

Supplementary Material

A novel notion of barycenter for probability distributions based on optimal weak mass transport

ELSA CAZELLES, FELIPE TOBAR, JOAQUÍN FONTBONA

A Additional mathematical background

A.1 p -Wasserstein distance

For $p = 2$ and $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ such that μ is absolutely continuous (*a.c.*) with respect to Lebesgue measure, the unique optimal plan is concentrated on the graph of a measurable map and Eq. (1) boils down to Monge's problem:

$$W_2(\mu, \nu) = \left(\min_{T \in \mathbb{T}(\mu, \nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \|x - T(x)\|^2 d\mu(x) \right)^{1/2}, \quad (14)$$

where $\mathbb{T}(\mu, \nu)$ is the set of measurable functions $T : \mathbb{R}^d \rightarrow \mathbb{R}^d$ such that $\nu = T\#\mu$. The *pushforward* operator $\#$ is defined such that for any measurable set $B \subset \mathbb{R}^d$, we have $\nu(B) = \mu(T^{-1}(B))$. In such a case, the optimal measurable map T in Eq. (14) is uniquely defined (see *e.g.* Th. 9.4 in [42]) and called *Monge map*.

A.2 Continuity of V

Theorem 6 ([9], Theorem 1.5). *Let $(\mu_n)_n \subset \mathcal{P}_2(\mathbb{R}^d)$ and $(\nu_n)_n \subset \mathcal{P}_1(\mathbb{R}^d)$. Then*

$$\begin{cases} \mu_n \rightarrow \mu & \text{in } W_2 \\ \nu_n \rightarrow \nu & \text{in } W_1 \end{cases} \implies \lim_n V(\mu_n | \nu_n) = V(\mu | \nu).$$

A.3 On the barycentric projection

For a given transport plan $\pi \in \Pi(\mu, \nu)$, with $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$, the associated barycentric projection is given by

$$S : x \mapsto \int_{\mathbb{R}^d} y d\pi_x(y).$$

First, for each $x \in \mathbb{R}^d$, $S(x)$ realises $\min_z \mathbb{E}_{Y \sim \pi_x} (\|z - Y\|^2)$. Second, this barycentric map S is actually optimal for the Monge's problem Eq. (14) between μ and $S\#\mu$, by Theorem 2.

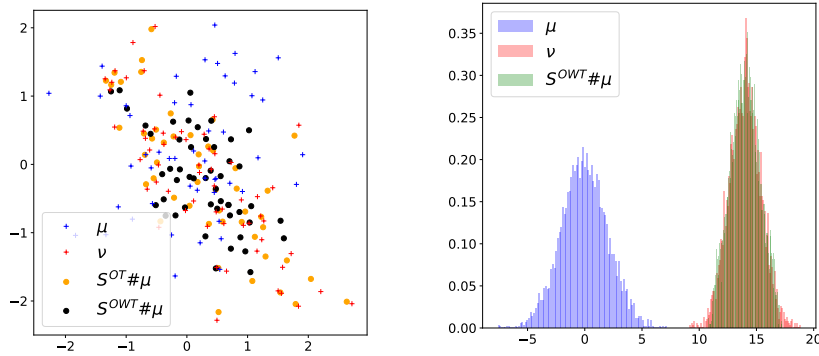


Figure 6: Example of pushforward measures constructed from barycentric projections for two measures μ and ν in two dimensions (left) and one dimension (right).

We next illustrate the differences between the optimal barycentric map and a barycentric map constructed from an OT plan in the classical Kantorovich formulation in Eq. (1). We sampled $r = 50$ observations X_i and $m = 60$ observations Y_i , each sets from a 2D Gaussian. We then defined the source and target distributions as $\mu = \frac{1}{r} \sum \delta_{X_i}$ and $\nu = \frac{1}{m} \sum \delta_{Y_i}$ respectively. Figure 6(left), shows these discrete distributions together with the pushforward measures $S^{OWT} \# \mu$ and $S^{OT} \# \mu$ constructed from the optimal weak plan π^{OWT} and an optimal plan π^{OT} respectively. The measure $S^{OT} \# \mu$ reasonably fits the target distribution ν , since when μ is a.c., $S^{OT} \# \mu = \nu$. In particular, if μ and ν had the same number of points, $S^{OT} \# \mu$ would have matched ν . Regarding the measure $S^{OWT} \# \mu$, recall that $V(\mu|\nu) = \inf_{\eta \leq_c \nu} W_2^2(\mu, \eta) = W_2^2(\mu, S^{OWT} \# \mu)$, and therefore $S^{OWT} \# \mu \leq_c \nu$. Lastly, we have $W_2^2(\mu, \nu) = 0.85$, and $V(\mu|\nu) = 0.52 \leq W_2^2(\mu, S^{OT} \# \mu) = 0.81$ as expected.

In Figure 6(right), we present an example in one dimension, where we sample 4000 observations from $\mathcal{N}(0, 2)$ (resp. $\mathcal{N}(14, 1.4)$) to construct the empirical source measure μ (resp. empirical target measure ν). The distributions μ and ν are presented in the form of histograms. The distribution resulting from the optimal weak transport map $S_\mu^\nu \# \mu$ is in convex order with ν .

B Proofs of Section 3

Proof of Proposition 1. Let $(\mu_m)_m \subset \mathcal{P}_2(\mathbb{R}^d)$ be a minimising sequence of $F(\mu) := \sum_{i=1}^n \lambda_i V(\mu|\nu_i)$ and let $M < \infty$ be such that $F(\mu_m) \leq M$ for all m . Then $(\mu_m)_m$ is tight. Indeed,

$$\begin{aligned} \int \|x\|^2 d\mu_m(x) &\leq 2 \sum_{i=1}^n \lambda_i \inf_{\pi \in \Pi(\mu_m, \nu_i)} \left[\int \|x - \int y d\pi_x(y)\|^2 d\mu_m(x) + \int \int \|y\|^2 d\pi_x(y) d\mu_m(x) \right] \\ &\leq 2M + 2 \int \int \|y\|^2 d\pi_x(y) d\mu_m(x) \leq 2M + 2 \sum_{i=1}^n \lambda_i \int \|y\|^2 d\nu_i(y), \end{aligned}$$

where the second inequality comes from Jensen's inequality. By Prokhorov's theorem, there exists a subsequence still denoted $(\mu_m)_m$ that weakly converges toward a probability measure μ^* . Recall that μ belongs to $\mathcal{P}_2(\mathbb{R}^d)$ since $\|\cdot\|^2$ is a l.s.c. function bounded from below and therefore $\int \|x\|^2 d\mu(x) \leq \liminf_m \int \|x\|^2 d\mu_m(x) < \infty$. By Theorem 6, we have that

$$F(\mu^*) = \sum_{i=1}^n \lambda_i \lim_m V(\mu_m|\nu_i) = \lim_m F(\mu_m) = \min_{\mu \in \mathcal{P}_2(\mathbb{R}^d)} F(\mu),$$

thus F admits at least a minimiser. \square

Proof of Lemma 1. By Strassen's theorem, we can build $X' \sim \mu'$ and $X \sim \mu$ in the same probability space, in such a way that $\mathbb{E}(X|X') = X'$. Denote by η^* the law η attaining $\inf_{\eta \leq_c \nu} W_2^2(\mu, \eta)$, and let $(X, Z) = (X, S_\mu^\nu(X))$ be the realisation of the optimal coupling for W_2 of μ and η^* , which can also be constructed in the same probability space due to its specific form. Then, by (the conditional version of) Jensen's inequality we have

$$V(\mu|\nu) = W_2^2(\mu, \eta^*) = \mathbb{E} [\mathbb{E} (\|X - S_\mu^\nu(X)\|^2 | X')] \geq \mathbb{E} \|X' - \mathbb{E}(S_\mu^\nu(X)|X')\|^2.$$

Recall now that $S_\mu^\nu(X) = \mathbb{E}(Y|X)$, where the conditional expectation is a measurable function only of X , constructed from the joint law $\pi^{\mu, \nu}$. Thus, for every nonnegative convex function ϕ , by applying twice Jensen's inequality we get

$$\mathbb{E}\phi(\mathbb{E}(S_\mu^\nu(X)|X')) \leq \mathbb{E}\phi(S_\mu^\nu(X)) = \mathbb{E}\phi(\mathbb{E}(Y|X)) \leq \mathbb{E}\phi(Y),$$

where $Y \sim \nu$. That is to say, the law η of the r.v. $\mathbb{E}(S_\mu^\nu(X)|X')$ satisfies $\eta \leq_c \nu$. It follows that

$$V(\mu|\nu) \geq W_2^2(\mu', \eta) \geq \inf_{\tilde{\eta} \leq_c \nu} W_2^2(\mu', \tilde{\eta}) = V(\mu'|\nu).$$

This immediately implies that μ' is a weak barycenter whenever μ is. In particular, if μ is a weak barycenter, then so is the Dirac mass supported on its mean. We then deduce that the set of minimisers of $\sum_i \lambda_i V(\mu|\nu_i)$ admits at least a Dirac mass δ_ω and

$$V(\delta_\omega|\nu_i) = \int \|x - \int y d\pi_x(y)\|^2 d\delta_\omega(x) = \|\omega - \mathbb{E}Y_i\|^2.$$

This implies that $\inf_\omega \sum \lambda_i V(\delta_\omega|\nu_i)$ is uniquely attained for $\bar{\omega} = \sum \lambda_i \mathbb{E}Y_i$. \square

Proof of Proposition 2. A probability measure μ is a weak barycenter if and only if $\sum_{i=1}^n \lambda_i V(\mu|\nu_i)$ is equal to the r.h.s. of (7). Let us suppose first that $\mathbb{E}Y_i = m$ for all $1 \leq i \leq n$, in which case the infimum in (5) is equal to 0. Then, μ is a weak barycenter if and only if $\mu \leq_c \nu_i$ for all $1 \leq i \leq n$ by definition of weak optimal transport (2), since in this case $V(\mu|\nu_i) = 0$. The general case can be reduced to the previous one, noting that

$$\begin{aligned} V(\mu|\nu_i) &= \inf_{\eta \leq_c \nu_i} W_2^2(\mu, \eta) \\ &= \inf_{\eta \leq_c \hat{\nu}_i} W_2^2(\hat{\mu}, \eta) + \|\mathbb{E}_\mu(X) - \mathbb{E}_{\nu_i}(Y_i)\|^2 \\ &= V(\hat{\mu}|\hat{\nu}_i) + \|\mathbb{E}_\mu(X) - \mathbb{E}_{\nu_i}(Y_i)\|^2, \end{aligned}$$

so that minimising $\sum_{i=1}^n \lambda_i V(\mu|\nu_i)$ over $\mu \in \mathcal{P}(\mathbb{R}^d)$ is equivalent to minimising $\sum_{i=1}^n \lambda_i V(\mu'|\hat{\nu}_i) + \sum_{i=1}^n \lambda_i \|\omega - \mathbb{E}_{\nu_i} Y_i\|^2$ over the two independent parameters (ω, μ') , with $\omega \in \mathbb{R}^d$ and $\mu' \in \mathcal{P}(\mathbb{R}^d)$ centered, taking μ as the law of $X = X' + \omega$ with $X' \sim \mu'$. \square

Proof of Lemma 2. Thanks to Prop. 3.3 in [5], we have that the (unique) Wasserstein barycenter verifies $\tilde{\mu} = \left(\sum_{i=1}^n \lambda_i T_{\tilde{\mu}}^{\nu_i} \right) \# \tilde{\mu}$ where $T_{\tilde{\mu}}^{\nu_i}$ is the optimal Monge map between $\tilde{\mu}$ and ν_i (see (14)). Moreover, from Proposition 4, a weak barycenter $\bar{\mu}$ also checks $\bar{\mu} = \left(\sum_{i=1}^n \lambda_i S_{\bar{\mu}}^{\nu_i} \right) \# \bar{\mu}$, where $S_{\bar{\mu}}^{\nu_i}$ is the optimal barycentric projection associated to $\bar{\pi}^i$ for $V(\bar{\mu}|\nu_i)$. Therefore, by Jensen's inequality applied twice,

$$\begin{aligned} W_2^2(\bar{\mu}, \tilde{\mu}) &\leq \iint \|x - y\|^2 d\bar{\mu}(x) d\tilde{\mu}(y) = \iint \left\| \sum_{i=1}^n \lambda_i S_{\bar{\mu}}^i(x) - \sum_{i=1}^n \lambda_i T_{\tilde{\mu}}^i(y) \right\|^2 d\bar{\mu}(x) d\tilde{\mu}(y) \\ &\leq \sum_{i=1}^n \lambda_i \iint \left\| T_{\bar{\mu}}^i(y) - z \right\|^2 d\bar{\pi}_x^i(z) d\bar{\mu}(x) d\tilde{\mu}(y) = \sum_{i=1}^n \lambda_i \iint \left\| T_{\bar{\mu}}^i(y) - z \right\|^2 d\tilde{\mu}(y) d\bar{\pi}^i(x, z) \\ &= \sum_{i=1}^n \lambda_i \iint \|y - z\|^2 d\nu_i(y) d\nu_i(z) = 2 \sum_{i=1}^n \lambda_i \iint (\mathbb{E}\|Y_i\|^2 - \|\mathbb{E}Y_i\|^2). \end{aligned}$$

\square

Proof of Theorem 3. Observe first that, by Theorem 2 and Strassen's theorem, solving the OWT problem (2) provides a unique (in law) coupling of three random variables (X, Y, Z) such that:

- i) (X, Y) has joint law $\pi^{\mu, \nu}$; in particular X and Y have the laws μ and ν respectively,
- ii) $Z = S_{\bar{\mu}}^{\nu}(X) = \mathbb{E}(Y|X)$ a.s., it has law η^* and it is optimally coupled to X in the sense of the optimal transport problem (1),
- iii) (Z, Y) is a martingale, that is $\mathbb{E}(Y|Z) = Z$ a.s..

Bringing all together we get the decomposition:

$$Y = Z + Y - Z = S_{\bar{\mu}}^{\nu}(X) + Y - \mathbb{E}(Y|X). \quad (15)$$

Now, by Lemma 1, if $X \sim \mu$ then the Dirac mass $\delta_{\mathbb{E}X}$ is a weak barycenter too. Thus we have on one hand:

$$\sum_{i=1}^n \lambda_i V(\mu|\nu_i) = \sum_{i=1}^n \lambda_i V(\delta_{\mathbb{E}X}|\nu_i). \quad (16)$$

Using Jensen's inequality, we see on the other hand that

$$\begin{aligned} V(\mu|\nu_i) &= \inf_{\eta \leq_c \nu_i} W_2^2(\mu, \eta) \\ &= \inf_{\eta \leq_c \nu_i} \mathbb{E}\|X - Z\|^2, \text{ with } (X, Z) \text{ an optimal coupling for } W_2^2 \text{ of } \mu \text{ and } \eta \\ &\geq \inf_{\eta \leq_c \nu_i} \|\mathbb{E}X - \mathbb{E}Z\|^2 = \inf_{\eta \leq_c \nu_i} W_2^2(\delta_{\mathbb{E}X}, \delta_{\mathbb{E}Z}) \\ &\geq \inf_{\tilde{\eta} \leq_c \nu_i} W_2^2(\delta_{\mathbb{E}X}, \tilde{\eta}) = V(\delta_{\mathbb{E}X}|\nu_i). \end{aligned}$$

Identity (16) thus implies $V(\mu|\nu_i) = V(\delta_{\mathbb{E}X}|\nu_i)$ for all i . Denoting by η_i the law η attaining $\inf_{\eta \leq \nu_i} W_2^2(\mu, \eta)$, and by (X, Z_i) the optimal coupling for W_2 of μ and η_i , we see that the latter can only occur if the equality case $\mathbb{E}\|X - Z_i\|^2 = \|\mathbb{E}X - \mathbb{E}Z_i\|^2$ in Jensen's inequality holds. This implies that $X - Z_i$ is deterministic for each i . Since $\mathbb{E}Z_i = \mathbb{E}Y_i$, we thus must have $X - Z_i = \mathbb{E}X - \mathbb{E}Y_i$. Taking $Z = Z_i$ and $Y = Y_i$ in Eq. (15), and noting that $S_\mu^{\nu_i}(X) = Z_i = X - (\mathbb{E}X - \mathbb{E}Y_i)$ the statement follows. \square

C Proofs of Section 3.3

Proof of Lemma 3. Let (μ_m) and (ν_m) , $m \in \mathbb{N}$, be two sequences in $\mathcal{P}_2(\mathbb{R}^d)$ respectively converging to μ and ν w.r.t. W_2 . Then, (μ_m) and (ν_m) are tight and thus the sequence $(\pi^m) := (\pi^{\mu_m, \nu_m})$ is tight too. Let π^{m_k} be a weakly convergent subsequence and π its limit. By Proposition 2.8 in [8] we have

$$\liminf_m V(\mu_m|\nu_m) = \liminf_m \int \|x - \int y d\pi_x^{m_k}\|^2 d\mu_{m_k}(x) \geq \int \|x - \int y d\pi_x\|^2 d\mu(x) \geq V(\mu|\nu).$$

However, we have $\lim_m V(\mu_m|\nu_m) = V(\mu|\nu)$ thanks to Theorem 6, hence $\int \|x - \int y d\pi_x\|^2 d\mu(x) = V(\mu|\nu)$. By uniqueness of the optimum for problem (2) we deduce that $\pi = \pi^{\mu, \nu}$. Since the same holds true for any weak limiting point of (π^m) , it follows that π^m weakly converges to $\pi^{\mu, \nu}$. Last, since $\int \|x\|^2 + \|y\|^2 d\pi^m(x, y) = \int \|x\|^2 d\mu_m(x) + \int \|y\|^2 d\nu_m(y)$, this quantity converges to $\int \|x\|^2 d\mu(x) + \int \|y\|^2 d\nu(y) = \int \|x\|^2 + \|y\|^2 d\pi(x, y)$, whence $W_2(\pi^m, \pi) \rightarrow 0$, and $(\mu, \nu) \in (\mathcal{P}_2(\mathbb{R}^d))^2 \mapsto \pi^{\mu, \nu} \in \mathcal{P}_2(\mathbb{R}^d \times \mathbb{R}^d)$ is continuous, as required (hence measurable).

We now establish the joint measurability of $(x, \nu) \in \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \mapsto S_\mu^\nu(x)$ for fixed μ . Notice this is a stronger statement than just measurability in the x variable, for each (μ, ν) . Write $\bar{B}(x, r)$ for the closed ball of radius $r > 0$ centered at x . One easily checks that the function

$$(x, \pi) \mapsto \Psi_r(x, \pi) := \frac{\int y \mathbf{1}_{\{(y, z): z \in \bar{B}(x, r)\}} d\pi(z, y)}{\int \mathbf{1}_{\{(y, z): z \in \bar{B}(x, r)\}} d\pi(z, y)}$$

is measurable w.r.t. the pair (x, π) , the two integrals being limits of integrals with respect to $d\pi(z, y)$, of some bounded continuous functions of (x, y, z) . Thus, $\limsup_{r \rightarrow 0} \Psi_r(x, \pi)$, $\liminf_{r \rightarrow 0} \Psi_r(x, \pi)$ and the function $\Phi(x, \pi) := \limsup_{r \rightarrow 0} \Psi_r(x, \pi) \mathbf{1}_{\{\limsup_{r \rightarrow 0} \Psi_r(x, \pi) = \liminf_{r \rightarrow 0} \Psi_r(x, \pi)\}}$ depend in a measurable way on (x, π) . It follows that $(x, \mu, \nu) \mapsto \Phi(x, \pi^{\mu, \nu})$ is measurable as the composition of two measurable functions. But notice that for each fixed $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ one has $\Psi_r(x, \pi^{\mu, \nu}) = \frac{\int_{\bar{B}(x, r)} [\int y d\pi_z(y)] d\mu(z)}{\mu(\bar{B}(x, r))}$ which, by the Lebesgue derivation theorem for Radon measures (*see e.g.* [16]), converges $d\mu(x)$ a.s. in x , to $\int y d\pi_x(y) = S_\mu^\nu(x)$. Thus, for each $\mu \in \mathcal{P}_2(\mathbb{R}^d)$,

$$S_\mu^\nu(x) = \Phi(x, \pi^{\mu, \nu}) \quad \text{for all } \nu \in \mathcal{P}_2(\mathbb{R}^d) \text{ and } d\mu(x) \text{ a.e. } x,$$

with $(x, \nu) \mapsto \Phi(x, \pi^{\mu, \nu})$ a measurable function. The conclusion follows. \square

Proof of Proposition 3. By Theorem 6.16 in [42], we know that there exists a sequence of discretely supported distributions $(\mathbb{Q}_n)_n \subset \mathcal{P}_2(\mathcal{P}_2(\mathbb{R}^d))$ of the form $\mathbb{Q}_n = \sum_{i=1}^n \lambda_i \delta_{\nu_i}$, with $(\lambda_i)_{1 \leq i \leq n}$ in the simplex, and such that $W_2^2(\mathbb{Q}, \mathbb{Q}_n) := \inf_{\pi \in \Pi(\mathbb{Q}, \mathbb{Q}_n)} \int W_2^2(\nu, \tilde{\nu}) d\pi(\nu, \tilde{\nu}) \rightarrow 0$. We set

$$L_n(\mu) := \int_{\mathcal{P}_2(\mathbb{R}^d)} V(\mu|\nu) d\mathbb{Q}_n(\nu) = \sum_{i=1}^n \lambda_i V(\mu|\nu_i).$$

We denote $\mu^n \in \mathcal{P}_2(\mathbb{R}^d)$ the minimiser of L_n . Let us prove that $(\mu^n)_n$ is tight. First, μ^n admits moments of order 2 thanks to Jensen's inequality:

$$\begin{aligned} \int \|x\|^2 d\mu^n(x) &\leq \sum_{i=1}^n \lambda_i \left[\int \|x - S_{\mu^n}^{\nu_i}(x)\|^2 d\mu^n(x) + \int \|S_{\mu^n}^{\nu_i}(x)\|^2 d\mu^n(x) \right] \\ &\leq \sum_{i=1}^n \lambda_i V(\mu^n | \nu_i) + \sum_{i=1}^n \lambda_i \int \|y\|^2 d\nu_i(y) \\ &\leq \sum_{i=1}^n \lambda_i V(\mu | \nu_i) + \sum_{i=1}^n \lambda_i \int \|y\|^2 d\nu_i(y) \quad \text{for some } \mu \in \mathcal{P}_2(\mathbb{R}^d) \text{ since } \mu^n \text{ minimises } L_n \\ &\leq 2 \int \|x\|^2 d\mu(x) + 3 \sum_{i=1}^n \lambda_i \int \|y\|^2 d\nu_i(y), \end{aligned}$$

where the last inequality comes from $V(\mu | \nu_i) = \int \|x - S_{\mu}^{\nu_i}(x)\|^2 d\mu(x) \leq 2 \int \|x\|^2 d\mu(x) + 2 \int \|S_{\mu}^{\nu_i}(x)\|^2 d\mu(x)$. Moreover, since $W_2^2(\mathbb{Q}, \mathbb{Q}_n) \rightarrow 0$, we have (Lemma 5.1.7 in [6]) that $\int \psi(\nu) d\mathbb{Q}_n(\nu) \rightarrow \int \psi(\nu) d\mathbb{Q}(\nu)$ for any function ψ such that $|\psi(\nu)| \leq a + bW_2^2(\nu, \nu_0)$, $a, b \geq 0$. In particular, choosing $\psi(\nu) = W_2^2(\nu, \delta_0) = \int \|y\|^2 d\nu(y)$, it implies that $\sum_{i=1}^n \lambda_i \int \|y\|^2 d\nu_i(y) \rightarrow \int \int \|y\|^2 d\nu(y) d\mathbb{Q}(\nu) < \infty$. Therefore $(\sum_{i=1}^n \lambda_i \int \|y\|^2 d\nu_i(y))_n$ is bounded and $(\mu^n)_n$ is tight. Thus by Prokhorov's theorem, there exists a subsequence, still denoted $(\mu^n)_n$, that converges towards $\bar{\mu}$.

Let us now prove that this particular $\bar{\mu}$ minimises the function $L : \mu \mapsto \int_{\mathcal{P}_2(\mathbb{R}^d)} V(\mu | \nu) d\mathbb{Q}(\nu)$. First, let $\eta \in \mathcal{P}_2(\mathbb{R}^d)$, still by Lemma 5.1.7 in [6] and since $V(\eta | \nu) \leq W_2^2(\eta, \nu)$, we get that $L(\eta) = \int V(\eta | \nu) d\mathbb{Q}(\nu) \geq \liminf_{n \rightarrow \infty} \int V(\eta | \nu) d\mathbb{Q}_n(\nu)$. Since for each n , the distribution μ^n minimises L_n , we have

$$\liminf_{n \rightarrow \infty} \int V(\eta | \nu) d\mathbb{Q}_n(\nu) \geq \liminf_{n \rightarrow \infty} \int V(\mu^n | \nu) d\mathbb{Q}_n(\nu). \quad (17)$$

Thanks to Fatou's Lemma for sequences of measures (\mathbb{Q}_n) (see [22]), we have that

$$\liminf_{n \rightarrow \infty} \int V(\mu^n | \nu) d\mathbb{Q}_n(\nu) \geq \int \liminf_{n \rightarrow \infty} V(\mu^n | \nu) d\mathbb{Q}(\nu) = \int V(\bar{\mu} | \nu) d\mathbb{Q}(\nu),$$

where the last equality comes from the lower semi-continuity of V (Theorem 2.9 in [8]). This proves that $\bar{\mu}$ minimises L . \square

D Proofs of Section 4

The proof of Theorem 4, on the continuity of $G : \mu \mapsto (\sum_{i=1}^n \lambda_i S_{\mu}^{\nu_i}) \# \mu$, leans on the two following technical lemmas.

Lemma 5. *Let $(\rho_m)_m$ be a given sequence and ν be a fixed law in $\mathcal{P}_2(\mathbb{R}^d)$. For each m , let $S_m := S_{\rho_m}^{\nu}$ denote the barycenter map associated with the optimal coupling $\pi^{\rho_m, \nu}$ for (2). Then, the sequence of laws $(S_m \# \rho_m)_m$ has uniformly integrable second moments.*

Proof of Lemma 5. Let (X_m, Y_m) be a pair of random variables (r.v.) with joint law $\pi^{\rho_m, \nu}$, defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Notice that $S_m \# \rho_m$ is the law of the r.v. $\mathbb{E}(Y_m | X_m)$. Then, for each $M, K \geq 0$ we have

$$\begin{aligned} \int_{\{\|x\|^2 \geq M\}} \|x\|^2 dS_m \# \rho_m(x) &= \mathbb{E}(\|\mathbb{E}(Y_m | X_m)\|^2 \mathbf{1}_{\{\|\mathbb{E}(Y_m | X_m)\|^2 \geq M\}}) \\ &\leq \mathbb{E}(\mathbb{E}(\|Y_m\|^2 | X_m) \mathbf{1}_{\{\|\mathbb{E}(Y_m | X_m)\|^2 \geq M\}}) \\ &= \mathbb{E}(\|Y_m\|^2 \mathbf{1}_{\{\|\mathbb{E}(Y_m | X_m)\|^2 \geq M, \|Y_m\|^2 \geq K\}}) \\ &\quad + \mathbb{E}(\|Y_m\|^2 \mathbf{1}_{\{\|\mathbb{E}(Y_m | X_m)\|^2 \geq M, \|Y_m\|^2 < K\}}) \\ &\leq \mathbb{E}(\|Y_m\|^2 \mathbf{1}_{\|Y_m\|^2 \geq K}) + \frac{K}{M} \mathbb{E}(\|\mathbb{E}(Y_m | X_m)\|^2), \end{aligned}$$

where we have used Jensen's inequality and the fact that $\mathbb{E}(Y_m|X_m)$ is measurable w.r.t. the σ -field generated by X_m . Applying Jensen's inequality to the last term again, and recalling that Y_m has law $\nu \in \mathcal{P}_2(\mathbb{R}^d)$, we deduce that

$$\sup_m \int_{\{\|x\|^2 \geq M\}} \|x\|^2 dS_m \# \rho_m(x) \leq \int_{\{\|x\|^2 \geq K\}} \|x\|^2 d\nu(x) + \frac{K}{M} \int \|x\|^2 d\nu(x), \quad (18)$$

which is smaller than a given $\varepsilon > 0$, by choosing $K > 0$ and then $M > 0$ large enough. \square

Lemma 6. *Let $(\rho_m)_n, \rho$ in $\mathcal{P}_2(\mathbb{R}^d)$ be such that $W_2(\rho_m, \rho) \rightarrow 0$. We have:*

- i) *For each $\nu \in \mathcal{P}_2(\mathbb{R}^d)$ the sequence of laws $((id, S_{\rho_m}^\nu) \# \rho_m)_m$ converges w.r.t. W_2 in $\mathcal{P}_2(\mathbb{R}^d \times \mathbb{R}^d)$ to $(id, S_\rho^\nu) \# \rho$.*
- ii) *There exists in some probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a sequence of r.v. $(X_m)_m$ of laws $(\rho_m)_m$ and a r.v. X of law ρ such that, for each $\nu \in \mathcal{P}_2(\mathbb{R}^d)$, the sequence $(X_m, S_{\rho_m}^\nu(X_m))_m$ (with laws $((id, S_{\rho_m}^\nu) \# \rho_m)_m$) converges in $\mathbb{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ to $(X, S_\rho^\nu(X))$ (with law $(id, S_\rho^\nu) \# \rho$).*

Proof of Lemma 6. For the entire proof, we fix a $\nu \in \mathcal{P}_2(\mathbb{R}^d)$ and we write $S_m := S_{\rho_m}^\nu$ and $S := S_\rho^\nu$ for simplicity.

i) By Theorem 2 and Part 1. of Theorem 1.5 in [9], $(S_m \# \rho_m)_m$ converges to $S \# \rho$ w.r.t. W_1 and, by Lemma 5, also with respect to W_2 . In particular, the sequence $((id, S_m) \# \rho_m)_m$ has tight marginals, and therefore it is tight too.

Let us identify its weak limiting points. For simplicity we rename $((id, S_m) \# \rho_m)_m$ a weakly convergent subsequence. By the previous discussion, its weak limit $d\hat{\rho}(x, z)$ clearly has first and second marginal laws equal to $d\rho(x)$ and $dS \# \rho(z)$ respectively. Moreover, $\int \|x\|^2 d\rho_m(x) + \int \|z\|^2 dS_m \# \rho_m(z) \rightarrow \int \|x\|^2 + \|z\|^2 d\hat{\rho}(x, z)$, hence $((id, S_m) \# \rho_m)_m$ converges to some $\hat{\pi}$ with respect to W_2 in $\mathcal{P}_2(\mathbb{R}^d \times \mathbb{R}^d)$.

Now, by the characterisation of optimisers in Theorem 2, we have $V(\rho_m|\nu) = W_2^2(\rho_m, S_m \# \rho_m) = \int \|x - S_m(x)\|^2 d\rho_m(x)$. Taking $m \rightarrow \infty$, and thanks to Theorem 6, we finally obtain

$$V(\rho|\nu) = W_2^2(\rho, S \# \rho) = \int \|x - z\|^2 d\hat{\pi}(x, z).$$

In particular, using again Theorem 2, we conclude that $d\hat{\pi}(x, z)$ must be of the form $(id, S) \# \rho$.

ii) By Skorohod's representation theorem, one can construct simultaneously in some probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a sequence of r.v. $(X_m)_m$ of laws $(\rho_m)_m$ and a r.v. X of law ρ such that $(X_m)_m$ converges \mathbb{P} -a.s. to X . Moreover, since the sequence $(\rho_m)_m$ converges w.r.t. W_2 in $\mathcal{P}_2(\mathbb{R}^d)$, it has uniformly integrable second order moments. It follows that the sequence of r.v. $(|X_m|^2)_m$ is uniformly integrable and, by the Vitali convergence theorem, that $(X_m)_m$ also converges to X in $\mathbb{L}^2(\Omega, \mathcal{F}, \mathbb{P})$.

Now, by Lemma 5, the sequence of r.v. $(|S_m(X_m)|^2)_n$ is uniformly integrable too. Thus, by the Vitali convergence theorem, the statement will follow by proving that $S_m(X_m)$ converges in \mathbb{P} -probability to $S(X)$.

For each $N \in \mathbb{N}$, let $y \mapsto (y)^N$ denote the truncation of a vector $y \in \mathbb{R}^d$ obtained by projecting it onto the centered ball of radius N , $(y)^N := (1 \wedge \frac{N}{|y|})y$, which is a 1-Lipschitz function bounded by N . By Theorem 2, the functions $S_m^N := (S_m)^N$ are then 1-Lipschitz and bounded uniformly in $m \in \mathbb{N}$. Therefore, by the Arzela-Ascoli theorem, their restrictions to each compact cylinder set R of \mathbb{R}^d defines a relatively compact set of functions, with respect to the uniform topology in $C(R, \mathbb{R}^d)$. It follows by a diagonal argument that some subsequence $(S_{m_k}^N)_k$ converges, uniformly on compact sets, to some continuous function \tilde{S} on \mathbb{R}^d . Since X_n converges a.s. to the finite value X , we deduce that \mathbb{P} -a.s. as $k \rightarrow \infty$,

$$(X_{m_k}, S_{m_k}^N(X_{m_k})) \rightarrow (X, \tilde{S}(X)).$$

Notice now that $(X_{m_k}, S_{m_k}^N(X_{m_k}))$ has the law $(id, (\cdot)^N \circ S_{m_k}) \# \rho_{m_k}$ for each k and thus, by part a) and continuity of the mapping $(x, y) \mapsto (x, (y)^N)$, the r.v. $(X, \tilde{S}(X))$ has the law $(id, (\cdot)^N \circ S) \# \rho$. Hence we deduce that

$$(X, \tilde{S}(X)) = (X, (S(X))^N)$$

\mathbb{P} - almost surely. The previous arguments can be applied not just to $(X_m)_m$ but to any subsequence of it. That is, we can similarly prove that any subsequence of $(X_m, (S_m(X_m))^N)_m$ has a subsequence that a.s. converges to $(X, (S(X))^N)$. This means that, for each $N \in \mathbb{N}$

$$(X_m, (S_m(X_m))^N) \rightarrow (X, (S(X))^N)$$

in \mathbb{P} - probability when $n \rightarrow \infty$. To conclude, by tightness we can find for each $\eta > 0$ some $N \in \mathbb{N}$ large enough so that $\mathbb{P}(|S(X)| \geq N) \leq \eta$ and $\mathbb{P}(|S_m(X_m)| \geq N) \leq \eta$ for all $m \in \mathbb{N}$, which yields for each $\varepsilon > 0$,

$$\mathbb{P}(|S_m(X_m) - S(X)| \geq \varepsilon) \leq 2\eta + \mathbb{P}(|(S_m(X_m))^N - (S(X))^N| \geq \varepsilon).$$

Thus $\limsup_m \mathbb{P}(|S_m(X_m) - S(X)| \geq \varepsilon) \leq 2\eta$ for arbitrary $\eta > 0$ or, equivalently, $\mathbb{P}(|S_m(X_m) - S(X)| \geq \varepsilon) \rightarrow 0$ as $m \rightarrow \infty$, which concludes the proof of b). \square

We can now proceed to the proof of continuity of $G : \mu \mapsto (\sum_{i=1}^n \lambda_i S_\mu^{\nu_i}) \# \mu$.

Proof of Theorem 4. Let $(\rho_m)_m, \rho$ in $\mathcal{P}_2(\mathbb{R}^d)$ such that $W_2(\rho_m, \rho) \rightarrow 0$. We need to prove that $W_2^2(G(\rho_m), G(\rho)) \rightarrow 0$. For each m , we write $S_m^i := S_{\rho_m}^{\nu_i}$ and $S^i := S_\rho^{\nu_i}$.

By Lemma 6.ii), there exists in some probability space a sequence $(X_m)_m$ of laws $(\rho_m)_m$ and a r.v. X of law ρ such that

$$(S_m^1(X_m), \dots, S_m^n(X_m)) \rightarrow (S^1(X), \dots, S^n(X)) \quad \text{in } \mathbb{L}^2(\mathbb{P}).$$

Therefore, $\sum_{i=1}^n \lambda_i S_m^i(X_m)$ converges to $\sum_{i=1}^n \lambda_i S^i(X)$ in $\mathbb{L}^2(\mathbb{P})$. Since $\sum_{i=1}^n \lambda_i S_m^i(X_m)$ has law $G(\rho_m)$ and $\sum_{i=1}^n \lambda_i S^i(X)$ has law $G(\rho)$, the proof is complete. \square

Proof of Proposition 4. As in [5], we easily see that

$$\sum_{i=1}^n \lambda_i \int \|x - S_\mu^{\nu_i}(x)\|^2 d\mu(x) = \sum_{i=1}^n \lambda_i \int \|\bar{S}(x) - S_\mu^{\nu_i}(x)\|^2 d\mu(x) + \int \|x - \bar{S}(x)\|^2 d\mu(x).$$

But $\int \|x - S_\mu^{\nu_i}(x)\|^2 d\mu(x) = W_2^2(\mu, S_\mu^{\nu_i} \# \mu)$ since from Thm 1.4 in [8] the barycentric map $S_\mu^{\nu_i}$ is an optimal map for the Monge problem between μ and $S_\mu^{\nu_i} \# \mu$. Moreover, by definition $G(\mu) = \bar{S} \# \mu$, therefore $\int \|x - \bar{S}(x)\|^2 d\mu(x) \geq W_2^2(\mu, G(\mu))$. Finally, since $S_\mu^{\nu_i} \# \mu \leq_c \nu_i$, we have that $\int \|\bar{S}(x) - S_\mu^{\nu_i}(x)\|^2 d\mu(x) \geq V(G(\mu)|\nu_i)$. This, recalling that $V(\mu|\nu_i) = W_2^2(\mu, S_\mu^{\nu_i} \# \mu)$, yields

$$\sum_{i=1}^n \lambda_i V(\mu|\nu_i) \geq \sum_{i=1}^n \lambda_i V(G(\mu)|\nu_i) + W_2^2(\mu, G(\mu)). \quad (19)$$

Therefore, if μ is a weak barycenter, we readily get that $\mu = G(\mu)$. \square

Proof of Proposition 5. As in the proof of Theorem 4, we denote S_k^i the optimal barycentric projection associated to $\pi^{k,i} \in \Pi(\mu_k, \nu_i)$. First, we easily have that $\mu_{k+1} \in \mathcal{P}_2(\mathbb{R}^d)$, indeed by Jensen's inequality

$$\int \|x\|^2 d\mu_{k+1}(x) = \iint \left\| \sum_{i=1}^n \lambda_i S_k^i(x) \right\|^2 d\mu_k(x) \leq \sum_{i=1}^n \lambda_i \int \|y\|^2 d\nu_i(y) < \infty.$$

Then $(\mu_k)_k$ is tight, with uniformly integrable 2-moments by Lemma 5. Therefore $(\mu_k)_k$ admits a convergent subsequence in W_2 . Let $\tilde{\mu}$ be a weak limit of a subsequence $(\mu_{k_j})_j$, then we have $W_2(\mu_{k_j}, \tilde{\mu}) \xrightarrow{j \rightarrow \infty} 0$. By continuity of G in Theorem 4, we get $W_2(\mu_{k_j+1}, G(\tilde{\mu})) \xrightarrow{j \rightarrow \infty} 0$.

Moreover, by Theorem 6 we have for $F(\mu) := \sum_{i=1}^n \lambda_i V(\mu|\nu_i)$ that $F(\mu_{k_j}) \rightarrow F(\tilde{\mu})$ and $F(\mu_{k_j+1}) \rightarrow F(G(\tilde{\mu}))$ as $j \rightarrow \infty$. Let us prove that these two limits coincide. From (19), we have

$$F(\mu_{k_j}) \geq \sum_{i=1}^n \lambda_i V(G(\mu_{k_j})|\nu_i) = \sum_{i=1}^n \lambda_i V(\mu_{k_j+1}|\nu_i) = F(\mu_{k_j+1}).$$

Iterating this inequality leads to $F(\mu_{k_j}) \geq F(\mu_{k_j+1}) \geq F(\mu_{k_j+2})$ which yields $F(\tilde{\mu}) = F(G(\tilde{\mu}))$ and then $\tilde{\mu} = G(\tilde{\mu})$, using inequality (19). Thus $(\mu_{k_j})_j$ converges w.r.t. W_2 to a probability distribution $\tilde{\mu}$ which is a fixed point of G . \square

Proof of Lemma 4. The proof is similar to that of [35, Lemma 3.8]. For the sake of clarity, we rewrite it in our setting. We assume that $x = \int S_{\bar{\mu}}^{\nu}(x) d\mathbb{Q}(\nu)$, $\bar{\mu}(x)$ -a.s. is not true, then

$$\begin{aligned} 0 &< \int \|x - \int S_{\bar{\mu}}^{\nu}(x) d\mathbb{Q}(\nu)\|^2 d\bar{\mu}(x) \\ &= \int \|x\|^2 d\bar{\mu}(x) - 2 \iint \langle x, S_{\bar{\mu}}^{\nu}(x) \rangle d\mathbb{Q}(\nu) d\bar{\mu}(x) + \int \left\| \int S_{\bar{\mu}}^{\nu}(x) d\mathbb{Q}(\nu) \right\|^2 d\bar{\mu}(x). \end{aligned}$$

Moreover, $S_{\bar{\mu}}^{\mu} \# \bar{\mu} \leq_c \mu$, therefore by Theorem 1.4 in [8], we get

$$\begin{aligned} \int V \left(\left[\int S_{\bar{\mu}}^{\nu} d\mathbb{Q}(\nu) \right] \# \bar{\mu} | \mu \right) d\mathbb{Q}(\mu) &\leq \int \left\| \int S_{\bar{\mu}}^{\nu} d\mathbb{Q}(\nu) - S_{\bar{\mu}}^{\mu} \right\|_{\mathbb{L}^2(\bar{\mu})}^2 d\mathbb{Q}(\mu) \\ &= \iint \|S_{\bar{\mu}}^{\nu}(x)\|^2 d\bar{\mu}(x) d\mathbb{Q}(\nu) - \int \left\| \int S_{\bar{\mu}}^{\nu} d\mathbb{Q}(\nu) \right\|^2 d\bar{\mu}(x). \end{aligned}$$

Finally, noticing that $\iint \|x - S_{\bar{\mu}}^{\nu}(x)\|^2 d\bar{\mu}(x) d\mathbb{Q}(\nu) = \int V(\bar{\mu} | \nu) d\mathbb{Q}(\nu)$, we hence get

$$\int V \left(\left[\int S_{\bar{\mu}}^{\nu} d\mathbb{Q}(\nu) \right] \# \bar{\mu} | \mu \right) d\mathbb{Q}(\mu) < \int V(\bar{\mu} | \nu) d\mathbb{Q}(\nu),$$

which is in contradiction with $\bar{\mu}$ weak barycenter of \mathbb{Q} . \square

In order to study the convergence of the iterative scheme in (10), we define the following objects:

$$L(\mu) := \frac{1}{2} \int V(\mu | \nu) d\mathbb{Q}(\nu) \quad (20)$$

$$H(\mu)(x) := - \int (S_{\bar{\mu}}^{\nu} - \text{id}) d\mathbb{Q}(\nu)(x) \quad x \in \mathbb{R}^d. \quad (21)$$

Moreover, we denote by $\{\mathcal{F}_k\}_k$ the filtration of the i.i.d. sample $\nu^k \sim \mathbb{Q}$, namely \mathcal{F}_{-1} is the trivial sigma-algebra and \mathcal{F}_{k+1} is the sigma-algebra generated by ν^0, \dots, ν^k and therefore μ_k in (10) is \mathcal{F}_k -measurable.

The next two Propositions are needed to prove Theorem 5.

Proposition 6. *The functions $\mu \in \mathcal{P}_2(\mathbb{R}^d) \mapsto \|H(\mu)\|_{\mathbb{L}^2(\mu)}^2$ and $\mu \in \mathcal{P}_2(\mathbb{R}^d) \mapsto L(\mu)$ are continuous w.r.t W_2 .*

Proof. Let us first assume that $(\rho_m)_m, \rho$ in $\mathcal{P}_2(\mathbb{R}^d)$ are such that $W_2(\rho_m, \rho) \rightarrow 0$. We want to prove that

$$\|H(\rho_m)\|_{\mathbb{L}^2(\rho_m)}^2 \rightarrow \|H(\rho)\|_{\mathbb{L}^2(\rho)}^2 \quad (22)$$

when $m \rightarrow \infty$. Consider the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and r.v.'s $(X_m)_m$ and X constructed in Lemma 6.ii), and recall that, for each $\nu \in \mathcal{P}_2(\mathbb{R}^d)$, the r.v.'s $(X_m, S_{\rho_m}^{\nu}(X_m))$ have law $(\text{id}, S_{\rho_m}^{\nu}) \# \rho_m$ for each m and converge in $\mathbb{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ to the r.v. $(X, S_{\rho}^{\nu}(X))$, which has the law $(\text{id}, S_{\rho}^{\nu}) \# \rho$. We next extend this construction in order to suitably randomise ν . More precisely, we enlarge the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ to the product space $(\bar{\Omega}, \bar{\mathcal{F}}, \bar{\mathbb{P}}) = (\Omega \times \mathcal{P}_2(\mathbb{R}^d), \mathcal{F} \otimes \mathcal{B}(\mathcal{P}_2(\mathbb{R}^d)), \mathbb{P} \otimes \mathbb{Q})$, that is, we add an independent random variable, called ν , taking values in $\mathcal{P}_2(\mathbb{R}^d)$ and which has distribution \mathbb{Q} .

Thanks to the measurability of the mappings $(x, \nu) \mapsto S_{\rho_m}^{\nu}$ and $(x, \nu) \mapsto S_{\rho}^{\nu}$ proven in Lemma 3, by replacing ν by ν in the previous objects we obtain random vectors $(X_m, S_{\rho_m}^{\nu}(X_m))$ and $(X, S_{\rho}^{\nu}(X))$ defined in $(\bar{\Omega}, \bar{\mathcal{F}}, \bar{\mathbb{P}})$ which have, conditionally on $\{\nu = \nu\}$, the laws $(\text{id}, S_{\rho_m}^{\nu}) \# \rho_m$ and $(\text{id}, S_{\rho}^{\nu}) \# \rho$ respectively. Moreover, ν is independent of the r.v. X, X_1, \dots, X_m under \mathbb{P} .

Now, by conditioning on $\{\nu = \nu\}$, using the convergence result in Lemma 6.ii) and the dominated convergence Theorem, we can easily check that $((X_m, S_{\rho_m}^{\nu}(X_m)))_m$ converges to $(X, S_{\rho}^{\nu}(X))$ in $\bar{\mathbb{P}}$ -probability. Furthermore, one can integrate w.r.t. \mathbb{Q} the bound (18) obtained for fixed ν in the proof of Lemma 5 and, denoting $\bar{\mathbb{E}}$ the expectation with respect to $\bar{\mathbb{P}}$, deduce that

$$\begin{aligned} \sup_m \bar{\mathbb{E}} \left(\|S_{\rho_m}^{\nu}(X_m)\|^2 \mathbf{1}_{\{\|S_{\rho_m}^{\nu}(X_m)\|^2 \geq M\}} \right) &\leq \int \int_{\{\|x\|^2 \geq K\}} \|x\|^2 d\nu(x) \mathbb{Q}(d\nu) \\ &\quad + \frac{K}{M} \int \int \|x\|^2 d\nu(x) \mathbb{Q}(d\nu), \end{aligned}$$

for each $M, K \geq 0$, where the r.h.s. is finite since $\mathbb{Q} \in \mathcal{P}_2(\mathcal{P}_2(\mathbb{R}^d))$. It follows that the sequence $((X_m, S_{\rho_m}^\nu(X_m)))_m$ has uniformly integrable second moments, and therefore converges also in $L^2(\bar{\Omega}, \bar{\mathcal{F}}, \bar{\mathbb{P}})$ to $(X, S_\rho^\nu(X))$, thanks to the Vitali convergence theorem. In particular, as m tends to infity, we have

$$\int V(\mu_m|\nu)d\mathbb{Q}(\nu) = \mathbb{E}\|X_m - S_{\rho_m}^\nu(X_m)\|^2 \rightarrow \mathbb{E}\|X - S_\rho^\nu(X)\|^2 = \int V(\mu|\nu)d\mathbb{Q}(\nu),$$

which proves the continuity of the function $\mu \in \mathcal{P}_2(\mathbb{R}^d) \mapsto L(\mu)$.

We observe now that $\bar{\mathbb{E}}(S_{\rho_m}^\nu(X_m)|X_m) = \int S_{\rho_m}^\nu(X_m)d\mathbb{Q}(\nu)$ and $\bar{\mathbb{E}}(S_\rho^\nu(X)|X) = \int S_\rho^\nu(X)d\mathbb{Q}(\nu)$, $\bar{\mathbb{P}}$ - a.s., Moreover, if \mathcal{F}_∞ denotes the σ -algebra generated by (X_1, X_2, \dots) , one has $\bar{\mathbb{E}}(S_{\rho_m}^\nu(X_m)|\mathcal{F}_\infty) = \bar{\mathbb{E}}(S_{\rho_m}^\nu(X_m)|X_m)$ and $\bar{\mathbb{E}}(S_\rho^\nu(X)|\mathcal{F}_\infty) = \bar{\mathbb{E}}(S_\rho^\nu(X)|X)$. Using the continuity in $L^2(\bar{\Omega}, \bar{\mathcal{F}}, \bar{\mathbb{P}})$ of the conditional expectation with respect to \mathcal{F}_∞ , we deduce that

$$X_m - \bar{\mathbb{E}}(S_{\rho_m}^\nu(X_m)|X_m) \rightarrow X - \bar{\mathbb{E}}(S_\rho^\nu(X)|X) \quad (23)$$

in $L^2(\bar{\Omega}, \bar{\mathcal{F}}, \bar{\mathbb{P}})$. We conclude that $\bar{\mathbb{E}}\|X_m - \bar{\mathbb{E}}(S_{\rho_m}^\nu(X_m)|X_m)\|^2 \rightarrow \bar{\mathbb{E}}\|X - \bar{\mathbb{E}}(S_\rho^\nu(X)|X)\|^2$ as $m \rightarrow \infty$, which is exactly the required convergence (22). \square

Proposition 7. For the sequence $(\mu_k)_k$ defined in (10), we have

$$\mathbb{E}(L(\mu_{k+1}) - L(\mu_k)|\mathcal{F}_k) \leq \gamma_k^2 L(\mu_k) - \gamma_k \|H(\mu_k)\|_{\mathbb{L}^2(\mu_k)}^2. \quad (24)$$

Proof. The arguments are similar to the ones used for the population Wasserstein barycenter iterative scheme in the proof of Proposition 3.2 in [10]. Let us set them for the present problem. Let $\nu \in \text{supp}(Q)$, then $([(1 - \gamma_k)\text{id} + \gamma_k S_{\mu_k}^{\nu^k}], S_{\mu_k}^\nu] \# \mu_k$ belongs to $\Pi(\mu_{k+1}, S_{\mu_k}^\nu \# \mu_k)$. Therefore we have

$$\begin{aligned} V(\mu_{k+1}|\nu) &\leq W_2^2(\mu_{k+1}, S_{\mu_k}^\nu \# \mu_k) \quad \text{since } S_{\mu_k}^\nu \# \mu_k \leq_c \nu \\ &\leq \int \|(1 - \gamma_k)x + \gamma_k S_{\mu_k}^{\nu^k}(x) - S_{\mu_k}^\nu(x)\|^2 d\mu_k(x) \\ &= \int \|x - S_{\mu_k}^\nu(x)\|^2 d\mu_k(x) - 2\gamma_k \int \langle x - S_{\mu_k}^\nu(x), x - S_{\mu_k}^{\nu^k}(x) \rangle d\mu_k(x) \\ &\quad + \gamma_k^2 \int \|x - S_{\mu_k}^{\nu^k}(x)\|^2 d\mu_k(x) \\ &= V(\mu_k|\nu) + \gamma_k^2 V(\mu_k|\nu^k) - 2\gamma_k \int \langle x - S_{\mu_k}^\nu(x), x - S_{\mu_k}^{\nu^k}(x) \rangle d\mu_k(x). \end{aligned}$$

Integrating with respect to ν , and divided by 2 we get

$$L(\mu_{k+1}) \leq L(\mu_k) + \frac{\gamma_k^2}{2} V(\mu_k|\nu^k) - \gamma_k \int \langle H(\mu_k)(x), x - S_{\mu_k}^{\nu^k}(x) \rangle d\mu_k(x).$$

We can then take the conditional expectation with respect to the filtration \mathcal{F}_k , knowing that μ_k is \mathcal{F}_k -measurable and that ν^k is independently sampled from \mathcal{F}_k , we have

$$\begin{aligned} \mathbb{E}(L(\mu_{k+1})|\mathcal{F}_k) &\leq L(\mu_k) + \frac{\gamma_k^2}{2} \int V(\mu_k|\nu)d\mathbb{Q}(\nu) - \gamma_k \int \langle H(\mu_k)(x), \int x - S_{\mu_k}^\nu(x)d\mathbb{Q}(\nu) \rangle d\mu_k(x) \\ &= L(\mu_k) + \gamma_k^2 L(\mu_k) - \gamma_k \int \langle H(\mu_k)(x), H(\mu_k)(x) \rangle d\mu_k(x) \\ &= (1 + \gamma_k^2)L(\mu_k) - \gamma_k \|H(\mu_k)\|_{\mu_k}^2. \end{aligned}$$

\square

Proof of Theorem 5. We will proceed in a similar way as in the proof of Theorem 1.4 in [10]. Let us first note that the set $K_{\mathbb{Q}}$ is compact in $\mathcal{P}_2(\mathbb{R}^d)$ w.r.t. W_2 (see [42]). Moreover, the sequence $(\mu_k)_k$ is a.s. included in this compact set, as can be seen by induction using the facts that the function $|\cdot|^{2+\epsilon}$

is convex and that $S_{\mu_k}^{\nu_k} \# \mu_k \leq_c \nu_k$, with $\nu^k \sim \mathbb{Q}$. Now let $\bar{\mu}$ be a weak population barycenter, i.e. $\bar{\mu}$ minimises L defined in (20), and write $\bar{L} := L(\bar{\mu})$. We introduce the sequences

$$h_k := L(\mu_k) - \bar{L} \geq 0 \quad \text{and} \quad \alpha_k := \prod_{i=1}^{k-1} \frac{1}{1 + \gamma_i^2}.$$

From condition (11), the sequence $(\alpha_k)_k$ converges to some $\alpha_\infty > 0$. By Proposition 7, we have

$$\begin{aligned} \mathbb{E}(h_{k+1} - (1 + \gamma_k^2)h_k | \mathcal{F}_k) &\leq \gamma_k^2 \bar{L} - \gamma_k \|H(\mu_k)\|_{\mathbb{L}_2(\mu_k)}^2 \leq \gamma_k^2 \bar{L} \\ \Rightarrow \mathbb{E}(\alpha_{k+1}h_{k+1} - \alpha_k h_k | \mathcal{F}_k) &\leq \alpha_{k+1} \gamma_k^2 \bar{L} \quad \text{upon multiplying by } \alpha_{k+1}. \end{aligned} \quad (25)$$

Defining now

$$\delta_k := \begin{cases} 1 & \text{if } \mathbb{E}(\alpha_{k+1}h_{k+1} - \alpha_k h_k | \mathcal{F}_k) > 0 \\ 0 & \text{otherwise,} \end{cases}$$

we deduce that

$$\begin{aligned} \sum_{k=1}^{\infty} \mathbb{E}(\delta_k (\alpha_{k+1}h_{k+1} - \alpha_k h_k)) &= \sum_{k=1}^{\infty} \mathbb{E}(\delta_k \mathbb{E}(\alpha_{k+1}h_{k+1} - \alpha_k h_k | \mathcal{F}_k)) \\ &\leq L(\bar{\mu}) \sum_{k=1}^{\infty} \alpha_{k+1} \gamma_k^2 \leq L(\bar{\mu}) \sum_{k=1}^{\infty} \gamma_k^2 < \infty. \end{aligned}$$

Since $h_k \alpha_k \geq 0$, by the quasi-martingale convergence theorem $(h_k \alpha_k)_k$ converges almost surely. Since $(\alpha_k)_k$ converges to $\alpha_\infty > 0$, then $(h_k)_k$ also converges almost surely to some $h_\infty \geq 0$. Taking expectations in Eq. (25) and summing we get

$$\mathbb{E}(\alpha_{k+1}h_{k+1}) \leq \mathbb{E}(\alpha_0 h_0) + \bar{L} \sum_{m=1}^k \alpha_{m+1} \gamma_m^2 \leq C.$$

Fatou's Lemma yields $\mathbb{E}(\alpha_\infty h_\infty) < \infty$, and since $\alpha_\infty < \infty$, we deduce that h_∞ is almost surely finite. Our goal now is to show that $h_\infty = 0$. From (25) we deduce as before that

$$\mathbb{E}(\alpha_{k+1}h_{k+1}) - \mathbb{E}(\alpha_0 h_0) \leq \bar{L} \sum_{m=1}^k \alpha_{m+1} \gamma_m^2 - \sum_{m=1}^k \alpha_{m+1} \gamma_m \mathbb{E} \left(\|H(\mu_m)\|_{\mathbb{L}_2(\mu_m)}^2 \right).$$

Taking liminf, using Fatou on the l.h.s. and monotone convergence on the r.h.s. we obtain that

$$-\infty < \mathbb{E}(\alpha_\infty h_\infty) - \mathbb{E}(\alpha_0 h_0) \leq C - \mathbb{E} \left(\sum_{m=1}^{\infty} \alpha_{m+1} \gamma_m \|H(\mu_m)\|_{\mathbb{L}_2(\mu_m)}^2 \right)$$

hence, in particular,

$$\sum_{k=1}^{\infty} \gamma_k \|H(\mu_k)\|_{\mathbb{L}_2(\mu_k)}^2 < +\infty \quad \text{a.s.} \quad (26)$$

Note that the conditions in Eqs. (26) and (11) imply that

$$\liminf_{t \rightarrow \infty} \|H(\mu_t)\|_{\mathbb{L}_2(\mu_t)}^2 = 0, \quad \text{a.s.} \quad (27)$$

Observe also that, by the compactness of $K_{\mathbb{Q}}$ and the continuity of L in Proposition 6, the set $\{\rho : L(\rho) \geq \bar{L} + \delta\} \cap K_{\mathbb{Q}}$ is W_2 -compact. Therefore, the function $\rho \mapsto \|H(\rho)\|_{\mathbb{L}_2(\rho)}^2$, also W_2 -continuous by Proposition 6, attains its minimum on that set. That minimum cannot be zero, as otherwise we would have obtained a fixed-point that is not a weak barycenter, contradicting our hypothesis. It follows that

$$\forall \delta > 0, \inf_{\{\rho : L(\rho) \geq \bar{L} + \delta\} \cap K_{\mathbb{Q}}} \|H(\rho)\|_{\mathbb{L}_2(\rho)}^2 > 0. \quad (28)$$

Since $\{\mu_k\}_k \subset K_{\mathbb{Q}}$ a.s., we deduce from the previous result the a.s. inclusions of events:

$$\begin{aligned} \{h_\infty \geq 2\delta\} &\subset \{\mu_t \in \{\rho : L(\rho) \geq L(\bar{\mu}) + \delta\} \cap K_{\mathbb{Q}} \forall t \text{ large enough}\} \\ &\subset \bigcup_{\ell \in \mathbb{N}} \left\{ \|H(\mu_t)\|_{\mathbb{L}_2(\mu_t)}^2 > 1/\ell : \forall t \text{ large enough} \right\} \subset \left\{ \liminf_{t \rightarrow \infty} \|H(\mu_t)\|_{\mathbb{L}_2(\mu_t)}^2 > 0 \right\}, \end{aligned}$$

where Eq. (28) was used to obtain the second inclusion. It follows using Eq. (27) that $\mathbb{P}(h_\infty \geq 2\delta) = 0$ for every $\varepsilon > 0$, hence $h_\infty = 0$ a.s. as claimed. In other words, $L(\mu_k) \rightarrow \bar{L}$ a.s. as $k \rightarrow \infty$.

We already established that $\{\mu_k\}_k \subset K_{\mathbb{Q}}$, hence the sequence is relatively compact. We finally conclude that the limit $\hat{\mu}$ of any convergent subsequence $\{\mu_{k_j}\}_{k_j}$ satisfies $L(\hat{\mu}) = \lim_j L(\mu_{k_j}) = \bar{L}$, whence, it is a weak barycenter.

Remark 2. Assumption (A) can be replaced by the following more general condition:

(A') \mathbb{Q} gives full measure to a W_2 -compact set $K_{\mathbb{Q}}$ which is “weakly geodesically closed”, in the sense that for any $\mu, \nu \in K_{\mathbb{Q}}$ and $t \in [0, 1]$, $((1-t)\text{id} + tS_\mu^\nu)\#\mu \in K_{\mathbb{Q}}$.

□

E Numerical results

E.1 Proximal algorithm for the computation of the OWT plan

This section is dedicated to the resolution of the OWT problem. Let $\mu = \sum_{i=1}^r a_i \delta_{x_i}$ and $\nu = \sum_{j=1}^m b_j \delta_{y_j}$, be two discrete measures, the OWT problem boils down to solving

$$\min_{\pi \in \mathbb{R}^{r \times m}} \underbrace{\sum_{i=1}^r a_i \|x_i - \left(\frac{\pi \mathbf{y}}{\mathbf{a}}\right)_i\|^2}_{f(\pi)} + \underbrace{1_{\Pi(\mu, \nu)}(\pi)}_{g(\pi)}, \quad (29)$$

where 1_C is the indicator function of the set C i.e.

$$1_C(\pi) = \begin{cases} \pi & \text{if } \pi \in C \\ \infty & \text{otherwise.} \end{cases}$$

The proximal algorithm to solve Eq. (29) then reads:

$$\pi^{\ell+1} = \text{prox}_{\theta_\ell g}(\pi^\ell - \theta_\ell \nabla f(\pi^\ell)). \quad (30)$$

As $\Pi(\mu, \nu)$ is a closed non-empty convex set, the proximal operator of g reduces to the Euclidean projection onto $\Pi(\mu, \nu)$:

$$\text{proj}_{\Pi(\mu, \nu)}(P) = \arg \min_{\pi \in \mathbb{R}^{r \times m}} \|P - \pi\|^2 = \arg \min_{\pi \in \mathbb{R}^{r \times m}} \langle \pi, -P \rangle + \frac{1}{2} \|\pi\|^2$$

where $\|\cdot\|$ is the Frobenius norm. This projection problem can be solved by Dykstra’s algorithm with alternate Bregman projections [20] or by stochastic dual approaches of OT regularised by an \mathbb{L}_2 norm [36]. This method is summarised in Algorithm 3. In particular, we used an accelerated version of Eq. (30) via FISTA [12] (with $\omega_\ell \in [0, 1]$ an extrapolation parameter and θ_ℓ the usual stepsize chosen by a line search) in order to compute the optimal plan π_μ^ν in the weak transport problem.

The optimal barycentric projection is then given by $S_\mu^\nu = \frac{\pi_\mu^\nu \mathbf{y}}{\mathbf{a}}$. We initialised the algorithm with a random matrix whose elements sum to 1. Observe that, from Algorithm 1, the K optimal barycentric projection computations can be parallelised for each step n .

Algorithm 3: Computation of the optimal weak plan

Output: π_μ^ν ;

Input: $\mu = \sum_{i=1}^r a_i \delta_{x_i}$ and $\nu = \sum_{j=1}^m b_j \delta_{y_j}$;

Initialise π_0 random matrix;

while not converge do

| $P_{\ell+1} := \pi_\ell + \omega_\ell(\pi_\ell - \pi_{\ell-1})$;

| $\pi_{\ell+1} := \text{proj}_{\Pi(\mu, \nu)}(P_{\ell+1} - \theta_\ell \nabla f(P_{\ell+1}))$;

end

With respect to the efficiency of this algorithm, Figure 7 shows a comparison of different settings for Eq. (30) in order to compute an optimal weak transport plan. For that purpose, we considered

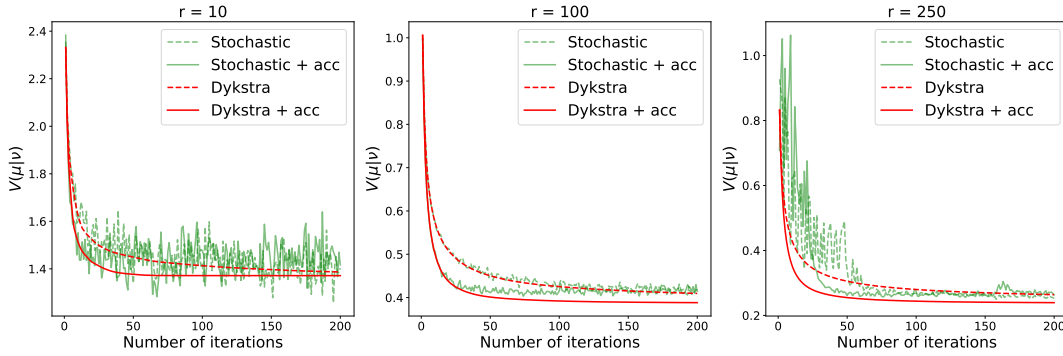


Figure 7: Convergence of the algorithm (3) in several settings for measures μ and ν supported on $r = m$ points.

two discrete distributions μ and ν each constructed from $r = m = 10, 100$ and 250 samples of two dimensional Gaussian measures. We illustrate the convergence for both the standard and accelerated versions of the proximal algorithm, as well as for the projection into $\Pi(\mu, \nu)$ via Dykstra’s algorithm or the stochastic dual approach. As expected, the accelerated version of Eq. (30) converges faster than the classical proximal algorithm, and the projection step is more stable with Dykstra’s algorithm. Moreover, the smaller the number of support points, the faster the convergence. We have also noted that the random initialisation does not affect the convergence towards the minimiser of Eq. (29).

E.2 Additional experiments

Gaussian distributions As in Section 5 of [5], we computed a weak barycenter between two 2D centered ellipses $E(\Sigma_i) = \{s \in \mathbb{R}^2 : x^t \Sigma_i^{-1} x = 1\}$ with covariances matrices

$$\Sigma_1 = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad \Sigma_2 = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix},$$

by considering 300 random observations for each ellipse. We then executed the iterations of Algorithm 1 until the difference of the objective function (i.e., the sum in Eq. (5)) between two successive iterations was smaller than $1e - 5$. This occurred at the 8th iteration, and the resulting weak barycenter was a circle within both ellipses. As we have access to the value of the weak barycenter problem (see Eq. (7)), we also compared the value of the objective function at the 8th iteration (that is $3.62e - 4$) to $\frac{1}{2} \sum_{i=1}^2 \|\mathbb{E}(Y_i)\|^2 - \|\frac{1}{2} \sum_{i=1}^2 \mathbb{E}(Y_i)\|^2$, with a plug-in estimator for $\mathbb{E}(Y_i)$. The approximated objective was equal to $3.21e - 4$, therefore, Algorithm 1 gave a satisfactory optimised weak barycenter.

Ellipse distributions ($r = 100$ & $K = 15$). We considered ellipse distributions with random center in $(-5, 5)$, random semi-major and semi-minor axes in $(6, 14)$. The results are presented in Fig. 8, where the same conclusions as in the Gaussian examples hold.

Pair-of-ellipses ($r \in (200, 300)$ & $K = 10$). In the same fashion, we considered distributions supported on two ellipses with random centers in $(-5, 5)$, random semi-major and semi-minor axes in $(1, 7)$ and $(7, 13)$ respectively. Fig. 9 shows the distributions (left) as well as the OT and OWT barycenters (right) computed from random samples of the distributions. Observe that, once again, the weak barycenter better preserved the structure of the distributions when computing Algorithm 2.

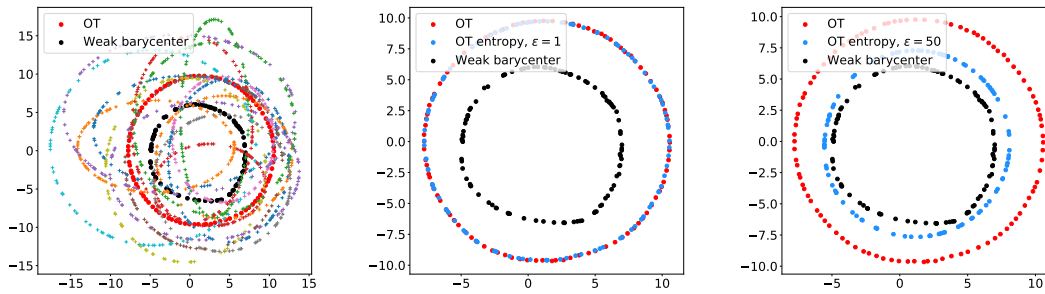


Figure 8: (left) Ellipse distributions and their OWT (black) and OT (red) barycenters computed with Algorithm 2. (center & right) Illustration of the weak (black), OT (red) and OT Sinkhorn (blue) barycenters for different values of $\varepsilon = 1, 50$.

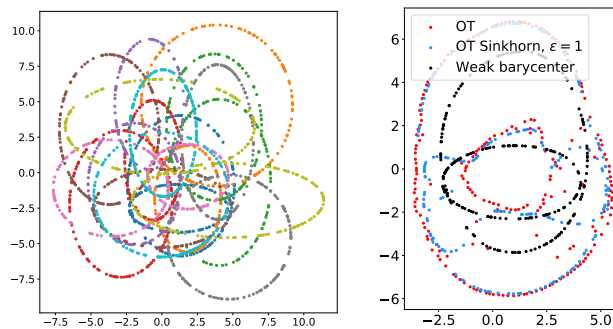


Figure 9: (left) Distributions supported on a pair-of-squares. (right) OWT (black), OT (red) and OT Sinkhorn for $\varepsilon = 1$ (blue) barycenters computed with Algorithm 2.

References

- [1] M. Agueh and G. Carlier. Barycenters in the Wasserstein space. *SIAM Journal on Mathematical Analysis*, 43(2):904–924, 2011.
- [2] A. Alfonsi, J. Corbetta, and B. Jourdain. Sampling of probability measures in the convex order and approximation of martingale optimal transport problems. *Available at SSRN 3072356*, 2017.
- [3] Jason M Altschuler and Enric Boix-Adsera. Wasserstein barycenters can be computed in polynomial time in fixed dimension. *J. Mach. Learn. Res.*, 22:44–1, 2021.
- [4] P.C. Alvarez-Esteban, E Del Barrio, J.A. Cuesta-Albertos, and C. Matrán. Wide consensus for parallelized inference. *arXiv: 1511.05350*, 2015.
- [5] P.C. Álvarez-Esteban, E. Del Barrio, J.A. 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.
- [6] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient flows: in metric spaces and in the space of probability measures*. Springer Science & Business Media, 2008.
- [7] M. Arjovsky, S. Chintala, and L. Bottou. Wasserstein generative adversarial networks. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70, pages 214–223. PMLR, 2017.
- [8] J. Backhoff-Veraguas, M. Beiglböck, and G. Pammer. Existence, duality, and cyclical monotonicity for weak transport costs. *Calculus of Variations and Partial Differential Equations*, 58(6):203, 2019.
- [9] J. Backhoff-Veraguas, M. Beiglböck, and G. Pammer. Weak monotone rearrangement on the line. *Electronic Communications in Probability*, 25, 2020.

- [10] J. Backhoff-Veraguas, J. Fontbona, G. Rios, and F. Tobar. Stochastic gradient descent in Wasserstein space. *arXiv preprint arXiv: 2201.04232*, 2022.
- [11] Y. Balaji, R. Chellappa, and S. Feizi. Robust optimal transport with applications in generative modeling and domain adaptation. *arXiv:2010.05862*, 2020.
- [12] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM journal on imaging sciences*, 2(1):183–202, 2009.
- [13] M. Beiglböck, P. Henry-Labordère, and F. Penkner. Model-independent bounds for option prices—a mass transport approach. *Finance and Stochastics*, 17(3):477–501, 2013.
- [14] J. Bigot, E. Cazelles, and N. Papadakis. Data-driven regularization of Wasserstein barycenters with an application to multivariate density registration. *Information and Inference: A Journal of the IMA*, 8(4):719–755, 2019.
- [15] J. Bigot and T. Klein. Characterization of barycenters in the Wasserstein space by averaging optimal transport maps. *ESAIM: Probability and Statistics*, 22:35–57, 2018.
- [16] M.W. Botsko. An elementary proof of Lebesgue’s differentiation theorem. *The American Mathematical Monthly*, 110(9):834–838, 2003.
- [17] S. Chewi, T. Maunu, P. Rigollet, and A.J. Stromme. Gradient descent algorithms for Bures-Wasserstein barycenters. *arXiv: 2001.01700*, 2020.
- [18] M. Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in Neural Information Processing Systems 26*, pages 2292–2300. 2013.
- [19] M. Cuturi and A. Doucet. Fast computation of Wasserstein barycenters. *Proceedings of the 31st International Conference on Machine Learning*, 32(2):685–693, 2014.
- [20] A. Dessein, N. Papadakis, and J.-L. Rouas. Regularized optimal transport and the ROT mover’s distance. *The Journal of Machine Learning Research*, 19(1):590–642, 2018.
- [21] D. Dvinskikh. Stochastic approximation versus sample average approximation for population Wasserstein barycenters. *arXiv e-prints*, pages arXiv–2001, 2020.
- [22] E.A. Feinberg, P.O. Kasyanov, and Y. Liang. Fatou’s lemma in its classical form and Lebesgue’s convergence theorems for varying measures with applications to Markov decision processes. *Theory of Probability & Its Applications*, 65(2):270–291, 2020.
- [23] R. Flamary, N. Courty, A. Gramfort, M.Z. Alaya, A. Boisbunon, S. Chambon, L. Chapel, A. Corenflos, K. Fatras, N. Fournier, L. Gautheron, N. T.H. Gayraud, H. Janati, A. Rakotomamonjy, I. Redko, A. Rolet, A. Schutz, V. Seguy, D.J. Sutherland, R. Tavenard, A. Tong, and T. Vayer. Pot: Python optimal transport. *Journal of Machine Learning Research*, 22(78):1–8, 2021.
- [24] R.C. Gentleman, V.J. Carey, D.M. Bates, B. Bolstad, M. Dettling, S. Dudoit, B. Ellis, L. Gautier, Y. Ge, J. Gentry, et al. Bioconductor: open software development for computational biology and bioinformatics. *Genome biology*, 5(10):R80, 2004.
- [25] N. Gozlan and N. Juillet. On a mixture of Brenier and Strassen theorems. *Proceedings of the London Mathematical Society*, 120(3):434–463, 2020.
- [26] N. Gozlan, C. Roberto, P.-M. Samson, and P. Tetali. Kantorovich duality for general transport costs and applications. *Journal of Functional Analysis*, 273(11):3327–3405, 2017.
- [27] J. Guyon and P. Henry-Labordere. *Nonlinear option pricing*. CRC Press, 2013.
- [28] F. Hahne, A.H. Khodabakhshi, A. Bashashati, C.-J. Wong, R.D. Gascoyne, A.P. Weng, V. Seyfert-Margolis, K. Bourcier, A. Asare, T. Lumley, et al. Per-channel basis normalization methods for flow cytometry data. *Cytometry Part A: The Journal of the International Society for Advancement of Cytometry*, 77(2):121–131, 2010.
- [29] R.P. Kertz and U. Rösler. Complete lattices of probability measures with applications to martingale theory. *Lecture Notes-Monograph Series*, pages 153–177, 2000.
- [30] T. Le Gouic and J.-M. Loubes. Existence and consistency of Wasserstein barycenters. *Probability Theory and Related Fields*, 168(3-4):901–917, 2017.

- [31] Y. LeCun. The MNIST database of handwritten digits. <http://yann.lecun.com/exdb/mnist/>, 1998.
- [32] L. Li, A. Genevay, M. Yurochkin, and J. Solomon. Continuous regularized wasserstein barycenters. *arXiv preprint arXiv:2008.12534*, 2020.
- [33] X. Nguyen. Convergence of latent mixing measures in finite and infinite mixture models. *The Annals of Statistics*, 41(1):370–400, 2013.
- [34] G. Peyré and M. Cuturi. Computational optimal transport. *Foundations and Trends® in Machine Learning*, 11(5-6):355–607, 2019.
- [35] G. Rios, J. Backhoff-Veraguas, J. Fontbona, and F. Tobar. Bayesian learning with Wasserstein barycenters. *arXiv preprint arXiv:1805.10833v4*, 2018.
- [36] V. Seguy, B.B. Damodaran, R. Flamary, N. Courty, A. Rolet, and M. Blondel. Large-scale optimal transport and mapping estimation. In *International Conference on Learning Representations (ICLR)*, 2018.
- [37] Moshe Shaked and J George Shanthikumar. *Stochastic orders*. Springer Science & Business Media, 2007.
- [38] J. Solomon, F. De Goes, G. Peyré, M. Cuturi, A. Butscher, A. Nguyen, T. Du, and L. Guibas. Convolutional Wasserstein distances: Efficient optimal transportation on geometric domains. *ACM Transactions on Graphics (TOG)*, 34(4):1–11, 2015.
- [39] M. Staib, S. Claiici, J. Solomon, and S. Jegelka. Parallel streaming wasserstein barycenters. In *Advances in Neural Information Processing Systems*, pages 2647–2658, 2017.
- [40] V. Strassen. The existence of probability measures with given marginals. *Annals of Mathematical Statistics*, 36(2):423–439, 1965.
- [41] C. Villani. *Topics in Optimal Transportation*. Number 58. American Mathematical Soc., 2003.
- [42] C. Villani. *Optimal Transport: Old and New*. Springer Berlin Heidelberg, 2008.
- [43] J. Ye, P. Wu, J. Z. Wang, and J. Li. Fast discrete distribution clustering using Wasserstein barycenter with sparse support. *IEEE Transactions on Signal Processing*, 65(9):2317–2332, 2017.
- [44] Y. Zemel and V. Panaretos. Fréchet means and Procrustes analysis in Wasserstein space. *Bernoulli*, 25(2):932–976, 2019.

Checklist

1. For all authors...
 - (a) Do the main claims made in the abstract and introduction accurately reflect the paper’s contributions and scope? **[Yes]**
 - (b) Did you describe the limitations of your work? **[Yes]** See Section 4 and Discussion 7.
 - (c) Did you discuss any potential negative societal impacts of your work? **[N/A]**
 - (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? **[Yes]**
2. If you are including theoretical results...
 - (a) Did you state the full set of assumptions of all theoretical results? **[Yes]**
 - (b) Did you include complete proofs of all theoretical results? **[Yes]** Each Theorem, Proposition and Lemma is carefully proved in Appendices B, C and D.
3. If you ran experiments...
 - (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? **[Yes]** The toolbox for implementing the algorithms and the OWT problem and replicate the experiments of this paper is included in the Supplement and will be made publicly available after anonymity is lifted.
 - (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? **[Yes]** See Appendix Sections 5 and E.1.
 - (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? **[No]**

- (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [No]
- 4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
 - (a) If your work uses existing assets, did you cite the creators? [Yes] See [23, 20, 12, 36].
 - (b) Did you mention the license of the assets? [N/A]
 - (c) Did you include any new assets either in the supplemental material or as a URL? [Yes]
A toolbox to reproduce our proposed model is provided in the supplemental material.
 - (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [N/A]
 - (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [N/A]
- 5. If you used crowdsourcing or conducted research with human subjects...
 - (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]
 - (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]
 - (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]