Aequat. Math.

https://doi.org/10.1007/s00010-024-01049-1

Aequationes Mathematicae

A note on the Radiant formula and its relations to the sliced Wasserstein distance

Gennaro Auricchio

Abstract. In this note, we show that the squared Wasserstein distance can be expressed as the average over the sphere of one dimensional Wasserstein distances. We name this new expression for the Wasserstein Distance *Radiant Formula*. Using this formula, we are able to highlight new connections between the Wasserstein distances and the Sliced Wasserstein distance, an alternative Wasserstein-like distance that is cheaper to compute.

Mathematics Subject Classification. 49Q22, 30L15.

Keywords. Optimal transport, Radiant formula, Sliced Wasserstein distance.

1. Introduction and basic notation

In modern mathematical language, the p-th power of the Wasserstein distance between two probability measures over $\mathbb{R}^d$, namely μ and ν, is defined as

$$W_p^p(\mu,\nu) = \min_{\pi\in\Pi(\mu,\nu)} \int_{\mathbb{R}^d\times\mathbb{R}^d} l_p^p(x,y)d\pi, \tag{1}$$

where $l_p^p(x,y) = \sum_{i=1}^d \mid x_i - y_i \mid^p$ and $\Pi(\mu,\nu)$ is the set of measures over $\mathbb{R}^d\times\mathbb{R}^d$ whose marginals are μ and ν, [1].

Due to its ability of capturing the weak topology of the space of probability measures, the family of W_p^p distances has found a natural home in many applied fields, such as Computer Vision [2,3], generative models [4–6], and clustering [7,8]. For this reason, much effort has been spent to find cheap ways to compute the value of W_p^p given two measures. When μ and ν are discrete measures, the minimization problem (1) can be cast as an Linear Programming (LP) problem. Due to the separability of the l_p^p cost functions, it is possible to lower the complexity of these LP problems, [9,10]; however, for many applied tasks, this is yet not enough to make W_p^p an efficient alternative to

Published online: 06 April 2024

Birkhäuser

other metrics. Therefore several cheap-to-compute alternatives to the Wasserstein distance have been proposed: some approaches rely on adding an entropy regularization term [11–13] to the objective of (1), while other approaches considers topological equivalent alternatives, like, for example, the Fourier Based metrics [14,15]. Another successful alternative is the *Sliced Wasserstein Distance* (SWD) [16,17]. The SWD computes the distance between two measures by comparing their projection on all possible affine 1-dimensional sub-spaces of $\mathbb{R}^d$. Since the Wasserstein distance between measures supported over a line can be computed through an explicit formula, the SWD can be computed without solving a minimization problem.

In this note, we propose a general methodology to relate the original Wasserstein distances to the SWDs. First, we show that, when both the probability measures are supported over $\mathbb{R}^2$, the W_2^2 distance can be represented as follows

$$W_2^2(\mu,\nu) = \frac{1}{\pi}\int_0^{2\pi}\bigg(\int_{\mathbb{R}} W_2^2(\mu_{|x_2^{(\theta)}}, \zeta^{(\theta)}_{|x_2^{(\theta)}})d\mu_2^{(\theta)}\bigg)d\theta,$$

where $\{\zeta^{(\theta)}\}_{\theta\in[0,2\pi]}$ is a suitable family of measures on $\mathbb{R}^2$ (see Theorem 2). We call this identity Radiant formula and use it to find equivalence bounds between the classic Wasserstein distances and their sliced counterparts. We then extend these results to the case $p \neq 2$ and use the Knothe-Rosenblatt heuristic transportation plan [18,19] to provide an upper bound on the absolute error between the SWD and W_p^p. Finally, we extend our results to any $\mathbb{R}^d$, with $d \geq 2$.

2. Our contribution

For the sake of clarity, we first introduce our results for measures supported over $\mathbb{R}^2$ and then extend our findings to the higher dimensional setting in a dedicated subsection. In what follows, we denote with $\mathcal{P}(\mathbb{R}^d)$ the set of Borel probability measures over $\mathbb{R}^d$.

The Radiant formula

As a starting point of our discussion, we show that any W_p distance can be computed by summing the averages two one-dimensional Wasserstein distances between μ and two suitable probability measures.

Proposition 1. *Let $\mu,\nu \in \mathcal{P}(\mathbb{R}^2)$ and $p \geq 1$. Then, there exists a couple of measures (ζ,η),*[1] *such that $\zeta \in \Pi(\nu_1,\mu_2)$, $\eta \in \Pi(\mu_1,\nu_2)$, and*

$$W_p^p(\mu,\nu) = \int_{\mathbb{R}} W_p^p(\mu_{|x_1},\eta_{|x_1})d\mu_1 + \int_{\mathbb{R}} W_p^p(\mu_{|x_2},\zeta_{|x_2})d\mu_2, \tag{2}$$

[1] These measures do depend on μ, ν, and p in general, however, to lighten up the notation, we drop these indexes.

where λ_i *is the marginal of* λ *on the* i*-th coordinate and* $\lambda_{|x_i}$ *is the conditional law of* λ *given* x_i*.*

Proof. Let $\gamma \in \mathcal{P}(\mathbb{R}^2 \times \mathbb{R}^2)$ be an optimal transportation plan between μ and ν. We then have that

$$\begin{aligned} W_p^p(\mu,\nu) &= \int_{\mathbb{R}^2\times\mathbb{R}^2} \Big(\mid x_1 - y_1 \mid^p + \mid x_2 - y_2 \mid^p \Big) d\gamma \\ &= \int_{\mathbb{R}^2\times\mathbb{R}} \mid x_1 - y_1 \mid^p df + \int_{\mathbb{R}\times\mathbb{R}^2} \mid x_2 - y_2 \mid^p dg, \end{aligned}$$

where f is the marginal of γ on (x_1, x_2, y_1) e and g is the marginal of γ over (x_1, x_2, y_2). Finally, let η and ζ be the marginals of γ over (x_1, y_2) and (y_1, x_2), respectively. Since γ is optimal, from [20], we have that

$$\int_{\mathbb{R}^2\times\mathbb{R}} \mid x_1 - y_1 \mid^p df = \int_{\mathbb{R}} W_p^p(\mu_{|x_2}, \zeta_{|x_2}) d\mu_2. \tag{3}$$

Similarly, we have that $\int_{\mathbb{R}^2\times\mathbb{R}} \mid x_2 - y_2 \mid^p dg = \int_{\mathbb{R}} W_p^p(\mu_{|x_1}, \eta_{|x_1}) d\mu_1$, which concludes the proof. □

Let us set $V_\theta = \{v_\theta, v_{\theta^\perp}\}$, where $v_\theta = (\cos(\theta), \sin(\theta))$ and $v_{\theta^\perp} = (-\sin(\theta), \cos(\theta))$. We notice that, for every $\theta \in [0, 2\pi]$, V_θ is the basis of $\mathbb{R}^2$ obtained by applying a θ-counterclockwise rotation of the canonical base $V = \{e_1, e_2\}$. In what follows, we denote with $(x_1^{(\theta)}, x_2^{(\theta)})$ the coordinates of $\mathbb{R}^2$ with respect to the base V_θ. Moreover, we denote with $\mu_1^{(\theta)}$ and $\mu_2^{(\theta)}$ the marginals of μ on $x_1^{(\theta)}$ and $x_2^{(\theta)}$, respectively. In this framework, given $p \in [1, \infty)$, the Sliced Wasserstein Distance is defined as follows

$$SW_p^p(\mu,\nu) = \frac{1}{2\pi} \int_0^{2\pi} W_p^p(\mu_1^{(\theta)}, \nu_1^{(\theta)}) d\theta. \tag{4}$$

Finally, we denote with $R_\theta : \mathbb{R}^2 \to \mathbb{R}^2$ the rotation that satisfies $R_\theta(e_1) = v_\theta$ and $R_\theta(e_2) = v_{\theta^\perp}$ and with $\mu^{(\theta)} := (R_\theta)_\# \mu$ the push-forward of μ through R_θ.[2] Notice that, according to our notation, the marginal of $\mu^{(\theta)}$ over the first coordinat coincides with $\mu_1^{(\theta)}$.

Theorem 2. *(The Radiant Formula) Let* $\mu, \nu \in \mathcal{P}(\mathbb{R}^2)$*. Then there exists a family of measures* $\{\zeta^{(\theta)}\}_{\theta\in[0,2\pi]}$ *such that, for every* $\theta \in [0, 2\pi]$*, it holds* $\zeta^{(\theta)} \in \Pi(\mu_2^{(\theta)}, \nu_1^{(\theta)})$ *and*

$$W_2^2(\mu,\nu) = \frac{1}{\pi} \int_0^{2\pi} \left(\int_{\mathbb{R}} W_2^2(\mu_{|x_2^{(\theta)}}, \zeta^{(\theta)}_{|x_2^{(\theta)}}) d\mu_2^{(\theta)} \right) d\theta, \tag{5}$$

where $\zeta^{(\theta)}_{|x_2^{(\theta)}}$ *is the conditional law of* $\zeta^{(\theta)}$ *given* $x_2^{(\theta)}$*. Moreover, if both* μ *and* ν *are absolutely continuous, the family* $\{\zeta^{(\theta)}\}_{\theta\in[0,2\pi]}$ *is unique.*

[2] We recall that $(R_\theta)_\#\mu$ is defined as $(R_\theta)_\#\mu(A) = \mu(R_\theta^{-1}(A))$ for every Borel set $A \subset \mathbb{R}^2$.

Proof. First, we notice that the W_2^2 distance between two measures μ and ν is preserved if we apply a rotation to $\mathbb{R}^2$. Indeed, if γ is an optimal transportation plan between μ and ν, the plan $(R_\theta, R_\theta)_\#\gamma$ is optimal between $(R_\theta)_\#\mu$ and $(R_\theta)_\#\nu$. This is due to the fact that $l_2^2(x,y) = l_2^2(R_\theta(x), R_\theta(y))$ for every $x, y \in \mathbb{R}^2$ and for every $\theta \in [0, 2\pi]$.

Given $\theta \in [0, 2\pi]$, Proposition 1 gives us a couple of measures, namely $\eta^{(\theta)}$ and $\zeta^{(\theta)}$ such that

$$W_p^p(\mu,\nu) = \int_{\mathbb{R}} W_p^p(\mu_{|x_1^{(\theta)}}, \eta^{(\theta)}_{|x_1^{(\theta)}})d\mu_1^{(\theta)} + \int_{\mathbb{R}} W_p^p(\mu_{|x_2^{(\theta)}}, \zeta^{(\theta)}_{|x_2^{(\theta)}})d\mu_2^{(\theta)}, \tag{6}$$

Since a $\frac{\pi}{2}$-counterclockwise rotation swaps the basis vectors in any V_θ base (i.e. v_θ and $v_{\theta^\perp}$), we have that $\zeta^{(\theta+\frac{\pi}{2})} = \eta^{(\theta)}$. Thus, we have

$$\int_{\mathbb{R}} W_p^p(\mu_{|x_1^{(\theta)}}, \eta^{(\theta)}_{|x_1^{(\theta)}})d\mu_1^{(\theta)} = \int_{\mathbb{R}} W_2^2(\mu_{|x_2^{(\theta+\frac{\pi}{2})}}, \zeta^{(\theta+\frac{\pi}{2})}_{|x_2^{(\theta+\frac{\pi}{2})}})d\mu_2^{(\theta+\frac{\pi}{2})}, \tag{7}$$

for each $\theta \in (0, 2\pi]$. By substituting (7) in (6), we find

$$W_2^2(\mu,\nu) = \int_{\mathbb{R}} W_2^2(\mu_{|x_2^{(\theta)}}, \zeta^{(\theta)}_{|x_2^{(\theta)}})d\mu_2^{(\theta)} + \int_{\mathbb{R}} W_2^2(\mu_{|x_2^{(\theta+\frac{\pi}{2})}}, \zeta^{(\theta+\frac{\pi}{2})}_{|x_2^{(\theta+\frac{\pi}{2})}})d\mu_2^{(\theta+\frac{\pi}{2})}. \tag{8}$$

Since (8) holds true for every $\theta \in [0, 2\pi]$, we can take the integral media and retrieve the radiant formula

$$\begin{aligned}
W_2^2(\mu,\nu) &= \frac{1}{2\pi}\int_{[0,2\pi]} W_2^2(\mu,\nu)d\theta \\
&= \frac{1}{2\pi}\int_{[0,2\pi]} \Big(\int_{\mathbb{R}} W_2^2(\mu_{|x_2^{(\theta)}}, \zeta^{(\theta)}_{|x_2^{(\theta)}})d\mu_2^{(\theta)} + \int_{\mathbb{R}} W_2^2(\mu_{|x_2^{(\theta+\frac{\pi}{2})}}, \zeta^{(\theta+\frac{\pi}{2})}_{|x_2^{(\theta+\frac{\pi}{2})}})d\mu_2^{(\theta+\frac{\pi}{2})} \Big) d\theta \\
&= \frac{1}{\pi}\int_{[0,2\pi]}\int_{\mathbb{R}} W_2^2(\mu_{|x_2^{(\theta)}}, \zeta^{(\theta)}_{|x_2^{(\theta)}})d\mu_2^{(\theta)} d\theta.
\end{aligned}$$

Finally, the uniqueness result follows from the uniqueness of the transportation plan between absolutely continuous probability measures, [21]. □

In Fig. 1, we give a visual example of the family $\{\zeta^{(\theta)}\}_{\theta\in[0,2\pi]}$ for two Dirac's deltas.

To prove Theorem 2, we made use of the rotation invariance property of W_2^2. This property, however, does not hold for W_p^p, which prevents us from expressing W_p^p using a radiant formula. However, we bypass this issue by defining a rotation-averaged version of the W_p distance as follows

$$RW_p^p(\mu,\nu) := \frac{1}{2\pi}\int_0^{2\pi} W_p^p(\mu^{(\theta)}, \nu^{(\theta)})d\theta. \tag{9}$$

We notice that $RW_2(\mu,\nu) = W_2(\mu,\nu)$ for every $\mu, \nu \in \mathcal{P}(\mathbb{R}^2)$.

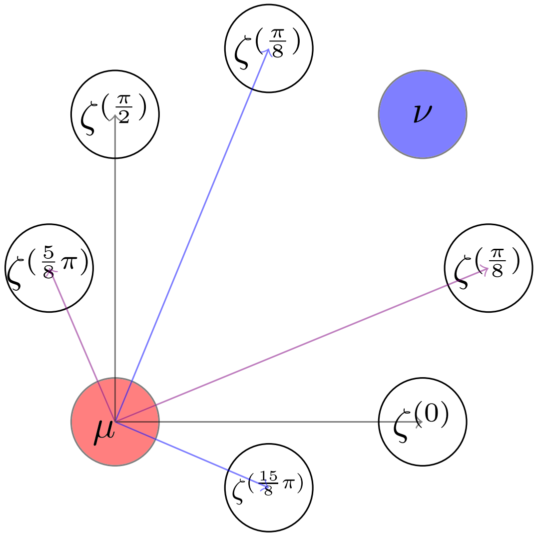

FIGURE 1. An example of the family $\{\zeta^{(\theta)}\}_{\theta\in[0,2\pi]}$ for two Dirac's deltas (we represent μ with the red dot and ν with the blue dot). The white dots represent a different $\zeta^{(\theta)}$ for different values of θ. The arrows represent the different flows $f^{(\theta)}$ and $g^{(\theta)}$. Arrows with the same colour are used to connect μ to $\zeta^{(\theta)}$ and $\zeta^{(\theta+\frac{\pi}{2})}$ (color figure online)

Proposition 3. *The function RW_p defined in (9) is a distance over $\mathcal{P}(\mathbb{R}^2)$, and it is invariant under rotation of the coordinates. Furthermore, it holds*

$$RW_p^p(\mu,\nu) \leq \frac{K_p}{2\pi} W_{1,p}^p(\mu,\nu) \leq n^{(\frac{p}{2}-1)_+} \frac{K_p}{2\pi} W_p^p(\mu,\nu), \tag{10}$$

where $(\circ)_+$ is the positive part function, $K_p := 2\int_0^{2\pi} \mid \cos(x) \mid^p dx$ and

$$W_{1,p}^p(\mu,\nu) := \min_{\pi\in\Pi(\mu,\nu)} \int_{\mathbb{R}^d\times\mathbb{R}^d} l_1^p(x,y)d\pi.$$

Thus, up to a constant, $RW_p^p(\mu,\nu)$ is dominated by W_p^p and $W_{1,p}$. Moreover, the first upper bound is tight.

Proof. We divide the proof the proposition into three pieces.

RW_p **is invariant under rotations** It follows from the fact that RW_p is defined as the average of the costs with respect to all the possible choices of coordinates. Indeed, given $\phi \in [0, 2\pi]$, let $\mu^{(\phi)} = (R_\phi)_\#\mu$ and $\nu^{(\phi)} = (R_\phi)_\#\nu$.

Then, it holds $(\mu^{(\phi)})^{(\theta)} = \mu^{(\theta+\phi)}$; thus

$$\begin{aligned} RW_p^p(\mu^{(\phi)}, \nu^{(\phi)}) &= \frac{1}{2\pi}\int_0^{2\pi} W_p^p(\mu^{(\phi+\theta)}, \nu^{(\phi+\theta)})d\theta \\ &= \frac{1}{2\pi}\int_0^{2\pi} W_p^p(\mu^{(\theta')}, \nu^{(\theta')})d\theta' = RW_p^p(\mu,\nu), \end{aligned}$$

where we used the change of variable $\theta' = \theta + \phi$.

RW_p is a distance First, notice that RW_p is symmetric since W_p is symmetric. Similarly, if $\mu = \nu$, we have that $W_p^p(\mu^{(\theta)}, \nu^{(\theta)}) = 0$ for every θ, thus $RW_p^p(\mu,\nu) = 0$. Conversely, since $W_p^p(\mu,\nu) \geq 0$, we have that $RW_p^p(\mu,\nu) = 0$ if and only if $W_p^p(\mu^{(\theta)}, \nu^{(\theta)}) = 0$ for almost every $\theta \in [0, 2\pi]$, hence $\mu = \nu$. To conclude, we prove the triangular inequality. Let μ, ν, and ζ be elements of $\mathcal{P}(\mathbb{R}^2)$. From the Minkowsky's inequality [21], we have that

$$\begin{aligned} RW_p(\mu,\nu) &= \left(\frac{1}{2\pi}\int_0^{2\pi} W_p^p(\mu^{(\theta)}, \nu^{(\theta)})d\theta\right)^{\frac{1}{p}} \\ &\leq \left(\frac{1}{2\pi}\int_0^{2\pi} (W_p(\mu^{(\theta)}, \zeta^{(\theta)}) + W_p(\zeta^{(\theta)}, \nu^{(\theta)}))^p d\theta\right)^{\frac{1}{p}} \\ &\leq \left(\frac{1}{2\pi}\int_0^{2\pi} W_p^p(\mu^{(\theta)}, \zeta^{(\theta)})d\theta\right)^{\frac{1}{p}} + \left(\frac{1}{2\pi}\int_0^{2\pi} W_p^p(\zeta^{(\theta)}, \nu^{(\theta)})\right)^{\frac{1}{p}} \\ &= RW_p(\mu,\zeta) + RW_p(\zeta,\nu), \end{aligned}$$

which concludes the second part of the proof.

RW_p is dominated by W_p and $W_{1,p}$ Let us consider $x, y \in \mathbb{R}^2$. Let ρ and ϕ be the polar coordinates of $x - y$, so that $x - y = \rho(\cos(\phi), \sin(\phi))$. We then have

$$x_1 - y_1 = \rho\cos(\phi) \qquad \text{and} \qquad x_2 - y_2 = \rho\sin(\phi).$$

We thus infer

$$| x_1 - y_1 |^p + | x_2 - y_2 |^p = \rho^p(| \cos(\phi) |^p + | \sin(\phi) |^p).$$

Let $\gamma \in \Pi(\mu,\nu)$ be an optimal transportation plan between μ and ν with respect to the p-th power of the Euclidean metric, that is $d(x,y) = || x - y ||_2^p$, so that

$$W_{1,p}^p(\mu,\nu) := \int_{\mathbb{R}^2\times\mathbb{R}^2} || x - y ||_2^p \, d\gamma.$$

Finally, we have

$$\begin{aligned}
RW_p^p(\mu,\nu) &= \frac{1}{2\pi}\int_0^{2\pi}\Big(\min_{\pi\in\Pi(\mu^{(\theta)},\nu^{(\theta)})} l_p^p(x,y)d\pi\Big)d\theta \\
&= \frac{1}{2\pi}\int_0^{2\pi}\Big(\min_{\pi\in\Pi(\mu,\nu)} l_p^p(R_\theta(x),R_\theta(y))d\pi\Big)d\theta \\
&\leq \frac{1}{2\pi}\int_0^{2\pi}\int_{\mathbb{R}^2\times\mathbb{R}^2}\rho^p(\mid\cos(\phi+\theta)\mid^p+\mid\sin(\phi+\theta)\mid^p)d\gamma d\theta \\
&= \frac{1}{2\pi}\int_{\mathbb{R}^2\times\mathbb{R}^2}\rho^p\int_0^{2\pi}(\mid\cos(\phi+\theta)\mid^p+\mid\sin(\phi+\theta)\mid^p)d\theta d\gamma \\
&= \frac{K_p}{2\pi}\int_{\mathbb{R}^2\times\mathbb{R}^2}\rho^p d\gamma = \frac{K_p}{2\pi}W_{1,p}^P(\mu,\nu),
\end{aligned}$$

where

$$K_p = 2\int_0^{2\pi}\mid\cos(\theta)\mid^p d\theta.$$

To conclude the proof, we recall the classic inequality

$$\mid\mid x-y\mid\mid_2^p \leq n^{(\frac{p}{2}-1)_+}\mid\mid x-y\mid\mid_p^p.$$

The tightness of the first inequality in (10) follows by considering two Dirac's delta. □

Since RW_p^p is a rotation invariant distance, we are able to express it through a radiant formula.

Theorem 4. *Let $p\geq 1$. Let μ and ν be two measures supported over $\mathbb{R}^2$. Then, there exists a family of measures $\{\zeta^{(\theta)}\}_{\theta\in[0,\pi]}$ such that*

$$RW_p^p(\mu,\nu) = \frac{1}{\pi}\int_0^{2\pi}\int_{\mathbb{R}} W_p^p(\mu_{|x_2^{(\theta)}},\zeta^{(\theta)}_{|x_2^{(\theta)}})d\mu_2^{(\theta)}d\theta. \tag{11}$$

Proof. Given any $\theta\in[0,2\pi]$, from Proposition 1, we have that there exists a couple of measures $\zeta^{(\theta)}$ and $\eta^{(\theta)}$ such that

$$W_p^p(\mu^{(\theta)},\nu^{(\theta)}) = \int_{\mathbb{R}} W_p^p(\mu^{(\theta)}_{|x_2^{(\theta)}},\zeta^{(\theta)}_{|x_2^{(\theta)}})d\mu_2^{(\theta)} + \int_{\mathbb{R}} W_p^p(\mu^{(\theta)}_{|x_1^{(\theta)}},\eta^{(\theta)}_{|x_1^{(\theta)}})d\mu_1^{(\theta)}.$$

By taking the average over θ, we conclude the thesis. □

Relation with the sliced Wasserstein distance

We now highlight how the Radiant Formula allows us to retrieve bounds on the Sliced Wasserstein distance in terms of the classic Wasserstein distance.

Theorem 5. *Given two probability measures $\mu,\nu\in\mathcal{P}(\mathbb{R}^2)$, we have*

$$SW_2^2(\mu,\nu) \leq \frac{1}{2}W_2^2(\mu,\nu).$$

Moreover, the bound is tight. Similarly, it holds

$$SW_p^p(\mu,\nu) \le n^{(\frac{p}{2}-1)_+}\frac{K_p}{\pi}W_p^p(\mu,\nu).$$

Proof. It follows from the convexity of the W_2^2 distance [21, Theorem 4.8]. Indeed, from the Radiant Formula (5) we know that

$$W_2^2(\mu,\nu) = \frac{1}{\pi}\int_0^{2\pi}\int_{\mathbb{R}} W_2^2(\mu_{|x_2^{(\theta)}}, \zeta^{(\theta)}_{|x_2^{(\theta)}})d\mu_2^{(\theta)}\,d\theta$$

then, following the notation in [21], if we set $\lambda = \mu_2^{(\theta)}$, $\mu = \mu_{|x_2^{(\theta)}}$ and $\nu = \zeta^{(\theta)}_{|x_2^{(\theta)}}$, we conclude

$$\begin{aligned} W_2^2(\mu,\nu) &= \frac{1}{\pi}\int_0^{2\pi}\int_{\mathbb{R}} W_2^2(\mu_{|x_2^{(\theta)}}, \zeta^{(\theta)}_{|x_2^{(\theta)}})d\mu_2^{(\theta)}d\theta \\ &\ge \frac{1}{\pi}\int_0^{2\pi} W_2^2(\mu_1^{(\theta)},\nu_1^{(\theta)})d\theta = 2\,SW_2^2(\mu,\nu) \end{aligned} \tag{12}$$

where the equality (12) comes from the fact that each $\zeta^{(\theta)} \in \Pi(\nu_1^{(\theta)},\mu_2^{(\theta)})$. To prove the tightness, it suffice to consider the measures $\mu = \delta_{(0,0)}$ and $\nu = \delta_{(1,1)}$.

By the same argument, we infer the bound on SW_p^p. □

Finally, we use the Knothe-Rosenblatt transportation plan [19,22] to bound the absolute error we commit by using the Sliced Wasserstein Distance over the Wasserstein Distance.

Theorem 6. *Given $\mu,\nu \in \mathcal{P}(\mathbb{R}^2)$, it holds*

$$\mid W_2^2(\mu,\nu) - SW_2^2(\mu,\nu)\mid \le \frac{1}{2\pi}\int_0^{2\pi}\int_{\mathbb{R}} W_2^2((\zeta_{KR}^{(\theta)})_{|y_1^{(\theta)}},\nu_{|y_1^{(\theta)}})d\nu_1^{(\theta)}d\theta, \tag{13}$$

where $\zeta^{(\theta)}_{KR}$ is the marginal of Knothe-Rosenblatt transportation plan on the coordinates $(x_2^{(\theta)}, y_1^{(\theta)})$. Furthermore, the upper bound in (13) is tight.

Proof. Let $V^{(\theta)}$ be a base of the space and let $\gamma^{(\theta)}_{KR}$ and $\zeta^{(\theta)}_{KR}$ be the Knothe-Rosenblatt transportation plan and its projection over $(x_2^{(\theta)}, y_1^{(\theta)})$, respectively.[3] Then, by definition of the Knothe-Rosenblatt transportation plan, we have

$$W_2^2(\mu,\nu) \le W^2_{KR}(\mu^{(\theta)},\nu^{(\theta)}) = W_2^2(\mu_1^{(\theta)},\nu_1^{(\theta)}) + \int_{\mathbb{R}} W_2^2((\zeta_{KR}^{(\theta)})_{|y_1^{(\theta)}},\nu_{|y_1^{(\theta)}})d\nu_1^{(\theta)}, \tag{14}$$

[3]Notice that the Knothe-Rosenblatt plan does depend on the choice of $V^{(\theta)}$.

where $W^2_{KR}(\mu^{(\theta)},\nu^{(\theta)})$ is the cost of the Knothe-Rosenblatt transportation plan according to the squared Euclidean distance. Since for every $\theta \in [0,2\pi]$ it holds $W_2^2(\mu_1^{(\theta)},\nu_1^{(\theta)}) \leq W_2^2(\mu,\nu)$, we infer

$$0 \leq W_2^2(\mu,\nu) - W_2^2(\mu_1^{(\theta)},\nu_1^{(\theta)}) \leq \int_{\mathbb{R}} W_2^2((\zeta_{KR}^{(\theta)})_{|y_1^{(\theta)}}, \nu_{|y_1^{(\theta)}}) d\nu_1^{(\theta)} \tag{15}$$

for any $\theta \in [0,2\pi)$. Finally, by taking the average over $[0,2\pi)$, we find

$$\frac{1}{2\pi}\int_0^{2\pi}\Big(W_2^2(\mu,\nu) - W_2^2(\mu_1^{(\theta)},\nu_1^{(\theta)})\Big)d\theta = \frac{1}{2\pi}\int_0^{2\pi}\int_{\mathbb{R}} W_2^2((\zeta_{KR}^{(\theta)})_{|y_1^{(\theta)}}, \nu_{|y_1^{(\theta)}}) d\nu_1^{(\theta)},$$

and, hence

$$|\ W_2^2(\mu,\nu) - SW_2^2(\mu,\nu)\ | \leq \frac{1}{2\pi}\int_0^{2\pi}\int_{\mathbb{R}} W_2^2((\zeta_{KR}^{(\theta)})_{|y_1^{(\theta)}}, \nu_{|y_1^{(\theta)}}) d\nu_1^{(\theta)},$$

which concludes the first part of the proof.

To prove the tightness, it suffice to consider the measures $\mu = \delta_{(0,0)}$ and any measure ν. In this case, the Knothe-Rosenblatt transportation plan between μ and ν is optimal, thus the inequality in (15) is an equality for every $\theta \in [0,2\pi]$. □

The extension to higher dimensional spaces

To conclude, we extend our results to the case in which the measures are supported over a higher dimensional space. We denote with d the dimension of the space, so that $\mu,\nu \in \mathcal{P}(\mathbb{R}^d)$. Moreover, let $\mathcal{R}_d$ the set of all the rotations from $\mathbb{R}^d$ to $\mathbb{R}^d$. Given $R \in \mathcal{R}_d$, let use define $e_i^{(R)} = R(e_i)$, where e_i is the i-th vector in the canonical base of $\mathbb{R}^d$. It is easy to see that $\{e_i^{(R)}\}_{i=1,\dots,d}$ is an orthonormal base of $\mathbb{R}^d$, we use $x_1^{(\theta)}, x_2^{(\theta)}, \dots, x_d^{(\theta)}$ to denote the coordinates of $\mathbb{R}^d$ with respect to the base $\{e_i^{(R)}\}$. Finally, we set ρ to be the uniform probability distribution over $\mathcal{R}_d$, since the set of rotation is identifiable as a compact set of the orthogonal matrices, this measure is well-defined.

Given a couple of measures μ and ν, the SWD is defined as follows

$$SW_p^p(\mu,\nu) = \frac{1}{|\ \mathbb{S}^{(d-1)}\ |}\int_{\mathbb{S}^{(d-1)}} W_p^p(\mu_v,\nu_v)dv, \tag{16}$$

where $\mathbb{S}^{(d-1)}$ is the set of all directions in $\mathbb{R}^d$ and μ_v is the marginal of μ over the span of $v \in \mathbb{S}^{(d-1)}$. In what follows, we consider the equivalent formulation

$$SW_p^p(\mu,\nu) = \int_{\mathcal{R}_d} W_p^p(\mu_1^{(R(e_1))}, \nu_1^{(R(e_1))})d\rho. \tag{17}$$

Indeed, given $v, v' \in \mathbb{S}^{(d-1)}$, let us define $T_v = \{R \in \mathcal{R}_d \text{ s.t. } R(e_1) = v\}$ and, similarly, $T_{v'} = \{R \in \mathcal{R}_d \text{ s.t. } R(e_1) = v'\}$. Then, it holds

$$\rho(T_v) = \rho(T_{v'}).$$

Infact, let $S \in \mathcal{R}_d$ be such that $S(v) = v'$, we then define $\mathcal{S} : T_v \to T_{v'}$ as follows

$$\mathcal{S}(R) = S \circ R.$$

It is easy to see that $\mathcal{S}$ is a bijection. Moreover, since it is induced by a rotation, its determinant is equal to 1, this combined with the fact that ρ is a uniform distribution, allows us to retrieve (17). Thus, in the following, when we refer to the SWD, we will refer to the one defined in (17).

First, it is easy to see that Proposition 1 and its proof can be easily adapted to the $\mathbb{R}^d$ case. In particular, we have the following.

Corollary 1. *For every $p \geq 1$ and for every $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ there exists a family of d measures $\{\zeta_i\}_{i=1,\dots,d} \in \mathcal{P}(\mathbb{R}^d)$ such that*

$$W_p^p(\mu,\nu) = \sum_{i=1}^{d} \int_{\mathbb{R}^{d-1}} W_p^p(\mu_{|x_{-i}}, (\zeta_i)_{|x_{-i}}) d\mu_{-i} \tag{18}$$

and $\mu_{-i} = (\zeta_i)_{-i}$ for every $i = 1, \dots, d$, where $x_{-i} \in \mathbb{R}^{d-1}$ is the vector x without the i-th coordinate and μ_{-i} is the marginal of the measure μ on all the coordinates but the i-th one, i.e. on $(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_d)$.

Proof. It follows from the same argument used in the proof of Proposition 1.

Indeed, let γ be the optimal transportation plan between μ and ν.

We then define ζ_i as the projection of γ on the coordinates $(x_1, \dots, x_{i-1}, y_i, x_{i+1}, \dots, x_d)$ for every $i \in \{1, 2, \dots, d\}$.

Then, using again the characterization showed in [20], we infer (18) and thus the thesis. $\square$

Using the characterization of Corollary 1, we are able to extend the Radiant Formula to the higher dimensional case. The same goes for the bounds presented in Theorems 5 and 6.

Theorem 7. *Given any couple of measures $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$, it holds*

$$W_2^2(\mu,\nu) = d \int_{\mathcal{R}_d} \int_{\mathbb{R}^{d-1}} W_2^2(\mu_{|x_{-1}^{(R)}}, (\zeta_{-1}^{(R)})_{|x_1^{(R)}}) d\mu_{-1}^{(R)} d\rho. \tag{19}$$

Furthermore, for every couple of measures $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$, it holds $\frac{1}{d} W_2^2(\mu,\nu) \geq SW_2(\mu,\nu)$ and

$$| W_2^2(\mu,\nu) - SW_2^2(\mu,\nu) | \leq \int_{\mathcal{R}_d} \int_{\mathbb{R}} W_2^2((\zeta_{KR}^{(R)})_{|y_1^{(R)}}, \nu_{|y_1^{(R)}}) d\nu_{(x_2^{(R)},\dots,x_d^{(R)})} d\rho, \tag{20}$$

where $\zeta_{KR}^{(R)}$ is the marginal of the Knothe-Rosenblatt transportation plan over the coordinates $(y_1^{(R)}, x_2^{(R)} \dots, x_d^{(R)})$ and $\nu_{(x_2^{(R)},\dots,x_d^{(R)})}$ is the marginal of ν over $(x_2^{(R)}, \dots, x_d^{(R)})$.

Both bounds (19) and (20) are tight.
Moreover the same bounds can be inferred for RW_p^p.

Proof. For the sake of simplicity, we just show how to extend the proof of Theorem 2 to prove identity (19). Indeed, the proof of (20) follows by applying the same argument to the proof of Theorem 6. The same goes for the bounds on RW_p^p.

For every $R \in \mathcal{R}_d$, Corollary 1 allows us find a set of d measures $\{\zeta_i^{(R)}\}_{i=1,\dots,d}$ such that

$$W_2^2(\mu,\nu) = \sum_{i=1}^{d} \int_{\mathbb{R}^{d-1}} W_2^2(\mu_{|x_{-i}^{(R)}}, (\zeta_i^{(R)})_{|x_{-i}^{(R)}}) d\mu_{-i}^{(R)}.$$

By taking the average with respect to ρ of both sides of the equation we get

$$W_2^2(\mu,\nu) = \int_{\mathcal{R}_d} W_2^2(\mu,\nu) d\rho = \int_{\mathcal{R}_d} \sum_{i=1}^{d} \int_{\mathbb{R}^{d-1}} W_2^2(\mu_{|x_{-i}^{(R)}}, (\zeta_i^{(R)})_{|x_{-i}^{(R)}}) d\mu_{-i}^{(R)} d\rho.$$

Let $S_i \in \mathcal{R}_d$ be a rotation such that $S_i(e_1) = e_i$ holds. Then we have $e_i^{(R)} = R(S_i(e_1))$ and, therefore,

$$\begin{aligned} W_2^2(\mu,\nu) &= \int_{\mathcal{R}_d} \sum_{i=1}^{d} \int_{\mathbb{R}^{d-1}} W_2^2(\mu_{|x_{-1}^{(R\circ S_i)}}, (\zeta_{-1}^{(R\circ S_i)})_{|x_{-1}^{(R\circ S_i)}}) d\mu_{-1}^{(R\circ S_i)} d\rho \\ &= d \int_{\mathcal{R}_d} \int_{\mathbb{R}^{d-1}} W_2^2(\mu_{|x_{-1}^{(R)}}, (\zeta_{-1}^{(R)})_{|x_{-1}^{(R)}}) d\mu_{-1}^{(R)} d\rho, \end{aligned}$$

where the last equality comes from the fact that every S_i induces a bijective map from $\mathcal{R}_d$ to $\mathcal{R}_d$ whose determinant is equal to 1 and ρ is a uniform distribution.

Using convexity, we retrieve the bound $\frac{1}{d} W_2^2(\mu,\nu) \geq SW_2^2(\mu,\nu)$. □

Finally, we study a more general class of Sliced Wasserstein distances. Let $\mathcal{H}_k$ be the set of all the k-dimensional hyper-plans in $\mathbb{R}^d$. Given $\mu,\nu \in \mathcal{P}(\mathbb{R}^d)$, we define the k-Sliced Wasserstein Distance as

$$\left(SW_2^{(k)}(\mu,\nu)\right)^2 = \int_{\mathcal{H}_k} W_2^2(\mu_H,\nu_H) d\mathbf{H}, \tag{21}$$

where μ_H (ν_H) is the marginal of μ (ν, respectively) over the hyper-plan $H \in \mathcal{H}_k$ and $\mathbf{H}$ is the uniform probability distribution over the space of k-dimensional hyper-plans. Roughly speaking, the k-Sliced Wasserstein Distance is a variant of SW that projects the two measures over all the k dimensional sub-spaces of $\mathbb{R}^d$ rather than on all the 1 dimensional sub-spaces. This class of metrics is a natural extension of the $\mathcal{S}_k$ metrics introduced in [23].

Theorem 8. *Let μ and ν be two probability measures over $\mathbb{R}^d$ and let k be an integer such that $k < d$.*

Then, it holds

$$SW_2^{(k)}(\mu,\nu) \leq \sqrt{\frac{k}{d}} W_2(\mu,\nu). \tag{22}$$

Moreover, the bound is tight.

Proof. Let us consider H, a k-dimensional hyper-plan.

Moreover, let $\mathcal{T} : \mathcal{R}_d \to \mathcal{H}_k$ be defined as

$$\mathcal{T}(R) = R(H_0)$$

where $H_0 \in \mathcal{H}_k$ is the k-dimensional hyper-space generated by $\{e_1, \dots, e_k\}$, i.e.

$$H_0 = \{x \in \mathbb{R}^d, \ \textbf{s.t.}\ x_d = x_{d-1} = \cdots = x_{d+1-k} = 0\}.$$

Given $H, H' \in \mathcal{H}_d$, let $S \in \mathcal{R}_d$ be a rotation such that $S(H) = H'$. hen, we have that for every $R' \in \mathcal{T}^{-1}(H)$, the rotation $R'' := S \circ R' \in \mathcal{T}^{-1}(H')$, indeed $R''(H_0) = H' = S(H) = S(R'(H_0))$.In particular, S induces a bijective map between $\mathcal{T}^{-1}(H)$ and $\mathcal{T}^{-1}(H')$.

Again, since the maps is induced by a rotation and $\mathbf{H}$ is a uniform distribution, we have that $\rho(\mathcal{T}^{-1}(H)) = \rho(\mathcal{T}^{-1}(H'))$. n particular, given a function $f : \mathcal{R}_d \to \mathbb{R}$, it holds

$$\int_{\mathcal{R}_d} f(R) d\rho = \int_{\mathcal{H}_k} f(\mathcal{T}(H)) d\mathbf{H}. \tag{23}$$

Given $R \in \mathcal{R}_d$, there exists a set of d measures in $\mathbb{R}^d$, namely $\zeta_i^{(R)}$ such that

$$W_2^2(\mu,\nu) = \sum_{i=1}^d \int_{\mathbb{R}^{d-1}} W_2^2(\mu_{|x_{-i}^{(R)}}, (\zeta_i^{(R)})_{|x_{-i}^{(R)}}) d\mu_{-i}^{(R)}.$$

For every $i \in \{1, 2, \dots, d\}$, let $S_i \in \mathcal{R}_d$ be such that $S_i(e_1) = e_i$.By definition, given $e_i^{(R)}$, it holds true that $e_i^{(R)} = S_i(e_1)^{(R)} = R(S_i(e_1)) = e_1^{(R\circ S_i)}$.

Thus, we get

$$\begin{aligned} W_2^2(\mu,\nu) &= \int_{\mathcal{R}_d} \sum_{i=1}^d \int_{\mathbb{R}^{d-1}} W_2^2(\mu_{|x_{-i}^{(R)}}, (\zeta_i^{(R)})_{|x_{-i}^{(R)}}) d\mu_{-i}^{(R)} d\rho \\ &= d \int_{\mathcal{R}_d} \int_{\mathbb{R}^{d-1}} W_2^2(\mu_{|x_{-1}^{(R)}}, (\zeta_1^{(R)})_{|x_{-1}^{(R)}}) d\mu_{-1}^{(R)} d\rho, \end{aligned}$$

where the last equality follows from the fact that, for every $i \in \{1, 2, \dots, d\}$, the function $\mathcal{S}_i : R \to R \circ S_i$ is bijective and its Jacobian's determinant is equal to 1.

Given $R \in \mathcal{R}_d$ and $\{e_1^{(R)}, \dots, e_d^{(R)}\}$ its related base of $\mathbb{R}^d$, we have that

$$l_2^2(x,y) = c_{1:k}^{(R)}(x,y) + c_{k:d}^{(R)}(x,y),$$

where $c_{1:k}^{(R)}(x,y) := \sum_{i=1}^k (x_i^{(R)} - y_i^{(R)})^2$ and $c_{k:d}^{(R)}(x,y) := \sum_{i=k+1}^d (x_i^{(R)} - y_i^{(R)})^2$. Again, by Corollary 1, we can then find a couple of measures $\Psi \in \Pi(\mu_{>k}^{(R)}, \nu_{\leq k}^{(R)})$ and $\Theta \in \Pi(\mu_{\leq k}^{(R)}, \nu_{>k}^{(R)})$ such that

$$
\begin{aligned}
W_2^2(\mu,\nu) = & \int_{\mathbb{R}^{d-k}} W^2_{c_{1:k}^{(R)}}(\mu_{|x_{k+1}^{(R)},\dots,x_d^{(R)}}, \Psi_{|x_{k+1}^{(R)},\dots,x_d^{(R)}}) d\mu_{>k}^{(R)} \\
& + \int_{\mathbb{R}^k} W^2_{c_{k:d}^{(R)}}(\mu_{|x_1^{(R)},\dots,x_k^{(R)}}, \Theta_{|x_1^{(R)},\dots,x_k^{(R)}}) d\mu_{\leq k}^{(R)},
\end{aligned}
$$

where $\mu_{\leq k}^{(R)}$ and $\mu_{>k}^{(R)}$ are the marginals of μ over the first k coordinates and the last $(d-k)$ coordinates, respectively.

Since, for every $R \in \mathcal{R}_d$ and for every $(x_{k+1}^{(R)}, \dots, x_d^{(R)})$, we have that $c_{1:k}$ is also separable, we can further split the Wasserstein distance $W^2_{c_{1:k}^{(R)}}(\mu_{|x_{k+1}^{(R)},\dots,x_d^{(R)}}, \Psi_{|x_{k+1}^{(R)},\dots,x_d^{(R)}})$ and obtain

$$
\begin{aligned}
& \int_{\mathbb{R}^{d-k}} W^2_{c_{1:k}^{(R)}}(\mu_{|x_{k+1}^{(R)},\dots,x_d^{(R)}}, \Psi_{|x_{k+1}^{(R)},\dots,x_d^{(R)}}) d\mu_{>k}^{(R)} \\
& \qquad = \sum_{i=1}^k \int_{\mathbb{R}^{d-1}} W_2^2(\mu_{|x_{-i}^{(R)}}, (\zeta_i^{(R)})_{|x_{-i}^{(R)}}) d\mu_{-i}^{(R)}.
\end{aligned}
$$

Thus, by convexity, we infer that

$$
\begin{aligned}
\sum_{i=1}^k \int_{\mathbb{R}^{d-1}} W_2^2(\mu_{|x_{-i}^{(R)}}, (\zeta_i^{(R)})_{|x_{-i}^{(R)}}) d\mu_{-i}^{(R)} &= \int_{\mathbb{R}^{d-k}} W_2^2(\mu_{|x_{-i}^{(R)}}, (\zeta_i^{(R)})_{|x_{\geq i}^{(R)}}) d\mu_{\geq i}^{(R)} \\
&\geq W_2^2(\mu_H, \nu_H),
\end{aligned}
$$

where $H = R(H_0)$. If we take the average over all the possible rotations, we get

$$
\int_{\mathcal{R}_d} \sum_{i=1}^k \int_{\mathbb{R}^{d-1}} W_2^2(\mu_{|x_{-i}^{(R)}}, (\zeta_i^{(R)})_{|x_{-i}^{(R)}}) d\mu_{-i}^{(R)} d\rho \geq \int_{\mathcal{R}_d} W_2^2(\mu_{\mathcal{T}(R)}, \nu_{\mathcal{T}(R)}) d\rho.
$$

Using identity (23), we get

$$
\int_{\mathcal{R}_d} \sum_{i=1}^k \int_{\mathbb{R}^{d-1}} W_2^2(\mu_{|x_{-i}^{(R)}}, (\zeta_i^{(R)})_{|x_{-i}^{(R)}}) d\mu_{-i}^{(R)} d\rho \geq \int_{\mathcal{H}_k} W_2^2(\mu_H, \nu_H) d\mathbf{H}.
$$

By the same argument used before, we can simplify the sum

$$
\begin{aligned}
k \int_{\mathcal{R}_d} \int_{\mathbb{R}^{d-1}} W_2^2(\mu_{|x_{-1}^{(R)}}, (\zeta_1^{(R)})_{|x_{-1}^{(R)}}) d\mu_{-1}^{(R)} d\rho &\geq \int_{\mathcal{H}_k} W_2^2(\mu_H, \nu_H) d\mathbf{H} \\
&= (SW_2^{(k)}(\mu,\nu))^2.
\end{aligned}
$$

Finally, we have that

$$W_2^2(\mu,\nu) = d\int_{\mathcal{R}_d}\int_{\mathbb{R}^{d-1}} W_2^2(\mu_{|x_{-1}^{(R)}}, (\zeta_1^{(R)})_{|x_{-1}^{(R)}})d\mu_{-1}^{(R)}d\rho \geq \frac{d}{k}(SW_2^{(k)}(\mu,\nu))^2,$$

which allows to conclude(22).

Lastly, the tightness property follows by considering two Dirac's deltas. □

3. Conclusion

In this note, we proposed an alternative representation of the Wasserstein distance and showcased how this formula has significant connections with a classic Sliced Wasserstein Distance and used them to prove bounds on Sliced Wasserstein Distances in term of the classic Wasserstein Distances. Furthermore, we used the Knothe-Rosenblatt heuristic to prove a bound over the absolute error.

The field of Sliced Wasserstein Distance keeps flourishing and producing different alternative versions [23–25]. We believe the Radiant Formula is a useful item to study the relationships between old Sliced-like distances and new ones.

Author contributions The authors confirm contribution to the paper as follows: study conception, design, and writing: Gennaro Auricchio. All authors reviewed the results and approved the final version of the manuscript.

Funding This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Data availability Data availability is not applicable to this article as no new data were created or analysed in this study.

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no Conflict of interest.

References

[1] Kantorovich, L.V.: On the translocation of masses. J. Math. Sci. **133**(4), 1381–1382 (2006)

[2] Solomon, J., De Goes, F., Peyré, G., Cuturi, M., Butscher, A., Nguyen, A., Du, T., Guibas, L.: Convolutional wasserstein distances: Efficient optimal transportation on geometric domains. ACM Trans. Gr. (ToG) **34**(4), 1–11 (2015)

[3] Gangbo, W., McCann, R.J.: Shape recognition via wasserstein distance. Q. Appl. Math. 705–737 (2000)

[4] Arjovsky, M., Chintala, S., Bottou, L.: Wasserstein generative adversarial networks. In: Proceedings of the 34th International Conference on Machine Learning, pp. 214–223 (2017)

[5] Yu, J., Lin, Z., Yang, J., Shen, X., Lu, X., Huang, T.S.: Generative image inpainting with contextual attention. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5505–5514 (2018)

[6] Pan, Z., Yu, W., Wang, B., Xie, H., Sheng, V.S., Lei, J., Kwong, S.: Loss functions of generative adversarial networks (gans): opportunities and challenges. IEEE Trans. Emerg. Top. Comput. Intell. **4**(4), 500–522 (2020)

[7] Chakraborty, S., Paul, D., Das, S.: Hierarchical clustering with optimal transport. Stat. & Probab. Lett. **163**, 108781 (2020)

[8] Auricchio, G., Bassetti, F., Gualandi, S., Veneroni, M.: Computing wasserstein barycenters via linear programming. In: Integration of Constraint Programming, Artificial Intelligence, and Operations Research: 16th International Conference, CPAIOR 2019, Thessaloniki, Greece, June 4–7, 2019, Proceedings 16, pp. 355–363. Springer (2019)

[9] Ling, H., Okada, K.: An efficient earth mover's distance algorithm for robust histogram comparison. IEEE trans. pattern anal. mach. intell. **29**(5), 840–853 (2007)

[10] Auricchio, G., Bassetti, F., Gualandi, S., Veneroni, M.: Computing kantorovich-wasserstein distances on d-dimensional histograms using $(d+1)$-partite graphs. Adv. Neural Inf. Process. Syst. **31** (2018)

[11] Altschuler, J., Niles-Weed, J., Rigollet, P.: Near-linear time approximation algorithms for optimal transport via sinkhorn iteration. Adv. Neural Inf. Process. Syst. **30** (2017)

[12] Schmitzer, B.: Stabilized sparse scaling algorithms for entropy regularized transport problems. SIAM J. Sci. Comput. **41**(3), 1443–1481 (2019)

[13] Cuturi, M.: Sinkhorn distances: lightspeed computation of optimal transport. Adv. Neural Inf. Process. Syst. **26** (2013)

[14] Auricchio, G., Codegoni, A., Gualandi, S., Toscani, G., Veneroni, M.: The equivalence of fourier-based and wasserstein metrics on imaging problems. Rend. Lincei **31**(3), 627–649 (2020)

[15] Auricchio, G., Codegoni, A., Gualandi, S., Zambon, L.: The fourier discrepancy function. Commun. Math. Sci. **21**(3), 627–639 (2023)

[16] Kolouri, S., Zou, Y., Rohde, G.K.: Sliced wasserstein kernels for probability distributions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5258–5267 (2016)

[17] Bonneel, N., Coeurjolly, D.: Spot: sliced partial optimal transport. ACM Trans. Gr. (TOG) **38**(4), 1–13 (2019)

[18] Knothe, H.: Contributions to the theory of convex bodies. Michigan Math. J. **4**(1), 39–52 (1957)
[19] Rosenblatt, M.: Remarks on a multivariate transformation. Ann. Math. Statist. **23**(3), 470–472 (1952). https://doi.org/10.1214/aoms/1177729394
[20] Auricchio, G.: On the pythagorean structure of the optimal transport for separable cost functions. Rend. Lincei Mat. Appl. **34**(4), 745–771 (2024)
[21] Villani, C.: Optimal Transport: Old and New, vol. 338. Springer, New York (2008)
[22] Knothe, H.: Contributions to the theory of convex bodies. Michigan Math. J. **4**(1), 39–52 (1957). https://doi.org/10.1307/mmj/1028990175
[23] Paty, F.-P., Cuturi, M.: Subspace robust wasserstein distances. In: International Conference on Machine Learning, pp. 5072–5081. PMLR (2019)
[24] Deshpande, I., Hu, Y.-T., Sun, R., Pyrros, A., Siddiqui, N., Koyejo, S., Zhao, Z., Forsyth, D., Schwing, A.G.: Max-sliced wasserstein distance and its use for gans. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 10648–10656 (2019)
[25] Nguyen, K., Ho, N., Pham, T., Bui, H.: Distributional sliced-wasserstein and applications to generative modeling. arXiv preprint arXiv:2002.07367 (2020)

Gennaro Auricchio
Computer Science Department
University of Bath
Bath 100190
England, UK
e-mail: ga647@bath.ac.uk

Received: July 21, 2023
Accepted: March 10, 2024