

# Exact solutions to Super Resolution on semi-algebraic domains in higher dimensions

Y. de Castro<sup>1</sup>, F. Gamboa<sup>2</sup>, D. Henrion<sup>3,4,5</sup>, J.-B. Lasserre<sup>3,4</sup>

Draft of February 9, 2015

## Abstract

We investigate the multi-dimensional Super Resolution problem on closed semi-algebraic domains for various sampling schemes such as Fourier or moments. We present a new semidefinite programming (SDP) formulation of the  $\ell_1$ -minimization in the space of Radon measures in the multi-dimensional frame on semi-algebraic sets.

While standard approaches have focused on SDP relaxations of the dual program (a popular approach is based on Gram matrix representations), this paper introduces an exact formulation of the primal  $\ell_1$ -minimization exact recovery problem of Super Resolution that unleashes standard techniques (such as moment-sum-of-squares hierarchies) to overcome intrinsic limitations of previous works in the literature. Notably, we show that one can exactly solve the Super Resolution problem in dimension greater than 2 and for a large family of domains described by semi-algebraic sets.

**Keywords:** super resolution; signed measure; semidefinite programming; total variation; semialgebraic domain.

## 1 Introduction

### 1.1 Super Resolution

The early formulation of the Super Resolution problem can be identified as the ability of faithfully reconstruct a high-dimensional sparse vector from the observation of a low-pass filter. This situation models important applications in imaging spectroscopy [HGTB94], image processing [PPK03], radar imaging [OBP94], or astronomy [MM05]. As a theoretical

---

<sup>1</sup>Département de Mathématiques (CNRS UMR 8628), Bâtiment 425, Faculté des Sciences d'Orsay, Université Paris-Sud 11, F-91405 Orsay Cedex, France.

<sup>2</sup>Institut de Mathématiques de Toulouse (CNRS UMR 5219). Université Paul Sabatier, 118 route de Narbonne, 31062 Toulouse, France.

<sup>3</sup>CNRS; LAAS; 7 avenue du colonel Roche, F-31400 Toulouse; France.

<sup>4</sup>Université de Toulouse; LAAS; F-31400 Toulouse; France.

<sup>5</sup>Faculty of Electrical Engineering, Czech Technical University in Prague, Technická 2, CZ-16626 Prague, Czech Republic.

baseline, suppose that one wants to reconstruct a vector  $x^*$  by solving a system of linear equations:

$$\mathbf{A}x = \mathbf{b} \quad \text{where } x \in \mathbb{R}^N, \mathbf{b} := \mathbf{A}x^*, \mathbf{b} \in \mathbb{R}^m \text{ and } m \ll N. \quad (1)$$

If the number  $s$  of non-vanishing components of  $x^*$  is small and the matrix  $\mathbf{A}$  enjoys some geometric property, namely the Null Space Property [CDD09], that depends only on its kernel and the sparsity  $s$ , then one can exactly reconstruct  $x^*$  by minimizing the  $\ell_1$ -norm within the affine subspace of all solutions of the linear system. Conditions on  $m$ ,  $N$ ,  $s$  and properties have been extensively studied, see for example [Don06b], [CDD09] and reference therein. Less than ten year ago, Super Resolution has seeded the ideas of compressed sensing theory [Don06a], [CT06], [CT07]. In this theory, the matrix  $\mathbf{A}$  is randomized and one is interested both in the construction of probability distributions allowing to show relevant properties, such as the Restricted Isometry Property [CT06], and the stability of the reconstruction process. This area of research is very fruitful and leads to many practical applications in signal and image processing, see for example [HS09], [TG07], or [CTL08]. To the best of our knowledge, the first mathematical works on Super Resolution are due to Donoho et al in the early ninety, see [DS89] and [Don92]. In these papers, the term *Super Resolution* appeared because the matrix  $\mathbf{A}$  is related to a discretization of some Fourier transform. As a matter of fact, when inverting a discrete Fourier transform, the separation of two close spikes in a sparse signal is made possible by minimizing the  $\ell_1$ -norm while a linear inversion method is not able to do so. Beside, in these years, many applied researchers were performing Fourier inversion of non negative sparse signals using some *entropy* regularization, see for example [SG87] and [Bur67]. At this time, the respective roles of sparsity and non negativity in regard of spikes separation from non linear Fourier inversion methods was not completely clear. An important step for understanding these roles has been taken by lifting the linear equation (1) up to the more abstract measure set up, see [Gas90], [GG96] and [DG96]:

$$\langle a_i, \mu \rangle = \int_X a_i(x) \mu(dx) = b_i \quad \text{where } \mu \in \mathcal{R}(X)_+ \text{ and } i = 1, \dots, m. \quad (2)$$

Here,  $(a_i(x))$  is a vector of continuous function defined on  $X$ , a given compact subset of  $\mathbb{R}^n$ ,  $\mathbf{b} = (b_i) \in \mathbb{R}^m$  and  $\mathcal{R}(X)_+$  is the set of all nonnegative Radon measures on  $X$ . In this frame, there exists very special points  $\mathbf{b}^*$  such that the set of all members of  $\mathcal{R}(X)_+$  satisfying (2) for  $\mathbf{b} = \mathbf{b}^*$  reduces to a singleton  $\{\mu_{\mathbf{b}^*}\}$ . Furthermore,  $\mu_{\mathbf{b}^*}$  is a discrete measure concentrated on very few points. Hence, if one deals with a  $\mathbf{b}$  close to a point  $\mathbf{b}^*$  the set of all solutions of (2) (or (1) with positivity constraint and a design matrix  $A$  discretizing the vectorial function  $(a_i(x))_{i=1}^m$ ) is very small. So that, two different methods for selecting a member of this set will lead to similar solutions. We refer to [Ana92], [GG96] and [DG96] where quantitative evaluations on the size of the set of solutions is performed and to [GG94] and [Lew96] for related evaluations in another context. Notice that the structure of points  $\mathbf{b}^*$  can be completely described in the case where the family of functions  $(a_i)$  is a Chebyshev system,  $T$ -system for short, this includes the case of discrete Fourier transform and moments, see [BE95] or [KS66] for an exhaustive overview on these systems of functions. A more involved situation is when (2) does not enjoy the non negativity assumption on the measure. This means that one wishes to solve the linear equation:

$$\langle a_i, \mu \rangle = \int_X a_i(x) \mu(dx) = b_i \quad \text{where } \mu \in \mathcal{R}(X) \text{ and } i = 1, \dots, m. \quad (3)$$

Here,  $\mathcal{R}(X)$  is the set of signed Radon measures on  $X$ . This is the frame of the present paper. Surprisingly, as shown by authors of this paper [dCG12] and under some assumptions on the family of functions  $(a_i)$ , there exists pairs  $(\mathbf{b}^*, \mu_{\mathbf{b}^*}) \in \mathbb{R}^m \times \mathcal{R}(X)$  such that  $b_i^* = \langle a_i, \mu_{\mathbf{b}^*} \rangle$ , for  $i = 1, \dots, m$ , and  $\mu_{\mathbf{b}^*}$  is the unique solution of (3) minimizing the total variation norm with  $\mathbf{b} = \mathbf{b}^*$ . Such  $\mu_{\mathbf{b}^*}$  are sparse in the sense that they are measures with finite support.

The study of the solutions to (1) and (3) uncovers that  $\ell_1$ -minimization faithfully reconstruct objects concentrated on very few points. However, the analysis in Super Resolution differs dramatically from Compressed Sensing. For instance, it is well known that sparse  $\ell_1$ -minimization cannot be successful in the ultrahigh-dimensional setting [Ver12] where  $N \gg \exp(m)$ . Indeed it was shown in [CGLP12] that one needs at least  $m \geq (\text{cst})s \log(N/s)$  measures to faithfully uncover  $s$  sparse vectors by  $\ell_1$ -minimization. Hence, the analysis of Compressed Sensing in terms of high-dimensional random geometry [CGLP12] cannot be extended to the space of measure. Moreover, observe that Compressed Sensing aims at recovering a sparse signal from random projection while, in Super Resolution, the sampling scheme is deterministic.

Admittedly their analysis differ but we can bridge the gap between (1) and (3) by considering their dual formulations. From the point of view of convex analysis, we see that the dual form of these linear programs aims at reconstructing a dual certificate [CT06, dCG12], i.e. an  $\ell_\infty$ -constraint linear combination of  $(a_i)$ , where  $(a_i)$  are the lines of  $\mathbf{A}$  in (1) and a family of continuous functions in (3). As pointed out by authors of this paper [dCG12], a parallel between Compressed Sensing and Super Resolution exists where the lines of  $\mathbf{A}$  are the evaluation of a vector of continuous function at some prescribed points. In this frame, Super Resolution can be seen as a Compressed Sensing problem where the dimension  $N$  goes to infinity. This analogy persists with the notion of dual certificate, i.e. a solution to the following dual program (4). Indeed, the dual programs given by the constraints (1) and (3) (while minimizing the  $\ell_1$ -norm and the total variation norm respectively) share the same expression:

$$\sup_{\mathbf{u} \in \mathbb{R}^m} \mathbf{b}^\top \mathbf{u} \quad \text{s.t.} \quad \|\mathbf{a}^\top \mathbf{u}\|_\infty \leq 1, \quad (4)$$

where  $\mathbf{a} = \mathbf{A}$  in Compressed Sensing (1) and  $\mathbf{a}(x) = (a_i(x))_{i=1}^m$  is a vector of continuous function in Super Resolution (3). Define a *dual certificate*  $\mathbf{P}$  as:

$$\mathbf{P} = \mathbf{a}^\top \mathbf{u}^*, \quad (5)$$

where  $\mathbf{u}^*$  is a solution to the dual program (4). From the duality properties, note that  $\mathbf{P}$  is a sub-gradient of the  $\ell_1$ -norm at a solution to the primal program (3). Hence, we can ensure that a target measure  $\mu^*$  is a solution to (3) if we are able to construct a dual polynomial (5) that interpolates the phases of the weights of  $\mu^*$  at its support points.

From a theoretical point of view, one of the main issue in Super Resolution consists in exhibiting such a *dual certificate*  $\mathbf{P}$ . In the Fourier frame, notice that an important construction, for target discrete measures whose support satisfy a *separable condition*, is given in the fundamental paper [CFG14] where a huge step has been taken. Indeed, the authors are the first to give a sharp condition on the support points of the target measure in order to warrant the existence of a dual certificate. Moreover, their proof is based on interpolating, by a Jackson kernel, the phases of the weights of the target measure at its support points and, henceforth, explicitly construct a dual certificate (5). In the Compressed Sensing frame,

observe that the same method has been investigated in [Kah11] using a Dirichlet kernel. In the present paper, we will not deal with this issue but rather with the practical resolution of the convex program (4).

In the Super Resolution frame, remark that the program (4) has finitely many variables but infinitely many constraints. This last point can be a severe limitation in practice. As a matter of fact, a difficult task is to construct a tractable program that deals with the  $\ell_\infty$ -constraint of (4). Standard formulations [CFG14] are based on Gram matrix representations, see below. Incidentally, these procedures cannot be extended to dimension greater than 2 or to semi-algebraic domains  $X$ . To cope with this issue, we consider a new parametrization of the primal program based on works of the authors [Las10]. Note that our method relies on infinitely many parameters but relaxations involving only a finite number of parameters are proved to lead to the exact solution of the primal program.

## 1.2 Previous works

During the last years, theoretical guarantees for exact recovery [BP13, dCG12, CFG14], bounds on the support recovery from inaccurate samplings [AdCG13, FG13], prediction of the Fourier coefficient from noisy observations [TBR13], and noise robustness [DP13] have been showed. These works prove that discrete measures can be recovered, in a robust manner, from few samples using an  $\ell_1$ -method.

From a numerical point of view, a solution to  $\ell_1$ -minimization is often computed using the dual program described by (4). Then, the  $\ell_\infty$ -norm constraint of the dual program (4) is equivalently formulated as a nonnegative constraint on (trigonometric) polynomials. This point of view unleashes Gram matrix representations [CFG14] or Toeplitz matrix representations [TBR13] to handle the constraint of non-negativity of (trigonometric) polynomials on domains. However, these formulations are limited to the frame of the real line and the torus in dimension one (see [Dum07] for instance) since they rely on the Fejér-Riesz theorem. As a matter of fact, the literature of Super Resolution has been focused on the fact that a semidefinite programming (SDP) formulation for the dual problem can be given as long as there exists a spectral factorization for a globally nonnegative trigonometric polynomial [CFG14] (Bounded Real Lemma using Fejér-Riesz theorem) or a spectral decomposition for semi-definite Toeplitz matrices [TBR13] (Caratheodory-Toeplitz theorem). Hence, except on the real line and the torus in dimension one, there is no exact SDP formulation of the Super Resolution problem for the dual form. However, relaxed SDP versions of the dual form in dimension greater than 2 are discussed in [XCV+13] and they have been used on the 2-sphere in [BDF15, BDF14].

## 1.3 Contribution

To the best of our knowledge, the present paper is the first to overcome this limitation and expand the scope of Super Resolution implementation to the *multi-dimensional* frame in general basic *semi-algebraic domains*. Indeed, we operate a smart method to tackle numerically the solution of equation (3) with minimal  $\ell_1$ -norm. This method uses both a re-parametrization in terms of moment sequences [Las10] and the so-called sum-of-squares

(SOS) decompositions of nonnegative multidimensional polynomials, used widely in systems control theory during the last decade, see for example [HG05]. Notice that, in the scope of Super Resolution, this technique is new and, contrary to other approaches, focuses on the primal program through a truncation of the moment sequences.

More precisely, given the real numbers  $b_i$ ,  $i = 1, \dots, m$ , consider the infinite-dimensional optimization problem:

$$\begin{aligned} \inf \quad & \|\mu\|_{TV} \\ \text{s.t.} \quad & \langle a_i, \mu \rangle = b_i, \quad i = 1, \dots, m \\ & \mu \in \mathcal{R}(X), \end{aligned} \tag{6}$$

where  $\|\cdot\|_{TV}$  is the total variation norm of measures (to be defined later). Notice that, under standard assumptions, problem (6) is feasible. That is

$$\exists \mu \in \mathcal{R}(X) \text{ such that for } i = 1, \dots, m, \quad \langle a_i, \mu \rangle = b_i. \tag{7}$$

Our main contribution concerns the numerical resolution of the total variation minimization problem (6). We extend the univariate ( $n = 1$ ) trigonometric SDP formulation of [CFG14] to a much more general SDP formulation in dimension  $n \geq 2$ , for measures supported on basic semialgebraic sets.

To this end, we use the Jordan decomposition of the signed measure  $\mu = \mu_+ - \mu_-$  as a difference of two nonnegative measures supported on  $X$  and we follow [Las10] to define a hierarchy of finite-dimensional primal-dual SDP problems:

- the primal problems correspond to SDP relaxations of the conditions that must satisfy finitely many moments of the two nonnegative measures on  $X$ ;
- the dual problems correspond to SDP strengthenings using SOS multipliers of the conditions that two distinguished polynomials are nonnegative on  $X$ .

The moment-SOS hierarchy is indexed by an integer  $k$ , called relaxation order, which is the (half of the) number of moments used to represent the measures in the primal problem, or equivalently, the (half of the) degree of the SOS representations of the polynomials in the dual problem. The larger is the relaxation order  $k$ , the larger is the size of the SDP problems, the number of variables and constraints growing polynomially in  $O(k^n)$ .

The primal SDP problem features the matrices of moments of the two nonnegative measures. If the rank of each moment matrix, as a function of  $k$ , stabilizes to a certain constant value, then the corresponding measure is atomic, with the number of atoms equal to the rank. Therefore, the total variation minimization problem (6) has been solved successfully, and this is certified by the polynomials solving the SDP problem. Numerical linear algebra can then be used to retrieve the support of the optimal measure.

In the sequel, we present some examples for which our method is the first to give an SDP formulation of the Super Resolution phenomena. As a matter of fact, our procedure encompasses a larger class of measurements than the class of standard moments discussed previously. Our numerical experiments are carried out with the Matlab interface GloptiPoly 3 which is designed to generate semidefinite relaxations of measure LP problems with polynomial data. So we assume that the functions  $a_i(x)$  in LP problem (8) are multivariate

polynomials, and for notational simplicity, we let  $a_i(x) := x^{\alpha_i} = x_1^{\alpha_i^1} \cdots x_n^{\alpha_i^n}$  where  $\alpha_i \in \mathbb{N}^n$  are given. Note that the choice of monomials is only motivated for notational simplicity, and that other choices of polynomials (e.g. Chebyshev polynomials) are typically preferable numerically<sup>1</sup>. SDP relaxations are then solved with SeDuMi or MOSEK, implementations of a primal-dual interior-point algorithm. For reproducibility purposes, our Matlab codes (using the public-domain interface GloptiPoly and the SDP solver SeDuMi) of the numerical examples presented next are available for download at

[homepages.laas.fr/henrion/software/tvsdp.tar.gz](http://homepages.laas.fr/henrion/software/tvsdp.tar.gz)

### 1.3.1 Disconnected domain

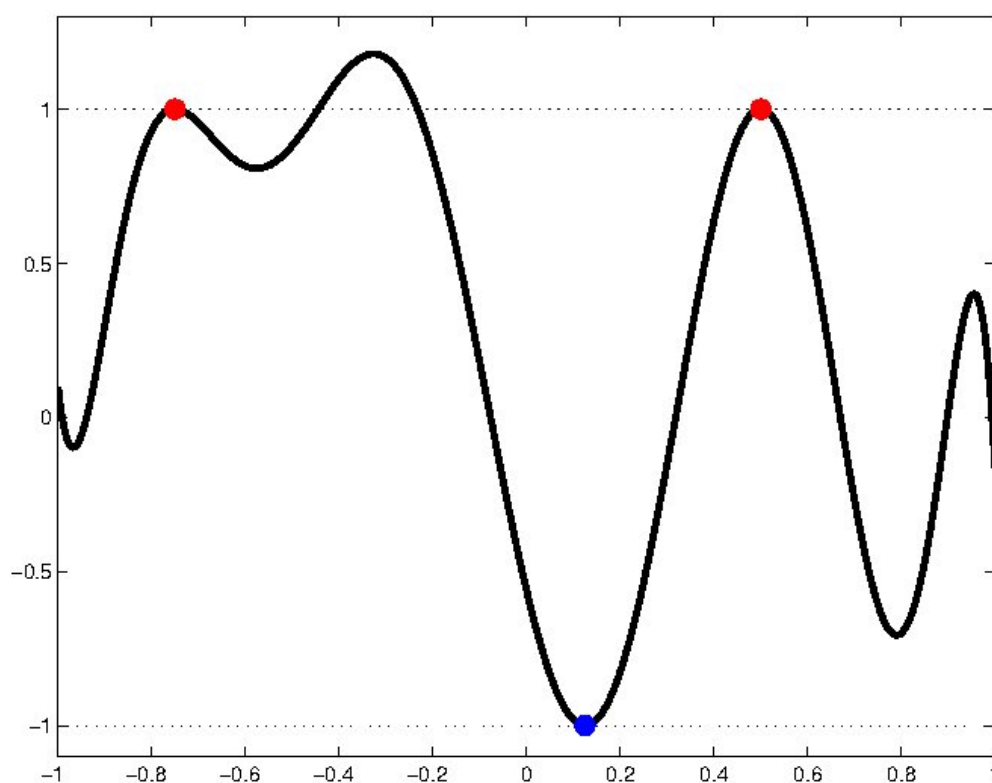


Figure 1: Degree 9 polynomial certificate for the univariate example, with 2 points (red) in the support of the positive part, and 1 point (blue) in the support of the negative part of the optimal measure.

We want to recover the measure:

$$\mu := \delta_{-3/4} + \delta_{1/2} - \delta_{1/8}$$

on the disconnected set  $X := [-1, -1/2] \cup [0, 1]$  which can be modeled as the polynomial superlevel set  $X = \{x \in \mathbb{R} : g_1(x) \geq 0\}$  for the choice:

$$g_1(x) := -(x+1)(x+1/2)x(x-1).$$

---

<sup>1</sup>A numerical analysis of the impact of the basis is however out of the scope of our work.

In LP (8) we let  $a_i(x) := x^i$  and  $b_i := (-3/4)^i + (1/2)^i - (1/8)^i$  for  $i = 0, 1, 2, \dots, 9$ . Solving the SDP relaxation of order  $k = 5$  on our standard PC takes 0.2 seconds, and optimality is certified from the solution of the primal moment problem with a rank 2 moment matrix for  $\mu_+$  and a rank 1 moment matrix for  $\mu_-$ , from which the 3 points can be extracted using numerical linear algebra. On Figure 1 we represent the degree 9 polynomial  $\sum_{i=0}^9 u_i x^i$  certifying optimality, constructed from the solution of the dual SOS problem. Indeed we can check that the polynomial attains the value +1 at the points  $x = -3/4$ , and  $x = 1/2$  (in red), it attains the value -1 at the point  $x = 1/4$  (in blue), while taking values between -1 and +1 on  $X$ . Notice in particular that the polynomial is larger than +1 around  $x = -1/4$ , but this point is not in  $X$ .

### 1.3.2 Low-pass filters in dimension greater than 3

In the Fourier frame, the recent SDP formulations of  $\ell_1$ -minimization in the space of complex valued measures are based on the Fejér-Riesz theorem. As a consequence, they cannot handle dimensions greater than 3. Observe that our procedure can bypass this limitation. For sake of readability, we present an example in dimension 2 although it can be extended to any dimension. We want to recover the measure

$$\mu := \delta_{(-1/2, 1/2)} + \delta_{(1/2, -1/2)} + \delta_{(1/2, 1/2)} + \delta_{(0,0)} - \delta_{(0, -1/2)} - \delta_{(1/2, 0)}$$

on the box  $X := [-1, 1]^2$ , from the knowledge of moments of degree up to 12, i.e.  $a_i(x) = x^i$  for  $i = 0, 1, \dots, 12$ . Solving the SDP relaxation of order  $k = 6$  on our standard PC takes less than 3 seconds, and optimality is certified from the solution of the primal moment problem with a rank 4 moment matrix for  $\mu_+$  and a rank 2 moment matrix for  $\mu_-$  from which the 6 points of the support of the optimal measure  $\mu$  can be extracted using numerical linear algebra with a relative accuracy around  $10^{-6}$ . On Figure 2 we represent the degree 12 polynomial certifying optimality, constructed from the solution of the dual SOS problem. Indeed we can check that the polynomial attains the value +1 at the 3 points  $x \in \{(-1/2, 1/2), (1/2, -1/2), (0, 0)\}$ , it attains the value -1 at the 2 points  $x \in \{(0, -1/2), (1/2, 0)\}$  (in blue), while taking values between -1 and +1 on  $X$ .

### 1.3.3 Localization of points on the sphere

Recent extensions of Super Resolution to spike deconvolution on the 2-sphere from spherical harmonic measurements has been investigated in [BDF15, BDF14]. In these paper, the authors give a sufficient condition for exact recovery using  $\ell_1$ -minimisation and they investigate spikes localization when the measurements are perturbed by additive noise.

From a numerical point of view, they used a relaxed version of the dual program (bounded real lemma in dimension  $d = 3$  and a Gram representation of the  $\ell_\infty$ -constraint appearing in the dual). Our work naturally extends to this frame and provides an exact formulation of the primal form.

For sake of numerical code simplicity, we have considered polynomials on  $\mathbb{R}^3$  restricted to the domain  $X$  given by the 2-sphere (note that one could have used homogenous spherical



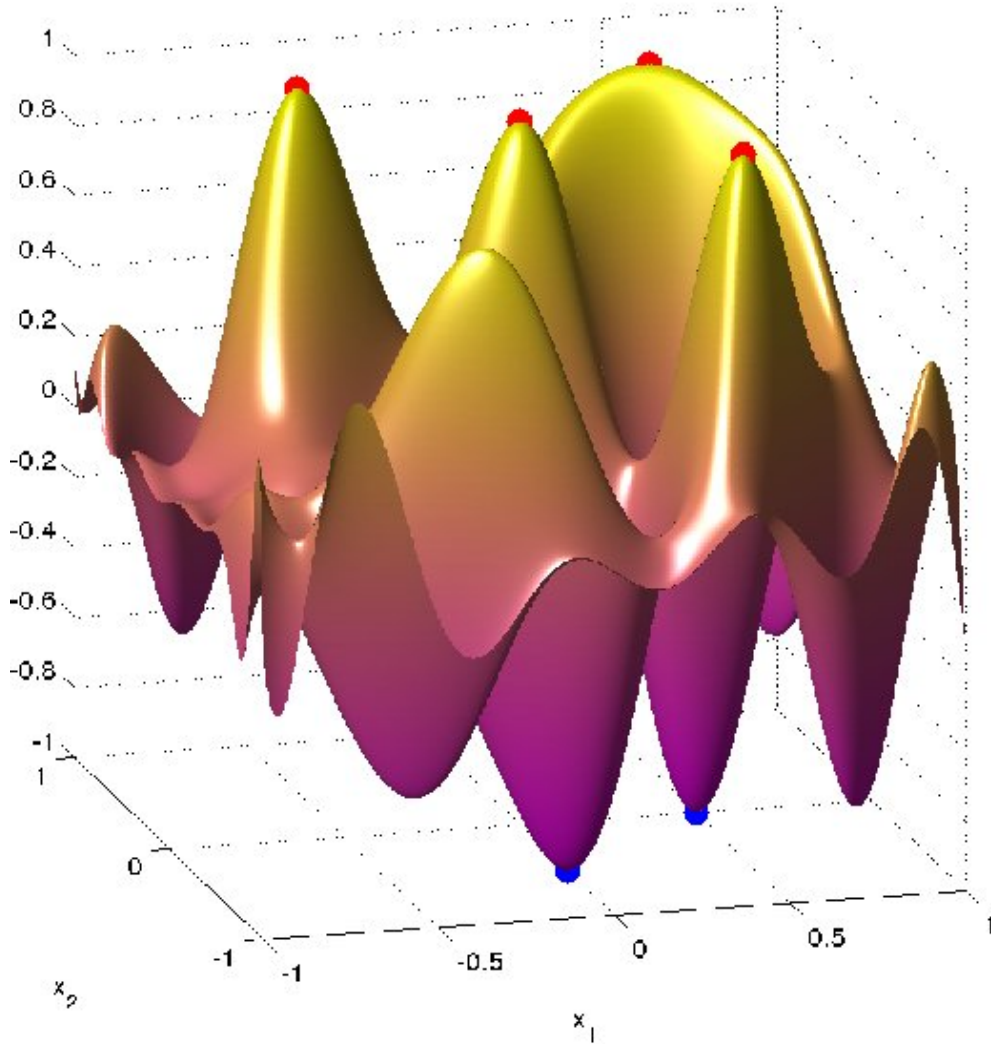


Figure 2: Degree 12 polynomial certificate for the bivariate example, with 4 points (red) in the support of the positive part, and 2 points (blue) in the support of the negative part of the optimal measure.

harmonics instead as in [BDF15]). We want to recover the measure

$$\mu := \delta_{(1,0,0)} + \delta_{(0,1,0)} + \delta_{(0,0,1)} - \delta_{(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}, 0)} - \delta_{(\frac{\sqrt{2}}{2}, 0, \frac{\sqrt{2}}{2})} - \delta_{(0, \frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2})}$$

that is supported on the positive orthant just for better visualization purposes. In LP (8), the  $a_i$  consist of 3-variate monomials of degree up to 5, i.e.  $m = 56$ . Solving the SDP relaxation of order  $k = 6$  on our standard PC takes less than 20 seconds, and optimality is certified from the solution of the primal moment problem with a rank 3 moment matrix for  $\mu_+$  and a rank 3 moment matrix for  $\mu_-$ , from which the 6 points can be extracted using numerical linear algebra. On Figure 3 we represent the degree 6 polynomial certifying optimality on the 2-sphere, constructed from the solution of the dual SOS problem. Indeed we can check that the polynomial attains the value +1 at the 3 prescribed points, it attains the value -1



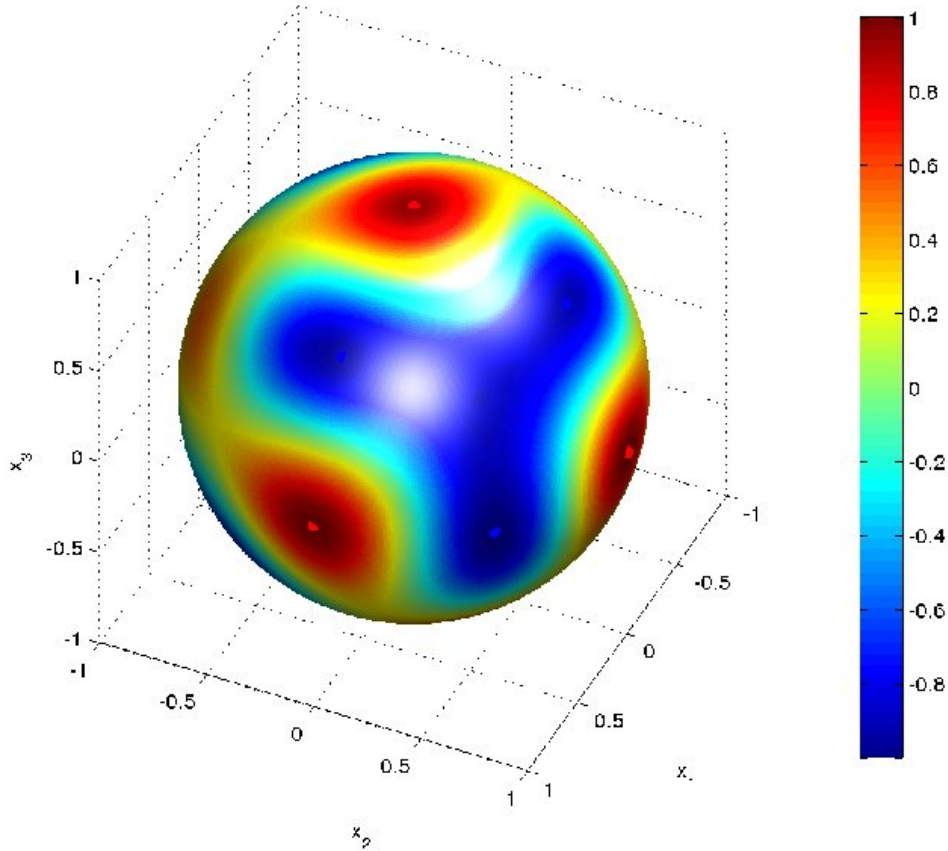


Figure 3: Degree 6 polynomial certificate for the sphere example, with 3 points (red) in the support of the positive part, and 3 points (blue) in the support of the negative part of the optimal measure.

(in red), at the 3 others prescribed points (in blue), while taking values between  $-1$  and  $+1$  on  $X$ .

From a theoretical point of view, the minimal separation condition appearing in [BDF15] requires a degree 2 polynomial. So our example satisfy the sufficient aforementioned condition.

## 2 Primal and dual LP formulation

### 2.1 General model and notation

Let  $n$  be a positive integer. Denote by  $\mathbb{R}[x]$  the set of all polynomials on  $\mathbb{R}^n$ , and for  $d \in \mathbb{N}$ ,  $\mathbb{R}_d[x]$  the set of all polynomials on  $\mathbb{R}^n$  with degree not greater than  $d$ . Further, we use the following notation:

- $X \subset \mathbb{R}^n$ , is a given closed basic semi-algebraic set:

$$X := \{x \in \mathbb{R}^n : g_j(x) \geq 0, j = 1, \dots, n_X\}$$

where  $g_j \in \mathbb{R}[x]$ ,  $j = 1, \dots, n_X$ , are given polynomials whose degrees are denoted by  $d_j$ ,  $j = 1, \dots, n_X$ . It is assumed that  $X$  is compact with an algebraic certificate of compactness. For example, one of the polynomial inequalities  $g_j(x) \geq 0$  should be of the form:

$$R^2 - \sum_{i=1}^n x_i^2 \geq 0,$$

for  $R$  a sufficiently large constant. Let  $\mathbf{g}_X := (g_j)_{j=1, \dots, n_X}$ .

- Let  $\mathbf{a} = (a_i)_{i=1}^m$  be a linearly independent family of polynomials of degree at most  $d$  on  $X$ . Notice that  $m \leq (1+d)^n$ .
- For monomials we use the multi-index notation

$$x^\alpha := \prod_{j=1}^n x_j^{\alpha_j}$$

for every  $x = (x_1, \dots, x_n) \in \mathbb{R}^n$  and  $\alpha = (\alpha_1, \dots, \alpha_n) \in \mathbb{N}^n$ .

- $\mathcal{C}(X)$ , the space of continuous functions on  $X$ , a Banach space when equipped with the sup-norm:

$$\|f\| = \sup_{x \in X} |f(x)|.$$

- $\mathcal{R}(X)$ , the space of signed Radon measures on  $X$ , a Banach space isometrically isomorphic to the topological dual  $\mathcal{C}(X)^*$  when equipped with the total variation norm:

$$\|\mu\|_{TV} = \sup_{\mathcal{P}} \sum_{E \in \mathcal{P}} |\mu(E)|,$$

where the supremum is taken over all partitions  $\mathcal{P}$  of  $X$  into a finite number of disjoint measurable subsets.

- $\mathcal{C}(X)_+ \subset \mathcal{C}(X)$  and  $\mathcal{R}(X)_+ \subset \mathcal{R}(X)$  the respective positive cones of nonnegative continuous functions on  $X$  and (nonnegative) Radon measures on  $X$ . We use the standard notation  $f \geq 0$  and  $\mu \geq 0$  for membership in  $\mathcal{C}(X)_+$  and  $\mathcal{R}(X)_+$ , respectively.
- To denote the integration of a function against a measure, we use the duality bracket:

$$\langle f, \mu \rangle = \int_X f d\mu,$$

for all  $f \in \mathcal{C}(X)$ , and  $\mu \in \mathcal{R}(X)$ .

With the usual Jordan decomposition:

$$\mu = \mu_+ - \mu_-$$

into a sum of two nonnegative Borel measures  $\mu_+, \mu_-$ , the optimization problem (6) can be rewritten equivalently as a linear programming (LP) problem in the convex cone  $\mathcal{R}(X)_+$ , namely:

$$\begin{aligned} p^* &= \inf \quad \langle 1, \mu_+ \rangle + \langle 1, \mu_- \rangle \\ \text{s.t.} \quad &\langle a_i, \mu_+ \rangle - \langle a_i, \mu_- \rangle = b_i, \quad i = 1, \dots, m \\ &\mu_+ \in \mathcal{R}(X)_+ \\ &\mu_- \in \mathcal{R}(X)_+. \end{aligned} \tag{8}$$

If  $\mathbf{a} = (a_i)_{i=1, \dots, m} \in \mathcal{C}(X)^m$  and  $\mathbf{b} = (b_i)_{i=1, \dots, m} \in \mathbb{R}^m$ , problem (8) is the dual of the following LP problem in the convex cone  $\mathcal{C}(X)_+$ :

$$\begin{aligned} d^* &= \sup \quad \mathbf{b}^\top \mathbf{u} \\ \text{s.t.} \quad &z_+(x) := 1 + \mathbf{a}^\top(x) \mathbf{u} \in \mathcal{C}(X)_+ \\ &z_-(x) := 1 - \mathbf{a}^\top(x) \mathbf{u} \in \mathcal{C}(X)_+ \end{aligned} \tag{9}$$

where the maximization is w.r.t.  $\mathbf{u} = (u_i)_{i=1, \dots, m} \in \mathbb{R}^m$ . Remark that LP problem (9) can be also written as:

$$\begin{aligned} d^* &= \sup \quad \mathbf{b}^\top \mathbf{u} \\ \text{s.t.} \quad &\|\mathbf{a}^\top(x) \mathbf{u}\|_\infty \leq 1. \end{aligned}$$

**Lemma 1** *There is no duality gap between primal LP (8) and dual LP (9), i.e.  $p^* = d^*$ .*

**Proof:** Define the vector  $r(\mu_+, \mu_-) \in \mathbb{R}^{m+1}$  by:

$$r(\mu_+, \mu_-) := (\langle 1, \mu_+ \rangle + \langle 1, \mu_- \rangle, \langle a_1, \mu_+ \rangle - \langle a_1, \mu_- \rangle, \dots, \langle a_m, \mu_+ \rangle - \langle a_m, \mu_- \rangle)$$

and the set

$$R := \{r(\mu_+, \mu_-) : (\mu_+, \mu_-) \in \mathcal{R}(X)_+ \times \mathcal{R}(X)_+\} \subset \mathbb{R}^{m+1}.$$

By [Bar02, Theorem 7.2],  $p^* = d^*$  provided that  $p^*$  is finite and  $R$  is closed. Finiteness of  $p^*$  follows from Assumption 7 and nonnegativity of the objective function  $\langle 1, \mu_+ \rangle + \langle 1, \mu_- \rangle$ . To prove that  $R$  is closed we have to show that for any sequence  $(\mu_+^n, \mu_-^n)_{n \in \mathbb{N}} \in \mathcal{R}(X)_+ \times \mathcal{R}(X)_+$  such that  $r(\mu_+^n, \mu_-^n) \rightarrow \mathbf{s} \in \mathbb{R}^{m+1}$  as  $n \rightarrow \infty$ , one has  $\mathbf{s} = r(\mu, \nu)$  for some finite measures  $\mu, \nu \in \mathcal{R}(X)_+$ . Since the supports of all the measures are contained in a compact set, and since  $\langle 1, \mu_+^n \rangle + \langle 1, \mu_-^n \rangle \rightarrow s_0$  all measures  $\mu_+^n, \mu_-^n$  are uniformly bounded. Therefore, from the weak-\* compactness (and weak-\* sequential compactness) of the unit ball (Banach-Alaoglu's Theorem), there is a subsequence  $(\mu_+^{n_k}, \mu_-^{n_k})_{k \in \mathbb{N}}$  that converges weakly-\* to an element  $(\mu, \nu) \in \mathcal{R}(X)_+ \times \mathcal{R}(X)_+$ . In particular, as all  $a_i$  are continuous,

$$\lim_{k \rightarrow \infty} r(\mu_+^{n_k}, \mu_-^{n_k}) = r(\mu, \nu),$$

which proves that  $R$  is closed. □

**Lemma 2** *For the dual LP problem (9) the supremum is attained.*

**Proof:** The feasibility set

$$U := \{\mathbf{u} \in \mathbb{R}^m : \|\mathbf{a}^\top(x)\mathbf{u}\|_\infty \leq 1\}$$

of the LP problem (9) is a closed convex subset of a finite-dimensional Euclidean space, and it contains the origin. Since the objective function in LP (9) is continuous on  $U$ , the optimum is attained if  $U$  is bounded. Suppose that  $U$  is not bounded. Then there exists a sequence  $(\mathbf{u}_n)_{n \in \mathbb{N}} \subset U$  such that  $\|\mathbf{u}_n\| \rightarrow \infty$  as  $n \rightarrow \infty$ . Write  $\mathbf{u}_n = \lambda_n \mathbf{v}_n$ , with  $\|\mathbf{v}_n\| = 1$ . Notice that  $0 < \lambda_n \rightarrow \infty$  and  $\mathbf{v}_n \in U$  because  $0 \in U$  and  $U$  is convex. Then

$$\|\mathbf{a}^\top(x)\mathbf{u}_n\|_\infty = \|\mathbf{a}^\top(x)\lambda_n\mathbf{v}_n\|_\infty = \lambda_n \|\mathbf{a}^\top(x)\mathbf{v}_n\|_\infty \leq 1,$$

so that  $\|\mathbf{a}^\top(x)\mathbf{v}_n\|_\infty \leq \lambda_n^{-1} \rightarrow 0$  as  $n \rightarrow \infty$ . Since  $\|\mathbf{v}_n\| = 1$ , there exists a subsequence  $n_k$  and  $\mathbf{v}$  with  $\|\mathbf{v}\| = 1$  such that  $\mathbf{v}_{n_k} \rightarrow \mathbf{v}$  as  $k \rightarrow \infty$  and  $\|\mathbf{a}^\top(x)\mathbf{v}\|_\infty = 0$ . By linear independence of  $\mathbf{a}$ , this implies that  $\mathbf{v} = 0$ , a contradiction.  $\square$

As a consequence of strong duality of Lemma 1, for any optimal primal-dual pair  $(\mu, u)$  we have the complementarity conditions

$$\langle z_+, \mu_+ \rangle = 0, \quad \langle z_-, \mu_- \rangle = 0$$

implying jointly with Lemma 2 that

$$\text{spt } \mu_+ \subset \{x \in X : \mathbf{a}^\top(x)\mathbf{u} = 1\}$$

and

$$\text{spt } \mu_- \subset \{x \in X : \mathbf{a}^\top(x)\mathbf{u} = -1\}$$

for some continuous function  $x \mapsto \mathbf{a}^\top(x)\mathbf{u}$ , and where  $\text{spt } \mu$  denotes the *support* of  $\mu$ , that is, the smallest closed set  $S \subset \mathbb{R}^n$  such that  $\mu(\mathbb{R}^n \setminus S) = 0$ .

**Lemma 3** *Problem (6) has an optimal atomic measure supported on at most  $2(m+1)$  points.*

**Proof:** Let  $\mu_+$  be a nonnegative measure solving problem (8) and let  $b_{+i} := \langle a_i, \mu_+ \rangle$ , with  $a_0 := 1$ ,  $i = 0, 1, \dots, m$ . If  $b_+ = 0$  then  $\mu_+ = 0$  is trivially atomic (with no atoms), so assume  $b_{+0} \neq 0$ , and consider the probability measure  $\bar{\mu}_+ := \mu_+/b_{+0}$  which satisfies the  $m$  equality constraints  $\langle a_i, \bar{\mu}_+ \rangle = \bar{b}_{+i} := b_{+i}/b_{+0}$ ,  $i = 1, \dots, m$ . From [Bar02, Proposition 9.4] there exists a probability measure  $\hat{\mu}_+$  satisfying the same equality constraints  $\langle a_i, \hat{\mu}_+ \rangle = \bar{b}_{+i}$  and which is supported on (at most)  $m+1$  points of  $X$ . The same reasoning can be applied to any nonnegative measure  $\mu_-$  solving problem (8), which has a discrete counterpart  $\hat{\mu}_-$  supported on (at most)  $m+1$  points of  $X$ . The result follows by considering the union of these two discrete supports, which consists of (at most)  $2(m+1)$  points of  $X$ .  $\square$

### 3 Primal and dual SDP formulation

Problem (8) is an instance of a generalized moment problem. As such it can be solved by a converging hierarchy of finite-dimensional primal-dual semidefinite programming (SDP) problems, as described comprehensively in [Las10]. In the sequel, we extract the key instrumental ingredients to the construction of the hierarchy.

### 3.1 Primal moment SDP

Recall from paragraph 2.1 that

$$X := \{x \in \mathbb{R}^n : g_j(x) \geq 0, j = 1, \dots, n_X\}$$

is a basic semi-algebraic set with a straightforward certificate of compactness, and let  $\mathbf{g}_X := (g_j)_{j=1, \dots, n_X}$  denote its defining polynomials. Given a measure  $\mu \in \mathcal{R}_+(X)$ , the real number

$$y_\alpha := \langle x^\alpha, \mu \rangle \tag{10}$$

is called its *moment* of order  $\alpha \in \mathbb{N}^n$ . Conversely, given a real valued sequence  $y := (y_\alpha)_{\alpha \in \mathbb{N}^n}$ , if identity (10) holds for all  $\alpha \in \mathbb{N}^n$ , we say that  $y$  has a representing measure  $\mu \in \mathcal{R}_+(X)$ . Equivalently, sequence  $y$  belongs to the infinite-dimensional moment cone

$$\mathcal{M}(X) := \{(y_\alpha)_{\alpha \in \mathbb{N}^n} : y_\alpha = \langle x^\alpha, \mu \rangle, \mu \in \mathcal{R}(X)_+\}.$$

In the sequel we describe a procedure to approximate this convex cone.

Given  $k \in \mathbb{N}$ , let  $\mathbb{R}[x]_k$  denote the space of real polynomials of degree at most  $k$ . Let us identify a polynomial  $p(x) = \sum_\alpha p_\alpha x^\alpha \in \mathbb{R}[x]_k$  with its vector  $p$  of coefficients in the monomial basis. Define the Riesz functional  $\ell_y$  as the linear functional acting on polynomials as follows:  $p \in \mathbb{R}[x]_k \mapsto \ell_y(p) = \sum_\alpha p_\alpha y_\alpha = p^\top y \in \mathbb{R}$ . Note that if sequence  $y$  has a representing measure  $\mu$ , then  $\ell_y(p) = \langle p, \mu \rangle$ . Define the *moment matrix* of order  $k$  as the Gram matrix of the quadratic form  $p \in \mathbb{R}[x]_k \mapsto \ell_y(p^2) \in \mathbb{R}$ , i.e. the matrix  $M_k(y)$  such that  $\ell_y(p^2) = p^\top M_k(y)p$ . By construction this matrix is symmetric and linear in  $y$ . Given a polynomial  $g \in \mathbb{R}[x]$ , define its localizing matrix of order  $k$  as the Gram matrix of the quadratic form  $p \in \mathbb{R}[x]_k \mapsto \ell_y(gp^2) \in \mathbb{R}$ , i.e. the matrix  $M_k(gy)$  such that  $\ell_y(gp^2) = p^\top M_k(gy)p$ . By construction this matrix is symmetric and linear in  $y$ . For  $j = 1, \dots, n_X$ , let  $k_j$  denote the smallest integer not less than half the degree of polynomial  $g_j$ , and let  $k_X := \max\{1, k_1, \dots, k_{n_X}\}$ . With these notations, and for  $k \geq k_X$ , define the finite-dimensional moment cone

$$\mathcal{M}_k(\mathbf{g}_X) := \{(y_\alpha)_{|\alpha| \leq 2k} : M_k(y) \succeq 0, M_{k-k_j}(g_j y) \succeq 0, j = 1, \dots, n_X\}$$

where  $\succeq 0$  means positive semidefinite.

Let  $y_+$  resp.  $y_-$  denote the sequence of moments

$$y_{+\alpha} := \int x^\alpha \mu_+(dx), \quad y_{-\alpha} := \int x^\alpha \mu_-(dx)$$

of  $\mu_+$  (resp.  $\mu_-$ ), indexed by  $\alpha \in \mathbb{N}^n$ . Primal measure LP (8) can be written as a primal moment LP:

$$\begin{aligned} p^* &= \min && y_{+0} + y_{-0} \\ &\text{s.t.} && \mathbf{A}(y_+, y_-) = \mathbf{b} \\ &&& y_+ \in \mathcal{M}(X) \\ &&& y_- \in \mathcal{M}(X) \end{aligned}$$

where the linear system of equations  $\mathbf{A}(y_+, y_-) = \mathbf{b}$  models the linear moment constraints. The moment relaxation of order  $k \geq \max\{k_X, d\}$  of the primal moment LP then reads:

$$\begin{aligned} p_k^* &= \min && y_{+0} + y_{-0} \\ &\text{s.t.} && \mathbf{A}(y_+, y_-) = \mathbf{b} \\ &&& y_+ \in \mathcal{M}_k(\mathbf{g}_X) \\ &&& y_- \in \mathcal{M}_k(\mathbf{g}_X) \end{aligned} \tag{11}$$

where the minimization is w.r.t. a vector  $(y_+, y_-)$  of moments of degree at most  $2k$ . For fixed  $k$ , problem (11) is a finite-dimensional linear programming problem in the convex cone of positive semidefinite matrices, i.e. an SDP problem. When  $k$  varies, the number of moments, as well as the size of the moment and localizing matrices in problem (11) are binomial coefficients growing in  $O(k^n)$ .

It can be shown that  $(p_k^*)$  is a monotonically nondecreasing converging sequence of lower bounds on  $p^*$ , i.e.  $p_{k+1}^* \geq p_k^*$  and  $\lim_{k \rightarrow \infty} p_k^* = p^*$ . However, in the context of solving LP (8), a more relevant result is the following:

**Theorem 1** *For a given relaxation order  $k \geq \max\{d, k_X\}$ , let  $(y_+^*, y_-^*)$  denote the solution of the moment SDP (11). If*

$$\text{rank } M_{k-k_X}(y_+^*) = \text{rank } M_k(y_+^*) \text{ and } \text{rank } M_{k-k_X}(y_-^*) = \text{rank } M_k(y_-^*) \tag{12}$$

*then  $p_k^* = p^*$  and LP (8) has an optimal solution  $(\mu_+^*, \mu_-^*)$  with  $\mu_+^*$  (resp.  $\mu_-^*$ ) atomic supported at  $r_+ := \text{rank } M_k(y_+^*)$ , (resp.  $r_- := \text{rank } M_k(y_-^*)$ ) points.*

**Proof:** By [Las10, Theorem 3.11],  $y_+^*$  (resp.  $y_-^*$ ) is the vector of moments up to order  $2k$ , of a measure  $\mu_+^*$  (resp.  $\mu_-^*$ ) supported on  $\text{rank } M_k(y_+^*)$  points (resp.  $\text{rank } M_k(y_-^*)$ ) points of  $X$ . Therefore  $(\mu_+^*, \mu_-^*)$  is a feasible solution of (8) with value  $p_k \leq p^*$ , which proves that  $(\mu_+^*, \mu_-^*)$  is an optimal solution of (8) and  $p_k = p^*$ .  $\square$

Given a moment matrix  $M_k(y_+^*)$  satisfying the rank constraint of Theorem 1, there is a numerical linear algebra algorithm that extracts the  $r_+$  points of the support of the corresponding atomic measure  $\mu_+^*$ , and similarly for  $\mu_-^*$ . The algorithm is described e.g. in [Las10, Section 4.3] and it is implemented in the Matlab toolbox GloptiPoly 3.

A certificate of optimality can be obtained by solving the dual problem to primal SDP problem (11), and this is described next.

### 3.2 Dual SOS SDP

For a given integer  $k$ , let  $\Sigma[x]_k \subset \mathbb{R}[x]_{2k}$  denote the space of SOS (sums of squares) polynomials of degree at most  $2k$ . If  $p \in \Sigma[x]_k$  this means that there exists  $q_j \in \mathbb{R}[x]_k$ ,  $j = 1, \dots, n_p$ , such that  $p = \sum_{j=1}^{n_p} q_j^2$ . Let

$$\mathcal{P}(X) := \{p \in \mathbb{R}[x] : p(x) \geq 0, \forall x \in X\}$$



denote the infinite-dimensional cone of nonnegative polynomials on  $X$ , and for  $k \geq k_X$ , define the finite-dimensional SOS cone, also called quadratic module

$$\mathcal{P}_k(\mathbf{g}_X) := \left\{ p_0 + \sum_{j=1}^{n_X} g_j p_j, p_j \in \Sigma[x]_{k-k_j}, j = 0, 1, \dots, n_X \right\} \subset \mathbb{R}[x]_{2k}.$$

Under the above assumptions on the polynomial family  $\mathbf{g}_X$  defining  $X$ , from Putinar's theorem, see e.g. [Las10, Theorems 2.14 and 3.8], it holds that  $\mathcal{P}(X) \cap \mathbb{R}[x]_\kappa$  is the closure of  $(\cup_{k \geq k_X} \mathcal{P}_k(\mathbf{g}_X)) \cap \mathbb{R}[x]_\kappa$  for all  $\kappa \geq k_X$ . Observe also that  $\mathcal{M}(X)$  (resp.  $\mathcal{M}_k(\mathbf{g}_X)$ ) is the dual cone to  $\mathcal{P}(X)$  (resp.  $\mathcal{P}_k(\mathbf{g}_X)$ ). Whereas testing whether a given polynomial belongs to  $\mathcal{P}(X)$  is a difficult task, testing whether a given polynomial belongs to  $\mathcal{P}_k(\mathbf{g}_X)$ , for a fixed  $k$ , amounts to solving an SDP problem.

Dual continuous function LP (9) can be written as a positive polynomial LP

$$\begin{aligned} d^* &= \max \mathbf{b}^\top \mathbf{u} \\ \text{s.t.} \quad & 1 + \mathbf{a}^\top(x) \mathbf{u} \in \mathcal{P}(X) \\ & 1 - \mathbf{a}^\top(x) \mathbf{u} \in \mathcal{P}(X) \end{aligned}$$

and its SOS strengthening of order  $k \geq k_X$  reads:

$$\begin{aligned} d_k^* &= \max \mathbf{b}^\top \mathbf{u} \\ \text{s.t.} \quad & 1 + \mathbf{a}^\top(x) \mathbf{u} \in \mathcal{P}_k(\mathbf{g}_X) \\ & 1 - \mathbf{a}^\top(x) \mathbf{u} \in \mathcal{P}_k(\mathbf{g}_X) \end{aligned} \tag{13}$$

where the maximization is w.r.t. a vector  $\mathbf{u} \in \mathbb{R}^m$ . It turns out that this is SOS problem (13) is an SDP problem dual to the moment problem (11):

**Lemma 4** *There is no duality gap between SDP problems (11) and (13), i.e.  $p_k^* = d_k^*$ , and both (11) and (13) have an optimal solution.*

**Proof:** We first show that (11) has an optimal solution. Recall that one of constraints  $g_j(x) \geq 0$  that define  $X$  states that  $M - \|x\|^2 \geq 0$  for some  $M > 1$ . From the constraint  $M_{k-k_j}(g_j y_+) \succeq 0$  one deduces that  $\ell_{y_+}(M - x_i^2) \geq 0$ , and  $\ell_{y_+}(M x_i^t - x_i^{t+2}) \geq 0$  for every  $t = 1, \dots, 2k - 2$ . Hence  $\ell_{y_+}(x_i^{2k}) \leq M^k y_{+0}$  for every  $i = 1, \dots, n$ . With similar arguments,  $\ell_{y_-}(x_i^{2k}) \leq M^k y_{-0}$  for every  $i = 1, \dots, n$ . By [Las10, Proposition 3.6]  $|y_{+\alpha}| \leq M^k y_{+0}$  and  $|y_{-\alpha}| \leq M^k y_{-0}$  for all  $\alpha \in \mathbb{N}_{2k}^n$ . Next, in a minimizing sequence  $(y_+^s, y_-^s)$ ,  $s \in \mathbb{N}$ , of (11) one has  $y_{+0}^s + y_{-0}^s \leq y_{+0}^1 + y_{-0}^1 =: \rho$  for all  $s$ , and so  $|y_{+\alpha}^s| \leq M^k \rho$  and  $|y_{-\alpha}^s| \leq M^k \rho$  for all  $\alpha \in \mathbb{N}_{2k}^n$ , and all  $s = 1, \dots$ . From this we deduce that there is a subsequence  $(y_+^{s_t}, y_-^{s_t})$ ,  $t \in \mathbb{N}$ , that converges to some  $(y_+^*, y_-^*)$  as  $t \rightarrow \infty$ , with value  $y_{+0}^* + y_{-0}^* = p_k^*$ . In addition by a simple continuity argument,  $M_k(y_+^*) \succeq 0$  and  $M_{k-k_j}(g_j y_+^*) \succeq 0$ ,  $j = 1, \dots, n_X$ . Similarly  $M_k(y_-^*) \succeq 0$  and  $M_{k-k_j}(g_j y_-^*) \succeq 0$ ,  $j = 1, \dots, n_X$ , which proves that  $(y_+^*, y_-^*)$  is an optimal solution of (11).

Next, the set of optimal solutions  $y^* := \{(y_+^*, y_-^*)\}$  of (11) is compact. This follows from  $|y_{+\alpha}^*| \leq M^k y_{+0}^* \leq M^k p_k^*$  and  $|y_{-\alpha}^*| \leq M^k y_{-0}^* \leq M^k p_k^*$  for all  $\alpha \in \mathbb{N}_{2k}^n$ . And so every sequence in  $y^*$  has a converging subsequence. From [Bar02, Chapter IV. Theorem 7.2] one also deduces that there is no duality gap between (11) and (13).

It remains to prove that (13) has an optimal solution. Consider a maximizing sequence  $(\mathbf{u}_t)_{t \in \mathbb{N}}$ , with  $\mathbf{b}^T \mathbf{u}_t \rightarrow p_k^* = p^*$  as  $t \rightarrow \infty$ . By feasibility in (13), one has  $\|\mathbf{a}(x)^T \mathbf{u}_t\|_\infty \leq 1$  for all  $t$  and therefore  $(\mathbf{u}_t) \subset \mathcal{U} := \{\mathbf{u} \in \mathbb{R}^n : \|\mathbf{a}(x)^T \mathbf{u}\|_\infty \leq 1\}$  and  $\mathcal{U}$  is compact (see the proof of Lemma 2). Therefore there exists  $\mathbf{u}^* \in \mathcal{U}$  and a subsequence  $(t_\ell)_{\ell \in \mathbb{N}}$  such that  $\mathbf{u}_{t_\ell} \rightarrow \mathbf{u}^* \in \mathcal{U}$  as  $\ell \rightarrow \infty$ . In particular  $\mathbf{b}^T \mathbf{u}^* = p^*$ . Moreover, since by Lemma 8 in the Appendix the convex cone  $\mathcal{P}_k(\mathbf{g}_X)$  is closed,  $1 - \mathbf{a}(x)^T \mathbf{u}_{t_\ell} \rightarrow 1 - \mathbf{a}(x)^T \mathbf{u}^* \in \mathcal{P}_k(\mathbf{g}_X)$ , which proves that  $\mathbf{u}^*$  is an optimal solution of (13).  $\square$

Assume that the rank conditions of Theorem 1 is satisfied at some relaxation order  $k$ , and let  $(\mu_+^*, \mu_-^*)$  denote the atomic measures optimal for problem (8), obtained from the solution of the primal SDP problem (11). Let  $\mathbf{u}^*$  denote an optimal solution of the dual SDP problem (13). The duality result of Lemma 4 implies that

$$\text{Supp } \mu_+^* \subset \{x \in X : \mathbf{a}^\top(x) \mathbf{u}^* = 1\}$$

and

$$\text{Supp } \mu_-^* \subset \{x \in X : \mathbf{a}^\top(x) \mathbf{u}^* = -1\}$$

so that the polynomial  $\mathbf{a}^\top(x) \mathbf{u}^*$  can be used as a certificate of optimality. We formulate this in the following dual to Theorem 1.

**Lemma 5** *Assume that the rank conditions (12) of Theorem 1 hold. Let us denote by  $\mathbf{u}^*$  the optimal solution of SOS SDP (13). Then the polynomial  $z_+^*(x) := 1 + \mathbf{a}^\top(x) \mathbf{u}^*$  vanishes at the  $r_+$  points of the support of  $\mu_+^*$ , and the polynomial  $z_-^*(x) := 1 - \mathbf{a}^\top(x) \mathbf{u}^*$  vanishes at the  $r_-$  points of the support of  $\mu_-^*$ .*

**Proof:** Let us denote by  $\{x_+^k\}_{k=1, \dots, r_+} \subset X$  the points of the support of the optimal measure  $\mu_+^*$ , computed from the moments  $y_+^*$  solving optimally moment SDP (11). By complementarity of the solutions of primal-dual SDP (11) and (13), it holds  $\langle z_+^*, \mu_-^* \rangle = 0$  and hence  $\langle z_+^*, \delta_{x_+^k} \rangle = z_+^*(x_+^k) = 0$  for each  $k = 1, \dots, r_+$ . The proof is similar for  $z_-^*$  and  $\mu_-^*$ .  $\square$

## 4 Discussion

We would like to point out that the developments in this paper were inspired by a previous work on optimal control for linear systems formulated as a primal LP (8) on measures and a dual LP on continuous functions (9), and solved numerically with primal-dual moment-SOS SDP hierarchies [CAHL13, CAHL14]. Formulating optimal control problems as moment problems was a classical research topic in the 1960s, where optimal control laws were sought in measures spaces (completions of Lebesgue spaces) to allow for oscillations and concentrations, see e.g. [Kra68] or the overview in [Fat99, Section III]. In the case of linear optimal control of an ordinary differential equation of order  $n$ , it was proved in [Neu64] that there is always an  $n$ -atomic optimal measure solving problem (8).

In practice, Theorem 1 should be used as follows:

- 1 Let  $k = \max\{d, k_X\}$ .

- 2 Solve SDP problem (11) and its dual (13) with a primal-dual algorithm.
- 3 If the rank condition (12) of Theorem 1 is satisfied, then extract the measure from the solution of (11) and the polynomial certificate from the solution of (13). Otherwise, let  $k = k + 1$ , and go to 1.

We conjecture that if the data  $\mathbf{a}, \mathbf{b}$  in problem (8) are generic, then there is a finite value of  $k$  for which the rank condition of Theorem 1 is satisfied. The rationale behind this assertion follows from a result by Nie [Nie14] on *generic* finite convergence for the moment-SOS SDP hierarchy for polynomial optimization over compact basic semi-algebraic sets. Translated in the present context for a fixed family of data  $\mathbf{a}$ , results in [Nie14] yield that there is a set of polynomials  $\{h_1, \dots, h_L\} \subset \mathbb{R}[\mathbf{u}]$ , such that, given a feasible solution  $\mathbf{u}$  of (9), if  $h_\ell(\mathbf{u}) \neq 0$  for all  $\ell = 1, \dots, L$ , then indeed

$$1 + \mathbf{a}^T(x)\mathbf{u} = p_0^1(x) + \sum_{j=1}^{n_X} p_j^1(x) g_j(x), \quad x \in \mathbb{R}^n,$$

and

$$1 - \mathbf{a}^T(x)\mathbf{u} = p_0^2(x) + \sum_{j=1}^{n_X} p_j^2(x) g_j(x), \quad x \in \mathbb{R}^n,$$

for some SOS polynomials  $p_j^k$ ,  $k = 1, 2$  and  $j = 1, \dots, n_X$ . So if the optimal solution  $\mathbf{u}^*$  of (9) satisfies  $h_\ell(\mathbf{u}^*) \neq 0$ ,  $\ell = 1, \dots, L$ , then  $d^* = d_k^*$  for some index  $k$  (i.e. finite convergence takes place). Similarly, by [Nie13] the rank-condition (12) of Theorem 1 also holds generically for polynomial optimization (which however is a context different from the present context). Put differently, finite convergence would not hold only if every optimal solution  $\mathbf{u}$  of (9) would be a zero of some polynomial of the family  $\{h_1, \dots, h_L\} \subset \mathbb{R}[\mathbf{u}]$ . But so far we have not proved that at least one optimal solution  $\mathbf{u}^*$  of (9) is not a zero of some of the polynomials  $h_\ell$ , at least for generic  $\mathbf{b}$ .

Of course, finite convergence occurs for trigonometric polynomials on  $X = [0, 2\pi]$ , which follows from the Fejér-Riesz theorem and this was exploited in the landmark paper [CFG14]. Similarly, but apparently not so well-known, the Fejér-Riesz theorem also holds in dimension  $n = 2$ . Indeed it follows from Corollary 3.4 in [Sch06] that every non-negative bivariate trigonometric polynomial can be written as a sum of squares of trigonometric polynomials<sup>2</sup>. So again for trigonometric polynomials on  $X = [0, 2\pi]^2$ , finite convergence of the hierarchy (13) takes place, i.e.,  $d_k^* = d^*$ . Note however that in contrast to the one-dimensional case, there is *no* explicit upper bound on the degrees of the sum of squares which are required, so that even in the two-dimensional Fourier case we do not have an a priori estimates on the smallest value of  $k$  for which  $d_k^* = d^*$  and for which we can guarantee that the rank condition of Theorem 1 is satisfied.

Generally speaking, even if our genericity conjecture is true, we do not have a priori estimates on the smallest value of  $k$  for which Theorem 1 holds. As mentioned above, this also true even in the two-dimensional case on  $[0, 2\pi]^2$  where finite convergence is guaranteed in all cases.

---

<sup>2</sup>We are grateful to Markus Schweighofer for providing this reference.

## 5 Appendix

We first recall some standard results of convex analysis.

**Lemma 6** ([FK94, Corollary I.1.3]) *Let  $C \subset \mathbb{R}^n$  be a closed convex cone with dual  $C^* = \{y : \langle x, y \rangle \geq 0, \forall x \in C\}$ . Then  $\text{int } C^* \neq \emptyset \Leftrightarrow C \cap (-C) = \{0\}$ .*

**Lemma 7** ([FK94, Corollary I.1.6]) *Let  $C \subset \mathbb{R}^n$  be a closed convex cone whose dual  $C^*$  has nonempty interior. Then for all  $y \in \text{int } C^*$ , the set  $\{x \in C : \langle x, y \rangle \leq 1\}$  is compact.*

**Lemma 8** *The convex cone  $\mathcal{P}_k(\mathbf{g}_X)$  is closed.*

**Proof:** Let  $\mathcal{S}_+^n$  be the convex cone of real symmetric matrices of size  $n$  that are positive semidefinite. Let  $\mathbb{N}_k^n$  be the set of  $n$ -dimensional integer vectors  $\alpha$  such that  $\sum_{i=1}^n \alpha_i \leq k$  and let  $v_k(x) := (x^\alpha)_{\alpha \in \mathbb{N}_k^n}$  be a vector of monomials of degree up to  $k$ . Next let  $v_k(x) v_k(x)^T = \sum_{\alpha \in \mathbb{N}_{2k}^n} x^\alpha A_{0\alpha}$  and

$$v_{k-v_j}(x) v_{k-v_j}(x)^T g_j(x) = \sum_{\alpha \in \mathbb{N}_{2k}^n} x^\alpha A_{j\alpha}, \quad j = 1, \dots, n_X,$$

for some appropriate real symmetric matrices  $A_{j\alpha}$ .

Consider a sequence  $(q_t)_{t \in \mathbb{N}} \subset \mathcal{P}_k(\mathbf{g}_X)$  such that  $q_t \rightarrow q \in \mathbb{R}[x]_{2k}$  as  $t \rightarrow \infty$ . That is

$$q_t(x) = p_{0t}(x) + \sum_{j=1}^{n_X} p_{jt}(x) g_j(x), \quad \forall x \in \mathbb{R}^n,$$

for some  $p_{jt} \in \Sigma[x]_{k-v_j}$ ,  $j = 0, \dots, n_X$ , for all  $t \in \mathbb{N}$ . More precisely, coefficient-wise

$$q_{t\alpha} = \langle Q_{0t}, A_{0\alpha} \rangle + \sum_{j=1}^{n_X} \langle Q_{jt}, A_{j\alpha} \rangle, \quad \forall \alpha \in \mathbb{N}_{2k}^n, \quad (14)$$

for some appropriate matrices  $Q_{jt} \in \mathcal{S}_+^{k-v_j}$ . Let  $y = (y_\alpha)_{\alpha \in \mathbb{N}_{2k}^n}$  be the moments  $y_\alpha := \int_X x^\alpha dx$  of the measure uniformly supported on  $X$ . Observe that since  $X$  has nonempty interior,

$$\int_X p(x) dx > 0 \quad \forall 0 \neq p \in \Sigma[x]_k,$$

and

$$\int_X p(x) g_j(x) dx > 0 \quad \forall 0 \neq p \in \Sigma[x]_{k-v_j}, \quad j = 1, \dots, n_X.$$

Put differently  $M_k(y) \succ 0$  and  $M_{k-v_j}(g_j y) \succ 0$ ,  $j = 1, \dots, n_X$ .

The convergence  $q_t \rightarrow q$  implies  $\langle q_t, y \rangle \rightarrow \langle q, y \rangle$  as  $t \rightarrow \infty$ . Hence there is some  $\eta$  such that  $\eta \geq \langle q_t, y \rangle$  for all  $t \in \mathbb{N}$ . This in turn implies

$$\begin{aligned} \eta \geq \langle q_t, y \rangle &= \langle p_{0t}, y \rangle + \sum_{j=1}^{n_X} \langle p_{jt} g_j, y \rangle \\ &= \langle Q_{0t}, M_k(y) \rangle + \sum_{j=1}^{n_X} \langle Q_{jt}, M_{k-v_j}(g_j y) \rangle. \end{aligned} \quad (15)$$

Therefore

$$\sup_t \langle Q_{0t}, M_k(y) \rangle \leq \eta, \quad \sup_t \langle Q_{jt}, M_{k-v_j}(g_j y) \rangle \leq \eta, \quad j = 1, \dots, n_X.$$

As  $0 \prec M_k(y) \in \text{int}(\mathcal{S}_+^k)^*$ , and  $0 \prec M_{k-v_j}(g_j y) \in \text{int}(\mathcal{S}_+^{k-v_j})^*$ ,  $j = 1, \dots, n_X$ , one may invoke Lemma 7 and conclude that the sequences  $(Q_{0t})_{t \in \mathbb{N}} \subset \mathcal{S}_+^k$  and  $(Q_{jt})_{t \in \mathbb{N}} \subset \mathcal{S}_+^{k-v_j}$  are norm-bounded. Therefore there is a subsequence  $(t_\ell)_{\ell \in \mathbb{N}}$  and matrices  $Q_0 \in \mathcal{S}_+^k$  and  $Q_j \in \mathcal{S}_+^{k-v_j}$ ,  $j = 1, \dots, n_X$ , such that

$$Q_{0t_\ell} \rightarrow Q_0, \quad Q_{jt_\ell} \rightarrow Q_j, \quad j = 1, \dots, n_X$$

as  $\ell \rightarrow \infty$ . Taking the limit for the subsequences  $(q_{t_\ell \alpha})_{\ell \in \mathbb{N}}$  and  $(Q_{jt_\ell})_{\ell \in \mathbb{N}}$  in (14) yields coefficient-wise

$$q_\alpha = \langle Q_0, A_{0\alpha} \rangle + \sum_{j=1}^{n_X} \langle Q_j, A_{j\alpha} \rangle, \quad \forall \alpha \in \mathbb{N}_{2k}^n,$$

which proves that  $q \in \mathcal{P}_k(\mathbf{g}_X)$ , the desired result.  $\square$

## References

- [Ana92] G. A. Anastassiou, *Weak convergence and the Prokhorov radius*, J. Math. Anal. Appl. 163(2):541-558, 1992.
- [AdCG13] J.-M. Azais, Y. de Castro, and F. Gamboa, *Spike detection from inaccurate samplings*, arXiv:1301.5873, Jan. 2013.
- [Bar02] A. Barvinok, *A course in convexity*, Graduate Studies in Mathematics, vol. 54, American Mathematical Society, Providence, RI, 2002.
- [BDF14] T. Bendory, S. Dekel, and A. Feuer, *Super-resolution on the sphere using convex optimization*, arXiv:1412.3282, Dec. 2014.
- [BDF15] T. Bendory, S. Dekel, and A. Feuer, *Exact recovery of Dirac ensembles from the projection onto spaces of spherical harmonics*, arXiv:1412.3284, Dec. 2014, to appear in Constructive Approximation, 2015.
- [BE95] P. B. Borwein and T. Erdélyi, *Polynomials and polynomial inequalities*, Springer Verlag, 1995.
- [BP13] K. Bredies and H. Pikkarainen, *Inverse problems in spaces of measures*, ESAIM: Control, Optimisation and Calculus of Variations, 19(1):190-218, 2013.
- [Bur67] J. P. Burg, *Maximum entropy spectral analysis.*, 37th Annual International Meeting., Society of Exploration Geophysics, 1967.
- [CFG14] E. J. Candes and C. Fernández-Granda, *Towards a mathematical theory of super-resolution*, Communications on Pure and Applied Mathematics, 67(6):906-956, 2014.

- [CT06] E. J. Candes and T. Tao, *Near-optimal signal recovery from random projections: universal encoding strategies?*, IEEE Trans. Inform. Theory 52(12):5406–5425, 2006.
- [CT07] E. J. Candes and T. Tao, *The Dantzig selector: statistical estimation when  $p$  is much larger than  $n$* , Ann. Statist. 35(6):2313–2351, 2007.
- [CGLP12] D. Chafaï, O. Guédon, G. Lécué, and A. Pajor, *Interaction between compressed sensing, random matrices and high dimensional geometry*, Panoramas et synthèses, 37, SMF, 2012.
- [CTL08] G.-H. Chen, J. Tang, and S. Leng, *Prior image constrained compressed sensing (piccs): a method to accurately reconstruct dynamic ct images from highly undersampled projection data sets*, Medical physics 35:660, 2008.
- [CAHL13] M. Claeys, D. Arzelier, D. Henrion, and J. B. Lasserre, *Moment LMI approach to LTV impulsive control*, IEEE Conf. Decision and Control, 5810–5815, 2013.
- [CAHL14] M. Claeys, D. Arzelier, D. Henrion, and J. B. Lasserre, *Measures and LMIs for impulsive nonlinear optimal control*, IEEE Trans. Autom. Control 59(5):1374–1379, 2014.
- [CDD09] A. Cohen, W. Dahmen, and R. DeVore, *Compressed sensing and best  $k$ -term approximation*, J. Amer. Math. Soc. 22(1):211–231, 2009.
- [dCG12] Y. de Castro and F. Gamboa, *Exact reconstruction using beurling minimal extrapolation*, Journal of Mathematical Analysis and Applications, 395(1):336–354, 2012.
- [Don92] D. L. Donoho, *Superresolution via sparsity constraints*, SIAM Journal on Mathematical Analysis 23(5):1309–1331, 1992.
- [Don06a] D. L. Donoho, *Compressed sensing*, IEEE Trans. Inform. Theory, 52(4):1289–1306, 2006.
- [Don06b] D. L. Donoho, *For most large underdetermined systems of linear equations the minimal  $l_1$ -norm solution is also the sparsest solution*, Comm. Pure Appl. Math. 59(6):797–829, 2006.
- [DS89] D. L. Donoho and P. B. Stark, *Uncertainty principles and signal recovery*, SIAM Journal on Applied Mathematics, 49(3):906–931, 1989.
- [DG96] P. Doukhan and F. Gamboa, *Superresolution rates in Prokhorov metric*, Canad. J. Math. 48(2):316–329, 1996.
- [Dum07] B. Dumitrescu, *Positive trigonometric polynomials and signal processing applications*, Springer Verlag, 2007.
- [DP13] V. Duval and G. Peyré, *Exact support recovery for sparse spikes deconvolution*, arXiv:1306.6909, June 2013.
- [FK94] J. Faraut and A. Korányi, *Analysis on symmetric cones*, Clarendon Press Oxford, 1994.



- [Fat99] H. O. Fattorini, *Infinite dimensional optimization and control theory*, Cambridge University Press, 1999.
- [FG13] C. Fernández-Granda, *Support detection in super-resolution*, arXiv:1302.3921, Feb. 2013.
- [GG94] F. Gamboa and É. Gassiat, *The maximum entropy method on the mean: applications to linear programming and superresolution*, Math. Programming 66(1):103–122, 1994.
- [GG96] F. Gamboa and É. Gassiat, *Sets of superresolution and the maximum entropy method on the mean*, SIAM J. Math. Anal. 27(4):1129–1152, 1996.
- [Gas90] É. Gassiat, *Problèmes des moments et concentration de mesure*, C. R. Acad. Sci. Paris Sér. I Math. 310(2):41–44, 1990.
- [HGTB94] T. D. Harris, R. D. Grober, J. K. Trautman, and E. Betzig, *Super-resolution imaging spectroscopy*, Applied spectroscopy 48(1):14A–21A, 1994.
- [HG05] D. Henrion and A. Garulli (Editors), *Positive polynomials in control*, Springer Verlag, 2005.
- [HS09] M. A. Herman and T. Strohmer, *High-resolution radar via compressed sensing*, IEEE Trans. Signal Processing 57(6):2275–2284, 2009.
- [Kah11] J.-P. Kahane, *Idempotents et échantillonnage parcimonieux*, C. R. Math. Acad. Sci. Paris 349(19-20):1073–1076, 2011.
- [KS66] S. Karlin and W. J. Studden, *Tchebycheff systems: With applications in analysis and statistics*, Pure and Applied Mathematics, Vol. XV, Interscience Publishers John Wiley & Sons, New York-London-Sydney, 1966.
- [Kra68] N. N. Krasovskii, *Theory of motion control* (in Russian), Nauka, Moscow, 1968.
- [Las10] J. B. Lasserre, *Moments, positive polynomials and their applications*, Imperial College Press Optimization Series, vol. 1, Imperial College Press, London, 2010.
- [Lew96] A. S. Lewis, *Superresolution in the Markov moment problem*, J. Math. Anal. Appl. 197(3):774–780, 1996.
- [MM05] D. Makovoz and F. R. Marleau, *Point-source extraction with mopex*, Publications of the Astronomical Society of the Pacific, 117(836):1113–1128, 2005.
- [Neu64] L. W. Neustadt, *Optimization, a moment problem, and nonlinear programming*, SIAM J. Control, 2(1):33–53, 1964.
- [Nie13] J. Nie, *Certifying Convergence of Lasserre’s Hierarchy via Flat Truncation*, Mathematical Programming, 142(1-2):485–510, 2013.
- [Nie14] J. Nie, *Optimality Conditions and Finite Convergence of Lasserre’s Hierarchy*, Mathematical Programming, 146(1-2):97–121, 2014.

- [OBP94] J. W. Odendaal, E. Barnard, and C. W. I. Pistorius, *Two-dimensional super-resolution radar imaging using the music algorithm*, IEEE Trans. Antennas and Propagation 42(10):1386–1391, 1994.
- [PPK03] S. C. Park, M. K. Park, and M. G. Kang, *Super-resolution image reconstruction: a technical overview*, IEEE Signal Processing Magazine, 20(3):21–36, 2003.
- [Sch06] C. Scheiderer, *Sums of squares on real algebraic surfaces*, Manuscripta Mathematica 119(4):395–410, 2006.
- [SG87] J. Skilling and S. F. Gull, *Prior knowledge must be used*, Maximum-entropy and Bayesian spectral analysis and estimation problems (Laramie, Wyo., 1983), Fund. Theories Phys., Reidel, Dordrecht, 1987.
- [TBR13] G. Tang, B. Narayan Bhaskar, and B. Recht, *Near minimax line spectral estimation*, arXiv:1303.4348, Mar. 2013.
- [TG07] Z. Tian and G. B. Giannakis, *Compressed sensing for wideband cognitive radios*, IEEE Int. Conf. Acoustics, Speech and Signal Processing, 2007.
- [Ver12] N. Verzelen, *Minimax risks for sparse regressions: Ultra-high dimensional phenomena*, Electronic Journal of Statistics 6:38–90, 2012.
- [XCV+13] W. Xu, J.-F. Cai, K. Vijay Mishra, M. Cho, and A. Kruger, *Precise Semidefinite Programming Formulation of Atomic Norm Minimization for Recovering  $d$ -Dimensional ( $d \geq 2$ ) Off-the-Grid Frequencies*, arXiv 1312.0485, Dec. 2013.