A multivariate generalization of Prony's method

Stefan Kunis[a,b,*], Thomas Peter[a], Tim Römer[a], Ulrich von der Ohe[a]

[a] *Osnabrück University, Institute of Mathematics*
[b] *Helmholtz Zentrum München, Institute of Computational Biology*

Abstract

Prony's method is a prototypical eigenvalue analysis based method for the reconstruction of a finitely supported complex measure on the unit circle from its moments up to a certain degree. In this note, we give a generalization of this method to the multivariate case and prove simple conditions under which the problem admits a unique solution. Provided the order of the moments is bounded from below by the number of points on which the measure is supported as well as by a small constant divided by the separation distance of these points, stable reconstruction is guaranteed. In its simplest form, the reconstruction method consists of setting up a certain multilevel Toeplitz matrix of the moments, compute a basis of its kernel, and compute by some method of choice the set of common roots of the multivariate polynomials whose coefficients are given in the second step. All theoretical results are illustrated by numerical experiments.

Keywords: frequency analysis, spectral analysis, exponential sum, moment problem, super-resolution
2010 MSC: 65T40, 42C15, 30E05, 65F30

1. Introduction

In this paper we propose a generalization of de Prony's classical method [10] for the parameter and coefficient reconstruction of univariate finitely supported complex measures to a finite number of variables. The method of de Prony lies at the core of seemingly different classes of problems in signal processing such as spectral estimation, search for an annihilating filter, deconvolution, spectral extrapolation, and moment problems. Thus we provide a new tool to analyze multivariate versions of a broad set of problems.

To recall the machinery of the classical Prony method let $\mathbb{C}_* := \mathbb{C} \setminus \{0\}$ and let $\hat{f}_j \in \mathbb{C}_*$ and pairwise distinct $z_j \in \mathbb{C}_*$, $j = 1, \ldots, M$, be given. Let δ_{z_j}

*Corresponding author
Email addresses: `skunis@uos.de` (Stefan Kunis), `petert@uos.de` (Thomas Peter), `troemer@uos.de` (Tim Römer), `uvonderohe@uos.de` (Ulrich von der Ohe)

denote the Dirac measure in z_j on $\mathbb{C}_*$ and let

$$\mu = \sum_{j=1}^{M} \hat{f}_j \delta_{z_j}$$

be a finitely supported complex measure on $\mathbb{C}_*$. By the Prony method the $\hat{f}_j$ and z_j are reconstructed from $2M+1$ moments

$$f(k) = \int_{\mathbb{C}_*} x^k \mathrm{d}\mu(x) = \sum_{j=1}^{M} \hat{f}_j z_j^k, \quad k = -M, \ldots, M.$$

Since the coefficients $\hat{p}_\ell \in \mathbb{C}$, $\ell = 0, \ldots, M$, of the (not a priori known) so-called Prony polynomial

$$p(Z) := \prod_{j=1}^{M} (Z - z_j) = \sum_{\ell=0}^{M} \hat{p}_\ell Z^\ell$$

fulfill the linear equations

$$\sum_{\ell=0}^{M} \hat{p}_\ell f(\ell - m) = \sum_{j=1}^{M} \hat{f}_j z_j^{-m} \sum_{\ell=0}^{M} \hat{p}_\ell z_j^\ell = \sum_{j=1}^{M} \hat{f}_j z_j^{-m} p(z_j) = 0, \quad m = 0, \ldots, M,$$

and are in fact the unique solution to this system with $\hat{p}_M = 1$, reconstruction of the z_j is possible by computing the kernel vector $(\hat{p}_1, \ldots, \hat{p}_{M-1}, 1)$ of the rank-M Toeplitz matrix

$$T := (f(k - \ell))_{\substack{\ell=0,\ldots,M \\ k=0,\ldots,M}} \in \mathbb{C}^{M+1 \times M+1}$$

and, knowing this to be the coefficient vector of p, compute the roots z_j of p. Afterwards, the coefficients $\hat{f}_j$ of f (that did not enter the discussion until now) can be uniquely recovered by solving a Vandermonde linear system of equations. When attempting to generalize this method to finitely supported complex measures on $\mathbb{C}_*^d$, it seems natural to think that the unknown parameters $z_j \in \mathbb{C}_*^d$ could be realized as roots of d-variate polynomials, and this is the approach we will follow here. As in the univariate case, the coefficients will be given as solution to a suitably constructed system of linear equations. However, for $d \geq 2$, an added difficulty lies in the fact that a non-constant polynomial always has uncountably many complex roots, so that a single polynomial cannot be sufficient to identify the parameters as its roots. A natural way to overcome this problem is to consider the common roots of a (finite) set of polynomials. These sets, commonly called algebraic varieties, are the subject of classical algebraic geometry and thus there is an immense body of algebraic literature on this topic from which we need only some basic notions as provided at the end of Section 2.

Our main results are presented in Section 3, which is divided into three parts. In Section 3.1 we prove sufficient conditions to guarantee parameter reconstruction for multivariate exponential sums by constructing a set of multivariate polynomials such that their common roots are precisely the parameters. In Section 3.2 we focus on the case that the parameters z_j lie on the d-dimensional torus, which allows us to prove numerical stability, provide some implications on the parameter distribution, and construct a single trigonometric polynomial localized at the parameters. Finally, we state a prototypical algorithm of the multivariate Prony method.

In Section 4 we discuss previous approaches towards the multivariate moment problem as can be found in [19, 1] for generic situations, in [30, 28, 11] based on projections of the measure, and in [7, 6, 5, 4] based on semidefinite optimization. Finally, numerical examples are presented in Section 5 and we conclude the paper with a summary in Section 6.

2. Preliminaries

Throughout the paper, the letter $d \in \mathbb{N}$ always denotes the dimension and we let

$$\mathbb{C}_*^d := (\mathbb{C}_*)^d = \{z \in \mathbb{C}^d : z_\ell \neq 0 \text{ for all } \ell = 1, \dots, d\}$$

be the domain for our parameters. For $z \in \mathbb{C}_*^d$, $k \in \mathbb{Z}^d$, we use the multi-index notation $z^k := z_1^{k_1} \cdot \ldots \cdot z_d^{k_d}$. We also let $\mathbb{T} := \{z \in \mathbb{C} : |z| = 1\}$ and the d-fold Cartesian product $\mathbb{T}^d$ is called *d-dimensional torus*. We start by defining the object of our interest, that is, multivariate exponential sums, as a natural generalization of univariate exponential sums.

Definition 2.1. *A function $f\colon \mathbb{Z}^d \to \mathbb{C}$ is a d-*variate exponential sum *if there is a finitely supported complex measure μ on $\mathbb{C}_*^d$, such that for all $k \in \mathbb{Z}^d$, $f(k)$ is the k-th moment of μ, that is, if there are $M \in \mathbb{N}$, $\hat{f}_1, \dots, \hat{f}_M \in \mathbb{C}_*$, and pairwise distinct $z_1, \dots, z_M \in \mathbb{C}_*^d$ such that with $\mu := \sum_{j=1}^M \hat{f}_j \delta_{z_j}$ we have*

$$f(k) = \int_{\mathbb{C}_*^d} x^k \mathrm{d}\mu(x) = \sum_{j=1}^M \hat{f}_j z_j^k$$

for all $k \in \mathbb{Z}^d$.

*In that case M, $\hat{f}_j$, and z_j, $j = 1, \dots, M$, are uniquely determined, and f is called M-*sparse*, the $\hat{f}_j$ are called* coefficients *of f, and z_j are called* parameters *of f. The set of parameters of f is denoted by Ω_f or, if there is no danger of confusion, simply by Ω.*

Remark 2.2. *Let $\hat{f}_j \in \mathbb{C}_*$ and pairwise distinct $t_j \in [0,1)^d$, $j = 1, \dots, M$, be given. Then the trigonometric moment sequence of $\tau = \sum_{j=1}^M \hat{f}_j \delta_{t_j}$,*

$$f\colon \mathbb{Z}^d \to \mathbb{C}, \quad k \mapsto \int_{[0,1)^d} \mathrm{e}^{2\pi \mathrm{i} kt} \mathrm{d}\tau(t) = \sum_{j=1}^M \hat{f}_j \mathrm{e}^{2\pi \mathrm{i} k t_j},$$

(where kt denotes the scalar product of k and t) is a d-variate exponential sum with parameters $\mathrm{e}^{2\pi \mathrm{i} t_j} = (\mathrm{e}^{2\pi \mathrm{i} t_{j,1}}, \dots, \mathrm{e}^{2\pi \mathrm{i} t_{j,d}}) \in \mathbb{T}^d$. This case will be analyzed in detail in Section 3.2.

Let $f\colon \mathbb{Z}^d \to \mathbb{C}$ be an M-sparse d-variate exponential sum with coefficients $\hat{f}_j \in \mathbb{C}_*$ and parameters $z_j \in \mathbb{C}_*^d$, $j = 1, \ldots, M$. Our objective is to reconstruct the coefficients and parameters of f given an upper bound n for M and a finite set of samples of f at a subset of $\mathbb{Z}^d$ that depends only on n, see also [26].

The following notations will be used throughout the paper. For $n \in \mathbb{N}$, let $I_n := \{0, \ldots, n\}^d$ and let $N := |I_n| = (n+1)^d$. The multilevel Toeplitz matrix

$$T_n(f) := (f(k-\ell))_{k,\ell \in I_n} \in \mathbb{C}^{N \times N},$$

which we also refer to as T_n, will play a crucial role in the multivariate Prony method. Note that the entries of T_n are sampling values of f on the grid $I_n - I_n = \{-n, \ldots, n\}^d$ having $(2n+1)^d$ points in total.

Next we establish the crucial link between the matrix T_n and the roots of multivariate polynomials. To this end, let

$$\Pi := \mathbb{C}[Z_1, \ldots, Z_d] = \{\sum_{k \in F} p_k Z_1^{k_1} \cdots Z_d^{k_d} : F \subset \mathbb{N}_0^d \text{ finite}, \, p_k \in \mathbb{C}_*\}.$$

denote the $\mathbb{C}$-algebra of d-variate polynomials and for $p = \sum_k p_k Z_1^{k_1} \cdot \ldots \cdot Z_d^{k_d} \in \Pi \setminus \{0\}$ let

$$\operatorname{maxdeg}(p) := \max\{\|k\|_\infty : p_k \neq 0\}.$$

The N-dimensional subvector space of d-variate polynomials of max-degree at most n is denoted by

$$\Pi_n := \{p \in \Pi \setminus \{0\} : \operatorname{maxdeg}(p) \leq n\} \cup \{0\} \cong \operatorname{span}\{\mathbb{C}^d \ni z \mapsto z^k : k \in I_n\},$$

and the *evaluation homomorphism* at $\Omega = \{z_1, \ldots, z_M\}$ will be denoted by

$$\mathcal{A}_n^\Omega \colon \Pi_n \to \mathbb{C}^M, \quad p \mapsto (p(z_1), \ldots, p(z_M)),$$

or simply by $\mathcal{A}_n$. Note that the representation matrix of $\mathcal{A}_n$ w.r.t. the canonical basis of $\mathbb{C}^M$ and the monomial basis of Π_n is given by the multivariate Vandermonde matrix

$$A_n = \left(z_j^k\right)_{\substack{j=1,\ldots,M \\ k \in I_n}} \in \mathbb{C}^{M \times N}.$$

The connection between the matrix T_n and polynomials that vanish on Ω lies in the observation that, using Definition 2.1, the matrix T_n admits the factorization

$$T_n = (f(k-\ell))_{k,\ell \in I_n} = P_n A_n^\top D_n A_n, \tag{2.1}$$

with $D_n = \operatorname{diag}(d)$, $d_j = z_j^{-n} \hat{f}_j$, $j = 1, \ldots, M$, and a permutation matrix $P_n \in \{0,1\}^{N \times N}$, see also [13] and its references for factorizations of this type. In particular, the kernel of A_n, corresponding to the polynomials in Π_n that vanish on Ω, is a subset of the kernel of T_n.

In order to deal with the multivariate polynomials encountered in this way we need some additional notation. The *zero locus* of a set $P \subset \Pi$ of polynomials is denoted by

$$V(P) := \{z \in \mathbb{C}^d : p(z) = 0 \text{ for all } p \in P\},$$

that is, $V(P)$ consists of the *common* roots of all the polynomials in P. For a set $\Omega \subset \mathbb{C}^d$,

$$I(\Omega) := \{p \in \Pi : p(z) = 0 \text{ for all } z \in \Omega\} = \bigcup_{n \in \mathbb{N}} \ker \mathcal{A}_n^\Omega$$

is the so-called *vanishing ideal* of Ω. Finally, for a set $P \subset \Pi$ of polynomials,

$$\langle P \rangle := \{\sum_{j=1}^{m} q_j p_j : m \in \mathbb{N},\, q_j \in \Pi,\, p_j \in P\}$$

is the *ideal generated by* P. Note that $V(P) = V(\langle P \rangle)$ always holds. Subsequently, we identify Π_n and $\mathbb{C}^N$ and switch back and forth between the matrix-vector and polynomial notation. In particular, we do not necessarily distinguish between $\mathcal{A}_n$ and its representation matrix A_n, so that e.g. "$V(\ker A_n)$" makes sense.

3. Main results

In the following two subsections, we study conditions on the degree n, and thereby on the number $(2n+1)^d$ of samples, such that the parameters z_j can be uniquely recovered and the polynomials used to identify them can be computed in a numerically stable way.

3.1. Complex parameters, polynomials, and unique solution

Our first result gives a simple but nonetheless sharp condition on the order of the moments such that the set of parameters Ω and the zero loci $V(\ker A_n)$ and $V(\ker T_n)$ are equal.

Theorem 3.1. *Let $f : \mathbb{Z}^d \to \mathbb{C}$ be an M-sparse d-variate exponential sum with parameters $z_j \in \mathbb{C}_*^d$, $j = 1, \ldots, M$. If $n \geq M$ then*

$$\Omega_f = V(\ker T_n(f)).$$

Moreover, if this equality holds for all M-sparse d-variate exponential sums f, then $n \geq M$.

Proof. Let $\Omega := \Omega_f = \{z_1, \ldots, z_M\}$. We start by proving $\Omega = V(\ker \mathcal{A}_n)$. Since $V(\ker \mathcal{A}_n) \supset V(\mathcal{A}_{n+1}) \supset \Omega$, it is sufficient to prove the case $n = M$. It is a simple fact that $I(\{z_j\}) = \langle Z_1 - z_{j,1}, \ldots, Z_d - z_{j,d} \rangle$, and that these ideals are pairwise comaximal, and hence we have

$$I(\Omega) = \prod_{j=1}^{M} I(\{z_j\}) = \langle \prod_{j=1}^{M} (Z_{\ell_j} - z_{j,\ell_j}) : \ell_j \in \{1, \ldots, d\} \rangle \subset \langle \ker \mathcal{A}_M \rangle \subset I(\Omega),$$

which implies $\langle \ker \mathcal{A}_M \rangle = I(\Omega)$. Thus we have $V(\ker \mathcal{A}_M) = V(\langle \ker \mathcal{A}_M \rangle) = V(I(\Omega)) = \Omega$ where the last equality holds because Ω is finite (and can easily be derived from the above).

Thus it remains to show that $\ker A_M = \ker T_M$. We proceed by proving $\operatorname{rank} A_M = M$. To simplify notation, we omit the subscript M on the matrices. Let $N := \dim \Pi_M$ and suppose that $A \in \mathbb{C}^{M\times N}$ has rank $r < M$. Let $\Omega' = \{z_1, \dots, z_r\}$ and w.l.o.g. let $A' \in \mathbb{C}^{r\times N}$, denoting the first r rows of A, be of rank r. Now the first part of the proof implies the contradiction $\Omega = V(\ker A) = V(\ker A') = \Omega'$.

Considering the factorization $T = PA^\top DA$ as in Equation (2.1) and applying Frobenius' rank inequality (see e.g. [17, 0.4.5 (e)]) yields

$$\operatorname{rank} A^\top D + \operatorname{rank} DA - \operatorname{rank} D \leq \operatorname{rank} A^\top DA = \operatorname{rank} T \leq \operatorname{rank} A$$

which implies $\operatorname{rank} T = \operatorname{rank} A = M$. The factorization clearly implies $\ker A \subset \ker T$ which together with the rank-nullity theorem $\dim\ker A = N - M = \dim\ker T$ yields the final result.

The converse follows from the fact that for $\Omega := \{(x_j, 1, \dots, 1) \in \mathbb{C}_*^d : j = 1, \dots, M\}$ with distinct $x_j \in \mathbb{C}_*$, any subset $B \subset \Pi_n$ such that $V(B) = \Omega$ (which holds, by assumption, for $B = \ker T_n$) necessarily contains a polynomial of (max-)degree at least M. ■

Example 3.2. *Let f be a 3-sparse 2-variate exponential sum with parameters $z_j \in \mathbb{C}_*^2$ and $\Omega = \{z_1, z_2, z_3\}$. The generating system of $I(\Omega)$ given in the proof of Theorem 3.1 is*

$$P := \left\{p_\ell : \ell \in \{1,2\}^3\right\}, \quad p_\ell(Z_1, Z_2) := \prod_{j=1}^{3} (Z_{\ell_j} - z_{j,\ell_j}).$$

We start by illustrating the generic case that no two coordinates are equal, i.e., $z_{j,\ell} \neq z_{i,\ell}$ if $j \neq i$ and $\ell = 1, 2$. The zero locus of each individual polynomial p_ℓ is illustrated in Figure 3.1, where each axis represents $\mathbb{C}$. The zero locus of each linear factor is a complex curve and illustrated by a single line. We note that the set P is redundant, i.e. the last three polynomials in the first row of Figure 3.1 are sufficient to recover the points uniquely as their common roots, but there is no obvious general rule which polynomials can be omitted.

Four other point configurations are shown in Figure 3.2. In the first three configurations coordinates of different points agree, which allows to remove some polynomials from P. In particular, the third point set which is collinear is generated already by $(Z_1 - z_{1,1})(Z_1 - z_{2,1})(Z_1 - z_{3,1})$ and $(Z_2 - z_{1,2})^3$. The fourth point set is generated either by the above set P of polynomials or by $\prod_{j=1}^{3}(Z_1 + Z_2 - z_{j,1} - z_{j,2})$ and $Z_1 - Z_2$, (which are not elements of P).

Remark 3.3. *Concerning the "natural" generator*

$$P = \left\{ \prod_{j=1}^{M} (Z_{\ell_j} - z_{j,\ell_j}) : \ell_j \in \{1, \dots, d\} \right\}$$

used in the proof above, we note that although the ideals $\langle P \rangle = \langle \ker \mathcal{A}_M \rangle$ coincide, the subvector space inclusion $\operatorname{span} P \subset \ker \mathcal{A}_M$ *is strict in general as*

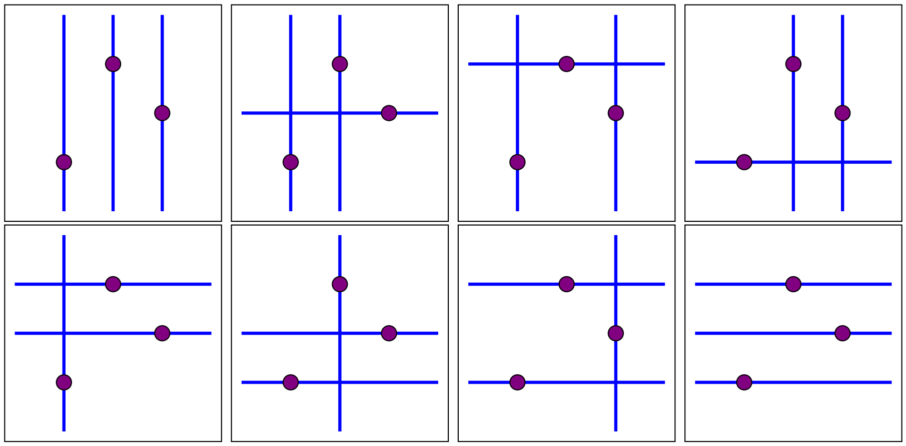


Figure 3.1: Zero sets of the polynomials in P, $d = 2$, $M = 3$, and for the case that no two coordinates are equal.

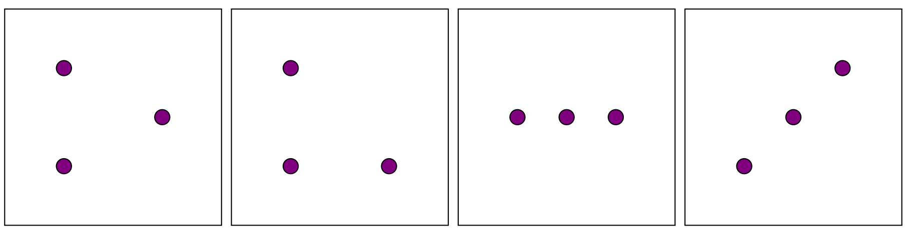


Figure 3.2: Point sets $\Omega \subset \mathbb{C}^2$, $d = 2$, $M = 3$.

can be seen for $d = 2$, $M = 2$, $z_1 = (0,0)$, $z_2 = (1,1)$ and the polynomial $Z_1 - Z_2 \in \ker \mathcal{A}_2$. Moreover, we have the cardinality $|P| = d^M$, at least for different coordinates $z_{j,\ell}$, and thus $|P| \gg M^d - M = \dim \ker \mathcal{A}_M$, i.e., the generator P contains many linear dependencies and is highly redundant for large M.

Finally, we would like to comment on the degree n and the total number of samples $(2n+1)^d$ with respect to the number of parameters M:

i) *A small degree $n \in \mathbb{N}$, $M < N < M + d$, and surjective $\mathcal{A}_n$ results in an uncountably infinite zero locus $V(\ker \mathcal{A}_n)$, since $\dim(\ker \mathcal{A}_n) \leq N - M < d$ and thus $I(\Omega)$ is generated by less than d polynomials.*

ii) *Increasing the degree results "generically" in a finite zero locus, cf. [1], but "generically" identifies spurious parameters since e.g. for $d = 2$ Bézout's theorem yields $|V(p,q)| \leq \deg(p)\deg(q)$ with equality in the projective setting (counting the roots with multiplicity), for coprime polynomials $p, q \in \ker \mathcal{A}_n$.*

Remark 3.4. *We discuss a slight modification of our approach. Instead of $I_n = \{0, \ldots, n\}^d = \{k \in \mathbb{N}_0^d : \|k\|_\infty \le n\}$ we take $J_n := \{k \in \mathbb{N}_0^d : \|k\|_1 \le n\}$ as index set and consider the matrix*

$$H_n(f) := (f(k+\ell))_{k,\ell \in J_n} \in \mathbb{C}^{\binom{n+d}{d} \times \binom{n+d}{d}}$$

instead of $T_n(f) = (f(k-\ell))_{k,\ell \in I_n}$. Theorem 3.1 also holds with T_n replaced by H_n with almost no change to the proof. In this way we need only $\binom{2n+d}{d}$ rather than $(2n+1)^d$ samples of f and also allow for arbitrary parameters $z_j \in \mathbb{C}^d$ instead of $z_j \in \mathbb{C}_*^d$. While T_n is a multilevel Toeplitz matrix, H_n is a submatrix of a multilevel Hankel matrix, and for the trigonometric setting discussed in the following subsection, it is more natural to consider the moments $f(k)$, $k \in \mathbb{Z}^d$, $\|k\|_\infty \le n$, than $f(k)$, $k \in \mathbb{N}_0^d$, $\|k\|_1 \le 2n$.

3.2. Parameters on the torus, trigonometric polynomials, and stable solution

We now restrict our attention to parameters $z_j \in \mathbb{T}^d$, hence $z_j = \mathrm{e}^{2\pi \mathrm{i} t_j}$ for a unique $t_j \in [0,1)^d$. In this case, $V(\ker \mathcal{A}_n)$ fulfills a 2^d-fold symmetry in the following sense. Let $p(z) = \sum_{k=0}^n \hat{p}_k z^k \in \ker \mathcal{A}_n$ and $z = (z_1, \ldots, z_d)^\top \in \mathbb{C}^d$ with $p(z) = 0$, then $z' = (\overline{z_1}^{-1}, z_2, \ldots, z_d)^\top$ is a root of the 1st-coordinate conjugate reciprocal polynomial

$$q(z) := \overline{\overline{z_1}^n p(\overline{z_1}^{-1}, z_2, \ldots, z_d)} = \sum_{k=0}^n \overline{\hat{p}_{n-k_1, k_2, \ldots, k_d}} z^k.$$

Since the roots $z \in \Omega \subset \mathbb{T}^d$ are self reciprocal $z' = z$, we have $q \in \ker \mathcal{A}_n$ and thus $z \in V(\ker \mathcal{A}_n)$ implies $z' \in V(\ker \mathcal{A}_n)$ for all choices of a conjugated reciprocal coordinate.

Moreover, we have the following construction of a so-called dual certificate [7, 6, 5, 4].

Theorem 3.5. *Let $d, n, M \in \mathbb{N}$, $n \ge M$, $t_j \in [0,1)^d$, $j = 1, \ldots, M$, $z_j := \mathrm{e}^{2\pi \mathrm{i} t_j}$, and $\Omega := \{z_j : j = 1, \ldots, M\}$ be given. Moreover, let $\hat{p}_\ell \in \mathbb{C}^N$, $\ell = 1, \ldots, N$, be an orthonormal basis with $\hat{p}_\ell \in \ker(T_n)^\perp$, $\ell = 1, \ldots, M$, and $p_\ell : \mathbb{C}_*^d \to \mathbb{C}$, $p_\ell(z) = \sum_{k \in I_n} \hat{p}_{\ell,k} z^k$, then $p : [0,1)^d \to \mathbb{C}$,*

$$p(t) = \frac{1}{N} \sum_{\ell=1}^M |p_\ell(\mathrm{e}^{2\pi \mathrm{i} t})|^2, \tag{3.1}$$

is a trigonometric polynomial of degree n and fulfills $0 \le p(t) \le 1$ for all $t \in [0,1)^d$ and $p(t) = 1$ if and only if $t = t_j$ for some $j = 1, \ldots, M$.

Proof. First note that every orthonormal basis $\hat{p}_\ell \in \mathbb{C}^N$, $\ell = 1, \ldots, N$, leads to

$$\sum_{\ell=1}^N |p_\ell(z)|^2 = \sum_{r,s=1}^N \overline{z^r} z^s \sum_{\ell=1}^N \overline{\hat{p}_{\ell,r}} \hat{p}_{\ell,s} = \sum_{r=1}^N |z^r|^2 = N$$

for $z \in \mathbb{T}^d$. Moreover, $\bar{z} = z^{-1}$ on $\mathbb{T}^d$ yields that p is indeed a trigonometric polynomial. Finally, Theorem 3.1 assures $\sum_{\ell=M+1}^N |p_\ell(z)|^2 = 0$ if and only if $z \in \Omega$. ∎

We proceed with an estimate on the condition number of the preconditioned matrix $T = T_n$.

Definition 3.6. *Let $M \in \mathbb{N}$ and $\Omega = \{e^{2\pi i t_j} : t_j \in [0,1)^d,\ j = 1,\ldots,M\}$, then*

$$\operatorname{sep}(\Omega) := \min_{r\in\mathbb{Z}^d,\ j\neq\ell} \|t_j - t_\ell + r\|_\infty$$

is the separation distance *of Ω. For $q > 0$, we say that Ω is q*-separated *if* $\operatorname{sep}(\Omega) > q$.

Theorem 3.7. *Let $d, n, M \in \mathbb{N}$, $t_j \in [0,1)^d$, $j = 1,\ldots,M$, $z_j := e^{2\pi i t_j}$, $q > 0$, and $\Omega := \{z_j : j = 1,\ldots,M\}$ be q-separated. Moreover, let $\hat{f}_j > 0$, then $n \geq 2dq^{-1}$ implies the condition number estimate*

$$\operatorname{cond}_2 WTW \leq \frac{(nq)^{d+1} + (2d)^{d+1}}{(nq)^{d+1} - (2d)^{d+1}} \cdot \frac{\max_j \hat{f}_j}{\min_j \hat{f}_j},$$

where the diagonal preconditioner $W = \operatorname{diag} w$, $w_k > 0$, $k \in I_n$, is well chosen. In particular, $\lim_{n\to\infty} \operatorname{cond}_2 WTW = \max_j \hat{f}_j / \min_j \hat{f}_j$.

Proof. First note, that the matrix T is hermitian positive semidefinite and define the condition number by $\operatorname{cond}_2 T := \|T\|_2 \|T^\dagger\|_2$, where $T^\dagger$ denotes the Moore-Penrose pseudoinverse. Let $D = \operatorname{diag} d$, $d_j = \hat{f}_j^{1/2}$, $j = 1,\ldots,M$, and $K = AW^2A^* \in \mathbb{C}^{M\times M}$, then we have

$$\operatorname{cond}_2 WTW = \operatorname{cond}_2 WA^*D^2AW = \operatorname{cond}_2 DAW^2A^*D = \frac{\max \hat{f}_j}{\min \hat{f}_j} \cdot \operatorname{cond}_2 K$$

and Corollary 4.7 in [21] yields the condition number estimate. The second claim follows since $\lim_{n\to\infty} \operatorname{cond}_2 K = 1$. ■

In summary, the condition

$$n \geq \max\{2dq^{-1}, M\} \tag{3.2}$$

allows for unique reconstruction of the parameters Ω and stability is guaranteed when computing the kernel polynomials from the given moments.

Remark 3.8. *Up to the constant $2d$, the condition $n > 2d/q$ in the assumption of Theorem 3.7 is optimal in the sense that equidistant nodes $t_j = j/m$, $j \in I_m$, $n < q^{-1} = m$, imply $A \in \mathbb{C}^{m^d\times n^d}$ and $\operatorname{rank} A = n^d < m^d = M$. Recently, the equivalent discrete variant of Ingham's inequality [20] has found sharp improvements in [30, Lemma 2.1], [22, Appendix A], and [24, Sect. 2.3] involving no preconditioning weights.*

Moreover, Theorem 3.1 asserts that the condition on the degree n with respect to the number ob parameters M is close to optimal in the specific setting. We briefly comment on the following typical scenarios for the point set Ω and the relation (3.2):

i) *quasi-uniform parameters* $t_j \in [0,1)^d$ *which fulfil* $\operatorname{sep}(\Omega) \approx C_d M^{-1/d}$ *imply* $\max\{2dq^{-1}, M\} = M$,
ii) *equidistant co-linear parameters, e.g.* $t_j = M^{-1}(j,\ldots,j)^\top$, *imply* $\operatorname{sep}(\Omega) \approx CM^{-1}$, *i.e., both terms are of similar size, and*
iii) *parameters* $t_j \in [0,1)^d$ *chosen at random from the uniform distribution, imply* $\mathbb{E}\operatorname{sep}(\Omega) = C'_d M^{-2}$, *cf. [32], and thus* $\max\{2dq^{-1}, M\} = C''_d M^2$.

Dropping the condition $n > M$ in (3.2) and restricting to the torus, we still get the following result on how much the roots of the polynomials in the kernel of T can deviate from the original set Ω.

Theorem 3.9. *Let* $d, n, M \in \mathbb{N}$, $t_j \in [0,1)^d$, $j = 1,\ldots,M$, $z_j := \mathrm{e}^{2\pi \mathrm{i} t_j}$, $q > 0$, *and* $\Omega := \{z_j : j = 1,\ldots,M\}$ *be q-separated, then* $n \geq 2dq^{-1}$ *implies*

$$\Omega \subset V(\ker T) \cap \mathbb{T}^d \subset \{z\mathrm{e}^{2\pi \mathrm{i} t} : z \in \Omega,\ \|t\|_\infty < 2d/n\}.$$

Proof. We prove the assertion by contradiction. Let $y \in V(\ker T) \cap \mathbb{T}^d$ be $2d/n$-separated from the point set $\Omega \subset \mathbb{T}^d$ and let $\hat{p}_\ell \in \mathbb{C}^N$, $\ell = 1,\ldots,N-M$, constitute a basis of $\ker T$. By definition, we have $p_\ell(y) = 0$ and thus the augmented Fourier Matrix

$$A_y := \begin{pmatrix} A \\ e_y \end{pmatrix}, \quad e_y := (\mathrm{e}^{2\pi \mathrm{i} k y})^\top_{k \in I_n},$$

fulfills $A_y \hat{p}_\ell = 0$, i.e., $\dim\ker A_y \geq N - M$. On the other hand, Corollary 4.7 in [21] implies $\operatorname{rank} A_y = M + 1$ and thus the contradiction $N = \dim\ker A_y + \operatorname{rank} A_y \geq N + 1$. ∎

3.3. Prototypical algorithm

Let f be an M-sparse d-variate exponential sum with pairwise distinct parameters $z_j \in \mathbb{C}_*^d$ and $n \geq M$ be an upper bound. Theorem 3.1 justifies the following prototypical formulation of the multivariate Prony method.

Algorithm 1 Multivariate Prony method.

Input: $d, n \in \mathbb{N}$,
$f(k)$, $k \in \{-n,\ldots,n\}^d$
Set up $T_n = (f(k-\ell))_{k,\ell \in I_n} \in \mathbb{C}^{N \times N}$
Compute $\ker T_n$
Compute $V(\ker T_n)$
Output: $V(\ker T_n) = \{z_1,\ldots,z_M\}$

The third step, i.e., the computation of the zero locus $V(\ker T_n)$, is beyond the scope of this paper and several methods can be found elsewhere, see e.g. [2, 25, 36, 37]. We further note that the number $(2n+1)^d$ of used samples scales as $\mathcal{O}(M^d)$ and that standard algorithms for computing the kernel of the matrix T_n have cubic complexity in n^d.

4. Other approaches

There are many variants of the one dimensional moment problem from Section 1, originating from such diverse fields as for example signal processing, electro engineering, and quantum chemistry, with as widespread applications as spectroscopy, radar imaging, or super-resolved optical microscopy, see e.g. the survey paper [27]. Variants of Prony's method with an increased stability or a direct computation of the parameters without the detour via polynomial coefficients include for example MUSIC [35], ESPRIT [33], the Matrix-Pencil method [18], the Approximate Prony method [29], the Annihilating Filter method [38], and methods relying on orthogonal polynomials [14].

Multivariate generalizations of these methods have been considered in [19, 1] and similar to our approach, the latter realizes the parameters as common roots of multivariate polynomials. However, both of these papers have an emphasis on the generic situation where e.g. the zero locus of two bivariate polynomials is finite. In this case, the total number of used moments for reconstruction might indeed scale as the number of parameters but no guarantee is given for a specific instance of the moment problem. Other multivariate generalizations [30, 28] decompose the multivariate moment problem into a series of univariate moment problems via projections of the measure. While again this approach typically works well, the necessary number of a-priori chosen projections for a signed measure scales as the number of parameters in the bivariate case [11]. We note that the subset

$$P_0 := \{\prod_{j=1}^{M} (Z_\ell - z_{j,\ell}) : \ell = 1, \dots, d\} \subset I(\Omega),$$

of the set of generators in the proof of Theorem 3.1 are exactly the univariate polynomials when projecting onto the d coordinate axes, see also the first and last zero locus in Figure 3.1.

A different approach to the moment problem from Section 1 has been considered in [7, 6, 5, 4] and termed 'super-resolution'. From a signal processing perspective, knowing the first moments is equivalent to sampling a low-pass version of the measure and restoring the high frequency information from these samples. With the notation of Remark 2.2 the measure τ with parameters $t_j \in [0,1)^d$ is the unique minimizer of

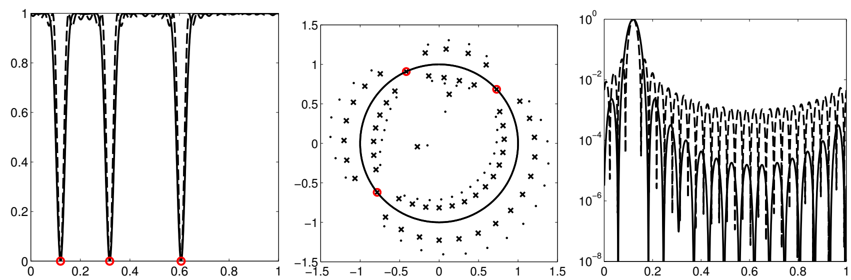
$$\min \|\nu\|_{\mathrm{TV}} \quad \text{s.t.} \quad \int_{[0,1)^d} \mathrm{e}^{2\pi \mathrm{i} k t} \mathrm{d}\nu(t) = f(k),\ k \in I_n,$$

provided the parameters fulfill a separation condition as in Section 3.2. This is proven via the existence of a so-called dual certificate [7, Appendix A] and becomes computationally attractive by recasting this dual problem as a semidefinite program. The program has just $(n+1)^2/2$ variables in the univariate case [7, Corollary 4.1], but at least we do not know an explicit bound on the number of variables in the multivariate case, see [12, Remark 4.17, Theorem 4.24, and Remark 4.26].

Finally note that there is a large body of literature on the related topic of reconstructing a multivariate sparse trigonometric polynomials from samples, see e.g. [3, 23, 15, 8, 31, 34, 16]. Translated to the situation at hand, all these methods heavily rely on the fact that the parameters $t_j \in [0,1)^d$ are located on a Cartesian grid with mesh sizes $1/m_1, \ldots, 1/m_d$ for some $m_1, \ldots, m_d \in \mathbb{N}$ and deteriorate if this condition fails [9]. Hence, these methods lack one major advantage of Prony's method, namely that the parameters $t_j \in [0,1)^d$ can, in principle, be reconstructed with infinite precision.

5. Numerical results

All numerical experiments are realized in MATLAB 2014a on an Intel i7, 12GByte, 2.1GHz, Ubuntu 14.04.



(a) Sum of squared absolute values of kernel polynomials on $\mathbb{T}$ (identified with $[0,1)$).

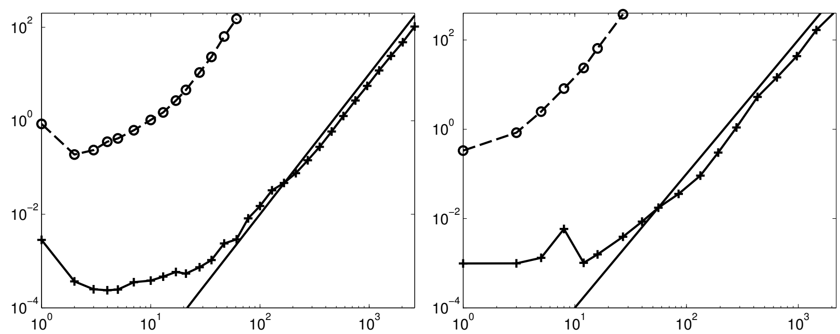
(b) Zero set of sum of squared absolute values of kernel polynomials on $\mathbb{C}$.

(c) Squared absolute values of the first polynomial orthogonal to the kernel.

Figure 5.1: Parameters $d=1$, $M=3$, $n=30$, $t_1=0.12$, $t_2=1/\pi$, and $t_3=\mathrm{e}^{-1/2}$. Dashed lines and $\times$ indicate no weighting, solid lines and $\cdot$ indicate triangular weights $w_k = \min\{k+1, n-k\}$, $k=0,\ldots,n-1$.

Example 5.1 ($d=1$)**.** *We consider the case $d=1$ with parameters on the 1-torus $\mathbb{T}$ that we identify with the interval $[0,1)$. For a 3-sparse exponential sum some of the associated (trigonometric) polynomials are visualized in Figure 5.1, where we start with the upper bound $n=30\geq 3$ and also indicate the effects of a preconditioner W according to Theorem 3.7 on the roots of the polynomials.*

The method introduced in [7] finds a polynomial of the form (3.1) *as a solution to a convex optimization problem, whereas we find such a polynomial with Prony's method. For this comparison we used the MATLAB code provided in [7] and modified it so that it runs for different problem sizes depending on the sparsity $M=1,\ldots,100$. This means that we used roughly $5M$ samples and random parameters $t_j \in [0,1)$, $j=1,\ldots,M$, satisfying the separation condition in [7]. We only measured the time for finding a polynomial of the form* (3.1)*, since the calculation of the roots is basically the same in both algorithms. In Figure 5.2 (a), where the times needed with* `cvx` *are depicted as circles and the times needed by Prony's method (computation of the kernel* $\ker T$ *and the dual* *certificate p, cf. Theorem 3.5) are depicted as crosses, we see that the solution via convex optimization takes considerably more time. Note that the end criterion of the convex optimization program is set to roughly* 10^{-6}*, therefore the solution accuracy does not increase beyond this point, whereas for Prony's method the solutions in this test are all in the order of machine accuracy,* 10^{-15}*.*



(a) Running times, $d = 1$. (b) Running times, $d = 2$.

Figure 5.2: Running time in seconds with respect to the number of parameters M for extracting a polynomial of form (3.1), convex optimization via `cvx` ($\circ$) and Prony's method (+).

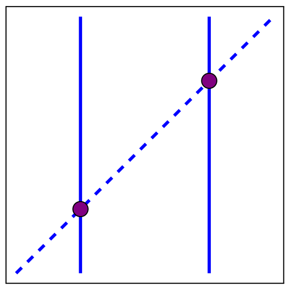
Example 5.2 ($d = 2$)**.** *We demonstrate our method to reconstruct the parameters from the moments* $f\colon \mathbb{Z}^2 \to \mathbb{C}$, $k \mapsto (1,1)^k + (-1,-1)^k$*. For moments of order* $|k| \leq n = 2$ *and the associated space of polynomials* Π_2 *with reverse lexicographical order on the terms, we get the* 9×9 *block Toeplitz matrix* $T = T_2 = (f(k-\ell))_{k,\ell \in I_2}$ *with the Toeplitz blocks* T', T'' *as follows:*

$$T = \begin{pmatrix} T' & T'' & T' \\ T'' & T' & T'' \\ T' & T'' & T' \end{pmatrix}, \quad T' = \begin{pmatrix} 2 & 0 & 2 \\ 0 & 2 & 0 \\ 2 & 0 & 2 \end{pmatrix}, \quad T'' = \begin{pmatrix} 0 & 2 & 0 \\ 2 & 0 & 2 \\ 0 & 2 & 0 \end{pmatrix}.$$

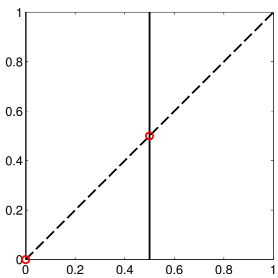
A vector space basis of $\ker T$ *is given by the polynomials*

$$p_1 = -1 + Z_1^2, \qquad p_2 = -Z_1 + Z_2, \qquad p_3 = -1 + Z_1 Z_2,$$
$$p_4 = -Z_1 + Z_1^2 Z_2, \qquad p_5 = -1 + Z_2^2, \qquad p_6 = -Z_1 + Z_1 Z_2^2,$$
$$p_7 = -1 + Z_1^2 Z_2^2.$$

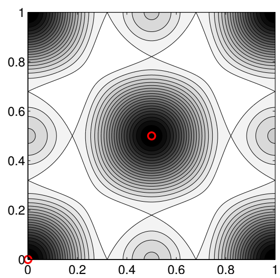
Since $p_3 = p_1 + Z_1 p_2$, $p_4 = Z_1 p_3$, $p_5 = Z_2 p_2 + p_3$, $p_6 = Z_1 p_5$, *and* $p_7 = (1 + Z_1 Z_2) p_3$*, we have* $\langle \ker T \rangle = \langle p_1, p_2 \rangle$ *and hence* $V(\ker T) = V(p_1, p_2) = \{(1,1), (-1,-1)\}$*. The zero loci of* p_1, p_2 *are depicted in Figure 5.3 (a) (in the style of Figure 3.1) resp. (b), where the torus* $\mathbb{T}^2$ *is identified with* $[0,1)^2$*. Note that we would typically expect the intersection of the zero locus of each polynomial with the torus to be finite, which is the case neither for* p_1 *nor* p_2*. In Figure 5.3 (c) the sum of the squared absolute values of an orthonormal basis of* $\ker T$ *is drawn.*



(a) See also Fig. 3.1, zero loci $V(p_1), V(p_2) \subset \mathbb{C}^2$.

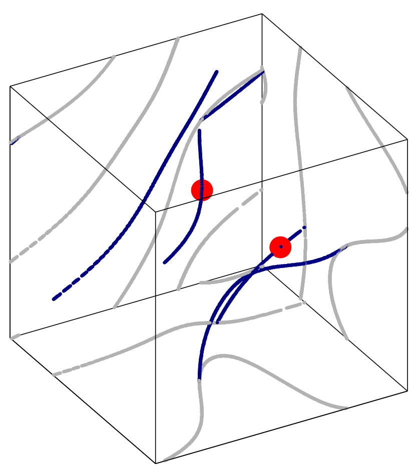


(b) Zeros of the kernel polynomial p_1 (lines) and p_2 (dashed) on $\mathbb{T}^2$.

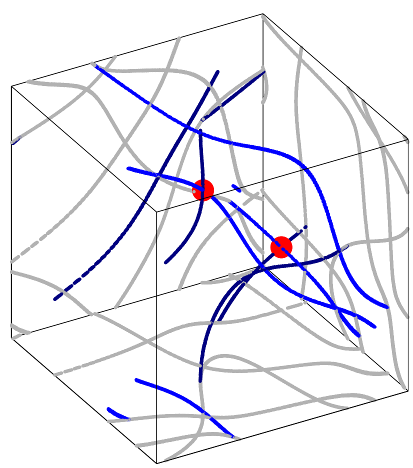


(c) Sum of squared absolute values of kernel polynomials.

Figure 5.3: Parameters $d = 2$, $M = 2$, $n = 2$, $t_1 = (0.0, 0.0)$ and $t_2 = (0.5, 0.5)$.



(a) Zeros of the kernel polynonial p_1 on $\mathbb{T}^3$.



(b) Zeros of the kernel polynomials p_1, p_2 on $\mathbb{T}^3$.

Figure 5.4: Parameters $d = 3$, $M = 2$, $n = 1$, $t_1 = (0.1, 0.3, 0.25)$ and $t_2 = (0.7, 0.8, 0.9)$.

Example 5.3 ($d = 3$). *Figure 5.4 depicts the intersection of* $\mathbb{T}^3$ *(identified with* $[0, 1)^3$*) and the zero loci of two polynomials that arise with the Prony method for* $M = 2$ *parameters choosing* $n = 1$ *(which is not an upper bound for* M*). This illustrates that, in the case* $d = 3$*, the zero locus of a single polynomial intersected with the torus can typically be visualized as a "one-dimensional" curve as suggested by the heuristic argument that a complex polynomial can be thought of as two real equations, which together with the three real equations that define* $\mathbb{T}^3$ *as a subset of* $\mathbb{C}^3 = \mathbb{R}^6$ *provides five equations, thus leaving one real degree of freedom.*

6. Summary

We suggested a multivariate generalization of Prony's method and gave sharp conditions under which the problem admits a unique solution. Moreover, we provided a tight estimate on the condition number for computing the kernel of the involved Toeplitz matrix of moments. Numerical examples were presented for spatial dimensions $d = 1, 2, 3$ and showed in particular that a so-called dual certificate in the semidefinite formulation of the moment problem can be computed much faster by solving an eigenvalue problem.

Beyond the scope of this paper, future research needs to address the actual computation of the common roots of the kernel polynomials, the stable reconstruction from noisy moments, and reductions both in the number of used moments as well as in computation time.

Acknowledgment. The authors thank S. Heider for the implementation of the approach [7] for the bivariate case and H. M. Möller for several enlightening discussions. The fourth author expresses his thanks to J. Abbott for warm hospitality during his visit in Genoa and numerous useful suggestions. Moreover, we gratefully acknowledge support by the DFG within the research training group 1916: Combinatorial structures in geometry and by the Helmholtz Association within the young investigator group VH-NG-526: Fast algorithms for biomedical imaging.

References

[1] F. Andersson, M. Carlsson, and M. V. de Hoop. Nonlinear approximation of functions in two dimensions by sums of exponential functions. *Appl. Comput. Harmon. Anal.*, 29:156–181, 2010.

[2] D. J. Bates, J. D. Hauenstein, A. J. Sommese, and C. W. Wampler. Bertini: Software for numerical algebraic geometry. Available at `bertini.nd.edu` with permanent doi: `dx.doi.org/10.7274/R0H41PB5`.

[3] M. Ben-Or and P. Tiwari. A deterministic algorithm for sparse multivariate polynomial interpolation. In *Proceedings of the twentieth annual ACM symposium on Theory of computing*, pages 301–309, 1988.

[4] T. Bendory, S. Dekel, and A. Feuer. Exact Recovery of Dirac Ensembles from the Projection Onto Spaces of Spherical Harmonics. *Constr. Approx.*, 42(2):183–207, 2015.

[5] T. Bendory, S. Dekel, and A. Feuer. Super-resolution on the sphere using convex optimization. *IEEE Trans. Signal Process.*, 64:2253–2262, 2015.

[6] E. J. Candès and C. Fernandez-Granda. Super-resolution from noisy data. *J. Fourier Anal. Appl.*, 19(6):1229–1254, 2013.

[7] E. J. Candès and C. Fernandez-Granda. Towards a mathematical theory of super-resolution. *Comm. Pure Appl. Math.*, 67(6):906–956, 2014.

[8] E. J. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inform. Theory*, 52:489–509, 2006.

[9] Y. Chi, L. L. Scharf, A. Pezeshki, and A. R. Calderbank. Sensitivity to basis mismatch in compressed sensing. *IEEE Trans. Signal Process.*, 59:2182–2195, 2011.

[10] B. G. R. de Prony. Essai éxperimental et analytique: sur les lois de la dilatabilité de fluides élastique et sur celles de la force expansive de la vapeur de l'alkool, a différentes températures. *Journal de l'école polytechnique*, 1(22):24–76, 1795.

[11] B. Diederichs and A. Iske. Parameter estimation for bivariate exponential sums. In *Sampling Theory and Applications (SampTA), 2015 International Conference on*, pages 493–497, 2015.

[12] B. Dumitrescu. *Positive trigonometric polynomials and signal processing applications*. Signals and Communication Technology. Springer, Dordrecht, 2007.

[13] R. L. Ellis and D. C. Lay. Factorization of finite rank Hankel and Toeplitz matrices. *Linear Algebra Appl.*, 173:19–38, 1992.

[14] F. Filbir, H. N. Mhaskar, and J. Prestin. On the problem of parameter estimation in exponential sums. *Constr. Approx.*, 35:323–343, 2012.

[15] M. Giesbrecht, G. Labahn, and W.-s. Lee. On the equivalence between Prony's and Ben-Or's/Tiwari's methods. *University of Waterloo Tech Report*, 23, 2002.

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

[17] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, New York, USA, 2nd edition, 2013.

[18] Y. Hua and T. K. Sarkar. Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise. *IEEE Trans. Acoust. Speech Signal Process.*, 38(5):814–824, 1990.

[19] T. Jiang, N. D. Sidiropoulos, and J. M. F. ten Berge. Almost-sure identifiability of multidimensional harmonic retrieval. *IEEE Trans. Signal Process.*, 49(9):1849–1859, 2001.

[20] V. Komornik and P. Loreti. *Fourier Series in Control Theory.* Springer-Verlag, New York, 2004.

[21] S. Kunis and D. Potts. Stability results for scattered data interpolation by trigonometric polynomials. *SIAM J. Sci. Comput.*, 29:1403–1419, 2007.

[22] W. Liao. MUSIC for multidimensional spectral estimation: stability and super-resolution. *IEEE Trans. Signal Process.*, to appear.

[23] Y. Mansour. Randomized interpolation and approximation of sparse polynomials. *SIAM J. Comput.*, 24:357–368, 1995.

[24] A. Moitra. Super-resolution, extremal functions and the condition number of Vandermonde matrices. In *Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing*, STOC '15, pages 821–830, New York, NY, USA, 2015. ACM.

[25] H. M. Möller and T. Sauer. H–bases for polynomial interpolation and system solving. *Adv. in Comp. Math.*, 12:335–362, 2000.

[26] T. Peter, G. Plonka, and R. Schaback. Reconstruction of multivariate signals via Prony's method. *Proc. Appl. Math. Mech.*, to appear.

[27] G. Plonka and M. Tasche. Prony methods for recovery of structured functions. *GAMM-Mitt.*, 37(2):239–258, 2014.

[28] G. Plonka and M. Wischerhoff. How many Fourier samples are needed for real function reconstruction? *J. Appl. Math. Comput.*, 42(1-2):117–137, 2013.

[29] D. Potts and M. Tasche. Parameter estimation for exponential sums by approximate Prony method. *Signal Processing*, 90(5):1631–1642, 2010.

[30] D. Potts and M. Tasche. Parameter estimation for multivariate exponential sums. *Electron. Trans. Numer. Anal.*, 40:204–224, 2013.

[31] H. Rauhut. Random sampling of sparse trigonometric polynomials. *Appl. Comput. Harmon. Anal.*, 22:16–42, 2007.

[32] M. Reitzner, M. Schulte, and C. Thäle. Limit theory for the Gilbert graph. *Preprint*, 2015.

[33] R. Roy and T. Kailath. ESPRIT—estimation of signal parameters via rotational invariance techniques. In *Signal Processing, Part II*, volume 23 of *IMA Vol. Math. Appl.*, pages 369–411. Springer, New York, 1990.

[34] M. Rudelson and R. Vershynin. On sparse reconstruction from Fourier and Gaussian measurements. *Comm. Pure Appl. Math.*, 61:1025–1045, 2008.

[35] R. O. Schmidt. Multiple emitter location and signal parameter estimation. *Antennas and Propagation, IEEE Transactions on*, 34(3):276–280, 1986.

[36] L. Sorber, M. Van Barel, and L. De Lathauwer. Numerical solution of bivariate and polyanalytic polynomial systems. *SIAM J. Numer. Anal.*, 52(4):1551–1572, 2014.

[37] H. J. Stetter. *Numerical Polynomial Algebra*. Society for Industrial and Applied Mathematics, Philadelphia, 2004.

[38] M. Vetterli, P. Marziliano, and T. Blu. Sampling signals with finite rate of innovation. *IEEE Trans. Signal Process.*, 50(6):1417–1428, 2002.