



Approximation and Interpolation of Singular Measures by Trigonometric Polynomials

Paul Catala¹ · Mathias Hockmann¹ · Stefan Kunis¹ · Markus Wageringel¹

Received: 28 March 2022 / Revised: 1 March 2024 / Accepted: 4 March 2024

© The Author(s) 2024

Abstract

Complex valued measures of finite total variation are a powerful signal model in many applications. Restricting to the d -dimensional torus, finitely supported measures can be exactly recovered from their trigonometric moments up to some order if this order is large enough. Here, we consider the approximation of general measures, e.g., supported on a curve, by trigonometric polynomials of fixed degree with respect to the 1-Wasserstein distance. We prove sharp lower bounds for their best approximation and (almost) matching upper bounds for effectively computable approximations when the trigonometric moments of the measure are known. A second class of sum of squares polynomials is shown to interpolate the indicator function on the support of the measure and to converge to zero outside.

Keywords Frequency analysis · Super resolution · Trigonometric approximation · Wasserstein distance

Mathematics Subject Classification 65T40 · 42A15

Communicated by Holger Rauhut.

✉ Stefan Kunis
stefan.kunis@uos.de

Paul Catala
paul.catala@uos.de

Mathias Hockmann
mathias.hockmann@uos.de

Markus Wageringel
markus.wageringel@uos.de

¹ Institute of Mathematics, Osnabrück University, Osnabrück, Germany

1 Introduction

Data science in general and more specifically signal and image processing relies on mathematical methods, with the fast Fourier transform as the most prominent example. Besides its favourable computational complexity, its success relies on the good approximation of smooth functions by trigonometric polynomials. Mainly driven by specific applications, functions with additional properties together with associated computational schemes have gained some attention: signals might for instance be sparse like in single molecule fluorescence microscopy [45], or live on some other lower dimensional structure like microfilaments, again in bio-imaging. Such properties are well modeled by measures, which can express the underlying structure through the geometry of their support, *e.g.* being discrete or singular continuous. This representation has in particular led to a better understanding of the sparse super-resolution problem [7, 11, 17], but has also proven useful in many more applications, such as phase retrieval in X-ray crystallography [3], or contour reconstruction in natural images [49]. In this work, we consider measures μ supported on the d -dimensional torus. The available data then consists of trigonometric moments of low to moderate order, *i.e.*

$$\hat{\mu}(k) = \int_{\mathbb{T}^d} e^{-2\pi i k x} d\mu(x), \quad k \in \{-n, \dots, n\}^d \tag{1.1}$$

for some $n \in \mathbb{N}$, and one asks for the reconstruction or approximation of μ from this partial information. In this context, our work focuses on approximations by trigonometric polynomials, more specifically on two types of asymptotic behaviours: after setting up some trigonometric polynomials q_n based on the knowledge of (1.1), we distinguish between pointwise convergence to the indicator function of $\text{supp } \mu$, *i.e.*

$$\lim_{n \rightarrow \infty} q_n(x) = \begin{cases} 1, & x \in \text{supp } \mu, \\ 0, & \text{else,} \end{cases}$$

and weak convergence, *i.e.*

$$\lim_{n \rightarrow \infty} \int_{\mathbb{T}^d} f(x) q_n(x) dx = \int_{\mathbb{T}^d} f(x) d\mu(x) \tag{1.2}$$

for all continuous test functions f . The latter is denoted by $q_n \rightharpoonup \mu$ and equivalent to convergence with respect to the Wasserstein distance for which we achieve quantitative rates.

Related work For discrete measures, there is a large variety of subspace methods that compute or approximate the parameters (positions and weights) of the measure, *e.g.*, Prony’s method [16, 30, 32, 53, 58], matrix pencil [17, 27, 46], ESPRIT [2, 40, 55, 57] or MUSIC [41, 60]. Except MUSIC, these methods realise the parameters as eigenvalues of specific moment matrices and are well understood in the univariate case [46]. In the multivariate case, an often used randomisation technique [17, 48] has only been discussed recently in a special case [23].

On the other hand, MUSIC [41, 60] as well as the variational methods [7, 9, 10, 54] build intermediate trigonometric polynomials which interpolate the value one at the support points and are smaller otherwise. If the measure is supported on a positive dimensional variety, the situation is more involved. Specific curves in a two-dimensional domain are identified by the kernel of moment matrices in [19, 49, 50], more general discussions can be found in [39, 62] where the support again is described implicitly by a trigonometric polynomial which takes the value one at the support and is smaller otherwise. Finally, Christoffel functions offer interesting guarantees both in terms of support identification [37] or approximation *on the support* [31, 43, 51], but, to the best of our knowledge, require regularity assumptions on the underlying measure, and only come with separate guarantees on and outside the support of the measure.

Contributions Following the approach of the seminal paper [44] to approximate a measure by using information about its trigonometric moments, we study easily computable trigonometric polynomials to approximate an arbitrary measure on the d -dimensional torus. In contrast to [44], we provide tight bounds on the pointwise approximation error as well as with respect to the 1-Wasserstein distance, the latter scaling inverse linearly with respect to the polynomial degree (up to a logarithmic factor). One of the main contributions of our work lies in the simple connection between approximation in the 1-Wasserstein distance and known results from approximation theory for Lipschitz functions. For example, we relate questions on the best approximation of measures by polynomials to best approximation results in $L^1(\mathbb{T}^d)$ and $C(\mathbb{T}^d)$. Additionally, we show analogously to classical approximation theory that near best approximations can be derived through convolution with certain kernels. As far as we know, these connections formulated in Sect. 3 where not considered before. On the other hand, we analyse in Sect. 4 the interpolation behaviour of a sum of squares polynomial, $p_{1,n}$, similarly suggested in [32, Thm. 3.5] and [49, Prop. 5.3] (and indeed closely related to the *rational* function in the MUSIC algorithm, see [60, Eq. (6)]). The main contribution of this section is not the invention of this polynomial $p_{1,n}$ but the analysis of its pointwise convergence to the indicator function of the Zariski closure of the support of the measure. This justifies to estimate the support of discrete measures or measures with support on an algebraic curve by considering points where $p_{1,n}$ is equal or close to its maximal value one. For instance, this might be used in future works to represent sparse objects in single molecule microscopy.

Organisation of the paper We summarize our main results in an informal way neither stating technical details nor making the involved constants explicit. After setting up the notations, Sect. 3 considers the approximation of measures by trigonometric polynomials. The convolution of the measure with polynomial kernels is studied in Sect. 3.1 and leads in Theorem 3.3 to the 1-Wasserstein upper bound

$$W_1(p_n, \mu) \leq \frac{c_1 \log n}{n},$$

where p_n denotes the convolution of the measure with the n -th Fejér kernel. This approximation comes with a saturation result in Sect. 3.2, Theorem 3.5, showing that

for each measure except the Lebesgue measure, there exists a constant c_2 with

$$W_1(p_n, \mu) \geq \frac{c_2}{n}.$$

This individual lower bound is complemented with a worst case lower bound for the best approximation p^* in Sect. 3.3, Theorem 3.6, showing

$$\sup_{\mu} W_1(p^*, \mu) \geq \frac{c_3}{n},$$

where the subsequent characterisation of the best approximation in the univariate case in Sect. 3.4, Theorem 3.9, and the following Example 3.10 show that the achieved constant c_3 is sharp.

In Sect. 4, we start by showing a specific sum of squares representation for p_n involving the moment matrix $(\hat{\mu}(k - \ell))_{k, \ell \in [n]}$. Setting all non-zero singular values in this representation to a constant yields the so-called signal polynomial $p_{1,n}$ which identifies the support of the measure in Sect. 4.1, Theorem 4.2, by

$$p_{1,n}(x) \begin{cases} = 1, & x \in \text{supp } \mu, \\ < 1, & \text{otherwise.} \end{cases}$$

As common to all subspace methods, this involves technical assumptions on the support of the measure and the degree n to be finite but large enough. For discrete measures the assumptions on the support are met and Sect. 4.2, Theorem 4.6, proves the pointwise convergence

$$p_{1,n}(x) \begin{cases} \leq 1 - c_4 n^2 \text{dist}(x, \text{supp } \mu)^2, & x \text{ close to } \text{supp } \mu, \\ \leq \frac{c_5}{n^2 \text{dist}(x, \text{supp } \mu)^2}, & \text{otherwise.} \end{cases}$$

A weak convergence result for discrete measures is proven in Theorem 4.9. Finally, Sect. 4.3, Theorem 4.10, proves pointwise convergence

$$p_{1,m+n}(x) \leq \frac{c_5(m, x)}{n}, \quad x \notin \text{supp } \mu,$$

also for positive dimensional support sets (which are generated in degree m). We end by illustrating the theoretical results by numerical examples in Sect. 5.

2 Preliminaries

Let $d \in \mathbb{N}$, $1 \leq p \leq \infty$ and let $|x - y|_p = \min_{k \in \mathbb{Z}^d} \|x - y + k\|_p$ denote the wrap-around p -norm on $\mathbb{T}^d = [0, 1)^d$. For $d = 1$ these wrap-around distances coincide and we denote them by $|x - y|_1$ to distinguish them from the absolute value. Throughout this paper, let μ, ν denote some complex Borel measures on \mathbb{T}^d with finite total variation

$\|\mu\|_{TV}$ and normalization $\mu(\mathbb{T}^d) = \nu(\mathbb{T}^d) = 1$.¹ This implies that the trigonometric moments as defined above are finite with $|\hat{\mu}(k)| \leq \|\mu\|_{TV}$ and $\hat{\mu}(0) = 1$. We denote the set of all such measures by \mathcal{M} and restrict to the real signed and nonnegative case by $\mathcal{M}_{\mathbb{R}}$ and \mathcal{M}_+ , respectively.

A function has Lipschitz-constant at most 1 if $f : \mathbb{T}^d \rightarrow \mathbb{C}$ admits $|f(x) - f(y)| \leq |x - y|_1$ for all $x, y \in \mathbb{T}^d$ and we denote this by the shorthand $\text{Lip}(f) \leq 1$. Using the dual characterisation by Kantorovich-Rubinstein, the 1-Wasserstein-distance of μ and ν is defined by

$$W_1(\nu, \mu) = \sup_{\text{Lip}(f) \leq 1} \left| \int_{\mathbb{T}^d} f(x) d(\nu - \mu)(x) \right|,$$

for any $\mu, \nu \in \mathcal{M}$. If $\mu, \nu \in \mathcal{M}_+$, this distance also admits the primal formulation

$$W_1(\nu, \mu) = \inf_{\pi} \int_{\mathbb{T}^{2d}} |x - y|_1 d\pi(x, y)$$

where the infimum is taken over all couplings π with marginals μ and ν , see e.g. [24, 52]. We note in passing that the 1-Wasserstein-distances for other p norms on \mathbb{T}^d are equivalent with lower and upper constant 1 and $d^{1-1/p}$, respectively. Moreover, the 1-Wasserstein distance defines a metric induced by the norm

$$\|\mu\|_{\text{Lip}^*} = \sup_{f: \text{Lip}(f) \leq 1, \|f\|_{\infty} \leq \frac{d}{2}} \left| \int_{\mathbb{T}^d} f(x) d\mu(x) \right|$$

which makes the space of complex-valued Borel measures with finite total variation a Banach space. By slight abuse of notation, we also write $W_1(p, \mu)$ in case the measure ν has density p , i.e., $d\nu(x) = p(x)dx$. Using the trigonometric moments from (1.1), we compute in Sect. 3 trigonometric approximations $q_n \in \mathcal{P}_n$ to the underlying measure, where we define the set \mathcal{P}_n of trigonometric polynomials with max-degree n as

$$\mathcal{P}_n = \left\{ p : \mathbb{T}^d \rightarrow \mathbb{C}, \quad x \mapsto p(x) = \sum_{k \in \{-n, \dots, n\}^d} \hat{p}(k) e^{2\pi i k x} \text{ for } \hat{p}(k) \in \mathbb{C} \right\}.$$

In Sect. 4, we additionally consider causal trigonometric polynomials where the coefficients of the polynomial are only nonzero at the nonnegative frequencies, i.e. for $k \in \{0, \dots, n\}^d$.

3 Approximation

We study in this section weakly convergent polynomial approximations of measures, i.e. approximations satisfying the property (1.2). The 1-Wasserstein distance (along

¹ The result that any complex measure has finite total variation can be found in [56, Thm. 6.4].

with all the Wasserstein distances) metrizes this notion of convergence for measures with equal mass [61, Thm.6.9], which allows us to provide both upper and lower bounds on the rates of convergence with respect to this distance of our constructions.

While our focus is principally on actually computable approximations, based on convolution with known kernels, we also turn in the last part of this section (Sect. 3.3 below) to more theoretical considerations on the best polynomial approximations with respect to the 1-Wasserstein distance, which, additionally to giving new perspectives on polynomial approximations of measures, also highlights the near-optimality of our constructions.

3.1 Approximation by Convolution and Upper Bounds

Similarly to standard approaches in approximation theory, one may derive easy-to-compute polynomial estimates for a measure μ , by considering the convolution of the latter with adequate kernels. For instance, given the first trigonometric moments of μ , the Fourier partial sums

$$S_n\mu(x) = (D_n * \mu)(x) = \sum_{k \in \mathbb{Z}^d, \|k\|_\infty \leq n} \hat{\mu}(k)e^{2\pi ikx},$$

which correspond to convolution with Dirichlet kernels², might serve as a sequence of approximations.

We focus in this section on yet another classical sequence of approximations, given by convolution with Fejér kernels $F_n : \mathbb{T}^d \rightarrow \mathbb{R}$ (by slight abuse of notation, we use the same notation for both the multivariate and univariate kernels), defined for $x = (x_1, \dots, x_d) \in \mathbb{T}^d$ as

$$F_n(x_1, \dots, x_d) = \prod_{i=1}^d F_n(x_i)$$

where, for any $x \in \mathbb{T}$,

$$\begin{aligned} F_n(x) &= \sum_{k=-n}^n \left(1 - \frac{|k|}{n+1}\right) e^{2\pi ikx} = \frac{1}{n+1} \left(\frac{\sin((n+1)\pi x)}{\sin(\pi x)}\right)^2 \\ &= \frac{1}{n+1} \left| \sum_{k=0}^n e^{2\pi ikx} \right|^2. \end{aligned} \tag{3.1}$$

The main object of study in this section is the trigonometric polynomial

$$p_n(x) = (F_n * \mu)(x) = \int_{\mathbb{T}^d} F_n(x-y) d\mu(y). \tag{3.2}$$

² On the univariate torus, the Dirichlet kernel is $D_n(x) = 1 + 2 \sum_{k=1}^n \cos(2\pi kx)$ and its multivariate version is given by $D_n(x_1, \dots, x_d) = D_n(x_1) \cdots D_n(x_d)$.

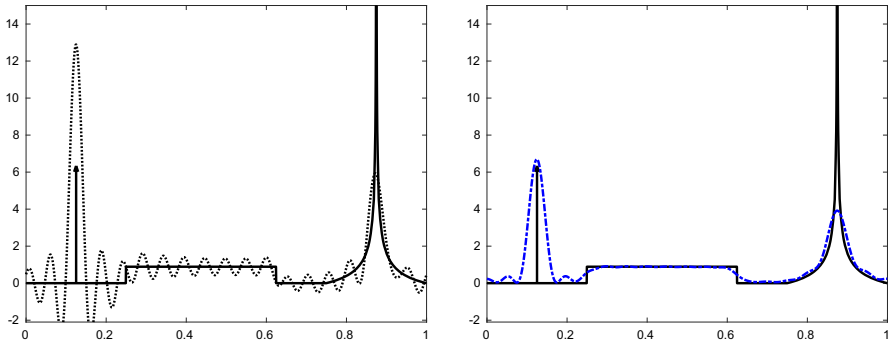


Fig. 1 The example measure (3.3), its approximations by the Fourier partial sum (left) and the convolution with the Fejér kernel (right). The weight $\frac{1}{3}$ of the Dirac measure in $\frac{1}{8}$ is displayed by an arrow of height $n/3$ for visibility

We give two illustrative examples in Example 3.1.

Example 3.1 Our first example for $d = 1$ is the measure

$$\mu = \frac{1}{3}\delta_{\frac{1}{8}} + \nu \in \mathcal{M}_+, \quad \frac{d\nu}{d\lambda}(x) = \frac{8}{9}\chi_{[\frac{1}{4}, \frac{5}{8}]}(x) + \frac{\sqrt{2}}{3} \left(\frac{1}{\sqrt{|x - \frac{7}{8}|}} - \sqrt{8} \right) \chi_{[\frac{3}{4}, 1]}(x), \quad (3.3)$$

where λ denotes the Lebesgue measure. Obviously, this measure μ has singular and absolutely continuous parts including an integrable pole at $x = \frac{7}{8}$.

Both the Fourier partial sums and the Fejér approximations for $n = 19$ are shown in the left and right panel of Fig. 1, respectively.

Our second example is a singular continuous measure for $d = 2$. We take $\mu = (2\pi r_0)^{-1}\delta_C \in \mathcal{M}_+$ as the uniform measure on the circle

$$C = \{x \in \mathbb{T}^2 : |x|_2 = r_0\}$$

for some radius $0 < r_0 < \frac{1}{2}$. The total variation of this measure is

$$\|\mu\|_{TV} = \hat{\mu}(0) = \int_{\mathbb{T}^2} d\mu(x) = \frac{1}{2\pi r_0} \int_C dx = 1.$$

Using a well-known representation of the Fourier transform of a radial function (cf. [22, p. 574]), we find

$$\hat{\mu}(k) = \int_{\mathbb{T}^2} e^{-2\pi i k x} d\mu(x) = \frac{1}{r_0} \int_0^\infty r J_0(2\pi r \|k\|_2) d\delta_{r_0}(r) = J_0(2\pi r_0 \|k\|_2) \quad (3.4)$$

for the trigonometric moments of μ , where J_0 denotes the 0-th order Bessel function of the first kind. These decay asymptotically with rate $\|k\|_2^{-1/2}$, cf. [22, Appendix B. 8]. The Fourier partial sum as well as the convolution with the Fejér kernel for $n = 29$

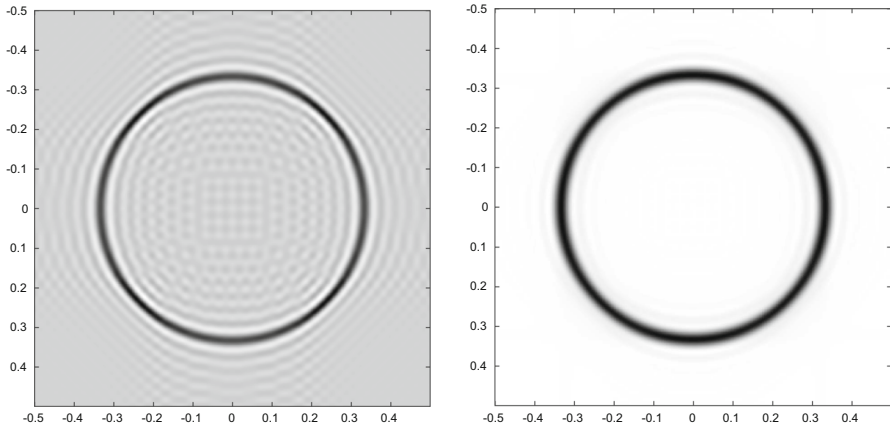


Fig. 2 Uniform measure on a circle of radius $r_0 = \frac{1}{3}$ and its approximations by the Fourier partial sum (left) and the convolution with the Fejér kernel (right), $n = 29$

are shown with maximal contrast in the left and right panel of Fig. 2, respectively. We observe that both approximators peak around the support of the measure and the approximation by convolution with the Fejér kernel produces less ringing than the Dirichlet kernel at the cost of a slightly thicker main lobe.

Of course, the construction and efficient evaluation of this approximation by p_n relies on the convolution theorem and the fast Fourier transform (FFT). Given the trigonometric moments $\hat{\mu}(k), k \in \{-n, \dots, n\}^d$, we multiply these with the Fourier coefficients of the Fejér kernel (3.1) in each dimension and use an inverse FFT to evaluate p_n on the equispaced grid $(2n + 1)^{-1}\{-n, \dots, n\}^d$. 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 [5, Lemma 1.6.4].

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

$$\frac{d}{\pi^2} \left(\frac{\log(n + 1)}{n + 1} + \frac{1}{n + 2} \right) \leq \int_{\mathbb{T}^d} F_n(x) |x|_1 dx \leq \frac{d}{\pi^2} \frac{\log(n) + 4}{n + 1}.$$

Proof First note that

$$\int_{\mathbb{T}^d} \prod_{s=1}^d F_n(x_s) \sum_{\ell=1}^d |x_\ell|_1 dx = \sum_{\ell=1}^d \int_{\mathbb{T}^d} \prod_{s=1}^d F_n(x_s) |x_\ell|_1 dx = d \int_{\mathbb{T}} F_n(x) |x|_1 dx,$$

where the second equality holds since $\int F_n(x_s) dx_s = 1$. Thus it is sufficient to consider the univariate case. The representation $F_n(x) = 1 + 2 \sum_{k=1}^n \left(1 - \frac{k}{n+1}\right) \cos(2\pi kx)$ gives

$$\int_{\mathbb{T}} F_n(x) |x|_1 dx = 2 \int_0^{1/2} \left(x + 2 \sum_{k=1}^n \left(1 - \frac{k}{n+1}\right) \cos(2\pi kx) x \right) dx$$

$$\begin{aligned}
 &= 2 \left[\frac{1}{8} + \sum_{k=1}^n \frac{(-1)^k - 1}{2\pi^2 k^2} - \sum_{k=1}^n \frac{(-1)^k - 1}{2(n+1)\pi^2 k} \right] \\
 &= 2 \left[\frac{1}{8} - \sum_{j=0}^{\lfloor \frac{n-1}{2} \rfloor} \frac{1}{\pi^2 (2j+1)^2} + \sum_{j=0}^{\lfloor \frac{n-1}{2} \rfloor} \frac{1}{(n+1)\pi^2 (2j+1)} \right]
 \end{aligned}$$

since $\int_0^{1/2} \cos(2\pi kx)x dx = ((-1)^k - 1)/(4\pi^2 k^2)$. Using that $\sum_{j=0}^\infty \frac{1}{(2j+1)^2} = \frac{\pi^2}{8}$, we obtain

$$\begin{aligned}
 \int_{\mathbb{T}} F_n(x)|x|_1 dx &= 2 \left[\frac{1}{\pi^2} \sum_{j=\lfloor \frac{n+1}{2} \rfloor}^\infty \frac{1}{(2j+1)^2} + \frac{1}{(n+1)\pi^2} \sum_{j=0}^{\lfloor \frac{n-1}{2} \rfloor} \frac{1}{2j+1} \right] \\
 &\leq \frac{2}{\pi^2} \left[\frac{1}{(2\lfloor \frac{n+1}{2} \rfloor + 1)^2} + \int_{\lfloor \frac{n+1}{2} \rfloor}^\infty \frac{1}{(2y+1)^2} dy + \frac{1 + \int_0^{\lfloor \frac{n-1}{2} \rfloor} \frac{1}{2y+1} dy}{n+1} \right] \\
 &\leq \frac{\frac{2}{(2\lfloor \frac{n+1}{2} \rfloor + 1)} + 1}{(2\lfloor \frac{n+1}{2} \rfloor + 1)\pi^2} + \frac{2 + \log(n)}{(n+1)\pi^2} \leq \frac{\log(n) + 4}{\pi^2(n+1)}.
 \end{aligned}$$

The lower bound follows similarly by bounding the series from the previous calculation by integrals from below. □

Theorem 3.3 *Let $d, n \in \mathbb{N}$ and $\mu \in \mathcal{M}$, then the measure with density p_n converges weakly to μ with*

$$W_1(p_n, \mu) \leq \frac{d}{\pi^2} \frac{\log(n) + 4}{n + 1} \cdot \|\mu\|_{\text{TV}},$$

which is sharp since $\mu \in \mathcal{M}_+$ implies $\|\mu\|_{\text{TV}} = \mu(\mathbb{T}^d) = 1$ and

$$\sup_{\mu \in \mathcal{M}} W_1(p_n, \mu) \geq \frac{d}{\pi^2} \left(\frac{\log(n+1)}{n+1} + \frac{1}{n+2} \right).$$

Proof We compute

$$\begin{aligned}
 W_1(p_n, \mu) &= \sup_{\text{Lip}(f) \leq 1} |\langle F_n * \mu, f \rangle - \langle \mu, f \rangle| \\
 &= \sup_{\text{Lip}(f) \leq 1} |\langle \mu, F_n * f - f \rangle| \\
 &\leq \sup_{\text{Lip}(f) \leq 1} \int_{\mathbb{T}^d} \int_{\mathbb{T}^d} F_n(x) |f(y-x) - f(y)| dx d\mu(y) \\
 &\leq \|\mu\|_{\text{TV}} \int_{\mathbb{T}^d} F_n(x) |x|_1 dx,
 \end{aligned}$$

and note that both inequalities become equalities when choosing $\mu = \delta_0$ and $f(x) = |x|_1$. Applying Lemma 3.2 gives the result. In particular, we remark in passing that $W_1(F_n, \delta_0) = \int_{\mathbb{T}^d} F_n(x)|x|_1 dx$. \square

Remark 3.4 Similar to classical results, the log-factor can be removed by choosing another convolution kernel, which then however does not allow for the representation later found in Lemma 4.1. The Jackson kernel, see [28, pp. 2 ff.],

$$J_{2m-2}(x) = \frac{3}{m(2m^2 + 1)} \frac{\sin^4(m\pi x)}{\sin^4(\pi x)}, \quad m \in \mathbb{N},$$

has degree $n = 2m - 2$, is normalised and satisfies by using $\frac{\sin(m\pi x)}{\sin(\pi x)} \leq \min(m, \frac{1}{2x})$

$$\begin{aligned} \int_{\mathbb{T}} J_n(x)|x|_1 dx &\leq \frac{6}{m(2m^2 + 1)} \left[\int_0^{1/2m} m^4 x dx + \int_{1/2m}^\infty \frac{1}{16x^3} dx \right] \\ &\leq \frac{3m}{2(2m^2 + 1)} \leq \frac{3}{2(n + 2)}. \end{aligned}$$

Analogously to Theorem 3.3, we get

$$W_1(J_n * \mu, \mu) \leq \frac{3}{2} \frac{d \cdot \|\mu\|_{TV}}{n + 2}, \tag{3.5}$$

which still is an approximate factor 6 worse than the lower bound in the univariate case (see Theorem 3.6). By numerical analysis or more detailed analysis of the above estimate, one can deduce that a factor 3 is due to the above estimate and a factor 2 seems to indicate that the Jackson kernel is not optimal. Moreover, upper and lower bound differ by a factor d in the multivariate case which might be due to the used norms or our proof techniques.

3.2 Saturation

Theorem 3.3 gives a worst case lower bound while, on the other hand, the Lebesgue measure is approximated by $F_n * \lambda = \lambda$ without any error. We may thus ask how well a measure $d\mu = w(x)dx$ with smooth (nonnegative) density might be approximated. For an introductory example, consider the univariate analytical density $w(x) = 1 + \cos(2\pi x)$. Since $F_n * w(x) - w(x) = \cos(2\pi x)/(n + 1)$, we achieve by testing with the Lipschitz function $f(x) = \cos(2\pi x)/(2\pi)$ that

$$W_1(F_n * w, w) \geq \frac{1}{2\pi(n + 1)} \int_{\mathbb{T}} \cos^2(2\pi x) dx = \frac{1}{4\pi(n + 1)}.$$

This effect is called saturation (e.g. cf. [5]). In greater generality, such a lower bound holds for each measure individually and can be inferred by a nice relationship between the Wasserstein distance and a discrepancy, cf. [18].

Theorem 3.5 For each individual measure $\mu \in \mathcal{M}$ different from the Lebesgue measure, there is a constant $c > 0$ such that we have for all $n \in \mathbb{N}$

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

Proof 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

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

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

$$\mathcal{D}(\mu, \nu) = \sup_{\|f\|_H \leq 1} \left| \int_{\mathbb{T}^d} f \, d(\mu - \nu) \right| = \sup_{\|f\|_H \leq 1} \left| \sum_{k \in \mathbb{Z}^d} \frac{\hat{f}(k)}{\hat{h}(k)} \widehat{h(k)\mu - \nu}(k) \right| = \|\widehat{h} \cdot \widehat{\mu - \nu}\|_{\ell^2}$$

and fulfils by the geometric-arithmetic inequality³

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

where $h(x) = \sum_{k \in \mathbb{Z}^d} \hat{h}(k) e^{2\pi i k x}$ 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 that quantifies the Lipschitz constant of any $f \in H$ with $\|f\|_H \leq 1$. For such a function f , the Cauchy-Schwarz inequality together with $|e^{2\pi i k y} - e^{2\pi i k (y+x)}|^2 = 2(1 - \cos(2\pi k x))$ gives

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

³ Explicitly, we used $\prod_{\ell=1}^d \left(1 - \frac{|k_\ell|}{n+1}\right) \leq \left[d^{-1} \sum_{\ell=1}^d \left(1 - \frac{|k_\ell|}{n+1}\right) \right]^d = \left[1 - \frac{\|k\|_1}{d(n+1)} \right]^d \leq 1 - \frac{\|k\|_1}{d(n+1)}$.

$$\begin{aligned}
 &= \|f\|_H^2 \sum_{k \in \mathbb{Z}^d} (2 - 2 \cos(2\pi kx)) |\hat{h}(k)|^2 \\
 &\leq 2 (K(x, x) - K(x, 0)),
 \end{aligned}$$

where $K(x, y) = \sum_{k \in \mathbb{Z}^d} |\hat{h}(k)|^2 e^{2\pi i k(x-y)} = (h * h)(x - y)$ denotes the so-called reproducing kernel⁴ of the space H . If this kernel is $K(x, y) = h^{[4]}(x_1 - y_1) \cdot \dots \cdot h^{[4]}(x_d - y_d)$ for some nonnegative univariate function $h^{[4]} \in C^2(\mathbb{T})$ being maximal in zero (and thus $(h^{[4]})'(0) = 0$), we find by a telescoping sum and the Taylor expansion

$$\begin{aligned}
 K(x, x) - K(x, 0) &= \prod_{\ell=1}^d h^{[4]}(0) - \prod_{\ell=1}^d h^{[4]}(x_\ell) \\
 &= \sum_{\ell=1}^d \left(h^{[4]}(0)^\ell \prod_{k=1}^{d-\ell} h^{[4]}(x_k) - h^{[4]}(0)^{\ell-1} \prod_{k=1}^{d-\ell+1} h^{[4]}(x_k) \right) \\
 &= \sum_{\ell=1}^d \left(h^{[4]}(0)^{\ell-1} \left(h^{[4]}(0) - h^{[4]}(x_{d-\ell+1}) \right) \prod_{k=1}^{d-\ell} h^{[4]}(x_k) \right) \\
 &\leq \sum_{\ell=1}^d \|h^{[4]}\|_\infty^{d-1} \left[h^{[4]}(0) - h^{[4]}(x_{d-\ell+1}) \right] \\
 &\leq \frac{1}{2} \|h^{[4]}\|_\infty^{d-1} \left\| \left(h^{[4]} \right)'' \right\|_\infty |x|_2^2.
 \end{aligned}$$

To make a specific choice, let $a \in (0, \frac{1}{8})$ be some irrational number and set

$$h^{[2]}(x) = \sum_{k \in \mathbb{Z}} (\chi_{[-a,a]} * \chi_{[-a,a]})(x + k), \quad x \in \mathbb{T}$$

as the periodisation of the convolution of the indicator function of $[-a, a]$ with itself. Based on this, we set $h^{[4]} = h^{[2]} * h^{[2]}$, and $h(x_1, \dots, x_d) = h^{[2]}(x_1) \cdot \dots \cdot h^{[2]}(x_d)$, which yields a valid kernel.⁵ Consequently, we derive that $f \in H$ with $\|f\|_H \leq 1$ satisfies $\text{Lip}(f) \leq c'_{d,a}$ for some constant $c'_{d,a} > 0$ depending on the dimension d and

⁴ Note that the assumptions $\hat{h}(k) \in \mathbb{R} \setminus \{0\}$ and $\hat{h}(k) = \hat{h}(-k)$ lead to $K(x, y) = \sum_{k \in \mathbb{Z}^d} |\hat{h}(k)|^2 e^{2\pi i k(x-y)} = \sum_{k \in \mathbb{Z}^d} |\hat{h}(k)|^2 \cos(2\pi k(x - y))$ and in particular K is real valued.

⁵ Note that the Fourier coefficients of $h^{[2]}$ agree with the Fourier transform of $\chi_{[-a,a]} * \chi_{[-a,a]}$ evaluated at integers by the Poisson summation formula, and analogously this holds for $h^{[4]}$ and the higher order spline obtained by threefold convolution of $\chi_{[-a,a]}$ with itself. By choosing $a < \frac{1}{8}$, $h^{[2]}$ and $h^{[4]}$ agree with these compactly supported convolutions on $[-\frac{1}{2}, \frac{1}{2}]$. One immediately gets $\widehat{h^{[4]}}(k) \in \mathcal{O}(k^{-4})$ by the convolution theorem of the Fourier transform and this indeed yields $h^{[4]} \in C^2(\mathbb{T})$ by [22, Prop. 3.3.12 or Ex. 2.4.1] meaning that the choice of $h^{[4]}$ is compatible with our previous assumptions on it. In particular, $h^{[4]}$ is maximal in zero since $h^{[2]}$ is even and nonnegative. Moreover, we directly have summability of $|\hat{h}(k)|^2$ for $k \in \mathbb{Z}^d$ such that K is a valid kernel.

the parameter a . This allows to conclude

$$W_1(p_n, \mu) \geq c'_{d,a}{}^{-1} \mathcal{D}(p_n, \mu) \geq \frac{c'_{d,a}{}^{-1}}{d(n+1)} \|h * (\mu - \lambda)\|_{L^2(\mathbb{T}^d)} =: \frac{c}{n+1}$$

for some $c \in \mathbb{R}$. Since a is irrational, we can directly see⁶ 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 constant c depending on the measure μ , the constant a , and the spatial dimension d . □

3.3 Best Approximation and Lower Bounds

After observing upper (Sect. 3.1) and lower bounds on the approximation by $p_n = F_n * \mu$ for individual measures μ (Sect. 3.2), one might ask whether an approximation rate faster than $\mathcal{O}(n^{-1})$ is possible by some general polynomial approximation. The following theorem shows that the answer to this question is negative as the best approximation by a normalised polynomial only yields a $\mathcal{O}(n^{-1})$ worst-case rate.

Theorem 3.6 *For any $d, n \in \mathbb{N}$ and for every $\mu \in \mathcal{M}$ there exists a polynomial with degree n of best approximation in the 1-Wasserstein distance among all polynomials with degree n . Moreover, we have*

$$\sup_{\mu \in \mathcal{M}} \min_{\substack{p \in \mathcal{P}_n \\ \hat{p}(0)=1}} \frac{W_1(p, \mu)}{\|\mu\|_{\text{TV}}} \geq \frac{1}{4(n+1)}.$$

Proof We have existence of a best approximation by polynomials in the Banach space of Borel measures with finite total variation.⁷ For the lower bound, we compute

$$\begin{aligned} \sup_{\mu \in \mathcal{M}} \min_{\substack{p \in \mathcal{P}_n \\ \hat{p}(0)=1}} W_1(p, \mu) &\geq \min_{\substack{p \in \mathcal{P}_n \\ \hat{p}(0)=1}} W_1(p, \delta_0) \\ &= \min_{\substack{p \in \mathcal{P}_n \\ \hat{p}(0)=1}} \sup_{f: \text{Lip}(f) \leq 1} \left| f(0) - \int_{\mathbb{T}^d} f(x) p(x) dx \right| \\ &= \min_{\substack{p \in \mathcal{P}_n \\ \hat{p}(0)=1}} \sup_{\substack{f: \|f\|_\infty \leq \frac{d}{2} \\ \text{Lip}(f) \leq 1}} \|f - \check{p} * f\|_\infty \\ &\geq \sup_{\substack{f: \|f\|_\infty \leq \frac{d}{2} \\ \text{Lip}(f) \leq 1}} \min_{p \in \mathcal{P}_n} \|f - p\|_\infty, \end{aligned}$$

⁶ We remark that $\hat{h}(k) = \prod_{\ell=1}^d \sin^2(2\pi k_\ell a) / (\pi^2 k_\ell^2) \neq 0$ for a irrational. Hence, $\|h * (\mu - \lambda)\|_{L^2(\mathbb{T}^d)} = 0$ implies by Parseval’s theorem that $\hat{\mu}(k) = \hat{\lambda}(k)$ for any $k \in \mathbb{Z}^d$. The latter is equivalent to $\mu = \lambda$.

⁷ For example, see [12, Thm. 3.1.1] and observe that the mentioned arguments also work for approximation by an affine linear subspace. The latter is needed here because of the constraint $\hat{p}(0) = 1$.

where we added a suitable constant to obtain the last equality⁸ and \check{p} denotes the reflection of p , i.e. $\check{p}(x) = p(-x)$ for all $x \in \mathbb{T}^d$. It remains to find the worst case error for the best approximation of a Lipschitz function by a trigonometric polynomial of degree n . While this is well-understood for $d = 1$ (cf. [1, 20]), we did not find a reference mentioning whether and how $d > 1$ is possible as well. Therefore, we show that the idea by [21] for the case $d = 1$ works also for $d > 1$ in our situation. A main ingredient of Fishers proof is the duality relation

$$\inf_{x \in Y \subset X} \|x_0 - x\| = \sup_{\substack{\ell \in X^* \\ \ell|_Y = 0, \|\ell\|_{X^*} \leq 1}} |\ell(x_0)|$$

for a Banach space X , $x_0 \in X$, with a subspace Y and dual space X^* . A second ingredient is given by the 1-periodic Bernoulli spline of degree 1, i.e.,⁹

$$\mathcal{B}_1(x) = \sum_{k \in \mathbb{Z} \setminus \{0\}} \frac{e^{2\pi i k x}}{2\pi i k} = \sum_{k=1}^{\infty} \frac{\sin(2\pi k x)}{\pi k} = \begin{cases} \frac{1}{2} - x, & x \in (0, 1), \\ 0, & x \in \{0, 1\}. \end{cases} \tag{3.6}$$

A Lipschitz continuous and 1-periodic function $f: \mathbb{T} \rightarrow \mathbb{R}$ with $\text{Lip}(f) \leq 1$ has a derivative f' almost everywhere and this derivative satisfies $\int_{\mathbb{T}} f'(s) = 0$ by the periodicity of f . Therefore, it follows that

$$\begin{aligned} (f' * \mathcal{B}_1)(t) &= \int_{\mathbb{T}} f'(s) \mathcal{B}_1(t - s) ds \\ &= - \int_0^t (t - s) f'(s) ds - \int_t^1 (t - s + 1) f'(s) ds \\ &= f(t) - \int_0^1 f(s) ds \end{aligned} \tag{3.7}$$

for $0 < t, s \leq 1$. The dual space of the space of continuous periodic functions is the space of periodic finite regular Borel measures equipped with the total variation norm and the duality formulation gives

$$\sup_{\substack{f: \mathbb{T}^d \rightarrow \mathbb{R} \\ \|f\|_{\infty} \leq \frac{d}{2}, \text{Lip}(f) \leq 1}} \min_{p \in \mathcal{P}_n} \|f - p\|_{\infty} = \sup_{\substack{f: \mathbb{T}^d \rightarrow \mathbb{R} \\ \|f\|_{\infty} \leq \frac{d}{2}, \text{Lip}(f) \leq 1}} \sup_{\substack{\hat{\mu}(k)=0, \|k\|_{\infty} \leq n \\ \|\mu\|_{\text{TV}} \leq 1}} \left| \int_{\mathbb{T}^d} f(x) d\mu(x) \right|.$$

Our main contribution to this result is the observation how to transfer the multivariate setting back to the univariate one. It is easy to verify that $f(x) = \frac{1}{d} \sum_{\ell=1}^d f_0(x_{\ell})$ for

⁸ By the constraint that $\int_{\mathbb{T}^d} p(x) dx = p_0 = 1$, we can add an arbitrary constant $c \in \mathbb{C}$ to f without changing the value of $f(0) - \int_{\mathbb{T}^d} f(x) p(x) dx$. By this, we can set $f(0) = 0$ and obtain $|f(x)| = |f(x) - f(0)| \leq |x|_1 \leq \frac{d}{2}$ such that we can restrict the supremum to all Lipschitz functions with $\|f\|_{\infty} \leq d/2$.

⁹ One can easily see that the Fourier series of $g(x) = \frac{1}{2} - x, x \in [0, 1)$, is given by the series in (3.6) By the Dirichlet-Jordan test, one directly obtains the convergence of the Fourier series towards $g(x)$ at $x \in (0, 1)$ and towards zero at the discontinuity point $x = 0$.

a univariate Lipschitz function f_0 , $\text{Lip}(f_0) \leq d$, $\|f_0\|_\infty \leq \frac{d}{2}$ fulfils the conditions for the Lipschitz function f . Additionally, $\mu^* = \frac{1}{d} \sum_{s=1}^d \mu_s$ with $\mu_s = \left(\otimes_{\ell \neq s} \lambda(x_\ell)\right) \otimes \mu_0^*(x_s)$,

$$\mu_0^*(x_s) = \frac{1}{2(n+1)} \sum_{j=0}^{2n+1} (-1)^j \delta_{j/(2n+2)}(x_s)$$

and λ being the Lebesgue measure on \mathbb{T} is admissible.¹⁰ Since this choice of μ_s integrates $\int g d\mu_s = 0$ if g is constant with respect to x_s (and the same holds for constant univariate functions integrated against μ_0^*), we obtain with (3.7)

$$\begin{aligned} \sup_{\substack{f: \mathbb{T}^d \rightarrow \mathbb{R} \\ \|f\|_\infty \leq \frac{d}{2}, \text{Lip}(f) \leq 1}} \min_{p \in \mathcal{P}_n} \|f - p\|_\infty &\geq \sup_{\substack{f_0: \mathbb{T} \rightarrow \mathbb{R} \\ \|f_0\|_\infty \leq \frac{d}{2}, \text{Lip}(f_0) \leq d}} \left| \frac{1}{d^2} \sum_{s, \ell=1}^d \int_{\mathbb{T}^d} f_0(x_\ell) d\mu_s(x) \right| \\ &= \sup_{\substack{f_0: \mathbb{T} \rightarrow \mathbb{R} \\ \|f_0\|_\infty \leq \frac{d}{2}, \text{Lip}(f_0) \leq d}} \left| \frac{1}{d^2} \sum_{\ell=1}^d \int_{\mathbb{T}} f_0(x_\ell) d\mu_0^*(x_\ell) \right| \\ &= \sup_{\substack{f_0: \mathbb{T} \rightarrow \mathbb{R} \\ \|f_0\|_\infty \leq \frac{d}{2}, \text{Lip}(f_0) \leq d}} \frac{1}{d} \left| \int_{\mathbb{T}} f_0'(s) \left(\int_{\mathbb{T}} \mathcal{B}_1(t-s) d\mu_0^*(t) \right) ds \right|. \end{aligned}$$

We denote $\mathcal{B}_{\mu^*}(s) = \int_{\mathbb{T}} \mathcal{B}_1(t-s) d\mu_0^*(t)$ and observe $\int_{\mathbb{T}} \mathcal{B}_{\mu^*}(s) ds = 0$. Moreover, μ_0^* has moments $\hat{\mu}_0^*(k) = 1$ for $k \in (n+1)(2\mathbb{Z}+1)$ and $\hat{\mu}_0^*(k) = 0$ otherwise. Together with the Fourier representation (3.6) of \mathcal{B}_1 where one rewrites the sum over odd integers as the difference between the sum over all nonzero integers and the sum of all nonzero even integers, this gives

$$\begin{aligned} \mathcal{B}_{\mu^*}(s) &= \sum_{k \in \mathbb{Z} \setminus \{0\}} \frac{e^{2\pi i k(n+1)s}}{2\pi i k(n+1)} - \sum_{k \in \mathbb{Z} \setminus \{0\}} \frac{e^{2\pi i 2k(n+1)s}}{2\pi i 2k(n+1)} \\ &= \frac{1}{n+1} \mathcal{B}_1((n+1)s) - \frac{1}{2n+2} \mathcal{B}_1((2n+2)s) \end{aligned}$$

¹⁰ Note that $\|\mu^*\|_{\text{TV}} \leq \frac{1}{d} \sum_{s=1}^d \|\lambda\|_{\text{TV}}^{d-1} \|\mu_0^*\|_{\text{TV}} = 1$ and

$$\begin{aligned} d \cdot \hat{\mu}^*(k) &= \sum_{s=1}^d \hat{\mu}_0^*(k_s) \prod_{\ell \neq s} \delta_{k_\ell, 0} \\ &= \sum_{s=1}^d \frac{1}{2(n+1)} \sum_{j=0}^{2n+1} e^{-2\pi i j \frac{n+1+k_s}{2n+2}} \prod_{\ell \neq s} \delta_{k_\ell, 0} = \sum_{s=1}^d \delta_{k_s, n+1+(2n+2)\mathbb{Z}} \prod_{\ell \neq s} \delta_{k_\ell, 0} = 0 \end{aligned}$$

for $\|k\|_\infty \leq n$. Within this calculation $\delta_{i,j}$ for indices $i, j \in \mathbb{Z}$ denotes the usual Kronecker delta being one if $i = j$ and zero if $i \neq j$.

$$= \begin{cases} \frac{1}{4} \frac{1}{n+1}, & (n+1)s - \lfloor (n+1)s \rfloor \in (0, \frac{1}{2}), \\ -\frac{1}{4} \frac{1}{n+1}, & (n+1)s - \lfloor (n+1)s \rfloor \in (\frac{1}{2}, 1), \\ 0, & (2n+2)s \in \{0, \dots, 2n+1\}. \end{cases}$$

Here, the last equality is a direct consequence of (3.6). Now, we choose f_0 by taking $f'_0(s) = d \cdot \text{sgn}(\mathcal{B}_{\mu^*}(s))$ and $f_0(0) = 0$ which is possible as it yields

$$\|f_0\|_\infty = \max_{x \in \mathbb{T}} |f_0(x) - f_0(0)| \leq \max_{x \in \mathbb{T}} d|x|_1 = \frac{d}{2} \quad \text{and} \quad \int_{\mathbb{T}} f'_0(s) ds = 0.$$

Finally, we end up with

$$\begin{aligned} \sup_{\substack{f: \mathbb{T}^d \rightarrow \mathbb{R} \\ \|f\|_\infty \leq \frac{d}{2}, \text{Lip}(f) \leq 1}} \min_{p \in \mathcal{P}_n} \|f - p\|_\infty &\geq \int_{\mathbb{T}} |\mathcal{B}_{\mu^*}(s)| ds \\ &= \frac{1}{n+1} \int_{\mathbb{T}} \left| \mathcal{B}_1((n+1)s) - \frac{1}{2} \mathcal{B}_1((2n+2)s) \right| ds \\ &= \frac{1}{n+1} \int_{\mathbb{T}} \left| \mathcal{B}_1(s) - \frac{1}{2} \mathcal{B}_1(2s) \right| ds = \frac{1}{4(n+1)} \end{aligned}$$

and this was the claim. □

Remark 3.7 (Information theoretic point of view) One should distinguish the above result on the best approximation by a polynomial with given degree n from the question of how well one can recover any measure given its low order trigonometric moments. While the polynomial approximation calculated in the framework of Theorem 3.6 is based on the knowledge of all moments, the latter information theoretic question would only consider the moments $\hat{\mu}(k)$ for $k \in \{-n, \dots, n\}^d$. A lower bound can be reformulated as the largest difference

$$\sup \left\{ W_1(\mu, \nu) : \mu, \nu \in \mathcal{M}, \hat{\nu}(k) = \hat{\mu}(k) \text{ for } k \in \{-n, \dots, n\}^d \right\} \quad (3.8)$$

between two measures, which have equal low order moments and cannot be distinguished by a recovery algorithm if no additional prior is known. If $\hat{\mu}$ and $\hat{\nu}$ are equal up to order n , then convolution with the Jackson kernel yields $J_n * \mu = J_n * \nu$, so that the triangle inequality for W_1 and Remark 3.4 give

$$W_1(\mu, \nu) \leq W_1(\mu, J_n * \mu) + W_1(\nu, J_n * \nu) \leq \frac{3d}{2} \frac{\|\mu\|_{\text{TV}} + \|\nu\|_{\text{TV}}}{n+2},$$

and thus (3.8) is at most of order $\mathcal{O}(n^{-1})$. This order is also optimal and this can be seen by choosing μ as the Lebesgue measure λ , ν being absolutely continuous with $d\nu(x_1, \dots, x_d) = [1 + \cos(2\pi(n+1)x_1)] d\lambda(x_1, \dots, x_d)$, and $f(x) = \cos(2\pi(n+1)x_1)$.

$1)x_1)/(2\pi(n + 1))$ in

$$\begin{aligned}
 W_1(\mu, \nu) &= \sup_{f: \text{Lip}(f) \leq 1} \int_{\mathbb{T}^d} f(x) \cos(2\pi(n + 1)x_1) dx \\
 &\geq \int_{\mathbb{T}^d} \frac{\cos^2(2\pi(n + 1)x_1)}{2\pi(n + 1)} dx = \frac{1}{4\pi(n + 1)}.
 \end{aligned}$$

This shows that the knowledge of the Fourier coefficients of a measure up to order n without any prior assumption on the measure only allows to approximate the measure with worst case error of order n^{-1} . This worst case error rate can be decreased if prior knowledge on the ground truth measure, e.g. sparsity (see [24]), is assumed.

3.4 Univariate Situation and Uniqueness of Best Approximation

On the univariate torus \mathbb{T} , the Wasserstein distance of two probability measures can be rewritten as a L^1 distance of their cumulative density functions (CDF) shifted by some constant depending on the measures, see [6]. We extend this to real signed measures belonging to $\mathcal{M}_{\mathbb{R}}$.

Lemma 3.8 (Wasserstein via CDF) *For any univariate $\mu, \nu \in \mathcal{M}_{\mathbb{R}}$, we have*

$$W_1(\mu, \nu) = \int_0^1 |\mu([0, x]) - \nu([0, x]) - c^*(\mu, \nu)| dx,$$

and $c^*(\mu, \nu) \in \mathbb{R}$ depends on μ, ν .

Proof For $\mu, \nu \in \mathcal{M}_+(\mathbb{T})$ this is [6, Thm. 3.7]. For $\mu, \nu \in \mathcal{M}_{\mathbb{R}}$, we can use the Jordan decomposition of any signed measure as a difference of nonnegative measures, in other words we write $\mu = \mu_+ - \mu_-, \nu = \nu_+ - \nu_-$ for $\mu_+, \mu_-, \nu_+, \nu_-$ being nonnegative measures on \mathbb{T} and rewrite

$$\begin{aligned}
 W_1(\mu, \nu) &= \sup_{f: \text{Lip}(f) \leq 1} \left| \int_{\mathbb{T}} f(x) [d\nu_+(x) + d\mu_-(x) - (d\nu_-(x) + d\mu_+(x))] \right| \\
 &= (\nu_+ + \mu_-)(\mathbb{T}) \sup_{f: \text{Lip}(f) \leq 1} \left| \int_{\mathbb{T}} f(x) \left[\frac{d\nu_+(x) + d\mu_-(x)}{(\nu_+ + \mu_-)(\mathbb{T})} - \frac{d\nu_-(x) + d\mu_+(x)}{(\nu_+ + \mu_-)(\mathbb{T})} \right] \right|
 \end{aligned} \tag{3.9}$$

and this allows to apply [6, Thm. 3.7].¹¹ This then gives

$$\begin{aligned}
 &(\nu_+ + \mu_-)(\mathbb{T}) \int_0^1 \left| \left(\frac{\nu_+ + \mu_- - (\nu_- + \mu_+)}{(\nu_+ + \mu_-)(\mathbb{T})} \right) ([0, x]) \right. \\
 &\quad \left. - c^* \left(\frac{\nu_+ + \mu_-}{(\nu_+ + \mu_-)(\mathbb{T})}, \frac{\nu_- + \mu_+}{(\nu_+ + \mu_-)(\mathbb{T})} \right) \right| dx
 \end{aligned}$$

¹¹ Note that by $0 = \nu(\mathbb{T}) - \mu(\mathbb{T}) = (\nu_+ + \mu_-)(\mathbb{T}) - (\mu_+ + \nu_-)(\mathbb{T})$ both measures in the integral in (3.9) are probability measures. Moreover, observe that $(\nu_+ + \mu_-)(\mathbb{T}) \geq \nu(\mathbb{T}) = 1 > 0$ and hence it is possible to normalise as stated.

$$= \int_0^1 |(v - \mu)([0, x]) - c^*(v, \mu)| dx$$

for the Wasserstein distance of μ and ν where the constant $c^*(v, \mu)$ depends again only on the two measures. □

The question of uniqueness of the best approximation can be equivalently characterised by the uniqueness of the best approximation in $L^1(\mathbb{T})$ and thus allows for the following theorem.

Theorem 3.9 (Best approximation in the univariate case) *If $d = 1$ and $\mu, \nu \in \mathcal{M}_{\mathbb{R}}$, we have*

$$W_1(\nu, \mu) = \inf_{c \in \mathbb{R}} \int_{\mathbb{T}} |(\mathcal{B}_1 * \nu)(t) - (\mathcal{B}_1 * \mu)(t) - c| dt \tag{3.10}$$

with \mathcal{B}_1 being the Bernoulli spline from (3.6). This allows to conclude that for any $n \in \mathbb{N}$, any real, normalised measure which does not give mass to atoms (i.e. $\mu(\{x\}) = 0$ for all $x \in \mathbb{T}$) admits a unique best approximation by a normalised polynomial of degree $n \in \mathbb{N}$ with respect to the 1-Wasserstein distance.

Proof Let $\mu, \nu \in \mathcal{M}_{\mathbb{R}}$ and \mathcal{B}_1 denote the Bernoulli spline of degree 1 from the proof of Theorem 3.6, then we have by (3.7)

$$\begin{aligned} W_1(\nu, \mu) &= \sup_{f: \text{Lip}(f) \leq 1} \left| \int_{\mathbb{T}} f(x) [d\nu(x) - d\mu(x)] \right| \\ &= \sup_{f: \text{Lip}(f) \leq 1} \left| \int_{\mathbb{T}} \int_{\mathbb{T}} f'(t) \mathcal{B}_1(x - t) [d\nu(x) - d\mu(x)] dt \right| \\ &= \sup_{f: \text{Lip}(f) \leq 1} \left| \int_{\mathbb{T}} f'(t) [(\mathcal{B}_1 * \nu)(t) - (\mathcal{B}_1 * \mu)(t)] dt \right|. \end{aligned}$$

Since the integral over f' is zero by the periodicity of f , any $c \in \mathbb{R}$ yields

$$\begin{aligned} \left| \int_{\mathbb{T}} f'(t) [(\mathcal{B}_1 * \nu)(t) - (\mathcal{B}_1 * \mu)(t)] dt \right| &= \left| \int_{\mathbb{T}} f'(t) [(\mathcal{B}_1 * \nu)(t) - (\mathcal{B}_1 * \mu)(t) - c] dt \right| \\ &\leq \inf_{c \in \mathbb{R}} \int_{\mathbb{T}} |(\mathcal{B}_1 * \nu)(t) - (\mathcal{B}_1 * \mu)(t) - c| dt. \end{aligned}$$

We proceed by computing explicitly

$$\begin{aligned} (\mathcal{B}_1 * \mu)(t) &= \int_{[0,t) \cup (t,1)} \mathcal{B}_1(t - x) d\mu(x) \\ &= \int_{[0,t)} \frac{1}{2} - (t - x) d\mu(x) + \int_{(t,1)} \frac{1}{2} - (t - x + 1) d\mu(x) \\ &= \left(\frac{1}{2} - t \right) (\mu([0, 1)) - \mu(\{t\})) \end{aligned}$$

$$\begin{aligned}
 & + \int_{[0,1)} x d\mu(x) - t\mu(\{t\}) - \mu([0, 1)) + \mu([0, t]) \\
 = & \frac{\mu([0, t]) + \mu([0, t])}{2} - \mu([0, 1)) \left(t + \frac{1}{2} \right) + \int_{[0,1)} x d\mu(x) \quad (3.11)
 \end{aligned}$$

for $t \in (0, 1)$ and

$$(\mathcal{B}_1 * \mu)(0) = \int_{[0,1)} x d\mu(x) - \frac{1}{2}\mu([0, 1)) + \frac{1}{2}\mu(\{0\}).$$

On the other hand, Lemma 3.8 and (3.11) yield

$$\begin{aligned}
 & \int_0^1 |(v - \mu)([0, x]) - c^*(v, \mu)| dx = W_1(v, \mu) \\
 & \leq \inf_{c \in \mathbb{R}} \int_0^1 \left| (v - \mu)([0, x]) - \frac{(v - \mu)(\{x\})}{2} - c \right| dx
 \end{aligned}$$

and thus equality (3.10) for measures that give mass to at most countably many atoms because in this case the set of x where the integrands from the upper and lower bounds disagree has Lebesgue measure zero. But in fact, this holds for every measure as the following argument shows.¹² At first, one might consider the case of a finite positive measure μ . For $n \in \mathbb{N}$, consider $N_n := \{x \in \mathbb{T} : \mu(\{x\}) \geq n^{-1}\}$ and observe that for any finite subset $N^* \subseteq N_n$

$$\infty > \mu(\mathbb{T}) \geq \mu(N^*) \geq \sum_{x \in N^*} \mu(\{x\}) \geq n^{-1} \cdot \#N^*$$

and hence $\#N^* \leq n \cdot \mu(\mathbb{T})$, which then implies that N_n is finite with $\#N_n \leq n \cdot \mu(\mathbb{T})$. Therefore, the set

$$N := \bigcup_{n=1}^{\infty} N_n = \{x \in \mathbb{T} : \mu(\{x\}) > 0\}$$

is countable and the general case follows by decomposing $\mu = \mu_+ - \mu_-$.

With this knowledge, the question of approximation of μ by p with degree n and $\hat{p}(0) = 1$ can be rewritten as

$$\min_{\substack{p \in \mathcal{P}_n \\ \hat{p}(0)=1}} W_1(p, \mu) = \min_{\substack{p \in \mathcal{P}_n \\ \hat{p}(0)=1}} \inf_{c \in \mathbb{R}} \int_{\mathbb{T}} |(\mathcal{B}_1 * p)(t) - (\mathcal{B}_1 * \mu)(t) - c| dt.$$

By the assumption of an atom-free measure μ we have that $\mathcal{B}_1 * \mu$ is continuous by (3.11), and hence there exists a unique best L^1 -approximation $\tilde{p} = \mathcal{B}_1 * p^* - c$ (see e.g. [12, Thm. 3.10.9]) which defines the unique best approximation p^* to μ uniquely by $\tilde{p} = \mathcal{B}_1 * p^* - c$ and the normalisation condition $\hat{p}^*(0) = 1$. □

¹² This argument was pointed out by a reviewer of this work.

Example 3.10 Uniqueness and non-uniqueness of L^1 approximation is discussed in some detail in [14, 47] and we note the following:

- (i) For $\mu = \frac{1}{2}\delta_0 - \frac{1}{2}\delta_{1/2} + \lambda \in \mathcal{M}_{\mathbb{R}}$ where λ is again the Lebesgue measure, one finds

$$(\mathcal{B}_1 * \mu)(t) = \frac{1}{2} \left(\mathcal{B}_1(t) - \mathcal{B}_1\left(t - \frac{1}{2}\right) \right) = \begin{cases} 0, & t = 0, \\ \frac{1}{4}, & t \in (0, \frac{1}{2}), \\ 0, & t = \frac{1}{2}, \\ -\frac{1}{4}, & t \in (\frac{1}{2}, 1). \end{cases}$$

For any normalised polynomial p , we have that the difference $\mathcal{B}_1 * p(t) - \mathcal{B}_1 * \mu(t)$ differs from $\int_0^t p(x)dx - \mu([0, t])$ by a constant except at the discontinuity points $t = 0, \frac{1}{2}$. But as they have Lebesgue measure zero, we can derive from Theorem 3.9

$$\min_{\substack{p \in \mathcal{P}_n \\ \tilde{p}(0)=1}} W_1(p, \mu) = \inf_{c \in \mathbb{R}} \int_{\mathbb{T}} |(\mathcal{B}_1 * p)(t) - (\mathcal{B}_1 * \mu)(t) - c| dt. \tag{3.12}$$

As proven in [47, Thm. 5.1], the function $\mathcal{B}_1 * \mu$ does not have a unique L^1 approximation for even n . Thus, μ does not admit a unique best approximation either.

- (ii) For $\mu = \delta_0$ one has $\mathcal{B}_1 * \mu = \mathcal{B}_1$ such that again (3.12) holds for this choice of μ . According to [47, Lem. 2.2] this function \mathcal{B}_1 with only one jump has a unique best L^1 -approximation given by the interpolation polynomial

$$\tilde{p}(x) = \sum_{j=1}^n \frac{1}{2n+2} \cot\left(\frac{j\pi}{2n+2}\right) \sin(2\pi jx).$$

Deconvolving $\tilde{p} = \mathcal{B}_1 * p^*$ gives

$$p^*(x) = 1 + \sum_{j=1}^n \frac{j\pi}{n+1} \cot\left(\frac{j\pi}{2n+2}\right) \cos(2\pi jx)$$

as the unique best approximation to δ_0 . Since the error of the best L^1 approximation of \mathcal{B}_1 is known from a theorem by Favard [20] (e.g. this is mentioned in [12, p. 213]), we can compute

$$\begin{aligned} W_1(\delta_0, p^*) &= \inf_{c \in \mathbb{R}} \int_{\mathbb{T}} |(\mathcal{B}_1 * p^*)(t) - (\mathcal{B}_1 * \delta_0)(t) - c| dt \\ &\leq \|\mathcal{B}_1 * p^* - \mathcal{B}_1\|_{L^1(\mathbb{T})} = \frac{1}{4(n+1)}. \end{aligned}$$

By comparison with Theorem 3.6, we notice that equality holds in this calculation and that the bound from Theorem 3.6 is sharp.

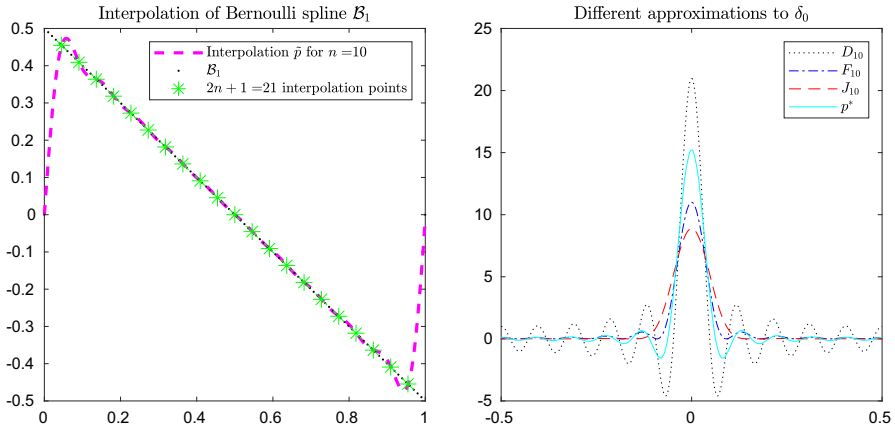


Fig. 3 Interpolation of \mathcal{B}_1 (left) and comparison of different polynomial approximations of degree $n = 10$ to δ_0 (right)

Table 1 Convergence rates of different trigonometric polynomials approximating the Dirac delta δ_0

Trigonometric polynomial	Sign of polynomial	$W_1(\delta_0, K_n)$	
Dirichlet D_n	Signed	$\leq \frac{\frac{4}{\pi^2} \log(n) + \mathcal{O}(1)}{4(n+1)}$	(Inequality (3.13))
Fejér F_n	Nonnegative	$\leq \frac{1}{\pi^2} \frac{\log(n)+4}{n+1}$	(Theorem 3.3)
Jackson J_n, n even	Nonnegative	$\leq \frac{3}{2} \frac{1}{n+2}$	(Remark 3.4)
Best approximation p^*	Signed	$= \frac{1}{4(n+1)}$	(Example 3.10 (ii))

Figure 3 and Table 1 summarize our findings on the approximation of δ_0 . The best approximation p^* as well as the Dirichlet kernel $D_n(x) = \sin((2n + 1)\pi x) / \sin(\pi x)$ are signed with small full width at half maximum (FWHM) but positive and negative oscillations at the sides. The latter might be seen as an unwanted artifact in applications. The approximations given by the Fejér and the Jackson kernel are nonnegative.

For completeness, we note that the Dirichlet kernel is the Fourier partial sum of δ_0 and allows for the estimate

$$W_1(\delta_0, D_n) \leq W_1(\delta_0, p^*) + W_1(p^*, D_n) \leq (1 + \|D_n\|_1) W_1(\delta_0, p^*) \leq \frac{\frac{4}{\pi^2} \log(n) + \mathcal{O}(1)}{4(n + 1)} \tag{3.13}$$

which relies on $W_1(p^*, D_n) = W_1(D_n * p^*, D_n * \delta_0) \leq \|D_n\|_1 W_1(\delta_0, p^*)$, the well known bound on the Lebesgue constant [5, Prop. 1.2.3], and Example 3.10 (ii).

Remark 3.11 We close by some remarks which are specific for the univariate setting:

- (i) We stress that Theorem 3.9 allows to compute the Wasserstein distance as an L^1 -distance for real signed univariate measures. Similarly, this allows to compute the

so-called star discrepancy $\|v([0, \cdot])\|_\infty$ as suggested in [44, eq. (2.1) and (2.2)]. However note that (3.11) has some additional term such that $v = \frac{1}{2}\delta_0 - \frac{1}{2}\delta_{1/2}$ with $v(\mathbb{T}) = 0$ gives

$$\|v([0, \cdot])\|_\infty = \frac{1}{2} \neq \frac{1}{4} = \|\mathcal{B}_1 * v\|_\infty$$

and thus [44, eq. (2.1) and (2.2)] needs some adjustment. More precisely, it seems that in the publication [44] a factor $\frac{1}{2}$ was lost since the k th Fourier coefficient of $v([0, \cdot])$ is $\frac{\hat{v}(k)}{ik}$ whereas $\hat{\mathcal{B}}_1(k) \cdot \hat{v}(k) = \frac{1}{2} \frac{\hat{v}(k)}{ik}$.

- (ii) In the univariate case, one can relate our work to a main result in [44]. As Theorem 3.9 reformulates the Wasserstein distance of two univariate measures in terms of the L^1 -distance of their convolution with the Bernoulli spline, one can view this Bernoulli spline as a kernel of type $\beta = 1$ following the notation of [44]. Thus, one can take $p = 1, p' = \infty$ in [44, Thm. 4.1] yielding that the Wasserstein distance between a measure μ and its trigonometric approximation is bounded from above by c/n . The latter agrees with our Remark 3.4 which additionally gives an explicit and small constant.
- (iii) The observation that the construction of p^* for δ_0 is possible via FFT's might lead to the idea to construct near-best approximations to any measure μ by interpolating $\mathcal{B}_1 * \mu$ by some \tilde{p} and to obtain the polynomial p of near best approximation which satisfies $\tilde{p} = \mathcal{B}_1 * p$ by dividing with the Fourier coefficients of the Bernoulli spline \mathcal{B}_1 . A first problem would be that the limited knowledge of moments only allows to interpolate the partial Fourier sum $S_n(\mathcal{B}_1 * \mu)$, which does not converge to $\mathcal{B}_1 * \mu$ uniformly as $n \rightarrow \infty$ for discrete μ . Secondly, the near-best approximation p cannot be expected to be nonnegative for a nonnegative measure μ which is another drawback compared to convolution with nonnegative kernels like the Fejér or Jackson kernel.

4 Interpolation

While Sect. 3 focuses on weak approximations of a measure μ , in particular via convolution with smooth kernels, we consider in this section another type of polynomial estimator, denoted by $p_{1,n}$ (4.4), which depends non-linearly on μ and is able to identify at a finite degree the support of μ , under some assumptions on the latter: more precisely, the main results of this section, stated in Theorems 4.6 and 4.10 below, are quantitative rates for the pointwise convergence

$$p_{1,n}(x) \xrightarrow{n \rightarrow \infty} \chi_{V_\mu}(x) = \begin{cases} 1, & x \in V_\mu, \\ 0, & \text{otherwise} \end{cases} \tag{4.1}$$

to the indicator function of the Zariski closure V_μ of the support, i.e. the smallest algebraic variety containing $\text{supp } \mu$. After discussing algebraic properties of this estimator (Sect. 4.1), we consider separately the case of discrete measures (Sect. 4.2) and general measures (Sect. 4.3).

In the following, let $[n] := \{0, \dots, n\}^d$ and $N := (n + 1)^d$. We use bold type to designate vectors (resp. matrices) of \mathbb{C}^N (resp. $\mathbb{C}^{N \times N}$) only (vectors of \mathbb{T}^d or \mathbb{N}^d are left in normal type). We write

$$\mathbf{e}_x^{(n)} := \left(e^{-2\pi i k x} \right)_{k \in [n]} \in \mathbb{C}^N$$

for the vector containing all d -variate trigonometric monomials up to max-degree n . Unlike previously, we consider in this section causal trigonometric polynomials [15], i.e. polynomials having zero coefficients at all negative frequencies. We often identify such a polynomial $p \in \langle e^{-2\pi i k \cdot}; k \in [n] \rangle$ with its vector of coefficients $\mathbf{p} \in \mathbb{C}^N$, i.e.

$$p(x) = \mathbf{e}_x^{(n)*} \mathbf{p} \quad \forall x \in \mathbb{T}^d.$$

Note that from Parseval’s theorem, $\|p\|_{L^2} = \|\mathbf{p}\|_2$. Note also that $|p|^2 \in \mathcal{P}_n$.

The key object of this section is the (truncated) moment matrix associated with the unknown measure μ , defined as

$$\mathbf{T}_n := \left(\hat{\mu}(k - \ell) \right)_{k, \ell \in [n]} \in \mathbb{C}^{N \times N}, \tag{4.2}$$

where $\hat{\mu}(k)$ are the trigonometric moments of μ (1.1).

4.1 Algebraic Considerations

It is well-known that the range and kernel of the matrix (4.2) reveal some of the structure of the measure hidden behind the moments, and methods that aim at recovering μ using purely algebraic manipulations on \mathbf{T}_n are often referred to as *subspace methods*, e.g. MUSIC [60], ESPRIT [55] or matrix pencils [26]. The starting point for these methods is often the singular value decomposition of \mathbf{T}_n , which we denote by

$$\mathbf{T}_n = \mathbf{U}_n \mathbf{\Sigma}_n \mathbf{V}_n^* = \sum_{j=1}^N \sigma_j^{(n)} \mathbf{u}_j^{(n)} \mathbf{v}_j^{(n)*},$$

where all matrices are of size $N \times N$, $\mathbf{u}_j^{(n)}$ and $\mathbf{v}_j^{(n)}$ are the j -th columns of \mathbf{U}_n and \mathbf{V}_n respectively (left and right singular vectors), and $\sigma_1^{(n)} \geq \sigma_2^{(n)} \geq \dots \geq \sigma_N^{(n)}$ are the diagonal entries of the diagonal matrix $\mathbf{\Sigma}_n$ (singular values). This decomposition is sometimes explicitly used to design estimators for the support of μ , such as MUSIC’s frequency estimation function [60], or Christoffel polynomials [43]. In fact, it is interesting as a motivating remark to see that the construction p_n from the previous section can also be expressed in terms of this singular value decomposition.

Lemma 4.1 *The polynomial estimator (3.2) fulfils*

$$p_n(x) = \frac{1}{N} \mathbf{e}_x^{(n)*} \mathbf{T}_n \mathbf{e}_x^{(n)} = \frac{1}{N} \sum_{j=1}^N \sigma_j^{(n)} \mathbf{u}_j^{(n)}(x) \overline{\mathbf{v}_j^{(n)}(x)}, \tag{4.3}$$

where, as explained above, $u_j^{(n)}(x) = e_x^{(n)*} u_j^{(n)}$ and $v_j^{(n)}(x) = e_x^{(n)*} v_j^{(n)}$.

Proof We have, for $n \in \mathbb{N}$ and $x \in \mathbb{T}^d$,

$$\begin{aligned} \frac{1}{N} e_x^{(n)*} T_n e_x^{(n)} &= \frac{1}{N} \sum_{k \in [n]} \sum_{l \in [n]} \hat{\mu}(k-l) e^{2\pi i k x} e^{-2\pi i l x} \\ &= \frac{1}{N} \int_{\mathbb{T}^d} \sum_{k \in [n]} \sum_{l \in [n]} e^{-2\pi i (k-l)y} e^{2\pi i k x} e^{-2\pi i l x} d\mu(y) \\ &= \int_{\mathbb{T}^d} \frac{1}{N} \left| \sum_{k \in [n]} e^{-2\pi i k (y-x)} \right|^2 d\mu(y) = \int_{\mathbb{T}^d} F_n(y-x) d\mu(y) \end{aligned}$$

where the last equality is a consequence of (3.1). Plugging in the singular value decomposition of T_n yields the second equality of the statement. \square

Note that if $\mu \in \mathcal{M}_{\mathbb{R}}$ (the set of real-valued measures), then the moment matrix T_n is Hermitian.

If $\mu \in \mathcal{M}_+$ (the set of nonnegative measures), then T_n is positive semi-definite, and we have in particular the sum of squares representation

$$p_n(x) = \frac{1}{N} \sum_{j=1}^N \sigma_j^{(n)} \left| v_j^{(n)}(x) \right|^2.$$

We now introduce polynomial estimators for the measure, which can be understood as the *unweighted* counterparts of p_n . Let $r_n := \text{rank } T_n$ and define signal- and noise-polynomials $p_{1,n}, p_{0,n} : \mathbb{T}^d \rightarrow [0, 1]$ respectively, by

$$p_{1,n}(x) = \frac{1}{N} \sum_{j=1}^{r_n} \left| v_j^{(n)}(x) \right|^2 \quad \text{and} \quad p_{0,n}(x) = \frac{1}{N} \sum_{j=r_n+1}^N \left| v_j^{(n)}(x) \right|^2. \tag{4.4}$$

This signal/noise terminology comes from the notions of signal and noise subspaces, which were initially introduced in [60] and are at the core of the aforementioned subspace methods in signal processing (we refer the interested reader to [42, Section 9.6] for an overview). Schematically speaking, they correspond to the spaces spanned by the vectors $(v_1^{(n)}, \dots, v_{r_n}^{(n)})$ (the *signal* space) and $(v_{r_n+1}^{(n)}, \dots, v_N^{(n)})$ (the *noise* space) respectively.

They are actually independent of the singular value decomposition itself, which ensures in particular that $p_{1,n}$ and $p_{0,n}$ are indeed well-defined.

The key idea of subspace methods, relating these spaces to the underlying measure μ , is that, given a causal polynomial $p \in \langle e^{-2\pi i k x}; k \in [n] \rangle$ that vanishes on $\text{supp } \mu$, one obtains using (4.2) that the k -th entry ($k \in [n]$) of $T_n p$ is given by

$$\sum_{l \in [n]} p_l \cdot \int_{\mathbb{T}^d} e^{-2\pi i (k-l)x} d\mu(x) = \int_{\mathbb{T}^d} e^{-2\pi i k x} p(x) d\mu(x) = 0, \tag{4.5}$$

and hence $\mathbf{p} \in \ker \mathbf{T}_n$. Thus, finding the common roots of all polynomials contained in the kernel of the matrix \mathbf{T}_n , may allow to identify the support of μ , or more accurately the smallest algebraic variety (the set of solutions of a polynomial system) that contains it, i.e. its Zariski closure V_μ . In what follows, we denote by $V(\ker \mathbf{T}_n)$ the set consisting of the common roots of all the polynomials in $\ker \mathbf{T}_n$, i.e.

$$V(\ker \mathbf{T}_n) := \{x \in \mathbb{T}^d : p(x) = \mathbf{e}_x^{(n)*} \mathbf{p} = 0 \text{ for all } \mathbf{p} \in \ker \mathbf{T}_n\}.$$

We begin in this section with qualitative, purely algebraic considerations about the polynomials (4.4). The next theorem shows that, under the condition that $V_\mu = V(\ker \mathbf{T}_n)$, $p_{0,n}$ and $p_{1,n}$ actually identify the set V_μ for finite n . Variants of this result can be found for the zero-dimensional and positive-dimensional ($d = 2, 3$) setting e.g. in [33] and [49, Propositions 5.2, 5.3], respectively.

Theorem 4.2 *Let $d, n \in \mathbb{N}$, $\mu \in \mathcal{M}$, and suppose $V(\ker \mathbf{T}_n) = V_\mu \subseteq \mathbb{T}^d$. Then $p_{0,n}(x) + p_{1,n}(x) = 1$ for all $x \in \mathbb{T}^d$. In particular, we have*

$$p_{1,n}(x) \begin{cases} = 1, & \text{if } x \in V_\mu, \\ < 1, & \text{otherwise.} \end{cases} \tag{4.6}$$

Proof We have

$$p_{1,n}(x) + p_{0,n}(x) = \frac{1}{N} \sum_{j=1}^N \left| v_j^{(n)}(x) \right|^2 = \frac{1}{N} \mathbf{e}_x^{(n)*} \mathbf{V}_n \mathbf{V}_n^* \mathbf{e}_x^{(n)} = \frac{1}{N} \mathbf{e}_x^{(n)*} \mathbf{e}_x^{(n)} = 1, \tag{4.7}$$

so in particular $p_{1,n}(x) \in [0, 1]$. Since $V(\ker \mathbf{T}_n) = V_\mu$ and $\ker \mathbf{T}_n = \langle \mathbf{v}_{r_n+1}^{(n)}, \dots, \mathbf{v}_N^{(n)} \rangle$, it follows that the polynomials $v_{r_n+1}^{(n)}, \dots, v_N^{(n)}$ vanish on V_μ , so $p_{1,n}(x) = 1$ for all $x \in V_\mu$. Conversely, if $x \in \mathbb{T}^d$ such that $p_{1,n}(x) = 1$, we claim that $x \in V_\mu$. Indeed, we have $1 - p_{1,n}(x) = \sum_{j=r_n+1}^N \left| v_j^{(n)}(x) \right|^2 = 0$, so it follows that x lies in the vanishing set of $v_{r_n+1}^{(n)}, \dots, v_N^{(n)}$, so $x \in V(\ker \mathbf{T}_n) = V_\mu$. \square

Remark 4.3 The hypothesis $V(\ker \mathbf{T}_n) = V_\mu$ in Theorem 4.2 is well-known in the theory of super-resolution [32, 58] or polynomial system solving [38], and is hard to check in practice. Note however that:

- (i) It is satisfied for all sufficiently large n if $\mu \in \mathcal{M}$ is finitely supported, see e.g. [35]. In particular, if μ is supported on r points $\{x_1, \dots, x_r\}$, this ensures that the rank of \mathbf{T}_n is equal to r (while in general, one only has $r_n \leq r$). The optimal n in that case depends on the geometry of the support, but it is sufficient to have $n + 1 > 6d / \min_{j \neq k} |x_j - x_k|_\infty$, see [34, Example 4.4] and [33, Cor. 2.10]. Similarly, the condition holds for sufficiently large n if $\mu \in \mathcal{M}_+$, see for example [39, Theorem 2.10] or [62, Proposition 4.10].

- (ii) If μ is neither finitely supported nor nonnegative, then $V(\ker T_n) = V_\mu$ can fail to hold for any $n \in \mathbb{N}$ (cf. [62, Example 4.9]). In this case, it is possible to rephrase the hypothesis in terms of a non-square moment matrix of suitable size (cf. [62, Theorem 4.3]) to obtain a statement similar to Theorem 4.2.
- (iii) Theorem 4.2 and the results below only deal with the Zariski closure V_μ , which coincides with $\text{supp } \mu$ only when the latter is the zero locus of a trigonometric polynomial, but is larger otherwise. In particular, we have $V_\mu = \mathbb{T}^d$ if $\text{supp } \mu$ has an interior point (with respect to the metric on \mathbb{T}^d) and in this case, $r_n = N$ and $p_{1,n} \equiv 1$. Beyond the scope of this paper, one might adapt the definition of $p_{1,n}$ by adequately thresholding the singular values of T_n , thus giving rise to algebraic approximations of the actual support.

Example 4.4 For $\mu = \delta_0$, we have $p_{1,n}(x) = F_n(x)/(n + 1)^d$ and the proof of Theorems 4.6 and 4.9 also show that $p_{1,n}$ is close to a sum of normalized Fejér kernels for arbitrary discrete measures. A singular continuous measure with support on the zero locus of a specific trigonometric polynomial is discussed as a numerical example in Sect. 5.

We conclude this subsection by stating a variational characterization of $p_{0,n}$, which will be an important tool in proof in the next sections.

Lemma 4.5 *If $\ker T_n \neq \{0\}$, we have that*

$$p_{0,n}(x) = \max \left\{ \frac{1}{N} \frac{|p(x)|^2}{\|p\|_2^2} : p \in \ker T_n \setminus \{0\} \right\}. \tag{4.8}$$

Proof As we assume $\ker T_n \neq \{0\}$, we have $r_n = \text{rank } T_n < N$ and find a matrix $V_0 = (v_{r_n+1}^{(n)}, \dots, v_N^{(n)}) \in \mathbb{C}^{N \times (N-r_n)}$ whose columns form an orthonormal basis of $\ker T_n$.

For fixed $x \in \mathbb{T}^d$, let $q_x := V_0 V_0^* e_x^{(n)} \in \ker T_n$ such that we identify this vector of coefficients with the polynomial satisfying $q_x(x) = e_x^{(n)*} q_x = \sum_{j=r_n+1}^N |v_j^{(n)}(x)|^2 = N p_{0,n}(x)$.

For all $p \in \ker T_n$, we have

$$q_x^* p = e_x^{(n)*} V_0 V_0^* p = e_x^{(n)*} p = p(x).$$

In particular, note that

$$\|q_x\|_2^2 = q_x^* q_x = e_x^{(n)*} V_0 V_0^* e_x^{(n)} = N p_{0,n}(x). \tag{4.9}$$

Therefore, by the Cauchy–Schwarz inequality, it follows that

$$|p(x)|^2 = |q_x^* p|^2 \leq \|q_x\|_2^2 \cdot \|p\|_2^2 = N p_{0,n}(x) \cdot \|p\|_2^2.$$

Hence, we have

$$p_{0,n}(x) \geq \max_{\mathbf{p} \in \ker \mathbf{T}_n \setminus \{\mathbf{0}\}} \frac{|p(x)|^2}{N \|\mathbf{p}\|_2^2} \geq \frac{|q_x(x)|^2}{N \|\mathbf{q}_x\|_2^2} = p_{0,n}(x),$$

if $\mathbf{q}_x \neq \mathbf{0}$. The first inequality also holds when $\mathbf{q}_x = \mathbf{0}$ such that the result follows due to (4.9) in this case. □

4.2 Zero-Dimensional Situation

We now come to the first main result of this section, stated in Theorem 4.6 below, which gives quantitative rates for the pointwise convergence (4.1) in the case where μ is a discrete measure. If the measure is given by

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

with (Zariski-closed) support $V_\mu = \text{supp } \mu = \{x_1, \dots, x_r\} \subset \mathbb{T}^d$ and complex weights $\mathbf{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_r)$, then the moment matrix allows for the Vandermonde factorisation

$$\mathbf{T}_n = \mathbf{A}_n^* \mathbf{\Lambda} \mathbf{A}_n, \quad \mathbf{A}_n = (e^{2\pi i k x_j})_{j=1, \dots, r; k \in [n]} \in \mathbb{C}^{r \times N},$$

which will be instrumental.

Theorem 4.6 (Pointwise convergence) *Let $\mu = \sum_{j=1}^r \lambda_j \delta_{x_j}$, $\lambda_j \in \mathbb{C} \setminus \{0\}$, $x_j \in \mathbb{T}^d$, and let $x \in \mathbb{T}^d$ such that $x \neq x_j$ for all $1 \leq j \leq r$. Let λ_{\min} and λ_{\max} be the minimal and maximal weights, in absolute value. If $n + 1 > 6d / \min_{j \neq \ell} |x_j - x_\ell|_\infty$, then*

$$p_{1,n}(x) \leq \frac{1}{(n + 1)^2} \cdot \frac{|\lambda_{\max}|}{|\lambda_{\min}|} \cdot \frac{9d^{d/2}}{4} \sum_{j=1}^r \frac{1}{|x - x_j|_\infty^2}.$$

In particular, this implies the pointwise convergence (4.1). Moreover, for $x \in \mathbb{T}^d$ such that $\min_j |x - x_j|_\infty \leq \sqrt{d} / (n + 1)$, one has

$$p_{1,n}(x) \leq \begin{cases} 1 - \frac{3^{d-1}(2d-1)}{2^d d^{2+d/2}} \cdot (n + 1)^2 \cdot \min_j |x - x_j|_\infty^2, & d > 1, \\ 1 - \frac{\pi^2}{2 \cdot 3^5} (n + 1)^2 \cdot \min_j |x - x_j|_\infty^2, & d = 1. \end{cases}$$

Proof The condition $n + 1 > 6d / \min_{j \neq \ell} |x_j - x_\ell|_\infty$ implies $\text{rank } \mathbf{T}_n = r$ and $V_\mu = V(\ker \mathbf{A}_n) = V(\ker \mathbf{T}_n)$, see [34, Exa. 4.4] and [33, Cor. 2.5, 2.10, and their proofs]. We have $p_{1,n} = 1 - p_{n,0}$ by Theorem 4.2 and since $\langle \mathbf{v}_{r+1}, \dots, \mathbf{v}_N \rangle = \ker \mathbf{T}_n = \ker \mathbf{A}_n$, it follows that $p_{1,n}$ does not depend on the weights λ_j , and we assume without

loss of generality $\lambda_j > 0$. Let then $T_n = V\Sigma V^*$ be the moment matrix of this nonnegative measure. One has, for any $x \in \mathbb{T}^d$,

$$\begin{aligned} p_{1,n}(x) &= \frac{1}{N} \sum_{j=1}^r |v_j^{(n)}(x)|^2 \leq \frac{1}{N} \sum_{j=1}^r \frac{\sigma_j^{(n)}}{\sigma_{\min}^{(n)}} |v_j^{(n)}(x)|^2 = \frac{1}{\sigma_{\min}^{(n)}} p_n(x) \\ &= \frac{1}{\sigma_{\min}^{(n)}} \sum_{j=1}^r |\lambda_j| F_n(x - x_j), \end{aligned}$$

where the last two equalities are (4.3) and (3.2), respectively. The final estimate follows from

$$F_n(x - x_j) \leq \frac{(n + 1)^{d-1}}{(n + 1) \sin^2(\pi|x - x_j|_\infty)} \leq \frac{(n + 1)^{d-2}}{4|x - x_j|_\infty^2}$$

where $\sin(\pi x) \geq 2x$ for $x \in [0, \frac{1}{2}]$ was used and $\sigma_{\min}^{(n)} \geq \frac{1}{9d^{d/2}}(n + 1)^d |\lambda_{\min}|$, see [34, Exa. 4.4].

We denote the $(r + 1)$ -th standard basis vector by $e_{r+1} = (0, \dots, 0, 1)^\top \in \mathbb{C}^{r+1}$. Regarding the second estimate, consider the Vandermonde matrix

$$\tilde{A}_{n,x} = [e_{x_1}^{(n)} \dots e_{x_r}^{(n)} e_x^{(n)}] \in \mathbb{C}^{N \times (r+1)}$$

and note that its pseudo-inverse gives rise to the Lagrange polynomial $\ell_{r+1}(y) = e_{r+1}^* \tilde{A}_{n,x}^\dagger e_y^{(n)}$, satisfying $\ell_{r+1}(x_j) = 0$ for $j = 1, \dots, r$ and $\ell_{r+1}(x) = 1$.¹³ We compute

$$\begin{aligned} \|\ell_{r+1}\|_{L^2}^2 &= \int_{\mathbb{T}^d} |e_{r+1}^* \tilde{A}_{n,x}^\dagger e_y^{(n)}|^2 dy \\ &= \int_{\mathbb{T}^d} |(\tilde{A}_{n,x}^{\dagger*} e_{r+1}, e_y^{(n)})|^2 dy = \|\tilde{A}_{n,x}^{\dagger*} e_{r+1}\|_2^2 \leq \sigma_{\min}(\tilde{A}_{n,x})^{-2} \end{aligned}$$

and use Lemma 4.5 to bound

$$1 - p_{1,n}(x) = p_{0,n}(x) = \max_p \frac{|p(x)|^2}{N \|p\|_{L^2}^2} \geq \frac{|\ell_{r+1}(x)|^2}{N \|\ell_{r+1}\|_{L^2}^2} \geq \frac{\sigma_{\min}(\tilde{A}_{n,x})^2}{N}.$$

The assertion follows from known estimates on the smallest singular value for the Vandermonde matrix with pairwise clustering nodes, see [25, Cor. 3.20] for $d > 1$ and [13, Cor. 4.2] for the univariate case $d = 1$. □

¹³ Recall that $\tilde{A}_{n,x}$ has full rank if $\min_j |x - x_j|_\infty > 0$ due to the separation of the nodes and [25, Corollary 3.20] or [13, Corollary 4.2] for the case $d = 1$. Hence, we have $\ell_{r+1}(x_j) = e_{r+1}^* \tilde{A}_{n,x}^\dagger e_{x_j}^{(n)} = e_{r+1}^* \tilde{A}_{n,x}^\dagger \tilde{A}_{n,x} e_j = e_{r+1}^* e_j = 0$ for $j = 1, \dots, r$ and analogously $\ell_{r+1}(x) = e_{r+1}^* e_{r+1} = 1$.

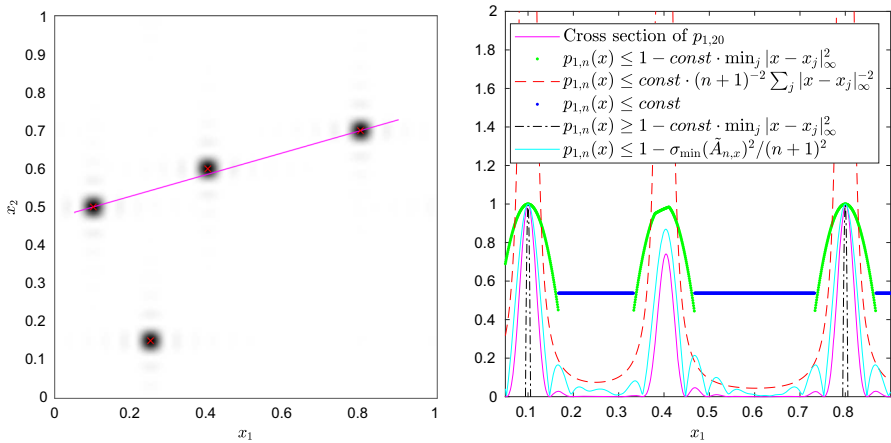


Fig. 4 Summary of the bounds on $p_{1,n}$ from Theorem 4.6 and Remark 4.7 for $d = 2, n = 20$, and a discrete measure μ supported on four points. The polynomial $p_{1,20}$ was evaluated on a grid in \mathbb{T}^2 and interpolated on the magenta cross section (left), while the bounds on $p_{1,20}$ on this cross section are displayed (right). We see that specifically the bound $1 - \sigma_{\min}(\tilde{A}_{n,x})^2/N$ from the proof of Theorem 4.6 reproduces the behaviour of $p_{1,n}$. The constant upper bound on $p_{1,n}$ away from the support of μ can be derived by using estimates for $\sigma_{\min}(\tilde{A}_{n,x})$ in the case of separated nodes

Remark 4.7 Actually, Theorem 4.6 shows the correct orders in n and $\min_j |x - x_j|_\infty^2$ in the upper bound of $p_{1,n}(x)$. First note that $1 - p_{1,n}$ and all its partial derivatives of order 1 vanish on x_1, \dots, x_r . For fixed $x \in \mathbb{T}^d$, and $j' = \operatorname{argmin}_j |x - x_j|_\infty$, the Taylor expansion at $x_{j'}$ thus gives $\xi \in \mathbb{T}^d$ such that

$$1 - p_{1,n}(x) = \frac{1}{2}(x - x_{j'})^\top H_x(\xi)(x - x_{j'}),$$

where $H_x(\xi) := (-\partial_s \partial_t p_{1,n}(\xi))_{1 \leq s, t \leq d}$ is the Hessian of $1 - p_{1,n}$ at ξ . Thus,

$$1 - p_{1,n}(x) \leq \frac{1}{2} \|H_x(\xi)\|_F \cdot |x - x_{j'}|_2^2 \leq \frac{d}{2} \max_{r,s} \|\partial_r \partial_s p_{1,n}\|_{L^\infty} \cdot d |x - x_{j'}|_2^2.$$

One may apply Bernstein’s inequality (see e.g. [12, Chapter 4]) to $y_s \mapsto p_{1,n}(y_1, \dots, y_d)$ and $y_r \mapsto \partial_s p_{1,n}(y_1, \dots, y_d)$ successively (both trigonometric polynomials of degree n), and obtain

$$2\pi^2 d^2 n^2 \cdot \min_j |x - x_j|_\infty^2$$

since $\|p_{1,n}\|_{L^\infty} = 1$. A bivariate visualisation of the bounds on $p_{1,n}$ is shown in Fig. 4.

In fact, in this discrete setting, normalizing $p_{1,n}$ differently even leads to a weak convergence result towards the empirical measure associated with the support points. This result, stated in Theorem 4.9 below, uses the following technical lemma.

Lemma 4.8 (Convergence of singular values) *Let $\mu = \sum_{j=1}^r \lambda_j \delta_{x_j}$ be a discrete complex measure whose weights are ordered non-increasingly with respect to their absolute value. Assume that $(n + 1) \min_{j \neq \ell} |x_j - x_\ell|_\infty > d$, then the singular values $\sigma_j^{(n)}$ of the moment matrix T_n fulfil*

$$\left| |\lambda_j| - \frac{\sigma_j^{(n)}}{N} \right| \leq \frac{1}{n + 1} \cdot \frac{|\lambda_1| (1 + \sqrt{e}) r}{2 \min_{j \neq \ell} |x_j - x_\ell|_\infty}, \quad j = 1, \dots, r.$$

Proof With the polar decomposition $\frac{1}{\sqrt{N}} A_n^* = P H$, where $P \in \mathbb{C}^{N \times r}$ is unitary and $H \in \mathbb{C}^{r \times r}$ is positive-definite, we have that $|\lambda_1| \geq \dots \geq |\lambda_r|$ are the singular values of the matrix $P \Lambda P^*$. Therefore, for the singular values of $T_n = A_n^* \Lambda A_n$, we obtain

$$\begin{aligned} \max_{1 \leq j \leq r} \left| \frac{\sigma_j^{(n)}}{N} - |\lambda_j| \right| &\leq \left\| \frac{1}{N} T_n - P \Lambda P^* \right\|_2 = \| H \Lambda H^* - \Lambda \|_2 \\ &\leq \| H \Lambda (H - I_r) \|_2 + \| (H - I_r) \Lambda \|_2 \\ &\leq |\lambda_1| (\| H \|_2 + 1) \| H - I_r \|_2 \\ &\leq |\lambda_1| (\| H \|_2 + 1) \left\| (H + I_r)^{-1} \right\|_2 \| H^2 - I_r \|_2 \\ &\leq |\lambda_1| \frac{\frac{1}{\sqrt{N}} \sigma_{\max}(A_n) + 1}{\frac{1}{\sqrt{N}} \sigma_{\min}(A_n) + 1} \left\| \frac{1}{N} A_n A_n^* - I_r \right\|_F, \end{aligned}$$

where the first inequality is due to [4, Theorem 2.2.8] and the last inequality is a consequence of $H = P^* P H = \frac{1}{\sqrt{N}} P^* A_n^*$ yielding

$$H^2 = H^* H = \frac{1}{N} A_n P P^* A_n^* = \frac{1}{\sqrt{N}} A_n P P^* P H = A_n \frac{1}{\sqrt{N}} P H = \frac{1}{N} A_n A_n^*.$$

Each entry of the matrix $\frac{1}{N} A_n A_n^* - I_r$ is a modified Dirichlet kernel and can be bounded uniformly by

$$\left\| \frac{1}{N} A_n A_n^* - I_r \right\|_F = \frac{1}{N} \left(\sum_{j=1}^r \sum_{l \neq j} \left| \sum_{k \in [n]} e^{2\pi i k (x_l - x_j)} \right|^2 \right)^{1/2} \leq \frac{r}{N} \cdot \frac{(n + 1)^{d-1}}{2 \min_{j \neq \ell} |x_j - x_\ell|_\infty}.$$

Moreover, since $(n + 1) \min_{j \neq \ell} |x_j - x_\ell|_\infty > d$, it follows from [35, Theorem 2.1] that

$$\frac{1}{\sqrt{N}} \sigma_{\max}(A_n) \leq \sqrt{\left(1 + \frac{1}{d} \right)^d} \leq \sqrt{e}.$$

□

Theorem 4.9 For $\lambda_j \in \mathbb{C} \setminus \{0\}$ and pairwise different $x_j \in \mathbb{T}^d$, $j = 1, \dots, r$, we have

$$\frac{p_{1,n}}{\|p_{1,n}\|_{L^1}} \rightarrow \tilde{\mu} = \frac{1}{r} \sum_{j=1}^r \delta_{x_j}$$

as $n \rightarrow \infty$.

Proof Throughout this proof we use that the Vandermonde matrix A_n has full rank r if n is sufficiently large. In particular, this implies $V(\ker T_n) = V_\mu$ and $\|p_{1,n}\|_{L^1} = r/N$. We define $\tilde{p}_n = F_n * \tilde{\mu}$ and observe that for any continuous function f on \mathbb{T}^d we have

$$\begin{aligned} & \left| \int_{\mathbb{T}^d} \frac{p_{1,n}(x)}{\|p_{1,n}\|_{L^1}} f(x) \, dx - \frac{1}{r} \sum_{j=1}^r f(x_j) \right| \\ & \leq \left| \int_{\mathbb{T}^d} \left(\frac{p_{1,n}(x)}{\|p_{1,n}\|_{L^1}} - \tilde{p}_n(x) \right) f(x) \, dx \right| + \left| \int_{\mathbb{T}^d} \tilde{p}_n(x) f(x) \, dx - \frac{1}{r} \sum_{j=1}^r f(x_j) \right| \\ & \leq \left\| \frac{N}{r} p_{1,n} - \tilde{p}_n \right\|_{L^1} \|f\|_{L^\infty} + \left| \int_{\mathbb{T}^d} f \, d(F_n * \tilde{\mu}) - \int_{\mathbb{T}^d} f \, d\tilde{\mu} \right|, \end{aligned}$$

so, by Theorem 3.3, it is enough to show that $\left\| \frac{N}{r} p_{1,n} - \tilde{p}_n \right\|_{L^1}$ converges to zero for $n \rightarrow \infty$. If n is sufficiently large, then by Lemma 4.1 we can write $\tilde{p}_n(x) = \frac{1}{N} e_x^{(n)*} \tilde{U} \tilde{\Sigma} \tilde{U}^* e_x^{(n)}$, where $\tilde{\Sigma} \in \mathbb{C}^{r \times r}$ denotes the diagonal matrix consisting of non-zero singular values, and $\tilde{U} \in \mathbb{C}^{N \times r}$ denotes the corresponding singular vector matrix of the moment matrix of $\tilde{\mu}$. As the signal polynomial $p_{1,n} = 1 - p_{0,n}$ only depends on the kernel of the moment matrix T_n of μ , which agrees with the kernel of A_n and with the kernel of the moment matrix of $\tilde{\mu}$, it follows by (4.4) that $p_{1,n}(x) = \frac{1}{N} e_x^{(n)*} \tilde{U} \tilde{U}^* e_x^{(n)}$ and thus

$$\left| \frac{N}{r} p_{1,n}(x) - \tilde{p}_n(x) \right| = \left| e_x^{(n)*} \tilde{U} \left(\frac{\mathbf{I}_r}{r} - \frac{\tilde{\Sigma}}{N} \right) \tilde{U}^* e_x^{(n)} \right| \leq \|e_x^{(n)*} \tilde{U}\|_2 \left\| \frac{1}{r} \mathbf{I}_r - \frac{1}{N} \tilde{\Sigma} \right\|_2.$$

Since $\int_{\mathbb{T}^d} \|e_x^{(n)*} \tilde{U}\|_2^2 \, dx = N \|p_{1,n}\|_{L^1} = r$ is constant, the result follows from Lemma 4.8. □

4.3 Positive-Dimensional Situation

For a measure μ whose support is contained in a non-trivial algebraic variety of any dimension, we derive a pointwise convergence rate $p_{1,n}(x) = \mathcal{O}(n^{-1})$ outside of the variety in Theorem 4.10 and together with Theorem 4.2 this proves (4.1) if $V(\ker T_n) = V_\mu$. It is not clear whether this is already optimal, as we found $\mathcal{O}(n^{-2})$ as an approximation rate in the case of a discrete measure.

Theorem 4.10 *Let $y \in \mathbb{T}^d$ and let $g \in \langle e^{2\pi i(k,x)} \mid k \in [m] \rangle$ be a trigonometric polynomial of max-degree m such that $g(y) \neq 0$ and g vanishes on $\text{supp } \mu$. Then*

$$\begin{aligned} p_{1,n+m}(y) &\leq 1 - \frac{(n+1)^d}{(n+m+1)^d} \cdot \frac{|g(y)|^2}{(F_n * |g|^2)(y)} \\ &\leq \frac{\|g\|_{L^2}^2}{|g(y)|^2} \frac{m(4m+2)^d}{n+1} + \frac{dm}{(n+m+1)}, \end{aligned}$$

for $n \in \mathbb{N}$, $n \geq m$.

Proof Set $N_n = (n+1)^d$ for $n \in \mathbb{N}$ and define the trigonometric polynomial $p(x) = e_{n,y}(x)g(x)$ of max-degree $n+m$, where $e_{n,y}(x) := e_x^{(n)*} e_y^{(n)}$. Furthermore, we define $f(x) := |g(x)|^2$. Then

$$|p(x)|^2 = N_n F_n(x-y)f(x),$$

for all $x \in \mathbb{T}^d$. On the other hand,

$$\|p\|_{L^2}^2 = N_n (F_n * f)(y).$$

The existence of a trigonometric polynomial g which vanishes on the support of the measure μ but not at $y \in \mathbb{T}^d$ shows already that $p \in \langle e^{2\pi i(k,x)} \mid k \in [n+m] \rangle$ satisfies these conditions as well and thus $\ker T_{n+m} \neq \{0\}$ by (4.5). This allows to use Lemma 4.5 in order to obtain

$$\begin{aligned} 1 - p_{1,n+m}(y) &\geq \frac{|p(y)|^2}{N_{n+m} \|p\|_{L^2}^2} \\ &= \frac{N_n}{N_{n+m}} \cdot \frac{f(y)}{(F_n * f)(y)} \\ &\geq \left(1 - \frac{m}{n+m+1}\right)^d \frac{1}{1+h_n}, \end{aligned} \tag{4.10}$$

where we define $h_n := \|F_n * f - f\|_{L^\infty} / f(y)$. This proves the first statement. For the second upper bound, we compute¹⁴

$$|(F_n * f - f)(x)| = \left| \sum_{k \in \{-m, \dots, m\}^d} \sum_{\substack{s \in \{0,1\}^d \\ 1 \leq |s| \leq d}} \frac{(-1)^{|s|} |k^s|}{(n+1)^{|s|}} \hat{f}(k) e^{2\pi i k x} \right|$$

¹⁴ We remind that the coefficients of the multivariate F_n can be written as $\prod_{\ell=1}^d (1 - \frac{|k_\ell|}{n+1}) = \sum_{\substack{s \in \{0,1\}^d \\ 0 \leq |s| \leq d}} \frac{(-1)^{|s|} |k^s|}{(n+1)^{|s|}}$ with the multi-index notation $k^s := k_1^{s_1} \dots k_d^{s_d}$ and $|s| = s_1 + \dots + s_d$.

$$\begin{aligned}
 &= \left| \sum_{k \in \{-m, \dots, m\}^d} \sum_{\substack{s \in \{0, 1\}^d \\ 1 \leq |s| \leq d}} \frac{(-1)^{|s|} |k^s|}{(n+1)^{|s|}} \int_{\mathbb{T}^d} |g(z)|^2 e^{2\pi i k(x-z)} dz \right| \\
 &\leq \sum_{\substack{s \in \{0, 1\}^d \\ 1 \leq |s| \leq d}} \left(\frac{m}{n+1} \right)^{|s|} (2m+1)^d \|g\|_{L^2}^2 \\
 &\leq \|g\|_{L^2}^2 \frac{m(4m+2)^d}{n+1}
 \end{aligned}$$

by using that $f = |g|^2$ is a trigonometric polynomial of degree m . Then it follows from (4.10) that

$$\begin{aligned}
 p_{1,n+m}(y) &\leq 1 - \left(1 - \frac{m}{n+m+1} \right)^d \left(1 - \frac{h_n}{1+h_n} \right) \\
 &\leq 1 - \left(1 - \frac{dm}{n+m+1} \right) \left(1 - \frac{h_n}{1+h_n} \right) \\
 &= \frac{h_n}{1+h_n} + \frac{dm}{(n+m+1)} - \frac{h_n}{1+h_n} \cdot \frac{dm}{(n+m+1)} \\
 &\leq \frac{\|g\|_{L^2}^2}{|g(y)|^2} \frac{m(4m+2)^d}{n+1} + \frac{dm}{(n+m+1)},
 \end{aligned}$$

since we can apply $\frac{h_n}{1+h_n} \leq h_n$. □

5 Numerical Examples

We illustrate in this section the asymptotic behaviour of p_n and $p_{1,n}$ for several types of singular measures, with respect to the 1-Wasserstein distance. We compute the distance using a semidiscrete optimal transport algorithm, described below. The code to reproduce the figures is available at <https://github.com/Paulcat/Measure-trigo-approximations>.

Our experiments focus on three examples on \mathbb{T}^2 : a discrete measure μ_d supported on 15 points, with (nonnegative) random amplitudes, a uniform measure μ_{cu} supported on the trigonometric algebraic curve

$$\cos(2\pi x) \cos(2\pi y) + \cos(2\pi x) + \cos(2\pi y) = \frac{1}{4}, \tag{5.1}$$

and a uniform measure μ_{ci} supported on the circle centered in $c_0 = (\frac{1}{2}, \frac{1}{2})$ with radius $r_0 = 0.3$.

The moments of μ_{cu} are computed numerically up to machine precision using *Arb* [29] with a parametrization of the implicit curve (5.1). It follows from (3.4) that the

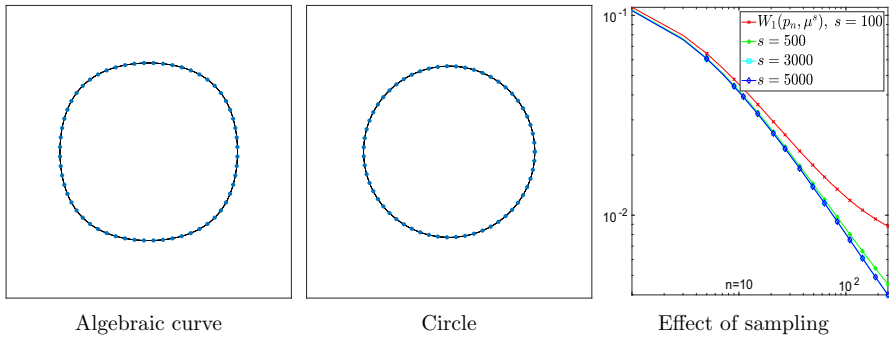


Fig. 5 The two example measures μ_{cu}^s (left) and μ_{ci}^s (middle) used in our numerical tests. In this display the two continuous measures are discretized using $s = 60$ samples. The amplitudes of the spikes in both measures are taken equal, and normalized. The last plot shows the 1-Wasserstein distance $W_1(F_n * \mu_{cu}, \mu_{cu}^s)$ for degrees $n = 1, \dots, 250$ and several values of s

trigonometric moments of the measure μ_{ci} are given by

$$\widehat{\mu}_{ci}(k) = e^{-2\pi i k c_0} J_0(2\pi r_0 \|k\|_2).$$

The polynomials p_n , $J_n * \mu$, and $p_{1,n}$ can be evaluated efficiently via the fast Fourier transform over a regular grid in \mathbb{T}^2 . For the polynomial $p_{1,n}$, the singular value decomposition of the moment matrix T_n can be computed at reduced cost by exploiting that T_n has Toeplitz structure and resorting only to matrix–vector multiplications which can be computed by means of the FFT.

To compute transport distances to the measure $\mu \in \{\mu_{cu}, \mu_{ci}\}$, let the curve $C = \text{supp } \mu \subset \mathbb{T}^d$ denote its support with arc-length L . Now let $s \in \mathbb{N}$, take a partition $C = \bigcup_{\ell=1}^s C_\ell$ into path-connected curves with measure $\mu(C_\ell) = s^{-1}$ and arc-length L_ℓ , and any $x_\ell \in C_\ell$, then

$$\begin{aligned} W_1\left(\frac{1}{s} \sum_{\ell=1}^s \delta_{x_\ell}, \mu\right) &= \sup_{f: \text{Lip}(f) \leq 1} \left| \sum_{\ell} \int_{C_\ell} [f(x) - f(x_\ell)] d\mu(x) \right| \\ &\leq \sum_{\ell=1}^s \int_{C_\ell} |x - x_\ell|_1 d\mu(x) \leq \sum_{\ell=1}^s \sqrt{d} L_\ell \mu(C_\ell) = \frac{\sqrt{d} \cdot L}{s}. \end{aligned}$$

We denote the resulting discrete measures by μ_{cu}^s and μ_{ci}^s , respectively (see Fig. 5). In our tests, we use $s = 3000$ samples, which offers a satisfactory tradeoff between computational time and accuracy for our range of degrees n . Indeed, the computational cost of evaluating the objective (5.2) or its gradient grows linearly in s , while for degrees up to $n = 250$, sampling beyond 3000 points has no effect on the output of our algorithm for computing $W_1(p_n, \mu^s)$, see Fig. 5.

Now let $\mu = \sum_{j=1}^s \lambda_j \delta_{x_j}$ refer to either μ_d , μ_{cu}^s or μ_{ci}^s . The semidiscrete optimal transport between a measure with density p and the discrete measure μ may be

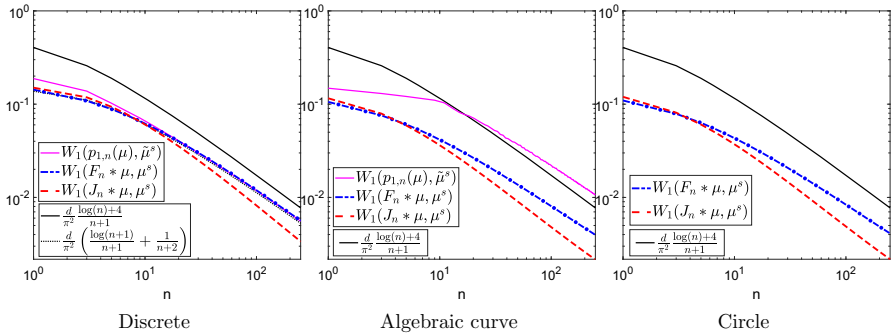


Fig. 6 Asymptotics of p_n and $p_{1,n}$. For $p_{1,n}$, the distance is computed with respect to the unweighted measure $\tilde{\mu}^s$, that is $\tilde{\mu}^s = \frac{1}{s} \sum_{j=1}^s \delta_{x_j}$ where $\{x_1, \dots, x_s\}$ is the support of μ

computed by solving the finite-dimensional optimization problem

$$\max_{w \in \mathbb{R}_+^s} f(w), \quad f(w) = \sum_{j=1}^s \lambda_j w_j + \sum_{j=1}^s \int_{\Omega_j(w)} (|x_j - y| - w_j) p(y) dy \quad (5.2)$$

where the Laguerre cells associated to the weight vector w are given by

$$\Omega_j(w) = \left\{ y \in \mathbb{T}^d : |x_j - y| - w_j \leq |x_k - y| - w_k, \quad k = 1, \dots, s \right\},$$

see e.g. [52, Sec. 5.2]. In our implementation, the density measure (and the Laguerre cells) are computed over a 502×502 grid. We use a BFGS algorithm to perform the maximization, using the Matlab implementation [59]; we stop the iterations when the change of value of the objective goes below 10^{-9} , or when the infinity norm $\|\nabla f\|_\infty$ goes below 10^{-5} . Note that this last condition has a geometrical interpretation since the j -th component of ∇f corresponds to the difference between the measure of the Laguerre cell $\Omega_j(w)$ and the amplitude λ_j . We set the limit number of iterations to 100.

In the discrete case, our numerical results (see Fig. 6) show that the Wasserstein distance $W_1(p_n, \mu^s)$ decreases at a rate close to the worst-case bound derived in Theorem 3.3. This is also the case for $W_1(p_{1,n}, \tilde{\mu}^s)$, which is coherent with the bound given in the proof of Theorem 4.9. In the positive dimensional cases, one would need to compute the Wasserstein distances for degrees larger than $n = 250$ to be able to reliably estimate a rate, but this would require better optimized algorithms, in the spirit for instance of [36], which goes beyond the scope of this paper. Still, our preliminary results seem to indicate that the rates for $F_n * \mu$ and $J_n * \mu$ in the positive dimensional situation are similar to the ones for discrete measures, but with better constants, see Fig. 6. For $p_{1,n}$ on the other hand, although the theory does not foresee weak convergence in that case, if it were to occur, our results indicate that the rate would then be worse than in the discrete case.

6 Summary and Outlook

We provided tight bounds on the pointwise approximation error as well as with respect to the 1-Wasserstein distance when approximating arbitrary measures by trigonometric polynomials. We recently generalised this also to the approximation with respect to the p -Wasserstein metric where stronger localised kernels are used [8]. Future work might address the truncation of the singular value decomposition in Sect. 4 if the support of the measure is only approximated by the zero set of an unknown trigonometric polynomial or the available trigonometric moments are disturbed by noise.

Acknowledgements We thank all reviewers for their careful reading, valuable comments, and substantial suggestions which helped to improve this work and its presentation considerably. We would like to highlight that the information theoretic question in Remark 3.7 was raised by one of them and that they pointed out a mistake in a former version of Theorems 3.9 and 3.3.

Funding Open Access funding enabled and organized by Projekt DEAL.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Akhiezer, N., Krein, M.: On the best approximation of periodic functions. Dokl. Akad. Nauk SSSR **15**, 107–112 (1937)
2. Andersson, F., Carlsson, M.: ESPRIT for multidimensional general grids. SIAM J. Mat. Anal. Appl. **39**(3), 1470–1488 (2018)
3. Beinert, R., Plonka, G.: Sparse phase retrieval of one-dimensional signals by Prony's method. Front. Appl. Math. Stat. (Math. Comput. Data Sci.) **3**(5), 1–10 (2017)
4. Björck, Å.: Numerical Methods in Matrix Computations. Texts in Applied Mathematics, Vol. 59. Springer, Cham, p. xvi+800 (2015). <https://doi.org/10.1007/978-3-319-05089-8>
5. Butzer, P.L., Nessel, R.J.: Fourier Analysis and Approximation. Pure and Applied Mathematics, Volume 1: One-Dimensional Theory, vol. 40, p. xvi+553. Academic Press, New York (1971)
6. Cabrelli, C.A., Molter, U.M.: The Kantorovich metric for probability measures on the circle. J. Comput. Appl. Math. **57**(3), 345–361 (1995). [https://doi.org/10.1016/0377-0427\(93\)E0213-6](https://doi.org/10.1016/0377-0427(93)E0213-6)
7. Candès, E., Fernandez-Granda, C.: Towards a mathematical theory of super-resolution. Commun. Pure Appl. Math. **67**(6), 906–956 (2013)
8. Catala, P., Hockmann, M., Kunis, S.: Sparse super resolution and its trigonometric approximation in the p -Wasserstein distance. Proc. Appl. Math. Mech. **22**(1), e202200125 (2023)
9. de Castro, Y., Gamboa, F.: Exact reconstruction using beurling minimal extrapolation. J. Math. Anal. Appl. **395**(1), 336–354 (2012)
10. de Castro, Y., Gamboa, F., Henrion, D., Lasserre, J.: Exact solutions to Super Resolution on semi-algebraic domains in higher dimensions. IEEE Trans. Inform. Theory **63**(1), 621–630 (2017)
11. Denoyelle, Q., Duval, V., Peyré, G.: Support recovery for sparse super-resolution of positive measures. J. Fourier Anal. Appl. **23**, 1153–1194 (2017)
12. DeVore, R.A., Lorentz, G.G.: Constructive approximation. Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], Vol 303. Springer, Berlin, pp x+449 (1993)

13. Diederichs, B.: Well-posedness of sparse frequency estimation. In: Numerical Analysis (2019). [arXiv:1905.08005](https://arxiv.org/abs/1905.08005)
14. Dryanov, D., Petrov, P.: Interpolation and L_1 -approximation by trigonometric polynomials and blending functions. *J. Approx. Theory* **164**(8), 1049–1064 (2012). <https://doi.org/10.1016/j.jat.2012.05.009>
15. Dumitrescu, B.A.: Positive Trigonometric Polynomials and Signal Processing Applications. Signals and Communication Technology. Springer, BÉrlin (2017)
16. de Prony, G. R.: Essai Expérimental et Analytique: Sur les Lois de la Dilatabilité des Fluides Élastiques et sur celles de la Force Expansive de la Vapeur de l'Eau et de la Vapeur de l'Alcool, à différentes températures'. *Journal de l'École Polytechnique Floréal et Plairial* 1.cahier **22**, 24–76 (1795)
17. Ehler, M., Kunis, S., Peter, T., Richter, C.: A randomized multivariate matrix pencil method for super-resolution microscopy. *Electron. Trans. Numer. Anal.* **51**, 63–74 (2019)
18. Ehler, M., Gräf, M., Neumayer, S., Steidl, G.: Curve based approximation of measures on manifolds by discrepancy minimization. *Found. Comput. Math.* **21**(6), 1595–1642 (2021). <https://doi.org/10.1007/s10208-021-09491-2>
19. Fatemi, M., Amini, A., Vetterli, M.: Sampling and reconstruction of shapes with algebraic boundaries. *IEEE Trans. Signal Process.* **64**(22), 5807–5818 (2016). <https://doi.org/10.1109/TSP.2016.2591505>
20. Favard, J.: Sur les meilleurs procédés d'approximation de certaines classes de fonctions par des polynômes trigonométriques. *Bull. Sci. Math.* **61**, 209–224 (1937)
21. Fisher, S.D.: Best approximation by polynomials. *Approx. Theory* **21**(1), 43–59 (1977). [https://doi.org/10.1016/0021-9045\(77\)90118-6](https://doi.org/10.1016/0021-9045(77)90118-6)
22. Grafakos, L.: Classical Fourier Analysis. Graduate Texts in Mathematics, vol. 249, 3rd edn., p. xviii+638. Springer, New York (2014). <https://doi.org/10.1007/978-1-4939-1194-3>
23. He, H., Kressner, D.: Randomized Joint Diagonalization of Symmetric Matrices (2022). [arXiv:2212.07248](https://arxiv.org/abs/2212.07248)
24. Hockmann, M., Kunis, S.: Short communication: weak sparse superresolution is well-conditioned. *SIAM J. Imaging Sci.* **16**(1), SC1–SC13 (2023). <https://doi.org/10.1137/22M1521353>
25. Hockmann, M., Kunis, S.: Sparse super resolution is Lipschitz continuous (2021). [arXiv:2108.11925](https://arxiv.org/abs/2108.11925) [math.NA]
26. Hua, Y., Sarkar, T.: Generalized pencil-of-function method for extracting poles of an EM system from its transient response. *IEEE Trans. Antennas Propag.* **37**(2), 229–234 (1989)
27. Hua, Y., Sarkar, T.K.: Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise. *IEEE Trans. Acoust. Speech Signal Process.* **38**(5), 814–824 (1990)
28. Jackson, D.: The theory of approximation. Vol. 11. American Mathematical Society Colloquium Publications. Reprint of the 1930 original. American Mathematical Society, Providence, RI, pp. viii+178 (1994)
29. Johansson, F.: Arb: efficient arbitrary-precision midpoint-radius interval arithmetic. *IEEE Trans. Comput.* **66**(8), 1281–1292 (2017). <https://doi.org/10.1109/TC.2017.2690633>
30. Jozs, C., Lasserre, J., Mourrain, B.: Sparse polynomial interpolation: compressed sensing, super resolution, or Prony? *Adv. Comput. Math.* **45**(3), 1401–1437 (2019)
31. Kroó, A., Lubinsky, D.: Christoffel functions and universality in the bulk for multivariate orthogonal polynomials. *Can. J. Math.* **65**(3), 600–620 (2012)
32. Kunis, S., Peter, T., Römer, T., Von der Ohe, U.: A multivariate generalization of Prony's method. *Linear Algebra Appl.* **490**, 31–47 (2016)
33. Kunis, S., Möller, H.M., Peter, T., von der Ohe, U.: Prony's method under an almost sharp multivariate Ingham inequality. *J. Fourier Anal. Appl.* **24**(5), 1306–1318 (2018). <https://doi.org/10.1007/s00041-017-9571-5>
34. Kunis, S., Nagel, D.: On the smallest singular value of multivariate Vandermonde matrices with clustered nodes. *Linear Algebra Appl.* **604**, 1–20 (2020). <https://doi.org/10.1016/j.laa.2020.06.003>
35. Kunis, S., Nagel, D., Strotmann, A.: Multivariate Vandermonde matrices with separated nodes on the unit circle are stable. *Appl. Comput. Harmon. Anal.* **58**, 50–59 (2022). <https://doi.org/10.1016/j.acha.2022.01.001>
36. Lakshmanan, R., Pichler, A., Potts, D.: Nonequispaced fast Fourier transform boost for the Sinkhorn algorithm. *Electron. Trans. Numer. Anal.* **58**, 289–315 (2023). https://doi.org/10.1553/etna_vol58s289
37. Lasserre, J., Pauwels, E.: The empirical Christoffel function with applications in data analysis. *Adv. Comput. Math.* **45**(3), 1439–1468 (2019)

38. Laurent, M.: Sums of Squares, moment matrices and optimization over polynomials. In: Putinar, M., Sullivant, S. (eds.) *Emerging Applications of Algebraic Geometry*, pp. 157–270. Springer, New York (2009). https://doi.org/10.1007/978-0-387-09686-5_7
39. Laurent, M., Rostalski, P.: The approach of moments for polynomial equations. In: *Handbook on semidefinite, conic and polynomial optimization*. Vol. 166. Internat. Ser. Oper. Res. Management Sci. Springer, New York, pp. 25–60 (2012). https://doi.org/10.1007/978-1-4614-0769-0_2
40. Li, W., Liao, W., Fannjiang, A.: Super-resolution limit of the ESPRIT algorithm. *IEEE Trans. Inf. Theory* **66**(7), 4593–4608 (2020)
41. Liao, W., Fannjiang, A.: MUSIC for single-snapshot spectral estimation: stability and super-resolution. In: *CoRR* (2014)
42. Manolakis, D., Ingle, V., Kogon, S.: *Statistical and Adaptive Signal Processing*. ARTECH, (2005)
43. Marx, S., Pauwels, E., Weisser, T., Henrion, D., Lasserre, J.: Tractable semi-algebraic approximation using Christoffel-Darboux kernel. *Constr. Approx.* **54**, 391–429 (2021)
44. Mhaskar, H.N.: Super-resolution meets machine learning: approximation of measures. *J. Fourier Anal. Appl.* **25**(6), 3104–3122 (2019). <https://doi.org/10.1007/s00041-019-09693-x>
45. Moerner, W., Fromm, D.: Methods of single-molecule fluorescence spectroscopy and microscopy. *Rev. Sci. Instr.* **74**(8), 3597–3619 (2003)
46. Moitra, A.: Super-resolution, extremal functions and the condition number of Vandermonde matrices. In: *STOC'15—Proceedings of the 2015 ACM Symposium on Theory of Computing*. ACM, New York, pp. 821–830 (2015)
47. Moskona, E., Petrushev, P., Saff, E.B.: The Gibbs phenomenon for best L_1 -trigonometric polynomial approximation. *Constr. Approx.* **11**(3), 391–416 (1995). <https://doi.org/10.1007/BF01208562>
48. Mourrain, B.: Polynomial-exponential decomposition from moments. *Found. Comput. Math.* **18**(6), 1435–1492 (2018)
49. Ongie, G., Jacob, M.: Off-the-grid recovery of piecewise constant images from few Fourier samples. *SIAM J. Imaging Sci.* **9**(3), 1004–1041 (2016)
50. Pan, H., Blu, T., Dragotti, P.: Sampling curves with finite rate of innovation. *IEEE Trans. Signal Process.* **62**(2), 458–471 (2014)
51. Pauwels, E., Putinar, M., Lasserre, J.-B.: Data analysis from empirical moments and the Christoffel function. *Found. Comput. Math.* **21**(1), 243–273 (2021). <https://doi.org/10.1007/s10208-020-09451-2>
52. Peyré, G., Cuturi, M.: Computational optimal transport: with applications to data science. *Found. Trends Mach. Learn.* **11**(5–6), 355–607 (2019). <https://doi.org/10.1561/22000000073>
53. Plonka, G., Tasche, M.: Prony methods for recovery of structured functions. *GAMM-Mitt.* **37**(2), 239–258 (2014)
54. Poon, C., Peyré, G.: Multi-dimensional sparse super-resolution. *SIAM J. Math. Anal.* **51**(1), 1–44 (2018)
55. Roy, R., Kailath, T.: ESPRIT-estimation of signal parameters via rotational invariance techniques. *IEEE Trans. Acoust. Speech Signal Process.* **37**(7), 984–995 (1989)
56. Rudin, W.: *Real and complex analysis*, 3rd edn., p. xiv+416. McGraw-Hill Book Co., New York (1987)
57. Sahnoun, S., Usevich, K., Comon, P.: Multidimensional ESPRIT for damped and undamped signals: algorithm, computations, and perturbation analysis. *IEEE Trans. Signal Proc.* **65**(22), 5897–5910 (2017)
58. Sauer, T.: Prony's method in several variables. *Numer. Math.* **136**, 411–438 (2017)
59. Schmidt, M.: *minFunc: unconstrained differentiable multivariate optimization in Matlab* (2005). <http://www.cs.ubc.ca/~schmidtm/Software/minFunc.html>
60. Schmidt, R.: Multiple emitter location and signal parameter estimation. *IEEE Trans. Antennas Propag.* **34**(3), 276–280 (1986)
61. Villani, C.: *Optimal transport—Old and new*. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], Vol. 338. Springer, Berlin, pp. xxii+973 (2009). <https://doi.org/10.1007/978-3-540-71050-9>
62. Wageringel, M.: Truncated moment problems on positive-dimensional algebraic varieties (2022). [arXiv:2203.01269](https://arxiv.org/abs/2203.01269) [math.AC]