Noname manuscript No.
(will be inserted by the editor)

On the equivalence between Fourier-based and Wasserstein metrics

G. Auricchio, A. Codegoni, S. Gualandi,
G. Toscani, M. Veneroni

Received: date / Accepted: date

Abstract We investigate properties of some extensions of a class of Fourier-based probability metrics, originally introduced to study convergence to equilibrium for the solution to the spatially homogeneous Boltzmann equation. At difference with the original one, the new Fourier-based metrics are well-defined also for probability distributions with different centers of mass, and for discrete probability measures supported over a regular grid. Among other properties, it is shown that, in the discrete setting, these new Fourier-based metrics are equivalent either to the Euclidean-Wasserstein distance W_2, or to the Kantorovich-Wasserstein distance W_1, with explicit constants of equivalence.

Keywords Fourier-based Metrics · Wasserstein Distance · Fourier Transform

Mathematics Subject Classification (2010) 60A10 · 60E15 · 42A38

1 Introduction

In computational applied mathematics, numerical methods based on Wasserstein distances achieved a leading role over the last years. Examples include the comparison of histograms in higher dimensions [6, 9, 22], image retrieval [21], image registration [4, 11], or, more recently, the computations of barycenters among images [7, 15]. Surprisingly, the possibility to identify the cost function in a Wasserstein distance, together with the possibility of representing images as histograms, led to the definition of classifiers able to mimic the human eye [16, 21, 24].

More recently, metrics which are able to compare at best probability distributions were introduced and studied in connection with machine learning,

Università degli Studi di Pavia, Dipartimento di Matematica ”F.Casorati”
E-mail: stefano.gualandi@unipv.it, marco.veneroni@unipv.it, giuseppe.toscani@unipv.it, andrea.codegoni01@universitadipavia.it, gennaro.auricchio01@universitadipavia.it

where testing the efficiency of new classes of loss functions for neural networks training has become increasingly important. In this area, the Wasserstein distance often turns out to be the appropriate tool [1, 5, 18]. Its main drawback, though, is that it suffers from high computational complexity. For this reason, attempts to use other metrics, which require a lower computational cost while maintaining a good approximation, have been object of recent research [28]. There, the theory of approximation in the space of wavelets was the main mathematical tool.

Following the line of thought of [28], we consider here an alternative to the approximation in terms of wavelets, which is furnished by metrics based on the Fourier transform. In terms of computational complexity, the price to pay for a dimension $N \gg 1$ of the data changes from a time $O(N)$ to the time $O(N \log N)$ required to evaluate the fast Fourier transform.

While this represents a worsening, with respect to the use of wavelets, in terms of computational complexity, there is an effective improvement with respect to the computational complexity required to evaluate Wasserstein-type metrics, which is of the order $O(N^3 \log N)$. Furthermore, from the point of view of the important questions related to the comparison of these metrics with Wasserstein metrics in problems motivated by real applications, we prove in this paper that in the case of probability measures supported on a bounded domain, one has a precise and explicit evaluation of the constants of equivalence among these Fourier-based metrics and the Wassertein ones, a result which is not present in [28].

The Fourier-based metrics considered in this paper were introduced in [19], in connection with the study of the trend to equilibrium for solutions of the spatially homogeneous Boltzmann equation for Maxwell molecules. Since then, many applications of these metrics have followed in both kinetic theory and probability [10, 12–14, 20, 25, 30]. All these problems deal with functions supported on the whole space $\mathbb{R}^d$, with $d \geq 1$, that exhibit a suitable decay at infinity which guarantees the existence of a suitable number of moments.

Given two probability measures $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$, $d \geq 1$, and a real parameter $s > 0$, the Fourier-based metrics d_s considered in [19] are given by

$$d_s(\mu, \nu) := \sup_{\boldsymbol{k} \in \mathbb{R}^d \setminus \{0\}} \frac{|\widehat{\mu}(\boldsymbol{k}) - \widehat{\nu}(\boldsymbol{k})|}{|\boldsymbol{k}|^s}, \tag{1.1}$$

where $\widehat{\mu}$ and $\widehat{\nu}$ are the Fourier transforms of the measures μ and ν, respectively. As usual, given a probability measure $\mu \in \mathcal{P}(\mathbb{R}^d)$, the Fourier transform of μ is defined by

$$\hat{\mu}(\boldsymbol{k}) := \int_{\mathbb{R}^d} e^{-i\boldsymbol{k}\cdot\boldsymbol{x}} d\mu(\boldsymbol{x}).$$

These metrics, for $s \geq 1$, are well-defined under the further assumption of boundedness and equality of some moments of the probability measures. Indeed, a necessary condition for d_s to be finite, is that moments up to $[s]$ (the integer part of s) are equal for both measures [19].

In dimension $d = 1$, similar metrics were introduced a few years later by Baringhaus and Grübel in connection with the characterization of convex combinations of random variables [8]. Given two probability measures $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$, $d \geq 1$, and two real parameters $s > 0$ and $p \geq 1$, the multi-dimensional version of these Fourier-based metrics reads

$$D_{s,p}(\mu, \nu) := \left(\int_{\mathbb{R}^d} \frac{|\widehat{\mu}(\boldsymbol{k}) - \widehat{\nu}(\boldsymbol{k})|^p}{|\boldsymbol{k}|^{(ps+d)}} d\boldsymbol{k} \right)^{1/p} . \tag{1.2}$$

The metrics defined by (1.1) and (1.2) belong to the set of ideal metrics [32], and have been shown to be equivalent to other common probability distances [19,30], including the Wasserstein distance $W_2(\mu, \nu)$ [14], given by

$$W_2(\mu, \nu) := \inf_{\pi \in \Pi(\mu,\nu)} \left\{ \int_{\mathbb{R}^d \times \mathbb{R}^d} |\boldsymbol{x} - \boldsymbol{y}|^2 \, d\pi(\boldsymbol{x}, \boldsymbol{y}) \right\}^{1/2} , \tag{1.3}$$

where the infimum is taken on the set $\Pi(\mu, \nu)$ of all probability measures on $\mathbb{R}^d \times \mathbb{R}^d$ with marginal densities μ and ν. However, in dimension $d > 1$ the constants of equivalence are not explicit [14], so that it is difficult to establish a comparison between these metrics' efficacy in applications.

An unpleasant aspect related to the application of the previous Fourier-based distances is related to its finiteness, that requires, for high values of s, a sufficiently high number of equal moments for the underlying probability measures. In the context of kinetic equations of Boltzmann type, where conservation of momentum and energy of the solution is a consequence of the microscopic conservation laws of binary interactions among particles, this requirement on d_s, with $2 < s < 3$, is clearly not restrictive. However, in order to apply the Fourier-based metrics outside of the context of kinetic equations, this requirement appears unnatural. To clarify this point, let us consider the case in which we want to compare the distance between two images. If we take two grey scale images and model them as probability distributions, there is no reason why these distributions possess the same expected value. The simplest example is furnished by two images consisting of a black dot, each one centered in a different point of the region, that can be modeled as two Dirac delta functions centered in two different points.

In this paper we improve the existing results concerning the evaluation of the constants in the equivalence relations between the Fourier-based metrics and the Wasserstein one, in a relevant setting with respect to applications. This equivalence is related to the comparison of two discrete measures and it is based on the properties of the Fourier transform in the discrete setting. To this extent, we consider a new version of these metrics, the *periodic Fourier-based metrics*, that play the role of the metrics (1.1) and (1.2) in the discrete setting. With our results, we show that the new family of Fourier-based metrics represents a fruitful alternative to the Wasserstein metrics, both from the theoretical and the computational points of view.

To weaken the restriction about moments, we further consider a variant of the Fourier metric d_2 that remains well-defined even for probability measures with different mean values.

The content of this paper is as follows. In Section 2 we introduce the notations and the basic concepts of measure theory and optimal transport. Furthermore, we define the Fourier-based metrics, we recall their main properties, and we introduce our extension. Then, in view of applications, in Section 3, we consider a discrete setting and we define and study the properties of the new family of periodic Fourier-based metrics, highlighting their explicit equivalence with the Wasserstein distance in various cases. Section 4 presents numerical results obtained comparing our implementation of the *periodic Fourier-based metrics* with the Wasserstein metrics as implemented in the POT library [17]. The concluding remarks are contained in Section 5.

2 An extension of Fourier-based metrics

In what follows, we briefly review some basic notions of optimal transport, together with the definition and some properties of Wassertein and Fourier-based metrics. The final goal is to extend the definition of the metrics (1.1) and (1.2) for the particular case $s = 2$, which allows for a direct and fruitful comparison between the Fourier-based metrics and the Wasserstein metric W_2 defined in (1.3). In what follows, we only present the notions that are necessary for our purpose. For a deeper insight on optimal transport, we refer the reader to [2, 3, 26, 31]. Likewise, we address the interested reader to [14] for an exhaustive review of the properties of the Fourier-based metrics and their connections with other metrics used in probability theory.

We work on the Euclidean space $\mathbb{R}^d$, endowed with the Borel $\sigma-$algebra $\mathcal{B}(\mathbb{R}^d)$. We use bold letters to denote vectors of $\mathbb{R}^d$. If $\boldsymbol{x} \in \mathbb{R}^d$, then x_i denotes its i-th coordinate. Given $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$, $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \sum_{i=1}^n x_i y_i$ is their scalar product and $|\boldsymbol{x}| = \langle \boldsymbol{x}, \boldsymbol{x} \rangle^{1/2}$ is the Euclidean norm (or modulus) of $\boldsymbol{x}$.

The set of probability measures on $\mathbb{R}^d$ is denoted by $\mathcal{P}(\mathbb{R}^d)$. For all $m \in \mathbb{N}$ we denote by $\mathcal{P}_m(\mathbb{R}^d)$ the set of probability measures with finite moments up to order m

$$\mathcal{P}_m(\mathbb{R}^d) := \left\{ \mu \in \mathcal{P}(\mathbb{R}^d) : \int_{\mathbb{R}^d} \boldsymbol{x}^\beta \, d\mu(\boldsymbol{x}) < +\infty, \ \forall \beta \in \mathbb{N}^d, \ |\beta| \le m \right\}.$$

Given $\mu \in \mathcal{P}(\mathbb{R}^d)$ and a Borel map $f : \mathbb{R}^d \to \mathbb{R}^d$, then the image measure (or push-forward) of μ by f is $f_\# \mu \in \mathcal{P}(\mathbb{R}^d)$, given by $f_\# \mu(A) = \mu(f^{-1}(A))$ for all $A \in \mathcal{B}(\mathbb{R}^d)$. Equivalently, for every continuous compactly supported function ϕ on $\mathbb{R}^d$, it holds

$$\int_{\mathbb{R}^d} \phi(y) \, d(f_\# \mu)(y) = \int_{\mathbb{R}^d} \phi(f(x)) \, d\mu(x).$$

Our first goal is to define the Fourier-based metrics d_s, in the range $1 < s \le 2$, on $\mathcal{P}(\mathbb{R}^d)$.

Definition 1 Given $\mu \in \mathcal{P}_1(\mathbb{R}^d)$, we say that

$$\boldsymbol{m_\mu} = \int_{\mathbb{R}^d} \boldsymbol{x} \, d\mu(\boldsymbol{x})$$

is the *center* of μ.

The center of a measure μ can be moved by resorting to a translation. Given $\mu \in \mathcal{P}_1(\mathbb{R}^d)$ and $\boldsymbol{\tau} \in \mathbb{R}^d$, we define the translated measure $\mu_{\boldsymbol{\tau}} \in \mathcal{P}_1(\mathbb{R}^d)$ by

$$\mu_{\boldsymbol{\tau}} = S^{\boldsymbol{\tau}}_{\#}\mu, \quad \text{where} \quad S^{\boldsymbol{\tau}}(\boldsymbol{x}) = \boldsymbol{x} + \boldsymbol{\tau}.$$

Lemma 1 *Given $\mu, \nu \in \mathcal{P}_1(\mathbb{R}^d)$, there exists a unique vector $\boldsymbol{\tau} \in \mathbb{R}^d$ such that*

$$\boldsymbol{m}_{\boldsymbol{\mu}} = \boldsymbol{m}_{\boldsymbol{\nu}_{\boldsymbol{\tau}}}.$$

Proof Let $\boldsymbol{\tau} = \boldsymbol{m}_{\boldsymbol{\mu}} - \boldsymbol{m}_{\boldsymbol{\nu}}$, then

$$\boldsymbol{m}_{\boldsymbol{\nu}_{\boldsymbol{\tau}}} = \int_{\mathbb{R}^d} \boldsymbol{x} d\nu_{\boldsymbol{\tau}}(\boldsymbol{x}) = \int_{\mathbb{R}^d} (\boldsymbol{x} + \boldsymbol{\tau}) d\nu(\boldsymbol{x}) = \boldsymbol{m}_{\boldsymbol{\nu}} + \boldsymbol{\tau} = \boldsymbol{m}_{\boldsymbol{\mu}}.$$

■

Let us recall now the definition of transport plan, and the consequent definition of Wasserstein Distance.

Definition 2 (Transport plan) Given two probability measures $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$, a vector $\pi \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ is called a transport plan between μ and ν if its marginals coincide with μ, ν, that is

$$\pi(A \times \mathbb{R}^d) = \mu(A) \qquad \forall A \in \mathcal{B}(\mathbb{R}^d), \tag{2.4}$$
$$\pi(\mathbb{R}^d \times B) = \nu(B) \qquad \forall B \in \mathcal{B}(\mathbb{R}^d). \tag{2.5}$$

We denote by $\Pi(\mu, \nu)$ the set of all transport plans between μ and ν.

Definition 3 (Wasserstein distance) Given $p \in \mathbb{N}$ and $\mu, \nu \in \mathcal{P}_p(\mathbb{R}^d)$, the Wasserstein distance of order p between μ and ν is defined as

$$W_p(\mu, \nu) := \inf_{\pi \in \Pi(\mu,\nu)} \left\{ \int_{\mathbb{R}^d \times \mathbb{R}^d} |\boldsymbol{x} - \boldsymbol{y}|^p \, d\pi(\boldsymbol{x}, \boldsymbol{y}) \right\}^{1/p}, \tag{2.6}$$

where $|\cdot|$ is a norm defined in $\mathbb{R}^d$.

In this paper, we consider only the Euclidean norm, and we focus on Wasserstein distances with exponents $p = 1$ and $p = 2$, namely

$$W_1(\mu, \nu) := \inf_{\pi \in \Pi(\mu,\nu)} \left\{ \int_{\mathbb{R}^d \times \mathbb{R}^d} |\boldsymbol{x} - \boldsymbol{y}| \, d\pi(\boldsymbol{x}, \boldsymbol{y}) \right\}, \tag{2.7}$$
$$W_2(\mu, \nu) := \inf_{\pi \in \Pi(\mu,\nu)} \left\{ \int_{\mathbb{R}^d \times \mathbb{R}^d} |\boldsymbol{x} - \boldsymbol{y}|^2 \, d\pi(\boldsymbol{x}, \boldsymbol{y}) \right\}^{1/2}. \tag{2.8}$$

The W_2 metric satisfies an explicit translation property (Remark 2.19, [16]). We give below a short proof of this property.

Lemma 2 *Let $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$, with centers $\boldsymbol{m_\mu}$ and $\boldsymbol{m_\nu}$, respectively. For any given pair of vectors $\boldsymbol{v}, \boldsymbol{w} \in \mathbb{R}^d$ we have*

$$W_2(\mu_{\boldsymbol{v}}, \nu_{\boldsymbol{w}})^2 = W_2(\mu, \nu)^2 + |\boldsymbol{v} - \boldsymbol{w}|^2 + 2\langle \boldsymbol{v} - \boldsymbol{w}, \boldsymbol{m_\mu} - \boldsymbol{m_\nu} \rangle. \tag{2.9}$$

In addition, if we choose $\boldsymbol{v} = -\boldsymbol{m_\mu}$ and $\boldsymbol{w} = -\boldsymbol{m_\nu}$ it holds

$$W_2(\mu_{-\boldsymbol{m_\mu}}, \nu_{-\boldsymbol{m_\nu}})^2 = W_2(\mu, \nu)^2 - |\boldsymbol{m_\mu} - \boldsymbol{m_\nu}|^2. \tag{2.10}$$

Proof Given a transport plan $\pi \in \Pi(\mu, \nu)$, we consider the transport plan

$$\tilde{\pi} := (S^{\boldsymbol{v}}, S^{\boldsymbol{w}})_{\#}\pi,$$

where $S^{\boldsymbol{v}}(\boldsymbol{x}) = \boldsymbol{x} + \boldsymbol{v}$, $S^{\boldsymbol{w}}(\boldsymbol{y}) = \boldsymbol{y} + \boldsymbol{w}$. $\tilde{\pi}$ is a transport plan between the translated measures $\mu_{\boldsymbol{v}}$ and $\nu_{\boldsymbol{w}}$. Then, by definition of push-forward, we get

$$\begin{aligned}
&\int_{\mathbb{R}^d \times \mathbb{R}^d} |\boldsymbol{x} - \boldsymbol{y}|^2 d\tilde{\pi}(\boldsymbol{x}, \boldsymbol{y}) \\
&\quad = \int_{\mathbb{R}^d \times \mathbb{R}^d} |(\boldsymbol{x} + \boldsymbol{v}) - (\boldsymbol{y} + \boldsymbol{w})|^2 d\pi(\boldsymbol{x}, \boldsymbol{y}) \\
&\quad = \int_{\mathbb{R}^d \times \mathbb{R}^d} (|\boldsymbol{x} - \boldsymbol{y}|^2 + |\boldsymbol{v} - \boldsymbol{w}|^2 + 2\langle \boldsymbol{x} - \boldsymbol{y}, \boldsymbol{v} - \boldsymbol{w} \rangle) d\pi(\boldsymbol{x}, \boldsymbol{y}) \\
&\quad = \int_{\mathbb{R}^d \times \mathbb{R}^d} |\boldsymbol{x} - \boldsymbol{y}|^2 d\pi(\boldsymbol{x}, \boldsymbol{y}) + |\boldsymbol{v} - \boldsymbol{w}|^2 + 2\langle \boldsymbol{m_\mu} - \boldsymbol{m_\nu}, \boldsymbol{v} - \boldsymbol{w} \rangle.
\end{aligned}$$

If π is an optimal transport plan between μ and ν, we have

$$\begin{aligned}
W_2(\mu_{\boldsymbol{v}}, \nu_{\boldsymbol{w}})^2 &\leq \int_{\mathbb{R}^d \times \mathbb{R}^d} |\boldsymbol{x} - \boldsymbol{y}|^2 d\tilde{\pi}(\boldsymbol{x}, \boldsymbol{y}) \\
&= W_2(\mu, \nu)^2 + |\boldsymbol{v} - \boldsymbol{w}|^2 + 2\langle \boldsymbol{v} - \boldsymbol{w}, \boldsymbol{m_\mu} - \boldsymbol{m_\nu} \rangle.
\end{aligned}$$

By repeating the previous argument with an optimal transport plan between $\mu_{\boldsymbol{v}}$, $\nu_{\boldsymbol{w}}$, we find

$$\begin{aligned}
W_2(\mu_{\boldsymbol{v}}, \nu_{\boldsymbol{w}})^2 &= \int_{\mathbb{R}^d \times \mathbb{R}^d} |\boldsymbol{x} - \boldsymbol{y}|^2 d\pi(\boldsymbol{x}, \boldsymbol{y}) + |\boldsymbol{v} - \boldsymbol{w}|^2 + 2\langle \boldsymbol{v} - \boldsymbol{w}, \boldsymbol{m_\mu} - \boldsymbol{m_\nu} \rangle \\
&\geq W_2(\mu, \nu)^2 + |\boldsymbol{v} - \boldsymbol{w}|^2 + 2\langle \boldsymbol{v} - \boldsymbol{w}, \boldsymbol{m_\mu} - \boldsymbol{m_\nu} \rangle.
\end{aligned}$$

Hence, we can conclude

$$W_2(\mu_{\boldsymbol{v}}, \nu_{\boldsymbol{w}})^2 = W_2(\mu, \nu)^2 + |\boldsymbol{v} - \boldsymbol{w}|^2 + 2\langle \boldsymbol{v} - \boldsymbol{w}, \boldsymbol{m_\mu} - \boldsymbol{m_\nu} \rangle.$$

■

The idea of using translation operators to compute the distance of probability measures with different centers can be used to properly modify the Fourier-based metrics d_s and $D_{s,p}$ defined in (1.1) and (1.2). Indeed, as briefly discussed in the introduction, the case $s \geq 1$ requires the probability measures to satisfy the further condition given below [19].

Proposition 1 (Proposition 2.6, [14]) *Let $\lfloor s \rfloor$ denote the integer part of $s \in \mathbb{R}$, and assume that the densities $\mu, \nu \in \mathcal{P}_s(\mathbb{R}^d)$ possess equal moments up to $\lfloor s \rfloor$ if $s \notin \mathbb{N}$, or equal moments up to $s-1$ if $s \in \mathbb{N}$. Then the Fourier-based distance $d_s(\mu, \nu)$ is well-defined. In particular, $d_2(\mu, \nu)$ is well-defined for two densities with the same center.*

The interest in the d_2 metric is related to its equivalence to the Euclidean Wasserstein distance W_2. A detailed proof in dimension $d \geq 1$ can be found in the review paper [14].

Theorem 1 (Proposition 2.12 and Corollary 2.17, [14]) *For any given pair of probability densities $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ such that $\boldsymbol{m_\mu} = \boldsymbol{m_\nu}$, the d_2 metric is equivalent to the Euclidean Wasserstein distance W_2, that is, there exist two positive bounded constants $c < C$ such that*

$$cW_2(\mu, \nu) \leq d_2(\mu, \nu) \leq CW_2(\mu, \nu). \tag{2.11}$$

The proof in [14] does not provide in general the explicit expression of the two constants c and C. The value of these constants is quite involved, and it is strongly dependent on higher moments of the densities.

The equivalence result of Theorem 1 can easily be extended to cover the case of probability measures with different centers of mass. To this aim it is necessary, in analogy with the property of Wasserstein distance W_2 stated in Lemma 2, to modify the Fourier-based metrics d_2 and $D_{2,p}$ in such a way to allow for probability measures with different centers of mass. We start by considering the case of the metric d_2.

Definition 4 (Translated Fourier-based Metric) We define the function $\mathcal{D}_2 : \mathcal{P}_2(\mathbb{R}^d) \times \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$ as:

$$\mathcal{D}_2(\mu, \nu) := \sqrt{d_2(\mu, \nu_{\boldsymbol{m}_\mu - \boldsymbol{m}_\nu})^2 + |\boldsymbol{m_\mu} - \boldsymbol{m_\nu}|^2}. \tag{2.12}$$

Owing to Remark 1 and Proposition 1, $\mathcal{D}_2(\mu, \nu)$ is well-defined for each pair of probability measures in $\mathcal{P}_2(\mathbb{R}^d)$, independently of their centers. Note that $\nu_{\boldsymbol{m_\mu} - \boldsymbol{m_\nu}}$, which is the translation of ν by $\boldsymbol{m_\mu} - \boldsymbol{m_\nu}$, has the same center as μ. One could give an equivalent definition of $\mathcal{D}_2$ by translating μ, instead of ν, or by translating both centers to $\mathbf{0}$.

Lemma 3 *Given $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ and $\boldsymbol{v}, \boldsymbol{w} \in \mathbb{R}^d$, then*

$$|\widehat{\mu_{\boldsymbol{v}}}(\boldsymbol{k}) - \widehat{\nu_{\boldsymbol{w}}}(\boldsymbol{k})| = |\hat{\mu}(\boldsymbol{k}) - \widehat{\nu_{\boldsymbol{w}-\boldsymbol{v}}}(\boldsymbol{k})| = |\widehat{\mu_{\boldsymbol{v}-\boldsymbol{w}}}(\boldsymbol{k}) - \hat{\nu}(\boldsymbol{k})|.$$

Therefore

$$d_2(\mu_{\boldsymbol{v}}, \nu_{\boldsymbol{w}}) = d_2(\mu, \nu_{\boldsymbol{w}-\boldsymbol{v}}) = d_2(\mu_{\boldsymbol{v}-\boldsymbol{w}}, \nu).$$

In particular, the function $(\mu, \nu) \to d_2(\mu, \nu_{\boldsymbol{m_\mu} - \boldsymbol{m_\nu}})$ is symmetric.

Proof By the translation property of the Fourier Transform, for all $\boldsymbol{v} \in \mathbb{R}^d$ we have the identity

$$\widehat{\mu_{\boldsymbol{v}}}(\boldsymbol{k}) = e^{-i\boldsymbol{v}\cdot\boldsymbol{k}}\hat{\mu}(\boldsymbol{k}).$$

Therefore

$$\begin{aligned}|e^{-i\boldsymbol{v}\cdot\boldsymbol{k}}\hat{\mu}(\boldsymbol{k}) - e^{-i\boldsymbol{w}\cdot\boldsymbol{k}}\hat{\nu}(\boldsymbol{k})| &= |e^{-i\boldsymbol{w}\cdot\boldsymbol{k}}(e^{-i(\boldsymbol{v}-\boldsymbol{w})\cdot\boldsymbol{k}}\hat{\mu}(\boldsymbol{k}) - \hat{\nu}(\boldsymbol{k}))| \\ &= |e^{-i(\boldsymbol{v}-\boldsymbol{w})\cdot\boldsymbol{k}}\hat{\mu}(\boldsymbol{k}) - \hat{\nu}(\boldsymbol{k})|.\end{aligned}$$

This shows that

$$\sup_{\boldsymbol{k}\in\mathbb{R}^d\setminus\{0\}} \frac{|e^{-i\boldsymbol{v}\cdot\boldsymbol{k}}\hat{\mu}(\boldsymbol{k}) - e^{-i\boldsymbol{w}\cdot\boldsymbol{k}}\hat{\nu}(\boldsymbol{k})|}{|\boldsymbol{k}|^2} = \sup_{\boldsymbol{k}\in\mathbb{R}^d\setminus\{0\}} \frac{|e^{-i(\boldsymbol{v}-\boldsymbol{w})\cdot\boldsymbol{k}}\hat{\mu}(\boldsymbol{k}) - \hat{\nu}(\boldsymbol{k})|}{|\boldsymbol{k}|^2}.$$

■

Lemma 3 implies the following theorem.

Theorem 2 *The function $\mathcal{D}_2$ defined in* (2.12) *is a distance over $\mathcal{P}_2(\mathbb{R}^d)$.*

Proof Clearly $\mathcal{D}_2(\mu,\nu) \geq 0, \forall \mu,\nu \in \mathcal{P}_2(\mathbb{R}^d)$, and $\mathcal{D}_2(\mu,\nu) = 0$ if and only if $\mu = \nu$. Symmetry follows from Lemma 3. Finally, both $d_2(\mu,\nu)$, in reason of the fact that it is a distance, and $|\boldsymbol{m}_\mu - \boldsymbol{m}_\nu|$ satisfy the triangular inequality. ■

An analogous extension can be done for the metric $D_{2,p}$ defined in (1.2).

Definition 5 Given $p \geq 1$, we define $\mathcal{D}_{2,p} : \mathcal{P}_2(\mathbb{R}^d) \times \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$ by

$$\mathcal{D}_{2,p}(\mu,\nu) := \sqrt{D_{2,p}(\mu, \nu_{\boldsymbol{m}_{\boldsymbol{\mu}}-\boldsymbol{m}_{\boldsymbol{\nu}}})^2 + |\boldsymbol{m}_{\boldsymbol{\mu}} - \boldsymbol{m}_{\boldsymbol{\nu}}|^2}.$$

$\mathcal{D}_{2,p}$ is a metric on $\mathcal{P}_2(\mathbb{R}^d)$.

It is remarkable that the result of Theorem 1 can be extended to the $\mathcal{D}_2$ metric.

Theorem 3 *The function $\mathcal{D}_2$ defined in* (2.12) *is equivalent to the W_2 distance.*

Proof Let $\mu,\nu \in \mathcal{P}_2(\mathbb{R}^d)$ and let μ^*,ν^* denote the two corresponding translated measures centered in $\mathbf{0}$. By Lemma 2, we have

$$W_2^2(\mu,\nu) = W_2^2(\mu^*,\nu^*) + |\boldsymbol{m}_{\boldsymbol{\mu}} - \boldsymbol{m}_{\boldsymbol{\nu}}|^2. \tag{2.13}$$

Owing to Theorem 1, there exist two constants $c, C \in (0,\infty)$ such that

$$c d_2(\mu^*,\nu^*) \leq W_2(\mu^*,\nu^*) \leq C d_2(\mu^*,\nu^*). \tag{2.14}$$

Using (2.13) in (2.14), we get

$$c d_2(\mu^*,\nu^*)^2 + |\boldsymbol{m}_{\boldsymbol{\mu}} - \boldsymbol{m}_{\boldsymbol{\nu}}|^2 \leq W_2(\mu,\nu)^2 \leq C d_2(\mu^*,\nu^*)^2 + |\boldsymbol{m}_{\boldsymbol{\mu}} - \boldsymbol{m}_{\boldsymbol{\nu}}|^2,$$

which can be rewritten as

$$\begin{aligned}\min\{c,1\}\big(d_2(\mu^*,\nu^*)^2 + |\boldsymbol{m}_{\boldsymbol{\mu}} - \boldsymbol{m}_{\boldsymbol{\nu}}|^2\big) &\leq W_2(\mu,\nu)^2 \\ &\leq \max\{1,C\}\big(d_2(\mu^*,\nu^*)^2 + |\boldsymbol{m}_{\boldsymbol{\mu}} - \boldsymbol{m}_{\boldsymbol{\nu}}|^2\big).\end{aligned}$$

Finally

$$\min\{c,1\}\,\mathcal{D}_2^2(\mu,\nu) \leq W_2^2(\mu,\nu) \leq \max\{1,C\}\,\mathcal{D}_2^2(\mu,\nu).$$

■

3 The Periodic Fourier-based metrics

In this section, we introduce a family of (Discrete) Periodic Fourier-based metrics suitable to measure the distance between discrete probability measures whose support is restricted to a given set of points, and we discuss their equivalence with the Wasserstein metrics. The main result is that in this case one obtains a precise estimation of the constants of equivalence.

Definition 6 (Regular grid) For $N \in \mathbb{N} \setminus \{0\}$, we define the regular grid

$$G_N := \left\{ \boldsymbol{x} \in \mathbb{R}^d : N\boldsymbol{x} \in \mathbb{Z}^d \cap [0, N)^d \right\}.$$

Note that $G_N \subset [0,1)^d$.

Definition 7 (Discrete Measure over a grid) We say that μ is a a discrete measure over G_N if its support is contained in G_N, that is, if μ has the form

$$\mu(\boldsymbol{x}) = \sum_{\boldsymbol{y} \in G_N} \mu_{\boldsymbol{y}} \delta(\boldsymbol{x} - \boldsymbol{y}), \tag{3.15}$$

where $\mu_{\boldsymbol{y}} \in \mathbb{R}, \mu_{\boldsymbol{y}} \geq 0$ for all $\boldsymbol{y} \in G_N$.

The Discrete Fourier transform of a discrete measure over G_N is given by

$$\hat{\mu}(\boldsymbol{k}) = \sum_{\boldsymbol{x} \in G_N} \mu_{\boldsymbol{x}} e^{-i\langle \boldsymbol{x}, \boldsymbol{k} \rangle}. \tag{3.16}$$

The periodicity of the complex exponential implies that $\hat{\mu}$ is $2\pi N$-periodic over all directions, so that it is sufficient to study $\hat{\mu}$ over a strict subset of $\mathbb{R}^d$, e.g., over $[0, 2\pi N]^d$. For instance, the value of the Fourier-based metric (1.1) is achieved by searching for the "sup" operator on the bounded set $[0, 2\pi N]^d$. Since

$$\frac{1}{|\boldsymbol{k}|^2} \geq \frac{1}{|\boldsymbol{k}'|^2}, \quad \forall \boldsymbol{k} \in (0, 2\pi N]^d, \, \forall \boldsymbol{k}' \in \mathbb{R}^d_+ \setminus [0, 2\pi N]^d$$

and the function

$$\boldsymbol{k} \to |\hat{\mu}(\boldsymbol{k}) - \hat{\nu}(\boldsymbol{k})|$$

is $2\pi N$-periodic, for any given constant $s > 0$ the Discrete Fourier-based metric can be defined as

$$d_s(\mu, \nu) = \sup_{\boldsymbol{k} \in [0, 2\pi N]^d \setminus \{0\}} \frac{|\hat{\mu}(\boldsymbol{k}) - \hat{\nu}(\boldsymbol{k})|}{|\boldsymbol{k}|^s}. \tag{3.17}$$

Definition 8 (Dilated Discrete Measures) Given a discrete measure μ over G_N and $\gamma \in \mathbb{R}$ such that $\gamma > 0$, the γ-dilated measure μ_γ is

$$\mu_\gamma(\boldsymbol{x}) = \sum_{\boldsymbol{y} \in G_N} \mu_{\boldsymbol{y}} \delta(\gamma \boldsymbol{x} - \boldsymbol{y}).$$

The Fourier transform of μ_γ is

$$\hat{\mu}_\gamma(\boldsymbol{k}) = \sum_{\boldsymbol{x} \in G_N} \mu_{\boldsymbol{x}} e^{-\frac{i}{\gamma}\langle \boldsymbol{k}, \boldsymbol{x} \rangle} = \hat{\mu}\left(\frac{\boldsymbol{k}}{\gamma}\right). \tag{3.18}$$

Therefore, if $\hat{\mu}$ is T-periodic, then $\hat{\mu}_\gamma$ is γT-periodic. Like the original metrics (1.1) [14], the metric (3.17) satisfies the dilation property

$$d_s(\mu_\gamma, \nu_\gamma) = \frac{1}{\gamma^s} d_s(\mu, \nu). \tag{3.19}$$

In particular, if we consider μ of the form (3.15), the Fourier transform of its $\frac{1}{N}$-dilation is 2π-periodic.

We recall the definition of the metrics (1.2):

$$D_{s,p}(\mu, \nu) := \left(\int_{\mathbb{R}^d} \frac{|\widehat{\mu}(\boldsymbol{k}) - \widehat{\nu}(\boldsymbol{k})|^p}{|\boldsymbol{k}|^{(sp+d)}} d\boldsymbol{k} \right)^{\frac{1}{p}},$$

where $s > 0$ and $p \geq 1$. As we did for the Fourier Based Metrics d_s, thanks to the periodicity of the Fourier transform, we can restrict the domain of integration to $[0, T]^d$. In this case, for any given choice of the parameters p and s, this distance is well-defined any time the integrand is integrable in a neighbourhood of the origin. This corresponds to requiring that $\frac{1}{|\boldsymbol{k}|^\gamma}$ is integrable on the d-dimensional ball $B_1(0) = \{\boldsymbol{k} \in \mathbb{R}^d : |\boldsymbol{k}| \leq 1\}$, that is, if and only if $\gamma < d$. This consideration suggests the following definition.

Definition 9 (The Periodic Fourier-based Metric) Let μ and ν be two probability measures over G_N. The (s, p, α)-Periodic Fourier-based Metric (or PFM) between μ and ν is defined as

$$f_{s,p}^{(\alpha)}(\mu, \nu) := \left(\frac{1}{|T|^d} \int_{[0,T]^d} \frac{|\widehat{\mu}(\boldsymbol{k}) - \widehat{\nu}(\boldsymbol{k})|^p}{|\boldsymbol{k}|^{sp+\alpha}} d\boldsymbol{k} \right)^{\frac{1}{p}}, \tag{3.20}$$

where $p, s, \alpha \in \mathbb{R}$ and T is the period of $\hat{\mu}$ and $\hat{\nu}$. When $\alpha = 0$ and $s \in \mathbb{N}$ we say that $f_{s,p} := f_{s,p}^{(0)}$ is *pure*.

As discussed in the introduction, in dimension $d = 1$ the continuous version of the metrics (3.20) has been considered in [8]. Recently, these metrics have been considered in relation with the problem of convergence toward equilibrium of a Fokker–Planck type equation modeling wealth distribution [29], where various properties of these metrics have been studied. As pointed out in [29], if μ and ν have equal r-moments, the function $|\hat{\mu}(\boldsymbol{k}) - \hat{\nu}(\boldsymbol{k})|$ behaves like $|\boldsymbol{k}|^{r+1}$ as $\boldsymbol{k} \to 0$. As a consequence, the value of $f_{s,p}^{(\alpha)}(\mu, \nu)$ is finite only if the following condition is verified

$$p(s - r - 1) + \alpha < d. \tag{3.21}$$

If s, p and α satisfy (3.21), and thus $f_{s,p}^{(\alpha)} < +\infty$, we say that $f_{s,p}^{(\alpha)}$ is feasible.

Proposition 2 *Let μ and ν be two probability measures over G_N. For any given constant $\gamma > 0$, the following dilation property holds*

$$f_{s,p}^{(\alpha)}(\mu_\gamma, \nu_\gamma) = \frac{1}{|\gamma|^{s+\frac{\alpha}{p}}} f_{s,p}^{(\alpha)}(\mu, \nu).$$

Proof Using relation (3.18) and the change of variables $\boldsymbol{k} = \gamma \boldsymbol{k}'$, we get

$$
\begin{aligned}
f_{s,p}^{(\alpha)}(\mu_\gamma, \nu_\gamma) &= \left(\frac{1}{|\gamma T|^d} \int_{[0,\gamma T]^d} \frac{|\hat{\mu}_\gamma(\boldsymbol{k}) - \hat{\nu}_\gamma(\boldsymbol{k})|^p}{|\boldsymbol{k}|^{sp+\alpha}} d\boldsymbol{k} \right)^{\frac{1}{p}} \\
&= \left(\frac{1}{|\gamma T|^d} \int_{[0,\gamma T]^d} \frac{|\hat{\mu}(\frac{\boldsymbol{k}}{\gamma}) - \hat{\nu}(\frac{\boldsymbol{k}}{\gamma})|^p}{|\boldsymbol{k}|^{sp+\alpha}} d\boldsymbol{k} \right)^{\frac{1}{p}} \\
&= \left(\frac{1}{|\gamma|^d} \frac{1}{|T|^d} \int_{[0,T]^d} \frac{|\hat{\mu}(\boldsymbol{k}') - \hat{\nu}(\boldsymbol{k}')|^p}{|\gamma|^{sp+\alpha} |\boldsymbol{k}'|^{sp+\alpha}} |\gamma|^d d\boldsymbol{k}' \right)^{\frac{1}{p}} \\
&= \frac{1}{|\gamma|^{s+\frac{\alpha}{p}}} \left(\frac{1}{|T|^d} \int_{[0,T]^d} \frac{|\hat{\mu}(\boldsymbol{k}') - \hat{\nu}(\boldsymbol{k}')|^p}{|\boldsymbol{k}'|^{sp+\alpha}} d\boldsymbol{k}' \right)^{\frac{1}{p}} \\
&= \frac{1}{|\gamma|^{s+\frac{\alpha}{p}}} f_{s,p}^{(\alpha)}(\mu, \nu).
\end{aligned}
$$

■

It is important to remark that, at difference with the metrics (1.2), the analogous of the dilation property (3.19) is true only for $\alpha = 0$, that is only for pure metrics. We show next that the $f_{s,p}^{(\alpha)}$ metrics satisfy various monotonicity properties with respect to the parameters p and s.

Proposition 3 *Let μ and ν be two probability measures over G_N, with moments equal up to r. If $t \le s$, then*

$$
f_{t,p}^{(\alpha)}(\mu, \nu) \le (\sqrt{d}|T|)^{(s-t)} f_{s,p}^{(\alpha)}(\mu, \nu),
$$

for any p and α for which the metric is feasible, i.e., for $p(s-r-1)+\alpha < d$.

Proof We compute

$$
\begin{aligned}
f_{t,p}^{(\alpha)}(\mu, \nu) &= \left(\frac{1}{|T|^d} \int_{[0,T]^d} \frac{|\hat{\mu}(\boldsymbol{k}) - \hat{\nu}(\boldsymbol{k})|^p}{|\boldsymbol{k}|^{tp+\alpha}} d\boldsymbol{k} \right)^{\frac{1}{p}} \\
&= \left(\frac{1}{|T|^d} \int_{[0,T]^d} \frac{|\boldsymbol{k}|^{p(s-t)}}{|\boldsymbol{k}|^{p(s-t)}} \frac{|\hat{\mu}(\boldsymbol{k}) - \hat{\nu}(\boldsymbol{k})|^p}{|\boldsymbol{k}|^{tp+\alpha}} d\boldsymbol{k} \right)^{\frac{1}{p}} \\
&= \left(\frac{1}{|T|^d} \int_{[0,T]^d} |\boldsymbol{k}|^{p(s-t)} \frac{|\hat{\mu}(\boldsymbol{k}) - \hat{\nu}(\boldsymbol{k})|^p}{|\boldsymbol{k}|^{sp+\alpha}} d\boldsymbol{k} \right)^{\frac{1}{p}} \\
&\le (\sqrt{d}|T|)^{(s-t)} f_{s,p}^{(\alpha)}(\mu, \nu).
\end{aligned}
$$

The last inequality is obtained resorting to the bound $|\boldsymbol{k}| \le \sqrt{d}|T|$. ■

Proposition 4 *Let μ and ν be two probability measures over G_N. If $\alpha = 0$ and $p \le q$, then*

$$
f_{s,p}(\mu, \nu) \le f_{s,q}(\mu, \nu).
$$

Proof We have

$$
\begin{aligned}
f_{s,p}(\mu,\nu) &= \left(\frac{1}{|T|^d} \int_{[0,T]^d} \frac{|\hat{\mu}(\boldsymbol{k}) - \hat{\nu}(\boldsymbol{k})|^p}{|\boldsymbol{k}|^{sp}} d\boldsymbol{k} \right)^{\frac{1}{p}} \\
&= \left(\left(\frac{1}{|T|^d} \int_{[0,T]^d} \frac{|\hat{\mu}(\boldsymbol{k}) - \hat{\nu}(\boldsymbol{k})|^p}{|\boldsymbol{k}|^{sp}} d\boldsymbol{k} \right)^{\frac{q}{p}} \right)^{\frac{1}{q}} \\
&\leq \left(\frac{1}{|T|^d} \int_{[0,T]^d} \left(\frac{|\hat{\mu}(\boldsymbol{k}) - \hat{\nu}(\boldsymbol{k})|^p}{|\boldsymbol{k}|^{sp}} \right)^{\frac{q}{p}} d\boldsymbol{k} \right)^{\frac{1}{q}} \\
&= f_{s,q}(\mu,\nu).
\end{aligned}
$$

The last inequality follows from Jensen's inequality. ■

Remark 1 By letting $p \to +\infty$, we get

$$
\lim_{p\to\infty} f_{s,p}(\mu,\nu) = f_{s,\infty}(\mu,\nu) := d_s(\mu,\nu).
$$

Thanks to Hölder inequality, for all $p < +\infty$ we have the bound

$$
f_{s,p}(\mu,\nu) \leq d_s(\mu,\nu). \tag{3.22}
$$

The results of this Section are preliminary to our main result, which deals with the equivalence of the pure metrics, for $p = 2$, with the Wasserstein metrics. For the sake of simplicity, and without loss of generality, in the next subsection we consider measures in dimension $d = 2$.

3.1 *Equivalence with the Wasserstein metric W_1*

We consider the two cases $s = 1$ and $s = 2$, in dimension $d = 2$, and we show that $f_{1,2}$ and $f_{2,2}$ are equivalent to W_1 and W_2, respectively.

We start with the case $s = 1$. For any $\mu, \nu \in \mathcal{P}(G_N)$, the PFM is

$$
f_{1,2}(\mu,\nu) = \left(\frac{1}{|T|^2} \int_{[0,T]^2} \frac{|\hat{\mu}(\boldsymbol{k}) - \hat{\nu}(\boldsymbol{k})|^2}{|\boldsymbol{k}|^2} d\boldsymbol{k} \right)^{\frac{1}{2}}. \tag{3.23}
$$

We have the following

Theorem 4 *For any pair of measures $\mu, \nu \in \mathcal{P}(G_N)$, we have the inequality*

$$
f_{1,2}(\mu,\nu) \leq W_1(\mu,\nu).
$$

Proof Let π be a transport plan between μ and ν. It holds

$$\begin{aligned}
|\hat{\mu}(\boldsymbol{k}) - \hat{\nu}(\boldsymbol{k})| &= \Big| \sum_{\boldsymbol{x},\boldsymbol{y} \in G_N} e^{-i\boldsymbol{k}\cdot\boldsymbol{x}} \pi(\boldsymbol{x},\boldsymbol{y}) - \sum_{\boldsymbol{x},\boldsymbol{y} \in G_N} e^{-i\boldsymbol{k}\cdot\boldsymbol{y}} \pi(\boldsymbol{x},\boldsymbol{y}) \Big| \\
&= \Big| \sum_{\boldsymbol{x},\boldsymbol{y} \in G_N} \big(e^{-i\boldsymbol{k}\cdot\boldsymbol{x}} - e^{-i\boldsymbol{k}\cdot\boldsymbol{y}}\big) \pi(\boldsymbol{x},\boldsymbol{y}) \Big| \\
&\leq \sum_{\boldsymbol{x},\boldsymbol{y} \in G_N} \big|e^{-i\boldsymbol{k}\cdot\boldsymbol{x}} - e^{-i\boldsymbol{k}\cdot\boldsymbol{y}}\big| \pi(\boldsymbol{x},\boldsymbol{y}) \\
&= \sum_{\boldsymbol{x},\boldsymbol{y} \in G_N} \big|1 - e^{i\boldsymbol{k}\cdot(\boldsymbol{x}-\boldsymbol{y})}\big| \pi(\boldsymbol{x},\boldsymbol{y}) \\
&\leq \sum_{\boldsymbol{x},\boldsymbol{y} \in G_N} \big|\boldsymbol{k}\cdot(\boldsymbol{x}-\boldsymbol{y})\big| \pi(\boldsymbol{x},\boldsymbol{y}) \\
&\leq |\boldsymbol{k}| \sum_{\boldsymbol{x},\boldsymbol{y} \in G_N} |\boldsymbol{x}-\boldsymbol{y}| \pi(\boldsymbol{x},\boldsymbol{y}).
\end{aligned}$$

Hence, if π is the optimal transport plan, we conclude with the inequality

$$|\hat{\mu}(\boldsymbol{k}) - \hat{\nu}(\boldsymbol{k})| \leq |\boldsymbol{k}| W_1(\mu,\nu). \tag{3.24}$$

Using inequality (3.24) into definition (3.23), we finally obtain the bound

$$f_{1,2}(\mu,\nu) \leq \left(\frac{1}{|T|^2} \int_{[0,T]^2} \frac{\big(|\boldsymbol{k}| W_1(\mu,\nu)\big)^2}{|\boldsymbol{k}|^2} d\boldsymbol{k} \right)^{\frac{1}{2}} = W_1(\mu,\nu). \tag{3.25}$$

■

Since $W_1(\mu,\nu) < +\infty$ for every $\mu,\nu \in \mathcal{P}(G_N)$, inequality (3.25) implies that $f_{1,2}$ is bounded in correspondence to any pair of probability measures over the grid G_N.

We now show that $f_{1,2}$ and W_1 satisfy a reverse inequality, thus concluding that the two metrics are equivalent.

Theorem 5 *For any pair of measures $\mu,\nu \in \mathcal{P}(G_N)$ it holds*

$$W_1(\mu,\nu) \leq \frac{T^2}{2\pi} f_{1,2}(\mu,\nu). \tag{3.26}$$

Proof Owing to the dual characterization of the W_1 distance (see [31], Chapter 5), there exists a 1-Lipschitz function ϕ such that

$$W_1(\mu,\nu) = \int_{\mathbb{R}^2} \phi(\boldsymbol{x}) d\mu(\boldsymbol{x}) - \int_{\mathbb{R}^2} \phi(\boldsymbol{x}) d\nu(\boldsymbol{x}).$$

Since μ and ν are discrete and supported on a subset of $[0,1]^2$, we can write

$$W_1(\mu,\nu) = \sum_{\boldsymbol{x} \in G_N} \phi(\boldsymbol{x}) \big(\mu_{\boldsymbol{x}} - \nu_{\boldsymbol{x}}\big).$$

Therefore, resorting to the fact that both the measures have the same mass, for any given constant $c \in \mathbb{R}$ we have

$$W_1(\mu,\nu) = \sum_{\boldsymbol{x}\in G_N} \big(\phi(\boldsymbol{x}) + c\big)\big(\mu_{\boldsymbol{x}} - \nu_{\boldsymbol{x}}\big).$$

The last identity permits to choose ϕ such that $\phi(\frac{N}{2},\frac{N}{2}) = 0$. Since ϕ is 1-Lipschitz, we conclude that

$$|\phi(\boldsymbol{x})| \leq \frac{\sqrt{2}}{2}, \qquad \forall \boldsymbol{x} \in G_N. \tag{3.27}$$

By Hölder inequality we obtain

$$W_1(\mu,\nu) \leq \Bigg(\sum_{\boldsymbol{x}\in G_N} |\phi(\boldsymbol{x})|^2\Bigg)^{\frac{1}{2}} \Bigg(\sum_{\boldsymbol{x}\in G_N} |\mu_{\boldsymbol{x}} - \nu_{\boldsymbol{x}}|^2\Bigg)^{\frac{1}{2}}.$$

Since

$$\sum_{\boldsymbol{x}\in G_N} |\mu_{\boldsymbol{x}} - \nu_{\boldsymbol{x}}|^2 = \frac{1}{|T|^2}\int_{[0,T]^2} A(\boldsymbol{k})B(\boldsymbol{k})d\boldsymbol{k}$$

where

$$A(\boldsymbol{k}) = \sum_{\boldsymbol{x}\in G_N} (\mu_{\boldsymbol{x}} - \nu_{\boldsymbol{x}})e^{-i<\boldsymbol{x},\boldsymbol{k}>}$$
$$B(\boldsymbol{k}) = \sum_{\boldsymbol{y}\in G_N} (\mu_{\boldsymbol{y}} - \nu_{\boldsymbol{y}})e^{+i<\boldsymbol{y},\boldsymbol{k}>}$$

we have

$$\sum_{\boldsymbol{x}\in G_N} |\mu_{\boldsymbol{x}} - \nu_{\boldsymbol{x}}|^2 = \frac{1}{|T|^2}\int_{[0,T]^2} \big|\hat{\mu}(\boldsymbol{k}) - \hat{\nu}(\boldsymbol{k})\big|^2 d\boldsymbol{k}.$$

Now using (3.27) we obtain

$$\begin{aligned} W_1(\mu,\nu) &\leq \frac{\sqrt{2}N}{2}\Bigg(\frac{1}{|T|^2}\int_{[0,T]^2} \big|\hat{\mu}(\boldsymbol{k}) - \hat{\nu}(\boldsymbol{k})\big|^2 d\boldsymbol{k}\Bigg)^{\frac{1}{2}} \\ &= \frac{\sqrt{2}N}{2}\Bigg(\frac{1}{|T|^2}\int_{[0,T]^2} |\boldsymbol{k}|^2 \frac{|\hat{\mu}(\boldsymbol{k}) - \hat{\nu}(\boldsymbol{k})|^2}{|\boldsymbol{k}|^2} d\boldsymbol{k}\Bigg)^{\frac{1}{2}}. \end{aligned}$$

Since $|\boldsymbol{k}|^2 \leq 2T^2$ and $T = 2\pi N$, we can finally conclude that

$$W_1(\mu,\nu) \leq \frac{T^2}{2\pi}\Bigg(\frac{1}{|T|^2}\int_{[0,T]^2} \frac{|\hat{\mu}(\boldsymbol{k}) - \hat{\nu}(\boldsymbol{k})|^2}{|\boldsymbol{k}|^2} d\boldsymbol{k}\Bigg)^{\frac{1}{2}} = \frac{T^2}{2\pi} f_{1,2}(\mu,\nu).$$

∎

In consequence of the previous estimates, it is immediate to show that the metrics d_s and W_1 are equivalent. This is proven in the following

Corollary 1 *For any pair of measures $\mu, \nu \in \mathcal{P}(G_N)$*

$$d_1(\mu,\nu) \le W_1(\mu,\nu) \le \frac{T^2}{2\pi} d_1(\mu,\nu).$$

Proof The first inequality is a consequence of bound (3.24). The second one follows from inequality (3.22). ■

3.2 *Equivalence with the Wasserstein metric W_2*

The aim of this Section is to show the equivalence of the Fourier-based metric $f_{2,2}$ and the Wasserstein metric W_2. Let $s = 2$. In this case, the PFM takes the form

$$f_{2,2}(\mu,\nu) = \left(\frac{1}{|T|^2} \int_{[0,T]^2} \frac{|\hat{\mu}(\boldsymbol{k}) - \hat{\nu}(\boldsymbol{k})|^2}{|\boldsymbol{k}|^4} d\boldsymbol{k} \right)^{\frac{1}{2}}.$$

Clearly, the distance between the two probability measures is well-defined only when μ and ν possess the same expected value. Since, in general this is not the case, we start by translating the measures, as done in Section 2, in order to satisfy this condition. The following proposition shows that, for probability measures with the same center, the topology induced by $f_{2,2}$ is not stronger than the topology induced by W_2.

Theorem 6 *For any pair of measures $\mu, \nu \in \mathcal{P}(G_N)$ such that $\boldsymbol{m_\mu} = \boldsymbol{m_\nu}$, it holds*

$$f_{2,2}(\mu,\nu) \le 2\sqrt{2} W_2(\mu,\nu). \tag{3.28}$$

In particular, $f_{2,2}(\mu,\nu) < \infty$.

Proof For any given pair of probability measures μ and ν in $\mathcal{P}(G_N)$, with centers $\boldsymbol{m_\mu} = \boldsymbol{m_\nu}$, we have

$$i\boldsymbol{k} \sum_{\boldsymbol{x} \in G_N} \boldsymbol{x} \mu_{\boldsymbol{x}} = i\boldsymbol{k} \sum_{\boldsymbol{y} \in G_N} \boldsymbol{y} \nu_{\boldsymbol{y}}.$$

For any transport plan π between μ and ν, we can rewrite the previous relations in the form

$$i\boldsymbol{k} \sum_{\boldsymbol{x},\boldsymbol{y} \in G_N} (\boldsymbol{x} - \boldsymbol{y}) \pi_{\boldsymbol{x},\boldsymbol{y}} = 0. \tag{3.29}$$

Using identity (3.29) we obtain

$$\begin{aligned}
\hat{\mu}(\boldsymbol{k}) - \hat{\nu}(\boldsymbol{k}) &= \sum_{\boldsymbol{x} \in G_N} \mu_{\boldsymbol{x}} e^{-i\boldsymbol{k}\cdot\boldsymbol{x}} - \sum_{\boldsymbol{y} \in G_N} \nu_{\boldsymbol{y}} e^{-i\boldsymbol{k}\cdot\boldsymbol{y}} \\
&= \sum_{\boldsymbol{x},\boldsymbol{y} \in G_N} \left(e^{-i\boldsymbol{k}\cdot\boldsymbol{x}} - e^{-i\boldsymbol{k}\cdot\boldsymbol{y}} - i\boldsymbol{k}\cdot(\boldsymbol{x}-\boldsymbol{y}) \right) \pi_{\boldsymbol{x},\boldsymbol{y}} \\
&= \sum_{\boldsymbol{x},\boldsymbol{y} \in G_N} e^{-i\boldsymbol{k}\cdot\boldsymbol{y}} \big(e^{-i\boldsymbol{k}\cdot(\boldsymbol{x}-\boldsymbol{y})} - 1 - i\boldsymbol{k}\cdot(\boldsymbol{x}-\boldsymbol{y}) \big) \pi_{\boldsymbol{x},\boldsymbol{y}} \\
&\quad + \sum_{\boldsymbol{x},\boldsymbol{y} \in G_N} i\boldsymbol{k}\cdot(\boldsymbol{x}-\boldsymbol{y})(e^{-i\boldsymbol{k}\cdot\boldsymbol{y}} - 1)\pi_{\boldsymbol{x},\boldsymbol{y}}.
\end{aligned}$$

Using that for all $\theta \in \mathbb{R}$

$$|e^{i\theta} - 1| \le |\theta|,$$
$$|e^{i\theta} - 1 - i\theta| \le \frac{\theta^2}{2}$$

we obtain

$$\begin{aligned}
|\hat{\mu}(\boldsymbol{k}) - \hat{\nu}(\boldsymbol{k})| &\le \frac{|\boldsymbol{k}|^2}{2} \sum_{\boldsymbol{x},\boldsymbol{y} \in G_N} |\boldsymbol{x} - \boldsymbol{y}|^2 \pi_{\boldsymbol{x},\boldsymbol{y}} + |\boldsymbol{k}|^2 \sum_{\boldsymbol{x},\boldsymbol{y} \in G_N} |\boldsymbol{x} - \boldsymbol{y}||\boldsymbol{y}| \pi_{\boldsymbol{x},\boldsymbol{y}} \\
&\le \frac{|\boldsymbol{k}|^2}{2} \sum_{\boldsymbol{x},\boldsymbol{y} \in G_N} |\boldsymbol{x} - \boldsymbol{y}|^2 \pi_{\boldsymbol{x},\boldsymbol{y}} \\
&\quad + |\boldsymbol{k}|^2 \left(\sum_{\boldsymbol{x},\boldsymbol{y} \in G_N} |\boldsymbol{y}|^2 \pi_{\boldsymbol{x},\boldsymbol{y}} \right)^{\frac{1}{2}} \left(\sum_{\boldsymbol{x},\boldsymbol{y} \in G_N} |\boldsymbol{x} - \boldsymbol{y}|^2 \pi_{\boldsymbol{x},\boldsymbol{y}} \right)^{\frac{1}{2}}.
\end{aligned}$$

In particular, if we take π as the optimal transportation plan between μ and ν for the cost $|\boldsymbol{x} - \boldsymbol{y}|^2$ we get

$$\begin{aligned}
\frac{|\hat{\mu}(\boldsymbol{k}) - \hat{\nu}(\boldsymbol{k})|}{|\boldsymbol{k}|^2} &\le \frac{W_2^2(\mu,\nu)}{2} + \left(\sum_{\boldsymbol{y} \in G_N} |\boldsymbol{y}|^2 \nu_{\boldsymbol{y}} \right)^{\frac{1}{2}} W_2(\mu,\nu) \\
&= W_2(\mu,\nu) \left(\frac{W_2(\mu,\nu)}{2} + \left(\sum_{\boldsymbol{y} \in G_N} |\boldsymbol{y}|^2 \nu_{\boldsymbol{y}} \right)^{\frac{1}{2}} \right).
\end{aligned}$$

Since

$$W_2(\mu,\nu) \le W_2(\mu,\delta) + W_2(\delta,\nu) \le \left(\sum_{\boldsymbol{x} \in G_N} |\boldsymbol{x}|^2 \mu_{\boldsymbol{x}} \right)^{\frac{1}{2}} + \left(\sum_{\boldsymbol{y} \in G_N} |\boldsymbol{y}|^2 \nu_{\boldsymbol{y}} \right)^{\frac{1}{2}},$$

and, as μ and ν are supported in $[0,1]^2$,

$$\sqrt{\sum_{\boldsymbol{x} \in G_N} |\boldsymbol{x}|^2 \mu_{\boldsymbol{x}}} \le \sqrt{2}, \quad \sqrt{\sum_{\boldsymbol{y} \in G_N} |\boldsymbol{y}|^2 \nu_{\boldsymbol{y}}} \le \sqrt{2},$$

we obtain (3.28):

$$\frac{|\hat{\mu}(\boldsymbol{k}) - \hat{\nu}(\boldsymbol{k})|}{|\boldsymbol{k}|^2} \le 2\sqrt{2} W_2(\mu,\nu).$$

■

We conclude by showing the validity of a reverse inequality, thus proving the equivalence between $f_{2,2}$ and W_2.

Theorem 7 *For any pair of measures $\mu, \nu \in \mathcal{P}(G_N)$, we have the inequality*

$$W_2^2(\mu,\nu) \le \frac{T^3}{\pi} f_{2,2}(\mu,\nu).$$

Proof Let π be the optimal transportation plan between μ and ν for the cost $|\boldsymbol{x}-\boldsymbol{y}|$, since $|\boldsymbol{x}-\boldsymbol{y}| \leq \sqrt{2}$ for all $\boldsymbol{x}, \boldsymbol{y} \in G_N \subset [0,1]^2$, it holds

$$W_2^2(\mu,\nu) \leq \sum_{\boldsymbol{x},\boldsymbol{y} \in G_N} |\boldsymbol{x}-\boldsymbol{y}|^2 \pi_{\boldsymbol{x},\boldsymbol{y}} \leq \sum_{\boldsymbol{x},\boldsymbol{y} \in G_N} \sqrt{2}|\boldsymbol{x}-\boldsymbol{y}| \pi_{\boldsymbol{x},\boldsymbol{y}} = \sqrt{2} W_1(\mu,\nu).$$

Then, by Theorem 5 and Proposition 3 with $t=1$ and $p=s=2$, we get

$$\sqrt{2} W_1(\mu,\nu) \leq \frac{\sqrt{2} T^2}{2\pi} f_{1,2}(\mu,\nu) \leq \frac{T^3}{\pi} f_{2,2}(\mu,\nu),$$

which, together with the last inequality, concludes the proof. ■

The previous bounds hold provided that μ and ν are centered in the same point. However, when $\boldsymbol{m_\mu} - \boldsymbol{m_\nu} \neq 0$, we can resort, as in Section 2, to the new metric

$$\mathcal{F}_{2,2}(\mu,\nu) := \sqrt{\left(f_{2,2}(\mu,\nu_{\boldsymbol{m_\mu}-\boldsymbol{m_\nu}})^2 + |\boldsymbol{m_\mu} - \boldsymbol{m_\nu}|^2\right)},$$

which is well-defined also for probability measures having different centers. This shows that we can generalize, similarly to Theorem 2 and Theorem 3, the equivalence of $\mathcal{F}_{2,2}$ and W_2 to measures which are not centered in the same point.

3.3 *Connections with other distances*

As discussed in [29], the case in which $s \leq 0$ leads to stronger metrics. In this case, we clearly loose relations like (3.28), that link from above the Wasserstein metric with the Fourier-based metric. An interesting case is furnished by choosing $s=0$ into (3.20). The metric in this case is defined by

$$\begin{aligned} f_{0,2}(\mu,\nu) &= \left(\frac{1}{|T|^d} \int_{[0,T]^d} |\hat{\mu}(\boldsymbol{k}) - \hat{\nu}(\boldsymbol{k})|^2 d\boldsymbol{k} \right)^{\frac{1}{2}} \\ &= \left(\sum_{\boldsymbol{x} \in G} |\mu(\boldsymbol{x}) - \nu(\boldsymbol{x})|^2 \right)^{\frac{1}{2}}, \end{aligned}$$

which defines the Total Variation distance between the probability measures μ and ν.

We remark that the distance above corresponds to the choice $\alpha = 0$, which does not require the measures to possess the same mass. In alternative one can choose a value $\alpha \in [0,2)$. However, if $\alpha > 0$, one obtains a distance between measures that requires that the two measures have the same mass. Note however that the choice of values of $\alpha > 0$ allows to obtain a sequence of metrics that interpolate between the Total Variation distance and the W_1 distance, namely a family of measures that move from a strong metric to a weaker one.

In the case $s < 0$ the Fourier-based metric (3.20) becomes

$$f_{s,2}(\mu,\nu) = \left(\frac{1}{|T|^d} \int_{[0,T]^d} |\boldsymbol{k}|^{2|s|} |\hat{\mu}(\boldsymbol{k}) - \hat{\nu}(\boldsymbol{k})|^2 d\boldsymbol{k} \right)^{\frac{1}{2}}.$$

In particular, when $-s = n \in \mathbb{N}_+$, we find that

$$f_{-n,2}(\mu, \nu) = \left(\frac{1}{|T|^d} \int_{[0,T]^d} |\boldsymbol{k}|^{2n} |\hat{\mu}(\boldsymbol{k}) - \hat{\nu}(\boldsymbol{k})|^2 d\boldsymbol{k} \right)^{\frac{1}{2}}.$$

This metric, by Fourier identity, controls the n-th derivative of the measures μ and ν.

4 Numerical Results

We run extensive numerical tests to compare the Wasserstein metrics W_1 and W_2 with the corresponding Periodic Fourier-based Metrics $f^0_{1,2}$ and $f^0_{2,2}$.

The goal of our tests is to compare empirically the distance values obtained with the different metrics, and to measure the runtime gain that we can achieve using the Fourier-based metrics. In the following paragraphs, we report the main conclusions of our tests.

Implementation details. We implemented our algorithms in Python 3.7, using the Fast Fourier Transform implemented in the Numpy library [23]. To compute the Wasserstein distances, we use the Python Optimal Transport (POT) library [17]. All the tests are executed on a MacBook Pro 13 equipped with a 2.5 GHz Intel Core i7 dual-core and 16 GB of Ram.

Dataset. As problem instances, we use the DOTmark benchmark [27], which contains 10 classes of gray scale images, each containing 10 different images. Every image is given in the data set at the following pixel resolutions: $32 \times 32, 64 \times 64, 128 \times 128, 256 \times 256$, and 512×512. Figure 1 shows the Classic, Microscopy, and Shapes images, respectively, at the highest pixel resolution (one class for each row).

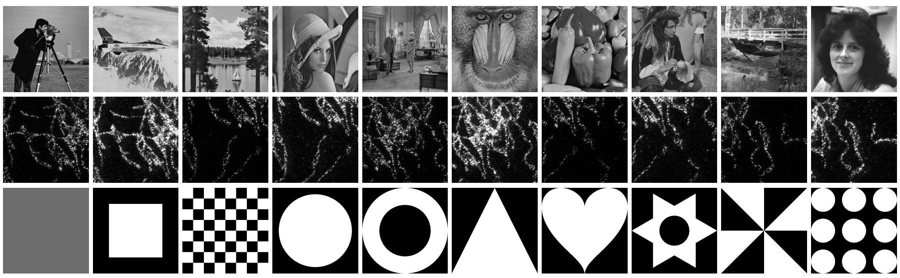

Fig. 1 DOTmark benchmark: Classic, Microscopy, and Shapes images.

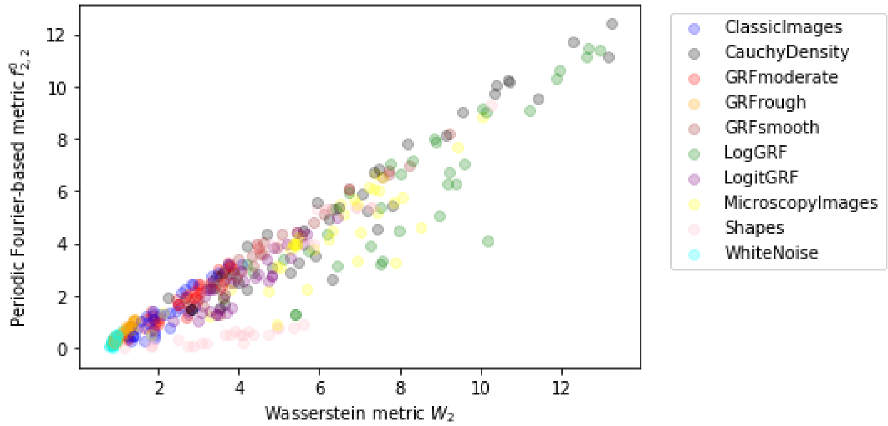

Fig. 2 Wasserstein metric W_2 versus Periodic Fourier-based metric $f^0_{2,2}$: Comparison of distance values for 450 pair of images of size 32×32.

Results. For each pair of images of the DOTmark dataset, we have computed the reciprocal distance values using the $W_1, W_2, f^0_{1,2}$ and $f^0_{2,2}$ metrics, and we have recorded the corresponding runtime in seconds.

The scatter plot in Figure 2 shows the relation between the W_2 and the $f^0_{2,2}$ distances for each pairs of images at pixel resolution 32×32. The plot shows that not only the two metrics are theoretically equivalent, as proofed in Theorem 6 and 7, but they also yields very similar values in practice. The only partial exception is the Shape class, which, however, contains artificial shape images. Note, however, that on the much more (application-wise) interesting Classic images, the two metrics return very close values.

Table 1 reports the averages and the standard deviations of the runtime, measured in seconds, at different image size. For each row and each metric, the averages are computed over 450 instances. The numerical results clearly show that the PFM metrics are orders of magnitude faster, and permits to compute the distance even for the largest 512×512 images in around 10 seconds. Note that using the POT library, we were unable to compute the W_1 and W_2 distances for images of size 256×256 and 512×512, due to memory issues.

Table 1 Runtime vs. Image size for different metrics: The runtime is measured in seconds and reported as "*Mean (StdDev)*". Each row gives the averages over 450 instances of pairwise distances.

	Averages Runtime in seconds			
Dimension	W_1	W_2	$f^0_{1,2}$	$f^0_{2,2}$
32×32	0.84 (0.30)	1.06 (0.32)	0.002 (10^{-4})	0.006 (10^{-4})
64×64	21.9 (7.96)	23.4 (8.49)	0.01 (10^{-3})	0.02 (10^{-3})
128×128	205.0 (45.9)	199.0 (45.0)	0.28 (0.07)	0.63 (0.16)
256×256			1.21 (0.40)	2.96 (0.94)
512×512			4.74 (1.32)	11.55 (2.84)

5 Conclusions

The Fourier-based metrics introduced in [19] and [8] are useful tools to measure the distance between pairs of probability distributions in terms of their Fourier transforms, and they represent an interesting alternative to the Wasserstein metric, in reason of their equivalence.

In this paper, we have shown that this equivalence can be precisely quantified when discrete probability measures are considered. In addition, our computational results shown the the Fourier metrics can be computed in matter of seconds even for very large images. Based on these results on Fourier metrics, it will be possible to design new numerical methods in computer imaging, having good theoretical convergence results with a lower computational cost than the the Wasserstein metric, which, nowadays, has still a heavy computational load.

Acknowledgements This paper was written within the activities of the GNFM of INDAM. The research was partially supported by the Italian Ministry of Education, University and Research (MIUR): Dipartimenti di Eccellenza Program (2018–2022) - Dept. of Mathematics "F. Casorati", University of Pavia. The PhD scholarship of Andrea Codegoni is founded by Sea Vision S.r.l..

References

1. J. Adler and S. Lunz. Banach Wasserstein GAN. In *Advances in Neural Information Processing Systems*, pages 6754–6763, (2018).
2. L. Ambrosio and N. Gigli. A userâĂŹs guide to optimal transport. In *Modelling and optimisation of flows on networks*, pages 1–155. Springer, (2013).
3. L. Ambrosio, N. Gigli, and G. Savaré. *Gradient flows: in metric spaces and in the space of probability measures.* Springer Science & Business Media, (2008).
4. S. Angenent, S. Haker, A. Tannenbaum, and L. Zhu. Optimal mass transport for registration and warping. *International Journal of computer vision*, 60(3):225–240, (2004).
5. M. Arjovsky, L. Bottou, and S. Chintala. Wasserstein GAN. *arXiv preprint arXiv:1701.07875*, (2017).
6. G. Auricchio, F. Bassetti, S. Gualandi, and M. Veneroni. Computing Kantorovich-Wasserstein distances on d-dimensional histograms using $(d+1)$-partite graphs. In *Advances in Neural Information Processing Systems*, pages 5793–5803, (2018).
7. G. Auricchio, F. Bassetti, S. Gualandi, and M. Veneroni. Computing Wasserstein Barycenters via Linear Programming. In *International Conference on Integration of*

Constraint Programming, Artificial Intelligence, and Operations Research, pages 355–363. Springer, (2019).
8. L. Baringhaus and R. Grübel. On a class of characterization problems for random convex combinations. *Annals of the Institute of Statistical Mathematics*, 49(3):555–567, (1997).
9. F. Bassetti, S. Gualandi, and M. Veneroni. On the computation of Kantorovich-Wasserstein distances between 2d-histograms by uncapacitated minimum cost flows. *arXiv preprint arXiv:1804.00445*, (2018).
10. M. Bisi, J.A. Carrillo, and G. Toscani. Decay rates in probability metrics towards homogeneous cooling states for the inelastic Maxwell model. *Journal of statistical physics*, 124(2-4):625–653, (2006).
11. N. Bonneel and D. Coeurjolly. SPOT: Sliced Partial Optimal Transport. *ACM Transactions on Graphics*, 38(4):1–13, (2019).
12. E.A. Carlen, M.C. Carvalho, and E. Gabetta. Central limit theorem for Maxwellian molecules and truncation of the Wild expansion. *Communications on Pure and Applied Mathematics*, 53(3):370–397, (2000).
13. E.A. Carlen, E. Gabetta, and G. Toscani. Propagation of smoothness and the rate of exponential convergence to equilibrium for a spatially homogeneous Maxwellian gas. *Communications in mathematical physics*, 199(3):521–546, (1999).
14. J.A. Carrillo and G. Toscani. Contractive probability metrics and asymptotic behavior of dissipative kinetic equations. *Prepublicacions del Centre de Recerca Matemàtica;730*, (2007).
15. M. Cuturi and A. Doucet. Fast computation of Wasserstein barycenters. In *International Conference on Machine Learning*, pages 685–693, (2014).
16. M. Cuturi and G. Peyré. Computational optimal transport. *Foundations and Trends® in Machine Learning*, 11(5-6):355–607, (2019).
17. R. Flamary and N. Courty. POT Python Optimal Transport library, 2017.
18. C. Frogner, C. Zhang, H. Mobahi, M. Araya, and T. Poggio. Learning with a Wasserstein loss. In *Advances in Neural Information Processing Systems*, pages 2053–2061, (2015).
19. G. Gabetta, G. Toscani, and B. Wennberg. Metrics for probability distributions and the trend to equilibrium for solutions of the Boltzmann equation. *Journal of statistical physics*, 81(5-6):901–934, (1995).
20. T. Goudon, S. Junca, and G. Toscani. Fourier-based distances and Berry-Esseen like inequalities for smooth densities. *Monatshefte für Mathematik*, 135(2):115–136, (2002).
21. L.J. Guibas, Y. Rubner, and C. Tomasi. The earth mover's distance as a metric for image retrieval. *International journal of computer vision*, 40(2):99–121, (2000).
22. H. Ling and K. Okada. An efficient Earth Mover's Distance algorithm for robust histogram comparison. *IEEE transactions on pattern analysis and machine intelligence*, 29(5):840–853, (2007).
23. Travis Oliphant. NumPy: A guide to NumPy. USA: Trelgol Publishing, 2006–. [Online; accessed ¡today¿].
24. O. Pele and M. Werman. Fast and robust earth mover's distances. In *2009 IEEE 12th International Conference on Computer Vision*, pages 460–467. IEEE, (2009).
25. A. Pulvirenti and G. Toscani. Asymptotic properties of the inelastic Kac model. *Journal of statistical physics*, 114(5-6):1453–1480, (2004).
26. F. Santambrogio. *Optimal transport for applied mathematicians*, volume 55. Springer, (2015).
27. J. Schrieber, D. Schuhmacher, and C. Gottschlich. Dotmark–A benchmark for Discrete Optimal Transport. *IEEE Access*, 5:271–282, 2017.
28. S. Shirdhonkar and D. W. Jacobs. Approximate earth moverâĂŹs distance in linear time. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2008.
29. M. Torregrossa and G. Toscani. Wealth distribution in presence of debts. A Fokker–Planck description. *arXiv preprint arXiv:1709.09858*, (2017).
30. G. Toscani and C. Villani. Probability metrics and uniqueness of the solution to the Boltzmann equation for a Maxwell gas. *Journal of statistical physics*, 94(3-4):619–637, (1999).
31. C. Villani. *Optimal transport: old and new.* Springer Science & Business Media, (2008).
32. V. M. Zolotarev. Probability metrics. *Theory of Probability & Its Applications*, 28(2):278–302, 1984.