Sparse polynomial interpolation: compressed sensing, super resolution, or Prony?

Cédric Josz, Jean-Bernard Lasserre, Bernard Mourrain

Noname manuscript No.
(will be inserted by the editor)

Sparse polynomial interpolation: compressed sensing, super resolution, or Prony?

Cédric Josz · Jean Bernard Lasserre · Bernard Mourrain

Abstract We show that the sparse polynomial interpolation problem reduces to a discrete super-resolution problem on the n-dimensional torus. Therefore the semidefinite programming approach initiated by Candès & Fernandez-Granda [7] in the univariate case (and later extended to the multivariate setting) can be applied. In particular, exact recovery is guaranteed provided that a geometric spacing condition on the "supports" holds and the number of evaluations are sufficiently many (but not many). It also turns out that the (compressed sensing) LP-formulation of ℓ_1-norm minimization is also guaranteed to provide exact recovery *provided that* the evaluations are made in a certain manner and even though the Restricted Isometry Property for exact recovery is not satisfied. (A naive compressed sensing LP-approach does not offer such a guarantee.) Finally we also describe the algebraic Prony method for sparse interpolation, which also recovers the exact decomposition but from less point evaluations and with no geometric spacing condition. We provide two sets of numerical experiments, one in which the super-resolution technique and Prony's method seem to cope equally well with noise, and another in which the super-resolution technique seems to cope with noise better than Prony's method, at the cost of an extra computational burden (i.e. a semidefinite optimization).

Keywords Linear programming · Prony's method · Semidefinite programming · super-resolution

1 Introduction

Suppose that we are given a *black-box* polynomial $g \in \mathbb{R}[\mathbf{z}]$, that is, g is unknown but given any "input" point $\mathbf{z} \in \mathbb{C}^n$, the black-box outputs the complex number $g(\mathbf{z})$.

A *sparse* polynomial is a polynomial $\mathbf{z} \mapsto g(\mathbf{z}) = \sum_\alpha g_\alpha \mathbf{z}^\alpha$ with only a few non-zero coefficients (g_α) and *sparse interpolation* is concerned with recovering the unknown monomials $(\mathbf{z}^\alpha)$ and coefficients (g_α) of a sparse polynomial via the sole knowledge of a few (and as few as possible) values of g at some points $(\mathbf{z}_k) \subset \mathbb{C}^n$ that one may choose at our convenience. Basically there are two possible methods to recover the coefficients of the unknown polynomial:

Cédric Josz
LAAS-CNRS, 7 avenue du Colonel Roche, BP 54200
31031 Toulouse Cédex 4, France
E-mail: cedric.josz@gmail.com

Jean Bernard Lasserre
LAAS-CNRS and Institute of Mathematics
7 avenue du Colonel Roche, BP 54200
31031 Toulouse Cédex 4, France E-mail: lasserre@laas.fr

Bernard Mourrain
INRIA, 2004 route des Lucioles
06902 Sophia Antipolis, France
E-mail: Bernard.mourrain@inria.fr

Naive LP

A naive "compressed sensing" LP-approach consists of solving $\min\{\|\mathbf{x}\|_1 : \mathbf{A}\mathbf{x} = b\}$ where $\mathbf{x}$ is the vector of coefficients of the unknown polynomial and $\mathbf{A}\mathbf{x} = b$ are linear constraints obtained from evaluations at some given points. In minimizing the ℓ_1 norm one expects to obtain a "sparse" solution to the undetermined system $\mathbf{A}\mathbf{x} = b$ (as in compressed sensing). However since the matrix $\mathbf{A}$ does not satisfy the sufficient *Restricted Isometry Property* (RIP), exact recovery is not guaranteed (at least by invoking results from compressed sensing).

Prony

The method goes back to the pioneer work of G. R de Prony [11] who was interested in recovering a sum of few exponential terms from sampled values of the function. Thus Prony's method is also a standard tool to recover a complex atomic measure from knowledge of some of its moments [24]. Briefly, in the univariate setting this purely algebraic method consists of two steps: (i) Computing the coefficients of a polynomial p whose roots form the finite support of the unknown measure. As p satisfies a recurrence relation it is the unique element (up to scaling) in the kernel of a (Hankel) matrix. (ii) The weights associated to the atoms of the support solve a Vandermonde system.

This algebraic method has then been used in the context of sparse polynomial interpolation. In the univariate case it consists in evaluating the black-box polynomial at values of the form φ^k for a finite number of pairs $(k, \varphi) \in \mathbb{N} \times \mathbb{C}$, fixed. A sequence of $2\,r$ evaluations allows to recover the decomposition exactly, where r is the number of terms of the sparse polynomial. The decomposition is obtained by computing a minimal recurrence relation between these evaluations, by finding the roots of the associate polynomial, which yields the exponents of the monomials and by solving a Vandermonde system which yields the coefficients of the terms in the sparse polynomial.

Since then, it has been extended to address numerical issues and to treat applications in various domains (particularly in signal processing). See e.g. [32], [34], [17], [4], [31], [30] and the many references therein. For instance, from an algorithmic point of view the approach has been improved by exploiting the Berlekamp-Massey algorithm [3], [26] and the structure of the involved matrices; see e.g. [21], [36], [18]. It has been applied to treat multivariate sparse interpolation problems [35] and [2]. It has also been extended to approximate data [16]. In particular it has been applied to sparse polynomial interpolation with noisy data in [38] to provide a way to recover a blackbox univariate polynomial exactly when some (but not all) of its evaluations are corrupted with noise (in the sipirit of error-decoding).

Generalizations of Prony's method to multivariate reconstruction problems have been developed more recently. Instead of computing a univariate polynomial from the moments and finding its roots, the algebraic structure of an Artinian Gorenstein algebra is computed and the decomposition is deduced from eigenvectors of multiplication operators. See e.g. [24] for a projection based method, [33] for methods which require the computation of an H-basis of the kernel ideal or more direct decomposition methods in [27] and [19] which we are going to use hereafter.

Super-resolution

The *super-resolution* problem consists in recovering the support of a sparse atomic (signed) measure on a compact set K, from known moments. Hereafter we will consider the particular case where K is the multi-dimensional torus $\mathbb{T}^n \subset \mathbb{C}^n$. In the work of Candès and Fernandez-Granda [7] it is shown that if the support atoms of the measure are well-separated then the measure is the unique solution of an infinite-dimensional convex optimization problem on a space of measures with the total variation as minimization criterion. In the univariate case its (truncated) moment matrix can be recovered by solving a single *Semidefinite Program* (SDP). The number of evaluations needed for exact recovery is then at most $4s$ if s is the number of atoms[1] (and in fact significantly less in all numerical examples provided). Interestingly, the total-variation minimization technique adapts nicely to noisy model and yields stable approximations of the weighted sum of Dirac measures, provided that the support atoms are well separated, see e.g. [6], [1], [14].

[1] As noted in Candès and Fernandez-Granda [7], with the proviso that the number of evaluations is larger than 128

An extension to the multivariate case has been proposed in [13] to recover weight sums of Dirac measures in $\mathbb{R}^n$, where now one needs to solve a *hierarchy* of semidefinite programs.

The existence and unicity of the solution of the total variation minimization problem relies on the existence of a dual certificate, that is, a polynomial with sup-norm reached at the points of the support of the measure. The relaxation into a hierarchy of semidefinite programs [13] yields a decomposition into a finite weighted sum of Dirac measures, provided that at some order of the hierarchy, a flat extension condition is satisfied at an optimal solution. Then the decomposition can be recovered from the moment matrix at this optimal solution by applying a Prony's like technique.

Contribution

• We first consider the naive LP-approach to sparse interpolation via ℓ_1-norm minimization $\min\{\|\mathbf{x}\|_1 : \mathbf{A}\mathbf{x} = b\}$ in the spirit of *compressed sensing* and we characterize optimal solutions via standard arguments of linear programming (LP). Interestingly, this characterization is a "formal analogue" in appropriate spaces of that in super-resolution (2.1) in some measure spaces. However as the matrix $\mathbf{A}$ does not satisfy the RIP there is no guarantee (at least by invoking results from compressed sensing) that an optimal solution is unique and corresponds to the unique sparse black-box polynomial g.

• We propose a variant of (multivariate) Prony's method in the context of sparse polynomial interpolation. In particular in the univariate case it only requires $r+1$ evaluations (instead of $2r$) where r is the number of monomials of the blackbox polynomial. It involves a Toeplitz matrix rather than a Hankel matrix and numerical experiments confirm the theoretical result regarding the number of evaluations.

• We then propose another approach which uses the fact that one has the choice of points $(\mathbf{z}_k) \subset \mathbb{C}^n$ at which evaluations of g can be done through the black-box and yields the following simple but crucial observation: By choosing $\mathbf{z}_k$ as some power $\mathbf{z}_0^\beta$ with $\beta \in \mathbb{N}^n$ (and where $\mathbf{z}_0 \in \mathbb{T}^n$ is fixed, arbitrary) the sparse polynomial p can be viewed as a signed atomic measure μ on $\mathbb{T}^n$ with finitely many atoms $(\mathbf{z}_0^\alpha) \subset \mathbb{T}^n$ associated with the nonzero coefficients of p (the signed weights of μ associated with each atom). In doing so the sparse interpolation problem is completely equivalent to a super-resolution problem on the multi-dimensional torus $\mathbb{T}^n$ and therefore the sparse polynomial is the unique optimal solution of a certain infinite-dimensional linear program on a space of measures, provided that a geometric condition of minimum spacing (between atoms of the support) is satisfied and sufficiently many evaluations are available. Notice that previous works on Prony's method (e.g. [38]) have also exploited (but in a different manner) evaluations at consecutive powers of a fixed element. In fact our view of a polynomial as a signed atomic measure on the torus is probably the shortest way to explain why Prony's method can be used for polynomial interpolation (as the original Prony's method can be interpreted directly as reconstructing an atomic measure on the complex plane from some of its moments [24]). To the best of our knowledge, this interpretation of a polynomial as a signed measure seems to be new.

We then relax this problem to a hierarchy of semidefinite programs. In principle, the convergence is only asymptotic (and guaranteed to be finite only in the cases $n = 1, 2$). However generic finite convergence results of polynomial optimization of Nie [29] seem to be also valid in our context as evidenced from our numerical experiments (and in those in De Castro et al. [13] in signal processing). The flat extension rank condition on moment matrices of Curto and Fialkow [10, Theorem 1.1] or its generalization in [25] can be extended to Toeplitz-like moment matrices [20], to test whether finite convergence takes place. In all our numerical experiments, finite convergence takes place and the coefficients and exponents of the unknown polynomial could be extracted. To give an idea, a univariate polynomial of degree 100 with 3 atoms can be recovered by solving a single SDP with associated 4×4 Toeplitz matrices and which only involves 4 evaluations. On the other hand if some atoms are close to each other then more information (i.e. evaluations) is needed as predicted by the spacing condition (and confirmed in some numerical experiments).

• In practice we reduce the number of measurements (i.e., evaluations) needed to retrieve a sparse polynomial when using super resolution. To do this we invoke a result (Lemma 1 of this paper) related to the full complex moment problem. It states that atomic measures on $\mathbb{C}^n$ with finitely many atoms are completely characterized by their moments $(\int \mathbf{z}^\alpha \, d\mu)_{\alpha \in \mathbb{N}^n}$, with no need of all moments $(\int \bar{\mathbf{z}}^\beta \mathbf{z}^\alpha \, d\mu)_{\alpha,\beta}$ involving conjugates. This result, which holds true in full generality, yields a simplified hierarchy

with significant computational savings. It is the subject of future work to determine whether this preserves the guarantee of "asymptotic" recovery of the original complete hierarchy; in our numerical experiments, finite convergence is always observed, and with fewer measurements than in the original method.

• A rigorous LP-approach. In fact the interpolation problem is even a *discrete* super-resolution problem (i.e. recovery of a discrete signal) where the atomic measure consists of finitely many atoms on a fixed grid $\{t/N\}_{t=0,\dots,N-1}$ as described in Candès and Fernandez-Granda [7, §1.4]. Therefore this fact also validates exact recovery via a (compressed sensing) LP-formulation of ℓ_1-minimization $\min_{\mathbf{x}}\{\|\mathbf{x}\|_1 : \mathbf{A}\mathbf{x} = b\}$ *provided* that the spacing condition is satisfied and evaluations (modeled by the constraints $\mathbf{A}\mathbf{x} = b$ are made in a certain manner on the torus $\mathbb{T}^n$, and not on a random sample of points in $\mathbb{R}^n$). Interestingly, this provides us with an important case of compressed sensing where exact recovery is guaranteed even though the RIP property is not satisfied. However from a practical side the SDP formulation is more efficient and elegant. Indeed for instance in the univariate case the size of the Toeplitz matrix involved is directly related to the number of atoms to recover whereas in the compressed sensing LP-approach, one has to fix *a priori* the length N of the vector $\mathbf{x}$ (which depends on the degree of the unknown polynomial, possibly very large) even if ultimately one is interested only in its few non zero entries (usually a very small number).

• Finally we provide a numerical comparison of the three approaches (LP, SDP and Prony) on a sample of problems and comment on their respective advantages and drawbacks. We clarify the relationship between Prony's method and super-resolution. In [7] Prony's method was briefly mentioned and neglected as sensitive to noise in the data (in contrast to super-resolution). We try to clarify this statement: actually, super-resolution *requires* Prony's method (or some variant of it) to extract relevant information from the output (the optimal solution) of the semidefinite program. In other words, super-resolution preprocesses the input data to Prony's method via a convex optimization procedure. We find that this sometimes helps to deal with noise in the context of polynomial interpolation, confirming the elegant theory of [7]. In some instances, super resolution does not perform well because of numerical issues present in current semidefinite programming solvers. To the best of our knowledge this drawback has not been discussed in the literature.

2 Notation, definitions and Preliminary results

2.1 Notation and definitions

Let $\mathbb{R}[\mathbf{x}]$ (resp. $\mathbb{R}[\mathbf{x}]_d$) denote the ring of real polynomials in the variables $\mathbf{x} = (x_1, \dots, x_n)$ (resp. polynomials of degree at most d), whereas $\Sigma[\mathbf{x}]$ (resp. $\Sigma[\mathbf{x}]_d$) denotes its subset of sums of squares (SOS) polynomials (resp. of SOS of degree at most $2d$). For every $\alpha = (\alpha_1, \dots, \alpha_n) \in \mathbb{N}^n$ the notation $\mathbf{x}^\alpha$ stands for the monomial $x_1^{\alpha_1} \cdots x_n^{\alpha_n}$ and for every $i \in \mathbb{N}$, let $\mathbb{N}^p_d := \{\beta \in \mathbb{N}^n : \sum_j \beta_j \leq d\}$ whose cardinal is $s(d) = \binom{n+d}{n}$. A polynomial $f \in \mathbb{R}[\mathbf{x}]$ is written $f = \sum_{\alpha \in \mathbb{N}^n} f_\alpha \, \mathbf{x}^\alpha$ with f_α almost all equal to zero, and f can be identified with its vector of coefficients $\mathbf{f} = (f_\alpha)$ in the canonical basis $(\mathbf{x}^\alpha)$, $\alpha \in \mathbb{N}^n$.

Denote by $\mathbb{R}[\mathbf{x}]^*_d$ the space of linear functionals on $\mathbb{R}[\mathbf{x}]_d$, identified with $\mathbb{R}^{s(d)}$. For a closed set $\mathbf{K} \subset \mathbb{R}^n$ denote by $C_d(\mathbf{K}) \subset \mathbb{R}[\mathbf{x}]_d$ the convex cone of polynomials of degree at most d that are nonnegative on $\mathbf{K}$, and for $f \in \mathbb{R}[\mathbf{x}]_d$, let

$$\|f\|_1 := \|\mathbf{f}\|_1 = \sum_{\alpha \in \mathbb{N}^n_d} |f_\alpha|.$$

Denote by $\mathcal{S}^t \subset \mathbb{R}^{t \times t}$ the space of real symmetric matrices, and for any $\mathbf{A} \in \mathcal{S}^t$ the notation $\mathbf{A} \succeq 0$ stands for $\mathbf{A}$ is positive semidefinite.

A real sequence $\mathbf{z} = (z_\alpha)$, $\alpha \in \mathbb{N}^n$, has a *representing measure* supported on a set $S \subset \mathbb{R}^n$ if there exists some finite Borel measure μ on S such that

$$z_\alpha = \int_S \mathbf{x}^\alpha \, d\mu(\mathbf{x}), \qquad \forall\, \alpha \in \mathbb{N}^n.$$

The space of finite Borel (signed) measures (resp. continuous functions) on $S \subset \mathbb{R}^n$ is denoted by $\mathscr{M}(S)$ (resp. $\mathscr{C}(S)$).

2.2 Super-resolution

Let $S \subset \mathbb{R}^n$ and suppose that μ is a signed atomic measure supported on a few atoms $(\mathbf{z}_i) \subset S$, $i = 1, \ldots, s$, i.e., $\mu = \sum_{k=1}^{s} w_i \, \delta_{\mathbf{z}_i}$. Super-resolution is concerned with retrieving the supports $(z_i) \subset S$ as well as the weights $(w_i) \subset \mathbb{R}$, from the sole knowledge of a few (and as few as possible) "moments" $(a_k = \int_S g_k \, d\mu)$, $k = 1, \ldots, m$, for some functions (g_k). One possible approach is to solve the convex optimization problem:

$$\rho \;=\; \inf_{\mu \in \mathscr{M}(S)} \{\, \|\mu\|_{TV} : \quad \int_S g_k \, d\mu \;=\; a_k, \quad k = 1, \ldots, m \,\} \tag{2.1}$$

where $\mathscr{M}(S)$ is the space of finite signed Borel measures on S equipped with the total-variation nom $\|\cdot\|_{TV}$. The dual of (2.1) reads:

$$\rho^* \;=\; \sup_{\lambda \in \mathbb{R}^m} \{\, \mathbf{a}^T \lambda : \| \sum_{k=1}^{m} \lambda_k \, g_k \|_\infty \;\leq\; 1 \,\}, \tag{2.2}$$

where $\|f\|_\infty = \sup_{\mathbf{x} \in S} |f(\mathbf{x})|$. (In fact and interestingly, both programs (2.1) and its dual (2.2) have already appeared in the sixties in a convex and elegant formulation of some bang-bang type optimal control problems; see Neustadt [28] and Krasovskii [23].) The rationale behind this approach is the analogy with *compressed sensing*. Indeed, the total variation norm $\|\mu\|_{TV}$ is the analogue for measures of the ℓ_1-norm for vectors[2].

In the univariate case when S is an interval (one may also consider the torus $\mathbb{T} \subset \mathbb{C}$) and the g_k's are the usual algebraic monomials $(\mathbf{x}^k)$, solving (2.1) then reduces to solving a single semidefinite program (SDP) and Candès and Fernandez-Granda [7] have shown that exact reconstruction is guaranteed provided that the (unknown) s supports (z_i) are sufficiently spaced and $m \geq \max[4s, 128]$. This approach was later generalized to arbitrary dimension and semi-algebraic sets in De Castro et al [13]; in contrast to the univariate case, one has to solve a hierarchy of semidefinite programs (instead of a single one). In the 2-dimensional and 3-dimensional examples treated in [13], exact recovery is obtained rapidly.

Alternatively one may also recover μ via the algebraic multivariate Prony method described in [27] and the references therein, and for which *no* minimum geometric separation of the supports is required. In addition, in the univariate case only $m = 2r$ moments are needed for exact recovery.

2.3 The multivariate Prony method

2.3.1 Hankel Prony

A multivariate Prony method has been proposed in [27,19] and is implemented in https://gitlab.inria.fr/mourrain/PolyExp. We refer to it in this paper as "Hankel Prony". It consists in two successive linear algebra operations.

Input

- Measurements $y_\alpha \in \mathbb{C}$ for $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{N}^n$ up to a degree $\|\alpha\|_1 = \sum_{i=1}^{n} \alpha_k \leqslant d$
- A threshold $\epsilon > 0$ to determine the numerical rank

Output Atomic measure μ

1. For $d_1 := \lfloor \frac{d}{2} \rfloor$ and $d_2 := \lceil \frac{d}{2} \rceil$ (where $\lfloor \cdot \rfloor$ and $\lceil \cdot \rceil$ denote the ceiling and floor of an integer), a singular value decomposition of a submatrix containing the measurements (i.e. $H_0 = (y_{\alpha+\beta})_{|\alpha| \leqslant d_1, |\beta| \leqslant d_2 - 1} = U \Sigma V^*$ where $(\cdot)^*$ stands for adjoint); the threshold $\epsilon > 0$ is used to determine the numerical rank r using the ratio of successive singular values. Precisely, the singular values in the diagonal matrix Σ are sorted in decreasing order and the rank is taken to be equal to the first instance when the ratio drops below the threshold. Multiplication matrices of size $r \times r$ can then be formed for each variable, i.e. $M_k = \Sigma_r^{-1} U_r H_k V_r$ where $H_k = (y_{\alpha+\beta+e_k})_{|\alpha| \leqslant d_1, |\beta| \leqslant d_2 - 1}$, Σ_r contains the r greatest

[2] To see this suppose that μ is the signed atomic measure $\sum_{i=1}^{s} \theta_i \delta_{\mathbf{x}_i}$. Then $\|\mu\|_{TV} = \|\theta\|_1$.

singular values in its diagonal, U_r is composed of the first r rows of the conjugate transpose of U, V_r is composed of the first r columns of the conjugate of V, and e_k denotes the row vector of size n of all zeros apart from 1 in position k.

2. An eigen-decompositon of a random linear combination of the multiplication matrices $\sum_{k=1}^{n} \lambda_k M_k = PDP^{-1}$ (for generic $\lambda_1, \dots, \lambda_n \in \mathbb{R}$) yields the atoms and the weights of the measure $\mu := \sum_{i=1}^{r} \omega_i \delta_{\xi_i}$. Precisely, the atoms are $\xi_i := ||P_i||_2^{-2}(P_i^* M_k P_i)_{1 \leqslant k \leqslant n}$ where P_i denotes the i^{th} column of P and the weights are

$$w_i := \frac{e_1 H_0 V_r P_i}{(\xi_i^\alpha)_{\|\alpha\|_1 \leqslant d_2 - 1} V_r P_i}. \tag{2.3}$$

We apply the above procedure to retrieve a measure from the output of the semidefinite optimization in super-resolution.

2.3.2 Toeplitz Prony

We now describe a new version of Prony's method, which we refer to as "Toeplitz Prony". In the setting of polynomial interpolation, Prony's method can be adapted to exploit the fact that we are interested in finding an atomic measure supported on the torus with real weights. As a result, fewer evaluations are necessary. For simplicity, we described this idea in the univariate setting, but it applies to the multivariate setting as well.

Following Kunis [24] say we are searching for a measure of finite support of the form $\mu = \sum_{k=1}^{r} \omega_k \delta_{\xi_k}$ where the weights ω_k are real and the support points ξ_k with coordinates of norm 1. Prony's method is based on the fact that the polynomial $p(z) = z^r - \sum_{k=1}^{r-1} p_k z^k := (z - \xi_1) \dots (z - \xi_r)$ satisfies $\int_{\mathbb{C}} q(z) p(z) d\mu = 0$ for any $q(z) \in \mathbb{C}[z]$. We consider instead the following relations

$$\begin{aligned} \int_{\mathbb{C}} \bar{z}^0 p(z) d\mu &= 0, \\ &\vdots \\ \int_{\mathbb{C}} \bar{z}^{r-1} p(z) d\mu &= 0, \end{aligned} \tag{2.4}$$

yielding

$$\begin{pmatrix} \sigma_0 & \dots & \sigma_{r-1} \\ \vdots & & \vdots \\ \overline{\sigma_{r-1}} & \dots & \sigma_0 \end{pmatrix} \begin{pmatrix} p_0 \\ \vdots \\ p_{r-1} \end{pmatrix} = \begin{pmatrix} \sigma_r \\ \vdots \\ \sigma_1 \end{pmatrix} \tag{2.5}$$

where $\sigma_k = \int z^k d\mu$ and $\overline{\sigma_k} = \int \overline{z}^k d\mu = \int z^{-k} d\mu$ since μ has real weights and the coordinates of its support points are of norm 1. Note that only $r + 1$ evaluations are needed and that the above matrix is a Toeplitz matrix, as opposed to the Hankel matrix of the Prony method. Both matrices have the same size, but to construct the Hankel matrix, $2r$ moments σ_k are needed.

The approach can be extended to the multivariate case, with Toeplitz like moment matrices. The rows are indexed by monomials and columns indexed by anti-monomials, that is, monomials with negative exponents. The entries of the matrix indexed by $(\alpha, -\beta)$ with $\alpha, \beta \in \mathbb{N}^n$ is $\sigma_{\alpha - \beta} = \int z^{\alpha - \beta} d\mu$. The same algorithm as in the Hankel Prony approach can then be used to obtain the decomposition of the measure from its moments. Note that the variant of Prony's method [33] (which also uses Toeplitz matrices) is computationally more demanding and thus not relevant here.

2.3.3 Advanced Prony

We now describe a more elaborate form of Prony's method, which we will refer to as "Advanced Prony". The multivariate Prony method decomposes a multi-index sequence $\sigma = (\sigma_\alpha)_{\alpha \in \mathbb{N}^n} \in \mathbb{C}^{\mathbb{N}^n}$, or equivalently a multivariate series, into a sum of polynomial-exponential sequences or series, from a finite set $\{\sigma_\alpha, \alpha \in A \subset \mathbb{N}^n\}$ of coefficients.

In the case of sparse interpolation, the coefficients σ_α of the series are the values $f(\zeta^\alpha)$ for $\alpha \in A \subset \mathbb{N}^n$. If $f = \mathbf{x}^\beta$ $\beta \in \mathbb{N}^n$, the corresponding series is the exponential series of ξ, where $\xi = \zeta^\beta$. Therefore if $f = \sum_{i=1}^{r} \omega_i \mathbf{x}^{\beta_i}$ is a sparse polynomial, the series $\sigma_\alpha = f(\zeta^\alpha)$ decomposes into a sum of r exponential series with weights ω_i and frequencies $\xi_i = \zeta^{\beta_i}$. The weights ω_i are the coefficients of the monomials of f and the frequencies ζ^{β_i} yield the exponents $\beta_i = \log_\zeta(\xi)$ of the monomials.

To compute this decomposition, we apply the following method. Subsets of monomials $A_0, A_1 \subset \mathbf{x}^{\mathbb{N}^n}$ are chosen adequately so that the rank of the Hankel matrix

$$H_0 = (\sigma_{\alpha_0+\alpha_1})_{\alpha_0\in A_0,\alpha_1\in A_1}$$

is the number of terms r. The Hankel matrices $H_i = (\sigma_{e_i+\alpha_0+\alpha_1})_{\alpha_0\in A_0,\alpha_1\in A_1}$ are also computed for $i = 1,\dots,n$ and (e_i) is canonical basis of $\mathbb{N}^n$. The subsets A_0, A_1 are chosen so that the monomial sets $\mathbf{x}^{A_0}$ and $\mathbf{x}^{A_1}$ contain a basis of the quotient algebra of the polynomials modulo the vanishing ideal of the points $\xi_1,\dots,\xi_r$.

Using Singular Value Decomposition [19] or a Gramm-Schmidt orthogonalization process [27], tables of multiplication by the variables in a basis of the associated Artinian Gorenstein algebra $\mathcal{A}_\sigma$ are deduced. The frequencies $\xi_i \in \mathbb{C}^n$, which are the points of the algebraic variety associated to $\mathcal{A}_\sigma$, are obtained by solving techniques from multiplication tables, based on eigenvector computation. The weights ω_i can then be deduced from the eigenvectors of these multiplication operators.

To compute this decomposition, only the evaluations $f(\zeta^\alpha)$ with $\alpha \in A = \cup_{i=1}^n e_0 + A_0 + A_1$ (where $e_0 = (0,\dots,0)$) are required.

Naturally, the "Advanced Prony" can be adapted to the Hankel and Toeplitz cases described in the two previous sections, yielding approaches which we will refer to as "Advanced H. Prony" and "Advanced T. Prony".

3 Sparse Interpolation

In §2.3 we have seen how to solve the sparse interpolation problem via Prony's method. We now consider two other approaches which both solve some convex optimization problem with a sparsity-inducing criterion.

3.1 A (naive) LP-approach to interpolation

Suppose that $g^* \in \mathbb{R}[\mathbf{x}]_d$ is an unknown polynomial of degree d and we can make a certain number of "black-box" evaluations $g^*(\mathbf{x}_k) = a_k$ at some points $(\mathbf{x}_k) \subset S$, $k = 1,\dots,s$, that we may choose to our convenience. Consider the following optimization problem $\mathbf{P}$:

$$\begin{aligned}\mathbf{P}: \rho = &\inf_{g\in\mathbb{R}[\mathbf{x}]_d} \{\, \|g\|_1 : \quad g(\mathbf{x}_k) \,=\, a_k, \quad k = 1,\dots,m \,\} &(3.1)\\ = &\inf_{g\in\mathbb{R}[\mathbf{x}]_d} \{\, \|g\|_1 : \quad \langle g, \delta_{\mathbf{x}_k}\rangle \,=\, a_k, \quad k = 1,\dots,m \,\} &(3.2)\end{aligned}$$

where $\delta_{\mathbf{x}_k}$ is the Dirac at the point $\mathbf{x}_k \in \mathbb{R}^n$, and $\langle\cdot,\cdot\rangle$ the duality bracket $\int_S f d\mu$ between $\mathscr{C}(S)$ and $\mathscr{M}(S)$. Equivalently $\mathbf{P}$ also reads:

$$\rho = \inf_g \{ \sum_{\alpha\in\mathbb{N}^n_d} |g_\alpha| : \sum_\alpha g_\alpha\, \mathbf{x}_k^\alpha \,=\, a_k, \quad k = 1,\dots,m \,\}, \tag{3.3}$$

or in the form of an LP as:

$$\rho = \inf_{g^+,g^-\geq 0} \{ \sum_{\alpha\in\mathbb{N}^n_d} (g^+_\alpha + g^-_\alpha) : \sum_\alpha g^+_\alpha\, \mathbf{x}_k^\alpha - g^-_\alpha \mathbf{x}_k^\alpha \,=\, a_k, \quad k = 1,\dots,m \,\} \tag{3.4}$$

which is an LP. Let $\mathbf{a} = (a_k)$, $k = 1,\dots,m$. The dual of the LP (3.1) is the LP:

$$\begin{aligned}\mathbf{P}^*: \rho = &\sup_{\lambda\in\mathbb{R}^m} \{\, \mathbf{a}^T\lambda : \quad |\sum_{k=1}^m \lambda_k\, \mathbf{x}_k^\alpha\,| \,\leq\, 1; \quad \alpha \in \mathbb{N}^n_d \,\}\\ = &\sup_{\lambda\in\mathbb{R}^m} \{\mathbf{a}^T\lambda : \quad \underbrace{|\langle \mathbf{x}^\alpha, \sum_{k=1}^m \lambda_k\, \delta_{\mathbf{x}_k}\rangle|}_{=|m_\lambda(\alpha)|} \,\leq\, 1; \quad \alpha \in \mathbb{N}^n_d \,\}\\ = &\sup_{\lambda\in\mathbb{R}^m} \{\mathbf{a}^T\lambda : \quad \|m_\lambda\|_\infty \,\leq\, 1 \,\}, &(3.5)\end{aligned}$$

where to every $\lambda \in \mathbb{R}^m$ is associated the vector $m_\lambda \in \mathbb{R}^{s(d)}$ defined by

$$m_\lambda(\alpha) \,:=\, \langle \mathbf{x}^\alpha, \sum_{k=1}^m \lambda_k\, \delta_{\mathbf{x}_k} \rangle \,=\, \sum_{k=1}^m \lambda_k\, \mathbf{x}_k^\alpha.$$

So in the dual $\mathbf{P}^*$ one searches for λ^*, equivalently the signed atomic measure $\mu^* := \sum_{k=1}^m \lambda_k^*\, \delta_{\mathbf{x}_k}$, as we also do in super-resolution (2.1) (but in $\mathbf{P}^*$ the support is known).

Lemma 3.1 *Let $(\hat{g}^+, \hat{g}^-)$ be an optimal solution of the naive LP (3.4) with associated polynomial $\mathbf{x} \mapsto \hat{g}(\mathbf{x}) := \hat{g}^+(\mathbf{x}) - \hat{g}^-(\mathbf{x})$, and let $\lambda^* \in \mathbb{R}^s$ be an optimal solution of its dual (3.5). Then:*

(i) $\hat{g}$ has at most m non-zero entries, out of potentially $s(d) = \binom{n+d}{n}$.

(ii) $-1 \le m_{\lambda^*}(\alpha) \le 1$ for all $\alpha \in \mathbb{N}^n_d$, and

$$\hat{g}_\alpha \,>\, 0 \quad \Rightarrow \quad \underbrace{m_\lambda(\alpha) = 1}_{\int \mathbf{x}^\alpha d\mu^* = 1}; \qquad \hat{g}_\alpha \,<\, 0 \quad \Rightarrow \quad \underbrace{m_\lambda(\alpha) = -1}_{\int \mathbf{x}^\alpha d\mu^* = -1}, \tag{3.6}$$

The proof is by standard arguments of LP and the Simplex algorithm.

Exact recovery. A natural issue is *exact recovery*, i.e., is there a value of m (with possibly $m \ll O(n^d)$) for which $\hat{g} = g^*$? And if yes, how small m must be?

A well-known and famous condition for exact recovery of sparse solution $\mathbf{x}^*$ to

$$\min_{\mathbf{x}} \{\, \|\mathbf{x}\|_1 : \quad \mathbf{A}\,\mathbf{x} \,=\, b \,\} \tag{3.7}$$

is the so-called *Restricted Isometry Property* (RIP) of the matrix $\mathbf{A}$ introduced in Candès and Tao [8] (see also Candès [5, Definition 1.1]) from which celebrated results of Candès et al. [9] in compressed sensing could be obtained. The interested reader is also referred to Fan and Kamath [15] for an interesting recent comparison of various algorithms to solve (3.7) (even in the case where the RIP does not hold).

It turns out that for Problem (3.3) the resulting (Vandermonde-like) matrix $\mathbf{A}$ does *not* satisfy the RIP property, and so exact recovery of the sparse polynomial as solution of (3.3) (and equivalently of the LP (3.4)) is *not* garanteed in general.

3.2 A formal analogy with super-resolution

Observe that (3.2) is the analogue in function spaces of the *super-resolution* problem in measure spaces. Indeed in both dual problems (2.2) and (3.5) one searches for a real vector $\lambda \in \mathbb{R}^m$. In the former it is used to build up a *polynomial* $h := \sum_{k=1}^m \lambda_k^* g_k$ uniformly bounded by 1 on S ($\|h\|_\infty \le 1$) while in the latter it is used to form an *atomic measure* $\sum_{k=1}^m \lambda_k^* \delta_{\mathbf{x}_k}$ whose moments (up to some order d) are uniformly bounded by 1 ($\|m_\lambda\|_\infty \le 1$).

Moreover, let (μ^*, λ^*) be a pair of optimal solutions to (2.1)-(2.2). Then $\mu^* = \mu^+ - \mu^-$ where μ^+ and μ^- are positive atomic measures respectively supported on disjoint sets $X_1, X_2 \subset S$, and each point of X_1 (resp. X_2) is a zero of the polynomial $\mathbf{x} \mapsto 1 - \sum_{k=1}^m \lambda_k^*\, g_k(\mathbf{x})$ (resp. $\mathbf{x} \mapsto 1 + \sum_{k=1}^m \lambda_k^*\, g_k(\mathbf{x})$). That is:

$$\sup_{\mathbf{x} \in S} |\sum_{k=1}^m \lambda_k^*\, g_k(\mathbf{x})| \,\le\, 1, \quad \text{and} \tag{3.8}$$

$$\mathbf{x}_j \in \mathrm{supp}(\mu^+) \,\Rightarrow\, \sum_{k=1}^m \lambda_k^*\, g_k(\mathbf{x}_j) \,=\, 1; \quad \mathbf{x}_j \in \mathrm{supp}(\mu^-) \,\Rightarrow\, \sum_{k=1}^m \lambda_k^*\, g_k(\mathbf{x}_j) \,=\, -1, \tag{3.9}$$

(compare with (3.6)).

3.3 Sparse interpolation as super-resolution

In §3.2 we have shown that the "compressed sensing" formulation (3.1) of the sparse interpolation problem could be viewed as a "formal analogue" in function spaces of the super-resolution problem in measure spaces.

In this section we show that sparse interpolation is in fact a true (as opposed to formal) super-resolution problem on the torus $\mathbb{T}^n \subset \mathbb{C}^n$, provided that evaluations are made at points chosen in a certain adequate manner. So let

$$\mathbf{x} \mapsto g(\mathbf{z}) \;=\; \sum_{\beta \in \mathbb{N}^n_d} g_\beta\, \mathbf{z}^\beta,$$

be the black-box polynomial with unknown real coefficients $(g_\beta) \subset \mathbb{R}$.

A crucial observation. Let $\mathbf{z}_0 \in \mathbb{T}^n$ (with $\mathbf{z}_0 \neq (1,\ldots,1)$) be fixed, e.g., of the form:

$$\mathbf{z}_0 \;:=\; (\exp(2i\pi/N), \ldots, \exp(2i\pi/N)), \tag{3.10}$$

for some arbitrary (fixed) integer N, or

$$\mathbf{z}_0 \;:=\; (\exp(2i\pi\theta_0), \ldots, \exp(2i\pi\theta_0)), \tag{3.11}$$

for some arbitrary (fixed) irrational $\theta_0 \in \mathbb{R}$. With the choice (3.11)

$$[\ \alpha, \beta \in \mathbb{Z}^n \text{ and } \alpha \neq \beta\ ] \quad \Rightarrow \quad \mathbf{z}_0^\alpha \;\neq\; \mathbf{z}_0^\beta.$$

whereas with the choice (3.10)

$$[\ \alpha, \beta \in \mathbb{Z}^n, \max_i \max[|\alpha_i|, |\beta_i|] < N, \text{ and } \alpha \neq \beta\ ] \quad \Rightarrow \quad \mathbf{z}_0^\alpha \;\neq\; \mathbf{z}_0^\beta.$$

Next for every $\alpha \in \mathbb{N}^n$:

$$\begin{aligned} a_\alpha \;:=\; g(\mathbf{z}_0^\alpha) \;=\; \sum_{\beta\in\mathbb{N}^n} g_\beta\,(z_0^\alpha)^\beta &= \sum_{\beta\in\mathbb{N}^n} g_\beta\,(z_{01}^{\alpha_1})^{\beta_1}\cdots(z_{0n}^{\alpha_n})^{\beta_n} \\ &= \sum_{\beta\in\mathbb{N}^n} g_\beta\,(z_{01}^{\beta_1})^{\alpha_1}\cdots(z_{0n}^{\beta_n})^{\alpha_n} \\ &= \sum_{\beta\in\mathbb{N}^n} g_\beta\,(z_0^\beta)^\alpha \;=\; \int_{\mathbb{T}^n} \mathbf{z}^\alpha\, d\mu_{g,\mathbf{z}_0}(\mathbf{z}), \end{aligned} \tag{3.12}$$

where $\mu_{g,\mathbf{z}_0}$ is the signed atomic-measure on $\mathbb{T}^n$ defined by:

$$\mu_{g,\mathbf{z}_0} \;:=\; \sum_{\beta\in\mathbb{N}^n} g_\beta\,\delta_{\xi_\beta} \quad (\text{and } \|\mu_{g,\mathbf{z}_0}\|_{TV} = \textstyle\sum_\beta |g_\beta| = \|g\|_1), \tag{3.13}$$

where $\xi_\beta := \mathbf{z}_0^\beta \in \mathbb{T}^n$, for all $\beta \in \mathbb{N}^n$ such that $g_\beta \neq 0$, and δ_ξ is the Dirac probability measure at the point $\xi \in \mathbb{C}^n$.

In other words: *Evaluating g at the point $\mathbf{z}_0^\alpha \in \mathbb{T}^n$ is the same as evaluating the moment $\int_{\mathbb{T}^n} \mathbf{z}^\alpha\, d\mu_{g,\mathbf{z}_0}$ of the signed atomic-measure $\mu_{g,\mathbf{z}_0}$. Therefore, the sparse interpolation problem is the same as recovering the finitely many unknown weights $(g_\beta) \subset \mathbb{R}$ and supports $(\mathbf{z}_0^\beta) \subset \mathbb{T}^n$ of the signed measure $\mu_{g,\mathbf{z}_0}$ on $\mathbb{T}^n$, from finitely many s moments of $\mu_{g,\mathbf{z}_0}$, that is, a super-resolution problem.*

Remark 3.2 *The n-dimensional torus $\mathbb{T}^n$ is one among possible choices but any other choice of a set $S \subset \mathbb{C}^n$ and $\mathbf{z}_0 \in \mathbb{C}^n$ (or $S \subset \mathbb{R}^n$ and $\mathbf{z}_0 \in \mathbb{R}^n$) is valid provided that $(\mathbf{z}_0^\alpha)_{\alpha\in\mathbb{N}^n} \subset S$. For instance $\mathbf{z}_0 \in (-1,1)^n$ and $S := [-1,1]^n$ is another possible choice.*

Let $\|\mathbf{z}\|_\infty$ denote the usual sup-norm $\sup_i |\mathbf{z}_i|$ in $\mathbb{C}^n$. Given $f_c \in \mathbb{N}$ define the set

$$\mathscr{A}_{f_c} := \{-f_c, -(f_c-1), \ldots, -1, 0, 1, \ldots, (f_c-1), f_c\}^n \subset \mathbb{Z}^n.$$

With the choice $S := \mathbb{T}^n$, $\mathbf{z}_0 \in \mathbb{T}^n$ as in (3.10) or in (3.11), and $f_c \in \mathbb{N}$, consider the optimization problems:

$$\rho_{f_c} = \inf_{\mu \in \mathscr{M}(\mathbb{T}^n)} \{ \|\mu\|_{TV} : \quad \int_{\mathbb{T}^n} \mathbf{z}^\alpha \, d\mu(\mathbf{z}) = a_\alpha, \quad \alpha \in \mathscr{A}_{f_c} \}, \tag{3.14}$$

where $a_\alpha = g(\mathbf{z}_0^\alpha)$ is obtained from the black-box polynomial $g \in \mathbb{R}[\mathbf{z}]$, and

$$\rho^*_{f_c} = \sup_{g \in \mathbb{C}[\mathbf{z}; \mathscr{A}_{f_c}]} \{ \Re(\mathbf{a}^T \mathbf{g}) : \|\Re(g(\mathbf{z}))\|_\infty \leq 1 \} \tag{3.15}$$

(where $\mathbf{a} = (a_\alpha)$). Notice that the super-resolution problem problem (3.14) has the following equivalent formulation in terms of an infinite dimensional LP

$$\rho_{f_c} = \inf_{\mu^+, \mu^- \in \mathscr{M}(\mathbb{T}^n)} \{ \int_{\mathbb{T}^n} 1 \, d(\mu^+ + \mu^-) : \int_{\mathbb{T}^n} \mathbf{z}^\alpha \, d(\mu^+ - \mu^-) = a_\alpha, \quad \alpha \in \mathscr{A}_{f_c} \}, \tag{3.16}$$

with same dual (3.15) as (3.14). Moreover $\rho_{f_c} = \rho^*_{f_c}$; the proof for $S = \mathbb{T}^n$ is very similar to the proof in De Castro et al. [13] for the case where $S \subset \mathbb{R}^n$ is a compact semi-algebraic set.

Theorem 3.3 *Let $g^* \in \mathbb{R}[\mathbf{z}]$, $\mathbf{x} \mapsto g^*(\mathbf{z}) := \sum_\beta g^*_\beta \mathbf{z}^\beta$, be an unknown real polynomial. Let $\Gamma := \{\beta \in \mathbb{N}^n : g^*_\beta \neq 0\}$ and $s := |\Gamma|$. Let $\mathbf{z}_0 \in \mathbb{T}^n$ be as in (3.11) or in (3.10) (in which case $N > \max_{i=1,\ldots,n} \max\{\beta_i : \beta \in \Gamma\}$), and let $a_\alpha = g(\mathbf{z}_0^\alpha)$, $\alpha \in \mathscr{A}_{f_c}$.*

Let $0 < \delta := \min\{ \|\mathbf{z}_0^\alpha - \mathbf{z}_0^\beta\|_\infty : \alpha, \beta \in \Gamma, \, \alpha \neq \beta \}$. There is a constant $C_n > 0$ (that depends only on the dimension[3] n) such that if $\delta > C_n / f_c$ then the optimization problem (3.14) has a unique optimal solution μ^* such that

$$\mu^* := \sum_{\alpha \in \Gamma} g^*_\beta \, \delta_{\mathbf{z}_0^\beta} \quad \textit{and} \quad \|\mu^*\|_{TV} = \sum_\alpha |g^*_\beta|. \tag{3.17}$$

In addition, there is no duality gap (i.e., $\rho_{f_c} = \rho^*_c$), (3.15) has an optimal solution $f^* \in \mathbb{C}[\mathbf{x}; \mathscr{A}_{f_c}]$, and

$$g^{*+}_\beta > 0 \Rightarrow \Re(f^*(\mathbf{z}_0^\beta)) = 1; \quad g^{*-}_\beta > 0 \Rightarrow \Re(f^*(\mathbf{z}_0^\beta)) = -1. \tag{3.18}$$

Proof Of course the measure μ^* in (3.17) is feasible for (3.14). From the definition of N and Γ, all points $(\mathbf{z}_0^\alpha) \subset \mathbb{T}^n$, $\alpha \in \Gamma$, are distinct whenever $\mathbf{z}_0$ is chosen as in (3.10) or in (3.11). Moreover, from [7, Theorem 1.2, 1.3] and the extensions in higher dimensions (see remark after Theorem 1.3 [7]), it follows that under the separation condition $\delta > C_n/f_c$, the optimal solution of (3.14) is unique and is the sparse measure on $\mathbb{T}^n$ that satisfies the moment conditions of (3.14), i.e., μ^*.

Next, write the optimal solution μ^* of (3.14) as $\mu^* = \mu^+ - \mu^-$ for two signed Borel measures $\mu^+, \mu^- \in \mathscr{M}(\mathbb{T}_n)_+$, i.e.,

$$\mu^+ = \sum_\beta g^{*+}_\beta \, \delta_{\mathbf{z}_0^\beta}; \quad \mu^- = \sum_\beta g^{*-}_\beta \, \delta_{\mathbf{z}_0^\beta}.$$

We have already mentioned that from [13], the optimal values of (3.14), (3.15) and (3.16) are the same, i.e., $\rho_{f_c} = \rho^*_{f_c}$, and therefore the measures μ^+ and μ^- are optimal solutions of (3.16). Let f^* be an optimal solution of (3.15). One relates μ^* and f^* has follows. As $\rho_{f_c} = \|\mu^*\|_{TV} = \rho^*_c = \Re(\mathbf{a}^T \mathbf{f}^*)$,

$$\begin{aligned} \|\mu^*\|_{TV} &= \int_{\mathbb{T}^n} d(\mu^+ + \mu^-) \\ &= \int_{\mathbb{T}^n} \underbrace{\Re(1 - f^*)}_{\geq 0} \, d\mu^+ \int_{\mathbb{T}^n} \underbrace{\Re(1 + f^*)}_{\geq 0} \, d\mu^- + \underbrace{\int_{\mathbb{T}^n} \Re(f^*) \, d(\mu^+ - \mu^-)}_{= \Re(\mathbf{a}^T \mathbf{f}^*)} \end{aligned}$$

it follow that μ^+ (resp. μ^-) is supported on the zeros of $\Re(1 - f^*)$ (resp. $\Re(1 + f^*)$) on $\mathbb{T}^n$. □

Therefore to recover s points one needs at most $(2s\, C_n + 1)^n$ evaluations.

[3] If $f_c \geq 128$ then $C_1 = 1.87$, $C_2 = 2.38$ as proved in Candès and Fernandez-Granda [7].

3.4 A hierarchy of SDP relaxations for solving the super-resolution problem

Recall that in the the super-resolution model described in Candès and Fernandez-Granda [7], one has to make evaluations in the multivariate case at all points $\mathbf{z}_0^\alpha$ with

$$\alpha \in \mathscr{A}_{f_c} = \{-f_c, -(f_c-1), \ldots, -1, 0, 1, \ldots, (f_c-1), f_c\}^n = \{\alpha \mid \|\alpha\|_\infty \leqslant f_c\} \subset \mathbb{Z}^n, \tag{3.19}$$

where $\|\alpha\|_\infty := \max\{|\alpha_1|, \ldots, |\alpha_n|\}$. This makes perfect sense in such applications as image reconstruction from measurements (typically 2-dimensional objects) of signal processing. However, for polynomial interpolation $|\mathscr{A}_{f_c}|$ is rapidly prohibitive if one consider polynomials of say $n = 10$ variables. Indeed, if $n = 10$ then the first order semidefinite program of the hierarchy entails matrix variables of size $1,024 \times 1024$. Bear in mind that currently, semidefinite programming solvers are limited to matrices of size a few hundred. Thus it is not possible to compute *even* the first order relaxation!

We propose to reduce the computional burden by using the one-norm truncation, i.e. $\|\alpha\|_1 := |\alpha_1| + \ldots + |\alpha_n|$, by making evaluations at all points $\mathbf{z}_0^\alpha$ with

$$\alpha \in \mathscr{A}^1_{f_c} := \{\alpha - \beta \mid \alpha, \beta \in \mathbb{Z}^n,\ \|\alpha\|_1, \|\beta\|_1 \leqslant f_c\}. \tag{3.20}$$

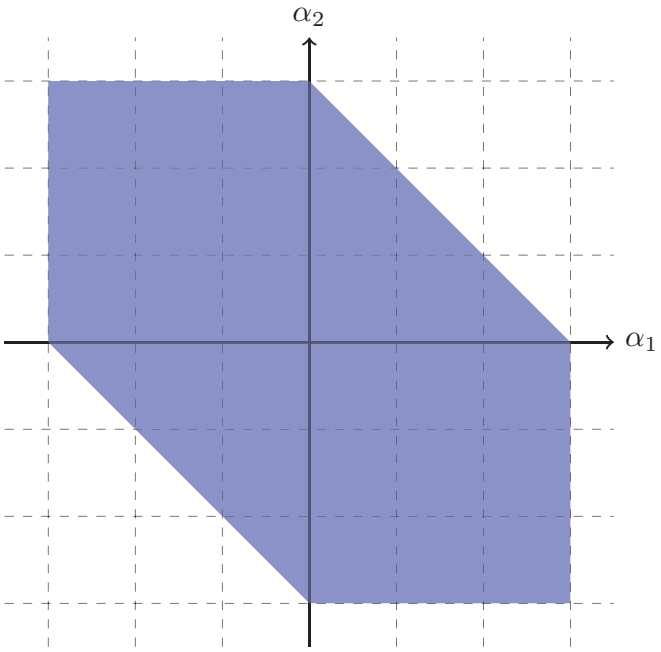

Fig. 1 Evaluations at $\alpha - \beta$ with $|\alpha_1| + |\alpha_2| \leqslant 3$ and $|\beta_1| + |\beta_2| \leqslant 3$ and $\alpha_1, \alpha_2, \beta_1, \beta_2 \in \mathbb{Z}$.

An illustration is provided in Figure 1. In addition:

$$\forall f_c \in \mathbb{N},\ \ \exists \tilde{f}_c \in \mathbb{N} :\ \ \forall l \geqslant \tilde{f}_c, \quad \mathscr{A}_{f_c} \subset \mathscr{A}^1_l. \tag{3.21}$$

Thus, all the theoretical results of [7] are preserved. To appreciate the gain in using $\mathscr{A}^1_{f_c}$, for 10 variables, the first order semidefinite program of the hierarchy entails matrix variables of size 11×11 (instead of $1,024 \times 1,024$) and 56 linear equalities. It is thus possible to compute the first order relaxation. The second order relaxation entails matrices of size 66×66 (instead of $59,049 \times 59,049$) and $1,596$ linear equalities.

Notice also that Toeplitz Prony of Section 2.3.2 uses $\mathscr{A}^1_{f_c}$ instead of $\mathscr{A}_{f_c}$ and is guaranteed to recover the optimal solution provided that f_c is large enough.

As the super-resolution problem (3.14) is concerned with atomic measures finitely supported on the multi-dimensional torus $\mathbb{T}^n$, we can adapt to the torus $\mathbb{T}^n$ the hierarchy of semidefinite programs defined in De Castro et al. [13] for solving super-resolution problems with atomic measures (finitely

supported) on semi-algebraic subsets of $\mathbb{R}^n$. For every fixed f_c, at step $d \geq f_c$ of the hierarchy, the semidefinite program to solve reads:

$$(\mathbf{P}_{d,f_c}) \qquad \begin{aligned} \rho_{d,f_c} = \inf_{y^+,y^-} \; & y_0^+ + y_0^- \\ \text{s.c.} \quad & y_\alpha^+ - y_\alpha^- = a_\alpha \;, \quad \forall \alpha \in \mathscr{A}_{f_c}^1 \\ & \mathbf{T}_d(\mathbf{y}^+) \succeq 0, \, \mathbf{T}_d(\mathbf{y}^-) \succeq 0, \end{aligned} \tag{3.22}$$

where the Hermitian matrix $\mathbf{T}_d(y^+)$ has its rows and columns indexed in $\{\alpha \mid \|\alpha\|_1 \leqslant d\}$ and $\mathbf{T}_d(\mathbf{y}^+)_{\alpha,\beta} = y^+_{\beta-\alpha}$, for every $\|\alpha\|_1, \|\beta\|_1 \leqslant d$, and similarly for the Hermitian matrix $\mathbf{T}_d(\mathbf{y}^-)$. In the univariate case $\mathbf{T}_d(\mathbf{y}^+)$ $\mathbf{T}_d(y^-)$ are Toeplitz matrices. When $\mathbf{y}^+$ is coming from a measure μ^+ on $\mathbb{T}^n$ then $y^+_\alpha = \int_{\mathbb{T}^n} \mathbf{z}^\alpha \, d\mu^+(\mathbf{z})$. Clearly, (3.22) is a relaxation of (3.16) and so $\rho_d \leq \rho_{f_c}$ for all $d \geq f_c$. Moreover $\rho_d \leq \rho_{d+1}$ for all d.

Note that with the above notations, the "Toeplitz Prony" method proposed in Section 2.3.2 consists in directly extracting a measure from the matrix $\mathbf{T}_d(a)$. In constrast, the super resolution approach consists of decomposing it into $\mathbf{T}_d(a) = \mathbf{T}_d(\mathbf{y}^+) - \mathbf{T}_d(\mathbf{y}^-)$, optimizing over $\mathbf{y}^+$ and $\mathbf{y}^-$, and then applying the "Toeplitz Prony" method to $\mathbf{T}_d(\mathbf{y}^+_*)$ and $\mathbf{T}_d(\mathbf{y}^-_*)$ at an optimal solution $(\mathbf{y}^+_*, \mathbf{y}^-_*)$ of (3.22).

Lemma 3.4 *For each $d \geq f_c$ the (complex) semidefinite program ($\boldsymbol{P}_{d,f_c}$) in* (3.22) *has an optimal solution $(\mathbf{y}^+, \mathbf{y}^-)$. In addition, if the rank conditions*

$$\mathit{rank}(\mathbf{T}_d(\mathbf{y}^+)) = \mathit{rank}(\mathbf{T}_{d-2}(\mathbf{y}^+)) \tag{3.23}$$

$$\mathit{rank}(\mathbf{T}_d(\mathbf{y}^-)) = \mathit{rank}(\mathbf{T}_{d-2}(\mathbf{y}^-)) \tag{3.24}$$

are satisfied then there exist two Borel atomic measures μ^+ and μ^- on $\mathbb{T}^n$ such that:

$$y^+_\alpha = \int_{\mathbb{T}^n} \mathbf{z}^\alpha \, d\mu^+(\mathbf{z}) \quad \textit{and} \quad y^-_\alpha = \int_{\mathbb{T}^n} \mathbf{z}^\alpha \, d\mu^-(\mathbf{z}), \quad \forall \alpha \in \mathscr{A}_d^1. \tag{3.25}$$

The support of μ^+ (resp. μ^-) consists of $\mathrm{rank}(\mathbf{T}_d(\mathbf{y}^+))$ *(resp.* $\mathrm{rank}(\mathbf{T}_d(\mathbf{y}^-))$*) atoms on $\mathbb{T}^n$ which can be extracted by a numerical algebra routine (e.g. the Prony method described in Section 2.3.1).*

In addition, if (3.23)-(3.24) *hold for an optimal solution of ($\boldsymbol{P}_{d,\tilde{f}_c}$) with $\tilde{f}_c$ as in* (3.21)*, then under the separation conditions of Theorem 3.3, the Borel measure $\mu^* := \mu^+ - \mu^-$ is the unique optimal solution of (3.14).*

Proof Consider a minimizing sequence $(\mathbf{y}^{+,\ell}, \mathbf{y}^{-,\ell})_{\ell \in \mathbb{N}}$ of (3.22). Since one minimizes $y_0^+ + y_0^-$ one has $y_0^{+,\ell} + y_0^{-,\ell} \leqslant y_0^{+,1} + y_0^{-,1} =: \rho$, for $\ell \geq 1$. The Toeplitz-like structure of $\mathbf{T}_d(\mathbf{y}^{+,\ell})$ and the psd constraint $\mathbf{T}_d(\mathbf{y}^{+,\ell}) \succeq 0$ imply $|y_\alpha^{+,\ell}| \leqslant \rho$ for all $\alpha \in \mathscr{A}_d^1$; and similarly $|y_\alpha^{-,\ell}| \leqslant \rho$ for all $\alpha \in \mathscr{A}_{f_c}^1$. Hence there is a subsequence (ℓ_k) and two vectors $\mathbf{y}^+ = (y^+_\alpha)_{\alpha \in \mathscr{A}_{f_c}^1}$ and $\mathbf{y}^- = (y^-_\alpha)_{\alpha \in \mathscr{A}_{f_c}^1}$, such that

$$\lim_{k \to \infty} \mathbf{y}^{+,\ell_k} = \mathbf{y}^+ \quad \text{and} \quad \lim_{k \to \infty} \mathbf{y}^{-,\ell_k} = \mathbf{y}^-.$$

In addition, from the above convergence it also follows that $(\mathbf{y}^+, \mathbf{y}^-)$ is a feasible solution of (3.22), hence an optimal solution of (3.22).

Next, in the univariate case, a Borel measure μ^+ and μ^- on $\mathbb{T}$ can always be extracted from the semidefinite positive Toeplitz matrices $\mathbf{T}_d(\mathbf{y}^+)$ and $\mathbf{T}_d(\mathbf{y}^-)$ respectively. This is true regardless of the rank conditions (3.23)-(3.24) and was proved in [37, P. 211]. In the multivariate case, $\mathbf{T}_d(\mathbf{y}^+)$ and $\mathbf{T}_d(\mathbf{y}^-)$ are Toeplitz-like matrices, and we may and will invoke the recent result [20, Theorem 5.2]. (Note that this is true for Toeplitz matrices, but *not* for general Hermitian matrices for which additional non-trivial conditions must be satisfied (see [20, Theorem 5.1]).) It implies that a Borel measure μ^+ (resp. μ^-) on $\mathbb{T}^n$ can be extracted from a multivariate semidefinite positive Toeplitz-like matrix $\mathbf{T}_d(\mathbf{y}^+)$ (resp. $\mathbf{T}_d(\mathbf{y}^-)$) if the rank condition (3.23) (resp. (3.24)) holds. Hence we have proved (3.25).

Finally the last statement follows from Theorem 3.3 and the fact that (3.14) and (3.16) have same optimal value and an optimal solution (μ^+, μ^-) of (3.16) provides and optimal solution $\mu^* = \mu^+ - \mu^-$ of (3.14). □

Asymptotics as d increases. In case the conditions (3.23)-(3.24) do not hold, we still have the following asymptotic result at an optimal solution.

Lemma 3.5 *Assume that f_c satisfies the conditions of Theorem 3.3 and let $\tilde{f}_c$ be as in* (3.21)*. For each $d \geq \tilde{f}_c$, let $(\mathbf{y}^{+,d}, \mathbf{y}^{-,d})$ be an optimal solution of (3.22). Then for each $\alpha \in \mathbb{Z}^n$,*

$$\lim_{d\to\infty} (y_\alpha^{+,d} - y_\alpha^{-,d}) \;=\; \int_{\mathbb{T}^n} \mathbf{z}^\alpha \, d\mu^*, \tag{3.26}$$

where the Borel signed measure μ^* on $\mathbb{T}^n$ is the unique optimal solution of (3.14) characterized in (3.17).

Proof As $\rho_{d,\tilde{f}_c} \leq \rho_{d+1,\tilde{f}_c} \leq \rho_{f_c}$ for all $d \geq \tilde{f}_c$ and $y_0^{+,d} + y_0^{-,d} = \rho_{d,\tilde{f}_c}$, it follows that $|y_\alpha^{+,d}| \leq \rho_{f_c}$ and $|y_\alpha^{-,d}| \leq \rho_{f_c}$ for all $\alpha \in \mathscr{A}_d^1$ and all $d \geq \tilde{f}_c$. By completing with zeros, one may and will consider all finite-dimensional vectors $\mathbf{y}^{+,d}$ and $\mathbf{y}^{-,d}$ as elements of a bounded set of ℓ_∞. Next, by weak-$\star$ sequential compactness of the unit ball of ℓ_∞, there exist infinite vectors $\mathbf{y}^+, \mathbf{y}^- \in \ell_\infty$, and a subsequence $(d_k)_{k\in\mathbb{N}}$ such that :

$$\lim_{k\to\infty} y_\alpha^{+,d_k} \;=\; y_\alpha^+; \quad \lim_{k\to\infty} y_\alpha^{-,d_k} \;=\; y_\alpha^-, \quad \forall \alpha \in \mathbb{Z}^n. \tag{3.27}$$

Moreover from the above convergence we also have $\mathbf{T}_d(\mathbf{y}^+) \succeq 0$ and $\mathbf{T}_d(\mathbf{y}^-) \succeq 0$ for all d. This in turn implies that $\mathbf{y}^+$ (resp. $\mathbf{y}^-$) is the moment sequence of a Borel measure μ^+ (resp. μ^-) on $\mathbb{T}^n$. In addition, the convergence (3.27) yields

$$a_\alpha \;=\; \lim_{k\to\infty} (y_\alpha^{+,d_k} - y_\alpha^{-,d_k}) \;=\; \int_{\mathbb{T}^n} \mathbf{z}^\alpha \, d(\mu^+ - \mu^-), \quad \forall \alpha \in \mathscr{A}_{\tilde{f}_c}^1,$$

and

$$\rho_{f_c} \geq \lim_{k\to\infty} \rho_{d_k,f_c} \;=\; \lim_{k\to\infty} (y_0^{+,d_k} + y_0^{-,d_k}) \;=\; \int_{\mathbb{T}^n} d(\mu^+ + \mu^-) \;\geq\; \|\mu^+ - \mu^-\|_{TV},$$

which proves that (μ^+, μ^-) is an optimal solution of (3.16). Therefore $\mu^* := \mu^+ - \mu^-$ is an optimal solution of (3.14) and thus unique when $\tilde{f}_c$ satisfies the condition of Theorem 3.3. This also implies that the limit y_α^+ (resp. y_α^-) in (3.27) is the same for all converging subsequences $(d_k)_{k\in\mathbb{N}^n}$ and therefore, for each $\alpha \in \mathbb{Z}^n$, the whole sequence $(y_\alpha^{+,d})_{d\in\mathbb{N}}$ (resp. $(y_\alpha^{-,d})_{d\in\mathbb{N}}$) converges to y_α^+ (resp. y_α^-), which yields the desired result (3.26). □

3.5 A rigorous compressed sensing LP approach

In this section we take advantage of an important consequence of viewing sparse interpolation as a super-resolution problem. Indeed when $\mathbf{z}_0$ is chosen as in (3.10) we *know* that the (unique) optimal solution μ^* of (3.14) is supported on the *a priori* fixed grid $(\exp(2i\pi k_1/N), \ldots, \exp(2i\pi k_n/N))$, where $0 \leq k_i \leq N$, $i = 1, \ldots, n$. That is, (3.14) is a *discrete* super-resolution problem as described in Candès and Fernandez-Granda [7]. Therefore solving (3.14) is also equivalent to solving the LP:

$$\min_{\mathbf{x}} \{ \|\mathbf{x}\|_1 : \; \mathbf{A}\,\mathbf{x} \;=\; \mathbf{b} \}$$

where $\mathbf{x} \in \mathbb{R}^{[0,1,\ldots,N]^n}$. The matrix $\mathbf{A}$ has its columns indexed by $\beta \in [0, 1, \ldots, N]^n$ and its rows indexed by $\alpha \in \mathscr{A}_{f_c}$, while $\mathbf{b} = (b_\alpha)_{\alpha\in\mathscr{A}_{f_c}}$ is the vector of black-box evaluations at the points $(\mathbf{z}_0^\alpha)$, $\alpha \in \mathscr{A}_{f_c}$. So

$$\mathbf{A}(\alpha, \beta) \;=\; (\mathbf{z}_0^\beta)^\alpha \;=\; z_{01}^{\beta_1\alpha_1} \cdots z_{0n}^{\beta_n\alpha_n}; \quad b_\alpha \;=\; g(\mathbf{z}_0^\alpha), \tag{3.28}$$

for all $\alpha \in \mathscr{A}_{f_c}$ and $\beta \in [0, 1, \ldots, N]^n$.

Proposition 3.6 *Under the conditions of Theorem 3.3, the LP $\rho = \min_{\mathbf{x}} \{ \|\mathbf{x}\|_1 : \; \mathbf{A}\,\mathbf{x} \;=\; \mathbf{b} \}$ with $\mathbf{A}$ and $\mathbf{b}$ as in (3.28) has a unique optimal solution which is the vector of coefficients of the polynomial g^* of Theorem 3.3.*

Proof Let $\mathbf{x} \in \mathbb{R}^{[0,\ldots,N]^n}$ be an admissible solution of the LP with $\mathbf{A}$ and $\mathbf{b}$ as in (3.28). Let $g \in \mathbb{R}[\mathbf{x}]$, $\mathbf{z} \mapsto g(\mathbf{z}) := \sum_\beta g_\beta \, \mathbf{z}^\beta$, be the polynomial with vector of coefficients

$$g_\beta \;=\; \mathbf{x}_\beta \; (= x_{\beta_1}, \ldots, x_{\beta_n}), \quad \forall \beta \in [0, \ldots, N]^n.$$

By construction one has $g(\mathbf{z}_0^\alpha) = a_\alpha = g^*(\mathbf{z}_0^\alpha)$ for all $\alpha \in \mathscr{A}_{f_c}$, where g^* is as in Theorem 3.3. The Borel measures ν^+ and ν^- on $\mathbb{T}^n$ defined by

$$\nu^+ := \sum_{\beta \in [0,\dots,N]^n} \max[0, x_\beta]\, \delta_{\mathbf{z}_0^\beta}; \quad \nu^- := \sum_{\beta \in [0,\dots,N]^n} -\min[0, x_\beta]\, \delta_{\mathbf{z}_0^\beta},$$

are a feasible solution of (3.16) and the Borel signed measure $\nu := \nu^+ - \nu^-$ satisfies $\|\nu\|_{TV} = \|\nu^+\| + \|\nu^-\|$. Hence $\|\nu\|_{TV} \geq \|\mu^*\|_{TV}$ where μ^* is the optimal solution of (3.14). So the optimal value ρ of the LP satisfies $\rho \geq \|\mu^*\|_{TV}$. On the other hand with g^* as in Theorem 3.3, let

$$\mathbf{x}^*_\beta \,(= x^*_{\beta_1}, \dots, x^*_{\beta_n}) \,:=\, g^*_\beta, \quad \forall \beta \in [0, \dots, N]^n.$$

Then $\|\mathbf{x}^*\|_1 = \|\mu^*\|_{TV} \leq \rho$ and so $\|\mathbf{x}^*\|_1 = \rho$, which proves that $\mathbf{x}^*$ is an optimal solution of the LP. Uniqueness follows from the uniqueness of solution to (3.14). □

4 Numerical experiments

In the problem of polynomial interpolation, we are not given a number of evaluations to begin with, i.e. f_c. Rather, we seek to recover a blackbox polynomial using the least number of evaluations. Thus, one could set $f_c = 1$, then compute a hierarchy of SDPs of order $d = 1, 2, \dots$. Next, set $f_c = 2$, and compute another hierarchy of order $d = 2, 3, \dots$. This leads to a hierarchy of hierarchies, which is costly from a computational perspective. Thus, we propose a single hierarchy where we choose to make all possible evaluations at each relaxation order. Therefore we have fixed $d = f_c$ in (3.22) and let f_c increase to see when we recover the desired optimal measure (polynomial g^*) of Theorem 3.3.[4]

In order to make a rigorous comparison with Prony's method, we use the same exact same procedure to extract the atomic measures from the output matrices of the semidefinite optimization as for Prony's method. For the super-resolution of order d, we use Prony with input measurements up to degree $2d$ (that way $d_1 = d_2 = d$ in Section 2.3.1) for each of the two Toeplitz matrices. In all numerical experiments, we use the threshold $\epsilon = 0.1$ for determining the rank of a matrix in its SVD decomposition. This threshold is also used to test the rank conditions (3.23)-(3.24).

4.1 Separation of the support

Initially, super resolution was concerned with signal processing where the measurements are given and fixed and we have no influence on them. In constrast, in the super resolution formulation of an polynomial interpolation problem, we can choose where we make the measurements, that is the points where we want to evaluate the blackbox polynomial. This can have a strong influence on the seperation condition which guarantees exact recovery on the signal (our blackbox polynomial). We illustrate this phenomenon on the following example. Suppose that we are looking for the blackbox polynomial

$$f(x) = 3x^{20} + x^{75} - 6x^{80} \tag{4.1}$$

whose degree we assume to be less than or equal to 100. We consider such a high degree in order to well illustrate the notion of the separation of the support. Below, we will consider more realistic polynomials, limited to degree 10. We now investigate two different ways of making evaluations and their impact on the separation of the support, crucial for super-resolution. Let us firstly evaluate the blackbox polynomial at the points

$$(e^{\frac{2\pi i}{101}})^0, (e^{\frac{2\pi i}{101}})^1, (e^{\frac{2\pi i}{101}})^2, \dots, (e^{\frac{2\pi i}{101}})^d \subset \mathbb{T} \tag{4.2}$$

at step d of the SDP hierarchy (i.e. $(\mathbf{P}_{d,d})$ in (3.22)). The proximity of points on the torus is thus directly related to the proximity of the exponents of the polynomial. It can be seen in the left part of Figure 2 that some of the point on the torus are very close to one another.

[4] In the univariate case, the optimal value of $(\mathbf{P}_{d,f_c})$ in (3.22) does not increase with d when $d > f_c$. Indeed, for any optimal solution of $(\mathbf{P}_{f_c,f_c})$, there exists a representing signed measure $\mu = \mu^+ - \mu^-$ on the torus. However, one may not be able to extract this measure.

Let us secondly evaluate in the blackbox polynomial at the points

$$(e^i)^0, (e^i)^1, (e^i)^2, \ldots, (e^i)^d \subset \mathbb{T} \tag{4.3}$$

at step d of the SDP hierarchy. The proximity of points on the torus is thus no longer related to the proximity of the exponents of the polynomial. It can be seen in the left part of Figure 3 that the points on the torus are nicely spread out. This is not guaranteed to be the case, but is expected to be true generically. In order to recover the blackbox polynomial once a candidate atomic measure is computed, we form a table of the integers $k = 1, \ldots, d$ modulo 2π. For each atom, we consider its argument and find the closest value in the table, yielding an integer k, i.e. the power of the monomial associated to the atom. The coefficient of the monomial is given by the weight of the atom.

We now provide numerical experiments. Table 1 and Table 2 show the optimal value and the number of atoms of the optimal measure μ at each order d. Graphical illustrations of the solutions appear in Figure 2 and Figure 3. The dual polynomials in the right hand of the figures illustrate why a higher degree is needed when the points are closer.

Table 1 Evaluation at roots of unity $e^{\frac{2ki\pi}{101}}$

Order $d = f_c$	$\|\mu\|_{TV}$	#supp(μ)
0	2.0000	1
1	7.6618	2
2	8.1253	3
3	8.3655	5
4	8.7240	7
5	8.9882	9
6	9.3433	11
7	9.5837	13
8	9.7993	17
9	9.9436	19
10	9.9978	20
11	10.0000	3

Table 2 Evaluation at $e^{k\mathbf{i}}$

Order $d = f_c$	$\|\mu\|_{TV}$	#supp(μ)
0	2.0000	1
1	8.7759	2
2	9.2803	3
3	10.0000	3

Remark 1 Before we move on to other examples, we note that naive LP with evaluations at random points on the real line requires about 50 evaluations on this example, compared with the 4 evaluations with super-resolution using multiple loops and in fact, the rigorous LP on the torus also requires 4 evaluations.

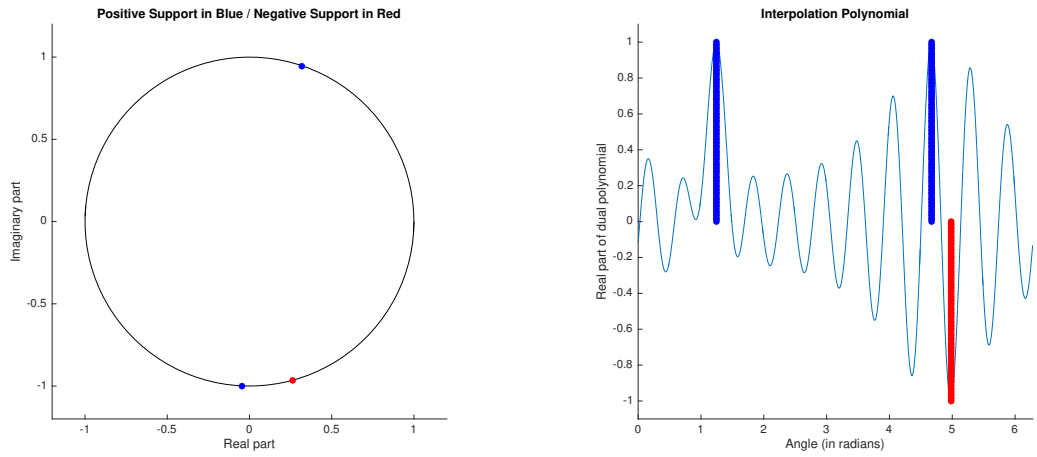

Fig. 2 Primal-dual solution of super-resolution at order 11 (using single loop)

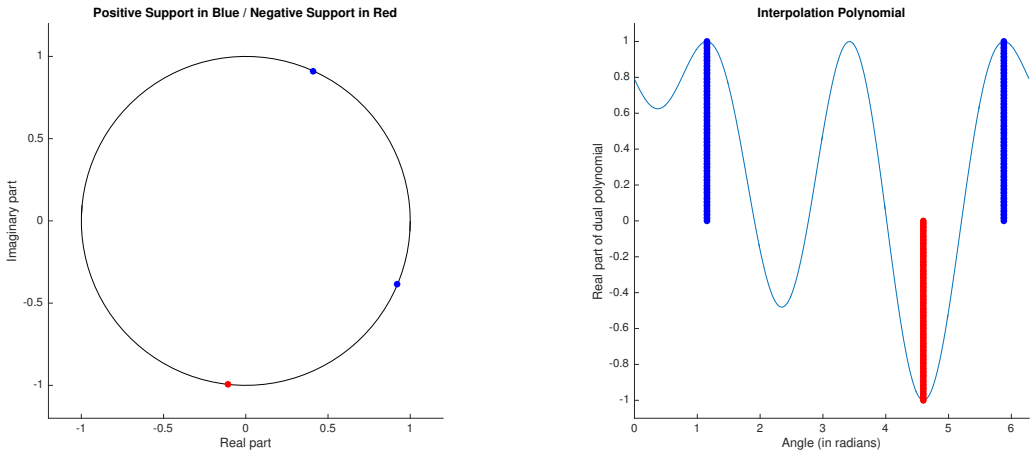

Fig. 3 Primal-dual solution of super-resolution at order 3 (using multiple loops)

4.2 Methodology for comparison

Our methodology for comparing the various approaches is as follows.

1. **Generation of the examples:** We define a random set of ten sparse polynomials with up to ten variables and up to degree ten (first column of Table 3). We believe that polynomials of higher degree are not realistic and are rarely used in numerical computations. For example, for a polynomial of n variables and k atoms, we generate the exponents β of the k monomials $\mathbf{x}^\beta$ randomly from $\mathbb{N}^n_d := \{\beta \in \mathbb{N}^n \mid \sum_{i=1}^n \beta_i \leqslant d\}$ and the associated non-zero coefficients g_β are drawns from a uniform distribution in the interval $[-10, 10]$.

2. **Results in the noiseless case:** We detect the minimum number of evaluations for each approach to recover the blackbox polynomial in the noiseless case and report the results in Table 3. We use evaluations at the points $e^{\alpha \mathbf{i}} = (e^{\mathbf{i}\alpha_1}, \ldots, e^{\mathbf{i}\alpha_n})$ with $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{Z}^n$ up to a certain degree $\sum_{k=1}^n |\alpha_k| \leqslant d$.The corresponding number of evaluations and degree (d) are reported in the columns *Rigorous LP*, *super-resolution*, and *Toeplitz Prony* of Table 3. In *Advanced T. Prony*, evaluations are made at different points. Thus, only the first three columns of Table 3 can be compared in presence of noise.

3. **Results in the presence of noise:** For each of the ten polynomials in the list of examples, we determine the maximum degree $d^{\max}$ for the evaluations $g(e^{i\alpha_1}, \ldots, e^{i\alpha_n})$ with $\alpha_1, \ldots, \alpha_n \in \mathbb{Z}$ and $\sum_{i=1}^n |\alpha_k| \leq d^{\max}$, among *Rigorous LP*, *super-resolution*, and *Toeplitz Prony* in Table 3. For example, for the first line of Table 3, that number is $d^{\max} := 2$ which corresponds to 3 evaluations in this univariate problem. As a result, we know that for these evaluations all three approaches return the correct sparse polynomial. We then add uniform noise to those evaluations, i.e.

$$g(e^{\mathbf{i}\alpha_1}, \ldots, e^{\mathbf{i}\alpha_n}) + \epsilon_\alpha \ , \quad \epsilon_\alpha \in \mathbb{C}, \quad \Re\epsilon_\alpha, \ \Im\epsilon_\alpha \in [-0.1, +0.1] \tag{4.4}$$

for all $|\alpha_1| + \ldots + |\alpha_n| \leqslant d^{\max}$ and $\alpha_1, \ldots, \alpha_n \in \mathbb{Z}$. Next, we run each approach ten times (with new noise every time) and report the average error in Table 4. The error is defined as the relative error in percentage of the output polynomial $\hat{g}(\mathbf{x}) = \sum_\alpha \hat{g}_\alpha \mathbf{x}^\alpha$ compared with the blackbox polynomial $g(\mathbf{x}) = \sum_\alpha g_\alpha \mathbf{x}^\alpha$ using the 2 norm of the coefficients, i.e.

$$100 \times \frac{\sqrt{\sum_\alpha (\hat{g}_\alpha - g_\alpha)^2}}{\sqrt{\sum_\alpha g_\alpha^2}}. \tag{4.5}$$

Note that in *Rigorous LP* and *super-resolution* the equalities associated to the evaluations are relaxed to inequalities, a functionality which is not possible in *Toeplitz Prony*. This allows for more robutness. Precisely, in *Rigorous LP*, we replace $\mathbf{Ax} - \mathbf{b} = 0$ by $-0.1 \leqslant \Re(\mathbf{Ax} - \mathbf{b}) \leqslant 0.1$ and

$-0.1 \leqslant \Im(\mathbf{Ax}-\mathbf{b}) \leqslant 0.1$, while in *super-resolution* we use a 2-norm ball of radius $0.1 \times \sqrt{2}$ (similar to the technique employed in [7]).

Blackbox Polynomial	Rigorous LP	Super Resolution	Toeplitz Prony	Advanced T. Prony
$-1.2x^4 + 6.7x^7$	2 (1)	3 (2)	3 (2)	3
$2.3x^6 + 5.6x^3 - 1.5x^2$	4 (3)	5 (4)	4 (3)	4
$-2.1x^3 + 5.4x^2 - 2.0x + 6.2x^5 - 5.2$	5 (4)	6 (5)	6 (5)	6
$0.8x_1x_2 - x_1x_2^2$	19 (3)	31 (4)	10 (2)	6
$-5.8x_1^2x_2^2 - 8.2x_1^2x_2^3 + 5.5x_1^3x_2 + 1.1$	10 (2)	19 (3)	19 (3)	13
$-7.2x_1x_2^2 + 1.8x_1^3x_2^2 + 2.6x_1^4x_2^5 + 6.2x_1x_2^5 + 2.5x_1$	10 (2)	19 (3)	19 (3)	14
$-3.5 + 8.1x_1^3x_2x_3$	7 (1)	28 (2)	28 (2)	9
$-1.2x_1^2x_2^2x_3^3 + 7.3x_1^2x_2 - 2.4x_2$	28 (2)	28 (2)	28 (2)	16
$-6.1x_1^2x_5 + 2.5x_2x_4 + 4.8x_3$	136 (2)	136 (2)	136 (2)	30
$2.9x_2x_3x_9^4x_{10} - 5.6x_1x_4^2x_7 - 4.1x_3x_5x_6^3x_8$	N. A.	1595 (2)	1595 (2)	65

Table 3 Minimum number of evaluations and degrees without noise (evaluations in the points $(e^{\mathbf{i}\alpha_1}, \ldots, e^{\mathbf{i}\alpha_n})$ for $|\alpha_1| + \ldots + |\alpha_n| \leqslant d$ and $\alpha_1, \ldots, \alpha_n \in \mathbb{Z}$ for the first three columns)

Blackbox Polynomial	Rigorous LP	Super Resolution	Toeplitz Prony
$-1.2x^4 + 6.7x^7$	4.18%	1.58%	0.61%
$2.3x^6 + 5.6x^3 - 1.5x^2$	1.94%	1.81%	0.85%
$-2.1x^3 + 5.4x^2 - 2.0x + 6.2x^5 - 5.2$	1.47%	1.40%	0.69%
$0.8x_1x_2 - x_1x_2^2$	3.23%	4.84%	2.26%
$-5.8x_1^2x_2^2 - 8.2x_1^2x_2^3 + 5.5x_1^3x_2 + 1.1$	1.13%	0.87%	1.29%
$-7.2x_1x_2^2 + 1.8x_1^3x_2^2 + 2.6x_1^4x_2^5 + 6.2x_1x_2^5 + 2.5x_1$	1.23%	1.08%	6.28%
$-3.5 + 8.1x_1^3x_2x_3$	0.79%	0.70%	0.50%
$-1.2x_1^2x_2^2x_3^3 + 7.3x_1^2x_2 - 2.4x_2$	2.19%	1.03%	1.39%
$-6.1x_1^2x_5 + 2.5x_2x_4 + 4.8x_3$	0.94%	1.15%	1.04%
$2.9x_2x_3x_9^4x_{10} - 5.6x_1x_4^2x_7 - 4.1x_3x_5x_6^3x_8$	N. A.	0.47%	0.46%

Table 4 Relative error in percentage with uniform noise between -0.1 and 0.1 for the real and imaginary parts on the measurements (evaluations in the points $(e^{i\alpha_1}, \ldots, e^{i\alpha_n})$ for $|\alpha_1| + \ldots + |\alpha_n| \leqslant d$ and $\alpha_1, \ldots, \alpha_n \in \mathbb{Z}$).

The *Advanced T. Prony* column of Table 3: the first r exponents $\alpha \in \mathbb{N}^n$ are chosen (where r is the number of monomials in the blackbox polynomias) for the monomials indexing the rows and the first r exponents $-\alpha$ with $\alpha \in \mathbb{N}^n$ are chosen for indexing the columns of the Toeplitz matrix. Since $g(e^{-\mathbf{i}\alpha}) = \overline{g(e^{\mathbf{i}\alpha})}$, the number of evaluations does not include the conjugate of known values of g. The number of monomials r is unknown but one could use *Advanced T. Prony* with $r = 1, 2, \ldots$ successively. We only report the result when setting r to the number of monomials in the blackbox polynomial. Note that in the other approaches in Table 3, we do *not* assume that the number of monomials is known. Same goes in the presence of noise.

4.3 Discussion

Disclaimer: In the sequel, we discuss various advantages and drawbacks of the three methods and of course the resulting conclusions should be interpreted with care as they are biased by the examples that we have considered.

• In the noiseless case, Rigorous LP generally requires the least number of evaluations compared with super-resolution and Toeplitz Prony as can be seen in Table 3. This is a remarkable situation where the compressed sensing approach (Rigorous LP) is guaranteed to recover the polynomial even if the RIP property does not hold. The classical result on Prony's method is that the number of evaluations to recover the blackbox polynomial is equal to twice the number of monomials in the blackbox polynomial (in the univariate case). Toeplitz Prony goes further: the number of evaluations is equal to the number of monomials plus one (in the univariate case, as explained in Section 2.3.2). For example, the third example in Table 3 requires 6 evaluations and is composed of 5 monomials.

• In terms of certification, in principle, super-resolution has to be applied with enough points ($\geqslant 128$ for $n = 1$, $\geqslant 512$ for $n = 2$ and more if the separation between the points is small [7]) to guarantee the existence of a dual certificate polynomial. Moreover, in the multivariate case, no bound on the order of the SDP relaxation is known to guarantee that the flat extension property is satisfied (rank conditions (3.23)-(3.24))[5]. In contrast, Toeplitz Prony requires evaluations at points $\alpha \in \mathbb{Z}^n$ with $|\alpha| \leq r$ where r is at most the number of monomials, in order to recover the decomposition of the sparse polynomial. In practice, the experimentations show that a small number of evaluations is sufficient to compute the decomposition in both methods.

• In terms of computational burden, among Rigorous LP, super-resolution, and Toeplitz Prony, the cheapest approach is Toeplitz Prony since it requires only two linear algbebra operations on matrices of size dependent on the number of monomials in the blackbox polynomials. super-resolution entails a heavy computational burden with the semidefinite optimization. Rigorous LP requires the longest setup time because a variable has to be created for each potential monomial in the blackbox polynomial, unlike the two other approaches. In particular, the setup time is too long on a standard laptop for the example with 10 variables (hence N. A. in Table 3 and Table 4). Howevever, after the setup step has been performed, computing the LP is fast and reliable.

• Concerning noise, it seems that the three methods perform more and less equally well even with the relatively large noise level that we have selected, namely 0.1 error on the evaluations. This is little bit surprising for the Prony method because it seems to be commonly admitted that Prony is not very robust to noise. This surprising relative robustness may be due to the large threshold $\epsilon = 0.1$ allowed in the rank determination of the SVD decomposition. Indeed, if we select a smaller threshold, we observe degradation of the results for Prony (and super resolution which relies on Prony for the extraction step after the optimization step). See table 5 below.

Blackbox Polynomial	Super Resolution				Toeplitz Prony			
	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-1}	10^{-2}	10^{-3}	10^{-4}
$-1.2x^4 + 6.7x^7$	1.58%	1.58%	1.58%	1.58%	0.79%	0.79%	0.79%	0.79%
$2.3x^6 + 5.6x^3 - 1.5x^2$	1.80%	1.80%	1.80%	1.80%	0.91%	1.15%	1.15%	1.15%
$-2.1x^3 + 5.4x^2 - 2.0x + 6.2x^5 - 5.2$	1.42%	1.44%	1.44%	1.44%	0.68%	0.68%	0.68%	0.68%
$0.8x_1x_2 - x_1x_2^2$	4.69%	17.25%	12.49%	25.50%	5.35%	137.90%	165.79%	226.29%
$-5.8x_1^2x_2^2 - 8.2x_1^2x_2^3 + 5.5x_1^3x_2 + 1.1$	0.93%	2.10%	42.69%	47.50%	1.22%	60.36%	53.86%	42.47%
$-7.2x_1x_2^2 + 1.8x_1^3x_2^2 + 2.6x_1^4x_2^5 + \ldots$	1.00%	17.56%	43.39%	67.55%	8.45%	89.45%	45.82%	33.87%
$-3.5 + 8.1x_1^3x_2x_3$	0.78%	0.78%	72.17%	62.02%	0.41%	96.50%	77.42%	83.27%
$-1.2x_1^2x_2^2x_3^3 + 7.3x_1^2x_2 - 2.4x_2$	1.09%	18.55%	86.55%	83.45%	3.59%	66.02%	39.77%	55.15%
$-6.1x_1^2x_5 + 2.5x_2x_4 + 4.8x_3$	0.85%	24.60%	105.84%	140.05%	1.68%	130.39%	86.77%	79.96%
$2.9x_2x_3x_9^4x_{10} - 5.6x_1x_4^2x_7 + \ldots$	0.54%	0.54%	136.20%	146.93%	6.45%	251.17%	119.65%	257.64%

Table 5 Same experiments as in Table 4 but with four different values of the rank threshold $\epsilon = 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}$.

5 Another efficient (*a priori* heuristic) approach

In sequel we propose still use the super-resolution hierarchy (3.22) but now by restricting the evaluations at points $\mathbf{z}_0^\alpha$ with

$$\alpha \in \mathscr{A}_{f_c}^2 := \{\alpha \in \mathbb{N}^n \mid \|\alpha\|_1 \leqslant f_c\} \subset \mathscr{A}_{f_c}^1 \subset \mathscr{A}_{f_c}. \tag{5.1}$$

This is illustrated in Figure 4. This restriction is first inspired by the fact that Hankel Prony descibed in Section 2.3.1 is guaranteed to work using only those evaluations. There is another more general inspiration coming from two mathematical results.

We provide a result valid in full generality for atomic measures (with finitely many atoms) which indeed suggests that in practice it may suffice to make evaluations at $\alpha \in \mathbb{N}$ (instead of $\alpha \in \mathbb{Z}$). The resulting semidefinite programs have Toeplitz matrices of the same dimension but include much less linear moment constraints. With 10 variables, the first order semidefinite program of the hierarchy entails matrix variables of size 11×11 and *only* 11 linear equalities (instead of 56)! The second order

[5] In a few cases where (3.23)-(3.24) are not satisfied, we are still able to a recover polynomial using the algorithm in Section 2.3.1.

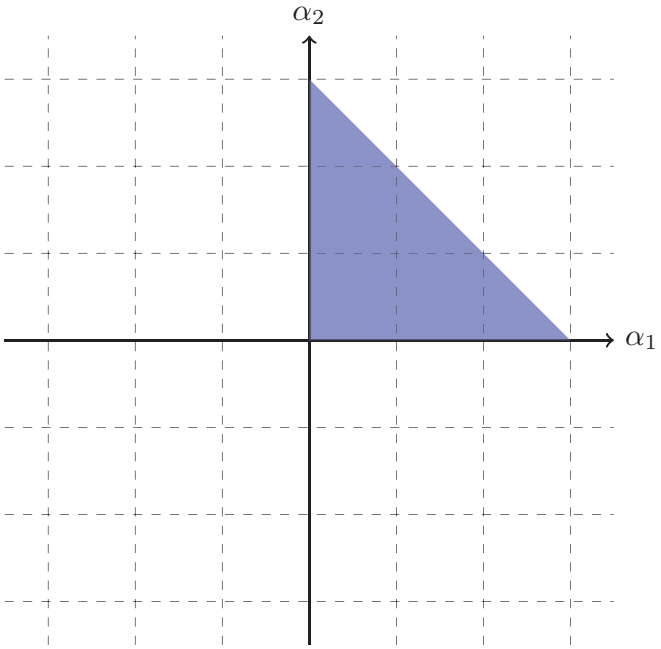

Fig. 4 Evaluations at α with $\alpha_1 + \alpha_2 \leqslant 3$ and $\alpha_1, \alpha_2 \in \mathbb{N}$.

relaxation entails matrix variables of size 66×66 and 66 linear equalities (instead of 1,596), and the third relaxation entails matrix variables of size 286×286 and 286 linear equalities (instead of 21,691), and so on. We first remind the reader of a well-known result.

Proposition 5.1 (Consequence of Stone-Weiestrass) *Let $(y_{\alpha,\beta})_{\alpha,\beta\in\mathbb{N}^n}$ denote a multi-indexed sequence of complex numbers and let $K \subset \mathbb{C}^n$ denote a compact set. If there exists a complex-valued finite Borel measure μ supported on K such that*

$$y_{\alpha,\beta} = \int_K z^\alpha \bar{z}^\beta d\mu \ , \quad \forall \alpha, \beta \in \mathbb{N}^n, \tag{5.2}$$

then μ is the unique complex-valued finite Borel measure to satisfy (5.2).

Proof Consider another such measure $\hat{\mu}$. Then

$$\int_K z^\alpha \bar{z}^\beta d(\mu - \hat{\mu}) = 0 \ , \quad \forall \alpha, \beta \in \mathbb{N}^n. \tag{5.3}$$

Thanks to the complex Stone-Weiestrass Theorem,

$$\int_{\mathbb{C}^n} \varphi d(\mu - \hat{\mu}) = 0 \ , \quad \forall \alpha, \beta \in \mathbb{N}^n. \tag{5.4}$$

for all function $\varphi : \mathbb{C}^n \longrightarrow \mathbb{C}$ continuous with respect to the sup-norm $\|\varphi\|_\infty := \sup_{z\in K} |\varphi(z)|$. Therefore $\mu = \hat{\mu}$.

In practice, whether it be interpolation or optimization, we are generally interested in *atomic* measures with finitely many atoms (in short, atomic measures in the sequel). The next result establishes that for such atomic measures we do not have to care about conjugates, which in view of Proposition 5.1, we find somewhat counter-intuitive.

Lemma 1 *Let $(y_\alpha)_{\alpha\in\mathbb{N}^n}$ denote a multi-indexed sequence of complex numbers. If there exists an atomic complex-valued measure μ such that*

$$y_\alpha = \int_{\mathbb{C}^n} z^\alpha d\mu \ , \quad \forall \alpha \in \mathbb{N}^n, \tag{5.5}$$

then μ is the unique atomic measure to satisfy (5.5).

Proof Let us write the measure μ as

$$\mu = \sum_{k=1}^{d} w_k \delta_{a(k)} \tag{5.6}$$

where $d \in \mathbb{N}$, and $w_1, \ldots, w_d \in \mathbb{C} \setminus \{0\}$, and $a(1), \ldots, a(d) \in \mathbb{C}^n$.

Consider another atomic measure $\hat{\mu}$ that satisfies (5.5), of the form

$$\hat{\mu} = \sum_{k=1}^{\hat{d}} \hat{w}_k \delta_{\hat{a}(k)} \tag{5.7}$$

where $\hat{d} \in \mathbb{N}$, and $\hat{w}_1, \ldots, \hat{w}_{\hat{d}} \in \mathbb{C} \setminus \{0\}$, and $\hat{a}(1), \ldots, \hat{a}(\hat{d}) \in \mathbb{C}^n$.

Consider the following truncated Hankel matrix

$$H_k(y) = (y_{\alpha+\beta})_{|\alpha|,|\beta| \leqslant k} \tag{5.8}$$

where $|\alpha| := \alpha_1 + \ldots + \alpha_n$. Thanks to Lemma 2, its rank is equal to d when $k \geqslant d-1$ and it is equal to $\hat{d}$ when $k \geqslant \hat{d}-1$. Thus $d = \hat{d}$. Moreover, when $k \geqslant d = \hat{d}$, Lemma 3 implies that

$$\text{span}\{\ v_k(a(1))\ ,\ \ldots\ ,\ v_k(a(d))\ \} \quad = \quad \text{span}\{\ v_k(\hat{a}(1))\ ,\ \ldots\ ,\ v_k(\hat{a}(d))\ \} \tag{5.9}$$

where $v_k(z) = (z^\alpha)_{|\alpha| \leqslant k}$.

We now reason by contradiction. Assume that one of the atoms of $\hat{\mu}$, say $\hat{a}(1)$, is distinct from the atoms of μ. Hence $\hat{a}(1), a(1), \ldots, a(d)$ are $d+1$ distinct points of $\mathbb{C}^n$. Lemma 2 implies that

$$v_k(\hat{a}(1))\ ,\ v_k(a(1))\ ,\ \ldots\ ,\ v_k(a(d)) \tag{5.10}$$

are linearly independent vectors if $k \geqslant d$. This contradicts equation (5.9). The atoms of μ and $\hat{\mu}$ thus coincide. Their weights satisfy

$$(w_1 - \hat{w}_1)v_k(a(1)) \ + \ \ldots \ + \ (w_d - \hat{w}_d)v_k(a(d)) = 0. \tag{5.11}$$

Again, thanks to Lemma 2, the vectors are linearly independent if $k \geqslant d-1$, thus $w_1 - \hat{w}_1 = \ldots = w_d - \hat{w}_d = 0$. This terminates the proof.

Numerical experiments:

Below, we replicate the experiments of Section 4.2 (with and without noise) but now we make evaluations in $\alpha_1, \ldots, \alpha_n \in \mathbb{N}$ instead of $\alpha_1, \ldots, \alpha_n \in \mathbb{Z}$.

Blackbox Polynomial	Rigorous LP	Super Resolution	Hankel Prony	Advanced H. Prony
$-1.2x^4 + 6.7x^7$	2 (1)	3 (2)	4 (3)	4
$2.3x^6 + 5.6x^3 - 1.5x^2$	4 (3)	5 (4)	6 (5)	6
$-2.1x^3 + 5.4x^2 - 2.0x + 6.2x^5 - 5.2$	5 (4)	6 (5)	10 (9)	10
$0.8x_1x_2 - x_1x_2^2$	10 (3)	15 (4)	10 (3)	7
$-5.8x_1^2x_2^2 - 8.2x_1^2x_2^3 + 5.5x_1^3x_2 + 1.1$	10 (3)	15 (4)	21 (5)	15
$-7.2x_1x_2^2 + 1.8x_1^3x_2^2 + 2.6x_1^4x_2^5 + 6.2x_1x_2^5 + 2.5x_1$	10 (3)	15 (4)	21 (5)	18
$-3.5 + 8.1x_1^3x_2x_3$	10 (2)	10 (2)	20 (3)	10
$-1.2x_1^2x_2^2x_3^3 + 7.3x_1^2x_2 - 2.4x_2$	20 (3)	20 (3)	20 (3)	16
$-6.1x_1^2x_5 + 2.5x_2x_4 + 4.8x_3$	21 (2)	21 (2)	56 (3)	28
$2.9x_2x_3x_9^4x_{10} - 5.6x_1x_4^2x_7 - 4.1x_3x_5x_6^3x_8$	66 (2)	66 (2)	286 (3)	58

Table 6 Minimum number of evaluations and degrees without noise (evaluations in the points $(e^{\mathbf{i}\alpha_1}, \ldots, e^{\mathbf{i}\alpha_n})$ for $\alpha_1 + \ldots + \alpha_n \leqslant d$ and $\alpha_1, \ldots, \alpha_n \in \mathbb{N}$ for the first three columns)

In Table 6, in the column *Advanced H. Prony*, the first r exponents $\alpha \in \mathbb{N}^n$ are chosen for the monomials indexing the rows and columns of the Hankel matrix, where r is the number of terms in the blackbox polynomial g. As in Section 4.2, for the first three columns of Table 6 and Table 7, we do

Blackbox Polynomial	Rigorous LP	Super Resolution	Hankel Prony
$-1.2x^4 + 6.7x^5$	2.32%	1.66%	0.97%
$2.3x^6 + 5.6x^3 - 1.5x^2$	1.71%	2.31%	3.33%
$-2.1x^3 + 5.4x^2 - 2.0x + 6.2x^5 - 5.2$	0.80%	1.64%	2.89%
$0.8x_1x_2 - x_1x_2^2$	14.91%	11.03%	52.14%
$-5.8x_1^2x_2^2 - 8.2x_1^2x_2^3 + 5.5x_1^3x_2 + 1.1$	0.73%	1.01%	2.13%
$-7.2x_1x_2^2 + 1.8x_1^3x_2^2 + 2.6x_1^4x_2^5 + 6.2x_1x_2^5 + 2.5x_1$	1.19%	12.30%	2.67%
$-3.5 + 8.1x_1^3x_2x_3$	0.82%	1.32%	0.93%
$-1.2x_1^2x_2^2x_3^3 + 7.3x_1^2x_2 - 2.4x_2$	3.29%	2.13%	16.99%
$-6.1x_1^2x_5 + 2.5x_2x_4 + 4.8x_3$	2.90%	1.64%	6.74%
$2.9x_2x_3x_9^4x_{10} - 5.6x_1x_4^2x_7 - 4.1x_3x_5x_6^3x_8$	107.87% (1) N.A. (2) N.A. (3)	161.36% (1) 2.12% (2) N.A. (3)	134.87% (1) 134.69% (2) 0.57% (3)

Table 7 Relative error in percentage with uniform noise between -0.1 and 0.1 for the real and imaginary parts on the measurements (evaluations in the points $(e^{i\alpha_1}, \ldots, e^{i\alpha_n})$ for $\alpha_1 + \ldots + \alpha_n \leqslant d$ and $\alpha_1, \ldots, \alpha_n \in \mathbb{N}$).

not assume anything to be known about the blackbox polynomial expect for the number of variables and an upper bound on the degree (i.e. 10).

In the presence of noise, the optimization step of super-resolution (before the second step of extraction) seems to behave as an efficient filter as it indeed reduces the error compared with Hankel Prony in 7 out of the 9 comparable instances of Table 7. However, sometimes, the semidefinite program does not provide a good output. Indeed, in the sixth example, among the ten trials there are two trials where the solver runs into numerical issues, which explains the large error of 12.30%.

6 Conclusion

We have addressed the sparse polynomial interpolation problem with three different approaches: compressed sensing, super resolution, and Prony's method. The common denominator of the three approaches is our view of a polynomial as a signed atomic measure where the atoms correspond to monomials and the weights to coefficients. Then, on the one hand we can invoke directly results from (discrete) super-resolution theory à la Candès & Fernandez-Granda [7] to show that the unknown black box polynomial is the unique solution of a certain LP on a measure space and also the unique solution of finite-dimensional linear program. On the other hand, invoking Kunis et al. [24] Prony's method can also be applied. To the best of our knowledge this unifying view of sparse interpolation is new and makes the numerical comparison of the three methods very natural. In our preliminary numerical experiments :

- Prony's method works well and better than expected in the presence of noise.
- Super-resolution acts in two steps: a first optimization step and then an extraction procedure applied to the optimal solution. The latter step is nothing less than Prony's method. We find that this optimization step sometimes helps significantly in the presence of noise.
- LP-Compressed sensing also works well but its set-up time is quite limiting.

Acknowledgement

The work of the first two authors was funded by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement 666981 TAMING).

Appendix

Lemma 2 *If $z^{(1)}, \ldots, z^{(d)}$ are distinct points of $\mathbb{C}^n$, then $v_{d-1}(z^{(1)}), \ldots, v_{d-1}(z^{(d)})$ are linearly independent vectors, where $v_d(z) := (z^\alpha)_{|\alpha \leqslant d}$.*

Proof Consider some complex numbers $c_1, \ldots, c_d$ such that

$$\sum_{k=1}^{d} c_k (z^{(k)})^\alpha = 0 \ , \quad \forall |\alpha| \leqslant d-1. \tag{6.1}$$

Given $1 \leqslant l \leqslant d$, define the Lagrange interpolation polynomial

$$L^{(l)}(z) := \prod_{\substack{1 \leqslant k \leqslant d \\ k \neq l}} \frac{z_{i(k)} - z^{(k)}_{i(k)}}{z^{(l)}_{i(k)} - z^{(k)}_{i(k)}} \tag{6.2}$$

where $i(k) \in \{1, \dots, n\}$ is an index such that $z^{(k)}_{i(k)} \neq z^{(l)}_{i(k)}$. It satisfies $L^{(l)}(z^{(k)}) = 1$ if $k = l$ and $L^{(l)}(z^{(k)}) = 0$ if $k \neq l$. The degree of $L^{(l)}(z) =: \sum_\alpha L^{(l)}_\alpha z^\alpha$ is equal to $d-1$. Thus we may multiply the equation in (6.1) by $L^{(l)}_\alpha$ to obtain

$$\sum_{k=1}^{d} c_k \; L^{(l)}_\alpha (z^{(k)})^\alpha = 0 \; , \quad \forall |\alpha| \leqslant d-1. \tag{6.3}$$

Summing over all $|\alpha| \leqslant d-1$ yields $\sum_{k=1}^d c_k \; L^{(l)}(z^{(k)}) = c_l = 0$.

Lemma 3 *If $u_1, \dots, u_d \in \mathbb{C}^n$ are linearly independent, and $c_1, \dots, c_d \in \mathbb{C}\backslash\{0\}$, then $\mathcal{R}(\sum_{i=1}^d c_i u_i u_i^T) = \mathcal{R}(\sum_{i=1}^d c_i u_i u_i^*) = \text{span}\{u_1, \dots, u_d\}$ where $\mathcal{R}$ denotes the range.*

Proof If $z \in \mathbb{C}^n$, then $(\sum_{i=1}^d c_i u_i u_i^T)z = \sum_{i=1}^d (c_i u_i^T z) u_i \in \text{span}\{u_1, \dots, u_d\}$ and $(\sum_{i=1}^d c_i u_i u_i^*)z = \sum_{i=1}^d (c_i u_i^* z) u_i \in \text{span}\{u_1, \dots, u_d\}$. Conversly, an element of the span $\sum_{i=1}^d \lambda_i u_i$ with $\lambda_1, \dots, \lambda_n \in \mathbb{C}$ belongs to the the range of $\sum_{i=1}^d c_i u_i u_i^T$ if there exists $z \in \mathbb{C}^n$ such that

$$\sum_{i=1}^{d} \lambda_i u_i = \left(\sum_{i=1}^{d} c_i u_i u_i^T \right) z$$

which is equivalent to each of the next three lines:

$$\sum_{i=1}^{d} [\lambda_i - (c_i u_i^T z)] u_i = 0, \tag{6.4}$$

$$\lambda_i = (c_i u_i)^T z \; , \; i = 1, \dots, d, \tag{6.5}$$

$$\lambda = (c_1 u_1 \dots c_d u_d)^T z. \tag{6.6}$$

Since $(c_1 u_1 \dots c_d u_d) \in \mathbb{C}^{n \times d}$ has rank d, its transpose has rank d. Thus there exists a desired $z \in \mathbb{C}^n$. Likewise, $\sum_{i=1}^d \lambda_i u_i$ belongs to the the range of $\sum_{i=1}^d c_i u_i u_i^*$ if there exists $z \in \mathbb{C}^n$ such that

$$\lambda_i = (c_i u_i)^* z \; , \; i = 1, \dots, d.$$

Since $(c_1 u_1 \dots c_d u_d) \in \mathbb{C}^{n \times p}$ has rank d, its conjugate transpose has rank d. Thus there exists a desired $z \in \mathbb{C}^n$.

References

1. Jean-Marc Azaïs, Yohann de Castro, and Fabrice Gamboa. Spike detection from inaccurate samplings. *Applied and Computational Harmonic Analysis*, 38(2):177–195, 2015.
2. Michael Ben-Or and Prasoon Tiwari. A deterministic algorithm for sparse multivariate polynomial interpolation. In *Proceedings of the Twentieth Annual ACM Symposium on Theory of Computing*, pages 301–309. ACM, 1988.
3. Elwyn R. Berlekamp. Nonbinary BCH decoding. *IEEE Transactions on Information Theory*, 14(2):242–242, 1968.
4. Gregory Beylkin and Lucas Monzón. On approximation of functions by exponential sums. *Applied and Computational Harmonic Analysis*, 19(1):17–48, 2005.
5. Emmanuel J. Candès. The restricted isometry property and its implications for compressed sensing *C.R. Acad. Sci. Paris Ser. I*, 346: 589–592, 2008.
6. Emmanuel J. Candès and Carlos Fernandez-Granda. Super-Resolution from Noisy Data. *Journal of Fourier Analysis and Applications*, 19(6):1229–1254, 2013.
7. Emmanuel J. Candès and Carlos Fernandez-Granda. Towards a Mathematical Theory of Super-resolution. *Communications on Pure and Applied Mathematics*, 67(6):906–956, 2014.
8. Emmanuel J. Candès and T. Tao. Decoding by linear programming. *IEEE Trans. Inform. Theory*, 51(12): 4203–4215, 2005. 2014.

9. Emmanuel J. Candès, J. Romberg and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics*, 59(8):1207–1223, 2006.
10. R. Curto and L. Fialkow. Truncated K-Moment Problems in Several Variables *J. Operator Theory*, 54 (2005), pp. 189–226.
11. Gaspard Riche de Baron de Prony. Essai experimental et analytique: Sur les lois de la dilatabilité de fluides élastique et sur celles de la force expansive de la vapeur de l'alcool, à différentes temperatures. *J. Ecole Polyt.*, 1:24–76, 1795.
12. Annie Cuyt and Wen-shin Lee. Sparse Interpolation and Rational Approximation. In Douglas Hardin, Doron Lubinsky, and Brian Simanek, editors, *Contemporary Mathematics*, volume 661, pages 229–242. American Mathematical Society, Providence, Rhode Island, 2016.
13. Y. De Castro, F. Gamboa, D. Henrion, and J.-B. Lasserre. Exact solutions to super-resolution on semi-algebraic domains in higher dimensions. *IEEE Trans. Inform. Theory*, 63:621–630, 2017.
14. Vincent Duval and Gabriel Peyré. Exact support recovery for sparse spikes deconvolution. *Foundations of Computational Mathematics*, 15(5):1315–1355, 2015.
15. Ya Yu Fan and C. Kamath. A comparison of Compressed Sensing and Sparse Recovery Algorithms Applied to Simulation Data. *Stat. Optim. Inform. Computing*, 4:194–213, 2016.
16. Mark Giesbrecht, George Labahn, and Wen-shin Lee. Symbolic–numeric sparse interpolation of multivariate polynomials. *Journal of Symbolic Computation*, 44(8):943–959, August 2009.
17. Gene Golub and Victor Pereyra. Separable nonlinear least squares: The variable projection method and its applications. *Inverse Problems*, 19(2):R1–R26, 2003.
18. Dima Yu Grigoriev, Marek Karpinski, and Michael F. Singer. Fast parallel algorithms for sparse multivariate polynomial interpolation over finite fields. *SIAM Journal on Computing*, 19(6):1059–1063, 1990.
19. J. Harmouch, H. Khalil, and B. Mourrain. Structured low rank decomposition of multivariate Hankel matrices. *Linear Algebra and its Applications*, April 2017.
20. Cédric Josz and Daniel K. Molzahn. Large Scale Complex Polynomial Optimization. `arXiv:1508.02068`
21. Erich Kaltofen and Yagati N. Lakshman. Improved Sparse Multivariate Polynomial Interpolation Algorithms. In *Proceedings of the International Symposium ISSAC'88 on Symbolic and Algebraic Computation*, ISSAC '88, pages 467–474, London, UK, 1989. Springer-Verlag.
22. Erich L. Kaltofen, Wen-shin Lee, and Zhengfeng Yang. Fast estimates of Hankel matrix condition numbers and numeric sparse interpolation. pages 130–136. ACM Press, 2011.
23. Nikolai N Krasovskii. *Theory of Motion Control*. Nauka, Moscow, 1968. (in Russian).
24. Stefan Kunis, Thomas Peter, Tim Römer, and Ulrich von der Ohe. A multivariate generalization of Prony's method. *Linear Algebra and its Applications*, 490:31–47, 2016.
25. Monique Laurent and Bernard Mourrain. A generalized flat extension theorem for moment matrices. *Archiv der Mathematik*, 93(1):87–98, 2009.
26. James Massey. Shift-register synthesis and BCH decoding. *IEEE transactions on Information Theory*, 15(1):122–127, 1969.
27. Bernard Mourrain. Polynomial-exponential decomposition from moments, 2016. hal-01367730, arXiv:1609.05720.
28. Lucien W. Neustadt. Optimization, a Moment Problem, and Nonlinear Programming. *Journal of the Society for Industrial and Applied Mathematics Series A Control*, 2(1):33–53, 1964.
29. Jiawang Nie. Optimality Conditions and Finite Convergence of Lasserre's Hierarchy *Math. Program. Ser. A*, 146(1-2): 97–121, 2014.
30. V. Pereyra and G. Scherer, editors. *Exponential Data Fitting and Its Applications*. Bentham Science Publishers, 2012.
31. Daniel Potts and Manfred Tasche. Nonlinear approximation by sums of nonincreasing exponentials. *Applicable Analysis*, 90(3-4):609–626, 2011.
32. Richard Roy and Thomas Kailath. ESPRIT-estimation of signal parameters via rotational invariance techniques. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 37(7):984–995, 1989.
33. Tomas Sauer. Prony's method in several variables. *Numerische Mathematik*, pages 1–28, 2016.
34. A. Lee Swindlehurst and Thomas Kailath. A performance analysis of subspace-based methods in the presence of model errors. I. The MUSIC algorithm. *IEEE Transactions on signal processing*, 40(7):1758–1774, 1992.
35. Richard Zippel. Probabilistic Algorithms for Sparse Polynomials. In *Proceedings of the International Symposiumon on Symbolic and Algebraic Computation*, EUROSAM '79, pages 216–226, London, UK, 1979. Springer-Verlag.
36. Richard Zippel. Interpolating polynomials from their values. *Journal of Symbolic Computation*, 9(3):375–403, 1990.
37. I. S. Iohvidov, Hankel and Toeplitz Matrices and Forms: Algebraic Theory. Birkhäuser Verlag, Boston, 1982.
38. M. T. Comer, E. L. Kaltofen, and C. Pernet. Sparse Polynomial Interpolation and Berlekamp/Massey Algorithms That Correct Outlier Errors in Input Values ISSAC '12: Proceedings of the 2012 international symposium on symbolic and algebraic computation, 2012, Grenoble, France. 2012.
39. Author, Article title, Journal, Volume, page numbers (year)
40. Author, Book title, page numbers. Publisher, place (year)