COMMUN. MATH. SCI.
Vol. 21, No. 3, pp. 627–639

THE FOURIER DISCREPANCY FUNCTION*

GENNARO AURICCHIO†, ANDREA CODEGONI‡, STEFANO GUALANDI§, AND LORENZO ZAMBON¶

Abstract. In this paper, we introduce the p-Fourier Discrepancy Functions, a new family of metrics for comparing discrete probability measures, inspired by the χ_r-metrics. Unlike the χ_r-metrics, the p-Fourier Discrepancies are well-defined for any pair of measures. We prove that the p-Fourier Discrepancies are convex, twice differentiable, and that their gradient has an explicit formula. Moreover, we study the lower and upper tight bounds for the p-Fourier Discrepancies in terms of the Total Variation distance.

Keywords. Fourier metrics; discrete discrepancy; tight bounds.

AMS subject classifications. 60E10; 60E15; 60A10; 94A17.

1. Introduction

Discrepancies are becoming omnipresent tool in every applied fields that require the comparison of probability measures. Examples include computer vision [1–8], supervised learning [9–14], and generative models [15–20]. Often the usage of these tools are bounded by their numerical complexity [21–24].

To mitigate these issues, in recent years, several studies have been devoted to introduce new discrepancies [25, 26] or to study the properties of the existing ones [27, 28]. A special role is played by the study of the relationships between different discrepancies, usually through bounds.

In particular, the problem of finding the tight bounds [29] in terms of the Total Variation has been particularly interesting for source coding [30–32].

A well-known family of distances between probability measures is given by the χ_r-metrics. They are defined as the L^p distance between the characteristic functions of two given measures weighted by the function $\|k\|^{-rp}$. Despite the appealing properties they enjoy, the use of these metrics is bounded by the fact that they are not well-defined unless the two measures we are comparing have equal moments up to the $\lceil r \rceil$-th one [33, 34]. This is a standard assumption in some applied fields, such as kinetic theory [35, 36]. In general, however, requiring two measures to have the same expectation is too restricting. In [37], the authors studied the χ_r-metrics in the specific framework of discrete measures supported over a regular grid. In this framework, they prove that some requirements about the measures can be dropped while still preserving the appealing properties of their continuous counterparts. However, these distances are defined through an integral, and for $r \geq 2$ some conditions on the moments are still required to ensure the finiteness of the integral. In this paper, we overcome this issue by introducing a discretized version of the χ_r-metrics, called Fourier Discrepancies.

*Received: January 05, 2022; Accepted (in revised form): July 04, 2022. Communicated by Lorenzo Pareschi.

†Department of Mathematics, University of Pavia, Pavia, PV, 27100 Italy (gennaro.auricchio@unipv.it).

‡Department of Mathematics, University of Pavia, Pavia, PV, 27100 Italy (andrea.codegoni01@universitadipavia.it).

§Department of Mathematics, University of Pavia, Pavia, PV, 27100 Italy (stefano.gualandi@unipv.it).

¶Department of Mathematics, University of Pavia, Pavia, PV, 27100 Italy (lorenzogianmar.zambon01@universitadipavia.it).

The paper is structured as follows. In Section 2, we recall the main notions about discrete probability measures and the Discrete Fourier Transform (DFT) [38]. In Section 3, we introduce a new family of distances between discrete probability measures, the p-Fourier Discrepancies. We show that they can be expressed as the square root of a bilinear form induced by a positive definite matrix, hence they are 1-homogeneous and convex. Moreover, we prove that the squared Fourier Discrepancy is twice differentiable and that both its gradient and Hessian have an explicit formula. In Section 4, we study the lower and upper tight bounds of the Fourier Discrepancy in terms of the Total Variation distance. In particular, we prove that the upper tight bound between any q-homogeneous and convex function and the Total Variation is attained in a finite set. We then present an open conjecture about the value of the upper tight bound of the Fourier Discrepancy. Finally, conclusions and future work are discussed in Section 5. For the sake of conciseness, we only report the essential proofs in the body of the paper and leave the others in the appendix.

2. Preliminaries

In this section, we state the framework of our work and fix our notation. Throughout the paper, we only consider one-dimensional discrete measures, but all the results may be extended to a multidimensional setting. Let us define the set $I_N \subset [0,1]$ as $I_N := \left\{0, \frac{1}{N}, \dots, \frac{N-1}{N}\right\}$. For the sake of simplicity, we will assume that N is an even number. A discrete measure μ on I_N is defined as

$$\mu := \sum_{j=0}^{N-1} \mu_j \delta_{\frac{j}{N}}, \tag{2.1}$$

where all the μ_j's are real values and, for any $k \in I_N$, δ_k is the Dirac's delta centered in k. We denote by $\mathcal{M}(I_N)$ the set of discrete measures over I_N and by $\mathcal{P}(I_N) := \{\mu \in \mathcal{M}(I_N) : \mu_j \geq 0, \sum_{j=0}^{N-1} \mu_j = 1\}$ the space of discrete probability measures.

Remark 2.1. Since any discrete measure supported on I_N is fully characterised by the $N-$uple of positive values $(\mu_0, \dots, \mu_{N-1})$, we refer to discrete measures and vectors interchangeably. Although this might lead to a slight abuse of notations, it allows us to express the Fourier Transform of a discrete measure through a linear operator.

Definition 2.1. *The Discrete Fourier Transform (DFT) of $\mu \in \mathcal{P}(I_N)$ is the $N-$dimensional vector $\hat{\mu} := (\hat{\mu}_0, \dots, \hat{\mu}_{N-1})$ defined as*

$$\hat{\mu}_k := \sum_{j=0}^{N-1} \mu_j e^{-2\pi i \frac{j}{N} k}, \qquad k \in \{0, \dots, N-1\}. \tag{2.2}$$

Remark 2.2. Since the complex exponential function $k \to e^{-2\pi i \frac{j}{N} k}$ is a $N-$periodic function for any integer j, we set $\hat{\mu}_k := \hat{\mu}_{mod_N(k)}$ for any $k \in \mathbb{Z}$, where $mod_N(k)$ is the $N-$modulo operation. In particular, $\hat{\mu}_{-k} = \hat{\mu}_{N-k}$ for any $k \in \{0, \dots, N-1\}$.

Remark 2.3. The DFT of a discrete measure can be expressed as a linear map:

$$(\hat{\mu}_0, \dots, \hat{\mu}_{N-1}) = \Omega \cdot (\mu_0, \dots, \mu_{N-1}), \tag{2.3}$$

where Ω is the $N \times N$ matrix defined as

$$\Omega := \begin{bmatrix} \omega_{0,0} & \omega_{0,1} & \cdots & \omega_{0,N-1} \\ \omega_{1,0} & \omega_{1,1} & \cdots & \omega_{1,N-1} \\ \vdots & \vdots & \ddots & \vdots \\ \omega_{N-1,0} & \omega_{N-1,1} & \cdots & \omega_{N-1,N-1} \end{bmatrix}, \tag{2.4}$$

and $\omega_{k,j}:=e^{-2\pi i\frac{j}{N}k}$. Since the matrix Ω is invertible, the DFT is a bijective function.

For a complete discussion about the Discrete Fourier Transform (DFT), we refer to [38].

3. The Fourier discrepancy function

In this section we introduce the p-Fourier Discrepancy Functions, a family of discrete versions of the metrics introduced in [37]. The p-Fourier Discrepancies inherit from their continuous counterparts the property of being bounded by the Wasserstein distance. We show that the Fourier Discrepancies are convex and have an explicit derivative.

Definition 3.1. *For any $p\geq 1$, the p-Fourier Discrepancy Function is defined as $\mathbb{F}_p:\mathcal{P}(I_N)\times\mathcal{P}(I_N)\to[0,+\infty)$, where*

$$\mathbb{F}_p^2(\mu,\nu):=\sum_{k=1}^{\frac{N}{2}-1}\frac{|\hat{\mu}_k-\hat{\nu}_k|^2}{|k|^{2p}}+\frac{|\hat{\mu}_{\frac{N}{2}}-\hat{\nu}_{\frac{N}{2}}|^2}{|N|^{2p}}. \tag{3.1}$$

Remark 3.1. It is easy to show that every $\mathbb{F}_p$ is a distance on $\mathcal{P}(I_N)$. In particular, unlike its continuous counterparts, $\mathbb{F}_p$ is finite even without requiring the two measures to have any equal moment.

Remark 3.2. Following [37], it is possible to prove that

$$\mathbb{F}_p\leq C_p W_1 \tag{3.2}$$

for any $p>\frac{3}{2}$, where W_1 is the 1-Wasserstein distance [39] and C_p is a constant that only depends on p.

For any $p\geq 1$, let us introduce the matrix $\mathbb{K}_p:=\mathrm{diag}(b_p)$, where the vector b_p is defined as

$$b_p:=\frac{1}{2}\left(1,1^{-2p},\dots,\left(\frac{N}{2}-1\right)^{-2p},\frac{2}{N^{2p}},\left(\frac{N}{2}-1\right)^{-2p},\dots,1^{-2p}\right). \tag{3.3}$$

Since $\hat{\mu}_k=\overline{\hat{\mu}_{N-k}}$, we can express the Fourier Discrepancy function as a quadratic form:

$$\mathbb{F}_p^2(\mu,\nu)=(\hat{\mu}-\hat{\nu})^T\mathbb{K}_p(\hat{\mu}-\hat{\nu})=(\mu-\nu)^T\mathbb{H}_p(\mu-\nu), \tag{3.4}$$

where $\mathbb{H}_p:=\Omega^T\mathbb{K}_p\Omega$ and Ω is the DFT matrix. Notice that we only consider the first $\frac{N}{2}$ frequencies as the last $\frac{N}{2}$ have the same magnitude, hence no information is lost by omitting them. Moreover, $\mathbb{H}_p$ is a symmetric and circulant matrix, since $(\mathbb{H}_p)_{i,j}=\mathrm{Re}((\hat{b}_p)_{i-j})$. Therefore, its eigenvalues can be explicitly computed [40], leading us to the following result.

Lemma 3.1. *For any $p\geq 1$, the matrix $\mathbb{H}_p$ is positive definite and its eigenvalues are given by*

$$\lambda_i=N\cdot(b_p)_i,\qquad i=0,\dots,N-1.$$

Since $\mathbb{H}_p$ is positive definite, there exists a matrix $\mathbb{L}_p$ such that $\mathbb{L}_p^T\mathbb{L}_p=\mathbb{H}_p$. We can then write $\mathbb{F}_p(\mu-\nu)=\|\mathbb{L}_p(\mu-\nu)\|_2$, where $\|\cdot\|_2$ is the l^2 norm. Hence, we have the following.

Theorem 3.1. *For any $p\geq 1$, the Fourier Discrepancy $\mathbb{F}_p$ is convex and $1-$homogeneous with respect to $\mu-\nu$.*

To conclude, we observe that we are able to explicitly compute the gradient and Hessian matrix of $\mathbb{F}_p^2$.

PROPOSITION 3.1. *For any $p\geq 1$ and for any probability measure ν, the function $L_{p,\nu}$: $\mathcal{P}(I_n)\to\mathbb{R}$, defined as $L_{p,\nu}(\mu):=\mathbb{F}_p^2(\mu,\nu)$, is twice differentiable. Moreover, its gradient and Hessian matrix are expressed through the explicit formulae:*

$$(\nabla L_{p,\nu})_l(\mu)=\frac{\partial L_{p,\nu}}{\partial \mu_l}(\mu)=2\sum_{j=0}^{N-1}(\mu_j-\nu_j)\cdot \mathrm{Re}\Big((\hat{b}_p)_{j-l}\Big) \tag{3.5}$$

and

$$(HL_{p,\nu})_{h,l}(\mu)=\frac{\partial^2 L_{p,\nu}}{\partial\mu_h\partial\mu_l}(\mu)=2\mathrm{Re}\Big((\hat{b}_p)_{h-l}\Big), \tag{3.6}$$

where $\hat{b}_p$ is the Fourier Transform of the vector b_p.

4. Tight bounds

In this section, we study the tight bounds for the p-Fourier Discrepancy in terms of the Total Variation distance. We recall that, for any pair of discrete measures supported on I_N, the Total Variation is defined as

$$TV(\mu,\nu):=\frac{1}{2}\sum_{j=0}^{N-1}|\mu_j-\nu_j|.$$

Following [32], for any given $\theta\in(0,1]$, we define the lower and the upper tight bounds, respectively $C_L(\theta)$ and $C_U(\theta)$, as

$$C_L(\theta):=\inf_{\mu,\nu:TV(\mu,\nu)=\theta}\mathbb{F}_p(\mu,\nu), \tag{4.1}$$

$$C_U(\theta):=\sup_{\mu,\nu:TV(\mu,\nu)=\theta}\mathbb{F}_p(\mu,\nu). \tag{4.2}$$

Due to the linearity of the DFT, we have that

$$\mathbb{F}_p^2(\mu,\nu)=\sum_{k=1}^{\frac{N}{2}-1}\frac{|\widehat{(\mu-\nu)}_k|^2}{|k|^{2p}}+\frac{|\widehat{(\mu-\nu)}_{\frac{N}{2}}|^2}{|N|^{2p}}, \tag{4.3}$$

we then set $\Delta:=\mu-\nu$ and express $\mathbb{F}_p$ as a function of Δ, rather than μ and ν. Analogously, we will often write $TV(\Delta)$ instead of $TV(\mu,\nu)$, as long as $\Delta=\mu-\nu$. We now introduce the set of null-sum measures over I_N, $\mathcal{O}(I_N)$, defined as $\mathcal{O}(I_N):=\big\{\Delta\in\mathcal{M}(I_N)\text{ s.t. }\sum_i\Delta_i=0\big\}$. Given any pair of probability measures μ and ν, it is easy to see $\mu-\nu\in\mathcal{O}(I_N)$. Up to a multiplicative constant, the converse is also true.

PROPOSITION 4.1. *Given any non-zero $\Delta\in\mathcal{O}(I_N)$ and $\theta\in(0,1]$, there exists $C>0$ and a pair of probability measures (μ,ν) such that*

$$\mu-\nu=C\cdot\Delta\quad\textit{and}\quad TV(\mu,\nu)=\theta.$$

REMARK 4.1. Thanks to Proposition 4.1, and for the 1-homogeneity of $\mathbb{F}_p$, we have that, for any $\theta\in[0,1)$

$$C_L(\theta)=\inf_{\substack{\Delta\in\mathcal{O}(I_N):\\ \Delta\neq 0}}\mathbb{F}_p\left(\frac{\theta}{TV(\Delta)}\Delta\right)=\theta\cdot\inf_{\substack{\Delta\in\mathcal{O}(I_N):\\ \Delta\neq 0}}\frac{\mathbb{F}_p(\Delta)}{TV(\Delta)}, \tag{4.4}$$

and, analogously,

$$C_U(\theta)=\theta\cdot \sup_{\substack{\Delta\in\mathcal{O}(I_N):\\ \Delta\neq 0}} \frac{\mathbb{F}_p(\Delta)}{TV(\Delta)}. \tag{4.5}$$

4.1. Lower tight bound. Let us define $\omega_k\in\mathbb{C}^N$ as the $k-$th column of the DFT matrix Ω. Since $\{\omega_k\}_{k=0,\dots,N-1}$ is an orthogonal basis of $\mathbb{C}^n$ [38], for any $\Delta\in\mathcal{O}(I_N)$ there exists a unique N-tuple of complex coefficients $\left(\lambda^{(k)}\right)_{k=0,\dots,N-1}$ such that

$$\Delta=\sum_{k=0}^{N-1}\lambda^{(k)}\omega_k.$$

We then define the set

$$\Xi:=\Big\{\Delta\in\mathcal{O}(I_N):\sum_{k=0}^{N-1}|\lambda^{(k)}|=1\Big\}, \tag{4.6}$$

and notice that Ξ is not empty, as we have that $\omega_{\frac{N}{2}}=(-1,+1,-1,+1,-1,\dots,+1)\in\Xi$. Finally, since both TV and $\mathbb{F}_p$ are 1-homogeneous functions, we rewrite (4.4) as

$$C_L(\theta)=\theta\cdot \inf_{\substack{\Delta\in\mathcal{O}(I_N):\\ \Delta\neq 0}} \frac{\mathbb{F}_p\left(\frac{\Delta}{\sum|\lambda^{(k)}|}\right)}{TV\left(\frac{\Delta}{\sum|\lambda^{(k)}|}\right)}\frac{\sum|\lambda^{(k)}|}{\sum|\lambda^{(k)}|}=\theta\cdot\inf_{\Delta\in\Xi}\frac{\mathbb{F}_p(\Delta)}{TV(\Delta)}. \tag{4.7}$$

We now state the main result of the section.

THEOREM 4.1. *The lower tight bound $C_L(\theta)$ is given by*

$$C_L(\theta)=2\theta N^{-p}, \tag{4.8}$$

and is attained at $\omega_{\frac{N}{2}}$.

Proof. To prove the theorem, we show that $\omega_{\frac{N}{2}}$ both minimizes the Fourier Discrepancy and maximizes the Total Variation over the set Ξ. This is enough to conclude $C_L(\theta)=\theta\frac{\mathbb{F}_p(\omega_{\frac{N}{2}})}{TV(\omega_{\frac{N}{2}})}$ which, through a simple computation, proves (4.8). For the sake of clarity, we divide the proof into two steps.

First step ($\omega_{\frac{N}{2}}$ maximizes TV over Ξ).
For any $\Delta\in\Xi$, we have

$$\begin{aligned}
TV(\Delta)&=TV\Big(\sum_{k=0}^{N-1}\lambda^{(k)}\omega_k\Big)=\frac{1}{2}\sum_{j=0}^{N-1}\Big|\sum_{k=0}^{N-1}\lambda^{(k)}(\omega_k)_j\Big|\\
&\leq\frac{1}{2}\sum_{j=0}^{N-1}\sum_{k=0}^{N-1}\Big|\lambda^{(k)}(\omega_k)_j\Big|=\frac{1}{2}\sum_{k=0}^{N-1}|\lambda^{(k)}|\sum_{j=0}^{N-1}|(\omega_k)_j|\\
&=\frac{N}{2}\sum_{k=0}^{N-1}|\lambda^{(k)}|=\frac{N}{2}.
\end{aligned}$$

We then conclude the first step of the proof by noticing that $TV(\omega_{\frac{N}{2}})=\frac{N}{2}$.

Second Step ($\omega_{\frac{N}{2}}$ *minimizes* $\mathbb{F}_p$ *over* Ξ). For any $j=0,\ldots,N-1$, the DFT of ω_j is given by

$$\widehat{(\omega_j)}_k=\sum_{l=0}^{N-1} e^{-i\frac{2\pi}{N}lk}(\omega_j)_l=\sum_{l=0}^{N-1} e^{-i\frac{2\pi}{N}l(k-j)}=N\delta_{k-j}.$$

From the linearity of the DFT, we infer

$$\widehat{\Delta}_k=\sum_{j=0}^{N-1}\lambda^{(j)}\widehat{(\omega_j)}_k=N\sum_{j=0}^{N-1}\lambda^{(j)}\delta_{k-j}=N\lambda^{(k)}, \tag{4.9}$$

therefore, for any $\Delta\in\mathcal{O}(I_N)$, we have

$$\mathbb{F}_p^2(\Delta)=N^2\left(\sum_{k=1}^{\frac{N}{2}-1}\frac{|\lambda^{(k)}|^2}{k^{2p}}+\frac{|\lambda^{(\frac{N}{2})}|^2}{|N|^{2p}}\right). \tag{4.10}$$

Finally, we conclude the proof by showing

$$\inf_{\Delta\in\Xi}\mathbb{F}_p(\Delta)=\mathbb{F}_p(\omega_{\frac{N}{2}})=N^{1-p}.$$

Let $\Delta\in\Xi$. From (4.9), we have that $\lambda^{(0)}=\frac{1}{N}\widehat{\Delta}_0=\frac{1}{N}\sum_j\Delta_j=0$. Moreover, since Δ is real, we have that $\widehat{\Delta}_k=\overline{\widehat{\Delta}_{N-k}}$ for any $k=1,\ldots,N-1$, hence $|\lambda^{(k)}|=|\lambda^{(N-k)}|$. Then, if we define

$$\gamma_j:=\begin{cases}2|\lambda^{(j)}| & j=1,\ldots,\frac{N}{2}-1,\\ |\lambda^{(\frac{N}{2})}| & j=\frac{N}{2},\end{cases}$$

the constraint (4.6) is written as

$$\sum_{j=1}^{\frac{N}{2}}\gamma_j=1,$$

while from (4.10) we obtain $\mathbb{F}_p^2(\Delta)=\sum_{k=1}^{\frac{N}{2}}\alpha_k\gamma_k^2$, with

$$\alpha_k:=\begin{cases}\left(\frac{N}{2}\right)^2k^{-2p} & k=1,\ldots,\frac{N}{2}-1,\\ N^{2-2p} & k=\frac{N}{2}.\end{cases}$$

Since the coefficient $\alpha_{\frac{N}{2}}$ is the lowest one, as long as $p\geq1$, the minimum of $\mathbb{F}_p$ is achieved when $\gamma_{\frac{N}{2}}=1$ and $\gamma_j=0$ for $j=1,\ldots,\frac{N}{2}-1$, and the proof is complete. ∎

4.2. Upper tight bound. We now show that it is possible to restrict the search space of the maximizer of (4.5) to a finite set with cardinality N. In particular, we prove that a similar restriction may be applied whenever we search for the upper tight bound between the Total Variation and any convex and $p-$homogeneous function of $\Delta\in\mathcal{O}(I_N)$. To accomplish that, we show that every $\Delta\in\mathcal{O}(I_N)$ can be written as a linear combination of simpler null-sum measures, namely $\eta_{i,j}$, defined as

$$\eta_{i,j}:=\delta_i-\delta_j,$$

for any $i,j\in\{0,\dots,N-1\}$ such that $i\neq j$. In particular, we have the following.

LEMMA 4.1. *Let Δ be a null-sum measure on I_N. Then, we can express Δ as $\Delta=TV(\Delta)\cdot\Delta'$, where Δ' is a convex combination of $\{\eta_{i_k,j_k}\}_k$ such that, for any $k\neq k'$, we have $i_k\neq j_{k'}$.*

This characterization allows us to restrict the set of possible maximizers of any convex and $p-$homogeneous function over the finite set $\{\eta_{i,j}\}_{i,j}$.

THEOREM 4.2. *Let $\mathbb{G}:\mathcal{O}(I_N)\to[0,+\infty)$ be a convex and $p-$homogeneous function. Then, there exist $i^\star,j^\star\in\{0,\dots,N-1\}$ such that, for any $\theta\in(0,1]$:*

$$\theta\cdot\eta_{i^\star,j^\star}=\underset{TV(\Delta)=\theta}{\operatorname{argmax}}\,\mathbb{G}(\Delta). \tag{4.11}$$

Proof. First, we notice that

$$(i^\star,j^\star):=\underset{i,j\in\{0,\dots,N-1\}}{\operatorname{argmax}}\mathbb{G}(\eta_{i,j}), \tag{4.12}$$

is well-defined as the maximum is taken over a finite set. Given any $\theta\in(0,1]$, let Δ be a null-sum measure such that $TV(\Delta)=\theta$. Lemma 4.1 allows us to write $\Delta=\theta\cdot\sum_k\lambda_k\eta_{i_k,j_k}$, with $\lambda_k\geq 0$ for any k and $\sum_k\lambda_k=1$.

Finally, from the p-homogeneity and the convexity of $\mathbb{G}$, we obtain:

$$\begin{aligned}\mathbb{G}(\Delta)&=\mathbb{G}\left(\theta\cdot\sum_k\lambda_k\eta_{i_k,j_k}\right)=\theta^p\cdot\mathbb{G}\left(\sum_k\lambda_k\eta_{i_k,j_k}\right)\\&\leq\theta^p\cdot\sum_k\lambda_k\mathbb{G}(\eta_{i_k,j_k})\leq\theta^p\cdot\sum_k\lambda_k\mathbb{G}(\eta_{i^\star,j^\star})\\&=\theta^p\cdot\mathbb{G}(\eta_{i^\star,j^\star})=\mathbb{G}(\theta\cdot\eta_{i^\star,j^\star}),\end{aligned}$$

which concludes the proof. ∎

Using the previous result we may recover the well-known upper tight bound between the l^p norm and the Total Variation. Indeed, since $||\eta_{i,j}||_p=2^{\frac{1}{p}}$ for any p, we find that the inequality $||\mu-\nu||_p\leq 2^{\frac{1}{p}}TV(\mu,\nu)$ is tight.

Since $\mathbb{F}_p:\mathcal{O}(I_N)\to[0,+\infty)$ is convex and $1-$homogeneous, we infer $C_U(\theta)=\theta\cdot\mathbb{F}_p(\eta_{i^\star,j^\star})$, for some $i^\star,j^\star\in\{0,\dots,N-1\}$. Therefore, to find the upper tight bound of $\mathbb{F}_p$ we only need to search over a finite set of points, which correspond to the differences between two Dirac's deltas. Since the DFT is linear, we have that $\widehat{\eta_{l,j}}=\mathcal{O}(I_N)_l-\Theta_j$, where $\Theta_k=\left(e^{i\frac{2\pi k}{N}0},e^{i\frac{2\pi k}{N}1},\dots,e^{i\frac{2\pi k}{N}(N-1)}\right)$ is the $k-$th column of the matrix Ω. Hence:

$$\mathbb{F}_p^2(\delta_l,\delta_j)=\mathbb{F}_p^2(\eta_{l,j})=\sum_{k=1}^{\frac{N}{2}-1}\frac{|(\Theta_l-\Theta_j)_k|^2}{|k|^{2p}}+\frac{|(\Theta_l-\Theta_j)_{\frac{N}{2}}|^2}{|N|^{2p}},$$

which boils down to

$$\mathbb{F}_p^2(\eta_{j,l})=\sum_{k=1}^{\frac{N}{2}-1}\frac{2-2\cos\left(\dfrac{2\pi|j-l|}{N}k\right)}{|k|^{2p}}+\frac{2-2\cos(\pi|j-l|)}{|N|^{2p}}, \tag{4.13}$$

for any $j,l\in\{0,\dots,N-1\}$. Finally, notice that $\mathbb{F}_p^2(\eta_{j,l})$ depends on j and l only through $d:=|j-l|$. Hence, we can further restrict to measures of the form $\eta_{0,d}$, with $d\in\{1,\dots,N-1\}$.

COROLLARY 4.1. *For every $p \geq 1$, there exists $d \in \{0,1,\dots,N-1\}$ such that*

$$C_U(\theta) = \theta \cdot \mathbb{F}_p(\eta_{0,d}).$$

Notice that, for any $d \in \{0,1,\dots,N-1\}$, we have $\mathbb{F}_p^2(\eta_{0,d}) = C - 2g_p(d)$, where C is a constant and $g_p : [0,N] \to \mathbb{R}$ is defined as:

$$g_p(d) := \sum_{k=1}^{\frac{N}{2}-1} \frac{\cos\left(\frac{2\pi d}{N} k\right)}{|k|^{2p}} + \frac{\cos(\pi d)}{|N|^{2p}}. \tag{4.14}$$

By studying the derivatives with respect to d, it is possible to show that $d^* = \frac{N}{2}$ is a local minimum for g_p. This leads us to the following open conjecture.

CONJECTURE 4.1. *For every $p \geq 1$ and $d \in \{0,1,\dots,N-1\}$, we have*

$$\mathbb{F}_p(\eta_{0,\frac{N}{2}}) \geq \mathbb{F}_p(\eta_{0,d}).$$

If our conjecture was true, we would have

$$C_U(\theta) = \theta \cdot \sqrt{\sum_{k=1}^{\frac{N}{2}-1} \frac{2 - 2(-1)^k}{|k|^{2p}} + \frac{2 - 2(-1)^{\frac{N}{2}}}{|N|^{2p}}}. \tag{4.15}$$

Notice that, for $p=1$, the value (4.15) converges to $\sqrt{\sum_{k=1}^{\infty} \frac{2-2(-1)^k}{k^2}} = \frac{\pi}{2}$ as $N \to \infty$.

We numerically verify that the conjecture is true for $p \in \{1, 1.5, 2\}$ and for any even N that ranges from 2 to 1000. In Figure 4.1, we report the graph of the function $d \to \mathbb{F}_p(\eta_{0,d})$ for $p \in \{1, 1.5, 2\}$ and $N \in \{10, 1000\}$.

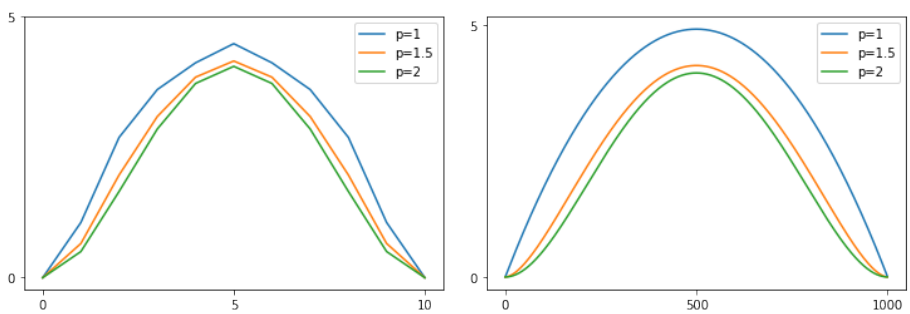

FIG. 4.1. *Plots of $\mathbb{F}_p(\eta_{0,d})$ for $p \in \{1, 1.5, 2\}$ and for $N=10$ (left), $N=1000$ (right). As conjectured, the maximum is attained at $d = \frac{N}{2}$.*

5. Conclusions and future work

In this paper, we introduced a new class of metrics between discrete probability measures, the p-Fourier Discrepancy Functions. For any $p \geq 1$, $\mathbb{F}_p$ is a well-defined distance induced by a bilinear form. It is convex, and its square is twice differentiable with explicit formulae for both the gradient and Hessian. Moreover, as Figure 4.1 shows, the Fourier Discrepancy between two Dirac's deltas depends on the distance between their supports. Most common discrepancies, such as the Total Variation or the Kullback-Leibler, do not enjoy this property, which is instead a feature of the Wasserstein distance. This is consistent with the bound (3.2) and with the equivalence between Fourier-based and Wasserstein distances [37]. In the last few years, the Wasserstein distance has been widely used in several applied fields because of its topological weakness and its ability to deal with the geometry of the underlying space [15]. However, its applicability, especially in higher dimensions, is bounded by the computational cost for both the distance and its gradient. On the other hand, the Fourier Discrepancy and its gradient are cheap to compute using the Fast Fourier Transform algorithm. We believe that the appealing properties of the Fourier Discrepancy make it a compelling alternative to the Wasserstein distance in several applied fields, such as machine learning [13, 41, 42], time series comparison [43], or barycenters computation [2, 4, 44]. Finally, the Fourier Discrepancy may be easily generalized to a multidimensional setting.

Acknowledgment. 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 funded by Sea Vision S.r.l..

Appendix A. Missing proofs.

Proof. **(Proof of Lemma 4.1.)** Let Δ be a null-sum measure. Without loss of generality, we can reorder the values of Δ as follows:

$$\Delta=(\alpha_1,\dots,\alpha_r,-\beta_1,\dots,-\beta_l,0,\dots,0),$$

where $r+l\leq N$, $\alpha_i,\beta_j>0$, $\alpha_i\leq\alpha_{i+1}$, $\beta_j\leq\beta_{j+1}$, for any i and j, and $\sum\alpha_i=\sum\beta_j$.

Without loss of generality, we assume that

$$\alpha_1\leq\beta_1.$$

Hence, we can write

$$\Delta=\alpha_1\eta_{0,r}+\Delta^{(1)},$$

where

$$\begin{aligned}\Delta^{(1)}&=(0,\alpha_2^{(1)},\dots,\alpha_r^{(1)},-\beta_1^{(1)},\dots,-\beta_l^{(1)},0,\dots,0)\\&:=(0,\alpha_2,\dots,\alpha_r,-(\beta_1-\alpha_1),-\beta_2,\dots,-\beta_l,0,\dots,0).\end{aligned}$$

Next, we compare $\alpha_2^{(1)}$ and $\beta_1^{(1)}$ and repeat the process until every entry vanishes. At the end, we find

$$\Delta=\lambda_1\eta_{0,r}+\dots+\lambda_k\eta_{r-1,N-1}=:\sum_k\lambda_k\eta_{i_k,j_k}. \tag{A.1}$$

Notice that each $\eta_{i,j}$ in (A.1) is such that $i<r$ and $j\geq r$ by construction, which implies $i\neq j$.

Since by hypothesis, for any $l=0,\dots,N-1$, all the l-th entries $(\eta_{i_k,j_k})_i$ have the same sign, we can write

$$|\Delta_l|=\Big|\sum_k\lambda_k(\eta_{i_k,j_k})_l\Big|=\sum_k\lambda_k|(\eta_{i_k,j_k})_l|.$$

Therefore:

$$TV(\Delta)=\frac{1}{2}\sum_l |\Delta_l| = \frac{1}{2}\sum_l\sum_k \lambda_k |(\eta_{i_k,j_k})_l|$$
$$=\frac{1}{2}\sum_k\sum_l \lambda_k |(\eta_{i_k,j_k})_l|$$
$$=\frac{1}{2}\sum_k \lambda_k \sum_l |(\eta_{i_k,j_k})_l| = \sum_k \lambda_k,$$

since $\sum_l |(\eta_{i,j})_l| = 2$ for any i,j. To conclude, it suffices to set

$$\Delta' := \frac{1}{TV(\Delta)}\Delta = \sum_k \widetilde{\lambda}_k \eta_{i_k,j_k},$$

where $\widetilde{\lambda}_k := \frac{\lambda_k}{\sum_l \lambda_l} > 0$, and $\sum_k \widetilde{\lambda}_k = 1$. ∎

Proof. **(Proof of Proposition 4.1.)** Let $C := \frac{\theta}{TV(\Delta)}$ and $\widetilde{\Delta} := C\cdot\Delta$, which are well-defined since $TV(\Delta)\neq 0$ for any non-zero Δ. Then, for the $1-$homogeneity of TV, we have that $TV(\widetilde{\Delta}) = \frac{\theta}{TV(\Delta)}\cdot TV(\Delta) = \theta$.

Let $\widetilde{\mu}$ and $\widetilde{\nu}$ be, respectively, the positive and negative part of $\widetilde{\Delta}$. Therefore, $\widetilde{\Delta} = \widetilde{\mu} - \widetilde{\nu}$ and $\widetilde{\mu}_i, \widetilde{\nu}_i \geq 0$ for any i. We have that

$$2\theta = \sum_i |\widetilde{\Delta}_i| = \sum_i \widetilde{\mu}_i + \sum_i \widetilde{\nu}_i, \tag{A.2}$$

and moreover, since $\widetilde{\Delta}$ is a null-sum measure:

$$0 = \sum_i \widetilde{\Delta}_i = \sum_i \widetilde{\mu}_i - \sum_i \widetilde{\nu}_i. \tag{A.3}$$

From (A.2) and (A.3), it follows easily that $\sum_i \widetilde{\mu}_i = \sum_i \widetilde{\nu}_i = \theta$.
We now define

$$\mu := \widetilde{\mu} + (1-\theta)\delta_0, \quad \nu := \widetilde{\nu} + (1-\theta)\delta_0.$$

We have that μ is a probability measure since $\mu_i \geq 0$ for any i and $\sum_i \mu_i = \sum_i \widetilde{\mu}_i + (1-\theta) = 1$. The same holds for ν. Moreover, $\mu - \nu = \widetilde{\Delta}$, hence $TV(\mu,\nu) = TV(\widetilde{\Delta}) = \theta$. ∎

Appendix B. Computing $\mathbb{F}_p(\eta_{j,l})$. Let us consider null-sum measures of the form $\eta_{l,j}$. We recall that $\eta_{l,j} := \delta_l - \delta_j$. Since

$$\widehat{\eta_{l,j}} = \Omega\cdot\eta_{l,j},$$

we have

$$\widehat{\eta_{l,j}} = \Theta_l - \Theta_j, \tag{B.1}$$

where Θ_k is the $k-$th column of the matrix Ω. By the definition of Ω we have

$$\Theta_l = \left(e^{i\frac{2\pi l}{N}0}, e^{i\frac{2\pi l}{N}1}, \ldots, e^{i\frac{2\pi l}{N}(N-1)}\right),$$

therefore, the value $\mathbb{F}_p^2(\eta_{l,j})$ is then given by

$$\mathbb{F}_p^2(\eta_{l,j}) = \sum_{k=1}^{\frac{N}{2}-1} \frac{|(\Theta_l - \Theta_j)_k|^2}{k^{2p}} + \frac{|(\Theta_l - \Theta_j)_{\frac{N}{2}}|^2}{|N|^{2p}}. \tag{B.2}$$

Let us now compute explicitly $|(\Theta_l - \Theta_j)_k|^2$ for a given k. We have

$$(\Theta_l - \Theta_j)_k = \cos\left(\frac{2\pi l}{N}k\right) - \cos\left(\frac{2\pi j}{N}k\right) + i\sin\left(\frac{2\pi l}{N}k\right) - i\sin\left(\frac{2\pi j}{N}k\right),$$

therefore,

$$\begin{aligned}
|(\Theta_l - \Theta_j)_k|^2 &= \left(\cos\left(\frac{2\pi l}{N}k\right) - \cos\left(\frac{2\pi j}{N}k\right)\right)^2 + \left(\sin\left(\frac{2\pi l}{N}k\right) - \sin\left(\frac{2\pi j}{N}k\right)\right)^2 \\
&= 2 - 2\left(\cos\left(\frac{2\pi l}{N}k\right)\cos\left(\frac{2\pi j}{N}k\right) + \sin\left(\frac{2\pi l}{N}k\right)\sin\left(\frac{2\pi j}{N}k\right)\right) \\
&= 2 - 2\cos\left(\frac{2\pi(j-l)}{N}k\right), \qquad \text{(B.3)}
\end{aligned}$$

where the equality in (B.3) comes from the following trigonometric identity:

$$\cos(\alpha - \beta) = \cos(\alpha)\cos(\beta) + \sin(\alpha)\sin(\beta).$$

Therefore,

$$\mathbb{F}_p^2(\eta_{j,l}) = \sum_{k=1}^{\frac{N}{2}-1} \frac{2 - 2\cos\left(\frac{2\pi|j-l|}{N}k\right)}{k^{2p}} + \frac{2 - 2\cos(\pi|j-l|)}{N^{2p}}. \tag{B.4}$$

REFERENCES

[1] S. Angenent, S. Haker, A. Tannenbaum, and L. Zhu, *Optimal mass transport for registration and warping*, Int. J. Comput. Vis., 60(3):225–240, 2004. 1

[2] F. Bassetti, S. Gualandi, and M. Veneroni, *On the computation of Kantorovich-Wasserstein distances between two-dimensional histograms by uncapacitated minimum cost flows*, SIAM J. Optim., 30(3):2441–2469, 2020. 1, 5

[3] G. Auricchio, F. Bassetti, S. Gualandi, and M. Veneroni, *Computing Wasserstein barycenters via linear programming*, in L.M. Rousseau and K. Stergiou (eds.), Integration of Constraint Programming, Artificial Intelligence, and Operations Research, Springer, Cham, 355–363, 2019. 1

[4] M. Cuturi and A. Doucet, *Fast computation of Wasserstein barycenters*, Int. Conf. Mach. Learn., 32(2):685–693, 2014. 1, 5

[5] N. Papadakis, *Optimal transport for image processing*, PhD thesis, Université de Bordeaux; Habilitation thesis, 2015. 1

[6] C.R. Vogel and M.E. Oman, *Iterative methods for total variation denoising*, SIAM J. Sci. Comput., 17(1):227–238, 1996. 1

[7] B. Fang, A. Guntuboyina, and B. Sen, *Multivariate extensions of isotonic regression and total variation denoising via entire monotonicity and Hardy–Krause variation*, Ann. Stat., 49(2):769–792, 2021. 1

[8] U. Ojha, Y. Li, J. Lu, A.A. Efros, Y.J. Lee, E. Shechtman, and R. Zhang, *Few-shot image generation via cross-domain correspondence*, Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 10743–10752, 2021. 1

[9] K. Janocha and W.M. Czarnecki, *On loss functions for deep neural networks in classification*, Schedae Inform., 25:49–59, 2016. 1

[10] Y. Bengio, I. Goodfellow, and A. Courville, *Deep Learning*, MIT Press, 2017. 1
[11] C.M. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006. 1
[12] J. Schmidhuber, *Deep learning in neural networks: An overview*, Neural Netw., 61:85–117, 2015. 1
[13] C. Frogner, C. Zhang, H. Mobahi, M. Araya-Polo, and T. Poggio, *Learning with a Wasserstein loss*, Proc. Int. Conf. Neur. Inf. Process. Syst., 28(2):2053–2061, 2015. 1, 5
[14] J. Yu, M. Wang, and D. Tao, *Semisupervised multiview distance metric learning for cartoon synthesis*, IEEE Trans. Image Process., 21(11):4636–4648, 2012. 1
[15] M. Arjovsky, S. Chintala, and L. Bottou, *Wasserstein generative adversarial networks*, Proc. Int. Conf. Mach. Learn., 34:214–223, 2017. 1, 5
[16] A.F. Ansari, J. Scarlett, and H. Soh, *A characteristic function approach to deep implicit generative modeling*, Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 7478–7487, 2020. 1
[17] C.-L. Li, W.-C. Chang, Y. Cheng, Y. Yang, and B. Póczos, *MMD GAN: Towards deeper understanding of moment matching network*, Proc. Int. Conf. Neur. Inf. Process. Syst., 31:2200–2210, 2017. 1
[18] Z. Wang, J. Li, and M. Enoh, *Removing ring artifacts in CBCT images via generative adversarial networks with unidirectional relative total variation loss*, Neural Comput. Appl., 31(9):5147–5158, 2019. 1
[19] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T.S. Huang, *Generative image inpainting with contextual attention*, Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 5505–5514, 2018. 1
[20] Z. Pan, W. Yu, B. Wang, H. Xie, V.S. Sheng, J. Lei, and S. Kwong, *Loss functions of generative adversarial networks (GANS): opportunities and challenges*, IEEE Trans. Emerg. Top. Comput. Intell., 4(4):500–522, 2020. 1
[21] G. Peyré and M. Cuturi, *Computational optimal transport: With applications to data science*, Found. Trends Mach. Learn., 11(5-6):355–607, 2019. 1
[22] D. Dvinskikh and D. Tiapkin, *Improved complexity bounds in Wasserstein barycenter problem*, International Conference on Artificial Intelligence and Statistics, PMLR, 1738–1746, 2021. 1
[23] R.E. Tarjan, *Dynamic trees as search trees via Euler tours, applied to the network simplex algorithm*, Math. Program., 78(2):169–177, 1997. 1
[24] Y. Nesterov, *Dual extrapolation and its applications to solving variational inequalities and related problems*, Math. Program., 109(2):319–344, 2007. 1
[25] J. Lin, *Divergence measures based on the Shannon entropy*, IEEE Trans. Inf. Theory, 37(1):145–151, 1991. 1
[26] N. Bonneel and D. Coeurjolly, *SPOT: Sliced Partial Optimal Transport*, ACM Trans. Graph., 38(4):1–13, 2019. 1
[27] H. Ling and K. Okada, *An efficient earth mover's distance algorithm for robust histogram comparison*, IEEE Trans. Pattern Anal. Mach. Intell., 29(5):840–853, 2007. 1
[28] G. Auricchio, F. Bassetti, S. Gualandi, and M. Veneroni, *Computing Kantorovich-Wasserstein distances on d-dimensional histograms using $(d+1)$-partite graphs*, Proc. Int. Conf. Neur. Inf. Process. Syst., 32:5798–5808, 2018. 1
[29] G.L. Gilardoni, *On the minimum f-divergence for given total variation*, Compt. Rend. Math., 343(11-12):763–766, 2006. 1
[30] I. Csiszár, *Two remarks to noiseless coding*, Inf. Control, 11(3):317–322, 1967. 1
[31] I. Csiszár, *Information-type measures of difference of probability distributions and indirect observation*, Stud. Sci. Math. Hung., 2:229–318, 1967. 1
[32] I. Sason, *Tight bounds for symmetric divergence measures and a refined bound for lossless source coding*, IEEE Trans. Inf. Theory, 61(2):701–707, 2014. 1, 4
[33] S.T. Rachev, *Probability Metrics and the Stability of Stochastic Models*, Wiley, 1991. 1
[34] S.T. Rachev, L.B. Klebanov, S.V. Stoyanov, and F. Fabozzi, *The Methods of Distances in the Theory of Probability and Statistics*, Springer, 10, 2013. 1
[35] J.A. Carrillo and G. Toscani, *Contractive probability metrics and asymptotic behavior of dissipative kinetic equations*, Riv. Mat. Univ. Parma, 7(6):75–198, 2007. 1
[36] L. Baringhaus and R. Grübel, *On a class of characterization problems for random convex combinations*, Ann. Inst. Stat. Math., 49(3):555–567, 1997. 1
[37] G. Auricchio, A. Codegoni, S. Gualandi, G. Toscani, and M. Veneroni, *The equivalence of Fourier-based and Wasserstein metrics on imaging problems*, Rend. Lincei-Mat. Appl., 31:627–649, 2020. 1, 3, 3.2, 5
[38] K.R. Rao and P.C. Yip, *The Transform and Data Compression Handbook*, CRC Press, 2018. 1, 2, 4.1
[39] C. Villani, *Optimal Transport: Old and New*, Springer Science & Business Media, 338, 2008. 3.2
[40] P.J. Davis, *Circulant Matrices*, Wiley, 1979. 3
[41] Y. Han, X. Liu, Z. Sheng, Y. Ren, X. Han, J. You, R. Liu, and Z. Luo, *Wasserstein loss-based deep*

object detection, Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 998–999, 2020. 5

[42] L. Hou, C.-P. Yu, and D. Samaras, *Squared earth movers distance loss for training deep neural networks on ordered-classes*, NIPS Workshop, 2017. 5

[43] Z. Zhang, P. Tang, and T. Corpetti, *Time adaptive optimal transport: A framework of time series similarity measure*, IEEE Access, 8:149764–149774, 2020. 5

[44] E. Anderes, S. Borgwardt, and J. Miller, *Discrete Wasserstein barycenters: Optimal transport for discrete data*, Math. Meth. Oper. Res., 84(2):389–409, 2016. 5