Provably convergent algorithm for free-support Wasserstein barycenter of continuous non-parametric measures

Qikun Xiang

Nanyang Technological University, Singapore

July 2, 2024

Joint work with:

Zeyi Chen (Nanyang Technological University & INSEAD)
Ariel Neufeld (Nanyang Technological University)

Outline

Optimal transport

- **Motivation:** given two probability measures $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$, determine the most economical way of transporting the mass from μ to ν under the cost $\mathbb{R}^d \times \mathbb{R}^d \ni (\boldsymbol{x}, \boldsymbol{y}) \mapsto \|\boldsymbol{x} - \boldsymbol{y}\|^2 \in \mathbb{R}$.

- **Monge's problem:**

$$\inf_{T:\mathbb{R}^d \to \mathbb{R}^d,\, T\sharp\mu=\nu} \left\{ \int_{\mathbb{R}^d} \left\|\boldsymbol{x} - T(\boldsymbol{x})\right\|^2 \mu(\mathrm{d}\boldsymbol{x}) \right\}. \tag{MP}$$

- **Kantorovich's (relaxed) problem:**

$$\inf_{\pi \in \Pi(\mu,\nu)} \left\{ \int_{\mathbb{R}^d \times \mathbb{R}^d} \|\boldsymbol{x} - \boldsymbol{y}\|^2 \, \pi(\mathrm{d}\boldsymbol{x}, \mathrm{d}\boldsymbol{y}) \right\}, \tag{KP}$$

  where $\Pi(\mu, \nu) := \{\pi \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d) : \pi$ has marginals μ and $\nu\}$ denotes the set of couplings.

- While (MP) can be infeasible as it does not allow mass to be split, (KP) is always feasible and an optimal coupling is always attained.

Properties of optimal transport

- For continuous $\mu \in \mathcal{P}_{2,\mathrm{ac}}(\mathbb{R}^d)$, (KP) and (MP) are related via **Brenier's theorem**.

Theorem (Brenier [1991])

Let $\mu \in \mathcal{P}_{2,\mathrm{ac}}(\mathbb{R}^d)$. Then, there exists a unique optimal coupling $\pi^\star$ for (KP). Moreover, there exists a convex lower semi-continuous $\varphi^\mu_\nu : \mathbb{R}^d \to \mathbb{R} \cup \{\infty\}$ such that $T^\mu_\nu := \nabla \varphi^\mu_\nu$ solves (MP) and $\pi^\star = [I_d, T^\mu_\nu]\sharp\mu$.

- We call φ^μ_ν the optimal Brenier potential and call T^μ_ν the optimal transport (OT) map.
- **Caffarelli's regularity theory** provides sufficient conditions for the regularity of φ^μ_ν.

Theorem (Caffarelli [1990, 1991, 1992, 1996])

Let $\mu, \nu \in \mathcal{P}_{2,\mathrm{ac}}(\mathbb{R}^d)$ be concentrated on bounded open sets $\mathcal{X}_\mu, \mathcal{X}_\nu$ that are uniformly convex and have $\mathcal{C}^2$ boundaries. If the density of μ (resp. ν) are positive on $\mathcal{X}_\mu$ (resp. $\mathcal{X}_\nu$) and belong to $\mathcal{C}^{k,\alpha}(\mathcal{X}_\mu)$ (resp. $\mathcal{C}^{k,\alpha}(\mathcal{X}_\nu)$), i.e., k times differentiable with α-Hölder partial derivatives, then it holds that $\varphi^\mu_\nu \in \mathcal{C}^{k+2,\alpha}(\mathcal{X}_\mu)$.

Wasserstein distance and Wasserstein barycenter

- The 2-Wasserstein distance between $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ is defined as

$$\mathcal{W}_2(\mu,\nu) := \left\{ \inf_{\pi \in \Pi(\mu,\nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \|\boldsymbol{x} - \boldsymbol{y}\|^2 \, \pi(\mathrm{d}\boldsymbol{x}, \mathrm{d}\boldsymbol{y}) \right\}^{\frac{1}{2}}.$$

- $\mathcal{W}_2(\cdot,\cdot)$ metrizes weak convergence in $\mathcal{P}_2(\mathbb{R}^d)$.

- Given $\nu_1, \ldots, \nu_K \in \mathcal{P}_2(\mathbb{R}^d)$, let $V(\mu) := \frac{1}{K} \sum_{k=1}^{K} \mathcal{W}_2(\mu, \nu_k)^2$. Then, $\bar{\mu} \in \mathcal{P}_2(\mathbb{R}^d)$ is called a Wasserstein barycenter of $\nu_1, \ldots, \nu_K$ [Agueh and Carlier 2011] if

$$\bar{\mu} \in \underset{\mu \in \mathcal{P}_2(\mathbb{R}^d)}{\arg\min}\, V(\mu).$$

Literature review

- **Applications of Wasserstein barycenter:** Bayesian statistics (e.g., Srivastava, Li, Dunson [2018]), unsupervised clustering (e.g., Ye et al. [2017]), pattern recognition (e.g., Tabak, Trigila, Zhao [2022]), etc.
- **Existing numerical methods for Wasserstein barycenter:**
  - **Parametric Wasserstein barycenter:** applicable when $\nu_1, \ldots, \nu_K$ are certain parametric measures (e.g., Gaussian); see, e.g., Álvarez-Esteban et al. [2016], Chewi et al. [2020].
  - **Discrete Wasserstein barycenter:** applicable when $\nu_1, \ldots, \nu_K$ are discrete measures; see, e.g., Cuturi and Doucet [2014], Benamou et al. [2015], and Anderes, Borgwardt, Miller [2016].
  - **Fixed-support methods:** restrict the support of the Wasserstein barycenter to a prespecified finite set and optimize over a finite number of probabilities; see, e.g., Staib et al. [2017], Claici, Chien, Solomon [2018], Dvurechenskii et al. [2018].
  - **Neural network-based methods:** parametrize φ_ν^μ and/or $\bar{\mu}$ with neural networks and optimize over neural network parameters; see, e.g., Fan, Taghvaei, Chen [2020], Li et al. [2020], Korotin et al. [2021].
- **Our contribution:** we propose a **provably convergent** algorithm for approximating the **free-support** Wasserstein barycenter of **non-parametric continuous** measures.

Outline

Preliminaries

- Álvarez-Esteban, del Barrio, Cuesta-Albertos, and Matrán [2016] have shown that the Wasserstein barycenter $\bar{\mu}$ is a fixed-point of $G(\mu) := \left[\frac{1}{K}\sum_{k=1}^{K} T_{\nu_k}^{\mu}\right]\sharp\mu$, i.e., $G(\bar{\mu}) = \bar{\mu}$.

Algorithm: Deterministic fixed-point scheme

Input: K input measures $\nu_1, \ldots, \nu_K \in \mathcal{P}_{2,\mathrm{ac}}(\mathbb{R}^d)$, initial measure $\mu_0 \in \mathcal{P}_{2,\mathrm{ac}}(\mathbb{R}^d)$.
Output: $(\mu_t)_{t\in\mathbb{N}_0}$.

```
for t = 0, 1, 2, ... do
    for k = 1, ..., K do
        Get the OT map T_{ν_k}^{μ_t}.
    μ_{t+1} ← [ (1/K) Σ_{k=1}^K T_{ν_k}^{μ_t} ] ♯ μ_t.
```

Theorem (Corollary 3.5 & Theorem 3.6 of Álvarez-Esteban et al. [2016])

$(\mu_t)_{t\geq 0}$ is tight and every $\mathcal{W}_2$-accumulation point of $(\mu_t)_{t\geq 0}$ is a fixed-point of G.
In particular, if G has a unique fixed-point, then $\mu_t \xrightarrow{\mathcal{W}_2} \bar{\mu}$ as $t \to \infty$.

- However, for general non-parametric ν_k, the computation of $T_{\nu_k}^{\mu_t}$ is intractable, and thus the deterministic fixed-point scheme **does not lead to a computationally tractable algorithm**.

Conceptual approach

- **Conceptual approach:** we estimate $T_{\nu_k}^{\mu_t}$ using samples from μ_t and ν_k and extend this scheme to a **stochastic fixed-point algorithm**.

Algorithm: Stochastic fixed-point algorithm **(conceptual)**

Input: K input measures $\nu_1, \ldots, \nu_K \in \mathcal{P}_{2,\mathrm{ac}}(\mathbb{R}^d)$, initial measure $\mu_0 \in \mathcal{P}_{2,\mathrm{ac}}(\mathbb{R}^d)$.
Output: $(\widehat{\mu}_t)_{t \in \mathbb{N}_0}$.

1 Initialize $\widehat{\mu}_0 \leftarrow \mu_0$.
2 **for** $t = 0, 1, 2, \ldots$ **do**
3   **for** $k = 1, \ldots, K$ **do**
4   Randomly generate $N_{t,k}$ i.i.d. samples $\{\boldsymbol{X}_{t+1,k,i}\}_{i=1:N_{t,k}}$ from $\widehat{\mu}_t$.
5   Randomly generate $N_{t,k}$ i.i.d. samples $\{\boldsymbol{Y}_{t+1,k,i}\}_{i=1:N_{t,k}}$ from ν_k.
6   Approximate $T_{\nu_k}^{\widehat{\mu}_t}$ with an estimator $\widehat{T}_{t+1,k} \approx T_{\nu_k}^{\widehat{\mu}_t}$ using the samples $\{\boldsymbol{X}_{t+1,k,i}\}_{i=1:N_{t,k}}$ and $\{\boldsymbol{Y}_{t+1,k,i}\}_{i=1:N_{t,k}}$.
7   Choose $\widehat{\mu}_{t+1} \in \mathcal{P}_{2,\mathrm{ac}}(\mathbb{R}^d)$ such that $\widehat{\mu}_{t+1} \approx \left[\frac{1}{K}\sum_{k=1}^{K} \widehat{T}_{t+1,k}\right]\sharp\widehat{\mu}_t$.

Settings

- Admissible support sets:

$$\mathcal{S} := \left\{\mathrm{cl}(\mathcal{Y}) : \mathcal{Y} \subset \mathbb{R}^d \text{ is open, bounded, uniformly convex, and has a } \mathcal{C}^2\text{-boundary}\right\}.$$

- Admissible fully supported probability measures:

$$\mathcal{M}_{\text{full}} := \left\{\rho \in \mathcal{P}_{2,\mathrm{ac}}(\mathbb{R}^d) : \mathrm{supp}(\rho) = \mathbb{R}^d, \text{ the density of } \rho \text{ is locally } \alpha\text{-Hölder and positive}\right\}.$$

- Admissible compactly supported probability measures:

$$\mathcal{M} := \left\{\mu \in \mathcal{P}_{2,\mathrm{ac}}(\mathbb{R}^d) : \mathrm{supp}(\mu) \in \mathcal{S}, \text{ the density of } \mu \text{ is } \alpha\text{-Hölder and positive on } \mathrm{supp}(\mu)\right\}.$$

- In particular, $\rho|_{\mathcal{X}}(\cdot) := \frac{\rho(\,\cdot \cap \mathcal{X})}{\rho(\mathcal{X})} \in \mathcal{M}$ for any $\rho \in \mathcal{M}_{\text{full}}$ and any $\mathcal{X} \in \mathcal{S}$.
- For any $\mu, \nu \in \mathcal{M}$, the regularity properties of φ_{ν}^{μ} is implied by Caffarelli's regularity theory.

Settings

- Family $(\mathcal{X}_r)_{r\in\mathbb{N}}$ of increasing admissible support sets used for adjusting the support of $(\widehat{\mu}_t)_{t\geq 0}$: $\mathcal{X}_r \in \mathcal{S}$, $\mathcal{X}_{r+1} \supseteq \mathcal{X}_r\ \forall r \in \mathbb{N}$, $\bigcup_{r\in\mathbb{N}} \mathcal{X}_r = \mathbb{R}^d$; e.g., the family of scaled unit balls.

- Plug-in OT map estimator $\widehat{T}^{\mu,m}_{\nu,n}(\,\cdot\,;\theta)$ estimates T^{μ}_{ν} based on m samples from μ and n samples from ν with smoothing parameter θ, subject to:
  - **Shape condition:** $\widehat{T}^{\mu,m}_{\nu,n}(\,\cdot\,;\theta) = \nabla\widehat{\varphi}^{\mu,m}_{\nu,n}(\,\cdot\,;\theta)$, $\widehat{\varphi}^{\mu,m}_{\nu,n}(\,\cdot\,;\theta) \in \mathcal{C}^{2,\alpha}(\mathbb{R}^d)$ is $\overline{\lambda}$-smooth, $\underline{\lambda}$-strongly convex (preserves the regularity of $\widehat{\mu}_t$);
  - **Growth condition:** $\mathbb{E}\left[\left\|\widehat{T}^{\mu,m}_{\nu,n}(\boldsymbol{x};\theta)\right\|^2\right] \leq u_1(\nu) + u_2(\nu)\|\boldsymbol{x}\|^2\ \forall \boldsymbol{x} \in \mathbb{R}^d$ where $u_1(\nu)$, $u_2(\nu)$ depend on ν (controls the approximation error of $\widehat{\mu}_{t+1} \approx \left[\frac{1}{K}\sum_{k=1}^{K} \widehat{T}_{t+1,k}\right]\sharp\widehat{\mu}_t$);
  - **Consistency condition:** for all $\epsilon > 0$, there exist $\overline{\theta}(\mu,\nu,\epsilon)$ and $\overline{n}(\mu,\nu,\epsilon)$ depending on μ,ν,ϵ, such that $\mathbb{E}\left[\left\|\widehat{T}^{\mu,m}_{\nu,n}(\,\cdot\,;\theta) - T^{\mu}_{\nu}\right\|^2_{\mathcal{L}^2(\mu)}\right] \leq \epsilon\ \ \forall\theta \geq \overline{\theta}(\mu,\nu,\epsilon),\ \forall m \geq \overline{n}(\mu,\nu,\epsilon),\ \forall n \geq \overline{n}(\mu,\nu,\epsilon)$ (controls the estimation error of $\widehat{T}_{t+1,k} \approx T^{\widehat{\mu}_t}_{\nu_k}$).

Concrete fixed-point algorithm

Algorithm: Stochastic fixed-point algorithm **(concrete)**

Input: K input measures $\nu_1, \ldots, \nu_K \in \mathcal{M}$, $\rho_0 \in \mathcal{M}_{\mathrm{full}}$, family $(\mathcal{X}_r)_{r\in\mathbb{N}}$ of increasing sets, plug-in OT map estimator $\widehat{T}^{\mu,m}_{\nu,n}(\,\cdot\,;\theta)$.

Output: $(\widehat{\mu}_t)_{t\in\mathbb{N}_0}$.

Initialize $\widehat{\rho}_0 \leftarrow \rho_0$.

for $t = 0, 1, 2, \ldots$ **do**

- Choose $R_t \in \mathbb{N}$.
- $\widehat{\mu}_t \leftarrow \widehat{\rho}_t|_{\mathcal{X}_{R_t}}$.
- **for** $k = 1, \ldots, K$ **do**
  - Choose $\Theta_{t,k} \in \mathbb{N}$ and $N_{t,k} \in \mathbb{N}$.
  - Randomly generate $N_{t,k}$ i.i.d. samples $\{\boldsymbol{X}_{t+1,k,i}\}_{i=1:N_{t,k}}$ from $\widehat{\mu}_t$.
  - Randomly generate $N_{t,k}$ i.i.d. samples $\{\boldsymbol{Y}_{t+1,k,i}\}_{i=1:N_{t,k}}$ from ν_k.
  - $\widehat{T}_{t+1,k} \leftarrow \widehat{T}^{\mu,m}_{\nu,n}(\,\cdot\,;\theta)\big|_{\mu=\widehat{\mu}_t,\,\nu=\nu_k,\,\theta=\Theta_{t,k},\,m=n=N_{t,k}}$.
- $\widehat{\rho}_{t+1} \leftarrow \left[\frac{1}{K}\sum_{k=1}^{K} \widehat{T}_{t+1,k}\right]\sharp\widehat{\rho}_t$.

Convergence analysis

Theorem (Convergence of the stochastic fixed-point algorithm)

*Let $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t\geq 0})$ be the filtered probability space induced by the stochastic fixed-point algorithm where $\mathcal{F}_t$ is the σ-algebra generated by all samples up to iteration t. Let $\beta > 0$. Then, one can **explicitly construct** $(\mathcal{F}_t)_{t\geq 0}$-adapted stochastic processes $(R_t)_{t\geq 0}$, $(\Theta_{t,k})_{t\geq 0,\, k=1:K}$, and $(N_{t,k})_{t\geq 0,\, k=1:K}$:*

$$\begin{aligned} R_0 &:= \bar{r}_2(\widehat{\rho}_0, \nu_1, \ldots, \nu_K, 1), \\ R_t &:= \max\left\{\bar{r}_1\left(\widehat{\rho}_t, t^{-(1+\beta)}\right), \bar{r}_2\left(\widehat{\rho}_t, \nu_1, \ldots, \nu_K, (t+1)^{-2(1+\beta)}\right)\right\} && \forall t \geq 1, \\ \Theta_{t,k} &:= \bar{\theta}\left(\widehat{\mu}_t, \nu_k, (t+1)^{-2(1+\beta)}\right) && \forall t \geq 0,\ \forall 1 \leq k \leq K, \\ N_{t,k} &:= \bar{n}\left(\widehat{\mu}_t, \nu_k, (t+1)^{-2(1+\beta)}\right) && \forall t \geq 0,\ \forall 1 \leq k \leq K, \end{aligned}$$

such that:

- *it holds $\mathbb{P}$-almost surely that $(\widehat{\mu}_t)_{t\geq 0}$ is tight and every $\mathcal{W}_2$-accumulation point of $(\widehat{\mu}_t)_{t\geq 0}$ is a fixed point of G;*
- *in particular, if G has a unique fixed-point, then $\widehat{\mu}_t \xrightarrow{\mathcal{W}_2} \bar{\mu}$ $\mathbb{P}$-almost surely as $t \to \infty$, where $\bar{\mu}$ is the **Wasserstein barycenter** of $\nu_1, \ldots, \nu_K$.*

Outline

Concrete plug-in OT map estimators

- **Recall:** a plug-in OT map estimator approximates T_ν^μ based on m samples $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_m$ from μ and n samples $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n$ from ν.
- We combine the results of Manole, Balakrishnan, Niles-Weed, and Wasserman [2021] and Taylor [2017] to get a shape-constrained convex least squares based OT map estimator.
- **Step 1:** solve the **discrete OT problem**: $\widehat{\pi}^\star \in \arg\min_{\widehat{\pi}\in\Pi(\breve{\mu}_m,\breve{\nu}_n)} \left\{ \int_{\mathbb{R}^d\times\mathbb{R}^d} \|\boldsymbol{x}-\boldsymbol{y}\|^2 \,\widehat{\pi}(\mathrm{d}\boldsymbol{x},\mathrm{d}\boldsymbol{y}) \right\}$, where $\breve{\mu}_m := \frac{1}{m}\sum_{i=1}^m \delta_{\boldsymbol{X}_i}$, $\breve{\nu}_n := \frac{1}{n}\sum_{j=1}^n \delta_{\boldsymbol{Y}_j}$ are empirical measures.
- **Step 2:** solve the **quadratically constrained quadratic programming (QCQP) problem**:

$$
\begin{aligned}
\underset{(\widetilde{\varphi}_i),\,(\widetilde{\boldsymbol{g}}_i)}{\text{minimize}} \quad & \sum_{i=1}^m\sum_{j=1}^n \widehat{\pi}^\star\big(\{(\boldsymbol{X}_i,\boldsymbol{Y}_j)\}\big)\|\widetilde{\boldsymbol{g}}_i + \underline{\lambda}\boldsymbol{X}_i - \boldsymbol{Y}_j\|^2 \\
\text{subject to} \quad & \widetilde{\varphi}_j \ge \widetilde{\varphi}_i + \langle \widetilde{\boldsymbol{g}}_i, \boldsymbol{X}_j - \boldsymbol{X}_i\rangle + \tfrac{1}{2(\overline{\lambda}-\underline{\lambda})}\|\widetilde{\boldsymbol{g}}_i - \widetilde{\boldsymbol{g}}_j\|^2 \qquad \forall 1\le i\le m,\ \forall 1\le j\le m, \\
& \|\widetilde{\boldsymbol{g}}_i + \underline{\lambda}\boldsymbol{X}_i\|^2 \le \overline{u}_0(\nu)^2 \qquad \forall 1\le i\le m,
\end{aligned}
$$

  where $\overline{u}_0(\nu)$ is the radius of $\mathrm{supp}(\nu)$. Let $(\widetilde{\varphi}_i^\star)_{i=1:m}$, $(\widetilde{\boldsymbol{g}}_i^\star)_{i=1:m}$ be the optimizer.

Concrete plug-in OT map estimators

- **Step 3:** for each $\boldsymbol{x} \in \mathbb{R}^d$, solve the **quadratic programming problem**:

$$\Delta := \left\{ \boldsymbol{w} = (w_1, \ldots, w_m)^\mathsf{T} : \sum_{i=1}^m w_i = 1,\ w_i \geq 0\ \forall 1 \leq i \leq m \right\} \subset \mathbb{R}^m,$$

$$\widetilde{\mathbf{G}}^\star := \begin{pmatrix} | & | & & | \\ \widetilde{g}_1^\star & \widetilde{g}_2^\star & \cdots & \widetilde{g}_m^\star \\ | & | & & | \end{pmatrix} \in \mathbb{R}^{d\times m},$$

$$v_i := \varphi_i^\star + \tfrac{1}{2(\overline{\lambda}-\underline{\lambda})}\|\boldsymbol{g}_i^\star\|^2 + \tfrac{\underline{\lambda}\overline{\lambda}}{2(\overline{\lambda}-\underline{\lambda})}\|\boldsymbol{X}_i\|^2 - \tfrac{\overline{\lambda}}{\overline{\lambda}-\underline{\lambda}}\langle \boldsymbol{g}_i^\star, \boldsymbol{X}_i\rangle \in \mathbb{R} \qquad \forall 1 \leq i \leq m,$$

$$\boldsymbol{v} := (v_1, \ldots, v_m)^\mathsf{T} \in \mathbb{R}^m,$$

$$\widehat{T}_{\mathrm{SCCLS}}(\boldsymbol{x}) := \underline{\lambda}\boldsymbol{x} + \widetilde{\mathbf{G}}^\star \widehat{\boldsymbol{w}}(\boldsymbol{x}),$$

$$\text{where } \widehat{\boldsymbol{w}}(\boldsymbol{x}) \in \operatorname*{arg\,max}_{\boldsymbol{w}\in\Delta} \left\{ \langle \widetilde{\mathbf{G}}^{\star\mathsf{T}}\boldsymbol{x} + \boldsymbol{v}, \boldsymbol{w}\rangle - \tfrac{1}{2(\overline{\lambda}-\underline{\lambda})}\|\widetilde{\mathbf{G}}^\star \boldsymbol{w}\|^2 \right\} \qquad \forall \boldsymbol{x} \in \mathbb{R}^d.$$

- $\widehat{T}_{\mathrm{SCCLS}}$ is uniquely defined and satisfies the **growth and consistency conditions**.

- However, $\widehat{T}_{\mathrm{SCCLS}}$ does not satisfy the **shape condition** as it lacks differentiability.

Smoothing

- We apply smoothing techniques to $\widehat{T}_{\mathrm{SCCLS}}$ to obtain two plug-in OT map estimators.
- **Kernel smoothing:**

$$\widehat{T}_{\mathrm{kern}}(\boldsymbol{x};\theta) := \int_{\mathbb{R}^d} \Psi_\theta(\boldsymbol{\eta}) \widehat{T}_{\mathrm{SCCLS}}(\boldsymbol{x}-\boldsymbol{\eta})\,\mathrm{d}\boldsymbol{\eta} \qquad \forall \boldsymbol{x} \in \mathbb{R}^d,$$

where $\Psi_\theta(\boldsymbol{\eta}) := (2\pi)^{-\frac{d}{2}}\theta^d \exp\big(-\theta^2\|\boldsymbol{\eta}\|^2/2\big)$ is the Gaussian kernel with covariance $\theta^{-2}\mathbf{I}$.

- **Softmax smoothing:**

$$\widehat{T}_{\mathrm{smax}}(\boldsymbol{x};\theta) := \underline{\lambda}\boldsymbol{x} + \widetilde{\mathbf{G}}^{\star}\widehat{\boldsymbol{w}}(\boldsymbol{x};\theta),$$

$$\text{where } \widehat{\boldsymbol{w}}(\boldsymbol{x};\theta) := \operatorname*{arg\,max}_{\boldsymbol{w}\in\Delta}\Big\{\langle\widetilde{\mathbf{G}}^{\star\mathsf{T}}\boldsymbol{x}+\boldsymbol{v},\boldsymbol{w}\rangle - \tfrac{1}{2(\overline{\lambda}-\underline{\lambda})}\|\widetilde{\mathbf{G}}^{\star}\boldsymbol{w}\|^2 - \tfrac{\eta(\boldsymbol{w})}{\theta}\Big\} \qquad \forall \boldsymbol{x} \in \mathbb{R}^d,$$

where $\eta(w_1,\ldots,w_m) := \log(m) + \sum_{i=1}^m w_i \log(w_i)$.

Properties of the smoothed plug-in OT map estimators

Theorem (Properties of smoothed plug-in OT map estimators)

Both $\widehat{T}_{\mathrm{kern}}(\,\cdot\,;\theta)$ *and* $\widehat{T}_{\mathrm{smax}}(\,\cdot\,;\theta)$ *satisfy the* ***shape, growth, and consistency conditions****. Specifically:*

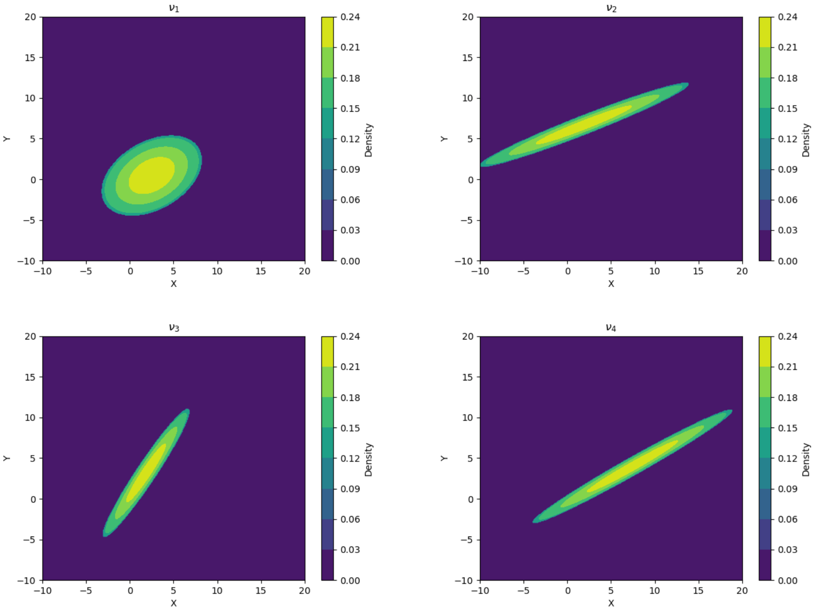
- $u_1(\nu) := 18\overline{u}_0(\nu)^2$, $u_2(\nu) := 2$, $\overline{n}(\mu,\nu,\epsilon) := \min\left\{n\in\mathbb{N} : C(\mu,\nu)\log(n)^2\kappa(n) \le \frac{\epsilon}{4}\right\}$ *for both* $\widehat{T}_{\mathrm{kern}}(\,\cdot\,;\theta)$ *and* $\widehat{T}_{\mathrm{smax}}(\,\cdot\,;\theta)$*;*
- $\overline{\theta}(\mu,\nu,\epsilon) := \left\lceil \left(\frac{4d}{\epsilon}\right)^{\frac{1}{2}}(\overline{\lambda}-\underline{\lambda})\right\rceil$ *for* $\widehat{T}_{\mathrm{kern}}(\,\cdot\,;\theta)$*;*
- $\overline{\theta}(\mu,\nu,\epsilon) := \left\lceil \frac{8}{\epsilon}\log\big(\overline{n}(\mu,\nu,\epsilon)\big)(\overline{\lambda}-\underline{\lambda})\right\rceil$ *for* $\widehat{T}_{\mathrm{smax}}(\,\cdot\,;\theta)$*,*

where $C(\mu,\nu)$ *is a constant in the* $\mathcal{L}^2(\mu)$ *estimation error bound of Manole et al. [2021], and*

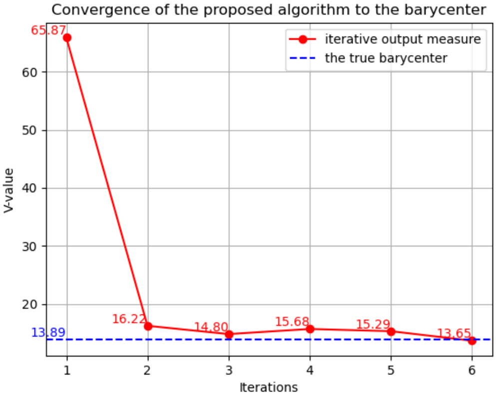
$$\kappa(n) := \begin{cases} n^{-\frac{1}{2}} & d\le 3, \\ n^{-\frac{1}{2}}\log(n) & d=4, \\ n^{-\frac{2}{d}} & d\ge 5 \end{cases} \qquad \forall n\in\mathbb{N}.$$

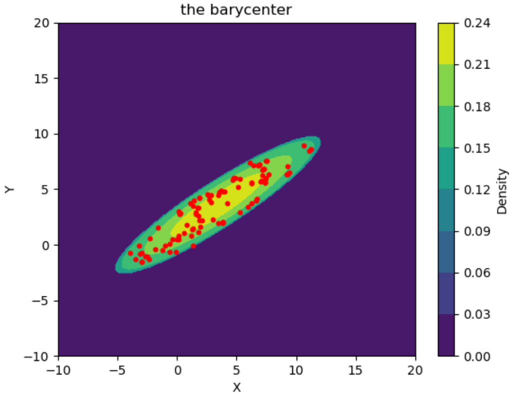
Outline

Preliminary numerical results: 2D Gaussian case



Preliminary numerical results: 2D Gaussian case





References

- M. Agueh and G. Carlier. Barycenters in the Wasserstein space. *SIAM Journal on Mathematical Analysis*, 43(2):904–924, 2011.
- P. C. Álvarez-Esteban, E. del Barrio, J. Cuesta-Albertos, and C. Matrán. A fixed-point approach to barycenters in Wasserstein space. *Journal of Mathematical Analysis and Applications*, 441(2):744–762, 2016.
- T. Manole, S. Balakrishnan, J. Niles-Weed, and L. Wasserman. Plugin estimation of smooth optimal transport maps. *Preprint, arXiv:2107.12364v2*, 2021.
- A. B. Taylor. Convex interpolation and performance estimation of first-order methods for convex optimization. PhD thesis, Catholic University of Louvain, Louvain-la-Neuve, Belgium, 2017.