IOP Publishing Inverse Problems
Inverse Problems **38** (2022) 125005 (19pp) https://doi.org/10.1088/1361-6420/ac9998

# On the uniqueness of solutions for the basis pursuit in the continuum

**Thomas Debarre**[1], **Quentin Denoyelle**[2,*] **and Julien Fageot**[3]

[1] EPFL, BIG, Switzerland
[2] Université Paris Cité, CNRS, MAP5, F-75006 Paris, France
[3] EPFL, LCAV, Switzerland

E-mail: thomas.debarre@epfl.ch, quentin.denoyelle@u-paris.fr and julien.fageot@epfl.ch

Received 4 March 2022, revised 12 September 2022
Accepted for publication 12 October 2022
Published 28 October 2022


**Abstract**

This paper studies the continuous-domain inverse problem of recovering Radon measures on the one-dimensional torus from low-frequency Fourier coefficients, where $K_c$ is the cutoff frequency. Our approach consists in minimizing the total-variation norm among all Radon measures that are consistent with the observations. We call this problem the basis pursuit in the continuum (BPC). We characterize the solution set of (BPC) in terms of uniqueness and describe its sparse solutions which are sums of few signed Dirac masses. The characterization is determined by the spectrum of a Toeplitz and Hermitian-symmetric matrix that solely depends on the observations. More precisely, we prove that (BPC) has a unique solution if and only if this matrix is neither positive definite nor negative definite. If it has both a positive and negative eigenvalue, then the unique solution is the sum of at most $2K_c$ Dirac masses, with at least one positive and one negative weight. If this matrix is positive (respectively negative) semi-definite and rank deficient, then the unique solution is composed of a number of Dirac masses equal to the rank of the matrix, all of which have nonnegative (respectively nonpositive) weights. Finally, in cases where (BPC) has multiple solutions, we demonstrate that there are infinitely many solutions composed of $K_c + 1$ Dirac masses, with nonnegative (respectively nonpositive) weights if the matrix is positive (respectively negative) definite.

Keywords: basis pursuit, total variation norm, uniqueness, Pisarenko decomposition, dual certificates

*Author to whom any correspondence should be addressed.

## 1. Introduction

In recent years, total-variation (TV)[4] regularization techniques have proved to be very fruitful to solve continuous-domain linear inverse problems with a sparsity prior. They provide a general framework for the recovery of sparse continuous-domain signals (e.g., Dirac streams or splines) from possibly corrupted finite-dimensional measurements. Such techniques rely on solid theoretical foundations [2–6], but also on many algorithmic advances [4, 7, 8], and have found various data-science applications [7, 9–11]. It is well known that their discrete-domain counterparts (i.e., $\ell^1$ regularization methods) lead to variational problems whose solutions are not necessarily unique [12]. Our goal in this paper is to provide a systematic study of the uniqueness and of the main properties of the solutions of the TV norm minimization problem when the low-frequency Fourier coefficients of the signal are prescribed.

### *1.1. Reconstruction via TV minimization*

Let $\mathbb{T} = \mathbb{R}/2\pi\mathbb{Z}$ be the 1D torus. We study the problem of recovering *real* Radon measures $w \in \mathcal{M}(\mathbb{T})$ defined on the domain $\mathbb{T}$ based on their low-frequency Fourier coefficients

$$\forall k \in \mathbb{Z},\ |k| \leqslant K_{\mathrm{c}}, \quad \widehat{w}[k] = \frac{1}{2\pi}\int_0^{2\pi} \mathrm{e}^{-\mathrm{i}kt}\,\mathrm{d}w(t) = y_k, \tag{1}$$

where $K_{\mathrm{c}} \in \mathbb{N}$ denotes the cutoff frequency and $\boldsymbol{y} = (y_{-K_{\mathrm{c}}}, \ldots, y_{-1}, y_0, y_1, \ldots, y_{K_{\mathrm{c}}}) \in \mathbb{C}^{2K_{\mathrm{c}}+1}$ are the observations. As the measures $w \in \mathcal{M}(\mathbb{T})$ are real, we must have $y_0 \in \mathbb{R}$ and for all $k \in \{1, \ldots, K_{\mathrm{c}}\}$, $y_k = \overline{y_{-k}}$. Therefore, the Toeplitz matrix

$$\mathbf{T}_{\boldsymbol{y}} = 2\pi \begin{pmatrix} y_0 & y_1 & \cdots & \cdots & y_{K_{\mathrm{c}}} \\ y_{-1} & y_0 & y_1 & \cdots & y_{K_{\mathrm{c}}-1} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ y_{-K_{\mathrm{c}}+1} & \cdots & y_{-1} & y_0 & y_1 \\ y_{-K_{\mathrm{c}}} & \cdots & \cdots & y_{-1} & y_0 \end{pmatrix} \in \mathbb{C}^{(K_{\mathrm{c}}+1)\times(K_{\mathrm{c}}+1)}, \tag{2}$$

which is at the core of our main contribution, is also Hermitian symmetric. Moreover, the observation vector $\boldsymbol{y} \in \mathbb{C}^{2K_{\mathrm{c}}+1}$ has only $2K_{\mathrm{c}}+1$ (real) degrees of freedom.

The recovery of a periodic Radon measure from finitely many observations is clearly an ill-posed problem. Therefore, we choose to formulate the reconstruction task as a regularized optimization problem using a sparsity prior. More precisely, in this paper, we consider the problem

$$\inf_{w\in\mathcal{M}(\mathbb{T}),\ \boldsymbol{\nu}(w)=\boldsymbol{y}} \|w\|_{\mathcal{M}}, \tag{3}$$

where $\boldsymbol{\nu} : \mathcal{M}(\mathbb{T}) \to \mathbb{C}^{2K_{\mathrm{c}}+1}$ is the measurement vector of Fourier coefficients

$$\forall w \in \mathcal{M}(\mathbb{T}), \quad \boldsymbol{\nu}(w) = (\widehat{w}[-K_{\mathrm{c}}], \ldots, \widehat{w}[0], \ldots, \widehat{w}[K_{\mathrm{c}}]), \tag{4}$$

and $\|\cdot\|_{\mathcal{M}}$ is the TV norm on the space of Radon measures $\mathcal{M}(\mathbb{T})$. The TV norm can be seen as an extension of the $\ell^1$ norm to the continuous domain. The choice of the TV norm promotes

[4] In this paper, the 'TV' is understood in the sense of measure theory and should not be confused with the concept of BV functions [1].

sparse continuous-domain reconstruction, and has recently received a lot of attention (see section 1.2). In the following, we name the problem (3), the *basis pursuit in the continuum* (BPC), in reference to its discrete-domain counterpart, the basis pursuit in [13].

### *1.2. Related works*

*Optimization over Radon measures*. The historical motivation to consider the TV norm as a regularization was to extend discrete $\ell^1$ regularization techniques, used in the theory of compressed sensing to recover sparse vectors [14–17], for the recovery of continuous-domain Dirac masses. The goal is to recover point sources, modeled as a sum of Dirac masses, from finitely many measurements. This question has received considerable attention in the 21st century, including methods that are not based on TV regularization, such as finite rate of innovation techniques [18] and Prony's methods [19, 20]. Several data-science problems can be formulated as a continuous-domain Dirac recovery problem, including radio-astronomy [21], super-resolution microscopy [7], or 3D image deconvolution [10].

The study of optimization problems over the space of Radon measures can be traced back to the pioneering works of Beurling [22], where Fourier-domain measurements were also considered. In the early 2010s, the works of De Castro and Gamboa [2], Candès and Fernandez-Granda [3, 23], and Bredies and Pikkarainen [4] considered optimization tasks of the form (3) (or its penalized version), with both theoretical analyses and novel algorithmic approaches to recover a sparse-measure solution, in the continuum [4, 7, 8, 10, 24–27]. The existence of sparse-measure solutions, i.e., solutions of the form $\sum_{k=1}^{K} a_k \delta_{x_k}$, where $K \in \mathbb{N}^*$, $a_k \in \mathbb{R}$, and $\delta_{x_k}$ is the Dirac mass at the location $x_k$, seems to have been proven for the first time in [22] and was later improved by Fisher and Jerome in [28]. Since then, a remarkable revival around TV optimization has occurred recently [8, 29–32].

*Algorithms for TV optimization*. The numerical resolution of optimization problems based on the TV norm has been largely studied in the literature. If we do not contribute to these computational aspects in this work, we briefly recall how they have been treated in practice. We can divide the different strategies to solve the BPC (3) numerically, or its penalized version (often called the BLASSO [4]), into three main approaches. A first one consists in discretizing the spatial domain, which converts the infinite-dimensional convex problem into a finite dimensional one, known as the basis pursuit [13] or the LASSO [33]. Many standard solvers exist to solve these problems, such as block-coordinate descent algorithms [34], homotopy algorithms [35], or proximal algorithms [36, 37]. A second one is based on reformulating the problem as a semidefinite program. This has been done for the BPC when the forward operator is a convolution with the ideal low-pass filter (which is the setup of this paper) in [23]. One can expect to recover the positions of the Dirac masses exactly since no discretization is performed. However, these techniques are limited to the case where the dual problem involves trigonometrical polynomials, and usually to the one-dimensional setting (although some extensions exist in higher dimensions [38, 39]). A last line of approaches tackles the BLASSO problem directly over the space of Radon measures. They involve for instance gradient descents and gradient flows [27], or the Frank–Wolfe algorithm [40] which leads to greedy methods [4, 7] that can achieve finite-time convergence in some cases [41].

*Uniqueness results for TV optimization*. It is well know that finite-dimensional $\ell^1$-regularization, of which the TV norm for Radon measures is the continuous-domain generalization, can lead to nonunique solutions [12, 42]. This is also the case for TV regularization. However, under some assumptions uniqueness may hold. Many uniqueness results for constrained or penalized TV-based optimization problems have thus been given in the literature, but from different perspectives than the one studied in this paper.

Usually, the underlying assumption is that the observations $\boldsymbol{y}$ are generated via a sparse measure $w_0$ and the question becomes whether it is possible to uniquely recover $w_0$, either exactly or in a stable way. In [2], de Castro and Gamboa introduced the concept of *extrema Jordan type measure* (see [2, definition 1]), which gives sufficient conditions on a given signed (with positive and negative weights) sparse measure to be the unique solution of a TV-based optimization problem. They also proved that when the input measure is nonnegative, $k$-sparse and the number of measurements is greater or equal than $2k+1$, then it is the unique solution of (3) if the measurement operator is defined from a T-system. Note that it has recently been proved [43–46] that in the nonnegativity setting, a $k$-sparse nonnegative measure can be uniquely recovered from at least $2k+1$ measurements and a nonnegativity constraint without the need for TV minimization.

Candès and Fernandez-Granda also studied the super-resolution problem of recovering a ground-truth sparse Radon measure $w_0$ from its low-frequency measurements. They have shown that if the minimal distance between the spikes of $w_0$ is large enough, then (3) has a unique solution, which is $w_0$ itself [3, theorem 1.2]. Duval and Peyré identified the so-called *nondegenerate source condition* [5, definition 5], under which the uniqueness of the reconstruction together with the recovery of the support of the underlying ground-truth sparse measure are shown. These results are based on the key notion of dual certificates, which also play an important role in our work. This notion has been introduced for discrete compressed sensing problems in [47] and connected to TV-based optimization problems in [2].

All these works are clearly related to this paper. However, the approach we propose here is different: we aim at characterizing the cases of uniqueness directly over the measurement vector $\boldsymbol{y}$, and we are agnostic to the ground-truth signal that generated it. The closest work in this direction is our recent publication [48]. We provide a full description of the solution set of non-periodic TV optimization problems with a regularization operator $\mathrm{D}^2$, where D is the derivative operator, and spatial sampling measurements.

*The problem of moments*. This problem, or its extension the *generalized problem of moments* [49, 50], is a classical one where one seeks to recover a measure from a sequence of its (potentially generalized) moments. This problem covers many applications in various fields, including Fourier-domain measurements, which are simply moments of trigonometrical polynomials. As a result, the tools developed in this domain can be harnessed, for example, to solve numerically (3) using semi-definite programming formulations by generalizing [3] to the multivariate case [38, 39]. Moreover, there are many existence results for the problem of moments. In [51, theorem 6.12], Curto and Fialkow prove the existence of a sparse nonnegative measure $w_0 \in \mathcal{M}(\mathbb{T})$ solution of the truncated trigonometric moment problem (as referred to in [51])

$$w \in \mathcal{M}(\mathbb{T}), \quad \boldsymbol{y} = \boldsymbol{\nu}(w), \tag{5}$$

if and only $\mathbf{T}_{\boldsymbol{y}}$ is positive semi-definite. Furthermore, solution $w_0$ has rank($\mathbf{T}_{\boldsymbol{y}}$) Dirac masses. Additionally, when $\mathbf{T}_{\boldsymbol{y}}$ is invertible, there exist infinitely many such $w_0$. When it is not, $w_0$ is unique. The contributions of [51] and their relation to ours are be discussed more precisely in remark 4.

### 1.3. Contributions

The existence of a solution to problem (3) is well established and can be obtained from the direct method in the calculus of variations. Moreover, it is known that there is always at least one sparse solution composed of at most $2K_{\mathrm{c}}+1$ Dirac masses [28]. However, the solution is in general not unique (the simplest case of nonuniqueness is with $K_{\mathrm{c}}=0$), and previous works

studying this question often start by assuming that the measurements come from a particular input signal. In this paper, we focus on

(a) Characterizing all the cases of uniqueness for (3);
(b) Describing the sparse solutions, e.g. the bounds on the number of Dirac masses and the signs of the weights;

from simple conditions depending *only* on the measurement vector $\boldsymbol{y}$, involving in particular the matrix $\mathbf{T}_{\boldsymbol{y}}$. The novelty of our approach thus lies in the fact that we are able to provide relatively deep information on the form of the sparse solutions of (3) and answer in all cases the question of uniqueness agnostically to the input signal that generated $\boldsymbol{y}$.

Our first contribution, theorem 1 in section 3.1, provides several equivalent conditions for the solution set of (3) to be composed of only nonnegative measures. We prove that if one of these conditions is not satisfied *and* $y_0 \geqslant 0$, then there is a unique solution composed of at most $2K_{\mathrm{c}}$ Dirac masses, with at least one positive and one negative weight. One of these equivalent conditions is used, in corollary 1, to formulate a simple criterion on the magnitudes of the entries of $\boldsymbol{y}$ which is a sufficient condition for uniqueness. Theorem 1 can of course be readily adapted to the nonpositive case.

Theorem 1 does not cover all the situations that may arise since it does not adjudicate the uniqueness of a solution to (3) when the solution set is composed of only nonnegative (or nonpositive) measures. This limitation is tackled in our main contribution, theorem 2, in section 3.2, which we state below. We recall that the Toeplitz matrix $\mathbf{T}_{\boldsymbol{y}}$ is given in (2).

**Theorem 1** *The solution set of problem* (3) *can be characterized as follows:*

(*a*) *If* $\mathbf{T}_{\boldsymbol{y}}$ *has at least one negative and one positive eigenvalue, then* (3) *has a unique solution composed of at most* $2K_{\mathrm{c}}$ *Dirac masses, with at least one positive and one negative weight;*
(*b*) *If* $\mathbf{T}_{\boldsymbol{y}}$ *is positive, respectively negative, semi-definite and* $\operatorname{rank}(\boldsymbol{T}_{\boldsymbol{y}}) < K_{\mathrm{c}} + 1$, *then* (3) *has a unique solution composed of exactly* $\operatorname{rank}(\boldsymbol{T}_{\boldsymbol{y}})$ *positive, respectively negative, Dirac masses;*
(*c*) *If* $\mathbf{T}_{\boldsymbol{y}}$ *is positive, respectively negative, definite* (*which implies that* $\operatorname{rank}(\boldsymbol{T}_{\boldsymbol{y}}) = K_{\mathrm{c}} + 1$), *then* (3) *has infinitely many solutions, and among sparse solutions none with less than* $K_{\mathrm{c}} + 1$ *Dirac masses and uncountably many of them composed of* $K_{\mathrm{c}} + 1$ *positive, respectively negative, Dirac masses.*

This theorem provides information on the form of the sparse solutions, from conditions on the spectrum of the Hermitian Toeplitz matrix $\mathbf{T}_{\boldsymbol{y}}$. Moreover, all possible scenarios are covered since all vectors $\boldsymbol{y} \in \mathbb{R} \times \mathbb{C}^{K_{\mathrm{c}}}$ fall in one single case of the theorem. Therefore, it also yields a final answer to the question of uniqueness.

In the case when $\mathbf{T}_{\boldsymbol{y}}$ is positive (or negative) semi-definite and rank deficient, our result can be derived from [51, theorem 6.12], once we admit that any nonnegative (or nonpositive) measures that match the observations is a solution of (3).[5] Our contributions are therefore closely related to this literature. However our framework is also different, and one of the novelty of our work lies in the bridges made between these related results in order to answer the question of uniqueness for (BPC). Consequently, we also choose to present, as much as possible, our contributions in a self-contained fashion. The proofs are based upon known tools from the field of point-source recovery:

[5] For a more detailed discussion on this matter, see remark 4.

- Dual certificates (see proposition 1);
- The Herglotz theorem (see proposition 4), which characterizes positive sparse measures in terms of their Fourier coefficients;
- The Carathéodory–Pisarenko–Fejér (CFP) decomposition [52–54] (see proposition 7), which leads to algorithms such as MUSIC [19].

## 2. Mathematical preliminaries

In section 2.1, we introduce the mathematical background of this paper. We also present the optimization problem of interest. In section 2.2, we then remind the main tools from duality theory that we use for studying the TV minimization problem.

### 2.1. Periodic Radon measures and BPC

*Radon measures*. Let $\mathcal{M}(\mathbb{T})$ be the space of periodic *real* Radon measures. By the Riesz–Markov theorem [55], $\mathcal{M}(\mathbb{T})$ is the continuous dual of the space $\mathcal{C}(\mathbb{T})$ of continuous periodic real functions endowed with the supremum norm $\|\cdot\|_\infty$. The TV norm on $\mathcal{M}(\mathbb{T})$ is the dual norm associated to $(\mathcal{C}(\mathbb{T}), \|\cdot\|_\infty)$ and is thus given by

$$\|w\|_{\mathcal{M}} = \sup_{\varphi\in\mathcal{C}(\mathbb{T}),\ \|\varphi\|_\infty\leqslant 1} \langle w,\ \varphi\rangle. \tag{6}$$

The normed space $(\mathcal{M}(\mathbb{T}), \|\cdot\|_{\mathcal{M}})$ is then a Banach space. The duality product between $w \in \mathcal{M}(\mathbb{T})$ and $\varphi \in \mathcal{C}(\mathbb{T})$ is given by

$$\langle w, \varphi\rangle = \frac{1}{2\pi}\int_0^{2\pi} \varphi(t)\mathrm{d}w(t). \tag{7}$$

The TV norm can be seen as the extension to the continuum of the $\ell^1$ norm for vectors, as for any sparse measure

$$w_{\boldsymbol{a},\boldsymbol{x}} = \sum_{k=1}^{K} a_k\delta_{x_k}, \quad \text{with} \quad K \in \mathbb{N}^*, \forall k \in \{1,\dots,K\},\ a_k \in \mathbb{R},\ x_k \in \mathbb{T}, \tag{8}$$

where the locations of the Dirac masses $x_k$ are pairwise distinct, we have

$$\|w_{\boldsymbol{a},\boldsymbol{x}}\|_{\mathcal{M}} = \frac{1}{2\pi}\|\boldsymbol{a}\|_1. \tag{9}$$

We also consider the set of nonnegative Radon measures $\mathcal{M}_+(\mathbb{T})$, which are Radon measures $w$ such that $\langle w,\ \varphi\rangle \geqslant 0$ for any positive continuous function $\varphi$ on $\mathbb{T}$. Similarly, we define the set of nonpositive Radon measures as $\mathcal{M}_-(\mathbb{T})$.

*Forward operator*. For any $k \in \mathbb{Z}$, we define $e_k : \mathbb{T} \to \mathbb{C}$ as $e_k(t) = \mathrm{e}^{\mathrm{i}kt}$. We consider the measurement operator given by

$$\boldsymbol{\nu} : w \in \mathcal{M}(\mathbb{T}) \mapsto (\widehat{w}[-K_\mathrm{c}], \dots, \widehat{w}[0], \dots \widehat{w}[K_\mathrm{c}]) \in \mathbb{C}^{K_\mathrm{c}} \times \mathbb{R} \times \mathbb{C}^{K_\mathrm{c}}, \tag{10}$$

where $K_\mathrm{c} \in \mathbb{N}$ is the cutoff frequency and for all $|k| \leqslant K_\mathrm{c}$,

$$\widehat{w}[k] = \frac{1}{2\pi}\int_0^{2\pi} \mathrm{e}^{-\mathrm{i}kt}\,\mathrm{d}w(t) = \langle w, e_k\rangle \tag{11}$$

is the $k$th Fourier coefficient of $w$. Note that we extended the duality product (7) to *complex* test functions $\varphi : \mathbb{R} \to \mathbb{C}$ as $\langle w, \varphi \rangle = \frac{1}{2\pi}\int_0^{2\pi} \overline{\varphi(t)} \mathrm{d}w(t)$. The operator $\boldsymbol{\nu}$ is a low-pass operator that only keeps the low-frequency Fourier coefficients.

*Optimization problem*. Let $\boldsymbol{y} = (y_{-K_c}, \ldots, y_0, \ldots, y_{K_c}) \in \mathbb{C}^{2K_c+1}$ such that[6]

$$y_0 \in \mathbb{R} \quad \text{and} \quad \forall k \in \{1, \ldots, K_c\},\ y_k = \overline{y_{-k}}, \tag{12}$$

be a given observations vector. We aim to solve the linear inverse problem $\boldsymbol{\nu}(w) = \boldsymbol{y}$. We introduce our terminology for solutions of this problem in the following definition.

**Definition 1.** Let $\boldsymbol{y} \in \mathbb{C}^{2K_c+1}$ given as in equation (12). We say that $w \in \mathcal{M}(\mathbb{T})$ is a *representing measure for* $\boldsymbol{y}$, if $w$ satisfies $\boldsymbol{\nu}(w) = \boldsymbol{y}$.

This linear inverse problem is ill-posed, as it has infinitely many representing measures. To tackle this issue, we choose to favor sparse solutions. Our approach consists in solving the following optimization problem

$$\min_{w \in \mathcal{M}(\mathbb{T}),\ \boldsymbol{\nu}(w) = \boldsymbol{y}} \|w\|_{\mathcal{M}}. \tag{BPC}$$

A solution of (BPC) (which is known to exist) therefore has the minimal TV norm among all representing measures for $\boldsymbol{y}$. As the TV norm is an extension of the $\ell^1$ norm, it is known to promote sparse solutions (composed of a sum of Dirac masses). We denote this problem the BPC as a tribute to its finite-dimensional counterpart, the basis pursuit [13].

### 2.2. Dual certificates for the BPC

The analysis of (BPC) benefits from the theory of duality for infinite-dimensional convex optimization, as exposed for instance by Ekeland and Temam in [56]. This line of research has proven to be fruitful to study optimization on measure spaces [2, 5, 23, 30, 48, 57]. We mostly rely on the concepts and results exposed in [5, 48], but very similar tools can be found elsewhere [2, 23]. Considering the dual problem to (BPC) and writing the optimality conditions that link the solutions of both problems[7] leads to the notion of *dual certificates*, which are continuous functions on $\mathbb{T}$ satisfying some conditions (see proposition 1 below). In particular, dual certificates enable to *certify* that some $w \in \mathcal{M}(\mathbb{T})$ is a solution of (BPC) and to localize its support. In the next definition, we introduce some notations that ease the related statements. We recall that a Radon measure $w \in \mathcal{M}(\mathbb{T})$ can be uniquely decomposed as $w = w_+ - w_-$, where $w_+$ and $w_-$ are nonnegative measures (Jordan decomposition).

**Definition 2.** Let $w \in \mathcal{M}(\mathbb{T})$. We define the *signed support* of $w$ as

$$\mathrm{supp}_{\pm}(w) = \mathrm{supp}(w_+) \times \{1\} \cup \mathrm{supp}(w_-) \times \{-1\}, \tag{13}$$

where $\mathrm{supp}(\tilde{w})$ is the support of $\tilde{w} \in \mathcal{M}(\mathbb{T})$.

Let $\eta \in \mathcal{C}(\mathbb{T})$. The positive and negative saturation sets of $\eta$ are given by

$$\mathrm{sat}_+(\eta) = \eta^{-1}(\{1\}) \quad \text{and} \quad \mathrm{sat}_-(\eta) = \eta^{-1}(\{-1\}), \tag{14}$$

respectively. Finally, we define the *signed saturation set* of $\eta$ as

$$\mathrm{sat}_{\pm}(\eta) = \mathrm{sat}_+(\eta) \times \{1\} \cup \mathrm{sat}_-(\eta) \times \{-1\}. \tag{15}$$

[6] These requirements on the observations vector $\boldsymbol{y}$ come from the fact that otherwise, the equation $\boldsymbol{\nu}(w) = \boldsymbol{y}$ has no solution in $\mathcal{M}(\mathbb{T})$, since $w$ is a real Radon measure.

[7] Note that dual certificates always exist for the (BPC) with Fourier measurements [3].

Now, we give two results which help us to characterize the cases of uniqueness for (BPC), in our main contribution theorem 2. The next proposition introduces formally the notion of dual certificates for the (BPC) problem.

**Proposition 1.** *There exists a function $\eta \in \mathcal{C}(\mathbb{T})$, which is a real trigonometrical polynomial of degree at most $K_c$, satisfying $\|\eta\|_\infty \leqslant 1$ and such that for any solution $w_0 \in \mathcal{M}(\mathbb{T})$ of (BPC), we have one of the following equivalent conditions:*

- $\langle w_0, \eta \rangle = \|w_0\|_{\mathcal{M}}$;
- $\mathrm{supp}_\pm(w_0) \subset \mathrm{sat}_\pm(\eta)$.

*Such a function $\eta$ is denoted as a dual certificate.*

Proposition 1 is stated in some other, nonetheless equivalent, form in [5, proposition 3], where dual certificates of the optimization problem (BPC) are studied[8]. A key role is played by the adjoint operator $\boldsymbol{\nu}^* : \mathbb{C}^{K_c} \times \mathbb{R} \times \mathbb{C}^{K_c} \to \mathcal{C}(\mathbb{T})$ (denoted by $\Phi^*$ in [5]), due to the fact that dual certificates must be in $\mathrm{Im}(\boldsymbol{\nu}^*)$. In our setup, we have that

$$\boldsymbol{\nu}^*(c_{-K_c}, \ldots, c_{-1}, c_0, c_1, \ldots, c_{K_c}) = \sum_{|k| \leqslant K_c} c_k e_{-k}. \tag{16}$$

The role of real trigonometric polynomials in proposition 1 is explained by the fact that the vectors $\boldsymbol{c} \in \mathbb{C}^{K_c} \times \mathbb{R} \times \mathbb{C}^{K_c}$ involved in our study are Hermitian symmetric because they belong to $\mathrm{Im}(\boldsymbol{\nu})$. We do not provide a detailed proof of proposition 1, since it has already been exposed elsewhere. It is for instance done in [48, propositions 1 and 2] in a different setting, but the arguments can be readily adapted.

An important consequence of proposition 1 for the study of uniqueness of (BPC), is the following proposition that can also be deduced from [5] (see also [2]).

**Proposition 2.** *If there exists a nonconstant[9] dual certificate for the problem (BPC), then it has a unique solution of the form $w = \sum_{k=1}^{K} a_k \delta_{x_k}$ with $K \leqslant 2K_c$ and $a_k \in \mathbb{R}$, $x_k \in \mathbb{T}$.*

For the sake of completeness, a proof of proposition 2 can be found in appendix A.

## 3. Toeplitz characterization of (BPC)

In this section, we present our contributions. We recall that all the main notations are introduced in section 2.1. In particular, the measurement vector $\boldsymbol{y} \in \mathbb{C}^{K_c} \times \mathbb{R} \times \mathbb{C}^{K_c}$ is given in equation (12). Theorem 1 in section 3.1 first provides several equivalent conditions which ensure that the solution set of (BPC) is solely composed of nonnegative (or nonpositive) measures. This can be proved not to hold when a simple criterion on the coefficient values of $\boldsymbol{y}$ is satisfied, see corollary 1. By the second part of theorem 1, this leads to the existence of a unique sparse solution composed of at most $2K_c$ Dirac masses, with at least one positive and one negative weight. Theorem 1 paves the way towards our main contribution, theorem 2 in section 3.2, which characterizes the solution set of (BPC) from simple condition on the spectrum of the Toeplitz and Hermitian symmetric matrix $\mathbf{T}_y$ introduced in (2). In particular, it gives all the cases where uniqueness holds, see corollary 2.

[8] Duval and Peyré consider more general measurement operators whose image can lie in a Hilbert space and exemplify their results for low-frequency measurements.

[9] In the more general case studied in [5], this corresponds to the nondegeneracy condition of the dual certificate.

### *3.1. A new criterion of uniqueness for (BPC)*

Before stating the main results of this section, let us first prove, in the next lemma, that the TV norm of a Radon measure upper-bounds its Fourier coefficients. Lemma 1 also provides elementary characterizations for nonnegative Radon measures that we use in theorem 1.

**Lemma 1.** *Let $w \in \mathcal{M}(\mathbb{T})$. Then,*

*(a) For any $k \in \mathbb{Z}$, we have $|\widehat{w}[k]| \leqslant \|w\|_{\mathcal{M}}$;*
*(b) $w \in \mathcal{M}_+(\mathbb{T})$ if and only if $\|w\|_{\mathcal{M}} = \widehat{w}[0]$;*

**Proof.** Let $k \geqslant 0$. Any $w \in \mathcal{M}(\mathbb{T})$ can be uniquely decomposed as $w = w_+ - w_-$ where $w_+, w_- \in \mathcal{M}_+(\mathbb{T})$ (Jordan decomposition). We then observe that

$$\begin{aligned} |\widehat{w}[k]| = |\langle w, e_k\rangle| &= |\langle w_+, e_k\rangle - \langle w_-, e_k\rangle| \leqslant |\langle w_+, e_k\rangle| + |\langle w_-, e_k\rangle| \\ &\leqslant \frac{1}{2\pi}\int_{\mathbb{T}} |e_k(t)|\mathrm{d}w_+(t) + \frac{1}{2\pi}\int_{\mathbb{T}} |e_k(t)|\mathrm{d}w_-(t) = \langle w_+, 1\rangle + \langle w_-, 1\rangle \\ &= \|w\|_{\mathcal{M}}. \end{aligned} \tag{17}$$

For (b), we observe that $\|w\|_{\mathcal{M}} = \|w_+\|_{\mathcal{M}} + \|w_-\|_{\mathcal{M}} = \langle w_+, 1\rangle + \langle w_-, 1\rangle = \widehat{w}_+[0] + \widehat{w}_-[0]$. Then,

$$w \in \mathcal{M}_+(\mathbb{T}) \;\Leftrightarrow\; w_- = 0 \;\Leftrightarrow\; \widehat{w}_-[0] = 0 \;\Leftrightarrow\; \|w\|_{\mathcal{M}} = \widehat{w}_+[0] = \widehat{w}[0], \tag{18}$$

which concludes the proof.

Next, we provide a lower bound on the minimal value of (BPC).

**Lemma 2.** *We have that*

$$\min_{w\in\mathcal{M}(\mathbb{T}),\ \boldsymbol{\nu}(w)=\boldsymbol{y}} \|w\|_{\mathcal{M}} \geqslant \max_{0\leqslant k\leqslant K_c} |y_k|. \tag{19}$$

**Proof.** We know that (BPC) has at least one solution $w_0 \in \mathcal{M}(\mathbb{T})$ that reaches its minimum value. Then, according to lemma 1, we have $\|w_0\|_{\mathcal{M}} \geqslant |\widehat{w_0}[k]| = |y_k|$ for all $-K_c \leqslant k \leqslant K_c$, which yields (19).

We can now state the main result of this section.

**Theorem 1.** *Let $\boldsymbol{y}$ be as in (12). We define $\varepsilon = \varepsilon(y_0) \in \{+, -\}$ such that $\varepsilon$ is $+$ if $y_0 \geqslant 0$, and $\varepsilon$ is $-$ if $y_0 < 0$. Then, the following conditions are equivalent:*

*(a) There exists $w_0 \in \mathcal{M}_\varepsilon(\mathbb{T})$ such that $\boldsymbol{\nu}(w_0) = \boldsymbol{y}$;*
*(b) The solution set of (BPC) is $\{w \in \mathcal{M}_\varepsilon(\mathbb{T}) :\ \boldsymbol{\nu}(w) = \boldsymbol{y}\}$;*
*(c) We have the following equality:*

$$|y_0| = \min_{w\in\mathcal{M}(\mathbb{T}),\ \boldsymbol{\nu}(w)=\boldsymbol{y}} \|w\|_{\mathcal{M}}. \tag{20}$$

*If the above equivalence is not satisfied, then (BPC) has a unique solution, composed of at most $2K_c$ Dirac masses, with at least one positive and one negative weight.*

**Proof.** Let us first prove that the items (a), (b) and (c) are equivalent. We assume that $y_0 \geqslant 0$. The case $y_0 < 0$ is similar.

(a) $\Rightarrow$ (c). Set $m = \min_{w\in\mathcal{M}(\mathbb{T}),\ \boldsymbol{\nu}(w)=\boldsymbol{y}} \|w\|_{\mathcal{M}}$. Let $w_0 \in \mathcal{M}_+(\mathbb{T})$ such that $\boldsymbol{\nu}(w_0) = \boldsymbol{y}$. By lemma 1, we have, for any $w \in \mathcal{M}(\mathbb{T})$, the equivalence

$$w \in \mathcal{M}_+(\mathbb{T}) \iff \|w\|_{\mathcal{M}} = \widehat{w}[0]. \tag{21}$$

As a result, we get that $m \leqslant \|w_0\|_{\mathcal{M}} = \widehat{w_0}[0] = y_0$. Finally, from lemma 2, we have that $m \geqslant y_0$, which yields $m = y_0$[10].

(c) $\Rightarrow$ (b). Let $w_0 \in \mathcal{M}(\mathbb{T})$ be a solution of (BPC) which is equivalent to $\min_{w\in\mathcal{M}(\mathbb{T}),\ \boldsymbol{\nu}(w)=\boldsymbol{y}} \|w\|_{\mathcal{M}} = \|w_0\|_{\mathcal{M}}$. Thus, by (c), this last equality can be equivalently rewritten as $y_0 = \|w_0\|_{\mathcal{M}}$, i.e. $\widehat{w_0}[0] = \|w_0\|_{\mathcal{M}}$. By lemma 1, it is equivalent to $w_0 \in \mathcal{M}_+(\mathbb{T})$. As $w_0$ satisfies the constraints $\boldsymbol{\nu}(w_0) = \boldsymbol{y}$, we finally obtain that the solution set of (BPC) is equal to $\{w \in \mathcal{M}_+(\mathbb{T}) :\ \boldsymbol{\nu}(w) = \boldsymbol{y}\}$.

(b) $\Rightarrow$ (a). This is a direct consequence of the fact that the solution set of (BPC) is non-empty.

Next, if the equivalence is not satisfied, this implies that there is no solution that is a non-negative or a nonpositive measure. Let $w_0 \in \mathcal{M}(\mathbb{T})$ be a solution and $\eta$ a dual certificate for problem (BPC), as given by proposition 1. Using proposition 2, we know that if the dual certificate is nonconstant, then the solution of (BPC) is unique and composed of at most $2K_c$ Dirac masses, with at least one negative and one positive weight. Let us assume by contradiction that $\eta$ is constant. As $w_0 \neq 0$, we have $\mathrm{supp}_{\pm}(w_0) \neq \emptyset$, hence $\mathrm{sat}_{\pm}(\eta) \neq \emptyset$. Since $\eta$ is constant, we thus have $\mathrm{sat}_{\pm}(\eta) = \mathbb{T} \times \{1\}$ or $\mathrm{sat}_{\pm}(\eta) = \mathbb{T} \times \{-1\}$, hence $w_0$ is nonnegative or nonpositive. This contradicts our initial assumption, which concludes the proof.

**Remark 1.** Theorem 1 shows that there exists a nonnegative representing measure for $\boldsymbol{y}$ if and only if the solution set of (BPC) is composed of all the nonnegative representing measures for $\boldsymbol{y}$. This suggests that minimizing the TV norm plays no role in this context and can be replaced by a simple nonnegativity constraint. This is consistent with recent results [43–46] which proved, in different setups, that when the observations $\boldsymbol{y}$ are generated by a nonnegative *sparse* measure, then replacing the TV norm with a nonnegativity constraint is enough to uniquely recover the input measure, provided that the number of measurements is sufficient. It is worth noting that, contrary to the cited works, we do not assume in theorem 1 that $\boldsymbol{y}$ is generated by a *sparse* measure that we would like to recover.

More importantly, theorem 1 proves that when the equivalence is not satisfied, then we must be in a case of uniqueness, which is, to the best of our knowledge, a new result.

From theorem 1 and lemma 2, we can deduce a simple criterion on the magnitude of the coefficients of $\boldsymbol{y}$ which ensures that the solution of (BPC) is unique. This criterion appears to be practically fruitful since uniqueness follows trivially for a large class of observation vectors $\boldsymbol{y}$.

**Corollary 1.** *If* $|y_0| < |y_{k_0}|$, *for some* $k_0 \neq 0$, *then* (BPC) *has a unique solution, composed of at most* $2K_c$ *Dirac masses, with at least one positive and one negative weight.*

**Proof.** By lemma 2, we have $\min_{w\in\mathcal{M}(\mathbb{T}),\ \boldsymbol{\nu}(w)=\boldsymbol{y}} \|w\|_{\mathcal{M}} \geqslant |y_{k_0}| > |y_0|$. This implies that the equivalent conditions of theorem 1 do not hold, as item (c) is not valid. We are thus in the uniqueness scenario of theorem 1.

[10] Note that we also proved that $w_0$ is solution of (BPC).

### *3.2. Characterization of solutions of (BPC) using $\mathbf{T}_y$*

In this section, we prove, in theorem 2, that the sign of the solutions of (BPC) is directly related to spectral properties of the matrix $\mathbf{T}_y$ formed from the observation vector $y$ defined as

$$\mathbf{T}_y = 2\pi \begin{pmatrix} y_0 & y_1 & y_2 & \cdots & y_{K_c} \\ y_{-1} & y_0 & y_1 & \cdots & y_{K_c-1} \\ y_{-2} & y_{-1} & y_0 & \cdots & y_{K_c-2} \\ \vdots & & & \ddots & \\ y_{-K_c} & y_{-K_c+1} & y_{-K_c+2} & \cdots & y_0 \end{pmatrix}. \tag{22}$$

More precisely, as $\mathbf{T}_y$ is Hermitian symmetric, one of the following three statements must hold:

- $\mathbf{T}_y$ has at least one negative and one positive eigenvalue;
- $\mathbf{T}_y$ is positive or negative semi-definite and rank deficient;
- $\mathbf{T}_y$ is positive or negative-definite.

In theorem 2, we prove that each of these scenarios leads to different properties of the solution set of (BPC), all expressed in terms of the nature of the sign of the solutions and of uniqueness. To this end, we first characterize the existence of a nonnegative representing measure for $y$ in terms of positive semi-definiteness of the matrix $\mathbf{T}_y$ in proposition 3. It can be obtained from [52, theorem 6.12], but we include for the sake of completeness.

**Proposition 3.** *The two following conditions are equivalent:*

(*a*) *There exists* $w_0 \in \mathcal{M}_+(\mathbb{T})$ *such that* $\boldsymbol{\nu}(w_0) = \boldsymbol{y}$*;*
(*b*) *The Hermitian matrix* $\mathbf{T}_y \in \mathbb{C}^{(K_c+1)\times(K_c+1)}$*, defined in* (22)*, is positive semi-definite.*

**Proof.** (a) $\Rightarrow$ (b). Let $w_0$ be a nonnegative representing measure for $y$. Consequently, by proposition 4 in appendix B, we have that $\sum_{k,\ell\in\mathbb{Z}} \widehat{w_0}[k-\ell] z_k \overline{z}_\ell \geqslant 0$, for any complex sequence $(z_k)_{k\in\mathbb{Z}}$ with finitely many nonzero terms. In particular, restricting to sequences such that $z_k = 0$ for $k < 0$ and $k > K_c$, we have that

$$\sum_{k,\ell=0}^{K_c} \widehat{w_0}[k-\ell] z_k \overline{z}_\ell \geqslant 0. \tag{23}$$

Since $\widehat{w_0}[k-\ell] = y_{k-\ell}$ for all $k, \ell \in \{0, \ldots, K_c\}$, we deduce that $\mathbf{T}_y$ is positive semi-definite.

(b) $\Rightarrow$ (a). We denote $\mathcal{P}_{K_c}(\mathbb{T}) = \mathrm{Span}(e_{-K_c}, e_{-K_c+1}, \ldots, e_{K_c})$ (where we recall that $e_k(t) = \mathrm{e}^{\mathrm{i}kt}$) the space of real trigonometric polynomials of degree at most $K_c$. Let us consider the linear mapping $\Phi : \mathcal{P}_{K_c}(\mathbb{T}) \to \mathbb{R}$ such that for all $|k| \leqslant K_c$, $\Phi(e_k) = y_k$. Then, $\Phi$ must be positive. Indeed, let $p \in \mathcal{P}_{K_c}(\mathbb{T})$ such that $p \geqslant 0$. According to proposition 5 in appendix B, $p$ can be written as $p = |q|^2$ for some complex trigonometric polynomial $q = \sum_{k=0}^{K_c} z_k e_k$ with $z_k \in \mathbb{C}$. This implies that $p = \sum_{k,\ell=0}^{K_c} z_k \overline{z}_\ell e_{k-\ell}$ and therefore that

$$\Phi(p) = \sum_{k,\ell=0}^{K_c} y_{k-\ell} z_k \overline{z}_\ell = \frac{1}{2\pi} \langle \mathbf{T}_y \boldsymbol{z}, \boldsymbol{z} \rangle \geqslant 0 \quad \text{where} \quad \boldsymbol{z} = (z_0, \ldots, z_{K_c}) \in \mathbb{C}^{K_c+1}. \tag{24}$$

Then, according to proposition 6 in appendix B, $\Phi$ can be extended to a positive, linear, and continuous functional from $\mathcal{C}(\mathbb{T})$ to $\mathbb{R}$. This implies by the Riesz–Markov theorem [55]

that there exists $w_0 \in \mathcal{M}_+(\mathbb{T})$ such that $\Phi = w_0$. Moreover, by construction of $\Phi$, $w_0$ satisfies $\boldsymbol{\nu}(w_0) = \boldsymbol{y}$. Consequently, item (a) is proved.

**Remark 2.** Proposition 3 can readily be adapted to the nonpositive case, since there exists a nonpositive representing measure for $\boldsymbol{y}$ if and only if $\mathbf{T}_{\boldsymbol{y}}$ is negative semi-definite.

Proposition 3 gives another equivalent condition to item (a) of theorem 1 which involves the matrix $\mathbf{T}_{\boldsymbol{y}}$. Building on this result, the statement on the nature of the solutions of (BPC) can be refined by leveraging the fact that $\mathbf{T}_{\boldsymbol{y}}$ is a Toeplitz matrix. Indeed, it is well known that a positive semi-definite and rank deficient Toeplitz matrix can be uniquely decomposed as the product $\mathbf{VDV}^*$, where $\mathbf{V}$ is a Vandermonde matrix whose columns are given by complex exponentials, $\mathbf{D}$ is a positive-definite diagonal matrix whose size is the rank of the starting Toeplitz matrix and $V^*$ is the conjugate transpose of $V$. This is the Pisarenko decomposition [52–54], also known as the CFP decomposition. This result is recalled in appendix C, for the sake of completeness. The next lemma is then the last missing piece towards theorem 2; it relates this decomposition of the Toeplitz matrix $\mathbf{T}_{\boldsymbol{y}}$ to the existence of a sparse representing measure for $\boldsymbol{y}$.

**Lemma 3.** *For all* $0 \leqslant K \leqslant K_{\mathrm{c}} + 1, \boldsymbol{a} = (a_1, \ldots, a_K) \in (\mathbb{R}_{>0})^K$ *and* $\boldsymbol{x} = (x_1, \ldots, x_K) \in \mathbb{T}^K$ *with pairwise distinct entries, the following statements are equivalent:*

(*a*) $\boldsymbol{y} = \boldsymbol{\nu}(w_{\boldsymbol{a},\boldsymbol{x}})$ *with* $w_{\boldsymbol{a},\boldsymbol{x}} = \sum_{k=1}^{K} a_k \delta_{x_k}$;

(*b*) $\mathbf{T}_{\boldsymbol{y}} = \mathbf{V}_{\boldsymbol{x}} \mathbf{D}_{\boldsymbol{a}} \mathbf{V}_{\boldsymbol{x}}^*$ *where* $\mathbf{V}_{\boldsymbol{x}} \in \mathbb{C}^{(K_c+1)\times K}$ *is the Vandermonde matrix whose kth column is given by* $\boldsymbol{e}_{\boldsymbol{K_c}}(x_k) = (1 \ \mathrm{e}^{\mathrm{i}x_k} \ \ldots \ \mathrm{e}^{\mathrm{i}K_c x_k})^T \in \mathbb{R} \times \mathbb{C}^{K_c}$ *and* $\mathbf{D}_{\boldsymbol{a}} \in \mathbb{R}^{K\times K}$ *the diagonal matrix with the entries of* $\boldsymbol{a}$ *on the diagonal.*

**Proof.** Let $K \leqslant K_{\mathrm{c}} + 1, \boldsymbol{a} \in (\mathbb{R}_{>0})^K$, and $\boldsymbol{x} \in \mathbb{T}^K$ be as in the statement of the lemma.

First, suppose that (a) holds. Then, for any $0 \leqslant m \leqslant K_{\mathrm{c}}$ and $0 \leqslant n \leqslant K_{\mathrm{c}} - m$, we have

$$y_m = \frac{1}{2\pi} \sum_{k=1}^{K} a_k \mathrm{e}^{-\mathrm{i}mx_k} = \frac{1}{2\pi} \sum_{k=1}^{K} a_k \mathrm{e}^{\mathrm{i}nx_k} \mathrm{e}^{-\mathrm{i}(m+n)x_k} = \frac{1}{2\pi} \left\langle (\mathrm{e}^{\mathrm{i}nx_k})_{1\leqslant k\leqslant K}, \, (a_k \mathrm{e}^{\mathrm{i}(m+n)x_k})_{1\leqslant k\leqslant K} \right\rangle. \tag{25}$$

We notice that $(\mathrm{e}^{\mathrm{i}nx_k})_{1\leqslant k\leqslant K}$ is the $(n+1)$th row of the matrix $\mathbf{V}_{\boldsymbol{x}}$ and $(a_k \mathrm{e}^{-\mathrm{i}(m+n)x_k})_{1\leqslant k\leqslant K}$ the $(m+n+1)$th column of $\mathbf{D}_{\boldsymbol{a}} \mathbf{V}_{\boldsymbol{x}}^*$. Therefore, $2\pi y_m$ is the $(n+1, m+n+1)$ entry of the matrix $\mathbf{V}_{\boldsymbol{x}} \mathbf{D}_{\boldsymbol{a}} \mathbf{V}_{\boldsymbol{x}}^*$. Hence, the elements of the $m$th upper diagonal of $\mathbf{V}_{\boldsymbol{x}} \mathbf{D}_{\boldsymbol{a}} \mathbf{V}_{\boldsymbol{x}}^*$ are $2\pi y_m$, and since $\mathbf{V}_{\boldsymbol{x}} \mathbf{D}_{\boldsymbol{a}} \mathbf{V}_{\boldsymbol{x}}^*$ is Hermitian symmetric, we get that $\mathbf{V}_{\boldsymbol{x}} \mathbf{D}_{\boldsymbol{a}} \mathbf{V}_{\boldsymbol{x}}^* = \mathbf{T}_{\boldsymbol{y}}$.

Conversely suppose that (b) holds. Then, the first line of $\mathbf{T}_{\boldsymbol{y}}$ gives the vector $2\pi \boldsymbol{y}$ and as demonstrated in (25) the first line of $\mathbf{V}_{\boldsymbol{x}} \mathbf{D}_{\boldsymbol{a}} \mathbf{V}_{\boldsymbol{x}}^*$ gives $2\pi \boldsymbol{\nu}(w_{\boldsymbol{a},\boldsymbol{x}})$, which implies that $\boldsymbol{y} = \boldsymbol{\nu}(w_{\boldsymbol{a},\boldsymbol{x}})$.

**Remark 3.** The statements of lemma 3 can once again readily be adapted to the case where the weights $\boldsymbol{a}$ are negative. In this case, the Toeplitz matrix $\mathbf{T}_{\boldsymbol{y}}$ is negative semi-definite.

We can now state our main result, which relies on theorem 1, proposition 3, proposition 7, and lemma 3.

**Theorem 2.** *The solution set of* (BPC) *can be characterized as follows:*

(*a*) *If* $\mathbf{T}_{\boldsymbol{y}}$ *has at least one negative and one positive eigenvalue, then* (BPC) *has a unique solution composed of at most* $2K_{\mathrm{c}}$ *Dirac masses, with at least one positive and one negative weight;*

(*b*) *If* $\mathbf{T}_{\boldsymbol{y}}$ *is positive, respectively negative, semi-definite and* $\operatorname{rank}(\boldsymbol{T}_{\boldsymbol{y}}) < K_{\mathrm{c}}+1$, *then* (BPC) *has a unique solution composed of exactly rank*$(\mathbf{T}_{\boldsymbol{y}})$ positive, respectively negative, Dirac masses;

(*c*) *If* $\mathbf{T}_{\boldsymbol{y}}$ *is positive, respectively negative, definite* (*which implies* $\operatorname{rank}(\boldsymbol{T}_{\boldsymbol{y}}) = K_{\mathrm{c}}+1$), *then* (BPC) *has infinitely many solutions. Moreover, there is no solution that is sum of less than* $K_{\mathrm{c}}+1$ *Dirac masses. Finally, there are uncountably many solutions composed of exactly* $K_{\mathrm{c}}+1$ *Dirac masses with positive, respectively negative, weights.*

**Proof.** If (a) holds, then by proposition 3 there is no nonnegative or nonpositive representing measure for $\boldsymbol{y}$. Hence, by theorem 1, (BPC) has a unique solution composed of at most $2K_{\mathrm{c}}$ Dirac masses, with at least one positive and one negative weight.

Suppose that the assumptions of (b) are satisfied. Without loss of generality, we can assume that $\mathbf{T}_{\boldsymbol{y}}$ is positive semi-definite. Then by the CFP decomposition, which is recalled in proposition 7 in appendix C, there exist unique sets $\{x_1,\dots,x_K\} \subset \mathbb{T}$ (the $x_k$ are pairwise distinct) and $\{a_1,\dots,a_K\} \subset \mathbb{R}_{>0}$ with $K = \operatorname{rank}(\mathbf{T}_{\boldsymbol{y}}) < K_{\mathrm{c}}+1$ such that

$$\mathbf{T}_{\boldsymbol{y}} = \mathbf{V}_{\boldsymbol{x}}\mathbf{D}_{\boldsymbol{a}}\mathbf{V}_{\boldsymbol{x}}^{*}, \tag{26}$$

with $\boldsymbol{x} = (x_1,\dots,x_K)$, $\boldsymbol{a} = (a_1,\dots,a_K)$, $\mathbf{V}_{\boldsymbol{x}} \in \mathbb{C}^{(K_{\mathrm{c}}+1)\times K}$ the Vandermonde matrix whose $k$th column is given by $\boldsymbol{e}_{K_{\mathrm{c}}}(x_k) = (1 \;\; \mathrm{e}^{\mathrm{i}x_k} \;\; \dots \;\; \mathrm{e}^{\mathrm{i}K_{\mathrm{c}}x_k})^{\mathrm{T}} \in \mathbb{C}^{K_{\mathrm{c}}+1}$ and $\mathbf{D}_{\boldsymbol{a}} \in \mathbb{R}^{K\times K}$ is a diagonal matrix with $\boldsymbol{a}$ on the diagonal. By lemma 3, equation (26) is equivalent to

$$\boldsymbol{y} = \boldsymbol{\nu}(w_{\boldsymbol{a},\boldsymbol{x}}) \quad \text{with} \quad w_{\boldsymbol{a},\boldsymbol{x}} = \sum_{k=1}^{K} a_k \delta_{x_k}. \tag{27}$$

Then, according to theorem 1 (equivalence between items (a) and (b)), the solution set of (BPC) is $\{w \in \mathcal{M}_+(\mathbb{T}) : \boldsymbol{\nu}(w) = \boldsymbol{y}\}$. Therefore, $w_{\boldsymbol{a},\boldsymbol{x}}$ is a solution of (BPC), which is composed of $K < K_{\mathrm{c}}+1$ positive Dirac masses. By lemma 3, the uniqueness of the CFP decomposition, and item (b) of theorem 1, it is the unique solution of (BPC) with less than $K_{\mathrm{c}}+1$ Dirac masses. By [2, theorem 2.1], it is the unique solution altogether, since it is possible to build a nonconstant dual certificate such that its signed saturation set is exactly $\operatorname{supp}_{\pm}(w_{\boldsymbol{a},\boldsymbol{x}})$, which is a sufficient condition of uniqueness by proposition 2.

Finally, suppose that the assumption of (c) is satisfied. Without loss of generality, we can assume that $\mathbf{T}_{\boldsymbol{y}}$ is positive-definite. Once again by proposition 7, there are uncountably many decompositions $\mathbf{T}_{\boldsymbol{y}} = \mathbf{V}_{\boldsymbol{x}}\mathbf{D}_{\boldsymbol{a}}\mathbf{V}_{\boldsymbol{x}}^{*}$ with $\boldsymbol{x} = (x_1,\dots,x_{K_{\mathrm{c}}+1}) \in \mathbb{T}^{K_{\mathrm{c}}+1}$ whose elements are pairwise distinct and $\boldsymbol{a} = (a_1,\dots,a_{K_{\mathrm{c}}+1}) \in (\mathbb{R}_{>0})^{K_{\mathrm{c}}+1}$. By lemma 3 and theorem 1 (equivalence between items (a) and (b)), this implies that for all these $(\boldsymbol{a},\boldsymbol{x})$ pairs, $w_{\boldsymbol{a},\boldsymbol{x}}$ is a solution of (BPC).[11] As a result, there are uncountably many solutions consisting of $K_{\mathrm{c}}+1$ Dirac masses with positive weights. There can be no solution consisting of less than $K_{\mathrm{c}}+1$ Dirac masses, since they would necessarily have positive weights and once again, by [[2], theorem 2.1], that solution would be the unique one, which we have proved to be false.

[11] As stated in proposition 7, one can arbitrarily choose one of the $x_k$ in $\mathbb{T}$ in the decomposition $\mathbf{T}_{\boldsymbol{y}} = \mathbf{V}_{\boldsymbol{x}}\mathbf{D}_{\boldsymbol{a}}\mathbf{V}_{\boldsymbol{x}}^{*}$, which proves that $t \in \mathbb{T} \mapsto 1$ is the unique dual certificate in this setup. Indeed any $(x_k, 1) \in \mathbb{T} \times \{\pm 1\}$ must be in the signed saturation set of any dual certificate.

**Remark 4.** Item (b) of theorem 2 can be deduced from well-grounded results in the literature. Indeed, it is known from a classical result [52, theorem 6.12] in the field of moment problems, that $\boldsymbol{y}$ has a unique nonnegative representing measure consisting of $\operatorname{rank}(\mathbf{T}_{\boldsymbol{y}})$ Dirac masses if and only if $\mathbf{T}_{\boldsymbol{y}}$ is positive semi-definite and rank deficient (i.e., $\operatorname{rank}(\mathbf{T}_{\boldsymbol{y}}) \leqslant K_{\mathrm{c}}$). Next, this measure is the unique solution of (BPC), since by [2, theorem 2.1], $\boldsymbol{y}$ is generated from a nonnegative measure composed of less than $K_{\mathrm{c}}+1$ Dirac masses. This can also be proved by the equivalence between items (a) and (b) of theorem 1. We also again recover the fact, mentioned in remark 1, that the TV norm plays no role in this context of nonnegativity. Concerning item (c) when $\mathbf{T}_{\boldsymbol{y}}$ is full rank, [52, theorem 6.12] states that there are infinitely many nonnegative representing measures composed of $K_{\mathrm{c}}+1$ Dirac masses. However, theorem 1 is, to the best of our knowledge, the first known result which proves that they are all solutions of (BPC).

**Remark 5.** Theorem 2 provides a sharper (and tight[12]) upperbound, $2K_{\mathrm{c}}$, on the number of Dirac masses of a sparse solution for (BPC) than representer theorems [58] which give the upperbound $2K_{\mathrm{c}}+1$. This improved bound arises when the assumptions of (a) hold. Note that in this context, the sparsity of the unique solution is no longer equal to the rank of $\mathbf{T}_{\boldsymbol{y}}$ since the CFP decomposition is valid only when $\mathbf{T}_{\boldsymbol{y}}$ is positive (or negative) semi-definite. This is confirmed by the fact that one can construct examples of a solution of (BPC) consisting of more than $K_{\mathrm{c}}+1$ Dirac masses[11], while the rank of $\mathbf{T}_{\boldsymbol{y}}$ is bounded by $K_{\mathrm{c}}+1$.

From theorem 2, we readily deduce all the situations of uniqueness for (BPC). They are summarized in the next corollary.

**Corollary 2.** *The problem* (BPC) *has a unique solution if and only if* $\mathbf{T}_{\boldsymbol{y}}$ *is neither positive nor negative definite.*

## 4. Conclusion

This paper deals with the linear continuous-domain inverse problem (BPC), where the goal is to recover a periodic Radon measure from its low-frequency Fourier coefficients $\boldsymbol{y}$ using a sparsity prior. We studied the question of uniqueness without making any assumptions of the ground-truth signal that generated $\boldsymbol{y}$. In this context, we proved that it can be characterized from simple conditions on the spectrum of the Toeplitz matrix $\mathbf{T}_{\boldsymbol{y}}$. We also demonstrated that this matrix contains information on the form of the sparse solutions of (BPC), namely their number of Dirac masses, or a bound thereon, and the signs of their weights.

## Acknowledgments

The authors are grateful to Christian Remling for his help regarding the extension of positive linear functionals in proposition 6. They also thank Shayan Aziznejad, Adrian Jarret, Matthieu Simeoni, and Michael Unser for interesting discussions. Julien Fageot is supported by the Swiss

[12] Consider the measure $w_0 = \sum_{k=0}^{2K_{\mathrm{c}}-1} (-1)^k \delta_{\frac{2\pi k}{2K_{\mathrm{c}}}}$, the measurements $\boldsymbol{y} = \boldsymbol{\nu}(w_0)$, and the function $\eta \in \mathcal{C}(\mathbb{T})$ defined as $\eta = t \mapsto \cos(K_{\mathrm{c}} t)$. Then, $\eta$ is a real trigonometric polynomial of degree $K_{\mathrm{c}}$ and $\operatorname{supp}_{\pm}(w_0) \subset \operatorname{sat}_{\pm}(\eta)$, which implies by proposition 1 that $w_0$ is a solution of (BPC).

National Science Foundation (SNSF) under Grants P2ELP2_181759 and P400P2_194364. The work of Thomas Debarre is supported by the SNSF under Grant 200020_184646/1.

## Data availability statement

No new data were created or analysed in this study.

## Appendix A. Proof of proposition 2

**Proof.** Let $\eta$ be a nonconstant dual certificate for (BPC) and $w_0 \in \mathcal{M}(\mathbb{T})$ a solution. Firstly, $\eta$ is a trigonometric polynomial of degree at most $K_c$ and so is its derivative $\eta'$. Since $\eta$ is nonconstant, $\eta'$ has at most $2K_c$ roots [59, page 150]. By proposition 1, we have $\mathrm{supp}_{\pm}(w_0) \subset \mathrm{sat}_{\pm}(\eta)$, hence any point in the support of $w_0$ is a root of $\eta'$. Consequently, $w_0$ is composed of at most $2K_c$ Dirac masses. Let $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_P) \in \mathbb{T}^P$ be the pairwise distinct roots of $\eta'$, with $P \leqslant 2K_c$. Then, we have $w_0 = \sum_{p=1}^{P} a_p \delta_{\tau_p}$, with $\boldsymbol{a} = (a_1, \ldots, a_P) \in \mathbb{R}^P$ (note that some weights may be equal to 0). Moreover, any other solution of (BPC) must be of the form $w_{\tilde{\boldsymbol{a}},\boldsymbol{\tau}} = \sum_{p=1}^{P} \tilde{a}_p \delta_{\tau_p}$, with $\tilde{\boldsymbol{a}} = (\tilde{a}_1, \ldots, \tilde{a}_P) \in \mathbb{R}^P$ (once again some weights may be equal to 0), where

$$\boldsymbol{\nu}(w_0) = \boldsymbol{y} = \boldsymbol{\nu}(w_{\tilde{\boldsymbol{a}},\boldsymbol{\tau}}). \tag{28}$$

Consider the matrix

$$\mathbf{M}_{\boldsymbol{\tau}} = \frac{1}{2\pi} \begin{bmatrix} \mathrm{e}^{\mathrm{i}K_c\tau_1} & \ldots & \mathrm{e}^{\mathrm{i}K_c\tau_P} \\ \vdots & \ddots & \\ \mathrm{e}^{\mathrm{i}\tau_1} & \ldots & \mathrm{e}^{\mathrm{i}\tau_P} \\ 1 & \ldots & 1 \\ \mathrm{e}^{-\mathrm{i}\tau_1} & \ldots & \mathrm{e}^{-\mathrm{i}\tau_P} \\ \vdots & \ddots & \\ \mathrm{e}^{-\mathrm{i}K_c\tau_1} & \ldots & \mathrm{e}^{-\mathrm{i}K_c\tau_P} \end{bmatrix} = \frac{1}{2\pi} \left[\mathrm{e}^{\mathrm{i}k\tau_p}\right]_{-K_c \leqslant k \leqslant K_c}^{1 \leqslant p \leqslant P} \in \mathbb{C}^{(2K_c+1)\times P}. \tag{29}$$

By definition of $\boldsymbol{\nu}$ and by equation (28), we get that $\boldsymbol{\nu}(w_0 - w_{\tilde{\boldsymbol{a}},\boldsymbol{\tau}}) = \mathbf{M}_{\boldsymbol{\tau}}(\boldsymbol{a} - \tilde{\boldsymbol{a}}) = \boldsymbol{0}$. The matrix $\mathbf{M}_{\boldsymbol{\tau}}$ is a Vandermonde-type matrix, which is therefore of full rank $P$, since $P \leqslant 2K_c$ and $\tau_1, \ldots, \tau_P$ are pairwise distinct. Hence the nullspace of $\mathbf{M}_{\boldsymbol{\tau}}$ is trivial and $\boldsymbol{a} = \tilde{\boldsymbol{a}}$, which prove the uniqueness of the solution $w_0$. □

## Appendix B. Trigonometric toolbox

This section is dedicated to known theoretical results (or easily deducible therefrom) that play a crucial role in our contributions in this paper.

A sequence $(a_k)_{k\in\mathbb{Z}}$ of complex numbers is *positive definite* if $a_0 \in \mathbb{R}_+$, $a_{-k} = \overline{a_k}$ for any $k \geqslant 0$, and for any sequence $(z_k)_{k\in\mathbb{Z}}$ of complex numbers with finitely many nonzero terms, we have

$$\sum_{k,\ell\in\mathbb{Z}} a_{k-\ell} z_k \overline{z_\ell} \geqslant 0. \tag{30}$$

**Proposition 4 (Herglotz theorem).** *A sequence $(a_k)_{k\in\mathbb{Z}}$ is positive-definite if and only if there exists a nonnegative measure $w \in \mathcal{M}_+(\mathbb{T})$ such that $\widehat{w}[k] = a_k$ for all $k \in \mathbb{Z}$.*

This theorem was obtained by Herglotz in [60]. For a modern exposition, we refer to [61, theorem 7.6]. The Herglotz theorem is an application of the Bochner theorem, which characterizes the Fourier transform of probability measures on locally Abelian groups $G$ [62], here with $G = \mathbb{T}$.

We recall the definition of the complex sinusoid functions $e_k(t) = \mathrm{e}^{\mathrm{i}kt}$ for $t \in \mathbb{T}$ and $k \in \mathbb{Z}$. For $K \geqslant 0$, we denote by $\mathcal{P}_K(\mathbb{T})$ the set of real trigonometric polynomial of degree at most $K$; i.e., functions of the form $p = \sum_{|k|\leqslant K} c_k e_k$ such that $c_0 \in \mathbb{R}$ and $c_{-k} = \overline{c_k} \in \mathbb{C}$ for any $1 \leqslant kv \leqslant K$.

**Proposition 5 (Fejér–Riesz theorem).** *Let $p = \sum_{|k|\leqslant K} c_k e_k \in \mathcal{P}_K(\mathbb{T})$ be a positive trigonometric polynomial of degree $K \geqslant 0$. Then, there exists a complex trigonometric polynomial $q = \sum_{k=0}^{K_c} z_k e_k$ such that $p = |q|^2$.*

The Fejér–Riesz theorem was conjectured by Fejér [63] and shown by Riesz [64]. See [65, page 26] for a recent exposition of this classical result. The next proposition deals with the extension of positive linear functionals from trigonometric polynomials to the space of continuous functions.

**Proposition 6.** *Let $K \geqslant 0$. Let $\Phi : \mathcal{P}_K(\mathbb{T}) \to \mathbb{R}$ be a linear and positive functional (i.e., $\Phi(p) \geqslant 0$ for any $p \geqslant 0$). Then, there exists an extension $\Phi : \mathcal{C}(\mathbb{T}) \to \mathbb{R}$ which is still linear and positive. Moreover, any such extension is continuous on $(\mathcal{C}(\mathbb{T}), \|\cdot\|_\infty)$.*

**Proof.** Let $E$ be an ordered topological vector space, $C$ its positive cone, and $M \subset E$. By [66, corollary 2 page 227], if $C \cap M$ contains an interior point of $C$, then any continuous, positive, and linear form over $M$ can be extended as a continuous, positive, and linear form over $E$.

We apply this result to $E = \mathcal{C}(\mathbb{T})$, whose positive cone is the space of positive continuous functions $\mathcal{C}_+(\mathbb{T})$, and to $M = \mathcal{P}_K(\mathbb{T})$. Then, $C \cap M = \mathcal{C}_+(\mathbb{T}) \cap \mathcal{P}_K(\mathbb{T})$ contains the constant function $p = 1$, which is an interior point of $\mathcal{C}_+(\mathbb{T})$ since $\{f \in \mathcal{C}(\mathbb{T}),\ \|f - 1\|_\infty < \frac{1}{2}\} \subset \mathcal{C}_+(\mathbb{T})$.

In our case, $\Phi$ is continuous over $(\mathcal{P}_K(\mathbb{T}), \|\cdot\|_\infty)$, since it is a linear functional over a finite-dimensional space. Hence, $\Phi$ is continuous, positive, and linear, and admits the desired extension.

## Appendix C. The CFP decomposition

In this appendix, we give, for the sake of completeness, the CFP decomposition which plays a major role in our main contribution theorem 2. Proposition 7, is a transcription from [54, theorem 1], up to slight adaptations of notations and reformulations. Before stating the result, let us define the column matrix $\boldsymbol{e}_K(x) = (1\ \mathrm{e}^{\mathrm{i}x}\ \ldots\ \mathrm{e}^{\mathrm{i}Kx})^{\mathrm{T}}$, for all $x \in \mathbb{T}$ and $K \in \mathbb{N}$.

**Proposition 7.** *Let $K \in \mathbb{N}$, $\mathbf{T} \in \mathbb{C}^{(K+1)\times(K+1)}$ be a Hermitian Toeplitz matrix, and $r = \mathrm{rank}(\mathbf{T})$. Then, the two following conditions are equivalent:*

(*a*) $\mathbf{T}$ *is positive semi-definite;*

(*b*) $\mathbf{T} = \mathbf{V}_{\boldsymbol{x}}\mathbf{D}_{\boldsymbol{a}}\mathbf{V}_{\boldsymbol{x}}^{*}$*, where* $\boldsymbol{x} = (x_1, \ldots, x_r) \in \mathbb{T}^r$ *has pairwise distinct elements,* $\mathbf{V}_{\boldsymbol{x}} \in \mathbb{C}^{(K+1)\times r}$ *is the Vandermonde matrix whose columns are* $\boldsymbol{e}_K(x_k)$ *for* $k \in \{1, \ldots, r\}$*, and* $\mathbf{D}_{\boldsymbol{a}} \in \mathbb{R}^{r\times r}$ *is a diagonal matrix with* $\boldsymbol{a} = (a_1, \ldots, a_r) \in (\mathbb{R}_{>0})^r$ *on the diagonal.*

*Moreover, when* $\mathbf{T}$ *is rank-deficient (i.e.* $r < K+1$*), then the decomposition in item* (*b*) *is unique up to any permutation applied to the coefficients of* $\boldsymbol{x}$ *and* $\boldsymbol{a}$*. When* $\mathbf{T}$ *has full rank, then there are uncountably many* $\{x_1, \ldots, x_{K+1}\} \subset \mathbb{T}$ *such that the decomposition holds. Note that* $x_{K+1}$ *can be arbitrarily chosen in* $\mathbb{T}$*.*

We remark that the decomposition given in [54, theorem 1] is as a sum of $r$-rank one matrices $a_k \boldsymbol{e}_K(x_k)\boldsymbol{e}_K(x_k)^*$. This can readily be shown to be equivalent to the decomposition in proposition 7.

Secondly, the result of [54], does not adjudicate on the uniqueness of the decomposition when the Toeplitz matrix $\mathbf{T}$ is of full rank. However, by carefully studying the proof of [54, theorem 1], one can see that the full-rank case is studied in detail (see in particular equations (10)–(12)). We summarized the conclusion at the end of proposition 7.

## ORCID iDs

Quentin Denoyelle https://orcid.org/0000-0003-2067-2659

## References

[1] Rudin L, Osher S and Fatemi E 1992 Nonlinear total variation based noise removal algorithms *Physica* D **60** 259–68
[2] de Castro Y and Gamboa F 2012 Exact reconstruction using Beurling minimal extrapolation *J. Math. Anal. Appl.* **395** 336–54
[3] Candès E and Fernandez-Granda C 2013 Super-resolution from noisy data *J. Fourier Anal. Appl.* **19** 1229–54
[4] Bredies K and Pikkarainen H 2013 Inverse problems in spaces of measures *ESAIM: Control Optim. Calc. Var.* **19** 190–218
[5] Duval V and Peyré G 2015 Exact support recovery for sparse spikes deconvolution *Found. Comput. Math.* **15** 1315–55
[6] Unser M, Fageot J and Ward J P 2017 Splines are universal solutions of linear inverse problems with generalized TV regularization *SIAM Rev.* **59** 769–93
[7] Boyd N, Schiebinger G and Recht B 2017 The alternating descent conditional gradient method for sparse inverse problems *SIAM J. Optim.* **27** 616–39
[8] Flinth A, de Gournay F and Weiss P 2020 On the linear convergence rates of exchange and continuous methods for total variation minimization *Math. Program.* **190** 221–57
[9] Courbot J-B, Duval V and Legras B 2020 Sparse analysis for mesoscale convective systems tracking *Signal Process. Image Commun.* **85** 115854
[10] Courbot J-B and Colicchio B 2021 A fast homotopy algorithm for gridless sparse recovery *Inverse Problems* **37** 025002
[11] Simeoni M 2021 Functional penalised basis pursuit on spheres *Appl. Comput. Harmon. Anal.* **53** 1–53
[12] Tibshirani R 2013 The lasso problem and uniqueness *Electron. J. Stat.* **7** 1456–90
[13] Chen S S, Donoho D L and Saunders M A 2001 Atomic decomposition by basis pursuit *SIAM Rev.* **43** 129–59
[14] Donoho D 2006 Compressed sensing *IEEE Trans. Inf. Theory* **52** 1289–306

[15] Elad M 2010 *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing* (Berlin: Springer)

[16] Foucart S and Rauhut H 2013 *A Mathematical Introduction to Compressive Sensing* vol 1 (Basel: Birkhäuser)

[17] Unser M, Fageot J and Gupta H 2016 Representer theorems for sparsity-promoting $\ell_1$ regularization *IEEE Trans. Inf. Theory* **62** 5167–80

[18] Vetterli M, Marziliano P and Blu T 2002 Sampling signals with finite rate of innovation *IEEE Trans. Signal Process.* **50** 417–1428

[19] Schmidt R 1986 Multiple emitter location and signal parameter estimation *IEEE Trans. Antennas Propag.* **34** 276–80

[20] Plonka G and Tasche M 2014 Prony methods for recovery of structured functions *GAMM-Mitt.* **37** 239–58

[21] Pan H, Blu T and Vetterli M 2016 Towards generalized FRI sampling with an application to source resolution in radioastronomy *IEEE Trans. Signal Process.* **65** 821–35

[22] Beurling A 1938 Sur les intégrales de Fourier absolument convergentes et leur applicationa une transformation fonctionnelle *9th Scandinavian Mathematical Congress* pp 345–66

[23] Candès E and Fernandez-Granda C 2014 Towards a mathematical theory of super-resolution *Commun. Pure Appl. Math* **67** 906–56

[24] Eftekhari A and Wakin M B 2013 Greed is super: a new iterative method for super-resolution *2013 IEEE Global Conf. Signal and Information Processing* (Piscataway, NJ: IEEE) p 631

[25] Elvira C, Gribonval R, Soussen C and Herzet C 2021 When does OMP achieve exact recovery with continuous dictionaries? *Appl. Comput. Harmon. Anal.* **51** 374–413

[26] Traonmilin Y, Aujol J-F and Leclaire A 2020 Projected gradient descent for non-convex sparse spike estimation *IEEE Signal Process. Lett.* **27** 1110–4

[27] Chizat L 2022 Sparse optimization on measures with over-parameterized gradient descent *Math. Program.* **194** 487–532

[28] Fisher S and Jerome J 1975 Spline solutions to $L^1$ extremal problems in one and several variables *J. Approx. Theory* **13** 73–83

[29] Azais J, de Castro Y and Gamboa F 2015 Spike detection from inaccurate samplings *Appl. Comput. Harmon. Anal.* **38** 177–95

[30] Duval V and Peyré G 2017 Sparse regularization on thin grids: I. The LASSO *Inverse Problems* **33** 055008

[31] Duval V and Peyré G 2017 Sparse spikes super-resolution on thin grids: II. The continuous basis pursuit *Inverse Problems* **33** 095008

[32] Poon C and Peyré G 2019 Multidimensional sparse super-resolution *SIAM J. Math. Anal.* **51** 1–44

[33] Tibshirani R 1996 Regression shrinkage and selection via the LASSO *J. R. Stat. Soc.* B **58** 267–88

[34] Tseng P 2001 Convergence of a block coordinate descent method for nondifferentiable minimization *J. Optim. Theory Appl.* **109** 475–94

[35] Osborne M R, Presnell B and Turlach B A 2000 A new approach to variable selection in least squares problems *IMA J. Numer. Anal.* **20** 389–403

[36] Combettes P L and Wajs V R 2005 Signal recovery by proximal forward–backward splitting *Multiscale Model. Simul.* **4** 1168–200

[37] Beck A and Teboulle M 2009 A fast iterative shrinkage-thresholding algorithm for linear inverse problems *SIAM J. Imaging Sci.* **2** 183–202

[38] de Castro Y, Gamboa F, Henrion D and Lasserre J-B 2016 Exact solutions to super resolution on semi-algebraic domains in higher dimensions *IEEE Trans. Inf. Theory* **63** 621–30

[39] Catala P, Duval V and Peyré G 2019 A low-rank approach to off-the-grid sparse superresolution *SIAM J. Imaging Sci.* **12** 1464–500

[40] Frank M and Wolfe P 1956 An algorithm for quadratic programming *Nav. Res. Log. Q.* **3** 95–110

[41] Denoyelle Q, Duval V, Peyré G and Soubies E 2019 The sliding Frank–Wolfe algorithm and its application to super-resolution microscopy *Inverse Problems* **36** 014001

[42] Ali A and Tibshirani R 2019 The generalized lasso problem and uniqueness *Electron. J. Stat.* **13** 2307–47

[43] Pouchol C and Verdier O 2020 The ML–EM algorithm in continuum: sparse measure solutions *Inverse Problems* **36** 035013

[44] Pouchol C and Verdier O 2020 Linear inverse problems with nonnegativity constraints through the beta-divergences: sparsity of optimisers (arXiv:2006.15845)

[45] Eftekhari A, Tanner J, Thompson A, Toader B and Tyagi H 2021 Sparse non-negative super-resolution—simplified and stabilised *Appl. Comput. Harmon. Anal.* **50** 216–80
[46] Eftekhari A, Bendory T and Tang G 2021 Stable super-resolution of images: theoretical study *Inf. Inference* **10** 161–93
[47] Candès E and Plan Y 2010 A probabilistic and RIPless theory of compressed sensing (arXiv:1011.3854)
[48] Debarre T, Denoyelle Q, Unser M and Fageot J 2021 Sparsest piecewise-linear regression of one-dimensional data *J. Comput. Appl. Math.* **46** 114044
[49] Lasserre J B 2008 A semidefinite programming approach to the generalized problem of moments *Math. Program.* **112** 65–92
[50] Lasserre J B 2009 *Moments, Positive Polynomials and their Applications* vol 1 (Singapore: World Scientific)
[51] Curto R E and Fialkow L 1991 Recursiveness, positivity and truncated moment problems *Houst. J. Math.* **17** 603–35
[52] Carathéodory C and Fejér L 1884–1940 Über den zusammenhang der extremen von harmonischen funktionen mit ihren koeffizienten und über den picard-landau'schen satz *Rend. Circ. Mat. Palermo* **32** 218–39
[53] Pisarenko V F 1973 The retrieval of harmonics from a covariance function *Geophys. J. Int.* **33** 347–66
[54] Yang Z and Xie L 2018 Frequency-selective Vandermonde decomposition of Toeplitz matrices with applications *Signal Process.* **142** 157–67
[55] Gray J 1984 The shaping of the Riesz representation theorem: a chapter in the history of analysis *Arch. Hist. Exact Sci.* **31** 127–87
[56] Ekeland I and Temam R 1976 *Convex Analysis and Variational Problems* (Philadelphia, PA: SIAM)
[57] Fernandez-Granda C 2016 Super-resolution of point sources via convex programming *Inf. Inference* **5** 251–303
[58] Fageot J and Simeoni M 2020 TV-based reconstruction of periodic functions *Inverse Problems* **36** 115015
[59] Powell M 1981 *Approximation Theory and Methods* (Cambridge: Cambridge University Press)
[60] Herglotz G 1911 Uber potenzreihen mit positivem, reelen teil im einheitskreis *Ber. Verhandl. Sachs Akad. Wiss. Leipzig, Math.-Phys. Kl.* **63** 501–11
[61] Katznelson Y 2004 *An Introduction to Harmonic Analysis* (Cambridge: Cambridge University Press)
[62] Rudin W 1962 *Fourier Analysis on Groups* vol 121967 (New York: Wiley)
[63] Fejér L 1916 Über trigonometrische polynome *J. Reine Angew. Math.* **146** 53–82
[64] Riesz F 1916 Über ein Problem des Herrn Carathéodory *J. Reine Angew. Math.* **146** 83–7
[65] Barry S 2005 *Orthogonal Polynomials on the Unit Circle* (Providence, RI: American Mathematical Society)
[66] Schaefer H 1971 Locally convex topological vector spaces *Topological Vector Spaces* (Berlin: Springer) pp 36–72