

Stable super-resolution of images: theoretical study

ARMIN EFTEKHARI[†]

Department of Mathematics and Mathematical Statistics, Umeå University, Umeå, Sweden

TAMIR BENDORY

Program in Applied and Computational Mathematics, Princeton University, Princeton, US

AND

GONGGUO TANG

Department of Electrical Engineering, Colorado School of Mines, Golden, Colorado, US

[†]Corresponding author. Email: armin.eftekhari@gmail.com

[Received on 12 October 2018; revised on 5 September 2020; accepted on 8 September 2020]

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 non-negative 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, i.e., a system of continuous functions that loosely behave like algebraic polynomials. Building upon the recent results for one-dimensional signals, we prove that this super-resolution algorithm is stable, in the generalized Wasserstein metric, to model mismatch (i.e., 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 Schiebinger *et al.* (2017, *Inf. Inference*, **7**, 1–30) 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 any imaging device, poorly separated sources are indistinguishable without using an appropriate localization technique that would take into account the sparse structure within the image.

This super-resolution problem of localizing point sources finds various applications in, for instance, astronomy [3], geophysics [4], chemistry, medicine, microscopy and neuroscience [5–11]. In this paper, we study the grid-free and non-negative super-resolution of two-dimensional (2-D) signals (i.e., images) in the presence of noise, extending the one-dimensional (1-D) results of [1].

Let x be a non-negative Borel measure supported on $\mathbb{T}^2 = [0, 1] \times [0, 1]$, and let $\{\phi_m\}_{m=1}^M$ be real-valued and continuous functions. We model the (possibly noisy) observations $\{y_{m,n}\}_{m,n=1}^M$ collected from

x as

$$y_{m,n} \approx \int_{\mathbb{I}^2} \phi_m(t) \phi_n(s) x(\mathrm{d}t, \mathrm{d}s). \quad (1.1)$$

More specifically, we assume that

$$\sum_{m,n=1}^M \left| y_{m,n} - \int_{\mathbb{I}^2} \phi_m(t) \phi_n(s) x(\mathrm{d}t, \mathrm{d}s) \right|^2 \leq \delta^2, \quad (1.2)$$

where $\delta \geq 0$ reflects the additive noise level. We do not impose a statistical model for 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

$$y[m, n] = y_{m,n}, \quad \Phi(t, s)[m, n] = \phi_m(t) \phi_n(s), \quad \forall m, n \in [M] := \{1, \dots, M\}, \quad (1.3)$$

we may rewrite (1.2) more compactly as

$$\left\| y - \int_{\mathbb{I}^2} \Phi(t, s) x(\mathrm{d}t, \mathrm{d}s) \right\|_{\mathrm{F}} \leq \delta, \quad (1.4)$$

where $\|\cdot\|_{\mathrm{F}}$ stands for the Frobenius norm. Often, ϕ_m and ϕ_n above are translated copies of a function ϕ , $\phi(t)\phi(s)$ 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 (1.4) is widely used as a model in imaging [12–14]. 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 $\{\phi_m\}_{m=1}^M = \{\cos(2\pi kt)\}_{k=0}^{f_c} \cup \{\sin(2\pi kt)\}_{k=1}^{f_c}$ with $M = 2f_c + 1$. It is also possible to collect observations using two different set of functions along t and s directions ($\{\phi_m(t)\}_m$ and $\{\psi_n(s)\}_n$). However, for the sake of clarity, we avoid this additional layer of complexity here.

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

$$\text{find a non-negative Borel measure } z \text{ on } \mathbb{I}^2 \text{ such that } \left\| y - \int_{\mathbb{I}^2} \Phi(t, s) z(\mathrm{d}t, \mathrm{d}s) \right\|_{\mathrm{F}} \leq \delta', \quad (1.5)$$

for some $\delta' \geq \delta$, which is reminiscent of nonnegative least squares in finite dimensions [15,16]. Once Program (1.5) is solved, the zeros of the optimal dual function can be used as estimates for the locations of the point sources. Alternatively, one may apply the Prony's method [17] or the matrix pencil approach [18] to a solution of Program (1.5), which is a measure on \mathbb{I}^2 , to locate the point sources.

Program (1.5) does not involve a grid on \mathbb{I}^2 and notably does not regularize z beyond non-negativity, thus radically deviating from the existing literature [2,13,14,19,20]. This paper establishes that in the noiseless setting $\delta = 0$, solving Program (1.5) precisely recovers the true measure x , provided that x is a non-negative sparse measure on \mathbb{I}^2 and under certain conditions on the imaging apparatus Φ . In addition, when $\delta > 0$ and x is an arbitrary nonnegative measure on \mathbb{I}^2 , solving Program (1.5) well approximates x . In particular, we establish that *any* non-negative measure supported on \mathbb{I}^2 that agrees with the observations y in the sense of Program (1.5) is near the true measure x .

This paper does not focus on the important question of how to numerically solve the infinite-dimensional Program (1.5) in practice. One straightforward approach would be to discretize the measure z on a fine uniform grid for \mathbb{I}^2 , thereby replacing Program (1.5) with a finite-dimensional convex feasibility program that can be solved with standard convex solvers. Moreover, a few recent papers

have proposed algorithms to directly solve Program (1.5) [21–23], i.e., these algorithms can be used to solve Program (1.5) without discretization. A comprehensive numerical comparison between these alternatives is of great importance and we leave that to a future study. This paper instead aims to provide theoretical justifications for the success of Program (1.5), thereby arguing that imposing non-negativity is theoretically enough for successful super-resolution. In other words, under mild conditions, the imaging device acts as an injective map on sparse non-negative measures and we can stably find its inverse map.

This work relies heavily on a recent work [1], which established that grid-free and non-negative super-resolution in 1-D can be achieved by solving the 1-D version of Program (1.5). In doing so, it removed the regularization required in prior work and substantially simplified the existing results. However, extending [1] to two dimensions is far from trivial and requires a careful design of a 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 contributions in detail, while proofs are deferred to Section 4 and the appendices.

Sparse measures without noise. Suppose that the measure x consists of K positive impulses located in \mathbb{I}^2 . In the absence of noise ($\delta = 0$), Proposition 2 below shows that solving Program (1.5) 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 \mathbb{I} . A Chebyshev system, or \mathcal{C} -system for short,¹ is a collection of continuous functions that loosely behave like algebraic monomials; see Definition 1. \mathcal{C} -system is a widely used concept in classical approximation theory [24–26] that also plays a pivotal role in some modern signal processing applications; see for instance [1, 2, 19]. In other words, Proposition 2 below establishes that the imaging operator Φ in (1.4) is an injective map from K -sparse non-negative measures on \mathbb{I}^2 , provided that $\{\phi_m\}_{m=1}^M$ form a \mathcal{C} -system on \mathbb{I} and $M \geq 2K + 1$.

In contrast to earlier results, no minimum separation between the impulses is necessary, Program (1.5) does not contain any explicit regularization to promote sparsity, and lastly $\{\phi_m\}_{m=1}^M$ need only to be continuous. We note that Proposition 2 is a non-trivial extension of the 1-D result in [1] to images. Indeed, the key concept of \mathcal{C} -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].

Arbitrary measure with noise. More generally, consider an arbitrary non-negative measure x supported on \mathbb{I}^2 . As detailed later, given $\varepsilon \in (0, 1/2]$, the measure x can always be approximated with a K -sparse and ε -separated non-negative measure, up to an error of $R(x, K, \varepsilon)$ in the generalized Wasserstein metric, denoted throughout by d_{GW} . This is true even if x itself is not ε -separated or not atomic at all. We may think of $R(x, K, \varepsilon)$ as the ‘model-mismatch’ of approximating x with a well-separated sparse measure, i.e., $R(x, K, \varepsilon) = d_{\text{GW}}(x, x_{K, \varepsilon}) = \min d_{\text{GW}}(x, \chi)$, where the minimum is taken over every non-negative, K -sparse and ε -separated measure χ .

In the presence of noise and numerical inaccuracies ($\delta \geq 0$), Theorem 11 below shows that solving Program (1.5) 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 (1.5) satisfies

$$d_{\text{GW}}(x, \hat{x}) \leq c_1 \delta + c_2(\varepsilon) + c_3 R(x, K, \varepsilon), \quad (1.6)$$

¹ It is also common to use T-system as the abbreviation of the Chebyshev system.

provided that $M \geq 2K + 2$, and the imaging apparatus and certain functions forms a \mathcal{C}^* -system, a natural generalization of the \mathcal{C} -system introduced earlier. The factors c_1, c_2, c_3 above are specified in the proof and depend chiefly on the measurement functions $\{\phi_m\}_{m=1}^M$; see (1.3). Note that the recovery error in (1.6) depends on the noise level δ , the separation ε and on how well x can be approximated with a K -sparse and ε -separated measure, similar to the 1-D results in [1]. In particular, as we will see later, when $\delta = \varepsilon = R(x, K, \varepsilon) = 0$, (1.6) reads as $d_{\text{GW}}(x, \hat{x}) = 0$, and Theorem 11 reduces to Proposition 2 for sparse and noise-free super-resolution.

We remark that Theorem 11 applies to any non-negative 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 $d_{\text{GW}}(x, \hat{x})$ does depend on how well x can be approximated with a well-separated sparse measure, which is reflected in the right-hand side of (1.6) and hidden in the factors c_1, c_2, c_3 therein. As emphasized earlier, no regularization other than non-negativity is used and $\{\phi_m\}_m$ need only be continuous.

As a concrete example of this general framework, we consider the case where $\{\phi_m\}_m$ are translated copies of a Gaussian ‘window’, i.e., copies of a Gaussian function. Building on the results from [1], we show in Section 2.3 that the conditions for both Proposition 2 and Theorem 11 are met for this important example. That is, solving Program (1.5) successfully and stably recovers an image that has undergone Gaussian blurring.

2. Main results

2.1 Sparse measure without noise

Let x be the non-negative atomic measure

$$x = \sum_{k=1}^K a_k \cdot \delta_{\theta_k}, \quad a_k > 0, \quad (2.1)$$

with K impulses located at $\Theta = \{\theta_k\}_{k=1}^K \subset \text{interior}(\mathbb{I}^2)$ and positive amplitudes $\{a_k\}_{k=1}^K$. Here, δ_{θ_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

$$y = \int_{\mathbb{I}^2} \Phi(\theta) x(d\theta) \in \mathbb{R}^{M \times M}. \quad (2.2)$$

To understand when solving Program (1.5) with $\delta' = 0$ successfully recovers the true measure x , recall the concept of \mathcal{C} -system [24].

DEFINITION 1 (\mathcal{C} -system). Real-valued and continuous functions $\{\phi_m\}_{m=1}^M$ form a \mathcal{C} -system on the interval \mathbb{I} , provided that the determinant of the $M \times M$ matrix $[\phi_m(\tau_k)]_{k,m=1}^M$ is positive for every (strictly) increasing sequence $\{\tau_k\}_{k=1}^M \subset \mathbb{I}$.

For example, the monomials $\{1, t, \dots, t^{M-1}\}$ form a \mathcal{C} -system on any closed interval of the real line. In fact, \mathcal{C} -system can be interpreted 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 \mathcal{C} -system $\{\phi_m\}_{m=1}^M$ has at most $M - 1$ distinct zeros on the interval \mathbb{I} . Or, given M distinct points, 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 \mathcal{C} -system. As an example in the context of super-resolution, translated copies of the Gaussian window e^{-t^2} form a \mathcal{C} -system on any interval of the real line, and so do many other windows [24]. As we will see later, the notion of \mathcal{C} -system allows us to design a non-negative polynomial with prescribed zeros on the interval \mathbb{I} , and this polynomial will play a key role in establishing the main results of this paper.

Proved in Section 4.2, the following result states that solving Program (1.5) successfully recovers x from the noise-free image y , provided that the measurement functions form a \mathcal{C} -system.

PROPOSITION 2 (Sparse measure without noise). Let x be a K -sparse non-negative measure supported on $\text{interior}(\mathbb{I}^2)$, specified by (2.1). Suppose that $M \geq 2K + 1$ and that the measurement functions $\{\phi_m\}_{m=1}^M$ form a \mathcal{C} -system on the interval \mathbb{I} . Lastly, for $\delta = 0$, consider the imaging operator Φ and the image $y \in \mathbb{R}^{M \times M}$ in (1.3) and (1.4). Then, x is the unique solution of Program (1.5) with $\delta' = 0$.

In words, Program (1.5) successfully localizes the K impulses present in the measure x from $(2K + 1)^2$ measurements, provided that the measurement functions $\{\phi_m\}_{m=1}^M$ form a \mathcal{C} -system on the interval \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 [13, 27, 28]. In addition, no regularization was imposed in Program (1.5) beyond non-negativity and the measurement functions only need to be continuous.

REMARK 3 (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

$$Q(\theta) = Q(t, s) = \sum_{m,n=1}^M b_{m,n} \phi_m(t) \phi_n(s) \quad (2.3)$$

would guarantee the success of Program (1.5) in the absence of noise. Known as the dual certificate for Program (1.5), this polynomial Q has to be non-negative 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 $S = \{s_k\}_{k=1}^K$ for short, the proof then constructs the polynomial Q by carefully combining non-negative 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 \mathcal{C} -system on the interval \mathbb{I} ; see Section 4.2 for the details. The basic idea of the proof is visualized in Fig. 1.

2.2 Arbitrary measure with noise

In this section, we present the main result of this paper. Theorem 11 below generalizes Proposition 2 to account for ① model mismatch, where x is not necessarily a well-separated sparse measure but might be close to one and ② imaging noise ($\delta \geq 0$). That is, Theorem 11 below addresses the stability of Program (1.5) to model mismatch and its robustness against imaging noise. Some preparation is necessary before presenting the result.

2.2.1 Separation. Unlike sparse and noise-free super-resolution in Proposition 2, a notion of separation will play a role in Theorem 11.

DEFINITION 4 (Separation). 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 $\text{sep}(x)$ denote the minimum separation between all impulses in x and the boundary of \mathbb{I}^2 . That is,

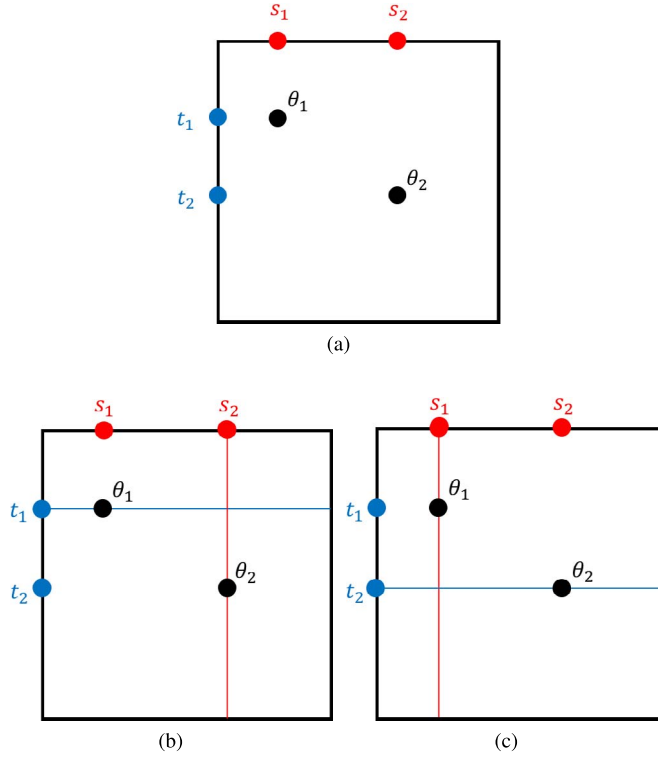


FIG. 1. This figure visualizes the idea behind the proof of Proposition 2, explained in Remark 3. As an example, consider the non-negative measure $x = a_1 \delta_{\theta_1} + a_2 \delta_{\theta_2}$. The locations of two impulses at $\theta_1 = (t_1, s_1) \in \mathbb{I}^2$ and $\theta_2 = (t_2, s_2) \in \mathbb{I}^2$ are shown with black dots in Fig. 1(a). For Program (1.5) to successfully recover x from the image y in (2.2), we should construct a non-negative polynomial Q of the form of (2.3) that has zeros exactly on θ_1 and θ_2 . We do so by combining a number of univariate polynomials in t and s . More specifically, consider a non-negative polynomial $q_{t_1}(t) = \sum_{m=1}^M b_m^{11} \phi_m(t)$ that vanishes only at t_1 . Likewise, consider similar non-negative polynomials $q_{t_2}(t)$, $q_{s_1}(s)$ and $q_{s_2}(s)$, which are zero only at t_2 , s_1 and s_2 , respectively. Figure 1(c) shows the zero set of the polynomial $q_{t_1}(t)q_{s_2}(s)$ as the union of blue and red lines. Similarly, Fig. 1(b) 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 non-negative polynomial of the form in (2.3) 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 measurement functions $\{\phi_m\}_{m=1}^M$ form a \mathcal{C} -system and $M \geq 2K + 1$, the existence of these univariate polynomials follows from standard results. We note that the construction of Q in this example is slightly different from the proof of Proposition 2, in order to simplify this presentation.

$\text{sep}(x)$ is the largest number ν such that

$$\begin{aligned} \nu &\leq |t_k - t_l|, & \nu &\leq |s_k - s_l|, & k \neq l, \, k, l \in [K], \\ \nu &\leq |t_k - 0|, & \nu &\leq |t_k - 1|, \\ \nu &\leq |s_k - 0|, & \nu &\leq |s_k - 1|. \end{aligned} \quad (2.4)$$

Naturally, if the measure x satisfies $\text{sep}(x) = \varepsilon$, we call x an ε -separated measure.

For example, for x in Fig. 1, we have $\text{sep}(x) = \min(t_2 - t_1, s_2 - s_1, t_1, 1 - t_2, s_1, 1 - s_2)$. We remark that our notion of separation is more restrictive than the one commonly used in the super-resolution literature [12,28,29], as it requires the point sources to be separated in both t and s directions by at least ν .

2.2.2 Generalized Wasserstein distance. As an error metric, we will use the generalized Wasserstein distance [30], which is closely related to the notion of unbalanced transport [31]. We first recall the total variation (TV) norm of a measure on \mathbb{I}^2 [32] is defined as $\|z\|_{\text{TV}} = \int_{\mathbb{I}^2} |z(dt)|$, akin to ℓ_1 -norm in finite dimensions. Recall also that the Wasserstein distance [32] for two non-negative measures z_1 and z_2 , supported on \mathbb{I}^2 , is defined as

$$d_W(z_1, z_2) = \inf \int_{\mathbb{I}^2 \times \mathbb{I}^2} \|\tau_1 - \tau_2\|_1 \cdot \gamma(d\tau_1, d\tau_2), \quad (2.5)$$

where the infimum is over every non-negative measure γ on $\mathbb{I}^2 \times \mathbb{I}^2$ that produces z_1 and z_2 as marginals, i.e.,

$$z_1(A_1) = \int_{A_1 \times \mathbb{I}^2} \gamma(d\tau_1, d\tau_2), \quad z_2(A_2) = \int_{\mathbb{I}^2 \times A_2} \gamma(d\tau_1, d\tau_2), \quad (2.6)$$

for all measurable sets $A_1, A_2 \subseteq \mathbb{I}^2$. If we were to think of z_1, z_2 as two piles of dirt, then $d_W(z_1, z_2)$ is the least amount of work needed to transform one pile to the other. The Wasserstein distance is defined only if the TV norms of the two measures are equal, i.e., $\|z_1\|_{\text{TV}} = \|z_2\|_{\text{TV}}$. The generalized Wasserstein distance extends d_W and allows for calculating the distance between non-negative measures with different TV norms.

DEFINITION 5 (Generalized Wasserstein distance). For two non-negative measures x_1 and x_2 supported on \mathbb{I}^2 , their generalized Wasserstein distance is defined as

$$d_{\text{GW}}(x_1, x_2) = \inf (\|x_1 - z_1\|_{\text{TV}} + d_W(z_1, z_2) + \|x_2 - z_2\|_{\text{TV}}), \quad (2.7)$$

where the infimum is over every pair of non-negative Borel measures z_1 and z_2 supported on \mathbb{I}^2 such that $\|z_1\|_{\text{TV}} = \|z_2\|_{\text{TV}}$.

Compared with (2.5), the two new terms in (2.7) gage the difference between the mass of x_1 and the mass of x_2 . Our choice of error metric d_{GW} is natural in the sense that any solution of Program (1.5) is itself a measure. However, note that controlling the error in the space of measures in our main result below does not immediately translate into controlling the error in the location of point sources, which may be estimated by, for example, applying the Prony's method [17] to a solution of Program (1.5).

For instance, when ε is infinitesimally small, the measure $\delta_{1/2} + \delta_{1/2+\varepsilon}$ has a small distance in the Wasserstein metric from the measure $2\delta_{1/2}$, but it is indeed impossible to distinguish the two impulses in the presence of noise in general, unless we impose additional structure on the noise. Nevertheless, in the limit of vanishing noise, it is possible to directly control the error in the location of point sources and we refer the reader to [33–35] for the details.

2.2.3 Model mismatch. Our main result, Theorem 11 below, bounds the recovery error $d_{\text{GW}}(x, \hat{x})$, where \hat{x} is a solution of Program (1.5). Note that, even though x is an arbitrary non-negative 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} . This sparse measure will play a key role in Theorem 11.

DEFINITION 6 (Residual). For a non-negative measure x supported on \mathbb{I}^2 , given an integer K and $\varepsilon \in (0, 1/2]$, there exists a K -sparse non-negative measure $x_{K,\varepsilon}$ that is ε -separated and well approximates x . More specifically, for any $\varepsilon \in (0, 1/2]$, there exists a K -sparse and ε -separated non-negative measure $x_{K,\varepsilon}$ such that

$$R(x, K, \varepsilon) := d_{\text{GW}}(x, x_{K,\varepsilon}) = \min d_{\text{GW}}(x, \chi), \quad (2.8)$$

where the minimum above is over every non-negative K -sparse and ε -separated measure χ supported on interior(\mathbb{I}^2).

In words, the residual $R(x, K, \varepsilon)$ can be thought of as the mismatch in modeling x with a well-separated sparse measure. Indeed, note that the minimum in (2.8) is achieved: we can limit the search in (2.8) to the (bounded) set of K -sparse and ε -separated measures with TV norm bounded by $\|x\|_{\text{TV}}$. This set is also closed, with respect to the weak topology imposed by d_{GW} , see [30, Theorem 13], and thus compact. Lastly, the objective function d_{GW} of (2.8) is a norm and thus continuous a fortiori, hence the claim.

2.2.4 Smoothness. For Program (1.5) to succeed in the general settings of this section, we also impose additional requirements on the imaging apparatus in the next two paragraphs. We assume in this section that the imaging apparatus is smooth in the following sense.

DEFINITION 7 (Smoothness). The imaging apparatus in (1.4) is L -Lipschitz-continuous if

$$\left\| \int_{\mathbb{I}^2} \Phi(\theta)(x_1(d\theta) - x_2(d\theta)) \right\|_{\text{F}} \leq L \cdot d_{\text{GW}}(x_1, x_2), \quad (2.9)$$

for every pair of measures x_1, x_2 supported on \mathbb{I}^2 .

It is often not difficult to verify the Lipschitz-continuity of Φ with respect to d_{GW} , as the following example demonstrates.

EXAMPLE 2.1 (Smoothness). As a toy example, suppose for simplicity that $\|x_1\|_{\text{TV}} = \|x_2\|_{\text{TV}}$, so that $d_{\text{GW}}(x_1, x_2) = d_{\text{W}}(x_1, x_2)$; see (2.7). Moreover, for clarity, let $m = 1$ and note that

$$\int \Phi(\theta)(x_1(d\theta) - x_2(d\theta)) = \int \phi(t)\phi(s)(x_1(dt, ds) - x_2(dt, ds)), \quad (2.10)$$

where ϕ is the measurement window. Let also L_ϕ denote the Lipschitz constant of $\phi(t)\phi(s)$ with respect to ℓ_1 -norm, i.e.,

$$|\phi(t_1)\phi(s_1) - \phi(t_2)\phi(s_2)| \leq L_\phi(|t_1 - t_2| + |s_1 - s_2|), \quad (2.11)$$

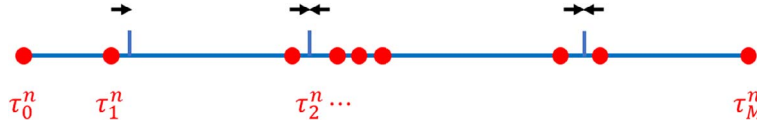


FIG. 2. This figure illustrates an example of an admissible sequence; see Definition 8. For a fixed n , the red dots form the increasing sequence $\{\tau_k^n\}_{k=0}^M$. Note that the endpoints of the interval are included in the sequence, i.e., $\tau_0^n = 0$ and $\tau_M^n = 1$. In view of Definition 8, as $n \rightarrow \infty$, the sequence $\{\tau_k^n\}_{k=1}^{M-1}$ converges to three distinct points on the interior of the interval, shown with blue bars. Moreover, all limit points in $(0, 1)$ have an even multiplicity except for one, which has a multiplicity of exactly one.

for every $t_1, t_2, s_1, s_2 \in \mathbb{I}$. Then, recalling the Kantorovich duality [32], we may write that

$$\begin{aligned} \left\| \int_{\mathbb{I}^2} \Phi(\theta) (x_1(d\theta) - x_2(d\theta)) \right\|_{\mathbb{F}} &= \left| \int_{\mathbb{I}} \phi(t) \phi(s) (x_1(dt, ds) - x_2(dt, ds)) \right| \quad (\text{see (2.10)}) \\ &= L_\phi \left| \int_{\mathbb{I}} \frac{\phi(t) \phi(s)}{L_\phi} (x_1(dt, ds) - x_2(dt, ds)) \right| \\ &\leq L_\phi \cdot \max_{\psi} \int \psi(\theta) x_1(d\theta) - \psi(\theta) x_2(d\theta) \\ &= L_\phi \cdot d_W(x_1, x_2) \quad (\text{Kantorovich duality}) \\ &= L_\phi \cdot d_{\text{GW}}(x_1, x_2), \quad (\text{simplifying assumption}) \end{aligned} \quad (2.12)$$

where the maximum in the third line above is over all 1-Lipschitz-continuous functions $\psi : \mathbb{I}^2 \rightarrow \mathbb{R}$ with respect to ℓ_1 -norm. We conclude that (2.9) holds with $L = L_\phi$ in this example.

2.2.5 \mathcal{C}^* -system. To study the stability of Program (1.5), we also need to modify the notion of \mathcal{C} -system in Definition 1. We begin with the definition of an admissible sequence, visualized in Fig. 2.

DEFINITION 8 (Admissible sequence). For a pair of integers K and M obeying $M \geq 2K + 1$, we say that $\{\{\tau_k^n\}_{k=0}^M\}_{n \geq 1} \subset \mathbb{I}$ is a (K, ε) -admissible sequence if

1. $\tau_0^n = 0$ and $\tau_M^n = 1$ for every n , i.e., the endpoints of $\mathbb{I} = [0, 1]$ are included in the increasing sequence $\{\tau_k^n\}_{k=0}^M$, for every n ;
2. as $n \rightarrow \infty$, the increasing sequence $\{\tau_k^n\}_{k=1}^{M-1}$ converges (element-wise) to an ε -separated finite subset of \mathbb{I} with at most K distinct points, where every element has an even multiplicity, except one element that appears only once.²

While \mathcal{C} -system in Definition 1 is a condition on all increasing sequences of length M , the \mathcal{C}^* -system below is a condition only on admissible sequences; these are the only sequences that matter in our analysis. Like a \mathcal{C} -system, a \mathcal{C}^* -system imposes certain requirements on a family of functions. Whereas the performance of Program (1.5) for sparse measures and in the absence of noise relates

² That is, every element is repeated an even number of times $(2, 4, \dots)$ except one element that appears only once.

to a certain \mathcal{C} -system in Proposition 2, the general performance of Program (1.5) relates to certain \mathcal{C}^* -systems, as we will see shortly in Theorem 11. The definition of \mathcal{C}^* -system below is immediately followed by its motivation.

DEFINITION 9 (\mathcal{C}^* -system). For an integer K and an even integer M obeying $M \geq 2K + 2$, real-valued functions $\{\phi_m\}_{m=0}^M$ form a $\mathcal{C}_{K,\varepsilon}^*$ -system on \mathbb{I} if every (K, ε) -admissible sequence $\{\{\tau_k^n\}_{k=0}^M\}_{n \geq 1}$ satisfies

1. the determinant of the $(M + 1) \times (M + 1)$ matrix $[\phi_m(\tau_k^n)]_{k,m=0}^M$ is positive for all sufficiently large n ;
2. moreover, all minors along the l th row of the matrix $[\phi_m(\tau_k^n)]_{k,m=0}^M$ approach zero at the same rate when $n \rightarrow \infty$. Here, l is the index of the element of the limit sequence that appears only once.³

REMARK 10 (Properties of \mathcal{C}^* -systems). ① Note that a $\mathcal{C}_{K,\varepsilon}^*$ -system on \mathbb{I} is also a $\mathcal{C}_{K',\varepsilon}^*$ -system for every integer $K' \leq K$. Indeed, this claim follows from the observation that every (K', ε) -admissible sequence is itself a (K, ε) -admissible sequence. ② Moreover, if $\{\phi_m\}_{m=0}^M$ form a $\mathcal{C}_{K,\varepsilon}^*$ -system on \mathbb{I} , then so do the scaled functions $\{c_m \phi_m\}_{m=0}^M$ for positive constants $\{c_m\}_{m=0}^M$.

Let us also offer some insight about \mathcal{C}^* -systems. In the proof of Proposition 2 for sparse and noise-free super-resolution, in order to construct a polynomial

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

with prescribed zeros on \mathbb{I} , we require that $\{\phi_m\}_{m=1}^M$ form a \mathcal{C} -system; see the discussion after Definition 1. On the other hand, for a given function ϕ_0 (which, in our context, will signify the stability to noise in super-resolution), in order to construct a polynomial

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

where equality holds at prescribed points in \mathbb{I} , it is natural to require $\{\phi_0\} \cup \{\phi_m\}_{m=1}^M$ to form a \mathcal{C} -system. The definition of \mathcal{C}^* -system above is based on the same idea but limited to admissible sequences, to ease the burden of verifying the conditions in Definition 9. In particular, Definition 9 will help exclude trivial polynomials, such as $0 \cdot \phi_0 + \sum_{m=1}^M b_m \phi_m$.

We remark that Definition 9 only considers admissible sequences to simplify the burden of verifying whether a family of functions form a \mathcal{C}^* -system.

To summarize, the widely used notion of \mathcal{C} -system in Definition 1 plays a key role in the analysis of sparse inverse problems in the absence of noise, whereas \mathcal{C}^* -system above was introduced in [1] and tailored for the stability analysis of sparse inverse problems.⁴ It was established in [1] that translated copies of the Gaussian window e^{-t^2} indeed form a \mathcal{C}^* -system, under mild conditions reviewed in

³ A non-negative sequence $\{u^n\}_{n \geq 1}$ approaches zero at the rate n^{-p} if $u_n = \Theta(n^{-p})$. See, for example, [36, p. 44].

⁴ Let us point out that we use the shorthand of \mathcal{C}^* -system here instead of \mathcal{T}^* -system used in [1].

Section 2.3 below. We suspect this to also hold for many other measurement windows with sufficiently fast decay.⁵

2.2.6 Main result. We are now ready to present the main result of this paper, which quantifies the performance of Program (1.5) in the general case where x is an arbitrary non-negative measure on \mathbb{I}^2 and in the presence of additive noise. Theorem 11, proved in Section 4.4, states that Program (1.5) approximately recovers x provided that certain \mathcal{C} - and \mathcal{C}^* -systems exist. As an example of this general result, Section 2.3 later specializes Theorem 11 for imaging under Gaussian blur.

THEOREM 11 (Arbitrary measure with noise). Consider a non-negative measure x supported on \mathbb{I}^2 . Consider also a noise level $\delta \geq 0$, the measurement functions $\{\phi_m\}_{m=1}^M$ and the image $y \in \mathbb{R}^{M \times M}$; see (1.3) and (1.4). We assume that the imaging apparatus is L -Lipschitz in the sense of (2.9).

For an integer K and $\varepsilon \in (0, 1/2]$, let $x_{K,\varepsilon}$ be a K -sparse and ε -separated non-negative measure on \mathbb{I}^2 that approximates x with the residual of $R(x, K, \varepsilon)$, in the sense of (2.8). In particular, let $\Theta = \{\theta_k\}_{k=1}^K = \{(t_k, s_k)\}_{k=1}^K \subset \text{interior}(\mathbb{I}^2)$ denote the support of $x_{K,\varepsilon}$, and set $T = \{t_k\}_{k=1}^K$ and $S = \{s_k\}_{k=1}^K$ for short.

With \hat{x} denoting a solution of Program (1.5) for $\delta' \geq (1 + L \cdot R(x, k, \varepsilon))\delta$, it holds that

$$d_{\text{GW}}(x, \hat{x}) \leq c_1 \delta + c_2(\varepsilon) + c_3 R(x, K, \varepsilon), \quad (2.13)$$

where d_{GW} is the generalized Wasserstein metric in (2.7). Above, $c_1, c_2(\varepsilon), c_3$ are specified explicitly in (4.23) and depend on the measure x , the separation ε and the measurement functions $\{\phi_m\}_{m=1}^M$. In particular, it holds that $c_2(0) = 0$.

The error bound in (2.13) holds if $M \geq 2K + 2$ and

1. $\{\phi_m\}_{m=1}^M$ form a \mathcal{C} -system on \mathbb{I} ,
2. $\{F_{T_\Omega}\} \cup \{\phi_m\}_{m=1}^M$ and $\{F_{S_\Omega}\} \cup \{\phi_m\}_{m=1}^M$ both form $\mathcal{C}_{K,\varepsilon}^*$ -systems on \mathbb{I} for every $\Omega \subseteq [K]$,
3. $\{F_{t_k}^+\} \cup \{\phi_m\}_{m=1}^M$ and $\{F_{t_k}^-\} \cup \{\phi_m\}_{m=1}^M$ both form $\mathcal{C}_{K,\varepsilon}^*$ -systems on \mathbb{I} for every $k \in [K]$,
4. $\{F_{s_k}^+\} \cup \{\phi_m\}_{m=1}^M$ form a $\mathcal{C}_{K,\varepsilon}^*$ -system on \mathbb{I} for every $k \in [K]$.

Above, for every index set $\Omega \subseteq [K]$ and $k \in [K]$, we define the functions $F_{T_\Omega}, F_{S_\Omega}, F_{t_k}^\pm, F_{s_k}^+ : \mathbb{I} \rightarrow \mathbb{R}$ as

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

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

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

⁵ The definition of \mathcal{C}^* -system here is slightly different from that in [1] but the difference is inconsequential.

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

An example of the functions in Theorem 11 appears in Fig. 3(e), where the purple graph is an example of F_{T_Ω} for $\Omega = \{t_1\}$, shown in the figure as F_{t_1} for brevity. Theorem 11 for image super-resolution is unique in a number of ways. The differences with prior work are further discussed in Section 3 and also summarized here. First, Theorem 11 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 [2,13,14,34].

Moreover, Theorem 11 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 11 is quantified with a natural metric between measures, i.e., the generalized Wasserstein metric, in contrast to prior work; see for example [37] that separately studies the error near and away from the impulses. Lastly, the measurement functions $\{\phi_m\}_m$ are required to be continuous rather than (several times) differentiable [2,34]. All this is achieved without the need to explicitly regularize for sparsity in Program (1.5).

Note also that, in practice, we often have an upper bound for the noise level δ and the model mismatch $R(x, K, \varepsilon)$, which would allow us to apply Theorem 11. This approach to quantifying stability against noise and model mismatch is common in model-based signal processing [38]. Several additional remarks are in order about Theorem 11.

REMARK 12 (Proof technique). For Program (1.5) to successfully recover a sparse measure in the absence of noise, we constructed a non-negative polynomial $Q(\theta)$, within the span of the measurement functions, which vanished 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, we need to construct a non-negative polynomial $Q(\theta)$ that is bounded away from zero far from the impulse locations Θ , i.e., $Q(\theta) \geq \bar{g} > 0$, for every θ far from Θ , where \bar{g} is a positive scalar. Letting $T = \{t_k\}_{k=1}^K$ and $S = \{s_k\}_{k=1}^K$ for short, the proof of Theorem 11 constructs Q by combining certain univariate polynomials, similar to the proof of Proposition 2 which was itself summarized earlier in Section 2.1, and illustrated in Fig. 1. Among these univariate polynomials, for example, the proof constructs a non-negative polynomial q_T such that $q_T(t) \geq 1$, for every t far from T . As shown in [1], such a univariate polynomial q_T exists if $\{F_T\} \cup \{\phi_m\}_{m=1}^M$ form a \mathcal{C}^* -system. In addition to Q , we also find it necessary to construct yet another non-negative polynomial Q^0 to control the recovery error near the impulse locations and thus complete the proof of Theorem 11, see Section 4.4 for more details. Figure 3 illustrates some of the key ideas in the proof of Theorem 11.

REMARK 13 (Recovery error). The bound on the recovery error $d_{\text{GW}}(x, \hat{x})$ in (2.13) depends on the noise level δ 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 ε -separated measure $x_{K,\varepsilon}$, with the residual of $R(x, K, \varepsilon)$; see (2.8). We might then think of a solution \hat{x} of Program (1.5) as an estimate for $x_{K,\varepsilon}$ and therefore an estimate for x , up to the residual $R(x, K, \varepsilon)$. Both the separation ε and the residual $R(x, K, \varepsilon)$ appear on the right-hand side of the error bound (2.13).

In particular, when $\delta = \varepsilon = R(x, K, \varepsilon) = 0$, we again obtain Proposition 2 for recovery of a K -sparse non-negative measure in the absence of noise. Note that, given a noise level δ , this work does not address the challenging problem of choosing the separation ε in order to minimize the right-hand side of (2.13). Intuitively, for large δ , we must choose the separation ε large enough to maintain stability against the

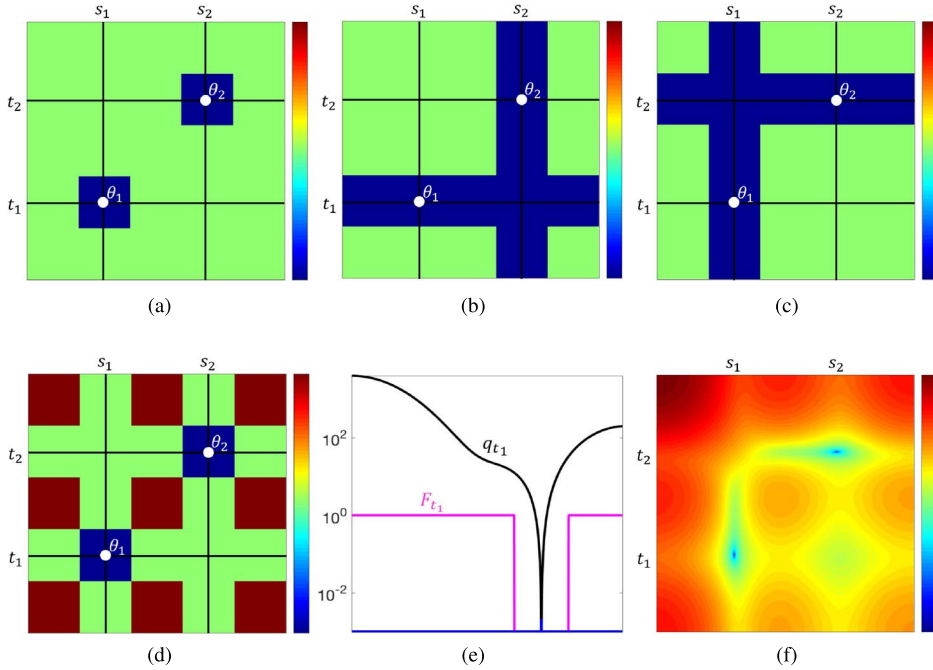


FIG. 3. This figure visualizes some of the principle ideas behind the proof of Theorem 11, explained in Remark 12. As an example, consider the non-negative measure $x = a_1 \delta_{\theta_1} + a_2 \delta_{\theta_2}$. The locations of two impulses at $\theta_1 = (t_1, s_1) \in \mathbb{I}^2$ and $\theta_2 = (t_2, s_2) \in \mathbb{I}^2$ are shown with white dots, and the black lines show the corresponding grid. For Program (1.5) to approximately recover x from the noisy image y given in (1.2), we need to construct a non-negative polynomial Q of the form in (2.3) that is zero at the impulse locations and large away from the impulses, to ensure stability. That is, we need $Q(\theta_1) = Q(\theta_2) = 0$ and also $Q(\theta) \geq \bar{g} > 0$ when θ is far from both θ_1, θ_2 , for a positive scalar \bar{g} . This lower bound for the polynomial Q is shown in Fig. 3(a) as a heat map, with warmer colors corresponding to larger values, i.e., blue corresponds to zero and green corresponds to \bar{g} . 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 $\sqrt{\bar{g}}$ away from t_1 . Likewise, consider similar non-negative functions $F_{t_2}(t), F_{s_1}(s), F_{s_2}(s)$ that are zero near and equal to $\sqrt{\bar{g}}$ away from t_2, s_1, s_2 , respectively. Figure 3(b) shows the heat map of $F_{t_1}(t)F_{s_2}(s)$, Fig. 3(c) shows the heat map of $F_{t_2}(t)F_{s_1}(s)$ and lastly Fig. 3(d) shows the heat map of their sum, i.e., $G(\theta) := F_{t_1}(t)F_{s_2}(s) + F_{t_2}(t)F_{s_1}(s)$. Note that G is zero near and larger than \bar{g} away from the impulse locations θ_1, θ_2 , as desired. It only remains to construct the univariate polynomials $q_{t_1}, q_{t_2}, q_{s_1}, q_{s_2} \in \text{span}(\phi_1, \dots, \phi_M)$ that satisfy the inequalities $q_{t_1} \geq F_{t_1}, q_{t_2} \geq F_{t_2}, q_{s_1} \geq F_{s_1}, q_{s_2} \geq F_{s_2}$, with equality at t_1, t_2, s_1, s_2 , respectively. Under the assumptions of Theorem 11, the existence of these univariate polynomials follows from [1]. In this fashion, we finally obtain a non-negative 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 \bar{g} away from the impulses, as desired. For example, for the Gaussian window detailed in Section 2.3 with the standard deviation $\sigma = 0.2$, Fig. 3(e) shows $q_{t_1}(t)$ and Fig. 3(f) shows the heat map of the dual certificate $Q^0(\theta)$, both in logarithmic scale. Yet, another polynomial Q^0 is needed to control the recovery error near the impulses and thus complete the proof of Theorem 11; see Section 4.4.

large noise level. In turn, a large ε leads to a large residual $R(x, K, \varepsilon)$; see (2.8). The correct balance between ε and $R(x, K, \varepsilon)$ depends on the particular choice of the measurement functions $\{\phi_m\}_m$ and is beyond the scope of this paper.

REMARK 14 (Minimum separation). Theorem 11 applies to *any* non-negative measure x . In particular, when x is an atomic measure, Theorem 11 applies *regardless* of the separation between the impulses

present in the measure x . However, it is crucial to note that the recovery error $d_{\text{GW}}(x, \hat{x})$ does indeed depend on the separation of x .

As an example, consider the atomic measure $x = \delta_{0.5} + \delta_{0.51}$. In order to apply Theorem 11, we can set $\varepsilon = 0.01$, so that $x_{K,\varepsilon} = x$ and $R(x, K, \varepsilon) = 0$. Now, the error bound in (2.13) reads as

$$d_{\text{GW}}(x, \hat{x}) \leq c_1(0.01) \cdot \delta + c_2(0.01), \quad (2.14)$$

where we have made explicit the dependence of c_1 on ε for emphasis. Alternatively, we may also apply Theorem 11 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, (2.13) reads as

$$d_{\text{GW}}(x, \hat{x}) \leq c_1(0.2) \cdot \delta + c_2(0.2) + c_3(\varepsilon) \cdot 0.19. \quad (2.15)$$

This work, however, does not address the optimal choice of separation ε as a function of noise level δ . That is, given δ , the choice of ε that would minimize the right-hand side of (2.13) is not studied here; see also [1].

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 non-negative measure as in (2.1) and $\{\phi_m\}_{m=1}^M$ are translated copies of a 1-D Gaussian window, i.e.,

$$\phi_m(t) = g_1(t - t'_m) := e^{-(t-t'_m)^2/\sigma^2}, \quad (2.16)$$

for $T' = \{t'_m\}_{m=1}^M \subset \mathbb{I}$ and standard deviation $\sigma > 0$. Recalling (1.2), note that

$$\begin{aligned} \int_{\mathbb{I}^2} \phi_m(t) \phi_n(s) x(dt, ds) &= \sum_{k=1}^K a_k \cdot g_1(t_k - t'_m) g_1(t_k - t'_n) \quad (\text{see (2.16)}) \\ &= \sum_{k=1}^K a_k \cdot g_2(\theta_k - \theta'_{m,n}) \\ &= \sum_{k=1}^K a_k \cdot e^{-\frac{\|\theta_k - \theta'_{m,n}\|_2^2}{\sigma^2}}, \end{aligned} \quad (2.17)$$

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 of the imaging device. 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

$$\int_{\mathbb{I}^2} \phi_m(t) \phi_n(s) x(dt, ds) = \sum_{k=1}^K a_k \cdot g_2(\theta_k - \theta'_{m,n}) = (g_2 \star x)(\theta'_{m,n}), \quad m, n \in [M], \quad (2.18)$$

where \star stands for convolution. Put differently, the integral above evaluates the Gaussian-blurred (or filtered) copy of measure x at locations $T' \times T'$.

Suppose first that there is no imaging noise and, consequently, $y_{m,n} = (g_2 \star x)(\theta'_{m,n})$ is the (m,n) th pixel of the image, for every $m, n \in [M]$. The Gaussian windows $\{\phi_m\}_{m=1}^M$, specified in (2.16), form a \mathcal{C} -system on \mathbb{I} for arbitrary $T' \subset \mathbb{I}$; see for instance [24, Example 5]. Therefore, in view of Proposition 2, the measure x is the unique solution of Program (1.5) with $\delta' = 0$, provided that $M \geq 2K+1$. This simple argument should be contrasted with the elaborate proofs of earlier 1-D results, for example Theorem 1.3 in [2].

In the presence of noise, i.e., when $\delta \geq 0$, [1, Lemma 23] establishes that all the families of functions in Theorem 11 are indeed $\mathcal{C}_{K,\varepsilon}^*$ -systems on \mathbb{I} , provided that the endpoints of \mathbb{I} are included in the sampling points T' . In fact, [1, Section 1.2] goes further and also evaluates the factors involved in the error bound for 1-D super-resolution; although arguably the result is suboptimal and there is a room for improvement. In principle, those results could be in turn used to evaluate c_1, c_2, c_3 in the error bound of Theorem 11, a direction which is not pursued here. The conclusion of this section is recorded below.

Gaussian window Consider a non-negative measure x supported on \mathbb{I}^2 . Consider also a noise level $\delta \geq 0$ and the family of measurement functions defined in (2.16). For an integer K and $\varepsilon \in (0, 1/2]$, let $x_{K,\varepsilon}$ be a K -sparse and ε -separated non-negative measure on \mathbb{I}^2 that approximates x in the sense of (2.8). With $M \geq 2K + 2$, let \hat{x} be a solution of Program (1.5) with

$$\delta' \geq \left(1 + \sqrt{\frac{2M}{\sigma^2 e}} R(x, k, \varepsilon)\right) \delta;$$

see (2.8). Then,

$$d_{\text{GW}}(x, \hat{x}) \leq c_1 \delta + c_2(\varepsilon) + c_3 R(x, K, \varepsilon), \quad (2.19)$$

where d_{GW} is the generalized Wasserstein metric in (2.7). Above, $c_1, c_2(\varepsilon), c_3$ are specified explicitly in (4.23) and depend on the true measure x , the separation ε and the sampling locations $T' = \{t'_m\}_{m=1}^M$ in (2.16). In particular, it holds that $c_2(0) = 0$.

3. Related work

The current wave of super-resolution research using convex optimization began with the two seminal papers of Candès and Fernandez-Granda [12,27]. 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, i.e., the ‘bandwidth’ of the sensing mechanism. Many papers extended this fundamental result to randomized models [29], support recovery analysis [37,39–42], denoising schemes [43,44], different geometries [45–49] and incorporating prior information [50]. Most of these works easily generalize to multidimensional signals. In addition, a special attention to multidimensional signals was given in a variety of papers; see for instance [51,52].

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 non-negative sparse super-resolution for atomic measures supported on a grid. In [13,14], 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 [53].

In this paper, we focus on the grid-free setting [54,55] in which the non-negative 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 [45, 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 11 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, i.e., \mathcal{C} -systems, does not naturally extend to higher dimensions.

Let us add that the low-noise regime for positive 1-D super-resolution was studied in [20]. 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 the noise level is small. Furthermore, the solution converges to the underlying positive measure if the signal-to-noise ratio scales like $\mathcal{O}(1/\text{sep}^2)$, where sep is the minimal separation between adjacent spikes. In contrast to our work, the framework of [20] builds upon smooth convolution kernel and uses a sparse-promoting regularizer, rather the feasibility problem considered in Program (1.5). In [33], 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 [56], which encodes the support of the measure as zeros of a designed polynomial. Such generalizations include methods such as MUSIC [57], Matrix Pencil [18] and ESPRIT [58], 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 [59]. The separation is not necessary for convex programs only for non-negative measures, like the model considered in this paper. The stability analysis of some of these methods, under a separation condition, is found in [60–65]. However, their extension to 2-D is not trivial and accordingly different methods were proposed [66–70]. 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 ε and $T = \{t_k\}_{k=1}^K \subset \mathbb{I}$, let us define the neighborhoods

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

$$T_\varepsilon := \bigcup_{k=1}^K t_{k,\varepsilon}, \quad (4.1)$$

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

$$\begin{aligned} \nu &\leq |t_k - t_l|, & k \neq l, \quad k, l \in [K], \\ \nu &\leq |t_k - 0|, & \nu \leq |t_k - 1|. \end{aligned} \quad (4.2)$$

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

$$\begin{aligned} \theta_{k,\varepsilon} &:= t_{k,\varepsilon} \times s_{k,\varepsilon} = \{\theta \in \mathbb{I}^2 : \|\theta - \theta_k\|_\infty \leq \varepsilon\} \subset \mathbb{I}^2, \\ \Theta_\varepsilon &:= \bigcup_{k=1}^K \theta_{k,\varepsilon} \subseteq T_\varepsilon \times S_\varepsilon, \end{aligned} \quad (4.3)$$

and let $\theta_{k,\varepsilon}^C$ and Θ_ε^C be the complements of these sets with respect to \mathbb{I}^2 . Above, $\|\theta\|_\infty = \max[|t|, |s|]$ is the ℓ_∞ -norm of $\theta = (t, s)$. Similarly, define the minimum separation of Θ , i.e., the smallest number ν for which both (4.2) holds and

$$\begin{aligned} \nu &\leq |s_k - s_l|, & k \neq l, \quad k, l \in [K], \\ \nu &\leq |s_k - 0|, & \nu \leq |s_k - 1|. \end{aligned} \quad (4.4)$$

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 (1.5) is successful if a certain dual certificate Q exist.

LEMMA 15 Let x be a K -sparse non-negative atomic measure supported on $\Theta \subset \text{interior}(\mathbb{I}^2)$; see (2.1). Then, x is the unique solution of Program (1.5) with $\delta' = 0$ if

- the $M^2 \times K$ matrix $[\phi_m(t_k)\phi_n(s_k)]_{m,n,k=1}^{m=M, n=M, k=K}$ has full column rank, and
- there exist real coefficients $\{b_{m,n}\}_{m,n=1}^M$ 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 $\text{interior}(\mathbb{I}^2)$ and vanishes only on Θ .

The following result, proved in Appendix A, states that the dual certificate required in Lemma 15 exists if the number of measurements M is large enough and the measurement functions $\{\phi_m\}_{m=1}^M$ form a \mathcal{C} -system on \mathbb{I} .

LEMMA 16 (Sparse measure without noise). Let x be a K -sparse non-negative atomic measure supported on $\text{interior}(\mathbb{I}^2)$. For $M \geq 2K + 1$, suppose that $\{\phi_m\}_{m=1}^M$ form a \mathcal{C} -system on \mathbb{I} . Then, the dual certificate Q prescribed in Lemma 15 exists.

Combining Lemmas 15 and 16 completes the proof of Proposition 2.

REMARK 17 (Proof technique of Lemma 16). The technical challenge in the proof of Lemma 16 is that \mathcal{C} -systems do not generalize to two dimensions. Indeed, to prove our claim, we effectively reduce the construction of the dual polynomial $Q(\theta)$ with $\theta = (t, s)$ into the construction of a number of univariate

polynomials in t and s . The key observation of the proof is the following. Suppose for simplicity that $\phi_1 \equiv 1$. 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 $q_T(t)$ of $\{\phi_m(t)\}_{m=1}^M$ is non-negative on \mathbb{I} and only vanishes on T . Similarly, consider a polynomial $q_S(s)$ of $\{\phi_n(s)\}_{n=1}^M$ that is non-negative on \mathbb{I} and only vanishes on S . Then, the polynomial $Q(\theta) = q_T(t) + q_S(s)$ is non-negative on \mathbb{I}^2 and vanishes only on $T \times S$.

In general $\Theta \subset (T \times S)$ and, consequently, the above Q will have unwanted zeros on $(T \times S) \setminus \Theta$. This issue may be addressed by replacing the $M^2 \times K$ matrix in Lemma 15 with a larger matrix of size $M^2 \times K^2$. Alternatively, we use in the proof a more nuanced argument to construct a different polynomial Q that vanishes exactly on Θ , without the unwanted zeros on $(T \times S) \setminus \Theta$; see Appendix A for details. This more nuanced argument also lends itself naturally to noisy super-resolution, as we will see later.

4.3 Geometric intuition for Proposition 2

Proposition 2 states that the imaging operator Φ in (1.4) is injective on all K -sparse non-negative 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 \mathcal{C} -system on \mathbb{I} . Here, we provide some geometric intuition about the role of the dual certificate in the proof of Proposition 2.

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

$$\mathcal{C} := \left\{ \int_{\mathbb{I}^2} \Phi(\theta) \chi(d\theta) : \chi \text{ is a non-negative measure on } \mathbb{I}^2 \right\} \subset \mathbb{R}^{M \times M}. \quad (4.5)$$

By the continuity of Φ and with an application of the dominated convergence theorem, it is easy to verify that \mathcal{C} is a closed convex cone, i.e., \mathcal{C} is a homogeneous and closed convex subset of $\mathbb{R}^{M \times M}$. When $\{\phi_m\}_{m=1}^M$ form a \mathcal{C} -system on \mathbb{I} , it is also not difficult to verify that $\{\Phi(t_l, s_l)\}_{l=1}^{M^2}$ are linearly independent matrices in $\mathbb{R}^{M \times M}$. This in particular implies that \mathcal{C} is a convex body, i.e., the interior of \mathcal{C} is not empty. Note also that $y \in \mathcal{C}$ because

$$y = \int_{\mathbb{I}^2} \Phi(\theta) x(d\theta) = \sum_{k=1}^M a_k \Phi(\theta_k). \quad (4.6)$$

For Program (1.5) to successfully recover the measure x , it suffices that

$$\mathcal{A} = \text{cone} \left(\{\Phi(\theta_k)\}_{k=1}^K \right) = \left\{ \sum_{k=1}^K \alpha_k \Phi(\theta_k) : \alpha_k \geq 0, \forall k \in [K] \right\}, \quad (4.7)$$

is a K -dimensional exposed face of the cone \mathcal{C} . See [71, Section 18] for the definition of exposed face. This in turn 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 \mathcal{C} at \mathcal{A} , i.e., when we can find b such that

$$\begin{cases} \langle b, c \rangle = 0, & \forall c \in \mathcal{A}, \\ \langle b, c \rangle > 0, & \forall c \in \mathcal{C} \setminus \mathcal{A}. \end{cases} \quad (4.8)$$

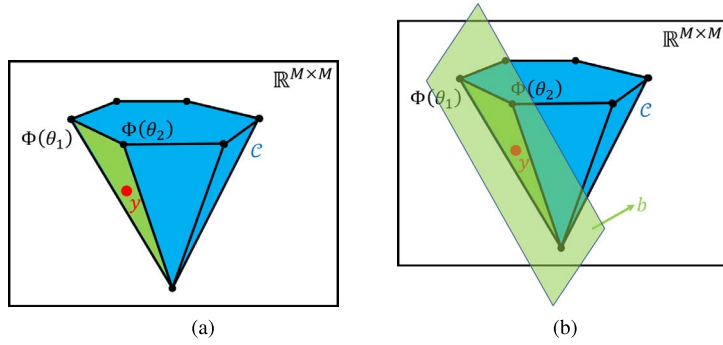


FIG. 4. This figure complements Section 4.3. Figure 4(a) shows the conic hull \mathcal{C} of the range of the imaging operator $\Phi : \mathbb{I}^2 \rightarrow \mathbb{R}^{M \times M}$. Note that the image y in (2.2) belongs to the cone \mathcal{C} ; see (4.6). For Program (1.5) to successfully recover the true measure x from the image y , it suffices that $\{\Phi(\theta_k)\}_{k=1}^K$ form a K -dimensional exposed face of the cone \mathcal{C} , where $\{\theta_k\}_{k=1}^K$ is the support of the measure x . This face is shown in green. In words, this condition is sufficient for x to be the unique solution of Program (1.5). Equivalently, it suffices to find a hyperplane, with normal vector b , that strictly supports \mathcal{C} on this face. This latter 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 the measure x .

Invoking (4.7), we find that (4.8) is equivalent to finding $b \in \mathbb{R}^{M \times M}$ such that

$$\begin{cases} \langle b, \Phi(\theta_k) \rangle = 0, & k \in [K], \\ \langle b, \Phi(\theta) \rangle > 0, & \theta \notin \{\theta_k\}_{k=1}^K. \end{cases} \quad (4.9)$$

In other words, for Program (1.5) to successfully recover the measure x , it suffices to find a ‘polynomial’

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

which vanishes on $\{\theta_k\}_{k=1}^K$ and is positive elsewhere on \mathbb{I}^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 \mathcal{C} , itself another convex cone in $\mathbb{R}^{M \times M}$, consists of the coefficients of all non-negative polynomials of $\{\phi_m \phi_n\}_{m,n=1}^M$ on \mathbb{I}^2 and in particular the coefficient vector b above belongs to an $(M^2 - K)$ -dimensional face of this polar cone [24]. We refer the reader to Fig. 4 for an illustration of the convex geometry underlying the problem of non-negative super-resolution.

4.4 Proof of Theorem 11 (arbitrary measure with noise)

In this section, we will prove the main result of this paper, i.e., Theorem 11. For an integer K and $\varepsilon \in (0, 1/2]$, let $x_{K,\varepsilon}$ be a K -sparse and ε -separated non-negative measure on \mathbb{I}^2 that approximates x with residual $R(x, K, \varepsilon)$, in the sense of (2.8). 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. Consider also the neighborhoods $\{\theta_{k,\varepsilon}\}_{k=1}^K \subseteq \mathbb{I}^2$ and $\Theta_\varepsilon = \cup_{k=1}^K \theta_{k,\varepsilon}$ defined in (4.3). Let $\{\theta_{k,\varepsilon}^C\}_{k=1}^K$ and Θ_ε^C denote the complements of these sets with respect to \mathbb{I}^2 .

Before turning to the details, let us outline the proof. We will show in Section 4.4.1 that the existence of certain dual certificates leads to a stable recovery of $x_{K,\varepsilon}$ with Program (1.5). Then, we show in Section 4.4.2 that these certificates exist under certain conditions on the imaging apparatus. Finally, in Section 4.4.3, we complete the proof of Theorem 11 by applying the triangle inequality to control $d_{\text{GW}}(x, \hat{x})$ as

$$d_{\text{GW}}(x, \hat{x}) \leq d_{\text{GW}}(x, x_{K,\varepsilon}) + d_{\text{GW}}(x_{K,\varepsilon}, \hat{x}) \leq R(x, K, \varepsilon) + d_{\text{GW}}(x_{K,\varepsilon}, \hat{x}). \quad (\text{see (2.8)})$$

4.4.1 Dual certificates. Lemmas 18 and 19 below show that Program (1.5) stably recovers the atomic measure $x_{K,\varepsilon}$ in the presence of noise, provided that certain dual certificates exist. The proofs, which appear in Appendices B and C, are standard and Lemmas 18 and 19 below are extensions of Lemmas 16 and 17, respectively, in [1]. In particular, Lemma 18 below controls the recovery error away from the support Θ of $x_{K,\varepsilon}$, whereas Lemma 19 controls the error near the support. The latter features an approximate dual certificate and shares some broad similarities with [72].

LEMMA 18 (Error away from the support). Let \hat{x} be a solution of Program (1.5) with $\delta' \geq \delta$, and set $h := \hat{x} - x_{K,\varepsilon}$ to be the error. Fix a positive scalar \bar{g} . Suppose that there exist real coefficients $\{b_{m,n}\}_{m,n=1}^M$ and a polynomial

$$Q(\theta) = Q(t, s) = \sum_{m,n=1}^M b_{m,n} \phi_m(t) \phi_n(s), \quad (4.10)$$

such that

$$Q(\theta) \geq G(\theta) := \begin{cases} 0, & \text{when there exists } k \in [K] \text{ such that } \theta \in \theta_{k,\varepsilon}, \\ \bar{g}, & \text{elsewhere in interior } (\mathbb{I}^2), \end{cases} \quad (4.11)$$

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

$$\int_{\Theta_\varepsilon^c} h(d\theta) \leq 2\|b\|_{\text{F}} \delta' / \bar{g}, \quad (4.12)$$

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

LEMMA 19 [Error near the support] Let \hat{x} be a solution of Program (1.5) with $\delta' \geq \delta$, and set $h := \hat{x} - x_{K,\varepsilon}$ to be the error. For $\alpha \in [0, 1]$, suppose that there exist real coefficients $\{b_{m,n}^0\}_{m,n=1}^M$ and a polynomial

$$Q^0(\theta) = Q^0(t, s) = \sum_{m,n=1}^M b_{m,n}^0 \phi_m(t) \phi_n(s), \quad (4.13)$$

such that

$$Q^0(\theta) \geq G^0(\theta) := \begin{cases} 1, & \text{when there exists } k \in [K] \text{ such that } \theta \in \theta_{k,\varepsilon} \text{ and } \int_{\theta_{k,\varepsilon}} h(d\theta) > 0, \\ -1, & \text{when there exists } k \in [K] \text{ such that } \theta \in \theta_{k,\varepsilon} \text{ and } \int_{\theta_{k,\varepsilon}} h(d\theta) \leq 0, \\ -1 + \alpha & \text{when there exists } k \in [K] \text{ such that } \theta = \theta_k \text{ and } \int_{\theta_{k,\varepsilon}} h(d\theta) \leq 0, \\ -1, & \text{elsewhere in interior}(\mathbb{I}^2), \end{cases} \quad (4.14)$$

where the equality holds on Θ . Then, it holds that

$$\sum_{k=1}^K \left| \int_{\theta_{k,\varepsilon}} h(d\theta) \right| \leq \alpha \|x_{K,\varepsilon}\|_{\text{TV}} + 2 \left(\|b^0\|_{\text{F}} + \frac{\|b\|_{\text{F}}}{\bar{g}} \right) \delta', \quad (4.15)$$

where $b^0 \in \mathbb{R}^{M \times M}$ is the matrix formed by the coefficients $\{b_{m,n}^0\}_{m,n=1}^M$, and the matrix b was introduced in Lemma 18.

By combining Lemmas 18 and 19, the next result bounds the error $d_{\text{GW}}(x_{K,\varepsilon}, \widehat{x})$. The proof is omitted as the steps are identical to those taken in [1, Lemma 18].

LEMMA 20 (Error in Wassertein metric). Suppose that the dual certificates Q and Q^0 in Lemmas 18 and 19 exist. Then,

$$d_{\text{GW}}(x_{K,\varepsilon}, \widehat{x}) \leq \left(\frac{8\|b\|_{\text{F}}}{\bar{g}} + 6\|b^0\|_{\text{F}} \right) \delta' + (\varepsilon + 3\alpha) \|x_{K,\varepsilon}\|_{\text{TV}}. \quad (4.16)$$

In order to apply Lemma 20, we must first show that the dual certificates Q and Q^0 specified in Lemmas 18 and 19 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 construct the dual certificates required in Lemmas 18 and 19, some preparation is necessary. For notational convenience, throughout we model $x_{K,\varepsilon}$ by (2.1), with x therein replaced with $x_{K,\varepsilon}$. Then, recall from Section 4.1 that $\Theta = \{\theta_k\}_{k=1}^K = \{(t_k, s_k)\}_{k=1}^K$ are the impulse locations of $x_{K,\varepsilon}$, and let $T = \{t_k\}_{k=1}^K$ and $S = \{s_k\}_{k=1}^K$ for short. In particular, note that $\Theta \subseteq T \times S$.

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 brevity. For a finite set of distinct points $T' \subset \mathbb{I}$ and $\varepsilon \in (0, \text{sep}(T'))$, let us also define the function $F_{T'} : \mathbb{I} \rightarrow \mathbb{R}$ as

$$F_{T'}(t) := \begin{cases} 0, & \text{when there exists } t' \in T' \text{ such that } t \in t'_\varepsilon, \\ 1, & \text{elsewhere on } \mathbb{I}, \end{cases} \quad (4.17)$$

where $t'_\varepsilon = \{t \in \mathbb{I} : |t - t'| \leq \varepsilon\}$ is the ε -neighborhood of t' . The following result, proved in Appendix D, states that the dual certificate prescribed in Lemma 18 exists if both $\{F_{T_\Omega}\} \cup \{\phi_m\}_{m=1}^M$ and $\{F_{S_\Omega}\} \cup \{\phi_m\}_{m=1}^M$ are \mathcal{C}^* -systems for every index set $\Omega \subseteq [K]$.

PROPOSITION 21 (Dual certificate for faraway error). Suppose that $\{\phi_m\}_{m=1}^M$ form a \mathcal{C} -system on \mathbb{I} with $M \geq 2K + 2$. For every index set $\Omega \subseteq [K]$, suppose also that $\{F_{T_\Omega}\} \cup \{\phi_m\}_{m=1}^M$ and $\{F_{S_\Omega}\} \cup \{\phi_m\}_{m=1}^M$ are both $\mathcal{C}_{K,\varepsilon}^*$ -systems on \mathbb{I} ; see (4.17). Then, the dual certificate Q , as specified in Lemma 18, exists with

$$\bar{g} = 2^{K-2}. \quad (4.18)$$

Next, for $t', s' \in \mathbb{I}$ and $\varepsilon \in (0, 1/2]$, let us define the functions $F_{t'}^\pm : \mathbb{I} \rightarrow \mathbb{R}$ and $F_{s'}^+ : \mathbb{I} \rightarrow \mathbb{R}$ as

$$F_{t'}^\pm(t) := \begin{cases} \pm 1, & \text{when } t \in t'_\varepsilon, \\ 0, & \text{elsewhere on } \mathbb{I}, \end{cases}$$

$$F_{s'}^+(s) := \begin{cases} 1, & \text{when } s \in s'_\varepsilon, \\ 0, & \text{elsewhere on } \mathbb{I}, \end{cases} \quad (4.19)$$

where t'_ε and s'_ε are the ε -neighborhoods of t' and s' , respectively. The following result, proved in Appendix E, states that the dual certificate, prescribed in Lemma 19, exists when certain \mathcal{C}^* -systems exist.

PROPOSITION 22 (Dual certificate for nearby error). For $M \geq 2K + 2$, suppose that $\{\phi_m\}_{m=1}^M$ form a \mathcal{C} -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 $\mathcal{C}_{K,\varepsilon}^*$ -systems on \mathbb{I} ; see (4.19). Then, the dual certificate Q^0 , as specified in Lemma 19, exists with $\alpha = \alpha(\varepsilon) = 1 - 1/q_{\max}^{\pi_*}(\varepsilon)$, where $q_{\max}^{\pi_*}(\varepsilon)$ is defined in (E2) and (E7). In particular, $\alpha(0) = 0$.

4.4.3 *Completing the proof of Theorem 11* Recall that the imaging apparatus is L -Lipschitz in the sense of (2.9). From this assumption and the triangle inequality, it follows that

$$\begin{aligned} \left\| y - \int_{\mathbb{I}^2} \Phi(\theta) x_{K,\varepsilon}(\mathrm{d}\theta) \right\|_{\mathbb{F}} &\leq \left\| y - \int_{\mathbb{I}^2} \Phi(\theta) x(\mathrm{d}\theta) \right\|_{\mathbb{F}} + \left\| \int_{\mathbb{I}^2} \Phi(\theta) (x(\mathrm{d}\theta) - x_{K,\varepsilon}(\mathrm{d}\theta)) \right\|_{\mathbb{F}} \\ &\leq \delta + L \cdot d_{\mathrm{GW}}(x, x_{K,\varepsilon}) \quad (\text{see (1.4, 2.9)}) \\ &= \delta + L \cdot R(x, K, \varepsilon) := \delta'. \quad (\text{see (2.8)}) \end{aligned} \quad (4.20)$$

In words, a solution \hat{x} of Program (1.5), with δ' specified above, can be thought of as an estimate of $x_{K,\varepsilon}$. Recall that we also constructed the prescribed dual certificates Q and Q^0 in Section 4.4.2; see Propositions 21 and 22. Consequently, Lemmas 18–20 are in force. The following argument thus

completes the proof of Theorem 11:

$$\begin{aligned}
d_{\text{GW}}(x, \hat{x}) &\leq d_{\text{GW}}(x, x_{K,\varepsilon}) + d_{\text{GW}}(x_{K,\varepsilon}, \hat{x}) \quad (\text{triangle inequality}) \\
&\leq R(x, K, \varepsilon) + \left(\frac{8\|b\|_{\text{F}}}{\bar{g}} + 6\|b^0\|_{\text{F}} \right) \delta' + (\varepsilon + 3\alpha(\varepsilon)) \|x_{K,\varepsilon}\|_{\text{TV}}. \quad (\text{see (2.8, 4.16)}) \\
&= R(x, K, \varepsilon) + \left(\frac{8\|b\|_{\text{F}}}{\bar{g}} + 6\|b^0\|_{\text{F}} \right) (\delta + L \cdot R(x, K, \varepsilon)) + (\varepsilon + 3\alpha(\varepsilon)) \|x_{K,\varepsilon}\|_{\text{TV}} \quad (\text{see (4.20)}) \\
&= \left(\frac{8\|b\|_{\text{F}}}{\bar{g}} + 6\|b^0\|_{\text{F}} \right) \delta + \left(\frac{8L\|b\|_{\text{F}}}{\bar{g}} + 6L\|b^0\|_{\text{F}} + 1 \right) R(x, K, \varepsilon) + (\varepsilon + 3\alpha(\varepsilon)) \|x_{K,\varepsilon}\|_{\text{TV}} \\
&= \left(2^{5-K} \|b\|_{\text{F}} + 6\|b^0\|_{\text{F}} \right) \delta + \left(2^{5-K} L\|b\|_{\text{F}} + 6L\|b^0\|_{\text{F}} + 1 \right) R(x, K, \varepsilon) \\
&\quad + (\varepsilon + 3\alpha(\varepsilon)) \|x_{K,\varepsilon}\|_{\text{TV}}. \quad (\text{see (4.18)}) \tag{4.21}
\end{aligned}$$

The first inequality above holds because the generalized Wasserstein distance d_{GW} in (2.7) indeed satisfies the triangle inequality; see [30, Proposition 5]. In the last line above, it would be more convenient to relate $\|x_{K,\varepsilon}\|_{\text{TV}}$ back to $\|x\|_{\text{TV}}$. To that end, we write that

$$\begin{aligned}
\|x_{K,\varepsilon}\|_{\text{TV}} &= d_{\text{GW}}(x_{K,\varepsilon}, 0) \quad (\text{see (2.7)}) \\
&\leq d_{\text{GW}}(x, x_{K,\varepsilon}) + d_{\text{GW}}(x, 0) \quad (\text{triangle inequality}) \\
&= d_{\text{GW}}(x, x_{K,\varepsilon}) + \|x\|_{\text{TV}} \quad (\text{see (2.7)}) \\
&\leq d_{\text{GW}}(x, 0) + \|x\|_{\text{TV}} \quad (\text{see (2.8)}) \\
&= 2\|x\|_{\text{TV}}. \quad (\text{see (2.7)}) \tag{4.22}
\end{aligned}$$

Finally, in light of (4.21, 4.22), the components of error bound in Theorem 11 are given explicitly as

$$\begin{aligned}
c_1 &= 32\|b\|_{\text{F}} + 6\|b^0\|_{\text{F}}, \\
c_2(\varepsilon) &= (\varepsilon + 3\alpha(\varepsilon)) \|x\|_{\text{TV}}, \\
c_3 &= 32L\|b\|_{\text{F}} + 6L\|b^0\|_{\text{F}} + 1. \tag{4.23}
\end{aligned}$$

5. Perspective

In this paper, we have shown that a simple convex feasibility program is guaranteed to robustly recover a sparse (non-negative) 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 over all sparse images and when we can stably find its inverse. These results build upon and extend a recent paper [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 [45–48,73]. 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 11 is rather ponderous. 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 11. Another interesting research direction is deriving the optimal separation ε (as a function of noise level δ) that minimizes the right-hand side of the error bound in (2.13). Such a result will provide the tightest error bound for Program (1.5).

This work has focused solely on the theoretical performance of Program (1.5). 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 non-negative measures, reduces the recovery error.

Acknowledgements

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. We are deeply indebted to the anonymous reviewers of this work for their careful and detailed comments. We would also like to sincerely thank Jean-Baptiste Seby for spotting an error in the earlier version of this manuscript.

Funding

Alan Turing Institute (EP/N510129/1 to A.E.); Turing Seed Funding (SF019 to A.E.); National Science Foundation (CCF-1704204 to G.T.); Defense Advanced Research Projects Agency Lagrange Program (N660011824020 to G.T.).

Data Availability Statement

No new data were generated or analyzed in support of this review.

REFERENCES

1. EFTEKHARI, A., TANNER, J., THOMPSON, A., TOADER, B. & TYAGI, H. (2019) Sparse non-negative super-resolution—simplified and stabilised. *Appl. Comput. Harmon. Anal.*
2. SCHIEBINGER, G., ROBEVA, E. & RECHT, B. (2017) Superresolution without separation. *Inf. Inference*, **7**, 1–30.
3. PUSCHMANN, K. G. & KNEER, F. (2005) On super-resolution in astronomical imaging. *Astron. Astrophys.*, **436**, 373–378.
4. KHAIDUKOV, V., LANDA, E. & MOSER, T. J. (2004) Diffraction imaging by focusing-defocusing: an outlook on seismic superresolution. *Geophysics*, **69**, 1478–1490.
5. BETZIG, E., PATTERSON, G. H., SOUGRAT, R., LINDWASSER, O. W., OLENYCH, S., BONIFACINO, J. S., DAVIDSON, M. W., LIPPINCOTT-SCHWARTZ, J. & HESS, H. F. (2006) Imaging intracellular fluorescent proteins at nanometer resolution. *Science*, **313**, 1642–1645.

6. HESS, S. T., GIRIRAJAN, T. P. K. & MASON, M. D. (2006) Ultra-high resolution imaging by fluorescence photoactivation localization microscopy. *Biophys. J.*, **91**, 4258–4272.
7. RUST, M. J., BATES, M. & ZHUANG, X. (2006) Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nat. Methods*, **3**, 793–796.
8. EKANADHAM, C., TRANCHINA, D. & SIMONCELLI, E. P. (2011) Neural spike identification with continuous basis pursuit. *Computational and Systems Neuroscience (CoSyNe), Salt Lake City, Utah*.
9. HELL, S. (2009) Primer: fluorescence imaging under the diffraction limit. *Nat. Methods*, **6**, 19.
10. TUR, R., ELDAR, Y. C. & FRIEDMAN, Z. (2011) Innovation rate sampling of pulse streams with application to ultrasound imaging. *IEEE Trans. Signal Process.*, **59**, 1827–1842.
11. SOLOMON, O., ELDAR, Y. C., MUTZAFI, M. & SEGEV, M. (2019) Sparcom: sparsity based super-resolution correlation microscopy. *SIAM J. Imaging Sci.*, **12**, 392–419.
12. CANDÈS, E. J. & FERNANDEZ-GRANDA, C. (2014) Towards a mathematical theory of super-resolution. *Commun. Pure Appl. Math.*, **67**, 906–956.
13. BENDORY, T. (2017) Robust recovery of positive stream of pulses. *IEEE Trans. Signal Process.*, **65**, 2114–2122.
14. MORGENSTERN, V. I. & CANDÈS, E. J. (2016) Super-resolution of positive sources: the discrete setup. *SIAM J. Imaging Sci.*, **9**, 412–444.
15. SLAWSKI, M. & HEIN, M. (2013) Non-negative least squares for high-dimensional linear models: consistency and sparse recovery without regularization. *Electron. J. Stat.*, **7**, 3004–3056.
16. FOUCART, S. & KOSLICKI, D. (2014) Sparse recovery by means of nonnegative least squares. *IEEE Signal Process. Lett.*, **21**, 498–502.
17. WEISS, L. & McDONOUGH, R. N. (1963) Prony's method, Z-transforms, and Pade approximation. *SIAM Rev.*, **5**, 145–149.
18. HUA, Y. & SARKAR, T. K. (1990) Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise. *IEEE Trans. Acoust. Speech Signal Process.*, **38**, 814–824.
19. DE CASTRO, Y. & GAMBOA, F. (2012) Exact reconstruction using Beurling minimal extrapolation. *J. Math. Anal. Appl.*, **395**, 336–354.
20. DENOYELLE, Q., DUVAL, V. & PEYRÉ, G. (2017) Support recovery for sparse super-resolution of positive measures. *J. Fourier Anal. Appl.*, **23**, 1153–1194.
21. EFTEKHARI, A. & THOMPSON, A. (2019) Sparse inverse problems over measures: equivalence of the conditional gradient and exchange methods. *SIAM J. Optim.*, **29**, 1329–1349.
22. BOYD, N., SCHIEBINGER, G. & RECHT, B. (2017) The alternating descent conditional gradient method for sparse inverse problems. *SIAM J. Optim.*, **27**, 616–639.
23. BREDIES, K. & PIKKARAINEN, H. K. (2013) Inverse problems in spaces of measures. *ESAIM Control Optim. Calc. Var.*, **19**, 190–218.
24. KARLIN, S. & STUDDEN, W. J. (1966) *Tchebycheff Systems: With Applications in Analysis and Statistics*. Pure and Applied Mathematics. University of California: Interscience Publishers.
25. KARLIN, S. (1968) *Total Positivity, Volume I*. Total Positivity. Stanford University Press.
26. KREIN, M. G., NUDELMAN, A. A. & LOUVISH, D. (1977) *The Markov Moment Problem and Extremal Problems*. Translations of Mathematical Monographs. American Mathematical Society.
27. CANDÈS, E. J. & FERNANDEZ-GRANDA, C. (2013) Super-resolution from noisy data. *J. Fourier Anal. Appl.*, **19**, 1229–1254.
28. BENDORY, T., DEKEL, S. & FEUER, A. (2016) Robust recovery of stream of pulses using convex optimization. *J. Math. Anal. Appl.*, **442**, 511–536.
29. TANG, G., BHASKAR, B. N., SHAH, P. & RECHT, B. (2013) Compressed sensing off the grid. *IEEE Trans. Inf. Theory*, **59**, 7465–7490.
30. PICCOLI, B. & ROSSI, F. (2014) Generalized Wasserstein distance and its application to transport equations with source. *Arch. Ration. Mech. Anal.*, **211**, 335–358.
31. CHIZAT, L., PEYRE, G., SCHMITZER, B. & VIALARD, F.-X. (2018) Scaling algorithms for unbalanced optimal transport problems. *Math. Comput.*, **87**, 2563–2609.

32. VILLANI, C. (2008) *Optimal Transport: Old and New*. Grundlehren der Mathematischen Wissenschaften. Berlin Heidelberg: Springer.
33. POON, C. & PEYRÉ, G. (2019) Multidimensional sparse super-resolution. *SIAM J. Math. Anal.*, **51**, 1–44.
34. DENOYELLE, Q., DUVAL, V. & PEYRÉ, G. (2017) Support recovery for sparse super-resolution of positive measures. *J. Fourier Anal. Appl.*, **23**, 1153–1194.
35. DUVAL, V. & PEYRE, G. (2015) Exact support recovery for sparse spikes deconvolution. *Found. Comput. Math.*, **15**, 1315–1355.
36. CORMEN, T. H., LEISERSON, C. E., RIVEST, R. L. & STEIN, C. (2009) *Introduction to Algorithms*. Computer Science. Cambridge, MA: MIT Press.
37. FERNANDEZ-GRANDA, C. (2013) Support detection in super-resolution. In *Proceedings of the 10th International Conference on Sampling Theory and Applications (SampTA 2013)* (pp. 145–148).
38. CANDÈS, E. J. & WAKIN, M. B. (2008) An introduction to compressive sampling. *IEEE Signal Process. Mag.*, **25**, 21–30.
39. LI, Q. & TANG, G. (2016) Approximate support recovery of atomic line spectral estimation: a tale of resolution and precision. *2016 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*. New York City, NY: IEEE, pp. 153–156.
40. DUVAL, V. & PEYRÉ, G. (2015) Exact support recovery for sparse spikes deconvolution. *Found. Comput. Math.*, **15**, 1315–1355.
41. AZAIS, J. M., DE CASTRO, Y. & GAMBOA, F. (2015) Spike detection from inaccurate samplings. *Appl. Comput. Harmon. Anal.*, **38**, 177–195.
42. BENDORY, T., BAR-ZION, A. D., ADAM, D., DEKEL, S. & FEUER, A. (2016) Stable support recovery of stream of pulses with application to ultrasound imaging. *IEEE Trans. Signal Processing*, **64**, 3750–3759.
43. BHASKAR, B. N., TANG, G. & RECHT, B. (2013) Atomic norm denoising with applications to line spectral estimation. *IEEE Trans. Signal Process.*, **61**, 5987–5999.
44. TANG, G., BHASKAR, B. N. & RECHT, B. (2015) Near minimax line spectral estimation. *IEEE Trans. Inf. Theory*, **61**, 499–512.
45. BENDORY, T., DEKEL, S. & FEUER, A. (2015) Exact recovery of Dirac ensembles from the projection onto spaces of spherical harmonics. *Constr. Approx.*, **42**, 183–207.
46. BENDORY, T., DEKEL, S. & FEUER, A. (2015) Super-resolution on the sphere using convex optimization. *IEEE Trans. Signal Process.*, **63**, 2253–2262.
47. BENDORY, T., DEKEL, S. & FEUER, A. (2014) Exact recovery of non-uniform splines from the projection onto spaces of algebraic polynomials. *J. Approx. Theory*, **182**, 7–17.
48. FILBIR, F. & SCHRÖDER, K. (2016) Exact recovery of discrete measures from Wigner D-moments. arXiv preprint arXiv:1606.05306.
49. DOSSAL, C., DUVAL, V. & POON, C. (2017) Sampling the Fourier transform along radial lines. *SIAM J. Numer. Anal.*, **55**, 2540–2564.
50. MISHRA, K. V., CHO, M., KRUGER, A. & XU, W. (2015) Spectral super-resolution with prior knowledge. *IEEE Trans. Signal Process.*, **63**, 5342–5357.
51. DE CASTRO, Y., GAMBOA, F., HENRION, D. & LASSERRE, J.-B. (2017) Exact solutions to super resolution on semi-algebraic domains in higher dimensions. *IEEE Trans. Inf. Theory*, **63**, 621–630.
52. XU, W., CAI, J.-F., MISHRA, K. V., CHO, M. & KRUGER, A. (2014) Precise semidefinite programming formulation of atomic norm minimization for recovering d-dimensional ($d \geq 2$) off-the-grid frequencies. *2014 Information Theory and Applications Workshop (ITA)*. New York City, NY: IEEE, pp. 1–4.
53. BENDORY, T. & EL-DAR, Y. C. (2015) Recovery of sparse positive signals on the sphere from low resolution measurements. *IEEE Signal Process. Lett.*, **22**, 2383–2386.
54. EFTEKHARI, A., ROMBERG, J. & WAKIN, M. B. (2013) Matched filtering from limited frequency samples. *IEEE Trans. Inf. Theory*, **59**, 3475–3496.
55. EFTEKHARI, A., ROMBERG, J. & WAKIN, M. B. (2011) A probabilistic analysis of the compressive matched filter. *Proceedings of the 9th International Conference on Sampling Theory and Applications (SampTA)*. New York, NY: IEEE.

56. STOICA, P., MOSES, R. L., WANG, Y., LI, J. & STOICA, P. (2005) *Spectral Analysis of Signals*, vol. 1. Upper Saddle River, NJ: Pearson Prentice Hall.
57. SCHMIDT, R. (1986) Multiple emitter location and signal parameter estimation. *IEEE Trans. Antennas Propag.*, **34**, 276–280.
58. ROY, R. & KAILATH, T. (1989) ESPRIT-estimation of signal parameters via rotational invariance techniques. *IEEE Trans. Acoust. Speech Signal Process.*, **37**, 984–995.
59. TANG, G. (2015) Resolution limits for atomic decompositions via Markov–Bernstein type inequalities. *2015 International Conference on Sampling Theory and Applications (SampTA)*. New York City, NY: IEEE, pp. 548–552.
60. LIAO, W. & FANNJIANG, A. (2016) Music for single-snapshot spectral estimation: stability and super-resolution. *Appl. Comput. Harmon. Anal.*, **40**, 33–67.
61. FANNJIANG, A. (2016) Compressive spectral estimation with single-snapshot ESPRIT: stability and resolution. arXiv preprint arXiv:1607.01827.
62. MOITRA, A. (2015) Super-resolution, extremal functions and the condition number of Vandermonde matrices. *Proceedings of the Forty-seventh Annual ACM Symposium on Theory of Computing*. New York City, NY: ACM, pp. 821–830.
63. LIAO, W. (2015) Music for multidimensional spectral estimation: stability and super-resolution. *IEEE Trans. Signal Process.*, **63**, 6395–6406.
64. EFTEKHARI, A. & WAKIN, M. B. (2015) Greed is super: a fast algorithm for super-resolution. arXiv preprint arXiv:1511.03385.
65. EFTEKHARI, A. & WAKIN, M. B. (2013) Greed is super: a new iterative method for super-resolution. *2013 IEEE Global Conference on Signal and Information Processing*. New York City, NY: IEEE, pp. 631–631.
66. SACCHINI, J. J., STEEDLY, W. M. & MOSES, R. L. (1993) Two-dimensional Prony modeling and parameter estimation. *IEEE Trans. Signal Process.*, **41**, 3127–3137.
67. ONGIE, G. & JACOB, M. (2016) Off-the-grid recovery of piecewise constant images from few Fourier samples. *SIAM J. Imaging Sci.*, **9**, 1004–1041.
68. PETER, T., PLONKA, G. & SCHABACK, R. (2017) Reconstruction of multivariate signals via Prony’s method. *Proc. Appl. Math. Mech.*, **490**, 31–47.
69. KUNIS, S., PETER, T., RÖMER, T. & VON DER OHE, U. (2016) A multivariate generalization of Prony’s method. *Linear Algebra Appl.*, **490**, 31–47.
70. ANDERSSON, F. & CARLSSON, M. (2018) ESPRIT for multidimensional general grids. *SIAM J. Matrix Anal. Appl.*, **39**, 1470–1488.
71. ROCKAFELLAR, R. T. (1970) *Convex Analysis*. Princeton Landmarks in Mathematics and Physics. Princeton, New Jersey: Princeton University Press.
72. CANDÈS, E. J. & PLAN, Y. (2011) A probabilistic and ripless theory of compressed sensing. *IEEE Trans. Inf. Theory*, **57**, 7235–7254.
73. DE CASTRO, Y. & MIJOLE, G. (2015) Non-uniform spline recovery from small degree polynomial approximation. *J. Math. Anal. Appl.*, **430**, 971–992.

A. Proof of Lemma 16

Consider the $M^2 \times K$ matrix

$$A = [\phi_m(t_k)\phi_n(s_k)]_{m,n,k=1}^{m=M,n=M,k=K}.$$

Note that A is a column submatrix of the $M^2 \times K^2$ matrix

$$B = [\phi_m(t_k)\phi_n(s_l)]_{m,n,k,l=1}^{m=M,n=M,k=K,l=K}.$$

Therefore, to show that A has full column rank, it suffices to show that B is non-singular. Note that B itself can be written as the Kronecker product of two $M \times K$ matrices, i.e.,

$$B = [\phi_m(t_k)]_{m,k=1}^{m=M,k=K} \otimes [\phi_n(s_l)]_{n,l=1}^{n=M,l=K} =: B_1 \otimes B_2,$$

where \otimes stands for Kronecker product. Since by assumption $\{\phi_m\}_{m=1}^M$ form a \mathcal{C} -system on \mathbb{I} and $M \geq 2K + 1 \geq K$, both B_1 and B_2 are non-singular. It follows that B , too, is non-singular, as claimed.

Next, we recall [1, Lemma 15], originally proved in [24, Theorem 5.1], stated below for convenience.

LEMMA 23 (Univariate polynomial of a \mathcal{C} -system). Consider a set $T' \subset \mathbb{I}$ of size K' . With $M \geq 2K' + 1$, suppose that $\{\phi_m\}_{m=1}^M$ form a \mathcal{C} -system on \mathbb{I} . Then, there exist coefficients $\{b_m\}_{m=1}^M$ such that the polynomial $q_{T'} = \sum_{m=1}^M b_m \phi_m$ is non-negative on \mathbb{I} and vanishes only on T' .

Recall that $\Theta = \{\theta_k\}_{k=1}^K = \{(t_k, s_k)\}_{k=1}^K$ are the impulse locations, and let us set $T = \{t_k\}_{k=1}^K$ and $S = \{s_l\}_{l=1}^K$ for short. For an index set $\Omega \subseteq [K]$, let $[K] \setminus \Omega$ denote its complement with respect to $[K]$. Let also $T_\Omega = \{t_k\}_{k \in \Omega}$ and $S_{[K] \setminus \Omega} = \{s_k\}_{k \in [K] \setminus \Omega}$. By assumption, $\{\phi_m\}_{m=1}^M$ form a \mathcal{C} -system on \mathbb{I} with $M \geq 2K + 1$. Therefore, for every index set $\Omega \subseteq [K]$ and in view of Lemma 23, there exist polynomials q_{T_Ω} and $q_{S_{[K] \setminus \Omega}}$ that are non-negative on \mathbb{I} and vanish only on T_Ω and $S_{[K] \setminus \Omega}$, respectively.

Let us form the polynomial

$$Q(\theta) = Q(t, s) = \sum_{\Omega \subseteq [K]} q_{T_\Omega}(t) \cdot q_{S_{[K] \setminus \Omega}}(s), \quad (\text{A.1})$$

where the sum is over all subsets of $[K]$. Evidently, Q is non-negative on \mathbb{I}^2 since each summand above is non-negative. We next verify that Q only vanishes on Θ . 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 $q_{T_\Omega}(t_k) = 0$. Or $k \in [K] \setminus \Omega$, in which case $q_{S_{[K] \setminus \Omega}}(s_k) = 0$. In both cases, the product vanishes, i.e., $q_{T_\Omega}(t_k) \cdot q_{S_{[K] \setminus \Omega}}(s_k) = 0$. Since the choice of Ω was arbitrary, it follows from (A.1) 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 $\theta = (t, s) \in T^C \times S^C \subseteq \Theta^C$, i.e., $t \in T^C$ and $s \in S^C$. For arbitrary index set $\Omega \subseteq [K]$, note that $q_{T_\Omega}(t) \cdot q_{S_{[K] \setminus \Omega}}(s) > 0$ by design. It follows from (A.1) 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 $\Omega_0 \subset [K]$ such that $t_k \in [K] \setminus \Omega_0$ and $s_l \in \Omega_0$. For such Ω_0 , it holds that $q_{T_{\Omega_0}}(t_k) \cdot q_{S_{[K] \setminus \Omega_0}}(s_l) > 0$. For instance, one can choose $\Omega_0 = \{s_l\}$ for which both $q_{T_{\{s_l\}}}(t_k)$ and $q_{S_{[K] \setminus \{s_l\}}}(s_l)$ are strictly positive. Consequently, $Q(\theta) > 0$ by (A.1) when $\theta \in \Theta^C \setminus (T^C \times S^C)$. In conclusion, Q is positive and vanishes only on Θ , as claimed. This completes the proof of Lemma 16.

B. Proof of Lemma 18

For notational convenience, throughout we model $x_{K,e}$ by (2.1), with x therein replaced with $x_{K,e}$. By feasibility of both \hat{x} and $x_{K,e}$ for Program (1.5), and after applying the triangle inequality, we observe that

$$\left\| \int_{\mathbb{I}^2} \Phi(\theta) h(d\theta) \right\|_{\mathbb{F}} \leq 2\delta', \quad (\text{B.1})$$

where $\theta = (t, s)$. Next, the existence of the dual certificate allows us to write that

$$\begin{aligned}
 \bar{g} \int_{\Theta_\varepsilon^C} h(d\theta) &\leq \int_{\Theta_\varepsilon^C} G(\theta) h(d\theta) \quad ((4.11) \text{ and the error } h \text{ is nonnegative on } \Theta_\varepsilon^C) \\
 &= \int_{\Theta_\varepsilon^C} G(\theta) h(d\theta) + \sum_{k=1}^K \int_{\theta_{k,\varepsilon}} G(\theta) h(d\theta) \quad (\text{see (4.3, 4.11)}) \\
 &= \int_{\mathbb{I}^2} G(\theta) h(d\theta) \quad (\text{see (4.3)}) \\
 &= \sum_{m,n=1}^M b_{m,n} \int_{\mathbb{I}^2} \phi_m(t) \phi_n(s) h(dt, ds) \quad (\text{see (4.10)}) \\
 &= \left\langle b, \int_{\mathbb{I}^2} \Phi(\theta) h(d\theta) \right\rangle \quad (\text{see (1.4)}) \\
 &\leq \|b\|_F \cdot \left\| \int_{\mathbb{I}^2} \Phi(\theta) h(d\theta) \right\|_F \quad (\text{Cauchy-Schwarz inequality}) \\
 &\leq \|b\|_F \cdot 2\delta', \quad (\text{see (B.1)})
 \end{aligned} \tag{B.2}$$

which completes the proof of Lemma 18. Above, we note that the matrix $b \in \mathbb{R}^{M \times M}$ is formed by the coefficients $\{b_{m,n}\}_{m,n=1}^M$.

C. Proof of Lemma 19

For notational convenience, throughout we model $x_{K,\varepsilon}$ by (2.1), with x therein replaced with $x_{K,\varepsilon}$. The existence of the dual certificate Q^0 allows us to write that

$$\begin{aligned}
 &\sum_{k=1}^K \left| \int_{\theta_{k,\varepsilon}} h(d\theta) \right| \\
 &= \sum_{k=1}^K \int_{\theta_{k,\varepsilon}} s_k h(d\theta) \quad \left(s_k := \text{sign} \left(\int_{\theta_{k,\varepsilon}} h(d\theta) \right) \right) \\
 &= \sum_{k=1}^K \int_{\theta_{k,\varepsilon}} (s_k - Q^0(\theta)) h(d\theta) + \sum_{k=1}^K \int_{\theta_{k,\varepsilon}} Q^0(\theta) h(d\theta) \\
 &= \sum_{k=1}^K \int_{\theta_{k,\varepsilon}} (s_k - Q^0(\theta)) h(d\theta) + \int_{\mathbb{I}^2} Q^0(\theta) h(d\theta) - \int_{\Theta_\varepsilon^C} Q^0(\theta) h(d\theta) \quad (\text{see (4.3)}) \\
 &= \sum_{s_k=1} \int_{\theta_{k,\varepsilon}} (1 - Q^0(\theta)) h(d\theta) + \sum_{s_k=-1} \int_{\theta_{k,\varepsilon}} (-1 - Q^0(\theta)) h(d\theta) \\
 &\quad + \int_{\mathbb{I}^2} Q^0(\theta) h(d\theta) - \int_{\Theta_\varepsilon^C} Q^0(\theta) h(d\theta)
 \end{aligned}$$

$$\begin{aligned}
&= \sum_{s_k=1} \int_{\theta_{k,\varepsilon}} (1 - Q^0(\theta)) h(d\theta) + \sum_{s_k=-1} \int_{\theta_{k,\varepsilon} \setminus \theta_k} (-1 - Q^0(\theta)) h(d\theta) \\
&\quad + \sum_{s_k=-1} \int_{\theta_k} (-1 + \alpha - Q^0(\theta)) h(d\theta) - \alpha \sum_{s_k=-1} \int_{\theta_k} h(d\theta) \\
&\quad + \int_{\mathbb{I}^2} Q^0(\theta) h(d\theta) - \int_{\Theta_\varepsilon^C} Q^0(\theta) h(d\theta). \tag{C.1}
\end{aligned}$$

Above, we separated the impulses based on the sign of the error near the impulses, and we also singled out the errors at the impulse locations corresponding to the negative sign. In view of (4.14), we can bound the last line above by

$$\begin{aligned}
& -\alpha \sum_{s_k=-1} \int_{\theta_k} h(d\theta) + \int_{\mathbb{I}^2} Q^0(\theta) h(d\theta) + \int_{\Theta_\varepsilon^C} h(d\theta) \quad (\text{see (4.14)}) \\
& \leq -\alpha \sum_{s_k=-1} \int_{\theta_k} h(d\theta) + \int_{\mathbb{I}^2} Q^0(\theta) h(d\theta) + 2\bar{g}^{-1} \|b\|_{\mathbb{F}} \delta' \quad (\text{see Lemma 18}) \\
& \leq \alpha \sum_{s_k=-1} \int_{\theta_k} x_{K,\varepsilon}(d\theta) + \int_{\mathbb{I}^2} Q^0(\theta) h(d\theta) + 2\bar{g}^{-1} \|b\|_{\mathbb{F}} \delta' \quad (h = \hat{x} - x_{K,\varepsilon} \text{ and } \hat{x} \geq 0) \\
& = \alpha \sum_{s_k=-1} a_k + \int_{\mathbb{I}^2} Q^0(\theta) h(d\theta) + 2\bar{g}^{-1} \|b\|_{\mathbb{F}} \delta' \quad (\text{see (2.1)}) \\
& \leq \alpha \|x_{K,\varepsilon}\|_{\text{TV}} + \int_{\mathbb{I}^2} Q^0(\theta) h(d\theta) + 2\bar{g}^{-1} \|b\|_{\mathbb{F}} \delta' \\
& = \alpha \|x_{K,\varepsilon}\|_{\text{TV}} + \left\langle b^0, \int_{\mathbb{I}^2} \Phi(\theta) h(d\theta) \right\rangle + 2\bar{g}^{-1} \|b\|_{\mathbb{F}} \delta' \quad (\text{see (1.4)}) \\
& \leq \alpha \|x_{K,\varepsilon}\|_{\text{TV}} + \|b^0\|_{\mathbb{F}} \cdot 2\delta' + 2\bar{g}^{-1} \|b\|_{\mathbb{F}} \delta', \quad (\text{Cauchy-Schwarz and B.1}) \tag{C.2}
\end{aligned}$$

which completes the proof of Lemma 19. Above, the matrix $b^0 \in \mathbb{R}^{M \times M}$ is formed by the coefficients $\{b_{m,n}^0\}_{m,n=1}^M$.

D. Proof of Proposition 21

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 21, which is summarized below for the convenience of the reader; see [1, Proposition 19].⁶

⁶ In the proof of [1, Proposition 19] and with the notation therein, F must be such that $\{F\} \cup \{\phi_j\}_{j=1}^m$ form a \mathcal{C}^* -system, but is otherwise arbitrary. Our Proposition 24 thus follows from [1, Proposition 19], after ① recalling the first property of \mathcal{C}^* -systems listed in Remark 10 and ② noting that the sum in the definition of \dot{q} in the proof of Proposition 19 is a continuous function of t because of the continuity of the functions $\{\phi_j\}_{j=1}^m$.

PROPOSITION 24 (Univariate polynomial of a \mathcal{C}^* -system). Consider a finite set $T' \subset \mathbb{I}$ of size no larger than K . For $M \geq 2K + 2$, suppose that $\{\phi_m\}_{m=1}^M$ form a \mathcal{C} -system on \mathbb{I} . Consider also $F' : \mathbb{R} \rightarrow \mathbb{R}$ and suppose that $\{F'\} \cup \{\phi_m\}_{m=1}^M$ form a $\mathcal{C}_{K,\varepsilon}^*$ -system on \mathbb{I} . Then, there exist real coefficients $\{b_m\}_{m=1}^M$ and a continuous polynomial $q_{T'} = \sum_{m=1}^M b_m \phi_m$ such that $q_{T'} \geq F'$ with equality holding on T' .

Let us now use Proposition 24 to complete the proof of Proposition 21. Fix an index set $\Omega \subseteq [K]$. By assumption, $\{F_{T_\Omega}\} \cup \{\phi_m\}_{m=1}^M$ form a $\mathcal{C}_{K,\varepsilon}^*$ -system on \mathbb{I} . Therefore, by Proposition 24, there exists a polynomial q_{T_Ω} such that

$$q_{T_\Omega} \geq F_{T_\Omega}, \quad (\text{D.1})$$

with equality holding on T_Ω . Likewise, by assumption, $\{F_{S_{[K]\setminus\Omega}}\} \cup \{\phi_m\}_{m=1}^M$ form a $\mathcal{C}_{K,\varepsilon}^*$ -system on \mathbb{I} and therefore there exists a polynomial $q_{S_{[K]\setminus\Omega}}$ such that

$$q_{S_{[K]\setminus\Omega}} \geq F_{S_{[K]\setminus\Omega}}, \quad (\text{D.2})$$

with equality holding on $S_{[K]\setminus\Omega}$.

As in the proof of Proposition 16, consider the polynomial

$$Q(\theta) = Q(t, s) = \sum_{\Omega \subseteq [K]} q_{T_\Omega}(t) \cdot q_{S_{[K]\setminus\Omega}}(s), \quad (\text{D.3})$$

where the sum is over all subsets of $[K]$. We next show that Q is the desired dual certificate, prescribed in Lemma 16. Recall the neighborhoods defined in (4.1, 4.3). Fix an index set $\Omega \subseteq [K]$ and $k \in [K]$. Assume that $k \in \Omega$. For every $\theta \in \theta_{k,\varepsilon} = t_{k,\varepsilon} \times s_{k,\varepsilon}$, it then holds that

$$\begin{aligned} q_{T_\Omega}(t) \cdot q_{S_{[K]\setminus\Omega}}(s) + q_{T_{[K]\setminus\Omega}}(t) \cdot q_{S_\Omega}(s) &\geq F_{T_\Omega}(t) \cdot F_{S_{[K]\setminus\Omega}}(s) + F_{T_{[K]\setminus\Omega}}(t) \cdot F_{S_\Omega}(s) \quad (\text{see (D.1, D.2)}) \\ &\geq 0 \cdot 1 + 1 \cdot 0 = 0, \quad (\text{see (4.17)}) \end{aligned}$$

and the equality in the first line above holds at least at $\theta_k = (t_k, s_k)$. In fact, the above statement holds also when $k \in [K] \setminus \Omega$. By summing up over all pairs $(\Omega, [K] \setminus \Omega)$ and then applying the above inequality, we arrive at

$$Q(\theta) = \sum_{\Omega \subseteq [K]} q_{T_\Omega}(t) \cdot q_{S_{[K]\setminus\Omega}}(s) \geq 0, \quad (\text{D.4})$$

which, to reiterate, holds for every $\theta \in \theta_{k,\varepsilon}$ and with equality at θ_k . The above bound is independent of k and we therefore conclude that

$$Q(\theta) \geq 0, \quad \theta \in \Theta_\varepsilon, \quad (\text{D.5})$$

with equality holding at Θ ; see (4.3).

On the other hand, consider $\theta \in \Theta_\varepsilon^C$, i.e., θ does not belong to any of the neighborhoods $\{\theta_{k,\varepsilon}\}_{k=1}^K$. We consider two cases below.

1. The first possibility is that

$$\theta = (t, s) \in T_\varepsilon^C \times S_\varepsilon^C \subseteq \Theta_\varepsilon^C.$$

In this case, in view of (D.1, D.2), it holds that

$$q_{T_\Omega}(t) \cdot q_{S_{[K]\setminus\Omega}}(s) \geq F_{T_\Omega}(t) \cdot F_{S_{[K]\setminus\Omega}}(s) \geq 1, \quad (\text{see (4.17)}) \quad (\text{D.6})$$

for every index set $\Omega \subseteq [K]$. By summing up over all index sets, it immediately follows that

$$\begin{aligned} Q(\theta) &= \sum_{\Omega \subseteq [K]} q_{T_\Omega}(t) \cdot q_{S_{[K] \setminus \Omega}}(s) \\ &\geq 2^K, \quad \theta \in T_\varepsilon^C \times S_\varepsilon^C. \end{aligned} \quad (\text{D.7})$$

2. The second possibility is that $\theta \in \Theta_\varepsilon^C \setminus (T_\varepsilon^C \times S_\varepsilon^C)$. In this case, there exists a distinct pair $k, l \in [K]$ such that $\theta = (t, s) \in t_{k, \varepsilon} \times s_{l, \varepsilon}$ and an index set $\Omega_0 \subset [K]$ such that $t_k \in [K] \setminus \Omega_0$ and $s_l \in \Omega_0$. It follows that

$$q_{T_{\Omega_0}}(t) \cdot q_{S_{[K] \setminus \Omega_0}}(s) \geq F_{T_{\Omega_0}}(t) \cdot F_{S_{[K] \setminus \Omega_0}}(s) \geq 1. \quad (\text{see (4.17)}) \quad (\text{D.8})$$

There are in fact 2^{K-2} such subsets of $[K]$ and we conclude that

$$\begin{aligned} Q(\theta) &= \sum_{\Omega \subseteq [K]} q_{T_\Omega}(t) \cdot q_{S_{[K] \setminus \Omega}}(s) \\ &\geq 2^{K-2}, \quad \theta \in \Theta_\varepsilon^C \setminus (T_\varepsilon^C \times S_\varepsilon^C). \end{aligned} \quad (\text{D.9})$$

By combining (D.7) and (D.9), we reach that

$$Q(\theta) \geq 2^{K-2} =: \bar{g}, \quad \theta \in \Theta_\varepsilon^C. \quad (\text{D.10})$$

Lastly, combining (D.5) and (D.10) completes the proof of Proposition 21.

E. Proof of Proposition 22

The proof is based on the same principles that appeared in Appendix D. Let us fix an arbitrary sign pattern $\pi \in \{\pm 1\}_{k=1}^K$. For every $k \in [K]$ and by assumption, $\{F_{s_k}^+\} \cup \{\phi_m\}_{m=1}^M$ form a $\mathcal{C}_{K, \varepsilon}^*$ -system on \mathbb{I} , see (4.19). Therefore, by Proposition 24, there exists for every $k \in [K]$ a polynomial $q_{s_k}^{\pi_k}$ such that

$$q_{s_k}^{\pi_k}(s) \geq \begin{cases} F_{s_k}^+(s) & \text{when } \pi_k = 1 \\ \varepsilon F_{s_k}^+(s) & \text{when } \pi_k = -1, \end{cases} \quad (\text{E.1})$$

for every $s \in \mathbb{I}$, with equality holding on $S = \{s_i\}_{i=1}^K$. When the sign pattern π contains at least one negative sign, we define for future use the normalized maximum

$$q_{\max}^\pi(\varepsilon) := \varepsilon^{-1} \max_{\pi_k = -1} \max_{s \in \mathbb{I}} q_{s_k}^{\pi_k}(s), \quad (\text{E.2})$$

where the inner maximum above is indeed achieved in view of the continuity of $q_{s_k}^{\pi_k}$ and the compactness of \mathbb{I} ; see Proposition 24. Note that $q_{\max}^\pi(\varepsilon) \geq 1$ for every $\varepsilon > 0$; see (4.19, E.1). When $\varepsilon = 0$, we choose the trivial polynomial $q_{s_k}^{\pi_k} = \varepsilon = 0$ for every k such that $\pi_k = -1$, and thus record that

$$q_{\max}^\pi(0) = 1. \quad (\text{E.3})$$

Likewise, for every $k \in [K]$, $\{F_{t_k}^{\pi_k}\} \cup \{\phi_m\}_{m=1}^M$ form a $\mathcal{C}_{K,\varepsilon}^*$ -system on \mathbb{I} by assumption; see (4.19). Therefore, by Proposition 24, there exists for every $k \in [K]$ a polynomial $q_{t_k}^{\pi_k}$ such that

$$q_{t_k}^{\pi_k}(t) \geq \begin{cases} F_{t_k}^{\pi_k}(t) & \text{when } \pi_k = 1 \\ \frac{F_{t_k}^{\pi_k}(t)}{\varepsilon q_{\max}^{\pi_k}(\varepsilon)} & \text{when } \pi_k = -1, \end{cases} \quad (\text{E.4})$$

for every $t \in \mathbb{I}$, with equality holding on $T = \{t_i\}_{i=1}^K$. In view of (E.1, E.4), and after recalling the definitions of $F_{s_k}^+$ and $F_{t_k}^\pm$ from (4.19), we observe that the product $q_{t_k}^{\pi_k}(t)q_{s_k}^{\pi_k}(s)$ satisfies

$$q_{t_k}^{\pi_k}(t)q_{s_k}^{\pi_k}(s) \geq \begin{cases} \pi_k & \text{when } \theta \in \theta_{k,\varepsilon} \\ -\frac{1}{q_{\max}^{\pi_k}(\varepsilon)} & \text{when } \theta = \theta_k \text{ and } \pi_k = -1 \\ -1 & \text{when } t \in t_{k,\varepsilon} \text{ and } \pi_k = -1 \\ 0 & \text{elsewhere in } \mathbb{I}^2, \end{cases} \quad (\text{E.5})$$

with equality holding at least on Θ . The third case above indicates that there is a stripe in \mathbb{I}^2 over which the product $q_{t_k}^{\pi_k}q_{s_k}^{\pi_k}$ is bounded below by -1 . Let us now consider the polynomial

$$Q^\pi(\theta) = Q^\pi(t, s) := \sum_{k=1}^K q_{t_k}^{\pi_k}(t) \cdot q_{s_k}^{\pi_k}(s).$$

We next establish that, for the appropriate choice of the sign pattern π , Q^π is indeed the desired dual certificate prescribed in Lemma 19. Recall that impulses are ε -separated and that, in particular, (4.2) holds for the choice of $\nu = \varepsilon$ therein. Then, we invoke (E.5) to write⁵ that

$$\begin{aligned} Q^\pi(\theta) &= \sum_{k=1}^K q_{t_k}^{\pi_k}(t) \cdot q_{s_k}^{\pi_k}(s) \\ &\geq \begin{cases} \pi_k & \text{when there exists } k \in [K] \text{ such that } \theta \in \theta_{k,\varepsilon} \\ -\frac{1}{q_{\max}^{\pi_k}(\varepsilon)} & \text{when there exists } k \in [K] \text{ such that } \theta = \theta_k \text{ and } \pi_k = -1 \\ -1 & \text{elsewhere in } \mathbb{I}^2, \end{cases} \end{aligned} \quad (\text{E.6})$$

and the equality holds at least on Θ . Finally, since our choice of the sign pattern π in the beginning of the proof was arbitrary, the existence of the dual certificate in Lemma 19 is also guaranteed, even though the sign pattern of the error near the point sources is unknown to us a priori. More specifically, let π_* denote the sign pattern specified by the error measure h in (4.14). Then, the dual certificate in Lemma 19 exists with

$$\alpha(\varepsilon) = 1 - 1/q_{\max}^{\pi_*}(\varepsilon). \quad (\text{E.7})$$

In particular, $\alpha(0) = 0$ by (E.3). This completes the proof of Proposition 22.