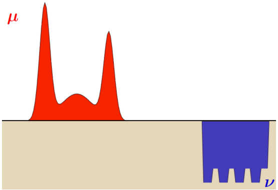
A Riemannian Block Coordinate Descent Method for Computing the Projection Robust Wasserstein Distance

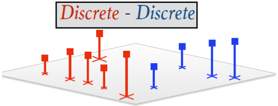
Minhui Huang [1] Shiqian Ma [2] Lifeng Lai [1]

[1]Department of Electrical and Computer Engineering
[2]Department of Mathematics
University of California, Davis

June 16, 2021

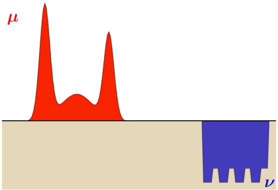
Wasserstein Distance

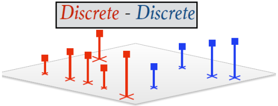




- The 2-Wasserstein distance: $\mathcal{W}(\mu, \nu) := \left(\inf_{\pi \in \Pi(\mu,\nu)} \int \|x - y\|^2 d\pi(x, y)\right)^{1/2}$.

Wasserstein Distance





- The 2-Wasserstein distance: $\mathcal{W}(\mu, \nu) := \left(\inf_{\pi \in \Pi(\mu,\nu)} \int \|x - y\|^2 d\pi(x, y)\right)^{1/2}$.
- Consider discrete empirical measures $\hat{\mu}_n = \sum_{i=1}^n r_i \delta_{x_i}, \hat{\nu}_n = \sum_{i=1}^n c_i \delta_{y_i}, r_i = c_j = \frac{1}{n}$.

$$\mathcal{W}^2(\hat{\mu}_n, \hat{\nu}_n) \;=\; \min_{\pi \in \Pi(\hat{\mu}_n, \hat{\nu}_n)} \langle C, \pi \rangle, \tag{1}$$

where $C_{ij} = \|x_i - y_j\|^2$ for $x_i, y_j \in \mathbb{R}^d$, $\Pi(\hat{\mu}_n, \hat{\nu}_n) = \{\pi \in \mathbb{R}_+^{n \times n} | \pi \mathbf{1} = r, \pi^\top \mathbf{1} = c\}$.

Projection Robust Wasserstein Distance (PRW)

- Sample complexity of Wasserstein Distance: $\mathbb{E}|\mathcal{W}(\mu, \nu) - \mathcal{W}(\hat{\mu}_n, \hat{\nu}_n)| = O(n^{-1/d})$.

Projection Robust Wasserstein Distance (PRW)

- Sample complexity of Wasserstein Distance: $\mathbb{E}|\mathcal{W}(\mu,\nu) - \mathcal{W}(\hat{\mu}_n, \hat{\nu}_n)| = O(n^{-1/d})$.
- Denote $\mathcal{G}_k = \{R \subset \mathbb{R}^d | dim(E) = k\}$ as the Grassmannian of k-dimensional subspace.

$$\begin{aligned} \mathcal{P}_k(\mu,\nu) &= \sup_{E\in\mathcal{G}_k} \mathcal{W}(\mathrm{Proj}_E \mu, \mathrm{Proj}_E \nu) \\ &= \max_{E\in\mathcal{G}_k} \min_{\pi\in\Pi(\mu,\nu)} \left(\int \|\mathrm{Proj}_E(x-y)\|^2 d\pi(x,y) \right)^{1/2}. \end{aligned} \tag{2}$$

Projection Robust Wasserstein Distance (PRW)

- Sample complexity of Wasserstein Distance: $\mathbb{E}|\mathcal{W}(\mu,\nu) - \mathcal{W}(\hat{\mu}_n, \hat{\nu}_n)| = O(n^{-1/d})$.
- Denote $\mathcal{G}_k = \{R \subset \mathbb{R}^d | dim(E) = k\}$ as the Grassmannian of k-dimensional subspace.

$$\begin{aligned} \mathcal{P}_k(\mu,\nu) &= \sup_{E \in \mathcal{G}_k} \mathcal{W}(\text{Proj}_E \mu, \text{Proj}_E \nu) \\ &= \max_{E \in \mathcal{G}_k} \min_{\pi \in \Pi(\mu,\nu)} \left(\int \|\text{Proj}_E(x-y)\|^2 d\pi(x,y) \right)^{1/2}. \end{aligned} \tag{2}$$

- Sample complexity of PRW: $\mathbb{E}|\mathcal{P}(\mu,\nu) - \mathcal{P}(\hat{\mu}_n, \hat{\nu}_n)| = O(n^{-1/k})$.

Projection Robust Wasserstein Distance (PRW)

- Sample complexity of Wasserstein Distance: $\mathbb{E}|\mathcal{W}(\mu,\nu) - \mathcal{W}(\hat{\mu}_n, \hat{\nu}_n)| = O(n^{-1/d})$.
- Denote $\mathcal{G}_k = \{R \subset \mathbb{R}^d | dim(E) = k\}$ as the Grassmannian of k-dimensional subspace.

$$\begin{aligned} \mathcal{P}_k(\mu,\nu) &= \sup_{E\in\mathcal{G}_k} \mathcal{W}(\mathrm{Proj}_E \mu, \mathrm{Proj}_E \nu) \\ &= \max_{E\in\mathcal{G}_k} \min_{\pi\in\Pi(\mu,\nu)} \left(\int \|\mathrm{Proj}_E(x-y)\|^2 d\pi(x,y) \right)^{1/2}. \end{aligned} \tag{2}$$

- Sample complexity of PRW: $\mathbb{E}|\mathcal{P}(\mu,\nu) - \mathcal{P}(\hat{\mu}_n, \hat{\nu}_n)| = O(n^{-1/k})$.
- Discrete distributions:

$$\mathcal{P}_k^2(\mu_n,\nu_n) := \max_{U\in\mathrm{St}(d,k)} \min_{\pi\in\Pi(\mu_n,\nu_n)} f(\pi,U) := \sum_{i,j=1}^{n} \pi_{ij} \|U^\top x_i - U^\top y_j\|^2, \tag{3}$$

where $\mathrm{St}(d,k) := \{U \in \mathbb{R}^{d\times k} \mid U^\top U = I_{k\times k}\}$ is the Stiefel manifold.

Riemannian Block Coordinate Descent Algorithm

- The dual problem of entropy regularized PRW:

$$\begin{aligned}
&\max_{U\in\mathcal{M},\alpha,\beta}\min_{\pi}\sum_{ij}\pi_{ij}\|U^\top x_i - U^\top y_j\|^2 - \eta H(\pi) + \alpha^\top(\pi\mathbf{1} - r) + \beta^\top(\pi^\top\mathbf{1} - c) \\
&= \max_{U\in\mathcal{M},\alpha,\beta} -\eta\sum_{ij}\exp\left(-\frac{\alpha_i + \beta_j + \|U^\top x_i - U^\top y_j\|^2}{\eta}\right) - \sum_i r_i\alpha_i - \sum_j c_j\beta_j.
\end{aligned} \tag{4}$$

Riemannian Block Coordinate Descent Algorithm

- The dual problem of entropy regularized PRW:

$$\max_{U\in\mathcal{M},\alpha,\beta}\min_{\pi}\sum_{ij}\pi_{ij}\|U^\top x_i-U^\top y_j\|^2-\eta H(\pi)+\alpha^\top(\pi\mathbf{1}-r)+\beta^\top(\pi^\top\mathbf{1}-c)$$
$$=\max_{U\in\mathcal{M},\alpha,\beta}-\eta\sum_{ij}\exp\left(-\frac{\alpha_i+\beta_j+\|U^\top x_i-U^\top y_j\|^2}{\eta}\right)-\sum_i r_i\alpha_i-\sum_j c_j\beta_j. \tag{4}$$

- Denote the above objective as $-g(\alpha,\beta,U)$, the RBCD algorithm:

$$\alpha^{t+1}:=\operatorname*{argmin}_{\alpha}\ g(\alpha,\beta^t,U^t) \tag{5a}$$
$$\beta^{t+1}:=\operatorname*{argmin}_{\beta}\ g(\alpha^{t+1},\beta,U^t) \tag{5b}$$
$$U^{t+1}:=\mathrm{Retr}_{U^t}(-\tau\,\mathrm{grad}_U\, g(\alpha^{t+1},\beta^{t+1},U^t)), \tag{5c}$$

Riemannian Block Coordinate Descent Algorithm

- The dual problem of entropy regularized PRW:

$$\max_{U\in\mathcal{M},\alpha,\beta}\min_{\pi}\sum_{ij}\pi_{ij}\|U^\top x_i - U^\top y_j\|^2 - \eta H(\pi) + \alpha^\top(\pi\mathbf{1}-r) + \beta^\top(\pi^\top\mathbf{1}-c)$$
$$= \max_{U\in\mathcal{M},\alpha,\beta} -\eta\sum_{ij}\exp\left(-\frac{\alpha_i+\beta_j+\|U^\top x_i - U^\top y_j\|^2}{\eta}\right) - \sum_i r_i\alpha_i - \sum_j c_j\beta_j. \tag{4}$$

- Denote the above objective as $-g(\alpha,\beta,U)$, the RBCD algorithm:

$$\alpha^{t+1} := \operatorname*{argmin}_{\alpha}\ g(\alpha,\beta^t,U^t) \tag{5a}$$
$$\beta^{t+1} := \operatorname*{argmin}_{\beta}\ g(\alpha^{t+1},\beta,U^t) \tag{5b}$$
$$U^{t+1} := \mathrm{Retr}_{U^t}(-\tau \mathrm{grad}\,_U g(\alpha^{t+1},\beta^{t+1},U^t)), \tag{5c}$$

- Convergence results:
  - Previous best complexity: $O\left((n^2d + n^2\epsilon^{-8})\epsilon^{-4}\right)$
  - RBCD complexity: $O\left(n^2d\epsilon^{-3}\right)$.

Numerical Results

- Fragmented Hypercube: consider a hypercube $\mu = \mathcal{U}([-1,1]^d)$ and a pushforward $\nu = T_{\#}\mu$ defined under the map $T(x) = x + 2\text{sign}(x) \odot (\sum_{k=1}^{k^*} e_k)$ with $\mathcal{W}(\mu,\nu)^2 = 4k^*$.
- Computational time:

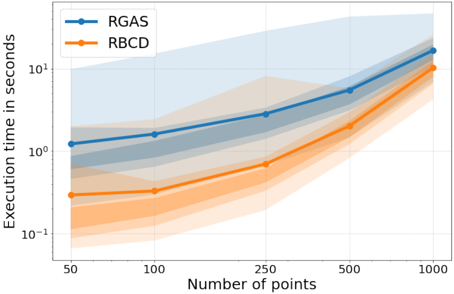


Figure: CPU time for calculating PRW. $d = 50$