Stable Super-Resolution of Images: A Theoretical Study

Armin Eftekhari, Tamir Bendory, and Gongguo Tang

October 16, 2018

Abstract

We study the ubiquitous super-resolution problem, in which one aims at localizing positive point sources in an image, blurred by the point spread function of the imaging device. To recover the point sources, we propose to solve a convex feasibility program, which simply finds a nonnegative Borel measure that agrees with the observations collected by the imaging device.

In the absence of imaging noise, we show that solving this convex program uniquely retrieves the point sources, provided that the imaging device collects enough observations. This result holds true if the point spread function of the imaging device can be decomposed into horizontal and vertical components, and if the translations of these components form a Chebyshev system, namely, a system of continuous functions that loosely behave like algebraic polynomials.

Building upon recent results for one-dimensional signals [1], we prove that this super-resolution algorithm is stable (in the generalized Wasserstein metric) to model mismatch (namely, when the image is not sparse) and to additive imaging noise. In particular, the recovery error depends on the noise level and how well the image can be approximated with well-separated point sources. As an example, we verify these claims for the important case of a Gaussian point spread function. The proofs rely on the construction of novel interpolating polynomials, which are the main technical contribution of this paper, and partially resolve the question raised in [2] about the extension of the standard machinery to higher dimensions.

1 Introduction

Consider an unknown number of point sources with unknown locations and amplitudes. An imaging mechanism provides us with a few noisy measurements from which we wish to estimate the locations and amplitudes of these sources. Because of the finite resolution of the imaging device, poorly separated sources are indistinguishable without using a proper localization algorithm that would take into account the particular structure of image. This super-resolution problem of localizing point sources finds various applications, for example in astronomy [3], geophysics [4], chemistry, medicine, microscopy and neuroscience [5, 6, 7, 8, 9, 10, 11]. Most of these applications involve images or higher dimensional signals. In this paper, we study the “grid-free” and positive super-resolution of two-dimensional (2-D) signals (namely, images) in the presence of noise, extending the one-dimensional (1-D) results of [1].

Let \( x \) be a nonnegative Borel measure supported on \( \mathbb{I}^2 = [0,1] \times [0,1] \), and let \( \{ \phi_m \}_{m=1}^M \) be real-valued and continuous functions. The (possibly noisy) observations \( \{ y_{m,n} \}_{m,n=1}^M \) collected from \( x \) are then given by

\[
y_{m,n} \approx \int_{\mathbb{I}^2} \phi_m(t) \phi_n(s) x(t,ds) dt.
\]

(1)

More specifically, we assume that

\[
\sum_{m,n=1}^M \left| y_{m,n} - \int_{\mathbb{I}^2} \phi_m(t) \phi_n(s) x(t,ds) \right|^2 \leq \delta^2,
\]

(2)
where $\delta \geq 0$ reflects the additive noise level. We do not impose any statistical model on the noise. If we define the matrices $y \in \mathbb{R}^{M \times M}$ and $\Phi(t, s) \in \mathbb{R}^{M \times M}$ such that

$$y[m, n] = y_{m,n}, \quad \Phi(t, s)[m, n] = \phi_m(t)\phi_n(s), \quad \forall m, n \in [M] := \{1, \cdots, M\}, \quad (3)$$

we may rewrite (2) more compactly as

$$\|y - \int_{I^2} \Phi(t, s) x(dt, ds)\|_F \leq \delta, \quad (4)$$

where $\| \cdot \|_F$ stands for the Frobenius norm. Often, $\phi_m$ and $\phi_n$ above are translated copies of a function $\phi$, and $\phi(t)\phi(s)$ is referred to as the point spread function of the imaging device and $y$ is the 2-D acquired signal that can be thought of as an image with $M^2$ pixels. We note that the tensor product model in (4) is widely used as a model in imaging [12, 13, 14]. For example, if the imaging device acts as an ideal low-pass filter with the cut-off frequency of $f_c$, then the corresponding choice is $\{\phi_m\}_{m=1}^M = \{\cos 2\pi k_1 f_c\}_{k=0}^{f_c} \cup \{\sin 2\pi k_1 f_c\}_{k=1}^{f_c}$ with $M = 2f_c + 1$.

In order to recover $x$, we suggest using the simple convex feasibility program

$$\text{find a nonnegative Borel measure } z \text{ on } \mathbb{R}^2 \text{ such that } \|y - \int_{I^2} \Phi(t, s) z(dt, ds)\|_F \leq \delta', \quad (5)$$

for some $\delta' \geq \delta$, which is reminiscent of nonnegative least squares in finite dimensions [15, 16].

Program (5) does not involve a grid on $I^2$, and notably does not regularize $z$ beyond nonnegativity, thus radically deviating from the existing literature [14, 13, 2, 17, 18]. This paper establishes that in the noiseless setting $\delta = 0$, solving Program (5) precisely recovers the true measure $x$, provided that $x$ is a nonnegative sparse measure on $I^2$ and under certain conditions on the imaging apparatus $\Phi$. More importantly, when $\delta > 0$ and $x$ is an arbitrary nonnegative measure on $I^2$, solving Program (5) well-approximates $x$. In particular, we establish that any nonnegative measure supported on $I^2$ that agrees with the observations $y$ in the sense of Program (5) is near the true measure $x$.

This paper does not focus on the important question of how to numerically solve the infinite-dimensional Program (5) in practice. One straightforward approach would be to discretize the measure $z$ on a fine uniform grid for $I^2$, thereby replacing Program (5) with a finite-dimensional convex feasibility program that can be solved with standard convex solvers. A few recent papers proposed algorithms to directly solve Program (5) and avoid discretization [19, 20, 21]. Moreover, there are other popular techniques to estimate sparse measures, for instance, using various generalizations of Prony’s method; see Section 3 for a survey of the related literature. A comprehensive numerical comparison between the alternatives for different noise levels is of great importance and we leave that to a future study. This paper aims to provide theoretical justifications for the success of Program (5), thereby arguing that imposing nonnegativity is enough for successful super-resolution. In other words, under mild conditions, the imaging device acts as an injective map on the set of sparse nonnegative measures and we can stably find its inverse.

This work builds and relies heavily on a recent work [1], which established that grid-free and positive super-resolution in 1-D can be achieved by solving the 1-D version of Program (5). In doing so, it removed the regularization required in prior work and substantially simplified the existing results. That being said, extending [1] to two dimensions is far from trivial and requires a careful design of new family of dual certificates, as will become clear in the next sections. Indeed, this work overcomes the technical obstacles noted in [2, Section 4] for extending the proof machinery to higher dimensions.

Before turning to the details, let us summarize the technical contributions of this paper. Section 2 presents these results in detail, while technical details are deferred to Section 4 and the appendices.

**Sparse measures without noise.** Suppose that $x$ consists of $K$ positive impulses located in $I^2$. In the absence of noise $\delta = 0$, Proposition 2 below shows that solving Program (5) with $\delta' = 0$ successfully recovers $x$ from the observations $y \in \mathbb{R}^{M \times M}$, provided that $M \geq 2K + 1$ and that $\{\phi_m\}_{m=1}^M$ form a Chebyshev system on $I$. A Chebyshev system, or T-system for short, is a collection of continuous functions that loosely behave like algebraic monomials; see Definition 1. T-system is a widely-used concept in classical approximation theory [22, 23, 24] that also plays a pivotal role in some modern signal processing applications; see for
instance [1, 17, 2]. In other words, Proposition 2 below establishes that the imaging operator $\Phi$ in (4) is an injective map from $K$-sparse nonnegative measures on $\mathbb{R}^2$, provided that $\{\varphi_m\}_{m=1}^M$ form a T-system on $I$ and $M \geq 2K + 1$.

In contrast with earlier results, no minimum separation between the impulses is necessary here. Program (5) does not contain any explicit regularization to promote sparsity, and lastly $\{\varphi_m\}_{m=1}^M$ need only to be continuous. We note that Proposition 2 is a nontrivial extension of the 1-D result in [1] to images. Indeed, the key concept of T-systems do not generalize to two or higher dimensions and proving Proposition 2 requires a novel construction of dual certificates to overcome the technical obstacles anticipated in [2, Section 4]; see also the general results below.

**Arbitrary measure with noise.** More generally, consider an arbitrary nonnegative measure $x$ supported on $\mathbb{R}^2$. As detailed later, given $\varepsilon \in (0, 1/2]$, $x$ can always be approximated with a $K$-sparse and $\varepsilon$-separated nonnegative measure, up to an error of $R(x, K, \varepsilon)$ in the generalized Wasserstein metric. This is true even if $x$ itself is not $\varepsilon$-separated or not atomic at all. We might think of $R(x, K, \varepsilon)$ as the “model-mismatch” of approximating $x$ with a well-separated sparse measure.

In the presence of noise and numerical inaccuracies, namely when $\delta \geq 0$, Theorem 5 below shows that solving Program (5) approximately recovers $x$ from the observations $y \in \mathbb{R}^{M \times M}$ in the generalized Wasserstein metric $d_{GW}$. In particular, a solution $\hat{x}$ of Program (5) satisfies

$$d_{GW}(x, \hat{x}) \leq c_1 \delta + c_2 \varepsilon + c_3 R(x, K, \varepsilon),$$

(6)

provided that $M \geq 2K + 2$, and the imaging apparatus and certain functions forms a $T^*$-system, a natural generalization of the T-system. The factors $c_1, c_2, c_3$ above are specified in the proof and depend chiefly on the measurement functions $\{\varphi_m\}_{m=1}^M$; see (3). Note that the recovery error in (6) depends on the noise level $\delta$, the separation $\varepsilon$, and on how well $x$ can be approximated with a $K$-sparse and $\varepsilon$-separated measure, similar to the 1-D results in [1]. Note also that, when $\delta = R(x, K, \varepsilon) = 0$, (6) reads $d_{GW}(x, \hat{x}) = 0$, and Theorem 5 reduces to Proposition 2 for sparse and noise-free super-resolution.

We remark that Theorem 5 applies to any nonnegative measure $x$, without requiring any separation between the impulses in $x$. In fact, $x$ might not be atomic at all. Of course, the recovery error $d_{GW}(x, \hat{x})$ does depend on how well $x$ can be approximated with a well-separated sparse measure, which is reflected in the right-hand side of (6) and hidden in the factors $c_1, c_2, c_3$ therein. As emphasized earlier, no regularization other than nonnegativity was used and $\{\varphi_m\}_{m}$ need only be continuous.

As a concrete example of this general framework, we consider the case where $\{\varphi_m\}_{m}$ are translated Gaussian windows. Building on the results from [1], we show in Section 2.3 that the conditions for both Proposition 2 and Theorem 5 are met for this important example. That is, solving Program (5) successfully and stably recovers an image that has undergone Gaussian blurring.

## 2 Main Results

### 2.1 Sparse Measure Without Noise

Let $x$ be a nonnegative atomic measure

$$x = \sum_{k=1}^K a_k \cdot \delta_{\theta_k}, \quad a_k > 0,$$

(7)

with $K \geq 1$ impulses located at $\Theta = \{\theta_k\}_{k=1}^K \subset \text{interior}(\mathbb{R}^2)$ with positive amplitudes $\{a_k\}_{k=1}^K$. Here, $\delta_{\theta_k}$ is the Dirac measure located at $\theta_k = (t_k, s_k)$. We first consider the case where there is no imaging noise ($\delta = 0$) and thus we collect the noise-free observations

$$y = \int_{\mathbb{R}^2} \Phi(t, s)x(dt, ds) \in \mathbb{R}^{M \times M}.$$

(8)

To understand when solving Program (5) with $\delta' = 0$ successfully recovers the true measure $x$, recall the concept of T-system [22]:


Definition 1. (T-system) Real-valued and continuous functions \( \{\phi_m\}_{m=1}^M \) form a T-system on the interval \( \mathbb{I} \), if the \( M \times M \) matrix \( [\phi_m(\tau_k)]_{k,m=1}^M \) is nonsingular for any increasing sequence \( \{\tau_k\}_{k=1}^M \subset \mathbb{I} \).

For example, the monomials \( \{1, t, \ldots, t^{M-1}\} \) form a T-system on any closed interval of the real line. In fact, T-system can be understood as a generalization of ordinary monomials. For instance, it is not difficult to verify that any “polynomial” \( \sum_{m=1}^M b_m \phi_m(t) \) of a T-system \( \{\phi_m\}_{m=1}^M \) has at most \( M - 1 \) distinct zeros on \( \mathbb{I} \). Or, given \( M \) distinct points on \( \mathbb{I} \), there exists a unique polynomial of \( \{\phi_m\}_{m=1}^M \) that interpolates these points. Note also that the linear independence of \( \{\phi_m\}_{m=1}^M \) is a necessary—but not sufficient—condition for forming a T-system.

In the context of super-resolution, for example, translated copies of the Gaussian window \( e^{-t^2} \) form a T-system on any interval, and so do many other windows [22]. As we will see later, the notion of T-system allows us to design a nonnegative polynomial with prescribed zeros on \( \mathbb{I} \), that plays a key role in the 2-D construction.

The following result, proved in Section 4.2, states that solving Program (5) successfully recovers \( x \) from the noise-free image \( y \), provided that the measurement functions form a T-system.

Proposition 2. (Sparse measure without noise) Let \( x \) be a \( K \)-sparse nonnegative measure supported on \( \text{interior}(\mathbb{I}^2) \); see (7). Let also \( M \geq 2K + 1 \) and suppose that the measurement functions \( \{\phi_m\}_{m=1}^M \) form a T-system on \( \mathbb{I} \). Lastly, let \( \delta = 0 \), consider the imaging operator \( \Phi \) and the image \( y \in \mathbb{R}^{M \times M} \) in (3) and (4). Then, \( x \) is the unique solution of Program (5) with \( \delta' = 0 \).

In words, Program (5) successfully localizes the \( K \) impulses present in \( x \) given only \( (2K + 1)^2 \) measurements, if the measurement functions \( \{\phi_m\}_{m=1}^M \) form a T-system on \( \mathbb{I} \). Note that no minimum separation is required between the impulses, in contrast to similar results for super-resolution with both signed and non-negative measures; see for instance [25, 26, 13]. In addition, no regularization was imposed in Program (5) beyond nonnegativity, and the measurement functions only need to be continuous.

Proof technique. Let us outline the proof of Proposition 2. Loosely speaking, a standard argument shows that the existence of a certain polynomial of the form

\[
Q(\theta) = Q(t,s) = \sum_{m,n=1}^M b_{m,n} \phi_m(t) \phi_n(s),
\]

would guarantee that Program (5) successfully recovers a sparse nonnegative measure in the absence of noise. Known as the dual certificate for Program (5), this polynomial \( Q \) has to be nonnegative on \( \mathbb{I}^2 \), with zeros only at the impulse locations \( \Theta = \{\theta_k\}_{k=1}^K = \{(t_k, s_k)\}_{k=1}^K \). Setting \( T = \{t_k\}_{k=1}^K \) and \( S = \{s_k\}_{k=1}^K \), the proof constructs \( Q \) by carefully combining nonnegative univariate polynomials with prescribed zeros on subsets of \( T \) and \( S \). In turn, such univariate polynomials exist if \( \{\phi_m\}_{m=1}^M \) form a T-system on \( \mathbb{I} \); see Section 4.2 for the details. The basic idea of the proof is visualized in Figure 1.

2.2 Arbitrary Measure With Noise

In this section, we present the main result of this paper. Theorem 5 below generalizes Proposition 2 to account for a) model mismatch, where \( x \) is not necessarily a well-separated sparse measure but might be close to one, and b) imaging noise \( \delta \geq 0 \). This result addresses the stability of Program (5) to model mismatch and its robustness against imaging noise. Some preparation is necessary before presenting the result.

Separation. Unlike sparse and noise-free super-resolution in Proposition 2, a notion of separation plays a role in the general result presented in Theorem 5. For an atomic measure \( x \) supported on \( \Theta = \{\theta_k\}_{k=1}^K = \{(t_k, s_k)\}_{k=1}^K \subset \mathbb{I}^2 \), let \( \text{sep}(x) \) be the minimum separation between all impulses in \( x \) and the boundary of \( \mathbb{I}^2 \).

Formally, \( \text{sep}(x) \) is the largest number \( \nu \) such that

\[
\nu \leq |t_k - t_l|, \quad \nu \leq |s_k - s_l|, \quad k \neq l, \ k, l \in [K],
\]

\[

\nu \leq |t_k - 0|, \quad \nu \leq |t_k - 1|,
\]

where \( K \) is an increasing sequence.
Figure 1: This figure explains the idea behind the proof of Proposition 2, see Section 2.1. As an example, consider the nonnegative measure \( x = a_1 \delta_{\theta_1} + a_2 \delta_{\theta_2} \). The locations of two impulses at \( \theta_1 = (t_1, s_1) \in \mathbb{I}^2 \) and \( \theta_2 = (t_2, s_2) \in \mathbb{I}^2 \) are shown with black dots in Figure 1a. For Program (5) to successfully recover \( x \) from the image \( y \) in (8), we should construct a nonnegative polynomial \( Q \) of the form in (9) that has zeros exactly on \( \theta_1 \) and \( \theta_2 \). We do so by combining a number of univariate polynomials in \( t \) and \( s \). More specifically, consider a nonnegative polynomial \( q_{t_1}(t) = \sum_{m=1}^{M} b_{11} \phi_m(t) \) that vanishes only at \( t_1 \). Likewise, consider similar nonnegative polynomials \( q_{s_2}(s) \), which are zero only at \( s_2 \), respectively. Figure 1b shows the zero set of the polynomial \( q_{t_1}(t)q_{s_2}(s) \) as the union of blue and red lines. Similarly, Figure 1c shows the zero set of the polynomial \( q_{t_2}(t)q_{s_1}(s) \).

Figure 1b shows the zero set of the polynomial \( q_{t_1}(t)q_{s_2}(s) \) as the union of blue and red lines. Similarly, Figure 1c shows the zero set of the polynomial \( q_{t_2}(t)q_{s_1}(s) \). Note that the intersection of these two zero sets is exactly \( \{\theta_1, \theta_2\} \). That is, \( q(\theta) = q_{t_1}(t)q_{s_2}(s) + q_{t_2}(t)q_{s_1}(s) \) is a nonnegative polynomial of the form in (9) that has zeros only at \( \{\theta_1, \theta_2\} \), as desired. It only remains now to construct the univariate polynomials \( q_{t_1}, q_{t_2}, q_{s_1}, q_{s_2} \) described above. When the imaging apparatus \( \{\phi_m\}_{m=1}^{M} \) forms a T-system and \( M \geq 2K + 1 \), the existence of these univariate polynomials follows from [1]. We note that the construction of \( Q \) in this example is slightly different from the more involved proof of Proposition 2 to simplify the presentation.

Naturally, if the measure \( x \) satisfies \( \text{sep}(x) \geq \varepsilon \), we call \( x \) an \( \varepsilon \)-separated measure. For example, for \( x \) in Figure 1a, we have \( \text{sep}(x) = \min(t_2 - t_1, s_2 - s_1, t_1, 1 - t_2, s_1, 1 - s_2) \).

This notion of separation is commonly used in the super-resolution literature; see [12, 26, 27], to name just a few.

**Generalized Wasserstein distance.** As an error metric, we use the generalized Wasserstein distance [28]. We first recall that the total-variation (TV) norm of a measure on \( \mathbb{I}^2 \) is defined as \( \|z\|_{TV} = \int_{\mathbb{I}^2} |dz| \), akin to \( \ell_1 \)-norm in finite dimensions. The standard Wasserstein distance for two nonnegative measures on \( \mathbb{I}^2 \),
frequently called the earth mover’s distance, is defined as
\[
d_w(z_1, z_2) = \inf \int_{\mathbb{R}^2 \times \mathbb{R}^2} |\tau_1 - \tau_2| \cdot \gamma(d\tau_1, d\tau_2),
\]
where the infimum is over every nonnegative measure \( \gamma \) on \( \mathbb{R}^2 \times \mathbb{R}^2 \) that produces \( z_1 \) and \( z_2 \) as marginals, that is,
\[
z_1(A_1) = \int_{A_1 \times \mathbb{R}^2} \gamma(d\tau_1, d\tau_2), \quad z_2(A_2) = \int_{\mathbb{R}^2 \times A_2} \gamma(d\tau_1, d\tau_2),
\]
for every measurable sets \( A_1, A_2 \subseteq \mathbb{R}^2 \). If we were to think of \( z_1, z_2 \) as two piles of dirt, then \( d_w(z_1, z_2) \) can be interpreted as the least amount of work needed to transform \( z_1 \) to \( z_2 \). The Wasserstein distance is defined only if the TV norms of the two measures are equal. The generalized Wasserstein distance extends \( d_w \) to allow for calculating the distance between nonnegative measures with different TV norms. It is defined by
\[
d_{GW}(x_1, x_2) = \inf (\|x_1 - z_1\|_{TV} + d_w(z_1, z_2) + \|x_2 - z_2\|_{TV}),
\]
where the infimum is over every pair of nonnegative Borel measures \( z_1, z_2 \) supported on \( \mathbb{I}^2 \) such that \( \|z_1\|_{TV} = \|z_2\|_{TV} \). The two new terms above gauge the mass difference between \( x_1 \) and \( x_2 \).

**Model mismatch.** Theorem 5 below bounds the recovery error \( d_{GW}(x, \hat{x}) \), where \( \hat{x} \) is a solution of Program (5). Even though \( x \) is an arbitrary nonnegative measure in this section, it can always be approximated with a well-separated sparse measure, up to some error with respect to the metric \( d_{GW} \). Indeed, given an integer \( K \) and \( \varepsilon \in (0, 1/2] \), there exists a \( K \)-sparse nonnegative measure \( x_{K,\varepsilon} \) that is \( \varepsilon \)-separated and approximates \( x \). More specifically, let us fix \( \lambda > 1 \) throughout. Then, for any \( \varepsilon \in (0, 1/2] \), there exists a \( K \)-sparse and \( \varepsilon \)-separated nonnegative measure \( x_{K,\varepsilon} \) such that
\[
R(x, K, \varepsilon) := d_{GW}(x, x_{K,\varepsilon}) \leq \lambda \inf d(x, \chi),
\]
where the infimum above is over every nonnegative \( K \)-sparse and \( \varepsilon \)-separated measure \( \chi \) supported on \( \text{interior}(\mathbb{I}^2) \). The case \( \lambda = 1 \) is excluded here because the infimum on the far-right of (14) might not be achieved. In what follows, the dependence of \( R(x, K, \varepsilon) \) on \( \lambda \) is suppressed to ease the notation. The residual \( R(x, K, \varepsilon) \) can be thought of as the mismatch in modelling \( x \) with a well-separated sparse measure.

For Program (5) to succeed in the general settings of this section, we impose additional assumptions on the imaging apparatus in the next two paragraphs.

**Smoothness.** We assume that the imaging operator \( \Phi \) in (4) is Lipschitz continuous, namely, there exists a constant \( L \geq 0 \) such that
\[
\left\| \int_{\mathbb{I}^2} \Phi(\theta)(x_1(d\theta) - x_2(d\theta)) \right\|_F \leq L \cdot d_{GW}(x_1, x_2),
\]
for every pair of measures \( x_1, x_2 \) supported on \( \mathbb{I}^2 \).

**T*-system.** To study the stability of Program (5), we also need to modify the notion of T-system in Definition 1. We begin with the next definition, which is immediately followed by an example.

**Definition 3. (Admissible sequence)** We say that \( \{\{\tau_n^k\}_{k=0}^M\}_{n \geq 1} \subset \mathbb{I} \) is a \((K, \varepsilon)\)-admissible sequence if
\begin{itemize}
  \item \( \tau_0^0 = 0 \) and \( \tau_M^0 = 1 \) for every \( n \), namely, the endpoints of \( \mathbb{I} = [0,1] \) are included in the sequence \( \{\tau_k^0\}_{k=0}^M \), for every \( n \). Moreover,
  \item as \( n \to \infty \), the sequence \( \{\tau_n^k\}_{k=1}^{M-1} \) converges (element-wise) to an \( \varepsilon \)-separated subset of \( \mathbb{I} \) with at most \( K \) unique points, where every element has an even multiplicity, except one element that appears only once.\(^1\)
\end{itemize}

\(^1\)That is, every element is repeated an even number of times (2, 4, \ldots) except one element that appears only once.
An example of admissible sequences is given in Figure 2. While T-system in Definition 1 is a condition on all increasing sequences of length $M$, the T*-system below is a condition only on admissible sequences; these are the only sequences that matter in our analysis.

Like a T-system, a T*-system imposes certain requirements on a family of functions: Whereas the performance of Program (5) for sparse measures and in the absence of noise relates to a certain T-system, the general performance of Program (5) relates to certain T*-systems, as we will see shortly in Theorem 5. The definition of T*-system below is immediately followed by its motivation.

**Definition 4. (T*-system)** For an even integer $M$, real-valued functions $\{\phi_m\}_{m=0}^M$ form a $T^*_K$-system on $I$ if any $(K, \varepsilon)$-admissible sequence $\{\{\tau_k^n\}_{k=0}^M\}_{n \geq 1}$ satisfies:

1. $\lim_{n \to \infty} \inf \|\phi_m(\tau_k^n)\|_{k,m=0}^M \geq 0$. That is, the determinant of $[\phi_m(\tau_k^n)]_{k,m=0}^M \in \mathbb{R}^{(M+1) \times (M+1)}$ is nonnegative in the limit of $n \to \infty$. Moreover,

2. all minors along the $I$th row of $[\phi_m(\tau_k^n)]_{k,m=0}^M$ approach zero at the same rate when $n \to \infty$. Here, $I$ is the index of the element of the limit sequence that appears only once.\(^2\)

Let us now provide some insight about T*-systems. In the proof of Proposition 2 for sparse and noise-free super-resolution, in order to generate a polynomial

$$\sum_{m=1}^M b_m \phi_m \geq 0,$$

with prescribed zeros on $I$, we assumed that $\{\phi_m\}_{m=1}^M$ form a T-system; see the discussion after Definition 1. On the other hand, to generate a polynomial

$$\sum_{m=1}^M b_m \phi_m \geq \phi_0,$$

with prescribed zeros on $I$, similarly $\{-\phi_0\} \cup \{\phi_m\}_{m=1}^M$ are required to form a T-system. The definition of T*-system above is based on the same idea but limited to admissible sequences, to ease the burden of verifying the conditions in Definition 4. In particular, Part 2 in Definition 4 excludes trivial polynomials, such as $0 \cdot \phi_0 + \sum_{m=1}^M m_m \phi_m$.

Let us take a moment to compare T- and T*-systems. An arbitrary polynomial of a T-system has a limited number of zeros, see the discussion after Definition 1. Polynomials of a T*-system, on the other hand, have no such property since the determinant in Part 1 of Definition 4 might vanish on $I$.

Instead, the notion of T*-system is designed to facilitate the construction of the necessary dual certificates for Program (5) in the presence of model mismatch and noise. In particular, as mentioned earlier, Part 2 in Definition 4 is designed to exclude trivial polynomials that do not qualify as dual certificates. Lastly, Definition 4 only considers admissible sequences to simplify the burden of verifying whether a family of functions form a T*-system.

To summarize, the widely-used notion of T-system in Definition 1 plays a key role in the analysis of sparse inverse problems in the absence of noise, whereas T*-system above is a new concept introduced in [1] and tailored for the *stability analysis* of sparse inverse problems. It was established in [1] that translated copies of the Gaussian window $e^{-t^2}$ indeed form a T*-system, under mild conditions on the translations specified in Section 2.3 below. We expect this to also hold for many other measurement windows with sufficiently fast decay.\(^3\)

We are now ready to present our main result about the performance of Program (5) in the general case where $x$ is an arbitrary nonnegative measure on $I^2$ and in the presence of additive noise. Theorem 5, proved in Section 4.4, states that Program (5) approximately recovers $x$ provided that certain T- and T*-systems exist. As an example of this very general result, Section 2.3 later applies Theorem 5 to the special case of imaging with Gaussian blur.

\(^2\)A nonnegative sequence $\{u^n\}_{n \geq 1}$ approaches zero at the rate $n^{-p}$ if $u_n = \Theta(n^{-p})$; see for example [29].

\(^3\)The definition of T*-system here is slightly different from that in [1] but the difference is inconsequential. The new definition here will help improve the dependence of recovery error bounds on number of observations both in one- and two-dimensional problems.
Theorem 5. (Arbitrary measure with noise) Consider a nonnegative measure $x$ supported on $\mathbb{I}^2$. Consider also a noise level $\delta \geq 0$, measurement functions $\{\phi_m\}_{m=1}^M$, the corresponding $L$-Lipschitz imaging operator $\Phi$, and the image $y \in \mathbb{R}^{M \times M}$.

For an integer $K$ and $\varepsilon \in (0, 1/2]$, let $x_{K,\varepsilon}$ be a $K$-sparse and $\varepsilon$-separated nonnegative measure on $\mathbb{I}^2$ that approximates $x$ in the sense of (14). In particular, let $\Theta = \{\theta_k\}_{k=1}^K = \{(t_k, s_k)\}_{k=1}^K \subset \text{interior}(\mathbb{I}^2)$ be the support of $x_{K,\varepsilon}$, and set $T = \{t_k\}_{k=1}^K$ and $S = \{s_k\}_{k=1}^K$ for short.

With $\tilde{x}$ denoting a solution of Program (5) for $\delta' \geq (1 + L \cdot R(x, k, \varepsilon))\delta$, it holds that

$$d_{GW}(x, \tilde{x}) \leq c_1 \delta + c_2 \varepsilon + c_3 R(x, K, \varepsilon),$$

(16)

where $d_{GW}$ is the generalized Wasserstein metric in (13). The coefficients $c_1, c_2, c_3$ above are specified explicitly in (40), and depend on the true measure $x$, the separation $\varepsilon$, and the imaging operator $\Phi$.

The error bound in (16) holds if the following requirements are met. For every index set $\Omega \subset [K]$ and $k \in [K]$, we define the functions

$$F_{T_0}(t) := \begin{cases} 0, & \text{when there exists } k \in \Omega \text{ such that } |t - t_k| \leq \varepsilon/2, \\ 1, & \text{elsewhere on } \text{interior}(\mathbb{I}), \end{cases}$$

$$F_{S_0}(s) := \begin{cases} 0, & \text{when there exists } k \in \Omega \text{ such that } |s - s_k| \leq \varepsilon/2, \\ 1, & \text{elsewhere on } \text{interior}(\mathbb{I}), \end{cases}$$

$$F_{T}^{k+}(t) := \begin{cases} \pm 1, & \text{when } |t - t_k| \leq \varepsilon/2, \\ 0, & \text{everywhere else on } \text{interior}(\mathbb{I}), \end{cases}$$

$$F_{S}^{k+}(s) := \begin{cases} 1, & \text{when } |s - s_k| \leq \varepsilon/2, \\ 0, & \text{everywhere else on } \text{interior}(\mathbb{I}), \end{cases}$$

for $t, s \in \mathbb{I}$, with an example given in Figure 3e.\footnote{The purple graph in Figure 3e is an example of $F_{\{t_1\}}$, denoted in the figure by $F_{t_1}$ to ease the notation.} With $M \geq 2K + 2$, we must have that

- $\{\phi_m\}_{m=1}^M$ form a $T$-system on $\mathbb{I}$,
- $\{F_{T_0}\} \cup \{\phi_m\}_{m=1}^M$ and $\{F_{S_0}\} \cup \{\phi_m\}_{m=1}^M$ both form $T_{K,\varepsilon}$-systems on $\mathbb{I}$ for every $\Omega \subset [K]$,
- $\{F_{T}^{k+}\} \cup \{\phi_m\}_{m=1}^M$ and $\{F_{T}^{k-}\} \cup \{\phi_m\}_{m=1}^M$ both form $T_{K,\varepsilon}$-systems on $\mathbb{I}$ for every $k \in [K]$,
- $\{F_{S}^{k+}\} \cup \{\phi_m\}_{m=1}^M$ form a $T_{K,\varepsilon}$-system on $\mathbb{I}$ for every $k \in [K]$.

Theorem 5 for image super-resolution is unique in a number ways. The differences with prior work are further discussed in Section 3 and also summarized here. First, Theorem 5 applies to arbitrary measures, not only atomic ones. In particular, for atomic measures, no minimum separation or limit on the density of impulses are imposed in contrast to earlier results [2, 13, 14, 30].
Moreover, Theorem 5 addresses both noise and model-mismatch in image super-resolution. Indeed, even in the 1-D case, stability was identified as a technical obstacle in earlier work [2]. In addition, the recovery error in Theorem 5 is quantified with a natural metric between measures, namely, the generalized Wasserstein metric, in contrast to prior work; see for example [31] that separately studies the error near and away from the impulses. Lastly, the measurement functions \( \{ \phi_m \} \) are required to be continuous rather than (several times) differentiable [2, 30]. All this is achieved without the need to explicitly regularize for sparsity in Program (5).

Several remarks are in order to clarify Theorem 5.

Proof technique. For Program (5) to successfully recover a sparse measure in the absence of noise, we needed to construct a nonnegative polynomial \( Q(\theta) \) in the span of measurement functions with zeros only at the impulse locations \( \Theta = \{ \theta_k \}_{k=1}^K = \{ (t_k, s_k) \}_{k=1}^K \); see the discussion after Proposition 2. For approximate recovery in the presence of model mismatch and noise, perhaps not surprisingly we need to construct a nonnegative polynomial \( Q(\theta) \) that is bounded away from zero far from the impulse locations \( \Theta \), namely,

\[
Q(\theta) \geq g > 0, \quad \forall \theta \text{ far from } \Theta,
\]

and a positive scalar \( g \). Letting \( T = \{ t_k \}_{k=1}^K \) and \( S = \{ s_k \}_{k=1}^K \) for short, the proof of Theorem 5 constructs \( Q \) by combining certain univariate polynomials, similar to the proof of Proposition 2 which was summarized earlier in Section 2.1 and Figure 1. Among these univariate polynomials, for example, the proof constructs a nonnegative polynomial \( q_T \) such that

\[
q_T(t) \geq 1, \quad \forall t \text{ far from } T.
\]

To that end, the proof requires that \( \{ F_T \} \cup \{ \phi_m \}_{m=1}^M \) form a T*-system. In addition to \( Q \), we also find it necessary to construct yet another polynomial \( Q^0 \) to complete the proof of Theorem 5; see Section 4.4 for more details. Figure 3 illustrates some of the ideas used in the proof of Theorem 5.
Theorem 5, the existence of these univariate polynomials follows from [1]. At last, we obtain a nonnegative polynomial $Q$ that is zero near and equal to $\theta$. This lower bound for polynomial $Q$ is shown in Figure 3 as a heat map, with warmer colors corresponding to larger values, namely blue corresponds to zero and green corresponds to $\theta$. Let us first express this lower bound on $Q$ in terms of univariate functions. To that end, consider a function $F_{t_1}(t)$ that is zero near and equal to $\sqrt{\mathcal{F}}$ away from $t_1$. Likewise, consider similar nonnegative functions $F_{t_2}(t), F_{s_1}(s), F_{s_2}(s)$ that are zero near and equal to $\sqrt{\mathcal{F}}$ away from $t_2, s_1, s_2$, respectively. Figure 3b shows the heat map of $F_{t_1}(t)F_{s_2}(s)$, Figure 3c shows the heat map of $F_{t_2}(t)F_{s_1}(s)$, and lastly Figure 3d shows the heat map of their sum, namely, $G(\theta) := F_{t_1}(t)F_{s_2}(s) + F_{t_2}(t)F_{s_1}(s)$. Note that $G$ is zero near and larger than $\mathcal{F}$ away from the impulse locations $\theta_1, \theta_2$, as desired.

It only remains to construct univariate polynomials $q_{t_1}, q_{t_2}, q_{s_1}, q_{s_2} \in \text{span}(\phi_1, \ldots, \phi_M)$ that satisfy the inequalities $q_{t_1} \geq F_{t_1}, q_{t_2} \geq F_{t_2}, q_{s_1} \geq F_{s_1}, q_{s_2} \geq F_{s_2}$, and equality holds at $t_1, t_2, s_1, s_2$, respectively. Under the conditions in Theorem 5, the existence of these univariate polynomials follows from [1]. At last, we obtain a nonnegative polynomial $Q(\theta) = q_{t_1}(t)q_{s_2}(s) + q_{t_2}(t)q_{s_1}(s)$ that is zero at the impulse locations and larger than $\mathcal{F}$ away from the impulses, as desired. For example, for the Gaussian window detailed in Section 2.3 with standard deviation $\sigma = 0.2$, Figure 3e shows $q_{t_1}(t)$ and Figure 3f shows the heat map of the dual certificate $Q(\theta)$, both in logarithmic scale. Yet another polynomial $Q^0$ is needed to complete the proof of Theorem 5, which is detailed in the proof; see Section 4.4.

Recovery error. The bound on the recovery error $d_{GW}(x, \tilde{x})$ in (16) depends on the noise level $\delta$ and on how well $x$ can be approximated with a well-separated sparse measure. More specifically, for any $\varepsilon \in (0, 1/2)$, $x$ can be approximated with a $K$-sparse and $\varepsilon$-separated measure $x_{K,\varepsilon}$, with a residual of $R(x, K, \varepsilon)$; see (14). We might then think of a solution $\tilde{x}$ of Program (5) as an estimate for $x_{K,\varepsilon}$ and therefore an estimate for $x$, up to the residual $R(x, K, \varepsilon)$. Both the separation $\varepsilon$ and the residual $R(x, K, \varepsilon)$ appear on the right-hand side of the error bound (16).

In particular, when $\delta, \varepsilon, R(x, K, \varepsilon) \to 0$, we again obtain Proposition 2 for recovery of $K$-sparse nonneg-
Theorem 5 applies to any nonnegative measure \( x \). In particular, when \( x \) is an atomic measure, Theorem 5 applies regardless of the separation between the impulses present in \( x \). However, the recovery error \( d_{GW}(x, \hat{x}) \) does depend on the separation of \( x \).

As an example, consider \( x = \delta_{0.5} + \delta_{0.51} \). In order to apply Theorem 5, we can set \( \varepsilon = 0.01 \), so that \( x_{K,\varepsilon} = x \) and \( R(x, K, \varepsilon) = 0 \). Now, the error bound in (16) reads as

\[
d_{GW}(x, \hat{x}) \leq C_1 \delta + 0.01 C_2, \tag{17}
\]

where \( C_1 = c_1(x, \varepsilon) \) and \( C_2 = c_2(x, \varepsilon) \) with \( \varepsilon = 0.01 \). Alternatively, we may also apply Theorem 5 by setting \( \varepsilon = 0.2 \), so that \( x_{K,\varepsilon} = \delta_{0.405} + \delta_{0.605} \) and \( R(x, K, \varepsilon) = 0.19 \). In this case, (16) reads as

\[
d_{GW}(x, \hat{x}) \leq C'_1 \delta + 0.2 C'_2 + 0.19 C'_3, \tag{18}
\]

where \( C'_1 = c_1(x, \varepsilon) \), \( C'_2 = c_2(x, \varepsilon) \), and \( C'_3 = c_3(x, \varepsilon) \) with \( \varepsilon = 0.2 \). Informally speaking, one would expect the bound on recovery error in (18) to be smaller than the bound in (17) and hence better, because resolving nearby impulses is more difficult and one would expect \( c_2(x, \varepsilon) \) to implode as \( \varepsilon \to 0 \). However, as mentioned in the previous paragraph, this work does not address the optimal choice of separation \( \varepsilon \) as a function of noise level \( \delta \). That is, given \( \delta \), the choice of \( \varepsilon \) that would minimize the right-hand of (16) is not studied here; see [1] for more details.

### Invariance

Although not mentioned in Theorem 5, as a sanity check, one might verify that the error bound on the right-hand side of (16) is invariant under scaling of the noise level \( \delta \) and the imaging operator \( \Phi \). If we replace \( \delta \) with \( \alpha \delta \) and \( \Phi \) with \( \alpha \Phi \) for a positive \( \alpha \), the right-hand side of (16) does not change; see [1] for more details.

### 2.3 Example with Gaussian Window

As an example of the general super-resolution framework presented in this paper, consider the case where \( x \) is a \( K \)-sparse nonnegative measure as in (7) and \( \{\phi_m\}_{m=1}^M \) are translated copies of a one-dimensional Gaussian window, namely,

\[
\phi_m(t) = g_1(t - t'_m) := e^{-\frac{(t - t'_m)^2}{\sigma^2}}, \tag{19}
\]

for \( T' = \{t'_m\}_{m=1}^M \subset \mathbb{I} \) and positive standard deviation \( \sigma \). Note that

\[
\int_{\mathbb{I}} \phi_m(t)\phi_n(s)\,x(dt, ds) = \sum_{k=1}^K a_k \cdot g_1(t_k - t'_m)g_1(t_k - t'_n) \\
= \sum_{k=1}^K a_k \cdot g_2(\theta_k - \theta'_{m,n}) \\
= \sum_{k=1}^K a_k \cdot e^{-\frac{\|\theta_k - \theta'_{m,n}\|^2}{\sigma^2}}, \tag{20}
\]

where \( \theta_k = (t_k, s_k), \theta'_{m,n} = (t'_m, t'_n) \) and \( g_2 \) is a 2-D Gaussian window, which can be thought of as the point-spread function. Note that we might also think of \( \{\theta'_{m,n}\}_{m,n=1}^M = T' \times T' \) as the “sampling points” in the sense that

\[
\int_{\mathbb{I}} \phi_m(t)\phi_n(s)\,x(dt, ds) = \sum_{k=1}^K a_k \cdot g_2(\theta_k - \theta'_{m,n}) = (g_2 \ast x)(\theta'_{m,n}), \quad m, n \in [M], \tag{21}
\]
Proposition 6. (Gaussian window) Consider a nonnegative measure $x$ supported on $\mathbb{I}^2$. Consider also a noise level $\delta \geq 0$ and measurement functions defined in (19). For an integer $K$ and $\varepsilon \in (0, 1/2]$, let $x_{K,\varepsilon}$ be a $K$-sparse and $\varepsilon$-separated nonnegative measure on $\mathbb{I}^2$ that approximates $x$ in the sense of (14). With $M \geq 2K + 2$, let $\hat{x}$ be a solution of Program (5) with
\[
\delta' \geq \left(1 + \frac{2M}{\sigma^2} R(x, k, \varepsilon)\right) \delta,
\]
see (14). Then it holds that
\[
d_{GW}(x, \hat{x}) \leq c_1\delta + c_2\delta + c_3 R(x, K, \varepsilon),
\]
where $d_{GW}$ is the generalized Wasserstein metric in (13). The coefficients $c_1, c_2, c_3$ above are specified explicitly in (40), and depend on the true measure $x$, the separation $\varepsilon$, and the imaging operator $\Phi$.

3 Related Work

The current wave of super-resolution research using convex optimization began with the two seminal papers of Candès and Fernandez-Granda [12, 25]. In those papers, the authors showed that a convex program with a sparse-promoting regularizer stably recovers a complex atomic measure from the low-end of its spectrum. This holds true if the minimal separation between any two spikes is inversely proportional to the maximal measured frequency, namely the “bandwidth” of the sensing mechanism. Many papers extended this fundamental result to randomized models [27], support recovery analysis [32, 33, 34, 31, 35], denoising schemes [36, 37], different geometries [38, 39, 40, 41, 42], and incorporating prior information [43]. Most of these works easily generalize to multi-dimensional signals. In addition, a special attention to multi-dimensional signals was given in a variety of papers; see for instance [44, 45].

The separation condition above is unnecessary for nonnegative measures, and this is the important regime on which this paper and most of this review focuses. There are a number of works that study nonnegative sparse super-resolution for atomic measures supported on a grid. In [14, 13], it was shown that for such 1-D or 2-D signals, stable reconstruction is possible without imposing a separation condition, but instead requiring a milder condition on the density of the impulses. In particular, the error grows exponentially fast as the density of the spikes increases. A similar result was derived for signals on the sphere [46].

In this paper, we focus on the grid-free setting in which the nonnegative measure is not necessarily supported on a predefined grid. This is the most general regime and requires more advanced machinery and algorithms. In [2], it was shown that in the absence of noise, a convex program with TV regularizer can recover—without imposing any separation—an atomic measure on the real line [2]. The same holds on other geometries as well [38, Section 5]. However, all these results have no stability guarantees, assume a differentiable point spread function and make use of a TV regularizer to promote sparsity. Our Proposition 2...
and Theorem 5 address all these shortcomings and solve the sparse (grid-free) image super-resolution problem in its most general form. The leap from the 1-D results of [1] to 2-D requires new techniques since the key technical ingredient, namely T-systems, does not extend to higher dimensions.

Let us add that the low-noise regime for positive 1-D super-resolution was studied in [18]. There, it was shown that a convex program with a sparse-promoting regularizer results in the same number of spikes as the original measure when noise is small. Furthermore, the solution converges to the underlying positive measure if the signal-to-noise ratio scales like $O(1/\text{sep}^2)$, where sep is the minimal separation between adjacent spikes. In contrast to our work, the framework of [18] builds upon smooth convolution kernel and uses a sparse-promoting regularizer, rather than the feasibility problem considered in Program (5). In [47], it was shown that the 2-D version of the same program enjoys similar properties for a pair of spikes.

Going back to signed measures, another line of work is based on various generalizations of Prony’s method [48], which encodes the support of the measure as zeros of a designed polynomial. Such generalizations include methods like MUSIC [49], Matrix Pencil [50], ESPRIT [51], to name a few. In 1-D and in the absence of noise, these methods are guaranteed to achieve exact recovery for a complex measure without enforcing any separation. This is not true for convex programs in which separation is a necessary condition, see [52]. The separation is not necessary for convex programs only for nonnegative measures, like the model considered in this paper. Stability analysis of some of these methods, under a separation condition, is found in [53, 54, 55, 56, 57]. However, their extension to 2-D is not trivial and accordingly different methods were proposed [58, 59, 60, 61, 62]. To the best of our knowledge, the stability of these algorithms for two or higher dimensions is not understood. That being said, we do not claim that convex programs are numerically superior over the Prony-like techniques and we leave comprehensive numerical study for future research.

4 Theory

4.1 Notation

At the risk of being redundant, let us collect here some of the notation used throughout this paper. For positive $\varepsilon$ and $T = \{t_k\}_{k=1}^K \subset \mathbb{I}$, let us define the neighbourhoods

$$t_{k,\varepsilon} := \{t \in \mathbb{I} : |t - t_k| \leq \varepsilon\} \subset \mathbb{I},$$

$$T_\varepsilon := \bigcup_{k=1}^K t_{k,\varepsilon},$$

and let $t_{k,\varepsilon}^C$ and $T_\varepsilon^C$ be the complements of these sets with respect to $\mathbb{I}$. Let us also sep($T$) denote the minimum separation of $T$, namely, the largest number $\nu$ for which

$$\nu \leq |t_k - t_l|, \quad k \neq l, \; k, l \in [K],$$

$$\nu \leq |t_k - 0|, \quad \nu \leq |t_k - 1|.$$  \hspace{1cm} (24)

Likewise, for positive $\varepsilon$ and $\Theta \subset \{\theta_k\}_{k=1}^K = \{(t_k, s_k)\}_{k=1}^K \subset \mathbb{I}^2$, we define the neighbourhoods

$$\theta_{k,\varepsilon} := t_{k,\varepsilon} \times s_{k,\varepsilon} = \{\theta \in \mathbb{I}^2 : \|\theta - \theta_k\|_\infty \leq \varepsilon\} \subset \mathbb{I}^2,$$

$$\Theta_\varepsilon := \bigcup_{k=1}^K \theta_{k,\varepsilon} \subseteq T_\varepsilon \times S_\varepsilon,$$

and let $\theta_{k,\varepsilon}^C$ and $\Theta_\varepsilon^C$ be the complements of these sets with respect to $\mathbb{I}^2$. Above, $\|\theta\|_\infty = \max[|t|, |s|]$ for $\theta = (t, s)$. Similarly, define the minimum separation of $\Theta$, namely the smallest number $\nu$ for which both (24) holds and

$$\nu \leq |s_k - s_l|, \quad k \neq l, \; k, l \in [K],$$

$$\nu \leq |s_k - 0|, \quad \nu \leq |s_k - 1|.$$  \hspace{1cm} (26)
4.2 Proof of Proposition 2 (Sparse Measure Without Noise)

The following standard result is an immediate extension of [1, Lemma 9] and, roughly speaking, states that Program (5) is successful if a certain dual certificate $Q$ exist.

Lemma 7. Let $x$ be a $K$-sparse nonnegative atomic measure supported on $\Theta \subset \text{interior}(I^2)$, see (7). Then $x$ is the unique solution of Program (5) with $\delta' = 0$ if

- the $M^2 \times K$ matrix $[\phi_m(t_k)\phi_n(s_k)]_{m=M,n=M,k=K}^{m=m,n=n,k=k}$ has full column rank, and

- there exist real coefficients $\{b_{m,n}\}_{m=1}^M$ and polynomial $Q(t,s) = \sum_{m,n=1}^M b_{m,n}\phi_m(t)\phi_n(s)$ such that $Q$ is nonnegative on $\text{interior}(I^2)$ and vanishes only on $\Theta$.

The following result, proved in Appendix A, states that the dual certificate required in Lemma 7 exists if the number of measurements $M$ is large enough and the measurement functions $\{\phi_m\}_{m=1}^M$ form a T-system on $I$. The technical difficulty arises from the fact that T-systems cannot be generalized to two dimensions. To prove this claim, we effectively reduce the construction of a polynomial $Q(\theta)$ with $\theta = (t,s)$ into the construction of a number of univariate polynomials in $t$ and $s$. The key observation is the following. Recall that $\Theta = \{\theta_{k}\}_{k=1}^K = \{(t_k,s_k)\}_{k=1}^K$ are the impulse locations and let $T = \{t_k\}_{k=1}^K$, $S = \{s_k\}_{k=1}^K$ for short, so that $\Theta \subseteq (T \times S)$. Suppose that a univariate polynomial $q_T$ is nonnegative on $I$ and only vanishes on $T$. Similarly, consider a polynomial $q_S$ that is nonnegative on $I$ and only vanishes on $S$. Then the polynomial $Q(\theta) = (q_T(t) + q_S(s))^2$ is nonnegative on $I^2$ and vanishes only on $T \times S$. However, in general $\Theta \subset (T \times S)$, and consequently $Q$ will have unwanted zeros on $(T \times S) \setminus \Theta$. A somewhat more nuanced argument is therefore required to construct a polynomial that vanishes exactly on $\Theta$ and not on $(T \times S) \setminus \Theta$, see Appendix A for the details.

Lemma 8. (Sparse measure without noise) Let $x$ be a $K$-sparse nonnegative atomic measure supported on $\text{int}(I^2)$. For $M \geq 2K + 1$, suppose that $\{\phi_m\}_{m=1}^M$ form a T-system on $I$. Then, the dual certificate $Q$ prescribed in Lemma 7 exists.

Combining Lemmas 7 and 8 completes the proof of Proposition 2.

4.3 Geometric Intuition for Proposition 2

Proposition 2 states that the imaging operator $\Phi$ in (4) is injective on all $K$-sparse nonnegative measures (such as $x$) provided that we take enough observations ($M \geq 2K + 1$) and the measurement functions $\{\phi_m\}_{m=1}^M$ form a T-system on $I$. Here, we provide some geometrical intuition for the role of the dual certificate. We mention that the dual certificate was derived using a variety of alternative arguments in the compressed sensing and super-resolution literature [63, 12].

Let us denote $\theta = (t,s)$ for short and consider the conic hull of the dictionary $\{\Phi(\theta)\}_{\theta \in I^2}$ defined as

$$C := \left\{ \int_{I^2} \Phi(\theta) \chi(d\theta) : \chi \text{ is a nonnegative measure on } I^2 \right\} \subset \mathbb{R}^{M \times M}. \tag{27}$$

By the continuity of $\Phi$ and with an application of the dominated convergence theorem, it is easy to verify that $C$ is a closed convex cone, namely $C$ is a homogeneous and closed convex subset of $\mathbb{R}^{M \times M}$. When $\{\phi_m\}_{m=1}^M$ form a T-system on $I$, it also not difficult to verify that $\{\Phi(t_l,s_l)\}_{l=1}^L$ are linearly independent matrices in $\mathbb{R}^{M \times M}$. This in particular implies that $C$ is a convex body, namely, the interior of $C$ is not empty. Note also that $y \in C$ because

$$y = \int_{I^2} \Phi(\theta) x(d\theta) = \sum_{k=1}^K a_k \Phi(\theta_k). \tag{28}$$

For Program (5) to successfully recover $x$, it is necessary that

$$\mathcal{A} = \text{cone} \left( \{\Phi(\theta_k)\}_{k=1}^K \right) = \left\{ \sum_{k=1}^K a_k \Phi(\theta_k) : a_k \geq 0, \forall k \in [K] \right\}, \tag{29}$$

14
is a $K$-dimensional face of the cone $C$. This happens if and only if we can find a hyperplane with normal vector $b \in \mathbb{R}^{M \times M}$ that strictly supports the cone $C$ at $A$, namely, when we can find $b$ such that

$$\begin{cases} 
\langle b, c \rangle = 0, & \forall c \in A, \\
\langle b, c \rangle > 0, & \forall c \in C \setminus A.
\end{cases} \quad (30)$$

Invoking (29), we find that (30) is equivalent to finding $b \in \mathbb{R}^{M \times M}$ such that

$$\begin{cases} 
\langle b, \Phi(\theta_k) \rangle = 0, & k \in [K], \\
\langle b, \Phi(\theta) \rangle > 0, & \theta \notin \{\theta_k\}_{k=1}^K.
\end{cases} \quad (31)$$

In other words, it is necessary to find a “polynomial”

$$Q(\theta) = Q(t, s) := \langle b, \Phi(\theta) \rangle = \sum_{m,n=1}^{M} b_{m,n} \phi_m(t) \phi_n(s),$$

that vanishes on $\{\theta_k\}_{k=1}^K$ and is positive elsewhere on $\mathbb{T}^2$. Building on the results in [1], we construct one such polynomial in Section 4.2 when $M \geq 2K + 1$. It is worth noting that the polar of the cone $C$, itself another convex cone in $\mathbb{R}^{M \times M}$, consists of the coefficients of all nonnegative polynomials of $\{\phi_m \phi_n\}_{m,n=1}^M$ on $\mathbb{T}^2$ and in particular the coefficient vector $b$ above belongs to an $(M^2 - K)$-dimensional face of this polar cone [22]. We refer the reader to Figure 4 for an illustration of the convex geometry underlying the problem of nonnegative super-resolution.

![Figure 4](image)

Figure 4: This figure complements Section 4.3. Figure 4a shows the conic hull $C$ of the trajectory of the imaging operator $\Phi : \mathbb{T}^2 \to \mathbb{R}^{M \times M}$. Note that the image $y$ belongs to the cone $C$, see (28). For Program (5) to successfully recover the true measure $x$ from image $y$, it is necessary that $\{\Phi(\theta_k)\}_{k=1}^K$ form a $K$-dimensional face of the cone $C$, where $\{\theta_k\}_{k=1}^K$ is the support of $x$. This face is shown in green. That is, this condition is necessary for $x$ to be the unique solution of Program (5). Equivalently, it is necessary to find a hyperplane, with normal vector $b$, that strictly supports $C$ on this face. This condition can be interpreted as finding a nonnegative polynomial of $\{\phi_m(t) \phi_n(s)\}_{m,n=1}^M$ with zeros exactly on the support of $x$. 

15
4.4 Proof of Theorem 5 (Arbitrary Measure with Noise)

In this section, we will prove the main result of this paper, Theorem 5. For an integer $K$ and $\varepsilon \in (0, 1/2]$, let $x_{K,\varepsilon}$ be a $K$-sparse and $\varepsilon$-separated nonnegative measure on $\mathbb{R}^d$ that approximates $x$ in the sense of (14). Let $\Theta = \{\theta_k\}_{k=1}^K = \{(t_k, s_k)\}_{k=1}^K \subseteq \text{interior}(\mathbb{I})^2$ be the support of $x_{K,\varepsilon}$, and set $T = \{t_k\}_{k=1}^K$ and $S = \{s_k\}_{k=1}^K$ for short. Consider also the neighbourhoods $\Theta_k = \{\theta_k\}_{k=1}^K \subseteq \mathbb{I}^2$ and $\Theta \varepsilon = \bigcup_{k=1}^K \Theta_k$ defined in (25). Let $\Theta^C_k$ and $\Theta^C \varepsilon$ denote the complements of these sets with respect to $\mathbb{I}^2$.

In the rest of this section, we first bound the error $d_{GW}(x_{K,\varepsilon}, \hat{x})$. Then we apply the triangle inequality to control $d_{GW}(x, \hat{x})$ by

$$d_{GW}(x, \hat{x}) \leq d_{GW}(x, x_{K,\varepsilon}) + d_{GW}(x_{K,\varepsilon}, \hat{x}) \leq R(x, K, \varepsilon) + d_{GW}(x_{K,\varepsilon}, \hat{x}),$$

(see (14)) thereby completing the proof of Theorem 5. To control $d_{GW}(x_{K,\varepsilon}, \hat{x})$, we will first show in Section 4.4.1 that the existence of certain dual certificates leads to stable recovery of $x_{K,\varepsilon}$ with Program (5). Then Section 4.4.2 proves that these certificates exist under certain conditions on the imaging apparatus.

4.4.1 Dual Certificates

Lemmas 9 and 10 below show that Program (5) stably recovers $x_{K,\varepsilon}$ in the presence of noise, provided that certain dual certificates exist. The proofs are standard and in particular Lemmas 9 and 10 below are immediate extensions, respectively, of Lemmas 5 and 7 in [1]. Both results were at length discussed in [1]. In short, Lemma 9 below controls the recovery error away from the support $\Theta$ of $x_{K,\varepsilon}$ and Lemma 10 below controls the error near the support.

Lemma 9. (Error away from the support) Let $\hat{x}$ be a solution of Program (5) with $\delta'$ specified in (39) and set $h := \hat{x} - x_{K,\varepsilon}$ to be the error. Fix a positive scalar $\bar{\delta}$. Suppose that there exist real coefficients $\{b_{m,n}\}_{m,n=1}^M$ and a polynomial

$$Q(\theta) = Q(t, s) = \sum_{m,n=1}^M b_{m,n}\phi_m(t)\phi_n(s),$$

such that

$$Q(\theta) \geq G(\theta) := \begin{cases} 0, & \text{when there exists } k \in [K]\text{ such that } \theta \in \Theta_k, \\ \bar{\delta}, & \text{elsewhere on } \text{interior}(\mathbb{I}^2), \end{cases}$$

where the equality holds on $\Theta = \{\theta_k\}_{k=1}^K$. Then we have that

$$\int_{\Theta^C} h(d\theta) \leq 2\|b\|_F \bar{\delta}/\bar{\delta},$$

(32)

where $b \in \mathbb{R}^{M \times M}$ is the matrix formed by the coefficients $\{b_{m,n}\}_{m,n=1}^M$.

Lemma 10. (Error near the support) Suppose that the dual certificate $Q$ in Lemma 9 exists. Suppose also that there exist real coefficients $\{b^0_{m,n}\}_{m,n=1}^M$ and a polynomial

$$Q^0(\theta) = Q^0(t, s) = \sum_{m,n=1}^M b^0_{m,n}\phi_m(t)\phi_n(s),$$

such that

$$Q^0(\theta) \geq G^0(\theta) := \begin{cases} 1, & \text{when there exists } k \in [K]\text{ such that } \theta \in \Theta_k, \text{ and } \int_{\Theta_k} h(d\theta) > 0, \\ -1, & \text{when there exists } k \in [K]\text{ such that } \theta \in \Theta_k, \text{ and } \int_{\Theta_k} h(d\theta) \leq 0, \\ 0, & \text{everywhere else on } \text{interior}(\mathbb{I}^2), \end{cases}$$

(33)

where the equality holds on $\Theta$. Then we have that

$$\sum_{k=1}^K \left| \int_{\Theta_k} h(d\theta) \right| \leq 2 (\|b\|_F + \|b^0\|_F) \delta',$$

(34)

where $b^0 \in \mathbb{R}^{M \times M}$ is the matrix formed by the coefficients $\{b^0_{m,n}\}_{m,n=1}^M$. 

16
By combining Lemmas 9 and 10, the next result bounds the error $d_{GW}(x_{K,\varepsilon}, \hat{x})$. The proof is omitted as it is identical to that of Lemma 18 in [1].

**Lemma 11. (Error in Wassertein metric)** Suppose that the dual certificates $Q$ and $Q^0$ in 9 and 10 exist. Then it holds that

$$d_{GW}(x_{K,\varepsilon}, \hat{x}) \leq \left( \left( 6 + \frac{2}{\bar{g}} \right) \|b\|_F + 6\|b^0\|_F \right) \delta' + \frac{\varepsilon}{2} \|x_{K,\varepsilon}\|_{TV}. \quad (35)$$

In order to apply Lemma 11, we must first show that the dual certificates $Q$ and $Q^0$ specified in Lemmas 9 and 10 exist. In the next section, we construct these certificates under certain conditions on the imaging apparatus.

**4.4.2 Existence of the Dual Certificates**

To prove the existence of the dual certificates required in Lemmas 9 and 10, some preparation is necessary. Recall that $\Theta = \{\theta_k\}_{k=1}^K = \{(t_k, s_k)\}_{k=1}^K$ are the impulse locations and let $T = \{t_k\}_{k=1}^K$ and $S = \{s_k\}_{k=1}^K$ for short. In particular, $\Theta \subseteq T \times S$. For an index set $\Omega \subseteq [K]$ and its complement $[K] \setminus \Omega$, we set $T_\Omega = \{t_k\}_{k \in \Omega}$ and $S_{[K] \setminus \Omega} = \{s_k\}_{k \in [K] \setminus \Omega}$. For a finite set $T' \subseteq \mathbb{I}$ and positive $\varepsilon \leq \text{sep}(T')$, let us define the function

$$F_{T'}(t) = \begin{cases} 0, & \text{when there exists } t' \in T' \text{ such that } t \in t'_\varepsilon, \\ 1, & \text{elsewhere on } \mathbb{I}, \end{cases} \quad (36)$$

where $t'_\varepsilon = \{t \in \mathbb{I} : |t - t'| \leq \varepsilon\}$. Loosely speaking, the following result, proved in Appendix B, states that the dual certificate required in Lemma 9 exists if both $\{F_{T_\Omega}\} \cup \{\phi_m\}_{m=1}^M$ and $\{F_{S_{[K] \setminus \Omega}}\} \cup \{\phi_m\}_{m=1}^M$ are $T^*$-systems for any index set $\Omega \subseteq [K]$.

**Proposition 12.** Suppose that $\{\phi_m\}_{m=1}^M$ form a $T$-system on $\mathbb{I}$ with $M \geq 2K + 2$. Let

$$\bar{g} = 2K^{-2}, \quad (37)$$

where $\bar{g}$ is the constant from Lemma 9. For every index set $\Omega \subseteq [K]$, suppose also that $\{F_{T_\Omega}\} \cup \{\phi_m\}_{m=1}^M$ and $\{F_{S_{[K] \setminus \Omega}}\} \cup \{\phi_m\}_{m=1}^M$ are both $T^*_K$-systems on $\mathbb{I}$, see (36). Then the dual certificate $Q$ specified in Lemma 9 exists.

For positive $\varepsilon \leq \text{sep}(\Theta)$ and $k \in [K]$, let us define the functions

$$F_T^{k+}(t) := \begin{cases} \pm 1, & \text{when } t \in t_{K,\varepsilon}, \\ 0, & \text{elsewhere on } \mathbb{I}, \end{cases}$$

$$F_S^{k+}(s) := \begin{cases} 1, & \text{when } s \in s_{K,\varepsilon}, \\ 0, & \text{elsewhere on } \mathbb{I}. \end{cases} \quad (38)$$

The following result, proved in Appendix C, states that the dual certificate required in 10 exists when certain $T^*$-systems exist.

**Proposition 13.** For $M \geq 2K + 2$, suppose that $\{\phi_m\}_{m=1}^M$ form a $T$-system on $\mathbb{I}$. For every $k \in [K]$, suppose also that $\{F_T^{k+}\} \cup \{\phi_m\}_{m=1}^M$, $\{F_T^{k-}\} \cup \{\phi_m\}_{m=1}^M$, and $\{F_S^{k+}\} \cup \{\phi_m\}_{m=1}^M$ are all $T^*_K$-systems on $\mathbb{I}$, see (38). Then the dual certificate $Q^0$ in Lemma 10 exists.

**4.4.3 Completing the Proof of Theorem 5**

Recall that the imaging operator $\Phi$ is $L$-Lipschitz, see (15). Using the triangle inequality, it follows that

$$\left\| y - \int_{\mathbb{I}^2} \Phi(\theta)x_{K,\varepsilon}(d\theta) \right\|_F \leq \left\| y - \int_{\mathbb{I}^2} \Phi(\theta)x(d\theta) \right\|_F + \left\| \int_{\mathbb{I}^2} \Phi(\theta)(x(d\theta) - x_{K,\varepsilon}(d\theta)) \right\|_F \leq \delta + L \cdot d_{GW}(x, x_{K,\varepsilon}) \quad (\text{see (4,15)})$$

$$= \delta + L \cdot R(x, K, \varepsilon) := \delta'. \quad (\text{see (14)})$$

\[ \text{(39)} \]
That is, a solution \( \hat{x} \) of Program (5) with \( \delta' \) specified above can be considered as an estimate of \( x_{K,\varepsilon} \). We also constructed the necessary dual certificates in the previous section by showing the existence of \( Q \) and \( Q^0 \) under the conditions specified in Propositions 12 and 13. Consequently, Lemmas 9, 10, and 11 hold. The following argument completes the proof of Theorem 5:

\[
d_{GW}(x, \hat{x}) \\
\leq d_{GW}(x, x_{K,\varepsilon}) + d_{GW}(x_{K,\varepsilon}, \hat{x}) \quad \text{(triangle inequality)} \\
\leq R(x, K, \varepsilon) + \left( \left( 6 + \frac{2}{g} \right) \|b\|_F + 6\|b^0\|_F \right) \delta' + \frac{\varepsilon}{2} \|x_{K,\varepsilon}\|_{TV} \quad \text{(see (14), (35))} \\
= R(x, K, \varepsilon) + \left( \left( 6 + \frac{2}{g} \right) \|b\|_F + 6\|b^0\|_F \right) (\delta + L \cdot R(x, K, \varepsilon)) + \frac{\varepsilon}{2} \|x_{K,\varepsilon}\|_{TV} \quad \text{(see (39))} \\
= \left( \left( 6 + \frac{2}{g} \right) \|b\|_F + 6\|b^0\|_F \right) \delta + \left( \left( 6 + \frac{2}{g} \right) L\|b\|_F + 6L\|b^0\|_F + 1 \right) R(x, K, \varepsilon) + \frac{\varepsilon}{2} \|x_{K,\varepsilon}\|_{TV} \\
\quad + \frac{\varepsilon}{2} \|x_{K,\varepsilon}\|_{TV}. \quad \text{(see (37))}
\]

Finally, the constants in Theorem 5 are given explicitly by:

\[
c_1 = 10\|b\|_F + 6\|b^0\|_F, \\
c_2 = \frac{\|x_{K,\varepsilon}\|_{TV}}{2}, \\
c_3 = 10L\|b\|_F + 6L\|b^0\|_F + 1.
\]

5 Perspective

In this paper, we have shown that a simple convex feasibility program is guaranteed to robustly recover a sparse (nonnegative) image in the presence of model mismatch and additive noise, under certain conditions on the imaging apparatus. No sparsity-promoting regularizer or separation condition is needed, and the techniques used here are arguably simple and intuitive. In other words, we have described when the imaging apparatus acts as an injective map on all sparse images and when we can stably find its inverse. These results build upon and extend a recent manuscript [1] which focuses on 1-D signals. The extension to images, however, requires novel constructions of interpolating polynomials, called dual certificates. In practice, many super-resolution problems appear in even higher dimensions. While we believe that similar results hold in any dimension, it is yet to be proven. Similarly, the super-resolution problem was studied in different non-Euclidean geometries for complex measures and under a separation condition [40, 64, 38, 39, 41]. It would be interesting to examine whether our results, which are based on the properties of Chebyshev systems, extend to these non-trivial geometries and to manifolds in general.

Verifying the conditions on the window for stable recovery in Theorem 5 is rather cumbersome. As an example, we have shown that the Gaussian window, a ubiquitous model of convolution kernels, satisfies those conditions. It is important to identify other such admissible windows and, if possible, simplify the conditions on the window in Theorem 5. Another interesting research direction is deriving the optimal separation \( \varepsilon \) (as a function of noise level \( \delta \)) that minimizes the right-hand side of the error bound in (16). Such a result will provide the tightest error bound for Program (5).

This work has focused solely on the theoretical performance of Program (5). It is essentially important to understand, even numerically, the pros and cons of the different localization algorithms suggested in the literature. For instance, it would be interesting to investigate whether the sparse-promoting regularizer, albeit not necessary for our analysis of nonnegative measures, reduces the recovery error.
Acknowledgements

AE is supported by the Alan Turing Institute under the EPSRC grant EP/N510129/1 and partially by the Turing Seed Funding grant SF019. GT is supported by the NSF grant CCF-1704204 and the DARPA Lagrange Program under ONR/SPAWAR contract N660011824020. AE would like to thank Jared Tanner and Bogdan Toader for their insightful feedback. TB would like to thank Amit Singer for his support and Nir Sharon for helpful discussions on Chebyshev systems.

References

[1] Armin Eftekhari, Jared Tanner, Andrew Thompson, Bogdan Toader, and Hemant Tyagi. Sparse non-negative super-resolution-simplified and stabilised. arXiv preprint arXiv:1804.01490, 2018.

[2] Geoffrey Schiebinger, Elina Robeva, and Benjamin Recht. Superresolution without separation. Information and Inference: A Journal of the IMA, 7(1):1–30, 2017.

[3] Klaus G Puschmann and Franz Kneer. On super-resolution in astronomical imaging. Astronomy & Astrophysics, 436(1):373–378, 2005.

[4] Valery Khaidukov, Evgeny Landa, and Tijmen Jan Moser. Diffraction imaging by focusing-defocusing: An outlook on seismic superresolution. Geophysics, 69(6):1478–1490, 2004.

[5] Eric Betzig, George H Patterson, Rachid Sougrat, O Wolf Lindwasser, Scott Olenych, Juan S Bonifacino, Michael W Davidson, Jennifer Lippincott-Schwartz, and Harald F Hess. Imaging intracellular fluorescent proteins at nanometer resolution. Science, 313(5793):1642–1645, 2006.

[6] Samuel T Hess, Thanu PK Girirajan, and Michael D Mason. Ultra-high resolution imaging by fluorescence photoactivation localization microscopy. Biophysical journal, 91(11):4258–4272, 2006.

[7] Michael J Rust, Mark Bates, and Xiaowei Zhuang. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (storm). Nature methods, 3(10):793–796, 2006.

[8] C Ekanadham, D Tranchina, and Eero P Simoncelli. Neural spike identification with continuous basis pursuit. Computational and Systems Neuroscience (CoSyNe), Salt Lake City, Utah, 2011.

[9] Stefan Hell. Primer: fluorescence imaging under the diffraction limit. Nature methods, 6(1):19, 2009.

[10] Ronen Tur, Yonina C Eldar, and Zvi Friedman. Innovation rate sampling of pulse streams with application to ultrasound imaging. IEEE Transactions on Signal Processing, 59(4):1827–1842, 2011.

[11] Oren Solomon, Yonina C Eldar, Maor Mutzafi, and Mordechai Segev. SPARCOM: Sparsity based super-resolution correlation microscopy. arXiv preprint arXiv:1707.09255, 2017.

[12] E.J. Candès and C. Fernandez-Granda. Towards a mathematical theory of super-resolution. Communications on Pure and Applied Mathematics, 67(6):906–956, 2014.

[13] Tamir Bendory. Robust recovery of positive stream of pulses. IEEE Transactions on Signal Processing, 65(8):2114–2122, 2017.

[14] Veniamin I. Morgenshtern and Emmanuel J. Candès. Super-resolution of positive sources: The discrete setup. SIAM Journal on Imaging Sciences, 9(1):412–444, 2016.

[15] Martin Slawski, Matthias Hein, et al. Non-negative least squares for high-dimensional linear models: Consistency and sparse recovery without regularization. Electronic Journal of Statistics, 7:3004–3056, 2013.

[16] Simon Foucart and David Koslicki. Sparse recovery by means of nonnegative least squares. 2014.

[17] Y. De Castro and F. Gamboa. Exact reconstruction using beurling minimal extrapolation. Journal of Mathematical Analysis and applications, 395(1):336–354, 2012.
(18) Quentin Denoyelle, Vincent Duval, and Gabriel Peyré. Support recovery for sparse super-resolution of positive measures. *Journal of Fourier Analysis and Applications*, 23(5):1153–1194, 2017.

(19) A. Eftekhari and A. Thompson. A bridge between past and present: Exchange and conditional gradient methods are equivalent. *arXiv preprint arXiv:1804.10243*, 2018.

(20) Nicholas Boyd, Geoffrey Schiebinger, and Benjamin Recht. The alternating descent conditional gradient method for sparse inverse problems. *SIAM Journal on Optimization*, 27(2):616–639, 2017.

(21) Kristian Bredies and Hanna Katriina Pikkarainen. Inverse problems in spaces of measures. *ESAIM: Control, Optimisation and Calculus of Variations*, 19(1):190–218, 2013.

(22) S. Karlin and W.J. Studden. *Tchebycheff systems: with applications in analysis and statistics*. Pure and applied mathematics. Interscience Publishers, 1966.

(23) S. Karlin. *Total Positivity*. Number v. 1 in Total Positivity. Stanford University Press, 1968.

(24) M. G. Krein, A. A. Nudelman, and D. Louvish. *The Markov Moment Problem And Extremal Problems*.

(25) E.J. Candès and C. Fernandez-Granda. Super-resolution from noisy data. *Journal of Fourier Analysis and Applications*, 19(6):1229–1254, 2013.

(26) Tamir Bendory, Shai Dekel, and Arie Feuer. Robust recovery of stream of pulses using convex optimization. *Journal of Mathematical Analysis and Applications*, 442(2):511–536, 2016.

(27) Gongguo Tang, Badri Narayan Bhaskar, Parikshit Shah, and Benjamin Recht. Compressed sensing off the grid. *IEEE transactions on information theory*, 59(11):7465–7490, 2013.

(28) Benedetto Piccoli and Francesco Rossi. Generalized Wasserstein distance and its application to transport equations with source. *Archive for Rational Mechanics and Analysis*, 211(1):335–358, 2014.

(29) T.H. Cormen, C.E. Leiserson, R.L. Rivest, and C. Stein. *Introduction to Algorithms*. Computer science. MIT Press, 2009.

(30) Quentin Denoyelle, Vincent Duval, and Gabriel Peyré. Support recovery for sparse super-resolution of positive measures. *Journal of Fourier Analysis and Applications*, 23(5):1153–1194, 2017.

(31) C. Fernandez-Granda. Support detection in super-resolution. *arXiv preprint arXiv:1302.3921*, 2013.

(32) Qiuwei Li and Gongguo Tang. Approximate support recovery of atomic line spectral estimation: A tale of resolution and precision. In *Signal and Information Processing (GlobalSIP), 2016 IEEE Global Conference on*, pages 153–156. IEEE, 2016.

(33) Vincent Duval and Gabriel Peyré. Exact support recovery for sparse spikes deconvolution. *Foundations of Computational Mathematics*, 15(5):1315–1355, 2015.

(34) J.M. Azais, Y. De Castro, and F. Gamboa. Spike detection from inaccurate samplings. *Applied and Computational Harmonic Analysis*, 38(2):177–195, 2015.

(35) Tamir Bendory, Avinoam David Bar-Zion, Dan Adam, Shai Dekel, and Arie Feuer. Stable support recovery of stream of pulses with application to ultrasound imaging. *IEEE Trans. Signal Processing*, 64(14):3750–3759, 2016.

(36) Badri Narayan Bhaskar, Gongguo Tang, and Benjamin Recht. Atomic norm denoising with applications to line spectral estimation. *IEEE Transactions on Signal Processing*, 61(23):5987–5999, 2013.

(37) G. Tang, B.N. Bhaskar, and B. Recht. Near minimax line spectral estimation. *IEEE Transactions on Information Theory*, 61(1):499–512, 2015.

(38) Tamir Bendory, Shai Dekel, and Arie Feuer. Exact recovery of Dirac ensembles from the projection onto spaces of spherical harmonics. *Constructive Approximation*, 42(2):183–207, 2015.
[39] Tamir Bendory, Shai Dekel, and Arie Feuer. Super-resolution on the sphere using convex optimization. *IEEE Transactions on Signal Processing*, 63(9):2253–2262, 2015.

[40] Tamir Bendory, Shai Dekel, and Arie Feuer. Exact recovery of non-uniform splines from the projection onto spaces of algebraic polynomials. *Journal of Approximation Theory*, 182:7–17, 2014.

[41] Frank Filbir and Kristof Schröder. Exact recovery of discrete measures from Wigner d-moments. *arXiv preprint arXiv:1606.05306*, 2016.

[42] Charles Dossal, Vincent Duval, and Clarice Poon. Sampling the fourier transform along radial lines. *SIAM Journal on Numerical Analysis*, 55(6):2540–2564, 2017.

[43] Kumar Vijay Mishra, Myung Cho, Anton Kruger, and Weiyu Xu. Spectral super-resolution with prior knowledge. *IEEE transactions on signal processing*, 63(20):5342–5357, 2015.

[44] Yohann De Castro, Fabrice Gamboa, Didier Henrion, and J-B Lasserre. Exact solutions to super resolution on semi-algebraic domains in higher dimensions. *IEEE Transactions on Information Theory*, 63(1):621–630, 2017.

[45] Weiyu Xu, Jian-Feng Cai, Kumar Vijay Mishra, Myung Cho, and Anton Kruger. Precise semidefinite programming formulation of atomic norm minimization for recovering d-dimensional (d ≥ 2) off-the-grid frequencies. In *Information Theory and Applications Workshop (ITA)*, 2014, pages 1–4. IEEE, 2014.

[46] Tamir Bendory and Yonina C Eldar. Recovery of sparse positive signals on the sphere from low resolution measurements. *IEEE Signal Processing Letters*, 22(12):2383–2386, 2015.

[47] Clarice Poon and Gabriel Peyré. Multi-dimensional sparse super-resolution. *arXiv preprint arXiv:1709.03157*, 2017.

[48] Petre Stoica, Randolph L Moses, et al. *Spectral analysis of signals*, volume 1. Pearson Prentice Hall Upper Saddle River, NJ, 2005.

[49] Ralph Schmidt. Multiple emitter location and signal parameter estimation. *IEEE transactions on antennas and propagation*, 34(3):276–280, 1986.

[50] Yingbo Hua and Tapan K Sarkar. Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 38(5):814–824, 1990.

[51] Richard Roy and Thomas Kailath. ESPRIT-estimation of signal parameters via rotational invariance techniques. *IEEE Transactions on acoustics, speech, and signal processing*, 37(7):984–995, 1989.

[52] Gongguo Tang. Resolution limits for atomic decompositions via markov-bernstein type inequalities. In *Sampling Theory and Applications (SampTA), 2015 International Conference on*, pages 548–552. IEEE, 2015.

[53] Wenjing Liao and Albert Fannjiang. Music for single-snapshot spectral estimation: Stability and super-resolution. *Applied and Computational Harmonic Analysis*, 40(1):33–67, 2016.

[54] Albert Fannjiang. Compressive spectral estimation with single-snapshot ESPRIT: Stability and resolution. *arXiv preprint arXiv:1607.01827*, 2016.

[55] Ankur Moitra. Super-resolution, extremal functions and the condition number of Vandermonde matrices. In *Proceedings of the forty-seventh annual ACM symposium on Theory of computing*, pages 821–830. ACM, 2015.

[56] Wenjing Liao. Music for multidimensional spectral estimation: stability and super-resolution. *IEEE Transactions on Signal Processing*, 63(23):6395–6406, 2015.

[57] Armin Eftekhari and Michael B Wakin. Greed is super: A fast algorithm for super-resolution. *arXiv preprint arXiv:1511.03385*, 2015.
the choice of nonnegative. We next verify that where the sum is over all subsets of \( \Omega \) in which case \( \Omega \geq T \). Let also \( I = S \) and vanish only on \( I \). Then, there exist t-s polynomials \( q_{T_1} \) and \( q_{S_{[K]\setminus\Omega}} \) that are nonnegative on \( I \) and vanish only on \( T_1 \) and \( S_{[K]\setminus\Omega} \), respectively.

Let us form the polynomial
\[
Q(\theta) = Q(t, s) = \sum_{\Omega \subseteq [K]} q_{T_1}(t) \cdot q_{S_{[K]\setminus\Omega}}(s),
\]
where the sum is over all subsets of \([K]\). Evidently, \( Q \) is nonnegative on \( \mathbb{R}^2 \) since each summand above is nonnegative. We next verify that \( Q \) only vanishes on \( \Theta \). To that end, consider \( \theta_k = (t_k, s_k) \in \Theta \) with \( k \in [K] \) and an index set \( \Omega \subseteq [K] \). There are two possibilities. Either \( k \in \Omega \), in which case \( q_{T_1}(t_k) = 0 \). Or \( k \in [K]\setminus\Omega \), in which case \( q_{S_{[K]\setminus\Omega}}(s_k) = 0 \). In both cases, the product vanishes, namely \( q_{T_1}(t_k) \cdot q_{S_{[K]\setminus\Omega}}(s_k) = 0 \). Since the choice of \( \Omega \) was arbitrary, it follows from (41) that \( Q(\theta_k) = Q(t_k, s_k) = 0 \) for every \( k \in [K] \).
On the other hand, suppose that \( \theta \in \Theta^c \). The first possibility is that \( \theta = (t, s) \in T^c \times S^c \subseteq \Theta^c \), namely, \( t \in T^c \) and \( s \in S^c \). For arbitrary index set \( \Omega \subseteq [K] \), note that \( q_{T_\Omega}(t) \cdot q_{S_{\Omega}}(s) > 0 \) by design. It follows from (41) that \( Q(\theta) = Q(t, s) > 0 \) when \( \theta \in T^c \times S^c \). The second possibility is that \( \theta = (t_k, s_l) \) with \( k \neq l \) and \( k, l \in [K] \). There always exists \( \Omega_0 \subseteq [K] \) such that \( t_k \in [K] \setminus \Omega_0 \) and \( s_l \in \Omega_0 \). For such \( \Omega_0 \), it holds that \( q_{T_\Omega}(t_k) \cdot q_{S_{\Omega}}(s_l) > 0 \). For instance, one can choose \( \Omega_0 = \{ s_l \} \) for which both \( q_{T_{\{s_l\}}}(t_k) \) and \( q_{S_{\Omega_0}}(s_l) \) are strictly positive. Consequently, \( Q(\theta) > 0 \) by (41) when \( \theta \in \Theta^c \setminus (T^c \times S^c) \). In conclusion, \( Q \) is positive and vanishes only on \( \Theta \), as claimed. This completes the proof of 8.

### B Proof of Proposition 12

The high-level strategy of the proof is again to construct the desired polynomial \( Q(\theta) \) in \( \theta = (t, s) \) by combining a number of univariate polynomials in \( t \) and \( s \). Each of these univariate polynomials is built using a 1-D version of Proposition 12, which is summarized below for the convenience of the reader, see [1, Proposition 11].

**Proposition 15.** Consider a finite set \( T' \subset \mathbb{I} \) of size \( K' \). For \( M \geq 2K' + 2 \), suppose that \( \{ \phi_m \}_{m=1}^M \) form a \( T \)-system on \( \mathbb{I} \). Consider also \( F' : \mathbb{R} \to \mathbb{R} \) and suppose that \( \{ F' \} \cup \{ \phi_m \}_{m=1}^M \) form an \( \mathbb{I} \)-system on \( \mathbb{I} \). Then, there exist real coefficients \( \{ b_m \}_{m=1}^M \) and a polynomial \( q_{T'} = \sum_{m=1}^M b_m \phi_m \) such that \( q_{T'} \geq F' \) with the equality holding on \( T' \).

Let us now use Proposition 15 to complete the proof of Proposition 12. Fix an index set \( \Omega \subseteq [K] \). By assumption, \( \{ F_{T_\Omega} \} \cup \{ \phi_m \}_{m=1}^M \) form an \( \mathbb{T}_{K,\varepsilon} \)-system on \( \mathbb{I} \) with \( M \geq 2K + 2 \). Therefore, by Proposition 15, there exists a polynomial \( q_{T_\Omega} \) such that

\[
q_{T_\Omega} \geq F_{\Omega},
\]

with the equality holding on \( T \). Likewise, by assumption, \( \{ F_{S_{\Omega}} \} \cup \{ \phi_m \}_{m=1}^M \) form an \( \mathbb{T}_{K,\varepsilon} \)-system on \( \mathbb{I} \) and therefore there exists a polynomial \( q_{S_{\Omega}} \) such that

\[
q_{S_{\Omega}} \geq F_{S_{\Omega}},
\]

with the equality holding on \( S \).

As in the proof of Proposition 8, consider the polynomial

\[
Q(\theta) = Q(t, s) = \sum_{\Omega \subseteq [K]} q_{T_\Omega}(t) \cdot q_{S_{\Omega}}(s),
\]

where the sum is over all subsets of \( [K] \). We next show that \( Q \) is the desired dual certificate. Recall the neighbourhoods defined in (23) and (25). Fix \( k \in [K] \) and assume that \( k \in \Omega \) for now. For \( \theta \in \theta_{k,\varepsilon} = t_k \times s_k, \) we then have that

\[
q_{T_\Omega}(t) \cdot q_{S_{\Omega}}(s) + q_{T_{\Omega_k}}(t) \cdot q_{S_{\Omega_k}}(s) \geq F_{T_\Omega}(t) \cdot F_{S_{\Omega}}(s) + F_{T_{\Omega_k}}(t) \cdot F_{S_{\Omega_k}}(s) \quad \text{(see (42), (43))}
\]

\[
\geq 0 \cdot 1 + 1 \cdot 0 = 0,
\]

and the equality in the first line above holds at \( \theta_k = (t_k, s_k) \). Note that the second inequality above holds also when \( k \in [K] \setminus \Omega \). By summing up over all pairs \((\Omega, [K] \setminus \Omega)\) and applying the above inequality, we find that

\[
Q(\theta) = \sum_{\Omega \subseteq [K]} q_{T_\Omega}(t) \cdot q_{S_{\Omega}}(s) \geq 2^{K-1} \cdot 0 = 0,
\]

which holds for \( \theta \in \theta_{k,\varepsilon} \) and with equality at \( \theta_k \). The above bound is independent of \( k \) and we therefore conclude that

\[
Q(\theta) \geq 0, \quad \theta \in \theta_{\varepsilon},
\]

with equality holding at \( \Theta \), see (25).
On the other hand, consider \( \theta \in \Theta^C \) that does not belong to \( \theta_{k,\varepsilon} \) for any \( k \in [K] \). There are two cases here. The first possibility is that
\[
\theta = (t, s) \in T^C \times S^C \subseteq \Theta^C.
\]
In this case, using (42) and (43), we have that
\[
q_{T \setminus I}(t) \cdot q_{S \setminus I}(s) \geq F_{T \setminus I}(t) \cdot F_{S \setminus I}(s) \geq 1,
\]
for every index set \( \Omega \subseteq [K] \). Therefore,
\[
Q(\theta) = \sum_{\Omega \subseteq [K]} q_{T \setminus I}(t) \cdot q_{S \setminus I}(s) \geq 2^K, \quad \theta \in T^C \times S^C.
\]
(48)
The second possibility is that \( \theta \in \Theta^C \setminus (T^C \times S^C) \). In this case, there exist \( k \neq l \) with \( k, l \in [K] \) such that \( \theta = (t, s) \in t_{k,\varepsilon} \times s_{l,\varepsilon} \) and an index set \( \Omega_0 \subseteq [K] \) such that \( t_k \in [K] \setminus \Omega_0 \) and \( s_l \in \Omega_0 \). It follows that
\[
q_{T \setminus I}(t) \cdot q_{S \setminus I}(s) \geq F_{T \setminus I}(t) \cdot F_{S \setminus I}(s) \geq 1.
\]
(49)
There are in fact \( 2^{K-2} \) such subsets of \( [K] \) and we conclude that
\[
Q(\theta) \geq 2^{K-2} = \exists \bar{g}, \quad \theta \in \Theta^C.
\]
(50)
By combining (48) and (50), we find that
\[
Q(\theta) \geq 2^{K-2} = \exists \bar{g}, \quad \theta \in \Theta^C.
\]
(51)
Combining (46) and (51) completes the proof of the proposition.

C Proof of Proposition 13

The proof is based on the same principles as the proof in Appendix B. Let us fix an arbitrary sign pattern \( \pi \in \{\pm 1\}^K \). For every \( k \in [K] \), by assumption, \( \{F^{k,\pi_k}_T\} \cup \{\phi_m\}_{m=1}^M \) form a \( T^*_k \)-system on \( \mathbb{I} \), see (38). Therefore, by Proposition 15, for every \( k \in [K] \) there exists a polynomial \( q^k_T \) satisfying
\[
q^k_T(t) \geq F^{k,\pi_k}_T(t), \quad t \in \mathbb{I},
\]
(52)
with the equality holding on \( T \). Likewise, for every \( k \in [K] \) and by assumption, \( \{F^{k,\pi^+}_S\} \cup \{\phi_m\}_{m=1}^M \) form a \( T^*_k \)-system on \( \mathbb{I} \). Therefore, there exists for every \( k \in [K] \) a polynomial \( q^k_S \) such that
\[
q^k_S(s) \geq F^{k}_S(s), \quad s \in \mathbb{I},
\]
(53)
with the equality holding on \( S \). Let us consider the polynomial
\[
Q^\pi(\theta) = Q^\pi(t, s) := \sum_{k=1}^K q^k_T(t) \cdot q^k_S(s).
\]
Recalling (38) and using (52) and (53) we observe that
\[
Q^\pi(\theta) \geq \sum_{k=1}^K F^{k,\pi_k}_T(t) \cdot F^{k}_S(s) \geq \begin{cases} \pi_k & \text{when } \theta \in \theta_{k,\varepsilon} \\ 0 & \text{elsewhere on } \mathbb{I}^2, \end{cases}
\]
(54)
and the equality holds on $\Theta$ (and in fact on its superset $T \times S$). Let $\pi^0$ be the sign pattern of $\{ \int_{\Theta_k} h(d\theta) \}_{k=1}^K$. Then, (54) implies that $Q^0 := Q_{\pi^0} \geq G^0$, see (33). This completes the proof of Proposition 13.