n}}{\varepsilon}\right)$
Dual extrapolation with area-convexity	This work	$\widetilde{O}\left(\frac{mn^2}{\varepsilon}\right)$

Reference.

[GDG19] Sergey Guminov, Pavel Dvurechensky, and Alexander Gasnikov. Accelerated alternating minimization. *arXiv preprint arXiv:1906.03622*, 2019.

[JST19] Arun Jambulapati, Aaron Sidford, and Kevin Tian. A direct $\tilde{O}(1/\varepsilon)$ iteration parallel algorithm for optimal transport. In *Advances in Neural Information Processing Systems*, pages 11359–11370, 2019.

[KTD+19] Alexey Kroshnin, Nazarii Tupitsa, Darina Dvinskikh, Pavel Dvurechensky, Alexander Gasnikov, and Cesar Uribe. On the complexity of approximating Wasserstein barycenters. In Kamalika Chaudhuri and Ruslan Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97, pages 3530–3540, 2019. arXiv:1901.08686.

[LHC+20] Tianyi Lin, Nhat Ho, Xi Chen, Marco Cuturi, and Michael I Jordan. Fixed-support wasserstein barycenters: Computational hardness and fast algorithm. 2020.

Preliminaries.

Given two histograms p, q from probability simplex Δ_n and ground cost $C \in \mathbb{R}_+^{n\times n}$, the optimal transport problem is formulated as follows

$$W(p,q) = \min_{X\in\mathcal{U}(p,q)} \langle C, X\rangle,$$

where X is a transport plan from transport polytope $\mathcal{U} = \{X \in \mathbb{R}_+^{n\times n}, X\mathbf{1} = p, X^\top\mathbf{1} = q\}$.

Let d be vectorized cost matrix of C, x be vectorized transport plan of X, $b = (p^\top, q^\top)$, and $A = \{0,1\}^{2n\times n^2}$ be an incidence matrix.

As $\sum_{i,j=1}^n X_{ij} = 1$, we following by the paper [JST19] rewrite the optimal transport problem as

$$W(p,q) = \min_{x\in\Delta_{n^2}} \max_{y\in[-1,1]^{2n}} \{d^\top x + 2\|d\|_\infty(\, y^\top A x - b^\top y)\}.$$

Problem Formulation.

We reformulate the problem of calculating Wasserstein barycenters of m histograms $q_1, q_2, \ldots, q_m \in \Delta_n$ as a saddle-point problem.

The Wasserstein barycenter problem is

$$p^* = \arg\min_{p\in\Delta_n} \frac{1}{m}\sum_{i=1}^m W(p, q_i).$$

The saddle-point formulation of this problem is

$$\min_{\mathbf{x}\in\mathcal{X}} \max_{\mathbf{y}\in\mathcal{Y}} \frac{1}{m}\left\{\boldsymbol{d}^\top\mathbf{x} + 2\|d\|_\infty\left(\mathbf{y}^\top\boldsymbol{A}\mathbf{x} - \boldsymbol{c}^\top\mathbf{y}\right)\right\},$$

where spaces

$\mathcal{X} \triangleq \underbrace{\Delta_{n^2}\times\ldots\times\Delta_{n^2}}_{m} \times \Delta_n$ and $\mathcal{Y} \triangleq [-1,1]^{2mn}$,

column vectors

$\mathbf{x} = (x_1^\top, \ldots, x_m^\top, p^\top)^\top \in \mathcal{X}$, $\mathbf{y} = (y_1^\top, \ldots, y_m^\top)^\top \in \mathcal{Y}$,

$\boldsymbol{d} = (d^\top, \ldots, d^\top, \mathbf{0}_n^\top)^\top$, $\boldsymbol{c} = (\mathbf{0}_n^\top, q_1^\top, \ldots, \mathbf{0}_n^\top, q_m^\top)^\top$, and

$\boldsymbol{A} = \left(\hat{A}\;\mathcal{E}\right) \in \{-1,0,1\}^{2mn\times(mn^2+n)}$ with block-diagonal matrix $\hat{A} = \mathrm{diag}\{A, A, \ldots, A\}$ of m blocks, and matrix

$\mathcal{E}^\top = \left(\left(-I_n\; 0_{n\times n}\right)\left(-I_n\; 0_{n\times n}\right)\cdots\left(-I_n\; 0_{n\times n}\right)\right).$

Numerical Experiments.

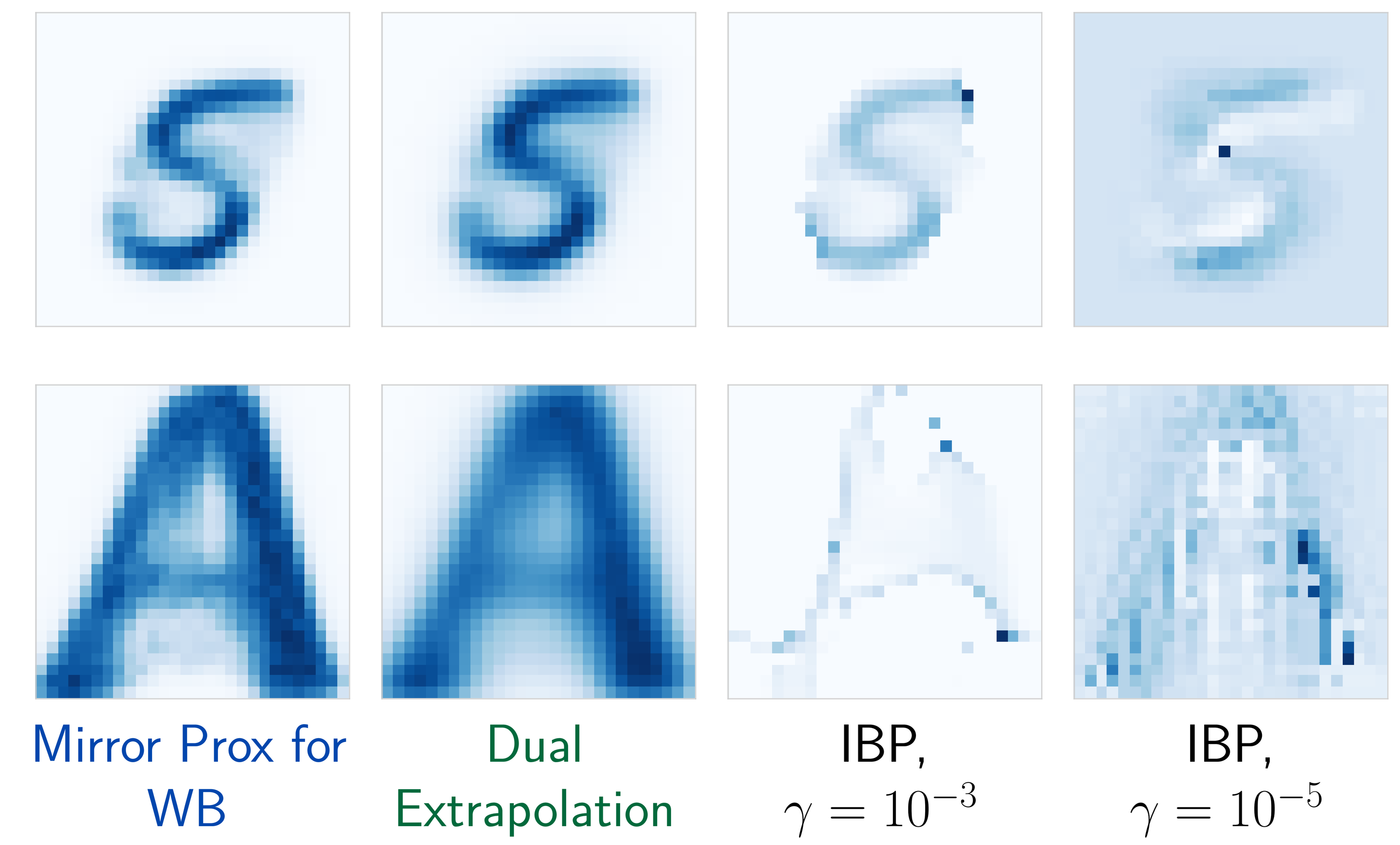

Fig.1. Wasserstein barycenters of hand-written digits '5' and letters 'A' computed by Mirror Prox for WB, and Dual Extrapolation for WB, and the IBP algorithm with small values of the regularizing parameter.

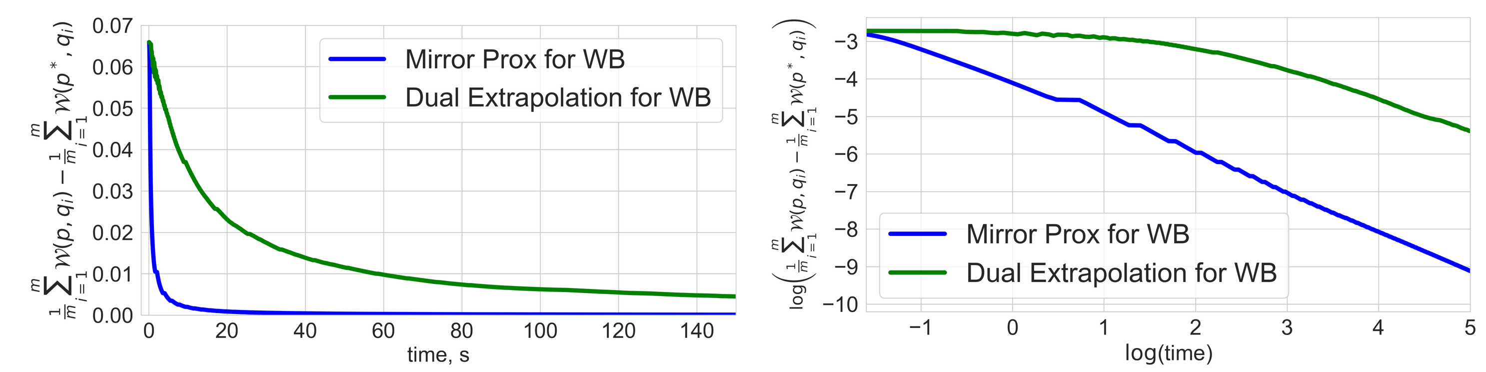

Fig.2. Convergence of Mirror Prox for WB and Dual Extrapolation for WB to the true barycenter of Gaussian measures w.r.t the function optimality gap $\frac{1}{m}\sum_{i=1}^m W(p, q_i) - \frac{1}{m}\sum_{i=1}^m W(p^*, q_i)$.

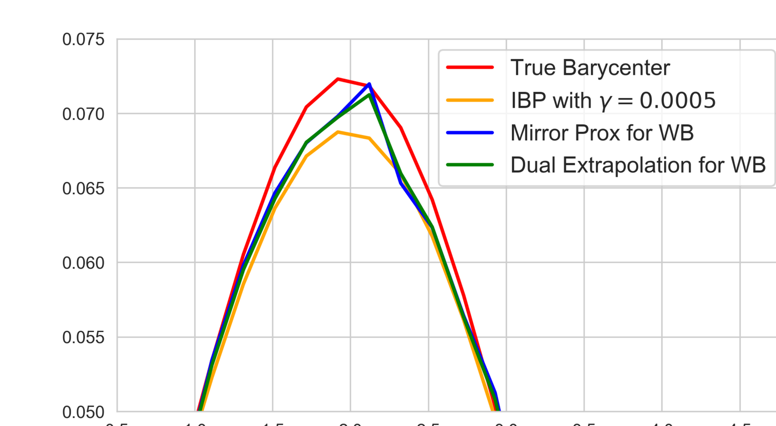

Fig.3. Convergence of the barycenters obtained by Mirror Prox for WB, and Dual Extrapolation for WB, and the IBP to the true barycenter of Gaussian measures.