

Spectral Properties of Infinitely Smooth Kernel Matrices in the Single Cluster Limit, with Applications to Multivariate Super-Resolution

Nuha Diab¹ and Dmitry Batenkov¹

¹Department of Applied Mathematics, Tel Aviv University, Tel Aviv, Israel

July 16, 2024

Abstract

We study the spectral properties of infinitely smooth multivariate kernel matrices when the nodes form a single cluster. We show that the geometry of the nodes plays an important role in the scaling of the eigenvalues of these kernel matrices. For the multivariate Dirichlet kernel matrix, we establish a criterion for the sampling set ensuring precise scaling of eigenvalues. Additionally, we identify specific sampling sets that satisfy this criterion. Finally, we discuss the implications of these results for the problem of super-resolution, i.e. stable recovery of sparse measures from bandlimited Fourier measurements.

1 Introduction

Super-Resolution Problem. The problem of super-resolution (SR) is to recover the fine details of a signal from inherently low resolution measurements in the frequency domain. The SR problem has a variety of applications in signal processing, imaging, optics, inverse scattering, and data analysis problems [9, 27]. In the last years, a particular SR model has received considerable attention, where the signal is modelled as a sparse measure [8]. Let $\mathbb{T}^d := [-\pi, \pi)^d \equiv (\mathbb{R} \bmod 2\pi)^d$, let $\delta(\cdot)$ denote the Dirac delta distribution, and consider the signal

$$f(t) := \sum_{j=1}^n \alpha_j \delta(t - x_j), \quad t, x_j \in \mathbb{T}^d, \quad \alpha_j \in \mathbb{C}, \quad (1)$$

which we aim to recover. We refer to $\{x_j\}$ as the "nodes" and to $\{\alpha_j\}$ as the "coefficients" or "amplitudes". The noisy measurements are given by

$$\hat{f}(\omega) := \sum_{j=1}^n \alpha_j e^{i\langle \omega, x_j \rangle} + \epsilon(\omega), \quad \omega \in \mathcal{G}_N := \{-N+1, \dots, -1, 0, 1, \dots, N-1\}^d \subset \mathbb{Z}^d, \quad |\epsilon(\omega)| \leq \epsilon. \quad (2)$$

Moreover, define $\text{SRF} := \frac{1}{N\Delta}$, where Δ represents the smallest distance (e.g. in the infinity norm) between the nodes. It is well-known that, at least in the one-dimensional case, this quantity controls the numerical stability of the problem. The super-resolution regime is when $\text{SRF} \gg 1$; while the

well-separated case is when $\text{SRF} \leq 1$. In the super-resolution regime, nodes are organized in clusters, with a distance of order Δ between any two nodes within a cluster.

One primary objective in theory of super-resolution is to define optimal bounds for reconstruction errors, commonly referred to as min-max error bounds, or the computational resolution limits [14, 28, 25, 12, 8]. These bounds are achieved by the (non-tractable) most effective algorithm under the worst-case scenario. Furthermore, these stability bounds serve to validate the optimality of different tractable solution methods, offering assurances regarding the methods' performance [26, 20, 21]. It has been showed previously that the min-max bounds are related to the smallest singular value of the Vandermonde matrix U of the system [12], which is defined as

$$U(x_1, \dots, x_n; \mathcal{G}_N) := [\exp i\langle \omega, x_j \rangle]_{\omega \in \mathcal{G}_N}^{j=1, \dots, n} \in \mathbb{C}^{|\mathcal{G}_N| \times n}. \quad (3)$$

Extensive research has been conducted on the one-dimensional scenario, covering both well-separated and super-resolution regimes [7, 25, 11, 33]. Let ℓ be the number of nodes in the largest cluster. Then, in the super-resolution regime, the smallest singular value of the Vandermonde matrix scales like $\text{SRF}^{1-\ell}$, resulting in “on-grid” min-max bounds to be of order $\text{SRF}^{2\ell-1}\epsilon$ [5, 25], (cf. similar off-grid bounds [8, 29]). In the multidimensional context, several results are available for the well-separated regime as well as the super-resolution regime [28, 23, 22, 16, 34], under specific conditions related to the unknown nodes. A key distinction between the one-dimensional and multidimensional cases lies in the spectrum of the Vandermonde matrix, wherein the geometry of the nodes plays a crucial role. As we shall show in this paper, and consistent with the observations in [22], it is not solely the distance between the nodes that determines the spectral properties of the Vandermonde matrix, at least in the near-colliding limit, but also the algebraic variety on which these nodes are situated.

Kernel matrices in the flat limit. Following the previous discussion, let us consider the Gramian matrix $D_N := \frac{1}{(2N)^d} U^* U$, which is the “kernel matrix” for the multidimensional Dirichlet kernel. Kernel matrices are of importance in various fields, including scattered data approximation and machine learning, see e.g. [1, 41] and references therein. In the context of super-resolution, if all the nodes $\mathcal{X} = \{x_j\}_{j=1}^n$ form a single cluster and $\text{SRF} \gg 1$, then the kernel matrix D_N can be considered in the so-called “flat limit”, a term introduced in [15] in the context of radial basis function (RBF) interpolation. The flat limit was investigated recently in various publications, from among those [1] and [41] being most relevant to our work. While [41] deals exclusively with RBF kernels and derives asymptotic eigenvalue scaling, [1] extends those results to arbitrary smooth kernels, albeit under the assumption of the points not lying on any low-dimensional algebraic variety. The approach from [41] is primarily based on Micchelli’s lemma [30], and we have adopted this approach previously in the one-dimensional case [7].

Contributions. The main result of this paper, Theorem 1, derives the asymptotic scaling of eigenvalues of a kernel matrix induced by an infinitely smooth kernel in the flat limit, generalizing the corresponding result of [1] to arbitrary geometry. As a corollary, we derive the asymptotic scaling of the eigenvalues of D_N , leading to the appropriate scaling of the singular values of the Vandermonde matrix (3), when the nodes are arranged within a single cluster. In this context, a “single cluster” is characterized by the condition $N\Delta \ll 1$. We therefore obtain a generalization of the results from [23] which considered mainly the “single line” geometry, as well as the extension of [7, Theorem 2.3] to the multivariate case. See also Remark 6.

In more detail, suppose that $x_j = \Delta y_j$ with $\epsilon = N\Delta$, and take $\epsilon \rightarrow 0$. We show that the number of eigenvalues of $D_N(\{x_1, \dots, x_n\})$ decaying like ϵ^{2k} , is precisely equal to a certain number t_k which

depends on the algebraic properties of the original nodes $\{y_j\}$ (see Theorem 5). In particular, when all the nodes belong to a one-dimensional affine subspace (a “single-line” geometry), then $t_k = 1$, which implies $\sigma_{\min}(U) \approx N^{\frac{d}{2}} \epsilon^{n-1}$ (as derived in [23]). On the other hand, when $\{y_j\}$ are in a general position, then, as long as $n \geq \binom{k+d}{d}$, we have $t_k = \binom{k+d-1}{d-1}$, where d is the dimension of the space. These numbers coincide for $d = 1$, but may differ drastically for $d > 1$. For example, if $d = 2$ then $\sigma_{\min}(U)$ may be as large as $N \epsilon^{\mathcal{O}(\sqrt{n})}$. Our results are slightly more general and provide the eigenvalue scaling for arbitrary (symmetric) sampling sets $\omega \in \mathcal{S}$, with the upper bounds the same as described above; however we can prove matching lower bounds only for certain \mathcal{S} satisfying the so-called Geometric Characterization Condition (in particular satisfied by \mathcal{G}_N), see Definition 8 below.

To demonstrate the relevance of our findings to the multivariate SR problem, in Section 6 we also report on a numerical study of reconstructing the signals from noisy measurements using the nonlinear least squares (NLS) method and the multidimensional ESPRIT method [36]. Our results indeed suggest that the asymptotic conditioning of the problem may vary from the “worst-case” single-line scenario (previously considered in [28]), to the “best-case” general position scenario. We are therefore confident that our findings will enhance the understanding of super-resolution stability in high-dimensional settings and can contribute to the analysis of current multivariate SR recovery methods such as [13, 32, 38, 39, 36]. Furthermore, we believe these results can serve as an important step towards establishing min-max error bounds in the general multidimensional scenario, when the nodes form multiple clusters (where some kind of multidimensional “confluent” Vandermonde matrices may be required, cf. [3, 2, 4, 6]).

Organization of the paper. In Section 2 we establish some notation and definitions. In Section 3 we present our main result, Theorem 1, which describes the asymptotic scaling of the eigenvalues of kernel matrices in the flat limit. In Section 4 we specialize the results to the Dirichlet kernel matrix, showing, in particular, that the bounds of Theorem 1 are tight. In Section 5 we derive the corresponding scaling of the singular values of the Vandermonde matrix, which requires additional technical arguments. Finally, numerical results are presented in Section 6.

Acknowledgements. We thank Konstantin Usevich, Benedikt Diederichs, and Weilin Li for helpful discussions. This work has been supported by the Israel Science Foundation, grant 1793/20, by a collaborative grant from the Volkswagen Foundation, and by the Ariane de Rothschild Women Doctoral Program.

2 Preliminaries

We adapt some definitions and notation from [1]. For a multi-index $\alpha = (\alpha_1, \dots, \alpha_d) \in \mathbb{Z}_+^d$, denote

$$|\alpha| := \sum_{j=1}^d \alpha_j, \quad \alpha! := \alpha_1! \cdots \alpha_d!,$$

and let for each $j = 0, 1, 2, \dots$

$$\mathbb{P}_j := \{\alpha \in \mathbb{Z}_+^d : |\alpha| \leq j\}, \quad \mathbb{H}_j := \{\alpha \in \mathbb{Z}_+^d : |\alpha| = j\} = \mathbb{P}_j \setminus \mathbb{P}_{j-1}, \quad \mathbb{P}_{-1} := \emptyset.$$

The cardinalities of these sets are given by:

$$p_j := \#\mathbb{P}_j = \binom{j+d}{d}, \quad h_j := \#\mathbb{H}_j = \binom{j+d-1}{d-1} = \#\mathbb{P}_j - \#\mathbb{P}_{j-1}. \quad (4)$$

For $\mathbf{x} \in \mathbb{R}^d$ and $\boldsymbol{\alpha} \in \mathbb{Z}_+^d$, the monomial \mathbf{x}^α is defined as $\mathbf{x}^\alpha := x_1^{\alpha_1} \cdots x_d^{\alpha_d}$, and its (total) degree is $|\boldsymbol{\alpha}|$. Let \prec be a fixed graded order on the set of multi-indices \mathbb{Z}_+^d (or, equivalently, on the set of monomials), for example a graded lexicographic order or a graded reverse lexicographic order. It will always be implied in the following.

Definition 1. Let $\Omega \subseteq \mathbb{R}^d$ be a fixed open set. A kernel function $\mathcal{K} : \Omega \times \Omega \rightarrow \mathbb{R}$ is said to belong to the class $\mathcal{C}^{(r,r)}(\Omega)$ (or $\mathcal{C}^{\mathbb{P}_r \times \mathbb{P}_r}(\Omega)$) when the partial derivatives

$$\mathcal{K}^{(\boldsymbol{\alpha}, \boldsymbol{\beta})}(\mathbf{x}, \mathbf{y}) := \frac{\partial^{|\boldsymbol{\alpha}|+|\boldsymbol{\beta}|}}{\partial x_1^{\alpha_1} \cdots \partial x_d^{\alpha_d} \partial y_1^{\beta_1} \cdots \partial y_d^{\beta_d}} \mathcal{K}(\mathbf{x}, \mathbf{y})$$

exist and are continuous on $\Omega \times \Omega$, for all $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{P}_r$. When $\mathcal{K} \in \mathcal{C}^{(r,r)}(\Omega)$ for all $r \in \mathbb{N}$, we write $\mathcal{K} \in \mathcal{C}^{(\infty, \infty)}(\Omega)$.

Definition 2. For a kernel $\mathcal{K} \in \mathcal{C}^{(r,r)}(\Omega)$, a finite set $\mathcal{X} \subset \Omega$ and $\epsilon > 0$ we define the scaled kernel $\mathcal{K}_\epsilon(\mathbf{x}, \mathbf{y}) := \mathcal{K}(\epsilon \mathbf{x}, \epsilon \mathbf{y})$ and the kernel matrix

$$K_\epsilon(\mathcal{X}) := [\mathcal{K}_\epsilon(\mathbf{x}, \mathbf{x}')]_{\mathbf{x}, \mathbf{x}' \in \mathcal{X}} \in \mathbb{R}^{|\mathcal{X}| \times |\mathcal{X}|}. \quad (5)$$

Definition 3 (Multivariate Vandermonde matrix). For a finite set $\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\} \subset \mathbb{R}^d$ of nodes, and an \prec -ordered set of multi-indices $\mathcal{A} = \{\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_m\} \subset \mathbb{Z}_+^d$ (i.e. $\boldsymbol{\alpha}_1 \prec \cdots \prec \boldsymbol{\alpha}_m$), we define the multivariate Vandermonde matrix as

$$V_{\mathcal{A}} = V_{\mathcal{A}}(\mathcal{X}) = [(\mathbf{x}_i)^{\boldsymbol{\alpha}_j}]_{\substack{1 \leq j \leq m \\ 1 \leq i \leq n}} \in \mathbb{R}^{n \times m}. \quad (6)$$

Notice that V is a real Vandermonde matrix, as opposed to the (transpose of the) complex Vandermonde matrix U defined in (3). Let us further denote $V_{\leq k} := V_{\mathbb{P}_k}(\mathcal{X})$ and $V_k := V_{\mathbb{H}_k}(\mathcal{X})$. The matrix $V_{\leq k} \in \mathbb{R}^{n \times p_k}$ can be partitioned into $k+1$ matrices $V_k \in \mathbb{R}^{n \times h_k}$ arranged by increasing degree:

$$V_{\leq k} = [V_0 \quad V_1 \quad \dots \quad V_k]. \quad (7)$$

Definition 4. Let $\mathcal{K} \in \mathcal{C}^{(r,r)}(\Omega)$, with $\Omega \subseteq \mathbb{R}^d$ and $\mathbf{0} \in \Omega$. Let $\mathcal{A}, \mathcal{B} \subset \mathbb{Z}_+^d$ be two sets of multi-indices satisfying $|\alpha|, |\beta| \leq r$ for all $\alpha \in \mathcal{A}, \beta \in \mathcal{B}$. The Wronskian matrix of \mathcal{K} is defined as:

$$W_{\mathcal{A}, \mathcal{B}}^{\mathcal{K}} = \left[\frac{\mathcal{K}^{(\alpha, \beta)}(\mathbf{0}, \mathbf{0})}{\alpha! \beta!} \right]_{\alpha \in \mathcal{A}, \beta \in \mathcal{B}}, \quad (8)$$

where the rows and columns are indexed by multi-indices in \mathcal{A} and \mathcal{B} according to the chosen ordering. In addition, denote by $W_{\leq k}^{\mathcal{K}} = W_{\mathbb{P}_k}^{\mathcal{K}} = W_{\mathbb{P}_k, \mathbb{P}_k}^{\mathcal{K}}$.

The final piece of notation is central to our discussion, describing a geometric property of \mathcal{X} .

Definition 5 (Discrete moment order, [40]). The discrete moment order of a set \mathcal{X} consisting of n pairwise distinct points, is the smallest number $m = \mu(\mathcal{X})$ such that

$$\ker(V_{\leq m}(\mathcal{X})^T) = \emptyset \iff \text{rank}(V_{\leq m}(\mathcal{X})) = n.$$

Note that such m exists (is finite). Indeed, if $\mathcal{X} = \{x_1, \dots, x_n\}$, then $\mu(\mathcal{X}) \leq n-1$ because the following Lagrange interpolation polynomials, each of total degree $n-1$, are linearly independent:

$$\left\{ \ell_i(x) = \prod_{j \neq i} \frac{H_j(x)}{H_j(x_i)} \right\}_{i=1}^n,$$

where H_j is some affine hyperplane containing x_j and not containing every other $x_k, k \neq j$. Cf. [40, Definition 3], [17] and [41].

3 Eigenvalues of smooth kernel matrices in the flat limit

In this section, we consider the kernel matrix $K_\epsilon = K_\epsilon(\mathcal{X})$ for a fixed infinitely smooth kernel \mathcal{K} and a fixed set $\mathcal{X} \subset \mathbb{R}^d$ of n pairwise distinct points. Our main interest is in the asymptotic order of the decay of the eigenvalues of K_ϵ in the flat limit $\epsilon \rightarrow 0$. We employ the standard $O(\cdot)$ notation.

Our main result gives an upper bound on the decay rates of all the eigenvalues.

Theorem 1. *Let $m = \mu(\mathcal{X})$ be the discrete moment order of the set \mathcal{X} . For any infinitely smooth kernel $\mathcal{K} \in \mathcal{C}^{(\infty, \infty)}(\Omega)$ with $\mathbf{0} \in \Omega$, and small enough ϵ , the eigenvalues of $K_\epsilon(\mathcal{X})$ split into $m + 1$ groups*

$$\lambda_{0,0} = O(1), \quad \{\lambda_{1,j}\}_{j=1}^{t_1} = O(\epsilon^2), \quad \dots, \quad \{\lambda_{m,j}\}_{j=1}^{t_m} = O(\epsilon^{2m}),$$

$$t_k := \text{rank}(V_{\leq k}(\mathcal{X})) - \text{rank}(V_{\leq k-1}(\mathcal{X})), \quad k = 1, 2, \dots, m.$$

Remark 1. *Theorem (1) is valid also for kernels in $\mathcal{C}^{(m+1, m+1)}$.*

Proof. We start by writing the Taylor expansion of K_ϵ around 0 as equation (55) in [1]:

$$K_\epsilon = V_{\leq m} \Delta_m W \Delta_m V_{\leq m}^T + \epsilon^{m+1} (V_{\leq m} \Delta_m W_1(\epsilon) + W_2(\epsilon) \Delta_m V_{\leq m}^T) + \epsilon^{2(m+1)} W_3(\epsilon) \quad (9)$$

where $W_i(\epsilon) = O(1)$, $W = W_{\leq m}^{\mathcal{K}}$ and $\Delta_m(\epsilon) = \text{diag}(1, \epsilon I_{h_1}, \dots, \epsilon^m I_{h_m}) \in \mathbb{R}^{p_m \times p_m}$. For the full derivation of (9), see section A in the Appendix.

Let $V = V_{\leq m}$ where $V = [V_0 \dots V_m] = [V_{\leq m-1} \quad V_m]$, and $t_k := \text{rank}(V_{\leq k}) - \text{rank}(V_{\leq k-1})$ as defined in the theorem. In [1] only the generic case $t_k = h_k - h_{k-1}$ is considered. Here we generalize Theorem 6.1 in [1] to the case where the multivariate Vandermonde matrices $V_{\leq k}$ are not necessarily full rank.

Consider the full QR decomposition of $V_{\leq m-1}$ with column pivoting, as elaborated in Section 5.4.2, specifically, formula (5.4.6) in the book [18]:

$$V_{\leq m-1} P_{m-1} = [Q_{\leq m-1} \quad Q_\perp] \begin{bmatrix} R_{11} & R_{12} \\ 0 & 0 \end{bmatrix} = Q \begin{bmatrix} R \\ 0 \end{bmatrix}$$

where $Q_{\leq m-1} \in \mathbb{R}^{n \times r_{m-1}}$, $Q_\perp \in \mathbb{R}^{n \times t_m}$, $R_{11} \in \mathbb{R}^{r_{m-1} \times r_{m-1}}$, $R \in \mathbb{R}^{r_{m-1} \times p_{m-1}}$, $r_{m-1} + t_m = n$ and P_{m-1} is a permutation matrix.

We have $[Q_{\leq m-1} \quad Q_\perp]^T V_{\leq m} P = \begin{bmatrix} R & Q_{\leq m-1}^T V_m \\ 0 & Q_\perp^T V_m \end{bmatrix}$ and $Q_\perp^T V_{\leq m} P = [0 \quad Q_\perp^T V_m]$ where $P := \begin{bmatrix} P_{m-1} & 0 \\ 0 & I_{h_m} \end{bmatrix}$.

Note that for every permutation matrix, we have $P^{-1} = P^T$.

Thus we can rewrite the expansion (9) as follows:

$$\begin{aligned} K_\epsilon &= V_{\leq m} P P^T \Delta_m P P^T W P P^T \Delta_m P P^T V_{\leq m}^T \\ &\quad + \epsilon^{m+1} (V_{\leq m} P P^T \Delta_m P P^T W_1(\epsilon) + W_2(\epsilon) P P^T \Delta_m P P^T V_{\leq m}^T) \\ &\quad + \epsilon^{2(m+1)} W_3(\epsilon) \\ K_\epsilon &= \tilde{V}_{\leq m} \tilde{\Delta}_m \tilde{W} \tilde{\Delta}_m \tilde{V}_{\leq m}^T + \epsilon^{m+1} (\tilde{V}_{\leq m} \tilde{\Delta}_m \tilde{W}_1(\epsilon) + \tilde{W}_2(\epsilon) \tilde{\Delta}_m \tilde{V}_{\leq m}^T) + \epsilon^{2(m+1)} W_3(\epsilon) \end{aligned}$$

where $\tilde{V}_{\leq m} := V_{\leq m} P$, $\tilde{\Delta}_m := P^T \Delta_m P = \text{diag}(\epsilon^{i_1} I_{h_{i_1}}, \dots, \epsilon^{i_m} I_{h_{i_m}}, \epsilon^m I_{h_m})$ due to Lemma 7, $\tilde{W} := P^T W P$, $\tilde{W}_1(\epsilon) := P^T W_1(\epsilon)$ and $\tilde{W}_2(\epsilon) := W_2(\epsilon) P$.

Multiplying by Q_{\perp}^T and its transpose we get:

$$\begin{aligned} Q_{\perp}^T K_{\epsilon} Q_{\perp} &= [0 \quad Q_{\perp}^T V_m] \tilde{\Delta} \tilde{W} \tilde{\Delta} [0 \quad Q_{\perp}^T V_m]^T \\ &\quad + \epsilon^{m+1} [0 \quad Q_{\perp}^T V_m] \tilde{\Delta} \tilde{W}_1 Q_{\perp} \\ &\quad + \epsilon^{m+1} Q_{\perp}^T \tilde{W}_2 \tilde{\Delta} [0 \quad Q_{\perp}^T V_m]^T \\ &\quad + (\epsilon^{m+1} Q_{\perp}^T W_3 Q_{\perp} \epsilon^{m+1}). \end{aligned}$$

We can also write the above as:

$$\begin{aligned} Q_{\perp}^T K_{\epsilon} Q_{\perp} &= O(\epsilon^{2m}) Q_{\perp}^T V_m \hat{W} V_m^T Q_{\perp} \\ &\quad + O(\epsilon^{2m+1}) (Q_{\perp}^T V_m \hat{W}_1 Q_{\perp} \\ &\quad + Q_{\perp}^T \hat{W}_2 V_m Q_{\perp}) \\ &\quad + O(\epsilon^{2(m+1)}), \end{aligned}$$

where \hat{W} is the lower right $h_m \times h_m$ sub-matrix of \tilde{W} . Thus, by the Courant-Fischer principle, K_{ϵ} has at least $t_m = \dim(Q_{\perp})$ eigenvalues which decay as fast as $O(\epsilon^{2m})$.

For any $k = 1, \dots, m-1$, let $V_{\leq k-1} P_{k-1} = [Q_{\leq k-1} \quad Q_{\perp, k-1}] \begin{bmatrix} R_{\leq k-1} \\ 0 \end{bmatrix}$ be the full QR decomposition of $V_{\leq k}$ with pivoting, where $Q_{\leq k-1} \in \mathbb{R}^{n \times [n - (t_k + \dots + t_m)]}$, $Q_{\perp, k-1} \in \mathbb{R}^{n \times [t_k + \dots + t_m]}$ and $R_{\leq k-1} \in \mathbb{R}^{[n - (t_k + \dots + t_m)] \times p_{k-1}}$.

We can apply the same argument on $Q_{\perp, k-1}$ and $V_{\leq k-1}$ by induction, then we get that there are at least $t_k + \dots + t_m$ eigenvalues which decay as fast as $O(\epsilon^{2k})$ and there are at least $t_{k+1} + \dots + t_m$ eigenvalues which decay as fast as $O(\epsilon^{2(k+1)})$. Therefore, there are at least t_k eigenvalues which decay as fast as $O(\epsilon^{2k})$. Since $\sum_{i=0}^m t_i = n$ (assume $t_0 = 0$) we get that there are exactly t_k eigenvalues decaying (at least) like ϵ^{2k} when ϵ goes to 0. \square

The bounds of Theorem 1 are not necessarily tight. As we show in the next lemma, the exactness of the asymptotic orders depends on the analytic properties of the kernel \mathcal{K} and the geometry of \mathcal{X} , therefore additional assumptions ought to be made on both of those objects in order to obtain matching lower bounds.

Lemma 1. *For $\mathcal{K} \in \mathcal{C}^{(\infty, \infty)}(\Omega)$ and $\mathcal{X} \subset \mathbb{R}^d$, there exists $C = C(\mathcal{X}, \mathcal{K}) \in \mathbb{R}$ such that*

$$\det(K_{\epsilon}(\mathcal{X})) = \epsilon^{2 \sum_{i=0}^m t_i} (C + O(\epsilon)), \quad \epsilon \rightarrow 0.$$

In particular, $C = \det(\tilde{R} W \tilde{R}^T)$ where $\tilde{R} \in \mathbb{R}^{n \times p_m}$ is full rank, and $W = W_{\leq m}^{\mathcal{K}} \in \mathbb{R}^{p_m \times p_m}$ is the Wronskian matrix of \mathcal{K} as in Definition 4.

Proof. Consider the matrices $Q_{\perp, k-1}$ defined in the proof of Theorem 1. Let $\mathcal{M}_k \subseteq \text{range}(Q_{\perp, k-1})$ and $\mathcal{M}_k \perp \text{range}(Q_{\perp, k})$ where $\mathcal{M}_k \in \mathbb{R}^{n \times t_k}$ and $Q_{\perp, -1} = \mathbb{R}^{n \times n}$. Define

$$\mathcal{Q}_{\perp} := (\mathcal{M}_0, \mathcal{M}_1, \dots, \mathcal{M}_m) \in \mathbb{R}^{n \times n}$$

It is easy to see that $\mathcal{Q}_{\perp}^T V_{\leq m}$ is upper triangular, with the block-diagonal part given by

$$\tilde{R} := \text{blkdiag}(\mathcal{Q}_{\perp}^T V_{\leq m}) = \text{diag}\{\mathcal{M}_0^T V_0, \dots, \mathcal{M}_m^T V_m\}.$$

Each diagonal block of \tilde{R} is full rank by definition, therefore \tilde{R} is full rank.

Therefore by [1, Lemma 6.4] we have

$$E_n^{-1} \mathcal{Q}_\perp^T V_{\leq m} \Delta_m = \tilde{R} + O(\epsilon)$$

where $E_n = \text{diag}\{1, \epsilon I_{t_1}, \dots, \epsilon^m I_{t_m}\} \in \mathbb{R}^{n \times n}$ and $\tilde{R} \in \mathbb{R}^{n \times p_m}$.

Thus, using (9), we get that

$$\begin{aligned} E_n^{-1} \mathcal{Q}_\perp^T K_\epsilon \mathcal{Q}_\perp E_n^{-1} &= (\tilde{R} + O(\epsilon)) W (\tilde{R}^T + O(\epsilon)) \\ &\quad + \epsilon^{m+1} ((\tilde{R} + O(\epsilon)) W_1(\epsilon) \mathcal{Q}_\perp E_n^{-1} \\ &\quad + E_n^{-1} \mathcal{Q}_\perp^T W_2(\epsilon) (\tilde{R}^T + O(\epsilon))) \\ &\quad + \epsilon^{2(m+1)} E_n^{-1} \mathcal{Q}_\perp^T W_3(\epsilon) \mathcal{Q}_\perp E_n^{-1} \\ &= \tilde{R} W \tilde{R}^T + O(\epsilon) \end{aligned}$$

where the last equality follows from $\epsilon^{2(m+1)} E_n^{-1} = O(\epsilon)$ and $\tilde{R} W \tilde{R}^T \in \mathbb{R}^{n \times n}$. This implies that $\epsilon^{-2 \sum_{j=0}^m j t_j} \det(K_\epsilon) = \det(\tilde{R} W \tilde{R}^T + O(\epsilon)) = \det(\tilde{R} W \tilde{R}^T) + O(\epsilon)$ (by Lemma (8) in the Appendix) with $C = \det(\tilde{R} W \tilde{R}^T)$. □

Corollary 1. *Suppose \mathcal{K} is symmetric and analytic (in both variables) in the neighborhood of $(\mathbf{0}, \mathbf{0})$. The scaling of the eigenvalues as given in Theorem 1 is exact, if and only if $\det(\tilde{R} W \tilde{R}^T) \neq 0$.*

Proof. By Theorem 1 and [1, Theorem 2.9], the eigenvalues of $K_\epsilon(\mathcal{X})$ satisfy

$$\lambda_{0,0} = \epsilon^0 (\tilde{\lambda}_{0,0} + O(\epsilon)), \quad \{\lambda_{1,j}\}_{j=1}^{t_1} = \{\epsilon^2 (\tilde{\lambda}_{1,j} + O(\epsilon))\}_{j=1}^{t_1}, \dots, \{\lambda_{m,j}\}_{j=1}^{t_m} = \{\epsilon^{2m} (\tilde{\lambda}_{m,j} + O(\epsilon))\}_{j=1}^{t_m},$$

where $t_k := \text{rank}(V_{\leq k}) - \text{rank}(V_{\leq k-1})$ and $m = \mu(\mathcal{X})$, and $\tilde{\lambda}_{s,j}$ do not depend on ϵ (but may be zero). By Lemma 1 we have that

$$|K_\epsilon(\mathcal{X})| = \epsilon^{2 \sum_{i=0}^m i t_i} \left\{ |\tilde{R} W \tilde{R}^T| + O(\epsilon) \right\}.$$

On the other hand, the determinant is the product of eigenvalues, therefore

$$\begin{aligned} |K_\epsilon(\mathcal{X})| &= \prod_{j=1}^n \lambda_j = \lambda_{0,0} \prod_{j=1}^{t_1} \lambda_{1,j} \dots \prod_{j=1}^{t_m} \lambda_{m,j} = (\tilde{\lambda}_{0,0} + O(\epsilon)) \prod_{j=1}^{t_1} \epsilon^2 (\tilde{\lambda}_{1,j} + O(\epsilon)) \dots \prod_{j=1}^{t_m} \epsilon^{2m} (\tilde{\lambda}_{m,j} + O(\epsilon)) \\ &= \epsilon^{2 \sum_{i=0}^m i t_i} \left\{ \prod_{s=1}^m \prod_{j=1}^{t_s} \tilde{\lambda}_{s,j} + O(\epsilon) \right\}. \end{aligned}$$

Now we just compare the two expressions above. In one direction, if the scaling is exact then $\tilde{\lambda}_{s,j} \neq 0$ for all $0 \leq s \leq m$ and $0 \leq j \leq t_s$, which implies $|\tilde{R} W \tilde{R}^T| \neq 0$. In the other direction, if there are $\ell \geq 1$ zeros $\tilde{\lambda}_k \in \{\tilde{\lambda}_{s,j}\}_{s=1, \dots, m}^{j=1, \dots, t_m}$ and $\tilde{\lambda}_k = 0$ for $k = 1, \dots, \ell$, then $|K_\epsilon| = \epsilon^{\ell+2 \sum_{i=0}^m i t_i} (\prod_{k=\ell+1}^m \tilde{\lambda}_k + O(\epsilon))$, forcing $|\tilde{R} W \tilde{R}^T| = 0$. □

4 Eigenvalues of the multidimensional Dirichlet kernel

In this section we consider the specific case of the multidimensional Dirichlet kernel, which is directly motivated by the problem of super-resolution as outlined in the Introduction. Let $\mathcal{S} \subset \mathbb{R}^d$ denote a finite set of frequencies, symmetric around the origin, where the spectral data (2) is available. We first derive a generic sufficient condition on \mathcal{S} which ensures that the scaling order of Theorem 1 is exact. Later we show that this condition is satisfied by $\mathcal{S} = \mathcal{G}_N$, the regular symmetric grid in \mathbb{R}^d of side length $2N + 1$, for large enough N .

Definition 6. For $\mathcal{S} \subset \mathbb{R}^d$, symmetric around $\mathbf{0} \in \mathbb{R}^d$, the (real-valued) Dirichlet kernel $\mathcal{D}_{\mathcal{S}}$ is defined as follows:

$$\mathcal{D}_{\mathcal{S}}(x) = \sum_{\omega \in \mathcal{S}} \exp(i\langle x, \omega \rangle).$$

By a slight abuse of notation, we also denote by $\mathcal{D}_{\mathcal{S}}$ the corresponding bi-multivariate kernel:

$$\mathcal{D}_{\mathcal{S}}(x, y) = \mathcal{D}_{\mathcal{S}}(x - y) = \sum_{\omega \in \mathcal{S}} \exp(i\langle x - y, \omega \rangle). \quad (10)$$

Given $\mathcal{X} \subset \mathbb{R}^d$ and $\epsilon > 0$, the corresponding Dirichlet kernel matrix and its flat limit version are denoted as:

$$\begin{aligned} D(\mathcal{X}, \mathcal{S}) &:= [\mathcal{D}_{\mathcal{S}}(x, x')]_{x, x' \in \mathcal{X}}, \\ D_{\epsilon}(\mathcal{X}, \mathcal{S}) &:= D(\epsilon\mathcal{X}, \mathcal{S}) = [\mathcal{D}_{\mathcal{S}}(\epsilon x, \epsilon x')]_{x, x' \in \mathcal{X}}. \end{aligned} \quad (11)$$

In addition, if $\mathcal{S} = \mathcal{S}_1 \times \dots \times \mathcal{S}_d$ is a tensor product grid, we have $\mathcal{D}_{\mathcal{S}}(x, y) = \prod_{j=1}^d \mathcal{D}_{\mathcal{S}_j}(x_j, y_j)$.

4.1 A nondegeneracy condition

Note that $\mathcal{D}_{\mathcal{S}}$ is a translation invariant, symmetric and analytic kernel. Recall the definition (8) of the Wronskian matrix for an arbitrary kernel. Set $W := W_{\leq m}^{\mathcal{D}_{\mathcal{S}}}$. By Corollary 1, in order to show tightness, it is sufficient to demonstrate that $C := \det(\tilde{R}W\tilde{R}^T) \neq 0$. In the following lemma, we show that this holds under a certain (more or less natural) non-degeneracy assumption on the set \mathcal{S} .

Lemma 2. For $W := W_{\leq m}$ as defined above, $\det(\tilde{R}W\tilde{R}^T) > 0$ holds whenever $\text{rank}(V_{\leq m}(\mathcal{S})) \geq n$.

Remark 2. Note that contrary to Section 3, the Vandermonde matrix $V_{\leq m}$ is here evaluated on the set \mathcal{S} (and not on \mathcal{X}), while the polynomial total degree is still $m = \mu(\mathcal{X})$.

Proof of Lemma 2. We first show that W is positive semidefinite. Directly from (8) and (10), we have

$$\begin{aligned} W &= \left[\frac{\mathcal{D}_{\mathcal{S}}^{(\alpha, \beta)}(0, 0)}{\alpha! \beta!} \right]_{\alpha, \beta \in \mathbb{P}_m} = \left[\sum_{\omega \in \mathcal{S}} \frac{(i\omega)^{\alpha} (-i\omega)^{\beta}}{\alpha! \beta!} \right]_{\alpha, \beta \in \mathbb{P}_m} \\ &= \left[\sum_{\omega = (\omega_1, \dots, \omega_d) \in \mathcal{S}} \frac{(i)^{|\alpha|} (-i)^{|\beta|} \omega_1^{\alpha_1 + \beta_1} \cdot \dots \cdot \omega_d^{\alpha_d + \beta_d}}{\alpha! \beta!} \right]_{\alpha, \beta \in \mathbb{P}_m} \\ &= FV_{\leq m}^T(\mathcal{S})V_{\leq m}(\mathcal{S})F^*, \end{aligned}$$

where $V_{\leq m}(\mathcal{S}) = [(\omega)^\alpha]_{\omega \in \mathcal{S}}^{\alpha \in \mathbb{P}_m}$ is the multivariate Vandermonde matrix and $F := \text{diag} \left(\frac{i! \alpha!}{\alpha!} \right)_{\alpha \in \mathbb{P}_m}$. Let $B := FV_{\leq m}^T(\mathcal{S})$, and consider $B = U\Sigma T^*$ the singular value decomposition of B . We therefore have $W = BB^* = U\Sigma^2 U^*$, implying that W is positive semidefinite (since W is Hermitian and has non-negative eigenvalues). Since \tilde{R} is full rank by Lemma 1, it follows that $\tilde{R}W\tilde{R}^T$ is positive semidefinite as well, implying that $\det(\tilde{R}W\tilde{R}^T) \geq 0$. To ensure $\det(\tilde{R}W\tilde{R}^T) > 0$, we require that $\tilde{R}W\tilde{R}^T$ be positive definite. The latter condition is satisfied when $\text{rank}(\tilde{R}W\tilde{R}^T) \geq n$, which is equivalent to $\text{rank}(W) = \text{rank}(V_{\leq m}(\mathcal{S})) \geq n$. \square

Remark 3. For $d = 1$, we have that $m = n - 1$ and $p_{n-1} = \binom{n}{1} = n$, meaning that the condition $\text{rank}(V_{\leq n-1}(\mathcal{S})) \geq n$ in Lemma (2) is satisfied as long as the sampling set \mathcal{S} contains at least n distinct points.

Remark 4. Given the condition $\text{rank}(V_{\leq m}(\mathcal{S})) \geq n$ in Lemma (2), we aim to find a sampling set \mathcal{S}_* with p_ℓ points where $p_{\ell-1} < n$ and $p_\ell \geq n$ such that $\text{rank}(V_{\leq \ell}(\mathcal{S}_*)) \geq n$ which is independent of $m = \mu(\mathcal{X})$ (the geometry of the nodes).

By applying Corollary 1, we immediately obtain the following result.

Lemma 3. Assume that the sampling set \mathcal{S} satisfies the condition $\text{rank}(V_{\leq m}(\mathcal{S})) \geq n$. Then the eigenvalues of $D_\epsilon(\mathcal{X}, \mathcal{S})$ split into $m + 1$ groups

$$\lambda_{0,0} = \epsilon^0(\tilde{\lambda}_{0,j} + O(\epsilon)), \quad \{\lambda_{1,j}\}_{j=1}^{t_1} = \{\epsilon^2(\tilde{\lambda}_{1,j} + O(\epsilon))\}_{j=1}^{t_1}, \dots, \{\lambda_{m,j}\}_{j=1}^{t_m} = \{\epsilon^{2m}(\tilde{\lambda}_{m,j} + O(\epsilon))\}_{j=1}^{t_m},$$

where $t_k := \text{rank}(V_{\leq k}(\mathcal{X})) - \text{rank}(V_{\leq k-1}(\mathcal{X}))$ and $\tilde{\lambda}_{s,j} \neq 0$ for all $0 \leq s \leq m$ and $0 \leq j \leq t_s$.

4.2 Geometric characterization condition and the uniform grid

The rank condition in Lemma 2 is nontrivial to verify for a given set \mathcal{S} . In this section we provide an explicit example of such a set, namely, the uniform grid \mathcal{G}_N in \mathbb{R}^d . Our approach is a straightforward application of well-known results in multivariate polynomial interpolation, adapted to our setting.

We first recall some definitions and results from [37].

Definition 7 ([37], p.194). Let \mathcal{P} be a linear subspace of the polynomial ring $\Pi := \mathbb{R}[x_1, \dots, x_d]$. The polynomial interpolation problem with respect to a (finite) set of linearly independent functionals $\Theta \subset \Pi'$ (Π' is the dual space) is said to be **poised** for \mathcal{P} , if for any $Y \in \mathbb{R}^{\#\Theta}$ there exists a unique $f \in \mathcal{P}$ such that $\Theta f = Y$.

Theorem 2 ([37] p.194). For $\mathcal{P} \subset \Pi$, and a finite set $\Theta \subset \Pi'$ as above, the following statements are equivalent:

1. The polynomial interpolation problem with respect to Θ is poised for \mathcal{P} .
2. $\dim \mathcal{P} = \#\Theta$ and $\Theta P = [\theta p : \theta \in \Theta, p \in P]$ satisfies $|\Theta P| \neq 0$ for any basis P of \mathcal{P} .

For $\Theta_{\mathcal{S}} := \{\delta_\xi\}_{\xi \in \mathcal{S}}$ the set of point evaluation functionals on \mathcal{S} , and the standard monomial basis P_n of polynomials of total degree $\leq n$, we have in particular $\dim P_n = \#\mathcal{S} = \binom{n+d}{d} = p_n$, and the matrix $\Theta_{\mathcal{S}} P = V := V_{\leq n}(\mathcal{S})$ is our Vandermonde matrix. Therefore, if we can find a sample set \mathcal{S} such that the polynomial interpolation problem with respect to $\Theta_{\mathcal{S}}$ is poised for P_n , this would immediately imply that $|V| \neq 0$. Examples of such situations are provided in [10, 17], and it turns out that there is a general sufficient condition which ensures poisedness.

Definition 8 ([10]). A lattice $\mathcal{J} := \{x_1, \dots, x_N\}$ of $N = p_n$ distinct nodes of \mathbb{R}^d is said to satisfy the **Geometric Characterization Condition (GC)** (of degree n) if corresponding to each node x_i there exist n distinct hyperplanes G_{i1}, \dots, G_{in} such that:

1. x_i does not lie on any of these hyperplanes, and
2. all the other nodes in \mathcal{J} lie on at least one of these hyperplanes.

The following result is fundamental for our purposes. We give a short proof for completeness.

Theorem 3 (Theorem 1 in [10]). Let $\mathcal{J} := \{x_1, \dots, x_N\}$ be a lattice of $N = p_n$ distinct nodes of \mathbb{R}^d . If \mathcal{J} satisfies the GC condition as in Definition 8, then the polynomial interpolation problem with respect to \mathcal{J} is poised for P_n .

Proof. Let $Y = \{y_i\}_{i=1, \dots, N}$ be an arbitrary vector of values. It can be directly verified that the following polynomial of total degree $\leq n$ interpolates Y on \mathcal{J} :

$$P(x) = \sum_{i=1}^N y_i \prod_{j=1}^n \frac{G_{ij}(x)}{G_{ij}(x_i)}.$$

Since Y was arbitrary, uniqueness follows. □

Next, we recall the definition of n -th order principal lattice [31].

Definition 9. Let $I_n = \{0, \frac{1}{n}, \frac{2}{n}, \dots, 1\}$. Further, denote by $X_i := e_i$ the simplex vertices with e_i being the i -th element in standard basis of \mathbb{R}^d , and furthermore put $X_0 := \mathbf{0} \in \mathbb{R}^d$. Then the n -th order principal lattice (corresponding to the coordinate simplex) is defined as

$$B(n, d) := \left\{ x \in \mathbb{R}^d \mid x = \sum_{i=0}^d \gamma_i X_i, \sum_{i=0}^d \gamma_i = 1, \gamma_i \in I_n \right\}.$$

Note that $B(n, d)$ contains precisely $p_n = \binom{n+d}{n}$ points.

It was proven in [10] Theorem 4, that $B(n, d)$ satisfies the GC condition which, by Theorem 3, guarantees poisedness [17]. This is almost what we want, and indeed as we will show, the uniform grid contains a scaled principal lattice.

Lemma 4. Let $J_n := \{0, 1, \dots, n\}$, and put

$$A(n, d) := n \cdot B(n, d) = \left\{ x \in \mathbb{R}^d \mid x = \sum_{i=0}^d \gamma_i X_i, \sum_{i=0}^d \gamma_i = n, \gamma_i \in J_n \right\} \subset \mathcal{G}_n.$$

Then the set $A(n, d)$ satisfies the GC condition of degree n .

Proof. The proof is very similar to the one in [10] for $B(n, d)$, and we provide it for completeness.

Let $\{\gamma_k : \mathbb{R}^d \rightarrow \mathbb{R}\}_{k=0, \dots, d}$ be the barycentric coordinate functions associated with the simplex X_0, \dots, X_d . Define the following hyperplanes $H_{km} := \{x \mid \gamma_k(x) = m\}$, $k = 0, \dots, d$, $m = 0, \dots, n$. These hyperplanes contain all the nodes in $A(n, d)$.

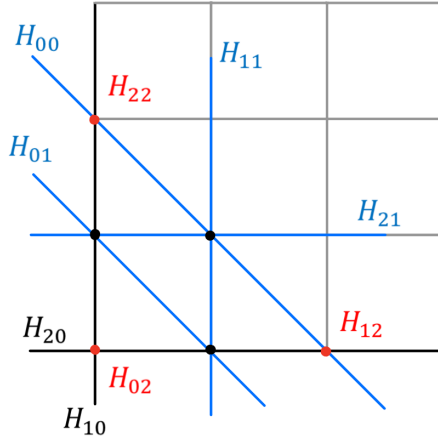


Figure 1: Demonstration of the proof of Lemma 4.

Now let $x \in A(n, d)$ be a lattice node, and let $\beta_k := \gamma_k(x)$. For each $k = 0, \dots, d$ with $\beta_k > 0$, the union of the hyperplanes $\{H_{kj}\}_{k=0, \dots, d}^{j=0, \dots, \beta_k-1}$ doesn't contain the node x . Since $\sum_{k=0}^d \beta_k = n$ we have exactly n hyperplanes in the set $\{H_{kj}\}_{k=0, \dots, d}^{j=0, \dots, \beta_k-1}$.

We still have to show why the union of $\{H_{kj}\}_{k=0, \dots, d}^{j=0, \dots, \beta_k-1}$ contains all the other nodes. Let $y \neq x \in A(n, d)$. Again, we have $\sum_{k=0}^d \gamma_k(y) = n$, therefore there is at least one $k \in \{0, \dots, d\}$ such that $\gamma_k(y) < \gamma_k(x) = \beta_k$. \square

To clarify the proof, we give an example for $d = 2, n = 2$. For the node $x = (0, 1)$ we have $\gamma_0(x) = 1, \gamma_1(x) = 0, \gamma_2(x) = 1$ because $x = 1 \cdot (0, 0) + 0 \cdot (1, 0) + 1 \cdot (0, 1)$. Following the proof, H_{00} and H_{20} are the hyperplanes which contain all the nodes except for x (see figure 1).

Remark 5. For the scaled principal lattice $A(n, d)$, the interpolation polynomial $P(x)$ of degree $\leq n$ for any $\{y_a\}_{a \in A(n, d)}$ is given by:

$$P(x) = \sum_{a \in A(n, d)} y_a \prod_{k=0}^d \prod_{m=0}^{\gamma_k(a)-1} \frac{\gamma_k(x) - m}{\gamma_k(a) - m}.$$

We are finally in a position to state and prove the main result of this section.

Theorem 4. Let $\mathcal{X} \subset \mathbb{R}^d$ be a finite set of n distinct points in \mathbb{R}^d , and let $m = \mu(\mathcal{X})$ be its discrete moment order. Let ℓ be the unique integer such that $p_{\ell-1} < n \leq p_\ell$. Then for every $N \geq \ell$ we have

$$\text{rank}(V_{\leq m}(\mathcal{G}_N)) \geq n,$$

so the conclusion of Lemma 3 holds for $\mathcal{S} = \mathcal{G}_N$.

Proof. Using Lemma 4 and Theorems 2 and 3, we get

$$\begin{aligned} |V_{\leq \ell}(A(\ell, d))| \neq 0 &\implies \text{rank}(V_{\leq \ell}(A(\ell, d))) = \binom{\ell + d}{d} = p_\ell \geq n \\ A(\ell, d) \subset \mathcal{G}_\ell &\implies \text{rank}(V_{\leq \ell}(\mathcal{G}_\ell)) \geq \binom{\ell + d}{d} \geq n. \end{aligned}$$

Now notice that since $\text{rank } V_{\leq m}(\mathcal{X}) = n$ we must have $p_m \geq n$ and therefore $m \geq \ell$. This implies

$$\text{rank}(V_{\leq m}(\mathcal{G}_N)) \geq \text{rank}(V_{\leq m}(\mathcal{G}_\ell)) \geq \text{rank}(V_{\leq \ell}(\mathcal{G}_\ell)) \geq n,$$

finishing the proof. \square

5 Vandermonde matrices in the super-resolution regime

In this section we come back to the super-resolution problem, and consider the Vandermonde matrices $U(x_1, \dots, x_n; \mathcal{G}_N)$ in (3) in the single cluster setting. We prove the multidimensional analogue of [7, Theorem 2.3], with the caveat that the geometry of the nodes stays fixed as $\text{SRF} \gg 1$. We start by defining the cluster geometry. Naturally, since the sampling set is an integer grid, we must restrict the nodes to $\mathbb{T}^d := [-\pi, \pi)^d \equiv (\mathbb{R} \bmod 2\pi)^d$ to avoid aliasing.

Definition 10. For $x, y \in \mathbb{T}^d$, we denote the wrap-around distance by

$$\|x - y\|_{\mathbb{T}^d} := \min_{r \in (2\pi\mathbb{Z})^d} \|x - y + r\|_\infty.$$

Definition 11. Let $\mathcal{X} = \{x_1, \dots, x_n\} \subset \mathbb{T}^d$. If for some $\tau > 1$ and $0 < \Delta < \pi/\tau$ we have

$$\forall x, y \in \mathcal{X}, x \neq y : \quad \Delta \leq \|x - y\|_{\mathbb{T}^d} \leq \tau\Delta,$$

then \mathcal{X} is said to form an (Δ, τ, n) -cluster.

In what follows we fix a node set $\mathcal{Y} = \{y_1, \dots, y_n\} \subset \mathbb{T}^d$ such that the (non-wrapped) minimal distance between any two nodes is

$$\rho = \rho(\mathcal{Y}) := \min_{i \neq j} \|y_i - y_j\|_\infty.$$

Now let $\mathcal{X} = \Delta\mathcal{Y}$ with $\Delta < \frac{1}{2\pi}$, so that \mathcal{X} forms an (Δ', τ', n) -cluster for $\Delta' = \rho\Delta$ and some $\tau' \leq \frac{2\pi}{\rho}$. (Otherwise, there exist $y, y' \in \mathcal{Y}$ such that $\|y - y'\|_{\mathbb{T}^d} \geq \frac{2\pi}{\rho}\Delta' = 2\pi\Delta$, a contradiction).

Further, let $\mathcal{G}_N = \{0, \pm 1, \dots, \pm N\}^d$ be the uniform symmetric grid of side length $2N + 1$.

Let $U := U(x_1, \dots, x_n; \mathcal{G}_N)$. Denote $\epsilon := N\Delta$. We have

$$U^*U = D(\mathcal{X}, \mathcal{G}_N) = D\left(\epsilon \frac{\mathcal{Y}}{N}, \mathcal{G}_N\right) = D_\epsilon\left(\frac{\mathcal{Y}}{N}, \mathcal{G}_N\right).$$

We further define the rescaled Dirichlet kernel

$$\begin{aligned} \mathcal{K}(x, y) &= \mathcal{K}_N(x, y) := \frac{1}{(2N)^d} \mathcal{D}_{\mathcal{G}_N}(x/N, y/N) \\ \implies K_\epsilon(\mathcal{Y}) &= \frac{1}{(2N)^d} D_\epsilon\left(\frac{\mathcal{Y}}{N}, \mathcal{G}_N\right) = D_N(\mathcal{X}). \end{aligned}$$

We have the following result.

Theorem 5. With the above notations, let $m = \mu(\mathcal{Y}) = \mu(\mathcal{X})$ be the discrete moment order of \mathcal{Y} .

For every $\epsilon_0 > 0, \alpha \in (0, 1)$ there exists N_0 such that for all $N \geq N_0$ and $\alpha\epsilon_0 < N\Delta < \epsilon_0$, the eigenvalues of D_N split into $m + 1$ groups

$$\lambda_{0,0}(D_N) \asymp (N\Delta)^0, \quad \{\lambda_{1,j}(D_N)\}_{j=1}^{t_1} \asymp (N\Delta)^2, \dots, \{\lambda_{m,j}(D_N)\}_{j=1}^{t_m} \asymp (N\Delta)^{2m},$$

where $t_k := \text{rank}(V_{\leq k}(\mathcal{X})) - \text{rank}(V_{\leq k-1}(\mathcal{X}))$, $k = 1, \dots, m$.

Proof. With the identification $D_N(\mathcal{X}) = K_\epsilon(\mathcal{Y})$ as above, we would like to apply Theorem 1 and Corollary 1 to the kernel $\mathcal{K} = \mathcal{K}_N$. However, \mathcal{K} depends on N , so we need to make sure that all the estimates in the proofs are uniform in N .

Definition 12 (sinc kernel). Let $\text{sinc}(x) := \begin{cases} \frac{\sin(x)}{x}, & \text{if } x \neq 0, \\ 1, & \text{if } x = 0. \end{cases}$. Given $\mathcal{X} \in \mathbb{R}^d$, the corresponding sinc kernel matrix is defined as follows:

$$\text{Sinc}(\mathcal{X}) := \left[\prod_{i=1}^d \text{sinc}(x_i - x'_i) \right]_{x, x' \in \mathcal{X}} \quad (12)$$

For $\epsilon > 0$, the flat limit version of the sinc kernel matrix is:

$$\text{Sinc}_\epsilon(\mathcal{X}) := \left[\prod_{i=1}^d \text{sinc}(\epsilon x_i - \epsilon x'_i) \right]_{x, x' \in \mathcal{X}} \quad (13)$$

Outline of the proof: we aim to prove that the eigenvalues of D_N converge to Sinc kernel when $N \rightarrow \infty$. The scaling of the eigenvalues of Sinc can be obtained using Corollary 1 (the constants $\tilde{\lambda}_{k,j}$ don't depend on N). First we show element-wise convergence of D_N to sinc and equicontinuity of D_N in $\epsilon < \epsilon_0$ to show uniform convergence of D_N to sinc (uniform in $\epsilon < \epsilon_0$ and independent of N). Then using the Bauer–Fike Theorem, we prove convergence of the eigenvalues.

We start with some definitions.

Theorem 6 (The Bauer–Fike Theorem). For a diagonalizable matrix A let E be it's eigenvectors matrix, $A = E\Lambda E^{-1}$ where Λ is a diagonal matrix. Let μ be an eigenvalue of A_N , then there exists $\lambda \in \Lambda(A)$ such that

$$|\lambda - \mu| \leq \kappa_p(E) \|A - A_N\|_p \quad (14)$$

where $\kappa_p(E) := \|E\|_p \|E^{-1}\|_p$ is the condition number of E associated with the norm $\|\cdot\|_p$.

Definition 13 (Element-wise convergence). Let $\{A_N\}_{N=1}^\infty$, where $A_N \in \mathbb{R}^{s \times s}$, be a sequence of matrices, and let $A \in \mathbb{R}^{s \times s}$ be a matrix. The sequence $\{A_N\}_{N=1}^\infty$ is said to converge element-wise to A if the following condition is satisfied:

For every $\delta > 0$, there exists a positive integer $M(\delta)$ such that for all $N \geq M(\delta)$, the inequality

$$|(A_N)_{i,j} - (A)_{i,j}| \leq \delta$$

holds for all $1 \leq i, j \leq s$.

Definition 14 (Equicontinuity). Let $\{f_n(x)\}$ be a sequence of functions. $\{f_n(x)\}$ is equicontinuous if for every $\delta > 0$, there exists $\nu > 0$ (independent of n) such that for all n and all $x, y \in [a, b]$ with $|x - y| < \nu$ we have $|f_n(x) - f_n(y)| < \delta$.

Definition 15 (Uniform convergence). A sequence of functions $\{f_n(x)\}$ converges uniformly to $f(x)$ on S if for every $\epsilon > 0$, there exists an integer N such that for all $n \geq N$ and for all $x \in S$, $|f_n(x) - f(x)| < \epsilon$.

Theorem 7. Let $C(J)$ be the space of continuous functions on a closed interval $J \subset \mathbb{R}$. A sequence in $C(J)$ is uniformly convergent if and only if it is equicontinuous and converges pointwise to a function.

Proof. The first direction is an immediate corollary of Theorem 7.25 in [35]. Since the sequence converges pointwise, we can replace the subsequence in the proof of (b) by the sequence itself. The second direction is Theorem 7.24. \square

Lemma 5. *For every $\epsilon > 0$ and $\mathcal{X} \in \mathbb{R}^d$, the Dirichlet kernel matrix $D_N(\mathcal{X})$ converges element-wise to the sinc kernel matrix $\text{Sinc}_\epsilon(\mathcal{Y})$ when $N \rightarrow \infty$.*

Proof.

$$(D_N(\mathcal{X}))_{i,j} = \frac{1}{(2N)^d} (D_\epsilon\left(\frac{\mathcal{Y}}{N}, \mathcal{G}_N\right))_{i,j} = \frac{1}{(2N)^d} \mathcal{D}_{\mathcal{G}_N}\left(\epsilon\left(\frac{y_i}{N} - \frac{y_j}{N}\right)\right) = \frac{1}{(2N)^d} \sum_{\omega \in \mathcal{G}_N} \exp\left(\imath \left\langle \epsilon\left(\frac{y_i}{N} - \frac{y_j}{N}\right), \omega \right\rangle\right) \quad (15)$$

Denote by $y := y_i - y_j$ and $y := (y^{(1)}, \dots, y^{(d)})$. First, we consider the case where $d = 1$:

$$\begin{aligned} \frac{1}{2N} \sum_{k=-N}^N \exp(\imath \epsilon k \frac{y}{N}) &= \frac{1}{2N} \left(1 + 2 \sum_{k=1}^N \cos\left(\epsilon k \frac{y}{N}\right)\right) = \frac{1}{2N} \frac{\sin\left(\left(N + \frac{1}{2}\right) \frac{\epsilon y}{N}\right)}{\sin\left(\frac{\epsilon y}{2N}\right)} \\ &\xrightarrow{N \rightarrow \infty} \frac{1}{2N} \frac{\sin\left(\left(1 + \frac{1}{2N}\right) \epsilon y\right)}{\frac{\epsilon y}{2N}} \xrightarrow{N \rightarrow \infty} \frac{\sin(\epsilon y)}{\epsilon y} = \text{sinc}(\epsilon y). \end{aligned}$$

For $d > 1$, using the above, we get

$$\frac{1}{(2N)^d} \sum_{\omega \in \mathcal{G}_N} \exp\left(\imath \left\langle \epsilon \left(\frac{y}{N}\right), \omega \right\rangle\right) = \prod_{i=1}^d \frac{1}{2N} \sum_{k=-N}^N \exp(\imath \epsilon k \frac{y^{(i)}}{N}) \xrightarrow{N \rightarrow \infty} \prod_{i=1}^d \text{sinc}(\epsilon y^{(i)}).$$

\square

We proved that each entry of $D_N(\mathcal{X})$ converges point-wise to each entry of $\text{Sinc}_\epsilon(\mathcal{Y})$. Now we show that this convergence is uniform for $\epsilon \in [0, \epsilon_0]$ for a given ϵ_0 . We show that using Theorem 7. First we need to show that the sequence $\{(D_N(\mathcal{X}))_{i,j}\}$ is equicontinuous.

For fixed $y \in \mathcal{Y}$, let $f_N(\epsilon) := \frac{1}{2N} \sum_{k=-N}^N \exp(\imath \epsilon k \frac{y}{N})$ and $f(\epsilon) := \text{sinc}(\epsilon y)$. We prove the following result.

Lemma 6. *Let $\epsilon_0 > 0$ be given. For every $\delta > 0$ there exists $\nu > 0$ (independent of N) such that for all N and all $\epsilon, \epsilon' \in [0, \epsilon_0]$ with $|\epsilon - \epsilon'| < \nu$ we have $|f_N(\epsilon) - f_N(\epsilon')| < \delta$.*

Proof. Let $\delta > 0$. we want to find $\nu(\delta) > 0$ such that for all N and $\epsilon, \epsilon' \in [0, \epsilon_0]$ with $|\epsilon - \epsilon'| < \nu$ we

have that $|f_N(\epsilon) - f_N(\epsilon')| < \delta$. Using that $|1 - e^{ix}| = 2|\sin(\frac{x}{2})|$ and $\sin(x) \leq x$ for $x \in [0, \frac{\pi}{2}]$ we get

$$\begin{aligned}
|f_N(\epsilon) - f_N(\epsilon')| &= \left| \frac{1}{2N} \sum_{k=-N}^N \exp(i\epsilon k \frac{y}{N}) - \frac{1}{2N} \sum_{k=-N}^N \exp(i\epsilon' k \frac{y}{N}) \right| \\
&= \left| \frac{1}{2N} \sum_{k=-N}^N \exp(i\epsilon k \frac{y}{N}) (1 - \exp(i(\epsilon - \epsilon') k \frac{y}{N})) \right| \\
&\leq \frac{1}{2N} \sum_{k=-N}^N |1 - \exp(i(\epsilon - \epsilon') k \frac{y}{N})| \\
&= \frac{1}{2N} \sum_{k=-N}^N 2|\sin((\epsilon - \epsilon') k \frac{y}{2N})| \\
&\leq \frac{2}{2N} \sum_{k=0}^N 2k|\epsilon - \epsilon'| \frac{y}{2N} = \frac{1}{N^2} \frac{N + N^2}{2} \nu y \\
&= \left(\frac{1}{2N} + \frac{1}{2} \right) y \nu \leq y \nu.
\end{aligned}$$

Thus we choose $\nu := \frac{\delta}{y}$. The same holds for $f_N(\epsilon) := \frac{1}{(2N)^d} \sum_{\omega \in \mathcal{G}_N} \exp(i\langle \epsilon(\frac{y}{N}), \omega \rangle)$. \square

Using Lemma 5 and 6 together with Theorem 7 we get the following Corollary.

Corollary 2. *Let ϵ_0 be given. For every $\delta > 0$ there exists $M = M(\epsilon_0, \delta)$ such that for all $N > M$*

$$|(D_N(\mathcal{X}))_{i,j} - (\text{Sinc}_\epsilon(\mathcal{Y}))_{i,j}| \leq \delta, \quad \forall \epsilon \in [0, \epsilon_0]. \quad (16)$$

Proof. We take ν in Definition 14 to be the minimum of all the $\nu_{i,j}^{(\ell)}$ according to the proof of Lemma 6, with $\ell = 1, \dots, d$. Note that for $y_{i,j}^{(\ell)} = 0$ we have $\nu_{i,j}^{(\ell)} = \infty$. \square

Corollary 3 (Uniform convergence of eigenvalues). *Let $\epsilon_0 > 0$ be given. For every $\delta' > 0$ there exists $M = M(\epsilon_0, \delta')$ s.t. for all $N > M$ there is an eigenvalue $\mu_k^{(N)}$ of $D_N(\mathcal{X})$ with*

$$|\lambda_k^{(\epsilon)} - \mu_k^{(N)}| \leq \delta', \quad \forall \epsilon \in [0, \epsilon_0] \quad (17)$$

where $\lambda_k^{(\epsilon)}$ is an eigenvalue of $\text{Sinc}_\epsilon(\mathcal{Y})$.

Proof. Since $\text{Sinc}_\epsilon(\mathcal{Y})$ is a symmetric matrix, the eigenvectors matrix E is orthogonal in the eigenvalue decomposition of $\text{Sinc}_\epsilon(\mathcal{Y})$ i.e $\text{Sinc}_\epsilon(\mathcal{Y}) = E\Lambda E^{-1}$. Thus we have $\kappa_2(E) = 1$. Using the Bauer–Fike theorem 6 with Corollary 2 and $\delta := \frac{\delta'}{n\sqrt{n}}$, we get:

$$\begin{aligned}
|\lambda - \mu| &\leq \|\text{Sinc}_\epsilon(\mathcal{Y}) - D_N(\mathcal{X})\|_2 \\
&\leq \sqrt{n} \|\text{Sinc}_\epsilon(\mathcal{Y}) - D_N(\mathcal{X})\|_\infty \\
&= \sqrt{n} \max_i \sum_{j=1}^n |(\text{Sinc}_\epsilon(\mathcal{Y}))_{i,j} - (D_N(\mathcal{X}))_{i,j}| \\
&\leq \sqrt{nn}\delta = \delta'.
\end{aligned}$$

\square

To complete the proof of Theorem 5, we choose $\delta' = O\left((\alpha\epsilon_0)^{2m}\right)$ in (17) and apply Corollary 1. The only remaining question is to show that $\det(RWR) \neq 0$. We will show that the limit $N \rightarrow \infty$ the rank of $W = \lim_{N \rightarrow \infty} \frac{1}{(2N)^d} W_N$ is full, and as a result, $\det\left(\tilde{R}W\tilde{R}^T\right)$ is uniformly bounded away from zero. By definition of the Wronskian, we write

$$\begin{aligned} \frac{1}{(2N)^d} W_N &= \frac{1}{(2N)^d} \left[\frac{\mathcal{D}_{\mathcal{G}_N}^{(\alpha,\beta)}\left(\frac{x}{N}, \frac{y}{N}\right)|_{x,y=0}}{\alpha!\beta!} \right]_{\alpha,\beta \in \mathbb{P}_m} = \left[\frac{1}{(2N)^d} \sum_{\omega \in \mathcal{G}_N} \frac{1}{N^{|\alpha|+|\beta|}} \frac{(i\omega)^\alpha (-i\omega)^\beta}{\alpha!\beta!} \right]_{\alpha,\beta \in \mathbb{P}_m} \\ &= \left[\frac{1}{(2N)^d} \sum_{\omega=(\omega_1, \dots, \omega_d) \in \mathcal{G}_N} \frac{(i)^{|\alpha|} (-i)^{|\beta|}}{\alpha!\beta!} \left(\frac{\omega_1}{N}\right)^{\alpha_1+\beta_1} \cdot \dots \cdot \left(\frac{\omega_d}{N}\right)^{\alpha_d+\beta_d} \right]_{\alpha,\beta \in \mathbb{P}_m} \\ &= \left[\prod_{j=1}^d \sum_{k=-N}^N \frac{1}{2N} \frac{(i)^{\alpha_j} (-i)^{\beta_j}}{\alpha_j!\beta_j!} \left(\frac{k}{N}\right)^{\alpha_j+\beta_j} \right]_{\alpha,\beta \in \mathbb{P}_m}. \end{aligned}$$

We get that each entry of $\frac{1}{(2N)^d} W_N$ is a Riemann sum, thus as $N \rightarrow \infty$ we get a multiple integral:

$$\lim_{N \rightarrow \infty} \prod_{j=1}^d \sum_{k=-N}^N \frac{1}{2N} \frac{(i)^{\alpha_j} (-i)^{\beta_j}}{\alpha_j!\beta_j!} \left(\frac{k}{N}\right)^{\alpha_j+\beta_j} = \frac{1}{2^d} \int_{[-1,1]^d} \frac{(i)^{|\alpha|} (-i)^{|\beta|}}{\alpha!\beta!} x^{\alpha+\beta} dx$$

Thus we get:

$$W = \lim_{N \rightarrow \infty} \frac{1}{(2N)^d} W_N = \left[\frac{1}{2^d} \int_{[-1,1]^d} \frac{(i)^{|\alpha|} (-i)^{|\beta|}}{\alpha!\beta!} x^{\alpha+\beta} dx \right]_{\alpha,\beta \in \mathbb{P}_m}.$$

It is immediately seen that W is none other than the Gram matrix for the monomials $\{x^\alpha\}_{\alpha \in \mathbb{P}_m}$:

$$W_{\alpha,\beta} = \langle x^\alpha, x^\beta \rangle = \frac{1}{2^d} \int_{[-1,1]^d} \frac{(i)^{|\alpha|} (-i)^{|\beta|}}{\alpha!\beta!} x^{\alpha+\beta} dx.$$

Since the set of monomials is linearly independent, we have that W is positive definite. Thus W is full rank, and as a result, $\det\left(\tilde{R}W\tilde{R}^T\right) > 0$. This completes the proof of Theorem 5. \square

Remark 6. *During the final stages of preparation of our manuscript, we became aware that Weilin Li is working on a related topic [24]. Relative to our results, more explicit conditions and constants are derived for the scaling of the smallest singular value of the multivariate Vandermonde matrix (we consider the entire spectrum); multi-cluster geometry is investigated; only cube and spherical sampling sets are considered; the decay rate in the case of uniform dilations of generic sets is not sharp; finally, a form of geometric characterization condition is used to describe the node geometry (as opposed to the sequence $\{t_k\}$ in our work).*

6 Numerical Experiments

In this section, we describe some numerical example to validate our results and investigate their applicability to the problem of super-resolution.

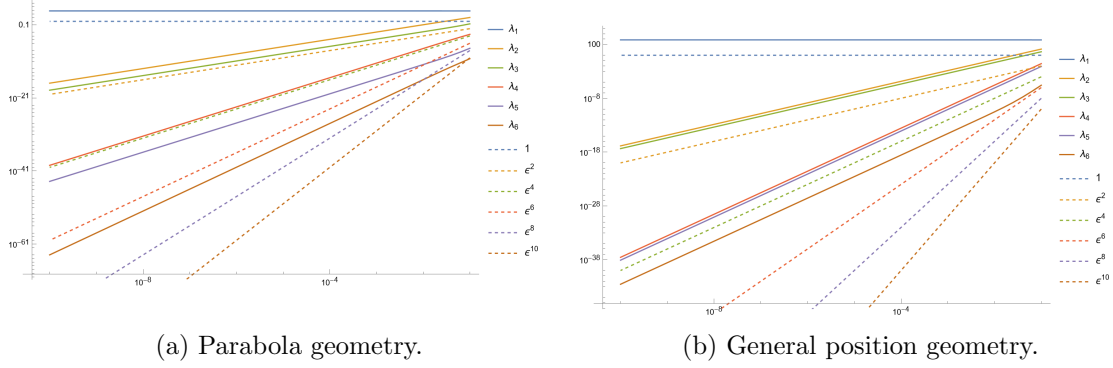


Figure 2: Eigenvalues of Dirichlet kernel with samples on the grid for the different node geometries. The x -axis is ϵ denoting the smallest distance between the nodes.

6.1 Dirichlet kernel eigenvalue scaling

We start with the numerical study of the eigenvalues of the Dirichlet matrix $D_\epsilon(\mathcal{X}, \mathcal{G}_n)$, confirming the scaling predicted by Lemma 3. Let $d = 2$, $n = 6$ and $\mathcal{X} = \{(x_i, y_i)\}_{i=1}^n$. We checked two scenarios:

1. All points satisfy $y_i = x_i^2$. We get that $\text{rank}(V_{\leq 3}) = 6$, where

$$\begin{aligned}
 V_{\leq 3} &= \begin{bmatrix} 1 & x_1 & y_1 & x_1^2 & x_1 y_1 & y_1^2 & x_1^3 & x_1^2 y_1 & x_1 y_1^2 & y_1^3 \\ 1 & x_2 & y_2 & x_2^2 & x_2 y_2 & y_2^2 & x_2^3 & x_2^2 y_2 & x_2 y_2^2 & y_2^3 \\ 1 & x_3 & y_3 & x_3^2 & x_3 y_3 & y_3^2 & x_3^3 & x_3^2 y_3 & x_3 y_3^2 & y_3^3 \\ 1 & x_4 & y_4 & x_4^2 & x_4 y_4 & y_4^2 & x_4^3 & x_4^2 y_4 & x_4 y_4^2 & y_4^3 \\ 1 & x_5 & y_5 & x_5^2 & x_5 y_5 & y_5^2 & x_5^3 & x_5^2 y_5 & x_5 y_5^2 & y_5^3 \\ 1 & x_6 & y_6 & x_6^2 & x_6 y_6 & y_6^2 & x_6^3 & x_6^2 y_6 & x_6 y_6^2 & y_6^3 \end{bmatrix} \\
 &= \begin{bmatrix} 1 & x_1 & x_1^2 & x_1^3 & x_1^4 & x_1^5 & x_1^3 & x_1^3 & x_1^4 & x_1^6 \\ 1 & x_2 & x_2^2 & x_2^3 & x_2^4 & x_2^5 & x_2^3 & x_2^3 & x_2^4 & x_2^6 \\ 1 & x_3 & x_3^2 & x_3^3 & x_3^4 & x_3^5 & x_3^3 & x_3^3 & x_3^4 & x_3^6 \\ 1 & x_4 & x_4^2 & x_4^3 & x_4^4 & x_4^5 & x_4^3 & x_4^3 & x_4^4 & x_4^6 \\ 1 & x_5 & x_5^2 & x_5^3 & x_5^4 & x_5^5 & x_5^3 & x_5^3 & x_5^4 & x_5^6 \\ 1 & x_6 & x_6^2 & x_6^3 & x_6^4 & x_6^5 & x_6^3 & x_6^3 & x_6^4 & x_6^6 \end{bmatrix} = [V_0 \mid V_1 \mid V_2 \mid V_3]
 \end{aligned}$$

We get that

$$\begin{aligned}
 t_0 &= \text{rank}(V_{\leq 0}) = 1, \\
 t_1 &= \text{rank}(V_{\leq 1}) - \text{rank}(V_{\leq 0}) = 3 - 1 = 2, \\
 t_2 &= \text{rank}(V_{\leq 2}) - \text{rank}(V_{\leq 1}) = 5 - 3 = 2, \\
 t_3 &= \text{rank}(V_{\leq 3}) - \text{rank}(V_{\leq 2}) = 6 - 5 = 1.
 \end{aligned}$$

The results in Figure 2a confirm the scaling of the eigenvalues.

2. The set of nodes in \mathcal{X} are in general positions. We have $\text{rank}(V_{\leq 2}) = 6$, where

$$V_{\leq 2} = \left[\begin{array}{c|cc|ccc} 1 & x_1 & y_1 & x_1^2 & x_1 y_1 & y_1^2 \\ 1 & x_2 & y_2 & x_2^2 & x_2 y_2 & y_2^2 \\ 1 & x_3 & y_3 & x_3^2 & x_3 y_3 & y_3^2 \\ 1 & x_4 & y_4 & x_4^2 & x_4 y_4 & y_4^2 \\ 1 & x_5 & y_5 & x_5^2 & x_5 y_5 & y_5^2 \\ 1 & x_6 & y_6 & x_6^2 & x_6 y_6 & y_6^2 \end{array} \right] = [V_0 \mid V_1 \mid V_2].$$

We get that

$$\begin{aligned} t_0 &= \text{rank}(V_{\leq 0}) = 1, \\ t_1 &= \text{rank}(V_{\leq 1}) - \text{rank}(V_{\leq 0}) = 3 - 1 = 2, \\ t_2 &= \text{rank}(V_{\leq 2}) - \text{rank}(V_{\leq 1}) = 6 - 3 = 3. \end{aligned}$$

The results in Figure 2b confirm the scaling of the eigenvalues also in this case.

6.2 Super-resolution

In this section, we consider the super-resolution problem for multidimensional sparse measures described in the Introduction, as it was the original motivation for our investigations. We reconstruct measures of the form (1) from the noisy measurements (2). We consider the asymptotic regime of constant N and $\Delta \rightarrow 0$.

6.2.1 Local Stability

Analogous to some previous studies in the one-dimensional case, e.g. [2, 4], we consider the “local stability” of the problem to be well-approximated by the error incurred by the nonlinear least squares (NLS) method applied to the noisy measurements:

$$\{\hat{x}_j^{NLS}, \hat{\alpha}_j^{NLS}\}_{j=1}^n = \arg \min_{\hat{\alpha}, \hat{x}} \frac{1}{2} \sum_{k \in \mathcal{G}_N} \left| \hat{f}(k) - \sum_{j=1}^n \hat{\alpha}_j e^{i\langle k, \hat{x}_j \rangle} \right|^2.$$

The initial values for the NLS method are taken to be the true parameter values (which are clearly unknown in practice). Our implementation of the NLS method is based on the Levenberg-Marquardt algorithm, with the complex residuals converted to real residuals by concatenating the real and imaginary parts. The results are shown in Figures 3 and 4. To measure the accuracy, we fitted the scaled error

$$\kappa := \max_{j=1, \dots, n} \|\hat{x}_j^{NLS} - x_j\|_{\infty} / \sigma$$

to the curve $\kappa \sim \Delta^m$, where Δ is the minimal distance between the nodes. It can be seen that the accuracy crucially depends on the geometry, and is indeed best for the random configuration, and worst for the single-line configuration (for which the scaling is precisely $\kappa \sim \Delta^{2n-2}$ for any d , as established in previous works on the subject).

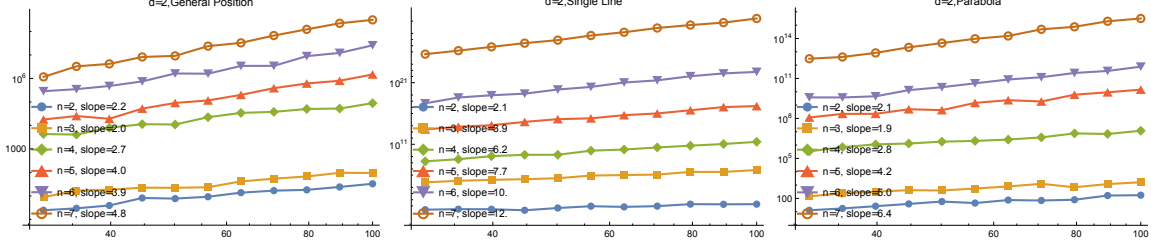


Figure 3: NLS reconstruction, general position/line/parabola, $d = 2$.

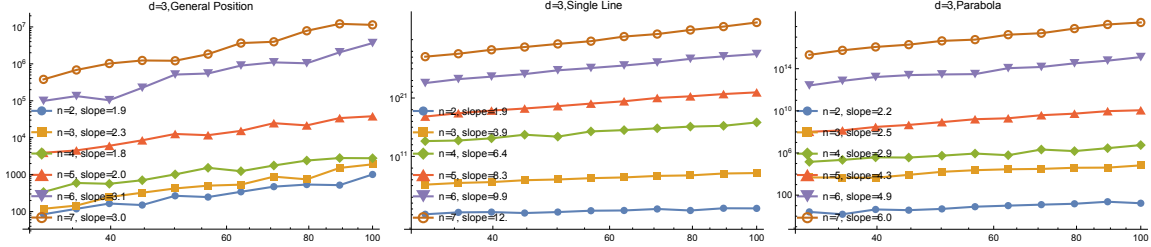


Figure 4: NLS reconstruction, general position/line/parabola, $d = 3$.

6.2.2 2D-ESPRIT

We have implemented the 2D ESPRIT algorithm according to [36, Section IV-A]. The experimental setup was as follows:

1. The number of nodes varied from $n = 2$ to $n = 6$.
2. The geometry was either random, a line $y = 5x$, or a parabola $y = 10x^2$.
3. The ESPRIT algorithm parameters were set as follows: $M_1 = M_2 = 40$, $L_1 = L_2 = 10$, $\beta_1 = \beta_2 = 0.5$.
4. The noise level was set to $\sigma = 10^{-20}$.

The results are shown in Figure 5. As for the NLS method, we fitted the scaled error

$$\kappa := \max_{j=1, \dots, n} \|\hat{x}_j^{ESPRIT} - x_j\|_{\infty} / \sigma$$

to the curve $\kappa \sim \Delta^m$, where Δ is the minimal distance between the nodes. As in the NLS case, it can be seen that the accuracy crucially depends on the geometry, providing best scaling for the random configuration, and worst scaling for the single-line configuration.

A Taylor Expansion

Let $\mathcal{K} \in \mathcal{C}^{(r,r)}(\Omega)$ where Ω is an open neighborhood of 0 and $r \geq m + 1$, where $m = \mu(\mathcal{X})$ is the discrete moment order of \mathcal{X} (Definition 5). We recall the Maclaurin expansion for bivariate functions from [1, section 5]. For $\epsilon x, \epsilon y$ such that $[0, \epsilon x], [0, \epsilon y] \subset \Omega$ (where $[0, x]$ is a line segment

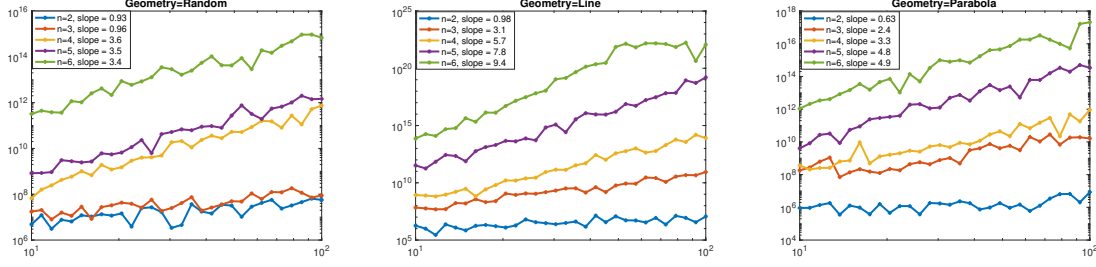


Figure 5: Accuracy of 2D ESPRIT for different geometries.

from 0 to x), we have:

$$\begin{aligned}
\mathcal{K}_\epsilon(x, y) &= \mathcal{K}(\epsilon x, \epsilon y) = \sum_{\alpha, \beta \in \mathbb{P}_m} \frac{(\epsilon x)^\alpha (\epsilon y)^\beta}{\alpha! \beta!} \mathcal{K}^{(\alpha, \beta)}(0, 0) \\
&+ \sum_{\alpha \in \mathbb{P}_m, \beta \in \mathbb{H}_{m+1}} \frac{(\epsilon x)^\alpha (\epsilon y)^\beta}{\alpha! \beta!} \mathcal{K}^{(\alpha, \beta)}(0, \theta_{\epsilon y, \alpha} \epsilon y) + \sum_{\alpha \in \mathbb{H}_{m+1}, \beta \in \mathbb{P}_m} \frac{(\epsilon x)^\alpha (\epsilon y)^\beta}{\alpha! \beta!} \mathcal{K}^{(\alpha, \beta)}(\eta_{\epsilon x, \beta} \epsilon x, 0) \\
&+ \sum_{\alpha, \beta \in \mathbb{H}_{m+1}} \frac{(\epsilon x)^\alpha (\epsilon y)^\beta}{\alpha! \beta!} \mathcal{K}^{(\alpha, \beta)}(\psi_{\epsilon x, \epsilon y} \epsilon x, \xi_{\epsilon x, \epsilon y} \epsilon y),
\end{aligned}$$

where $\{\theta_{\epsilon y, \alpha}\}_{\alpha \in \mathbb{P}_m} \subset [0, 1]$ depend on ϵy , $\{\eta_{\epsilon x, \beta}\}_{\beta \in \mathbb{P}_m} \subset [0, 1]$ depend on ϵx and $\psi_{\epsilon x, \epsilon y}, \xi_{\epsilon x, \epsilon y}$ depend on both ϵx and ϵy .

Let $\mathbb{P}_m = \{\alpha_1, \dots, \alpha_{p_m}\}$. Then we can write the above expansion as follows:

$$\begin{aligned}
\mathcal{K}(\epsilon x, \epsilon y) &= [(\epsilon x)^{\alpha_1}, \dots, (\epsilon x)^{\alpha_{p_m}}] W_{\leq m} [(\epsilon y)^{\alpha_1}, \dots, (\epsilon y)^{\alpha_{p_m}}]^T \\
&+ \epsilon^{m+1} [(\epsilon x)^{\alpha_1}, \dots, (\epsilon x)^{\alpha_{p_m}}] \mathbf{w}_{1,y}(\epsilon) + \epsilon^{m+1} \mathbf{w}_{2,x}(\epsilon)^T [(\epsilon y)^{\alpha_1}, \dots, (\epsilon y)^{\alpha_{p_m}}]^T \\
&+ \epsilon^{2(m+1)} w_{3,x,y}, \\
\mathbf{w}_{1,y}(\epsilon) &= \left[\sum_{\beta \in \mathbb{H}_{m+1}} \frac{\mathcal{K}^{(\alpha_1, \beta)}(0, \theta_{\epsilon y, \alpha_1} \epsilon y)}{\alpha_1! \beta!} y^\beta, \dots, \sum_{\beta \in \mathbb{H}_{m+1}} \frac{\mathcal{K}^{(\alpha_{p_m}, \beta)}(0, \theta_{\epsilon y, \alpha_{p_m}} \epsilon y)}{\alpha_{p_m}! \beta!} y^\beta \right]^T \\
\mathbf{w}_{2,x}(\epsilon) &= \left[\sum_{\alpha \in \mathbb{H}_{m+1}} \frac{\mathcal{K}^{(\alpha, \beta)}(\eta_{\epsilon x, \beta}(\epsilon x), 0)}{\alpha! \beta!} x^\alpha \right]_{\beta \in \mathbb{P}_m}^T \\
w_{3,x,y}(\epsilon) &= \sum_{\alpha, \beta \in \mathbb{H}_{m+1}} \frac{\mathcal{K}^{(\alpha, \beta)}(\psi_{\epsilon x, \epsilon y}(\epsilon x), \xi_{\epsilon x, \epsilon y}(\epsilon y))}{\alpha! \beta!} (\epsilon x)^\alpha (\epsilon y)^\beta
\end{aligned}$$

where $\mathbf{w}_{2,x}, \mathbf{w}_{1,y} : [0, \epsilon_0] \rightarrow \mathbb{R}^{p_m}$ are bounded vector functions with $\epsilon_0 > 0$ such that $\epsilon_0 x, \epsilon_0 y \in \Omega$, $\epsilon \in [0, \epsilon_0]$ and $w_{3,x,y}$ is bounded.

Let $\epsilon_0 > 0$ such that $\epsilon_0 x_i \in \Omega$ for $x_i \in \mathcal{X}$. Thus for $0 \leq \epsilon \leq \epsilon_0$, the scaled kernel matrix K_ϵ admits the following expansion:

$$K_\epsilon = V_{\leq m} \Delta_m W \Delta_m V_{\leq m}^T + \epsilon^{m+1} (V_{\leq m} \Delta_m W_1(\epsilon) + W_2(\epsilon) \Delta_m V_{\leq m}^T) + \epsilon^{2(m+1)} W_3(\epsilon) \quad (18)$$

where $W_1(\epsilon) := [\mathbf{w}_{1,x_1}(\epsilon), \dots, \mathbf{w}_{1,x_n}(\epsilon)]$, $W_2(\epsilon) := [\mathbf{w}_{2,x_1}(\epsilon), \dots, \mathbf{w}_{2,x_n}(\epsilon)]^T$ and $W_3(\epsilon) = [w_{3,x_i,x_j}(\epsilon)]_{i,j=1}^{n,n}$.

B Auxiliary Lemmas

Lemma 7. Let $D \in \mathbb{R}^{k \times k}$ be a $k \times k$ diagonal matrix and P a permutation matrix corresponding to a permutation σ , then

$$PDP^T = \text{diag}\{d_{\sigma(1)}, \dots, d_{\sigma(k)}\}, \quad D = \text{diag}\{d_1, \dots, d_k\}.$$

Proof. If P is a permutation matrix corresponding to a permutation σ , then for every vector $v \in \mathbb{R}^k$

$$Pv = \begin{bmatrix} v_{\sigma(1)} \\ \vdots \\ v_{\sigma(k)} \end{bmatrix}$$

and then for every matrix $A = \begin{bmatrix} R_1 \\ \vdots \\ R_k \end{bmatrix} \in \mathbb{R}^{k \times k}$, we have

$$PA = \begin{bmatrix} R_{\sigma(1)} \\ \vdots \\ R_{\sigma(k)} \end{bmatrix}.$$

Now let $A = D = \text{diag}\{d_1, \dots, d_k\}$, and take $j = 1, \dots, k$. Then $R_{\sigma(j)} = d_{\sigma(j)}e_{\sigma(j)}^T$. Therefore

$$R_{\sigma(j)}P^T = d_{\sigma(j)}(Pe_{\sigma(j)})^T = d_{\sigma(j)}e_j^T.$$

Stacking all the rows we get $PDP^T = \text{diag}\{d_{\sigma(1)}, \dots, d_{\sigma(k)}\}$. □

Lemma 8. Let $M, B \in \mathbb{R}^{n \times n}$. For $\epsilon > 0$ we have

$$\det(M + \epsilon B) = \det(M) + O(\epsilon), \quad \epsilon \ll 1. \quad (19)$$

Proof. Recall formula (0.8.12.3) for sum of matrices in [19]:

$$\det(M + \epsilon B) = \sum_{k=0}^n \epsilon^k \text{tr}(\text{adj}_k(M)C_k(B)), \quad (20)$$

where $\text{tr}(\cdot)$ is the trace of matrix, $\text{adj}_k(\cdot)$ is the k -th adjugate of matrix (0.8.12) and $C_k(\cdot)$ is the k -th compound (0.8.1). Using the following properties: $\text{adj}_0(M) = \det(M)$ and $C_0(M) = 1$, we have

$$\det(M + \epsilon B) = \epsilon^0 \text{tr}(\text{adj}_0(M)C_0(B)) + O(\epsilon) = \det(M) + O(\epsilon), \quad (21)$$

completing the proof. □

References

- [1] Simon Barthelmé and Konstantin Usevich. “Spectral properties of kernel matrices in the flat limit”. In: SIAM Journal on Matrix Analysis and Applications 42.1 (2021), pp. 17–57. DOI: 10.1137/19M129677X. URL: <https://doi.org/10.1137/19M129677X>.
- [2] D. Batenkov and Y. Yomdin. “On the Accuracy of Solving Confluent Prony Systems”. In: SIAM J. Appl. Math. 73.1 (2013), pp. 134–154. DOI: 10.1137/110836584.
- [3] Dmitry Batenkov. “Decimated Generalized Prony Systems”. Aug. 3, 2013. arXiv: 1308.0753 [math]. URL: <http://arxiv.org/abs/1308.0753> (visited on 06/01/2014).
- [4] Dmitry Batenkov. “Stability and Super-Resolution of Generalized Spike Recovery”. In: Applied and Computational Harmonic Analysis 45.2 (Sept. 1, 2018), pp. 299–323. ISSN: 1063-5203. DOI: 10.1016/j.acha.2016.09.004.
- [5] Dmitry Batenkov, Laurent Demanet, Gil Goldman, and Yosef Yomdin. “Conditioning of Partial Nonuniform Fourier Matrices with Clustered Nodes”. In: SIAM Journal on Matrix Analysis and Applications 44.1 (Jan. 1, 2020), pp. 199–220. ISSN: 0895-4798. DOI: 10/ggjwzb.
- [6] Dmitry Batenkov and Nuha Diab. “Super-Resolution of Generalized Spikes and Spectra of Confluent Vandermonde Matrices”. In: Applied and Computational Harmonic Analysis 65 (July 1, 2023), pp. 181–208. ISSN: 1063-5203. DOI: 10.1016/j.acha.2023.03.002.
- [7] Dmitry Batenkov, Benedikt Diederichs, Gil Goldman, and Yosef Yomdin. “The Spectral Properties of Vandermonde Matrices with Clustered Nodes”. In: Linear Algebra and its Applications 609 (Jan. 15, 2021), pp. 37–72. ISSN: 0024-3795. DOI: 10.1016/j.laa.2020.08.034.
- [8] Dmitry Batenkov, Gil Goldman, and Yosef Yomdin. “Super-Resolution of near-Colliding Point Sources”. In: Information and Inference: A Journal of the IMA 10.2 (June 15, 2021), pp. 515–572. DOI: 10.1093/imaiai/iaaa005.
- [9] M. Bertero and C. de Mol. “III Super-Resolution by Data Inversion”. In: Progress in Optics 36 (1996), pp. 129–178.
- [10] K. C. Chung and T. H. Yao. “On Lattices Admitting Unique Lagrange Interpolations”. In: SIAM Journal on Numerical Analysis 14.4 (1977), pp. 735–743. DOI: 10.1137/0714050.
- [11] Annie Cuyt and Wen-shin Lee. “How to Get High Resolution Results from Sparse and Coarsely Sampled Data”. In: Applied and Computational Harmonic Analysis (Oct. 2018). ISSN: 1063-5203. DOI: 10/ggb5cv.
- [12] Laurent Demanet and Nam Nguyen. “The Recoverability Limit for Superresolution via Sparsity”. In: arXiv preprint arXiv:1502.01385 (2015). arXiv: 1502.01385.
- [13] Benedikt Diederichs, Mihail N. Kolountzakis, and Effie Papageorgiou. “How Many Fourier Coefficients Are Needed?” In: Monatshefte für Mathematik (Oct. 21, 2022). ISSN: 0026-9255, 1436-5081. DOI: 10.1007/s00605-022-01792-0. (Visited on 11/08/2022).
- [14] D.L. Donoho. “Superresolution via Sparsity Constraints”. In: SIAM Journal on Mathematical Analysis 23.5 (1992), pp. 1309–1331.
- [15] T.A. Driscoll and B. Fornberg. “Interpolation in the Limit of Increasingly Flat Radial Basis Functions”. In: Computers & Mathematics with Applications 43.3-5 (Feb. 2002), pp. 413–422. ISSN: 08981221. DOI: 10.1016/S0898-1221(01)00295-4.

- [16] Hernán García, Camilo Hernández, Mauricio Junca, and Mauricio Velasco. “Approximate Super-Resolution of Positive Measures in All Dimensions”. In: Applied and Computational Harmonic Analysis (Feb. 11, 2020). ISSN: 1063-5203. DOI: 10/ggkz93.
- [17] Mariano Gasca and Thomas Sauer. “Polynomial Interpolation in Several Variables”. In: Advances in Computational Mathematics 12.4 (Mar. 1, 2000), pp. 377–410. ISSN: 1572-9044. DOI: 10.1023/A:1018981505752.
- [18] Gene H Golub and Charles F Van Loan. Matrix Computations. Vol. 3. JHU Press, 2013.
- [19] Roger A Horn and Charles R Johnson. Matrix analysis. Cambridge university press, 2013.
- [20] Rami Katz, Nuha Diab, and Dmitry Batenkov. “Decimated Prony’s Method for Stable Super-Resolution”. In: IEEE Signal Processing Letters 30 (2023), pp. 1467–1471. ISSN: 1558-2361. DOI: 10.1109/LSP.2023.3324553.
- [21] Rami Katz, Nuha Diab, and Dmitry Batenkov. On the Accuracy of Prony’s Method for Recovery of Exponential Sums. Apr. 16, 2024. DOI: 10.48550/arXiv.2302.05883. arXiv: 2302.05883 [cs, math]. preprint.
- [22] Stefan Kunis and Dominik Nagel. “On the Smallest Singular Value of Multivariate Vandermonde Matrices with Clustered Nodes”. In: Linear Algebra and its Applications 604 (Nov. 1, 2020), pp. 1–20. ISSN: 0024-3795. DOI: 10.1016/j.laa.2020.06.003. (Visited on 07/14/2020).
- [23] Stefan Kunis, Dominik Nagel, and Anna Strotmann. “Multivariate Vandermonde Matrices with Separated Nodes on the Unit Circle Are Stable”. In: Applied and Computational Harmonic Analysis 58 (May 1, 2022), pp. 50–59. ISSN: 1063-5203. DOI: 10.1016/j.acha.2022.01.001.
- [24] Weilin Li. Nonharmonic Multivariate Fourier Transforms and Matrices: Condition Numbers and Hyperplane Sampling. 2024. Pre-published.
- [25] Weilin Li and Wenjing Liao. “Stable super-resolution limit and smallest singular value of restricted Fourier matrices”. In: Applied and Computational Harmonic Analysis 51 (2020), pp. 118–156. ISSN: 1063-5203. DOI: 10.1016/j.acha.2020.10.004.
- [26] Weilin Li, Wenjing Liao, and Albert Fannjiang. “Super-Resolution Limit of the ESPRIT Algorithm”. In: IEEE Transactions on Information Theory (2020). ISSN: 1557-9654. DOI: 10/ggrnpw.
- [27] J. Lindberg. “Mathematical concepts of optical superresolution”. In: Journal of Optics 14 (2012).
- [28] Ping Liu and Hai Zhang. “A Mathematical Theory of Computational Resolution Limit in Multi-Dimensional Spaces”. In: Inverse Problems (Sept. 7, 2021). ISSN: 0266-5611, 1361-6420. DOI: 10.1088/1361-6420/ac245b. (Visited on 09/11/2021).
- [29] Ping Liu and Hai Zhang. “A Mathematical Theory of the Computational Resolution Limit in One Dimension”. In: Applied and Computational Harmonic Analysis 56 (Jan. 1, 2022), pp. 402–446. ISSN: 1063-5203. DOI: 10.1016/j.acha.2021.09.002.
- [30] Charles A. Micchelli. “Interpolation of Scattered Data: Distance Matrices and Conditionally Positive Definite Functions”. In: Constructive Approximation 2.1 (Dec. 1, 1986), pp. 11–22. ISSN: 1432-0940. DOI: 10.1007/BF01893414.
- [31] R. A. Nicolaidis. “On a Class of Finite Elements Generated by Lagrange Interpolation”. In: SIAM Journal on Numerical Analysis 9.3 (1972), pp. 435–445. DOI: 10.1137/0709039.

- [32] Juan Manuel Peña and Tomas Sauer. “Stabilized Recovery and Model Reduction for Multivariate Exponential Polynomials”. In: Journal of Symbolic Computation (Mar. 2024), p. 102313. ISSN: 07477171. DOI: 10.1016/j.jsc.2024.102313.
- [33] Markus Petz, Gerlind Plonka, and Nadiia Derevianko. “Exact Reconstruction of Sparse Non-Harmonic Signals from Their Fourier Coefficients”. In: Sampling Theory, Signal Processing, and Data Analysis 19.1 (May 2021), p. 7. ISSN: 2730-5724. DOI: 10.1007/s43670-021-00007-1.
- [34] Clarice Poon and Gabriel Peyré. “MultiDimensional Sparse Super-Resolution”. In: SIAM Journal on Mathematics 51.1 (Jan. 2019), pp. 1–44. ISSN: 0036-1410, 1095-7154. DOI: 10.1137/17M1147822.
- [35] Walter Rudin et al. Principles of mathematical analysis. Vol. 3. McGraw-hill New York, 1964.
- [36] Souleymen Sahnoun, Konstantin Usevich, and Pierre Comon. “Multidimensional ESPRIT for Damped and Undamped Signals: Algorithm, Computations, and Perturbation Analysis”. In: IEEE Transactions on Signal Processing 65.22 (Nov. 15, 2017), pp. 5897–5910. ISSN: 1053-587X, 1941-0476. DOI: 10.1109/TSP.2017.2736512.
- [37] Tomas Sauer. “Polynomial Interpolation in Several Variables: Lattices, Differences, and Ideals”. In: Studies in Computational Mathematics. Vol. 12. Elsevier, 2006, pp. 191–230. ISBN: 978-0-444-51844-6. DOI: 10.1016/S1570-579X(06)80009-1.
- [38] Tomas Sauer. “Prony’s Method in Several Variables”. en. In: Numerische Mathematik 136.2 (June 2017), pp. 411–438. ISSN: 0029-599X, 0945-3245. DOI: 10.1007/s00211-016-0844-8.
- [39] Tomas Sauer. “Prony’s Method in Several Variables: Symbolic Solutions by Universal Interpolation”. en. In: Journal of Symbolic Computation 84 (Jan. 2018), pp. 95–112. ISSN: 07477171. DOI: 10.1016/j.jsc.2017.03.006.
- [40] Robert Schaback. “Multivariate interpolation by polynomials and radial basis functions”. In: Constructive approximation 21.3 (2005), pp. 293–317. ISSN: 0176-4276.
- [41] Andrew J Wathen and Shengxin Zhu. “On spectral distribution of kernel matrices related to radial basis functions”. In: Numerical Algorithms 70.4 (2015), pp. 709–726.