Towards an FFT for measures

Anonymous Author(s)
Affiliation
Address
email

Abstract

Complex measures recently became a well-established data model. We discuss the adaptation of the ubiquitous fast Fourier transform to measures, which involves their approximation by a multivariate trigonometric polynomial respecting normalization and non-negativity if applicable. The achieved approximation results, with respect to the Wasserstein-1 distance, are sharp up to logarithmic factors. The Fourier transform of atomic measures is shown to be computed up to logarithmic factors in linear time with respect to the problem size. The inverse Fourier transform is in general more involved but can be replaced by the easily computed approximation for typical applications.

1 Introduction

To quote from [12]: "These days, it is almost beyond belief that there was a time before digital technology... Much of this magic is due to a family of algorithms that collectively go by the name *fast Fourier transform*. Indeed the FFT is perhaps the most ubiquitous algorithm used today to analyze and manipulate digital or discrete data." Much of the success of the FFT is due to the fact that trigonometric polynomials well approximate smooth functions and that algorithms as well as their implementations are efficient.

During the last two decades and mainly driven by the specific applications, several new aspects came into focus: While data might live in high spatial dimensions it often has additional properties that allow for its approximation by tailored computational schemes. Primal classical examples being solutions of the electronic Schrödinger equation or multivariate kink functions, which both belong to function spaces with dominating mixed smoothness, see e.g. [14]. Such functions are well approximated by trigonometric polynomials with frequencies on a hyperbolic cross, and, together with a spatial discretization on a sparse grid, gave rise to a so-called hyperbolic cross FFT [7]. Even more general, compressed sensing and sparse expansions also come with several variants of FFTs, see e.g. [6].

Here, we are interested in yet another generalized FFT which operates on measures. At the current stage, we would like to impose no further restriction than the measure living on the d-dimensional periodic unit cube. We put some emphasis on singular measures, which includes discrete and singular continuous measures for $d > 1$. Our main object of study is a certain proxy for the measure which comes with an approximation guarantee, is easily computable, and seems useful in typical applications, for instance to estimate the Wasserstein distance between two measures. The overarching concept is to trade exactness for efficiency: instead of precise computations up to machine precision, the proposed methods guarantee a certain target accuracy.

Submitted to 35th Conference on Neural Information Processing Systems (NeurIPS 2021). Do not distribute.

2 Preliminaries

Let $d \in \mathbb{N}$ denote the spatial dimension and $|x-y| = \min_{k\in\mathbb{Z}^d} \|x-y+k\|_1$ the wrap-around 1-norm on $\mathbb{T}^d = [0,1)^d$, then a function has Lipschitz-constant one, $\varphi \in \mathrm{Lip}(\mathbb{T}^d)$, if $|f(x)-f(y)| \leq |x-y|$ for all $x, y \in \mathbb{T}^d$. Please note that replacing the 1-norm by another p-norm just restricts the class of functions with Lipschitz-constant one further. Throughout this paper, let μ, ν denote some complex measures on $\mathbb{T}^d$ with finite total variation and normalization $\mu(\mathbb{T}^d) = \nu(\mathbb{T}^d) = 1$. The Fourier coefficients of μ are given by

$$\hat{\mu}(k) = \int_{\mathbb{T}^d} \mathrm{e}^{-2\pi \mathrm{i} k x} \mathrm{d}\mu(x), \qquad k \in \mathbb{Z}^d,$$

and these are finite with $|\hat{\mu}(k)| \leq \|\mu\|_{\mathrm{TV}}$ and $\hat{\mu}(0) = 1$. Using the dual characterisation by Kantorovich-Rubinstein, the Wasserstein-1-distance of μ and ν is given by

$$W_1(\nu,\mu) = \inf_{\pi} \int_{\mathbb{T}^{2d}} |x-y| \mathrm{d}\pi(x,y) = \sup_{\mathrm{Lip}(\varphi)\leq 1} \left| \int_{\mathbb{T}^d} \varphi(x)\, \mathrm{d}(\nu-\mu)\,(x) \right|,$$

where the infimum is taken over all couplings π with marginals μ and ν, respectively. By slight abuse of notation, we also write $W_1(p,\mu)$ in case the measure ν has density p, i.e., $\mathrm{d}\nu(x) = p(x)\mathrm{d}x$.

Now let

$$F_n(x) = \sum_{k=-n}^{n} \left(1 - \frac{|k|}{n+1}\right) \mathrm{e}^{2\pi \mathrm{i} k x} = \frac{1}{n+1} \left(\frac{\sin(n+1)\pi x}{\sin \pi x}\right)^2$$

denote the univariate Fejér kernel and $F_n(x_1,\ldots,x_d) = F_n(x_1)\cdot\ldots\cdot F_n(x_d)$ the multivariate Fejér kernel, respectively. The main object of study now is the approximation

$$p_n(x) = (F_n * \mu)\,(x) = \int_{\mathbb{T}^d} F_n(x-y)\mathrm{d}\mu(y), \tag{1}$$

an example is given in the following Figure 1. A similar construction can be found in [11], which however differs with respect to the constructed approximation and the metric for the approximation error.

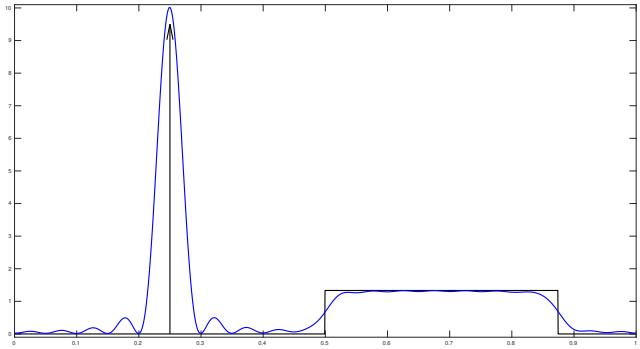

Figure 1: The measure $\mu = \frac{1}{2}\delta_{\frac{1}{4}} + \nu$ with $\mathrm{d}\nu = \frac{4}{3}\chi_{[\frac{1}{2},\frac{7}{8}]}\mathrm{d}\lambda$ and λ being the Lebesgue measure is approximated by the trigonometric polynomial density p_{19}.

3 Results

We start by noting that the suggested approximation preserves non-negativity and normalization.

Theorem 3.1. *Let $d, n \in \mathbb{N}$ and the measure μ be non-negative, then the finite moment matrix*

$$T_n := \left(\hat{\mu}(k-\ell)\right)_{k,\ell\in[n]}, \qquad [n] = \{0,\ldots,n\}^d, \tag{2}$$

is positive semi-definite. In particular, the approximation fulfills

$$p_n(x) = \frac{e_n(x)^* T_n e_n(x)}{(n+1)^d} \geq 0, \qquad e_n(x) = \left(\mathrm{e}^{2\pi \mathrm{i} k x}\right)_{k\in[n]}$$

and $\|p_n\|_{L^1} = \|\mu\|_{\mathrm{TV}} = 1$.

Proof. Let $q \in \mathbb{C}^{(n+1)^d}$, then direct computation shows

$$q^* T_n q = \sum_{k,l \in [n]} \overline{q_k} \left(\int_{\mathbb{T}^d} \mathrm{e}^{-2\pi \mathrm{i}(k-\ell)y} \mathrm{d}\mu(y) \right) q_\ell = \int_{\mathbb{T}^d} \left| \sum_{k \in [n]} q_k \mathrm{e}^{2\pi \mathrm{i} k y} \right|^2 \mathrm{d}\mu(y) \geq 0.$$

Choosing $q = e_n(x)$ yields the second claim and by interchanging the order of integration and noting that the value of the inner integral is independent of y also

$$\|p_n\|_{L^1} = \frac{1}{(n+1)^d} \int_{\mathbb{T}^d} \int_{\mathbb{T}^d} \left| \sum_{k \in [n]} \mathrm{e}^{2\pi \mathrm{i} k(y-x)} \right|^2 \mathrm{d}x \mathrm{d}\mu(y) = \mu(\mathbb{T}^d) = \|\mu\|_{\mathrm{TV}}.$$

□

Our next goal is a quantitative approximation result, for which we need the following preparatory lemma. This result can be found in qualitative form e.g. in [1, Lemma 1.6.4].

Lemma 3.2. *Let $n \in \mathbb{N}$, then we have*

$$\int_{\mathbb{T}} F_n(x)|x| \mathrm{d}x \leq \frac{\log(n+1)}{n}.$$

Proof. Using $F_n(x) \leq n+1$ and $F_n(x) \leq (4(n+1)x^2)^{-1}$, we obtain

$$2\int_0^{1/4n} x F_n(x) \mathrm{d}x + 2\int_{1/4n}^{1/2} x F_n(x) \mathrm{d}x \leq \frac{n+1}{16n^2} + \frac{\log(8n)}{2(n+1)} \leq \frac{\log(n+1)}{n}.$$

□

We note in passing, that a finer analysis allows for the estimate

$$\frac{\log(n+1)}{\pi^2 n} + \frac{c}{n} \leq \int_{\mathbb{T}} F_n(x)|x| \mathrm{d}x \leq \frac{\log(n+1)}{\pi^2 n} + \frac{C}{n}.$$

for some absolute constants $C, c \in \mathbb{R}$.

Theorem 3.3. *Let $d, n \in \mathbb{N}$, then*

$$W_1(p_n, \mu) \leq \frac{d \cdot \|\mu\|_{\mathrm{TV}} \cdot \log(n+1)}{n}$$

which is sharp up to a small multiplicative constant for $\mu = \delta_0$.

Proof. We compute

$$\begin{aligned}
W_1(p_n, \mu) &= \sup_{\mathrm{Lip}(\varphi) \leq 1} \left| \int_{\mathbb{T}^d} p_n(x)\varphi(x) \mathrm{d}x - \int_{\mathbb{T}^d} \varphi(y) \mathrm{d}\mu(y) \right| \\
&= \sup_{\mathrm{Lip}(\varphi) \leq 1} \left| \int_{\mathbb{T}^d} \left[\int_{\mathbb{T}^d} \prod_{s=1}^d F_n(x_s - y_s)\varphi(x) \mathrm{d}x - \varphi(y) \right] \mathrm{d}\mu(y) \right| \\
&\leq \sup_{\mathrm{Lip}(\varphi) \leq 1} \int_{\mathbb{T}^d} \int_{\mathbb{T}^d} \prod_{s=1}^d F_n(x_s) \left| \varphi(x+y) - \varphi(y) \right| \mathrm{d}x \mathrm{d}|\mu|(y) \\
&\leq \int_{\mathbb{T}^d} \int_{\mathbb{T}^d} \prod_{s=1}^d F_n(x_s) \sum_{\ell=1}^d |x_\ell| \mathrm{d}x \mathrm{d}|\mu|(y) \\
&= |\mu|(\mathbb{T}^d) \sum_{\ell=1}^d \int_{\mathbb{T}^d} \prod_{s=1}^d F_n(x_s)|x_\ell| \mathrm{d}x \\
&= d \cdot |\mu|(\mathbb{T}^d) \int_{\mathbb{T}} F_n(x)|x| \mathrm{d}x
\end{aligned}$$

which yields the result by applying Lemma 3.2. Regarding the sharpness, we see that the two inequalities become equalities for $\mu = \delta_0$ and $\varphi(x) = |x|$ and Lemma 3.2 is sharp up to a multiplicative constant. □

An almost matching lower bound is given as follows.

Theorem 3.4. *For any measure μ on $\mathbb{T}^d$ being not the Lebesgue measure, there is a constant $c > 0$ such that for all $n \in \mathbb{N}$ holds*

$$W_1(p_n, \mu) \geq \frac{c}{n+1}.$$

Proof. We rely on a nice relationship between the Wasserstein distance and a discrepancy as outlined in [4]. Let $\hat{h} \in \ell^2(\mathbb{Z}^d)$, $\hat{h}(k) \in \mathbb{R} \setminus \{0\}$, $\hat{h}(k) = \hat{h}(-k)$, and consider the reproducing kernel Hilbert space

$$H = \{\varphi \in L^2(\mathbb{T}^d) : \sum_{k \in \mathbb{Z}^d} |\hat{h}(k)|^{-2} |\hat{\varphi}(k)|^2 < \infty\}, \qquad \|\varphi\|_H^2 = \sum_{k \in \mathbb{Z}^d} |\hat{h}(k)|^{-2} |\hat{\varphi}(k)|^2.$$

Given two measures μ, ν, their discrepancy (which depends also on the space H) is defined by

$$\mathcal{D}(\mu, \nu) = \sup_{\|\varphi\|_H \leq 1} \left| \int_{\mathbb{T}^d} \varphi \, \mathrm{d}(\mu - \nu) \right|$$

and fulfills the geometric-arithmetic inequality

$$\begin{aligned}
\mathcal{D}(p_n, \mu)^2 &= \sum_{k \in \mathbb{Z}^d} |\hat{h}(k)|^2 |\hat{\mu}(k) - \hat{p}(k)|^2 \\
&= \sum_{\|k\|_\infty \leq n} |\hat{h}(k)|^2 \left| 1 - \prod_{\ell=1}^{d} \left(1 - \frac{|k_\ell|}{n+1} \right) \right|^2 |\hat{\mu}(k)|^2 + \sum_{\|k\|_\infty > n} |\hat{h}(k)|^2 |\hat{\mu}(k)|^2 \\
&\geq \sum_{\|k\|_\infty \leq n} |\hat{h}(k)|^2 \left| \frac{\|k\|_1}{d(n+1)} \right|^2 |\hat{\mu}(k)|^2 + \sum_{\|k\|_\infty > n} |\hat{h}(k)|^2 |\hat{\mu}(k)|^2 \\
&= \sum_{\|k\|_\infty \leq n} |\hat{h}(k)|^2 \left| \frac{\|k\|_1}{d(n+1)} \right|^2 |\hat{\mu}(k) - \hat{\lambda}(k)|^2 + \sum_{\|k\|_\infty > n} |\hat{h}(k)|^2 |\hat{\mu}_k - \hat{\lambda}(k)|^2 \\
&\geq \frac{1}{d^2(n+1)^2} \|h * (\mu - \lambda)\|_{L^2(\mathbb{T}^d)}^2
\end{aligned}$$

where $h(x) = \sum_{k \in \mathbb{Z}^d} \hat{h}(k) \mathrm{e}^{2\pi \mathrm{i} k x}$ and λ denotes the Lebesgue measure with $\hat{\lambda}(0) = 1$ and $\hat{\lambda}(k) = 0$ for $k \in \mathbb{Z}^d \setminus \{0\}$.

Our second ingredient is a Lipschitz estimate: If $\varphi \in H$ with $\|\varphi\|_H \leq 1$, then

$$\begin{aligned}
|\varphi(y) - \varphi(y+x)|^2 &= \left| \sum_{k \in \mathbb{Z}^d} \hat{\varphi}(k) \left(\mathrm{e}^{2\pi \mathrm{i} k y} - \mathrm{e}^{2\pi \mathrm{i} k (y+x)} \right) \right|^2 \\
&\leq \|\varphi\|_H^2 \sum_{k \in \mathbb{Z}^d} \left| \mathrm{e}^{2\pi \mathrm{i} k y} - \mathrm{e}^{2\pi \mathrm{i} k (y+x)} \right|^2 |\hat{h}(k)|^2 \\
&\leq 2 \left(K(0) - K(x) \right),
\end{aligned}$$

where $K(x) = \sum_{k \in \mathbb{Z}^d} |\hat{h}(k)|^2 \mathrm{e}^{2\pi \mathrm{i} k x} = (h * h)(x)$ denotes the so-called reproducing kernel of the space H. If this kernel is $K(x_1, \ldots, x_d) = h^{[4]}(x_1) \cdot \ldots \cdot h^{[4]}(x_d)$ for some univariate function

$h^{[4]} \in C^2(\mathbb{T})$, $\left(h^{[4]}\right)'(0) = 0$, we find by a telescoping sum and direct calculation

$$\begin{aligned} K(0) - K(x) &= \prod_{\ell=1}^{d} h^{[4]}(0) - \prod_{\ell=1}^{d} h^{[4]}(x_\ell) \\ &\leq \sum_{\ell=1}^{d} \left(h^{[4]}(0)^\ell \prod_{k=1}^{d-\ell} h^{[4]}(x_k) - h^{[4]}(0)^{\ell-1} \prod_{k=1}^{d-\ell+1} h^{[4]}(x_k) \right) \\ &\leq \sum_{\ell=1}^{d} \|h^{[4]}\|_\infty^{d-1} \left[h^{[4]}(0) - h^{[4]}(x_\ell) \right] \\ &\leq \frac{1}{2} \|h^{[4]}\|_\infty^{d-1} \left\| \left(h^{[4]}\right)'' \right\|_\infty |x|^2. \end{aligned}$$

To make a specific choice, let $a \in (0, \frac{1}{8})$ be some irrational number and set $h^{[2]} = \chi_{[-a,a]} * \chi_{[-a,a]}$ as the convolution of the indicator function on $[-a, a]$ with itself, $h^{[4]} = h^{[2]} * h^{[2]}$, and $h(x_1, \ldots, x_d) = h^{[2]}(x_1) \cdot \ldots \cdot h^{[2]}(x_d)$. Since the space of Lipschitz test functions is at least as large as the reproducing kernel Hilbert space, we derive

$$W_1(p_n, \mu) \geq \frac{1}{2} \cdot \left(\frac{\sqrt{3}}{4} \right)^{d-1} a^{1-\frac{3}{2}d} \mathcal{D}(p_n, \mu) \geq \frac{1 \cdot \left(\frac{\sqrt{3}}{4} \right)^{d-1} a^{1-\frac{3}{2}d}}{d(n+1)} \|h * (\mu - \lambda)\|_{L^2(\mathbb{T}^d)}.$$

Since a is irrational, we can directly see by Parseval's theorem that $\|h * (\mu - \lambda)\|_{L^2(\mathbb{T}^d)} = 0$ if and only if $\mu = \lambda$. For $\mu \neq \lambda$, we obtain the statement with a positive c depending on μ, a, and d. □

Remark 3.5. *Classical approximation theory offers more, which we shortly illustrate for the univariate case $d = 1$: As seen above, the Lebesgue measure is approximated by $F_n * \lambda = \lambda$ without any error. For the suggested approximation, we may thus ask how well a measure $\mathrm{d}\mu = w(x)\mathrm{d}x$ with smooth (non-negative) density might be approximated. If we choose the analytical density $w(x) = 1 + \cos(2\pi x)$, then $F_n * w(x) - w(x) = \cos(2\pi x)/(n+1)$ and by testing with the Lipschitz function $\varphi(x) = \cos(2\pi x)/(2\pi)$, we see that*

$$\begin{aligned} W_1(F_n * w, w) &= \sup_{\mathrm{Lip}(\varphi) \leq 1} \left| \int_{\mathbb{T}} (F_n * w(x) - w(x)) \, \varphi(x) \mathrm{d}x \right| \\ &\geq \frac{1}{2\pi(n+1)} \int_{\mathbb{T}} \cos^2(2\pi x) \mathrm{d}x = \frac{1}{4\pi(n+1)}. \end{aligned}$$

This phenomenon is known as saturation and might be cured by replacing the Fejér kernel by other kernels, i.e., the so-called Jackson kernel (being almost $F_{n/2}^2$) improves Theorem 3.3 by getting rid of the log-factor. Using kernels K_n with stronger localization and 'smoother' Fourier coefficients, e.g. higher powers of the Fejér kernel, allows to improve the rate beyond n^{-1} if the measure has a smooth density w. This can be seen from partial integration

$$\begin{aligned} W_1(K_n * w, w) &= \sup_{\mathrm{Lip}(\varphi) \leq 1} \left| \int_{\mathbb{T}} (K_n * \varphi(y) - \varphi(y)) \, w(y) \mathrm{d}y \right| \\ &= \sup_{\psi, \mathrm{Lip}(\psi') \leq 1} \left| \int_{\mathbb{T}} (K_n * \psi(y) - \psi(y)) \, w'(y) \mathrm{d}y \right| \end{aligned}$$

and the above arguments. However note that from a practical perspective, this asks for a-priori smoothness assumptions on the measure to choose a suitable kernel.

4 Fourier transforms and applications

We briefly discuss the actual computations and their complexities for atomic measures. This is followed by some sample applications of Fourier transforms and we end by numerical examples that confirm our error estimates.

4.1 Fourier transform of atomic measures

Let

$$\mu = \sum_{j=1}^{m} \lambda_j \delta_{x_j}$$

be given by its weights $\lambda_j \in \mathbb{C}$ and nodes $x_j \in \mathbb{T}^d$, $j = 1, \dots, m$, then the computation of the Fourier coefficients

$$\hat{\mu}(k) = \int_{\mathbb{T}^d} \mathrm{e}^{-2\pi \mathrm{i} k x} \mathrm{d}\mu(x) = \sum_{j=1}^{m} \lambda_j \mathrm{e}^{-2\pi \mathrm{i} k x_j}, \qquad |k| \le n,$$

is known as adjoint nonequispaced discrete Fourier transform. Its naive computation takes $\mathcal{O}(n^d \cdot m)$ floating point operations while a faster computation $\hat{\mu}_a$ takes $\mathcal{O}(n^d \log n + m |\log \varepsilon|^d)$ floating point operations and guarantees an accuracy

$$\max_{|k| \le n} |\hat{\mu}(k) - \hat{\mu}_a(k)| \le \varepsilon \sum_{j=1}^{m} |\lambda_j|,$$

see e.g. [8]. Efficient implementations are available `http://www.nfft.org`, including references to other packages, wrappers for julia and python, and a derivation of the fast algorithms.

4.2 Inverse Fourier transform

Given the Fourier coefficients

$$\hat{\mu}(k) = \int_{\mathbb{T}^d} \mathrm{e}^{-2\pi \mathrm{i} k x} \mathrm{d}\mu(x), \qquad |k| \le n,$$

of a complex measure μ, we suggest to compute the approximation

$$\tilde{p} := \left(p_n\left(\frac{j}{2n+1}\right)\right)_{|j| \le n}, \qquad p_n\left(\frac{j}{2n+1}\right) = \sum_{|k| \le n} w(k) \hat{\mu}(k) \mathrm{e}^{2\pi \mathrm{i} k j/(2n+1)}, \tag{3}$$

where $w_k := \prod_{s=1}^{d} (1 - |k_s|/(n+1))$ are the Fourier coefficients of the tensor product Fejér kernel. This computation employs one scaling and one fast Fourier transform and thus takes $\mathcal{O}(n^d \log n)$ floating point operations. The transform is invertible and comes with the approximation guarantee in Theorem 3.3.

For atomic measures, there is a large zoo of methods that compute or approximate the parameters of the measure, e.g., parametric methods like Prony's method, matrix pencil, ESPRIT, and MUSIC or non-parametric methods like TV-minimization, BLASSO. The support $\operatorname{supp} \mu = \{x_1, \dots, x_m\} \subset \mathbb{T}^d$ can e.g. be recovered perfectly by solving an eigenvalue problem for the moment matrix (2) as soon as $\operatorname{rank} T_n = m$. This is generically the case as soon as $n^d > m$, see e.g. [9, 5] for details. However note that in all cases, these methods are more expensive, ruling them out for large m or n.

4.3 Convolutions, marginals, and derivatives

All typical applications of Fourier transforms rely on the diagonalization of translation invariant linear operators. Such operators are convolutions, defined for two complex measures μ, ν spectrally via

$$\widehat{\mu * \nu}(k) = \hat{\mu}(k) \cdot \hat{\nu}(k), \qquad k \in \mathbb{Z}^d.$$

If the measures are represented as in (3), then an FFT followed by the multiplication of the Fourier coefficients yields the Fourier coefficients of the convolution product. For example, the partial derivatives is translation invariant and can be approximated via

$$\partial_x \mu = \partial_x \delta_0 * \mu \approx F_n' * \mu.$$

Similarly, let $\mu \in \mathcal{M}(\mathbb{T}^{d+d'})$, then the marginal $\mu_1(A) := \mu(A \times \mathbb{T}^{d'})$, $A \subset \mathbb{T}^d$, has Fourier coefficients

$$\hat{\mu}_1(k) = \int_{\mathbb{T}^d} \mathrm{e}^{-2\pi \mathrm{i} k x} \mathrm{d}\mu_1(x) = \int_{\mathbb{T}^{d+d'}} \mathrm{e}^{-2\pi \mathrm{i} k x} \mathrm{e}^{-2\pi \mathrm{i} 0 y} \mathrm{d}\mu(x, y) = \hat{\mu}(k, 0)$$

and we hope that this proves useful in a fully spectral formulation of the Sinkhorn algorithm.

4.4 Accuracy

To illustrate the Wasserstein rate of Theorem 3.3, we implement the semidiscrete optimal transport between a discrete measure $\mu := \sum \lambda_j \delta_{x_j}$ and the corresponding polynomial density p_n defined in (1). To this end, we solve the dual optimal transport problem

$$\sup_{f,g \in \mathcal{C}(\mathbb{T}^d)} \sum_{j=1}^{m} \lambda_j f(x_j) + \int_{\mathbb{T}^d} g(y) p_n(y) \mathrm{d}y \quad \text{s.t.} \quad \forall x, y \in \mathbb{T}^d, \quad f(x) + g(y) \leq |x-y|.$$

It is known, see e.g. [10], that this problem may be equivalently formulated as the following finite dimensional optimization problem

$$\max_{f \in \mathbb{R}_+^m} \sum_{j=1}^{m} \lambda_j f_j + \sum_j \int_{L_j(f)} (|x_j - y| - f_j) p_n(y) \mathrm{d}y \tag{4}$$

where $L_j(f)$ are the so-called Laguerre cells associated with the weights f, given by

$$L_j(f) \overset{\text{def.}}{=} \left\{ y \in \mathbb{T}^d;\ \forall i \neq j,\ |x_j - y| - f_j \leq |x_i - y| - f_i \right\}.$$

In our experiments, we compute the Laguerre cells over a 500×500 grid, and use a BFGS scheme to solve the maximization (4). We use Mark Schmidt's implementation for this step [13]. We evaluate $W_1(p_n, \mu)$ for a collection of discrete measures with random positions x_j, random positive amplitudes λ_j such that $\lambda_{\max}/\lambda_{\min} \simeq 6$, and random sparsities $2 \leq s \leq 10$. Our results are displayed in Figure 2, and show that the empirical rate actually matches the upper bound we provide in Theorem 3.3 in that case.

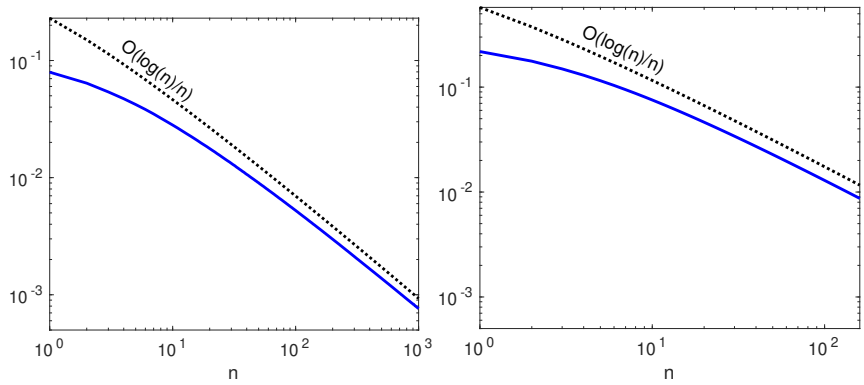

Figure 2: Average value of $W_1(p_n, \mu)$ over 100 random tests, in one (left) and two (right) dimensions.

4.5 Optimal transport

Given arbitrary measures μ, ν, we propose to approximate the Wasserstein distance $W_1(\mu, \nu)$ between μ and ν by $W_1^{(n)} := W_1(\tilde{p}_n, \tilde{q}_n)$, where $\tilde{p}_n = ((F_n * \mu)(x_j))_{j \in [N]}$ and $\tilde{q}_n = ((F_n * \nu)(x_j))_{j \in [N]}$ are the approximations of μ and ν respectively, evaluated on the grid points $x_j := N^{-1} j \in \mathbb{T}^d$, $j \in [N]$, for some integer $N \geq 2n+1$. The distance $W_1^{(n)}$ may be estimated accurately using the Sinkhorn algorithm [3], which consists in the alternation

$$\alpha^{(n)} = \tilde{p}_n \oslash K_\lambda \beta^{(n)}$$
$$\beta^{(n)} = \tilde{q}_n \oslash K_\lambda \alpha^{(n)},$$

where each iterates involve the updated value of $\alpha^{(n)}$ (resp, $\beta^{(n)}$), until some convergence criterion is met. Here the symbol $\oslash$ denotes pointwise division and the kernel matrix K_λ is defined as

$$(K_\lambda)_{ij} \overset{\text{def.}}{=} \exp(-\lambda |x_i - x_j|), \quad i, j \in [N],$$

where $\lambda > 0$ is the (entropic) regularization parameter, see [3] for more details. The approximate transport plan between the two polynomial densities $\tilde{p}_n$ and $\tilde{q}_n$ is then $\Pi^{(n)} := \mathrm{diag}(\alpha^{(n)}) K_\lambda \, \mathrm{diag}(\beta^{(n)})$, and the (debiased) Sinkhorn divergence approximates the Wasserstein distance, see e.g. [2].

Let Π be the optimal transport plan between μ and ν (for W_1). In our experiment, we set μ and ν to be discrete, so that in particular the coupling Π is also sparse. Figure 3 displays the evolution of $W_1(\Pi^{(n)}, \Pi)$ with n, computed using the semi-discrete procedure described in section 4.4. We scale the regularization parameter λ linearly in n, so that the problem becomes less and less regularized as the approximations $\tilde{p}_n$ and $\tilde{q}_n$ get sharper. At each step of the Sinkhorn algorithm, an estimate of the marginal $\tilde{p}_n$ (resp. $\tilde{q}_n$) is given by $\alpha^{(n)} \odot (K_\lambda \beta^{(n)})$ (resp. $\beta^{(n)} \odot (K_\lambda \alpha^{(n)})$. We use these estimates as a stopping criterion, ending the iterations when the mean relative error between $\tilde{p}_n$, $\tilde{q}_n$ and their estimates goes below 10^{-5}. Our results show that $\Pi^{(n)}$ converges towards Π empirically at the same rate than its marginals (up to multiplication by a scalar), see Figure 3.

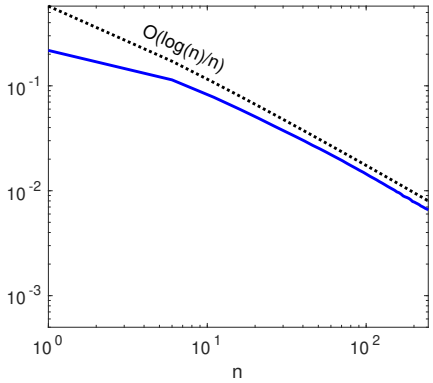

Figure 3: Average value of $W_1(\Pi^{(n)}, \Pi)$ over 50 random tests: the input marginals μ and ν are taken of same sparsity, randomly selected in $[\![2, 8]\!]$, with random amplitudes (with $\lambda_{\max}/\lambda_{\min} \simeq 6$) and random positions.

Since multiplication with K acts as convolution, we can formulate the Sinkhorn algorithm for arbitrary measures μ, ν by using the approximations $\tilde{p}$ from (3), analogously associate $\tilde{q}$ to ν and iterate on the Fourier coefficients of the dual potentials as

$$\hat{\alpha} = F_n \left(\tilde{p} \oslash F_n^{-1} D \hat{\beta} \right)$$
$$\hat{\beta} = F_n \left(\tilde{q} \oslash F_n^{-1} D \hat{\alpha} \right)$$

where D denotes some analytically known diagonal matrix. In particular, this easily allows to increase the parameter λ and n along the Sinkhorn iterations. We leave this study to future research.

5 Conclusion and outlooks

This paper introduced a simple proxy to approximate an arbitrary measure on the d-dimensional torus. We provided bounds on the approximation error that are sharp up to a logarithmic factor. Switching between trigonometric moments and spatial samples on a computational grid is done via the fast Fourier transform. For discrete measures with known parameters, the computation of the trigonometric moments is done by a nonequispaced FFT. The reconstruction of the parameters from these moments is costly so that we suggested to evaluate the proxy at the computational grid instead. As a proof of concept, we included a small example where an optimal transport plan between two measures is computed from their approximations and inherits their approximation error.

References

[1] P. L. Butzer and R. J. Nessel. *Fourier Analysis and Approximation I*. Birkhäuser Verlag, 1971.

[2] G. Carlier, V. Duval, G. Peyré, and B. Schmitzer. Convergence of entropic schemes for optimal transport and gradient flows. *SIAM J. Math. Anal.*, 49(2):1385–1418, 2017.

[3] M. Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in Neural Information Processing Systems*, 2013.

[4] M. Ehler, M. Gräf, S. Neumayer, and G. Steidl. Curve based approximation of measures on manifolds by discrepancy minimization. *Found. Comp. Math.*, to appear.

[5] M. Ehler, S. Kunis, T. Peter, and C. Richter. A randomized multivariate matrix pencil method for superresolution microscopy. *Electron. Trans. Numer. Anal.*, 51:63–74, 2019.

[6] A. Gilbert, P. Indyk, M. Iwen, and L. Schmidt. Recent developments in the sparse Fourier transform: A compressed Fourier transform for big data. *IEEE Signal Proc. Mag.*, 31(5):91–100, 2014.

[7] K. Hallatschek. Fouriertransformation auf dünnen Gittern mit hierarchischen Basen. *Numer. Math.*, 63:83–97, 1992.

[8] J. Keiner, S. Kunis, and D. Potts. Using NFFT 3 – a software library for various nonequispaced fast Fourier transforms. *ACM Trans. Math. Software*, 36(4):Art. 19, 30, 2009.

[9] S. Kunis, H. M. Möller, T. Peter, and U. von der Ohe. Prony's method under an almost sharp multivariate Ingham inequality. *J. Fourier Anal. Appl.*, 24(5):1306–1318, 2018.

[10] Q. Mérigot. A multiscale approach to optimal transport. *Comput. Graph. Forum*, 30(5):1583–1592, 2011.

[11] H. N. Mhaskar. Super-resolution meets machine learning: approximation of measures. *J. Fourier Anal. Appl.*, 25(6):3104–3122, 2019.

[12] D. Rockmore. The FFT: an algorithm the whole family can use. *Computing in Science Engineering*, 2(1):60–64, 2000.

[13] M. Schmidt. minFunc: Unconstrained differentiable multivariate optimization in Matlab. `http://www.cs.ubc.ca/~schmidt/Software/minFunc.html`, 2005.

[14] H. Yserentant. *Regularity and Approximability of Electronic Wave Functions*. Lecture Notes in Mathematics. Springer-Verlag, Berlin, 2010.