On the Stable Resolution Limit of Total Variation Regularization for Spike Deconvolution

Maxime Ferreira Da Costa, *Member, IEEE*, and Yuejie Chi, *Senior Member, IEEE*

Abstract—The stability of spike deconvolution, which aims at recovering point sources from their convolution with a point spread function (PSF), is known to be related to the separation between those sources. When the observations are noisy, it is critical to ensure *support stability*, where the deconvolution does not lead to spurious, or oppositely, missing estimates of the point sources. In this paper, we study the resolution limit of stably recovering the support of two closely located point sources using the Beurling-LASSO estimator, which is a convex optimization approach based on total variation regularization. We establish a sufficient separation criterion between the sources, depending only on the PSF, above which the Beurling-LASSO estimator is guaranteed to return a stable estimate of the point sources, with the same number of estimated elements as that of the ground truth. Our result highlights the impact of PSF on the resolution limit in the noisy setting, which was not evident in previous studies of the noiseless setting. Towards the end, we show that the same resolution limit applies to resolving two close-located sources in conjunction of other well-separated sources.

Index Terms—Spectral estimation, super-resolution, resolution limits, convex optimization.

I. INTRODUCTION

IN ITS classical formulation, the super-resolution problem consists of recovering a stream of localized temporal events, modeled as one-dimensional point sources (or spikes), characterized by their positions and amplitudes, from distorted and noisy observations. This problem finds a myriad of applications in applied and experimental sciences, such as spectrum and modal analysis, radar, sonar, optical imaging, wireless communications and sensing systems. The distortion is often characterized by a shift-invariant point spread function (PSF), acting as a low-pass band-limited filter, on the stream of spikes to recover, in accordance to the physical limitation of the measurement device involved in the acquisition of the point sources [1].

The problem, known as spike deconvolution, comes with a handful of statistical challenges. Of particular interest to

this paper is the *support stability* of the reconstruction in the presence of additive noise, defined as the capability of a given estimator to return the exact same number of point sources as that of the ground truth, without spurious or missing elements. This paper studies the support stability of the Beurling-LASSO estimator [2] to reconstruct two closely located point sources. Despite its apparent simplicity, this setup is of importance both in theory and in practice. In theory, it allows us to develop a deeper insight on the fundamental notion of *resolution limit* – the minimal distance above which two point sources are said to be distinguishable. In practice, it models the separation of a weak moving target from a strong clutter in radar [3], and accurate counting of the number of molecules in super-resolution fluorescence microscopy [4].

The Beurling-LASSO estimator is a convex optimization approach with the total variation (TV) regularization, which has been shown to provide exact reconstruction of the point sources in the absence of noise, whenever the point sources are sufficiently separated [2], [5], [6]. The TV regularization, applied to measures, can be regarded as a continuous analog of the standard ℓ_1 regularization for finite-dimensional vectors, but is advantageous by overcoming the basis mismatch issue [7]. In this paper, we show that the Beurling-LASSO estimator can also stably reconstruct the support of the two close-located point sources provided that they are separated by a distance that can be computed exactly using a formula depending only on the PSF, revealing the impact of PSF on the stability of spike deconvolution. Our result can be extended to a multi-source setting containing a mixture of two close-located sources and other well-separated sources.

A. Observation Model

We consider a scenario where there are only two point sources to recover. Denoting by $\mathcal{M}(\mathbb{R})$ the set of complex Radon measures over the reals, the signal to resolve is modeled as a measure $\nu_\star \in \mathcal{M}(\mathbb{R})$ of the form

$$\nu_\star(\tau) = c_1 \delta(\tau - \tau_1) + c_2 \delta(\tau - \tau_2), \quad (1)$$

where $\delta(\cdot)$ is the Dirac measure, $\tau_1, \tau_2 \in \mathbb{R}$ are the time-domain locations of the two spikes and $c_1, c_2 \in \mathbb{C} \backslash \{0\}$ are their non-zero associated complex amplitudes. The continuous-time signal $x(\tau)$ resulted from the convolution of the ground truth measure $\nu_\star(\tau)$ with the PSF $g(\tau)$ writes as

$$\begin{aligned} x(\tau) &= (g * \nu_\star)(\tau) \\ &= c_1 \, g(\tau - \tau_1) + c_2 \, g(\tau - \tau_2), \quad \forall \tau \in \mathbb{R}, \quad (2) \end{aligned}$$

where $*$ denotes linear convolution. Furthermore, because of the needs of digital processing, one typically takes discrete-time measurements. An idealistic, yet credible approximation of many super-resolution problems encountered in practice is to consider measurements drawn from uniform sampling of the Fourier transform of $x(\tau)$. Let $\mathcal{F}(\cdot)$ be the Fourier transform of a measure in $\mathcal{M}(\mathbb{R})$, defined as

$$\mathcal{F}(\mu)(f) = \int_{\mathbb{R}} e^{-i2\pi f\tau} \mathrm{d}\mu(\tau), \quad \forall \mu \in \mathcal{M}(\mathbb{R}), \ \forall f \in \mathbb{R}, \ \text{a.e.}.$$

The Fourier-domain counterpart of the observation model (2) becomes

$$X(f) = G(f) \cdot \mathcal{F}(\nu_\star)(f), \quad \forall f \in \mathbb{R}, \ \text{a.e.},$$

where $X = \mathcal{F}(x)$, $G = \mathcal{F}(g)$ are the Fourier transforms of the signal $x(\tau)$ and the PSF $g(\tau)$, respectively. We assume that the PSF $g(\tau)$ is band-limited, with a bandwidth of $B > 0$. Therefore, $G(f) = 0$ for every f outside the interval $\left(-\frac{B}{2}, \frac{B}{2}\right)$. We further assume an odd number $N = 2n + 1$ of measurements[1] are taken uniformly over the bandwidth $\left(-\frac{B}{2}, \frac{B}{2}\right)$. Therefore, the observation vector is given by $\boldsymbol{x} = \{x_k = X(kB/N)\}_{k=-n}^{n} \in \mathbb{C}^N$, corresponding to measuring $X(f)$ at frequencies $\{kB/N\}_{k=-n}^{n} \subset \left(-\frac{B}{2}, \frac{B}{2}\right)$.

For convenience, we introduce a normalized measure $\mu_\star \in \mathcal{M}(\mathbb{R})$ as $\mu_\star(t) = \frac{N}{B}\nu_\star(Nt/B)$ for all $t \in \mathbb{R}$, which by combining with (1) can be rewritten as,

$$\mu_\star(t) = c_1\delta(t - t_1) + c_2\delta(t - t_2), \tag{3}$$

where $t_1 = B\tau_1/N$ and $t_2 = B\tau_2/N$ are the normalized locations of the point sources. The observations $\boldsymbol{x}$ are linked to $\mu_\star$ by the linear relation

$$\boldsymbol{x} = \Phi_g(\mu_\star). \tag{4}$$

Here, the measurement operator $\Phi_g : \mathcal{M}(\mathbb{R}) \mapsto \mathbb{C}^N$ maps every $\mu \in \mathcal{M}(\mathbb{R})$ to the vector $\Phi_g(\mu) \in \mathbb{C}^N$ defined by

$$\Phi_g(\mu) = \mathrm{diag}(\boldsymbol{g}) \left[\mathcal{F}(\mu)(-n), \mathcal{F}(\mu)(-n+1), \dots, \mathcal{F}(\mu)(n)\right]^\top, \tag{5}$$

where $\boldsymbol{g} = \{g_k = G(kB/N)\}_{k=-n}^{n} \in \mathbb{C}^N$ is the vector obtained by sampling the Fourier transform of the PSF $g(\tau)$ at frequencies $\{kB/N\}_{k=-n}^{n}$. Furthermore, notice that the observation operator Φ_g is invariant with respect to integer shifts of the underlying measure $\mu_\star$. Thus, one can only hope to identify $\mu_\star$ over the set of Radon measure defined over the torus $\mathbb{T} \sim \mathbb{R}/\mathbb{Z}$, denoted as $\mathcal{M}(\mathbb{T})$. Without loss of generality, the delays t_1, t_2 are normalized within the unit interval, i.e. $t_1, t_2 \in [-\frac{1}{2}, \frac{1}{2})$.[2]

In the presence of noise or measurement errors, we assume $\boldsymbol{x}$ is corrupted by an additive term $\boldsymbol{w}$. The observations are given as

$$\boldsymbol{z} = \boldsymbol{x} + \boldsymbol{w} = \Phi_g(\mu_\star) + \boldsymbol{w}, \tag{6}$$

---

[1]An odd number of measurements is considered only for clarity and simplification purposes, and does not affect the generality of the results presented in this paper.

[2]Since $t_i = B\tau_i/N$, $i = 1, 2$, and assuming $\tau_i \in [-T/2, T/2)$, where T is the time window of interest, then the ambiguity constraint $t_i \in [-1/2, 1/2)$ suggests that the number of measurements should be greater than the time-bandwidth product, i.e. $N \geq T \cdot B$, to avoid aliasing.

where $\|\boldsymbol{w}\|_2 \leq \eta$ is assumed to be bounded for some noise level $\eta > 0$.

B. Reconstruction Using the Total Variation Minimization Framework

In the absence of noise, the super-resolution problem is defined as recovering $\mu_\star$ from the observations $\boldsymbol{x}$ and the PSF $g(\tau)$, yielding a linear inverse problem over the set of measures. Clearly, there are many possible measures that lead to the same observations, making the problem ill-posed. It is therefore necessary to impose structures on the measure of interest, where one of the most widely used structures is a sparsity prior. More precisely, one seeks for the measure $\widehat{\mu}_\natural$ with minimal support that is consistent with the observations $\boldsymbol{x}$ given in (4). Denoting by $\|\cdot\|_0$ the "pseudo-norm" counting the potentially infinite cardinality of the support of a measure in $\mathcal{M}(\mathbb{T})$, the optimal estimator $\widehat{\mu}_\natural$ for the super-resolution problem can be reformulated as the output of the optimization program

$$\widehat{\mu}_\natural = \arg\min_{\mu \in \mathcal{M}(\mathbb{T})} \|\mu\|_0 \quad \text{s.t.} \quad \boldsymbol{x} = \Phi_g(\mu), \tag{7}$$

which is known to be unique and equal to the ground truth $\mu_\star$ as long as the number of measurements N is at least twice as large as the number of spikes to recover [8].

However, on the computational front, the estimator (7) is infeasible due to the combinatorial aspects inherent to the definition of $\|\cdot\|_0$. Instead, a convex relaxation of the estimator (7) is proposed in [2] to recover the measure. This is achieved by relaxing the cost function by a convex surrogate known as total variation (TV), denoted as $\|\cdot\|_{\mathrm{TV}}$, whose formal definition will be discussed later. The total variation minimization of measures, equivalent to the atomic norm [9], [10], is a versatile framework that can be adapted to solve a variety of linear inverse problems over continuous dictionaries. The resulting TV estimator, denoted as $\widehat{\mu}_0$, is given by

$$\widehat{\mu}_0 = \arg\min_{\mu \in \mathcal{M}(\mathbb{T})} \|\mu\|_{\mathrm{TV}} \quad \text{s.t.} \quad \boldsymbol{x} = \Phi_g(\mu), \tag{8}$$

which is a convex program over the set of Radon measures, and can be computed efficiently by solving an associated semidefinite program (see e.g. [11]). In the presence of noisy observations of the form (6), the Beurling-LASSO estimator $\widehat{\mu}_\lambda$ [12], also known as the atomic norm denoiser [11], can be used to recover the ground truth. It can be understood as an extension of the celebrated LASSO estimator over the set of measures, which aims to estimate the ground truth measure by minimizing a sum of the TV norm of the measure and the squared Euclidean norm of the measurement residual (4), so that the estimate $\widehat{\mu}_\lambda$ is written as

$$\widehat{\mu}_\lambda = \widehat{\mu}_\lambda(\boldsymbol{z}) = \arg\min_{\mu \in \mathcal{M}(\mathbb{T})} \frac{1}{2} \|\boldsymbol{z} - \Phi_g(\mu)\|_2^2 + \lambda \|\mu\|_{\mathrm{TV}}, \tag{9}$$

where $\lambda > 0$ is a regularization parameter drawing a trade-off between the TV norm of the estimate, as well as its fidelity to the observations.
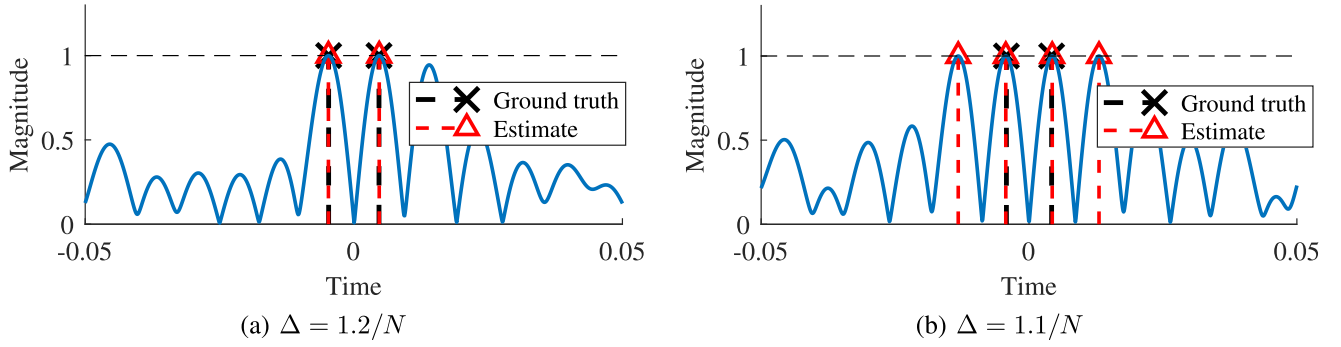
Fig. 1. Illustration of support stability of the Beurling-LASSO estimator for reconstructing two point sources with different separations Δ when the PSF is the ideal low-pass filter $g(\tau) = \mathrm{sinc}(\pi\tau)$ and for a number of samples $N = 129$ with SNR $= 40$dB. The locations of point sources are estimated as the peaks of the magnitude of the dual polynomial $\Phi_g^*(\widehat{\boldsymbol{p}}_\lambda)$, where $\widehat{\boldsymbol{p}}_\lambda$ is the solution of (14). (a): $\Delta = 1.2/N$, the estimator returns exactly two spikes closely located to the ground truth and is support stable. (b): $\Delta = 1.1/N$, the estimator returns two additional spurious spikes and is not support stable.

C. Resolution Limit of Spike Deconvolution

An important question for practical operations is the resolution of spike deconvolution, where one would like to ensure that the reconstructed measure is as close as possible to the ground truth. It has been known for decades that the separation between the spikes,

$$\Delta = |t_2 - t_1|_{\mathbb{T}} \triangleq \inf_{\ell \in \mathbb{Z}} |t_2 - t_1 + \ell|, \qquad (10)$$

which measures the distance over the torus $\mathbb{T}$, plays an important role – the smaller the separation, the more challenging it is to resolve them. For example, the Rayleigh limit (see e.g. [13]) is a classical *empirical* criterion to characterize the *resolution limit* the super-resolution problem, i.e., the minimal separation between two point sources, above which those sources are said to be distinguishable. In recent years, there has been a resurge of interest in a formal characterization of this limit - both in terms of achievability and impossibility. In particular, stability has been shown to be related to the asymptotic behaviors of the condition number of Vandermonde matrices with nodes on the unit circle [14]–[16], which diverges below a critical separation of the spikes. This phase transition induces the existence of a resolution limit under which point sources cannot be resolved in the presence of noise in the asymptotic regime where N tends to infinity [17], regardless of the algorithm used for reconstruction.

In this paper, we are interested in understanding the robustness of an estimator in the presence of noise. Among the many figures of merit to quantify this robustness, an important criterion is the *support stability* of the estimator, defined below when specialized to the two-spike setting.

Definition 1 (Support Stability): Consider the observations $\boldsymbol{z} = \Phi_g(\mu_\star) + \boldsymbol{w}$. An estimator $\widehat{\mu} = \widehat{\mu}(\boldsymbol{z})$ based on $\boldsymbol{z}$ is said to be support stable for a given ground truth measure $\mu_\star$ of the form (3) if there exists $\eta > 0$ such that for all $\boldsymbol{w}$ with $\|\boldsymbol{w}\|_2 < \eta$, the estimate $\widehat{\mu}$ is a measure containing two spikes, i.e.

$$\widehat{\mu}(\boldsymbol{z}) = \widehat{c}_1 \delta(t - \widehat{t}_1) + \widehat{c}_2 \delta(t - \widehat{t}_2),$$

and if the estimated parameters satisfy, up to a permutation Π of the indices: $\left|t_k - \widehat{t}_{\Pi(s)}\right|_{\mathbb{T}} = \mathcal{O}(\|\boldsymbol{w}\|_2)$ and $\left|c_s - \widehat{c}_{\Pi(s)}\right| = \mathcal{O}(\|\boldsymbol{w}\|_2)$ for $s = 1, 2$ in the limit of $\|\boldsymbol{w}\|_2 \to 0$.

This notion, introduced in [6], characterizes the capability of an estimator to output a measure containing the exact same number of spikes as that of the ground truth, when the signal-to-noise ratio (SNR) is large enough. As an example, Fig. 1 plots the reconstruction of a ground truth measure containing two spikes using the Beurling-LASSO estimator at SNR $= 40$dB under different separations when the PSF is the ideal low-pass filter. In this illustration, when $\Delta = 1.2/N$, the estimator returns exactly two spikes closely located to the ground truth; on the other hand, when $\Delta = 1.1/N$, the estimator returns additional spurious spikes that are not consistent with the ground truth, and therefore, is no longer support stable.

D. Overview of the Main Result

This paper studies the support stability of the Beurling-LASSO estimator (9) for the reconstruction of a two-spike measure of the form (1). We show that, the Beurling-LASSO estimator is support stable if the separation Δ is greater than $\gamma^\star/N$, which can be calculated exactly using the knowledge of the PSF $g(t)$. Our main contribution can be informally summarized in the following statement.

Theorem 2 (main Result, Informal Statement): Suppose that the PSF $g(\tau)$ satisfies some mild regularity conditions and is band-limited within $\left(-\frac{B}{2}, \frac{B}{2}\right)$. There exists a constant $\gamma^\star$, depending only on $g(\tau)$, such that if a measure $\mu_\star$ of the form (3) verifies
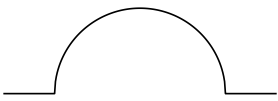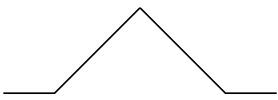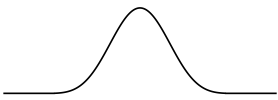
$$\Delta := |t_2 - t_1|_{\mathbb{T}} > \frac{\gamma^\star}{N},$$

then the Beurling-LASSO estimator is support stable when N is sufficiently large.

The complete statement of the theorem (c.f. Theorem 6) provides the formula of the stable resolution limit $\gamma^\star$, which can be computed for an arbitrary PSF verifying the hypotheses of the theorem. One highlight of our result is that it links the stable resolution limit directly to the PSF, which is not apparent in the study of the noise-free setting, where the resolution limit of exact recovery is independent of the PSF. This provides a quantitative means to evaluate and compare the choices of different PSFs in imaging and sensing applications.

TABLE I

VALUES OF THE MINIMAL SEPARATION $\gamma^\star$ FOR COMMONLY ENCOUNTERED POINT SPREAD FUNCTIONS. HEREIN, $J_0(\cdot)$ DENOTES THE BESSEL FUNCTION OF THE FIRST KIND, AND $\psi_{\tau_0}(\cdot)$ REFERS TO THE PROLATE SPHEROIDAL WAVE FUNCTION (PSWF) OF ORDER 0 FOR THE TEMPORAL CONCENTRATION BAND $[-\tau_0, \tau_0]$ [18], [19], I.E., THE FUNCTION $g(\cdot)$ WITH A FREQUENCY BAND $\left(-\frac{1}{2}, \frac{1}{2}\right)$ AND WITH $\|g\|_{L_2} = 1$ WHICH MAXIMIZES THE INTEGRAL $\int_{-\tau_0}^{\tau_0} |g(\tau)|^2 d\tau$

Point spread function	Fourier transform	$\gamma^\star$
Ideal low-pass: $\operatorname{sinc}(\pi\tau)$		1.132
Circular low-pass: $J_0(\pi\tau)/\sqrt{\pi\tau}$		1.253
Triangular low-pass: $\operatorname{sinc}(\pi\tau/2)^2$		1.449
Truncated Gaussian: $e^{-\frac{\tau^2}{2\sigma^2}} * \operatorname{sinc}(\pi\tau)$		(see Fig. 2 (a))
Prolate spheroidal wave function: $\psi_{\tau_0}(\tau)$		(see Fig. 2 (b))

The spectrum of a typical PSF has a decaying shape. A slower decay usually is associated with a smaller resolution limit and better super resolution capabilities for PSFs with the same bandwidth.

For illustration purposes, Table I lists the approximate values of $\gamma_\star$ associated with commonly encountered PSFs, such as ideal low-pass filters, circular low-pass filters, triangular low-pass filters, truncated Gaussian functions, and prolate spheroidal wave functions. Since we focus on the case of only two spikes, the separation condition $\gamma^\star$ is smaller than those in [5], [20], [21] for the noise-free setting, which allows more spikes. In addition, Fig. 2 illustrates how the stable resolution limit $\gamma^\star$ increases while the temporal concentration of the truncated Gaussian function and the prolate spheroidal wave function degenerates. Finally, Fig. 3 compares the constant $\gamma^\star$ predicted by Theorem 2 with the empirical success rates of the Beurling-LASSO estimator for different PSFs, which corroborates the findings of our theory. Our result holds even in the presence of additional point sources, as long as they are well-separated, see Section V.

E. Connections to Related Works

The resolution limits of spike deconvolution have been studied extensively, including but not limited to [22]–[27]. The performance of the TV estimator (8) has been studied in the noiseless setting with respect to the separation condition [2],

[6], [21]. Exact recovery of the TV estimator (8) is first guaranteed in [2], for an arbitrary number of spikes, given a separation $\Delta > 4/N$ under the proviso that the number of observations N is large enough, which has been later improved to $\Delta \geq 2.56/N$ in [28]. On the other hand, it is known that TV-regularization can fail whenever $\Delta < 2/N$ [21]. Furthermore, experimental evidence suggest the existence of a phase transition on the success of (8) whenever the minimal separation between any pair of spikes crosses the threshold $\Delta = 2/N$ in the limit of N tending to infinity. With the extra assumption that the number of spikes is exactly two, as in (3), it is shown in [6] that a separation $\Delta > 1/N$ is necessary to guarantee exact recovery.

The support stability of the Beurling-LASSO estimator is studied in [6] under the non-degenerate source condition (c.f. Def. 4), however it is unclear and challenging to establish when this condition will hold for general sources. In [29], [30], the support stability of reconstructing positive sources is considered without imposing a minimal separation condition. Our main theorem in this paper is achieved essentially via verifying the non-degenerate source condition for the two-spike case with arbitrary signs and the presence of additional well-separated spikes, which is already quite technical and non-trivial.

Furthermore, when the PSF is the ideal low-pass filter, and under additive white Gaussian noise, the stability of TV
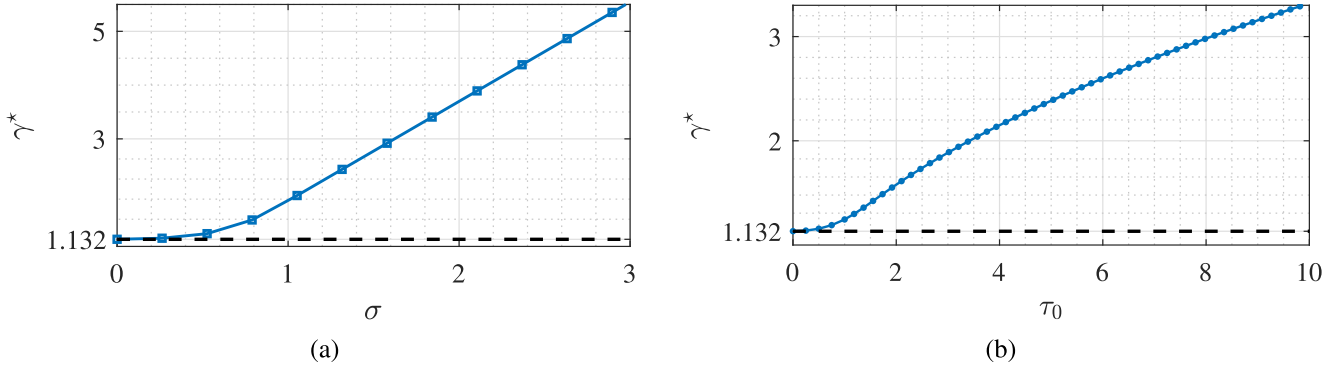
Fig. 2. The stable resolution limit $\gamma^\star$ for (a) a truncated Gaussian PSF for different values of the parameter σ, and (b) the prolate spheroidal wave function of order zero ψ_{τ_0} for different widths of the concentration band $[-\tau_0, \tau_0]$.
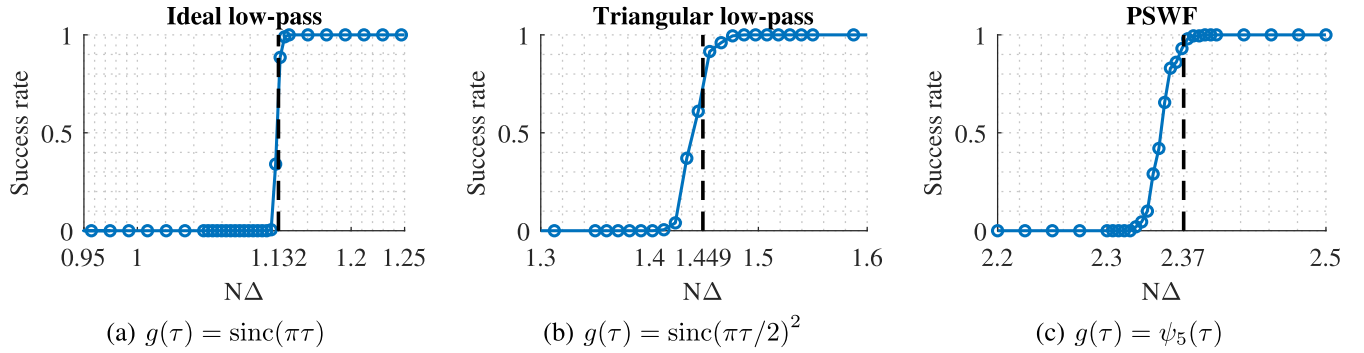

Fig. 3. Empirical success rates for the Beurling-LASSO estimator to return a measure with two point sources, for three different point spread functions, under additive white Gaussian noise, as a function of the separation parameter $N\Delta$. The support stability threshold $\gamma^\star$ predicted by Theorem 2 is shown in a black dashed line. Here, we set $N = 101$, $\mathrm{SNR} = 60\mathrm{dB}$. The results are averaged over 200 trials.

regularization is studied under various metrics. For example, the stability of an estimator $\widehat{\mu}$ in the observation domain is studied in [11], [20], which looks at bounds on $\|\Phi_g(\widehat{\mu} - \mu_\star)\|_2^2$. The performance of support detection has been examined in [31]–[33], which quantifies the residual of $\widehat{\mu}$ outside the support of $\mu_\star$, however these guarantees do not ensure the estimate $\widehat{\mu}$ contains the same number of spikes as the ground truth $\mu_\star$. A trade-off between the separation of the spikes and the error of the parameters is highlighted in [34] without resorting to the non-degenerate source condition. However, the required separation for the result in [34] to hold is quite large and assumes a Gaussian noise.

F. Organization of the Paper

The rest of this paper is organized as follows. Section II provides some prerequisites on spike deconvolution using the Beurling-LASSO estimator, including background literature. Section III states formally the main theorem of this paper including all technical details. Section IV proves the main theorem and Section V discusses the extension of the main theorem to the reconstruction of multiple point sources. Finally, we conclude in Section VI.

II. PREREQUISITES

In this section, we discuss the prerequisites on super resolution using total variation regularization, which are useful to the presentation and analysis of the main result in this paper.

A. Mathematical Notations

The transpose and adjunction of a vector $\boldsymbol{v}$ is denoted as $\boldsymbol{v}^\top$ and $\boldsymbol{v}^*$ respectively. The adjoint of a linear operator Φ is written as Φ^*. Vectors of a dimension $N = 2n + 1$ are indexed between $-n$ and n, so that $\boldsymbol{v} = [v_{-n}, \ldots, v_n]^\top$. For any $t \in \mathbb{T}$, we introduce the atomic vector $\boldsymbol{a}(t) \in \mathbb{C}^N$ and its derivative $\dot{\boldsymbol{a}}(t) \in \mathbb{C}^N$ as

$$\boldsymbol{a}(t) = \left[e^{-2\pi i(-n)t}, \ldots, e^{-2\pi int}\right]^\top,$$

$$\dot{\boldsymbol{a}}(t) = \frac{\mathrm{d}\boldsymbol{a}(t)}{\mathrm{d}t} = -2\pi i \left[-n e^{-2\pi i(-n)t}, \ldots, n e^{-2\pi int}\right]^\top$$
$$= -2\pi i \operatorname{diag}(-n, \ldots, n)\boldsymbol{a}(t),$$

where $\operatorname{diag}(-n, \ldots, n)$ is a diagonal matrix with diagonal entries $-n, \ldots, n$. Similarly, we define the ℓth order derivative of $\boldsymbol{a}(t)$ as $\boldsymbol{a}^{(\ell)}(t)$. To every vector $\boldsymbol{q} \in \mathbb{C}^N$, we associate the trigonometric polynomial $Q(t)$ of degree n such that

$$Q(t) = \boldsymbol{a}(t)^* \boldsymbol{q} = \sum_{k=-n}^{n} q_k e^{2\pi ikt}, \quad \forall t \in \mathbb{T}.$$

Its derivative satisfies $Q'(t) = \dot{\boldsymbol{a}}(t)^* \boldsymbol{q}$ for all $t \in \mathbb{T}$. The real part and conjugate of a complex number u is denoted as $\Re(u)$ and $\bar{u}$, and the sign of a non-zero complex number is given by $\operatorname{sgn}(u) = u/|u|$. For any two vectors $\boldsymbol{z}, \boldsymbol{p} \in \mathbb{C}^N$, we denote by $\langle \cdot, \cdot \rangle_\mathbb{R}$ the real inner product $\langle \boldsymbol{z}, \boldsymbol{p} \rangle_\mathbb{R} = \Re\{\boldsymbol{z}^* \boldsymbol{p}\}$, and we denote by $\boldsymbol{z} \odot \boldsymbol{p} \in \mathbb{C}^N$ their element-wise product.

The vector space of continuous functions from $\mathbb{T}$ to $\mathbb{C}$, denoted as $\mathcal{C}(\mathbb{T})$, is endowed with the supremum norm $\|\cdot\|_\infty$. The total variation norm $\|\cdot\|_{\mathrm{TV}}$, defined as the dual norm of $\|\cdot\|_\infty$, is given as

$$\|\mu\|_{\mathrm{TV}} = \sup_{\substack{h\in\mathcal{C}(\mathbb{T}) \\ \|h\|_\infty \leq 1}} \Re\left[\int_\mathbb{T} \overline{h(t)} \mathrm{d}\mu(t)\right], \quad \forall \mu \in \mathcal{M}(\mathbb{T}).$$

Given two functions g and h that depend on N, we use the classical Landau's notations $g = \mathcal{O}(h)$, $g = o(h)$ and $g = \omega(h)$ to denote $\lim_{N\to\infty}\left|\frac{g(N)}{h(N)}\right| = C$ for some $C \in \mathbb{R}$, $\lim_{N\to\infty}\left|\frac{g(N)}{h(N)}\right| = 0$, and $\lim_{N\to\infty}\left|\frac{g(N)}{h(N)}\right| = \infty$, respectively.

B. Tightness of Total Variation Minimization

In the noiseless setting, the TV estimator (8) is said to be tight if its output $\widehat{\mu}_0$ is equal to the output $\widehat{\mu}_\natural$ of the estimator (7). As for many other convex optimization-based methods for solving inverse problems, the Lagrangian duality theory can be leveraged to derive tightness guarantees. The Lagrange dual problem associated to (8) reads

$$\mathcal{D}_0(\boldsymbol{x}) = \arg\max_{\boldsymbol{p}\in\mathbb{C}^N} \langle \boldsymbol{x}, \boldsymbol{p}\rangle_\mathbb{R}$$
$$\text{subject to } \|\Phi_g^*(\boldsymbol{p})\|_\infty \leq 1, \quad (11)$$

where the adjoint of the operator Φ_g in (5) is given by

$$\Phi_g^*:\ \mathbb{C}^N \to \mathcal{M}(\mathbb{T})$$
$$\boldsymbol{p} \mapsto \Phi_g^*(\boldsymbol{p})(t) = \sum_{k=-n}^n p_k \bar{g}_k e^{i2\pi kt},$$
$$= \boldsymbol{a}(t)^* \mathrm{diag}(\overline{\boldsymbol{g}})\boldsymbol{p}, \quad \forall t \in \mathbb{T}. \quad (12)$$

In other words, Φ_g^* associates any $\boldsymbol{p} \in \mathbb{C}^N$ with a trigonometric polynomial $Q(t) = \boldsymbol{a}(t)^* \boldsymbol{q}$ of degree at most n, where $\boldsymbol{q} = \mathrm{diag}(\overline{\boldsymbol{g}})\boldsymbol{p}$. Moreover, as the restriction of the feasible set of (11) to the span of the operator Φ_g is compact, the set of solutions $\mathcal{D}_0(\boldsymbol{x})$ is non-empty as long as $\boldsymbol{x}$ is a consistent observation under the observation model (4).

It is now well understood that the tightness of TV regularization is characterized by the existence of a so-called *dual certificate* [2]: a function lying in the feasible set of the dual program (11), and satisfying certain extremal interpolation properties. Considering an input measure with only two point sources of the form (1), the corresponding dual certificate is defined as follows.

Definition 3 (Dual Certificate [2]): A vector $\boldsymbol{p} \in \mathbb{C}^N$ is said to be a dual certificate for the optimization problem (8) with an input $\mu_\star$ of the form (3) if and only if the trigonometric polynomial $Q(t) = \boldsymbol{a}(t)^* \boldsymbol{q}$ with a coefficient vector $\boldsymbol{q} = \mathrm{diag}(\overline{\boldsymbol{g}})\boldsymbol{p}$ verifies the conditions

$$Q(t_1) = \mathrm{sgn}(c_1), \quad (13\text{a})$$
$$Q(t_2) = \mathrm{sgn}(c_2), \quad (13\text{b})$$
$$|Q(t)| < 1, \quad \forall t \in \mathbb{T}\backslash\{t_1, t_2\}. \quad (13\text{c})$$

It can easily be verified that any dual certificate $\boldsymbol{p}$ achieves dual optimality with a dual objective $\langle \boldsymbol{x}, \boldsymbol{p}\rangle_\mathbb{R} = \|\mu_\star\|_{\mathrm{TV}}$. In fact, by a duality argument, if such a certificate exists,

$\mu_\star$ is the unique solution of (8). As a result, showing the existence of a dual certificate of a given instance of the total variation program (8) provides a constructive approach to prove the tightness of the TV estimator $\widehat{\mu}_0$. We refer the reader to [2], [35] for further discussion of this property.

C. Stability of the Beurling-LASSO Estimator

Moving to the noisy case, where we aim to recover $\mu_\star$ from the noisy observations (6), it becomes necessary to invoke the Beurling-LASSO estimator (9). Due to the existence of noise, it is no longer possible to recover $\mu_\star$ perfectly. However, we hope the estimator is stable, so that the estimate $\widehat{\mu}_\lambda$ is close to the ground truth measure $\mu_\star$ when the noise $\boldsymbol{w}$ and the regularization parameter λ are small enough. More precisely, we are interested in the support stability as defined in Def. 1, which is a quite strong metric carrying the desirable notion of maintaining a faithful estimate of each individual spike without incurring spurious or missing spikes. The *support stability* of the Beurling-LASSO estimator $\widehat{\mu}_\lambda$ in (9) has been studied for a broad class of measurement operators in [6]. The results, again, are derived from an analysis of the Lagrange dual problem of the estimator $\widehat{\mu}_\lambda$, which is given as

$$\widehat{\boldsymbol{p}}_\lambda = \widehat{\boldsymbol{p}}_\lambda(\boldsymbol{z}) = \arg\max_{\boldsymbol{p}\in\mathbb{C}^N} \langle \boldsymbol{z}, \boldsymbol{p}\rangle_\mathbb{R} - \frac{\lambda}{2}\|\boldsymbol{p}\|_2^2$$
$$\text{subject to } \|\Phi_g^*(\boldsymbol{p})\|_\infty \leq 1, \quad (14)$$

Note that, contrary to (11), the solution $\widehat{\boldsymbol{p}}_\lambda$ is unique for every $\boldsymbol{z}$ and $\lambda > 0$, as (14) can equivalently be interpreted as the projection of $\lambda^{-1}\boldsymbol{z}$ onto the convex feasible set. As the regularization parameter λ tends to 0, the output $\widehat{\boldsymbol{p}}_\lambda(\boldsymbol{x})$ of the dual problem (14) applied on the ground truth observations $\boldsymbol{x} = \Phi_g(\mu)$ converges towards the element $\widehat{\boldsymbol{p}}_{\min}$ of the solution set $\mathcal{D}_0(\boldsymbol{x})$ with the minimal norm so that [6]

$$\lim_{\lambda\to 0} \widehat{\boldsymbol{p}}_\lambda(\boldsymbol{x}) = \arg\min\{\|\boldsymbol{p}\|_2\ :\ \boldsymbol{p}\in\mathcal{D}_0(\boldsymbol{x})\} := \widehat{\boldsymbol{p}}_{\min}. \quad (15)$$

The minimal norm solution $\widehat{\boldsymbol{p}}_{\min}$ to the dual of the TV estimator (8) therefore encodes the behavior of the Beurling-LASSO estimator when both the noise level $\|\boldsymbol{w}\|_2$ and λ tend to 0. In fact, the support stability of Beurling-LASSO can be guaranteed for a measure $\mu_\star$ if some algebraic properties on $\widehat{\boldsymbol{p}}_{\min}$ can be verified. These properties introduced in [6] as *non-degenerate source conditions* are recalled in the following definition.

Definition 4 (Non-Degenerate Source Condition [6]): A measure $\mu_\star$ of the form (3) is said to verify the non-degenerate source condition with respect to the measurement operator Φ_g if and only if the trigonometric polynomial $\widehat{Q}_{\min}(t) = \boldsymbol{a}(t)^* \widehat{\boldsymbol{q}}_{\min}$ with coefficients $\widehat{\boldsymbol{q}}_{\min} = \mathrm{diag}(\bar{\boldsymbol{g}})\widehat{\boldsymbol{p}}_{\min}$, where $\widehat{\boldsymbol{p}}_{\min}$ is defined in (15):

1) verifies the dual certificate conditions in (13),
2) has non-vanishing second derivatives at the source locations, or equivalently

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}\left|\widehat{Q}_{\min}(t_s)\right| < 0, \quad s = 1, 2. \quad (16)$$

The first condition in the above definition essentially requires that the minimal norm solution $\widehat{\boldsymbol{p}}_{\min}$ is a valid dual

certificate for the TV estimator (8) with noiseless input x and, therefore, guarantees its tightness. In view of (15), this ensures that the low-noise limit of the Beurling-LASSO estimator (9) can recover the ground truth measure $\mu_\star$, which is a natural requirement for support stability. The second condition adds an additional constraint, which enforces the certificate to have a strictly concave modulus around the location of the point sources, which, roughly speaking, ensures no spurious spikes will be introduced when adding a little bit of noise. Together, these conditions are used to guarantee the support stability of the Beurling-LASSO estimator in the low-noise regime in [6], that is recalled in the following proposition.

Proposition 5 (Support Stability of Beurling-LASSO [6]): Suppose that the ground truth measure $\mu_\star$ verifies the non-degenerate source condition with respect to the sampling operator Φ_g in Def. 4. Then there exists $\alpha > 0$ such that the Beurling-LASSO estimator $\widehat{\mu}_\lambda(z)$ applied to the noisy measurements $z = \Phi_g(\mu_\star) + w$ with a regularization parameter $\lambda = \alpha^{-1} \|w\|_2$ is support stable in the sense of Def. 1.

Proposition 5 suggests a constructive approach to prove the support stability of the Beurling-LASSO estimator, namely, by verifying the minimal norm solution $\widehat{p}_{\min}$ associated with a measure $\mu_\star$ satisfies the non-degenerate source condition. However, the original work [6] does not provide explicit means to verify this condition. Subsequent works [29], [30] studied the special case of positive sources. Nevertheless, it remains unclear when these conditions are verified for general sources and PSFs. The main theorem, presented in the next section, is built upon verifying the non-degenerate source condition for a two-spike measure with arbitrary coefficients and PSFs satisfying some mild regularity conditions.

III. MAIN RESULT

This section formally introduces the main contribution of this paper, which is to provide a sufficient separation condition between the two spikes of a measure of the form (3) to guarantee the support stability. The provided bound depends only on the PSF g, and more specifically on its auto-correlation function and successive derivatives. The presented result is achieved via verifying the non-degenerate source condition presented in Def. 4, whose proof will be detailed in Section IV.

We denote by $\kappa = \mathcal{K}(g) \in L_2$ the auto-correlation of the PSF $g \in L_2$, defined as

$$\mathcal{K}(g)(\tau) = \int_{\mathbb{R}} \overline{g(u)} g(\tau + u) \mathrm{d}u, \quad \forall \tau \in \mathbb{R}. \quad (17)$$

Denote two auxiliary functions $u_\beta, v_\beta \in L_2$ defined for every $\beta > 0$ as

$$u_\beta(\tau) = \kappa\left(\tau - \frac{\beta}{2}\right) + \kappa\left(\tau + \frac{\beta}{2}\right), \quad (18a)$$

$$v_\beta(\tau) = \kappa\left(\tau - \frac{\beta}{2}\right) - \kappa\left(\tau + \frac{\beta}{2}\right). \quad (18b)$$

The above two functions describe the auto-correlations of a signal produced by two point sources separated by a distance β with same and opposite signs, respectively. When g is real, the κ is even, and we point out that the functions u_β, v_β are

even and odd, respectively. We are now ready to state the main theorem of this paper.

Theorem 6 (Stable Resolution Limit of Beurling-LASSO): Suppose that the PSF g satisfies the following regularity conditions (H1)-(H4).

(H1) $g \in L_2$ is real and non-zero, *i.e.* $\exists \tau \in \mathbb{R}, \ g(\tau) \neq 0$.

(H2) $G = \mathcal{F}(g) \in L_2$ is band-limited within B, *i.e.*

$$G(f) = 0, \quad \forall |f| > B/2.$$

(H3) G is bounded, *i.e.*

$$\sup_{f \in \mathbb{R}} |G(f)| = \sup_{f \in [-B/2, B/2]} |G(f)| < \infty.$$

(H4) For $\kappa = \mathcal{K}(g) \in L_2$ and its first three derivatives, $\kappa^{(\ell)}$ for $\ell = 0, 1, 2, 3$, and for any odd integer $N \in 2\mathbb{Z} + 1$, we have that

$$S_\ell(N) \triangleq \sup_{t \in \left[-\frac{1}{2}, \frac{1}{2}\right]} \left| \sum_{\substack{k \in \mathbb{Z} \\ k \neq 0}} \kappa^{(\ell)}\left(\frac{N}{B}(t + k)\right) \right| < \infty,$$

$$(19a)$$

$$\lim_{N \to \infty} S_\ell(N) = 0. \quad (19b)$$

Let $\gamma^\star$, depending only on the PSF g, be defined as $\gamma^\star = \max\{\gamma_1^\star, \gamma_2^\star, \gamma_3^\star\} > 0$ with

$$\gamma_1^\star = B \sup_{\beta > 0} \left\{ \sup_{\tau \geq 0} |\widetilde{s}_\beta(\tau)| > \widetilde{s}_\beta\left(\frac{\beta}{2}\right) \right\}, \quad (20a)$$

$$\gamma_2^\star = B \sup_{\beta > 0} \left\{ \sup_{\tau \geq 0} |\widetilde{r}_\beta(\tau)| > \widetilde{r}_\beta\left(\frac{\beta}{2}\right) \right\}, \quad (20b)$$

$$\gamma_3^\star = B \sup_{\beta > 0} \left\{ -\kappa''(0)^2 + \kappa''(\beta)^2 - \kappa'(\beta)\kappa'''(\beta) \geq 0 \right\},$$

$$(20c)$$

where the intermediate functions $\widetilde{s}_\beta(\tau)$, $\widetilde{r}_\beta(\tau)$ are further defined, for any $\beta > 0$ and $\tau \in \mathbb{R}$ as

$$\widetilde{s}_\beta(\tau) = (-\kappa''(0) - \kappa''(\beta)) v_\beta(\tau) - \kappa'(\beta) u_\beta'(\tau), \quad (21a)$$

$$\widetilde{r}_\beta(\tau) = (-\kappa''(0) + \kappa''(\beta)) u_\beta(\tau) + \kappa'(\beta) v_\beta'(\tau). \quad (21b)$$

Then there exists $N_0 \in \mathbb{N}$ such that, for every $N \geq N_0$ and every $\mu_\star$ of the form (3) with

$$\Delta := |t_1 - t_2|_{\mathbb{T}} > \frac{\gamma^\star}{N}, \quad (22)$$

there exists $\alpha > 0$ such that the Beurling-LASSO estimator $\widehat{\mu}_\lambda(\Phi_g(\mu_\star) + w)$ with the regularization parameter $\lambda = \alpha^{-1} \|w\|_2$ is support stable.

Theorem 6 provides an explicit means to compute $\gamma^\star$, based on the evaluation of (20), for a given PSF satisfying the regularity conditions. The key quantities, $\gamma_k^\star$, $k = 1, 2, 3$, are suprema of continuous functions where the complexity of the computation essentially depends on the variations and smoothness of κ. It is worth noticing that these quantities are also independent of the bandwidth B through a re-scaling the PSF via a transform $g(\tau) \leftarrow g(c\tau)$ for some $c > 0$, and can therefore be computed for a PSF with unit bandwidth $B = 1$. Table I provides several examples of stable resolution limits for PSFs frequently encountered in practice. The separation

condition (22) can be equivalently interpreted in terms of the delays τ_1, τ_2 of the unnormalized measure $\nu_\star$ as in (1) as

$$B\,|\tau_2 - \tau_1| > \gamma^\star, \tag{23}$$

provided the spikes to be localized are in the interval $(-N/2B, N/2B)$. In addition, our results allow arbitrary coefficients of the spikes, as long as they are sufficiently separated above the resolution limit.

If we impose a strictly stronger assumption that the PSF decays "reasonably fast", i.e, there exist constants $C_\ell > 0$, $\ell = 0, 1, 2, 3$ and $\delta > 0$ such that

$$\left|g^{(\ell)}(\tau)\right| \le \frac{C_\ell}{1 + |\tau|^{1+\delta}}, \quad \forall \tau \in \mathbb{R}, \ \ell = 0, 1, 2, 3, \tag{24}$$

then the assumptions (H3) and (H4) are automatically verified, which greatly eases the applicability of Theorem 6. Although (24) can be verified for the majority of the PSFs encountered in practical applications, it excludes some PSFs of theoretical interest, such as the ideal low-pass function $g(\tau) = \mathrm{sinc}\,(\pi\tau)$, and the truncated Gaussian kernel.

IV. PROOF OF THEOREM 6

This section aims to prove Theorem 6. Recall Proposition 5, which provides a sufficient condition to establish the support stability of the Beurling-LASSO estimator (9). The proof essentially consists of deriving a sufficiently large constant $\gamma^\star$ above which a ground truth measure $\mu_\star$ verifies the non-degenerate source condition (c.f. Definition 4) outlined in Proposition 5 when N is large enough. The proof is divided into four major steps.

1) Using invariance properties of the non-degenerate source condition, we start by reducing the problem to reconstructing a "canonical" measure $\mu_\star$ exhibiting useful symmetry and simplifying the ensuing calculations.
2) Next, we introduce a so-called minimal vanishing derivative polynomial $Q_V(t) = \boldsymbol{a}(t)^* \mathrm{diag}\,(\bar{\boldsymbol{g}})\,\boldsymbol{p}_V$, where $\boldsymbol{p}_V$ is the solution to a simple linear system depending on the parameters of the canonical measure $\mu_\star$. Leveraging Lemma 8 [6], we show that it is enough to show $Q_V(t)$ satisfies (13) and (16).
3) We next study the limiting behavior of $Q_V(t)$, where it converges towards a band-limited function $\mathcal{Q}_V(Nt/B)$ as $N \to \infty$. Furthermore, we show that, if $\mathcal{Q}_V(\tau)$ verifies the conditions in (65), then $Q_V(t)$ satisfies (13) and (16), provided that N is large enough.
4) Finally, we derive a sufficient separation condition $\gamma_\star$ above which the limit function $\mathcal{Q}_V(\tau)$ verifies the conditions (65) whenever $N\Delta > \gamma_\star$. We conclude on the statement of Theorem 6 by applying Proposition 5.

A. Canonical Reduction of the Problem

We start by reducing the problem to a simpler form without loss of generality by exploiting the following invariances of the non-degenerate source condition.

Lemma 7 (Invariances of the Non-Degenerate Source Condition): Suppose that a measure $\mu_1 \in \mathcal{M}(\mathbb{T})$ of the form (3) verifies the non-degenerate source condition given in Def. 4

with respect to the sampling operator Φ_g. The follow statements hold:

- *Translation invariance:* for any $t_0 \in \mathbb{T}$, the measure $\mu_2(t) = \mu_1(t - t_0)$ for all $t \in \mathbb{T}$ verifies the non-degenerate source condition for the same operator.
- *Scaling invariance:* for any $c > 0$ and $\theta \in [0, 2\pi)$, the measure $\mu_2(t) = ce^{i\theta}\mu_1(t)$ for all $t \in \mathbb{T}$ verifies the non-degenerate source condition for the same operator.
- *Time reversal invariance:* the measure $\mu_2(t) = \mu_1(-t)$ for all $t \in \mathbb{T}$ verifies the non-degenerate source condition for the same operator.

A proof of Lemma 7 is provided in Appendix A. Leveraging this result, we conclude that the non-degenerate source conditions depends only on the separation $\Delta \triangleq |t_2 - t_1|_{\mathbb{T}}$ between the two point sources and the angle $\theta \triangleq \arg(c_2/c_1) \bmod \pi$ between their complex amplitude. For any $\Delta \in (0, \frac{1}{2}]$ and $\theta \in [0, \pi]$, we define the *canonical* measure $\mu_\star(\Delta, \theta)$ as

$$\mu_\star(\Delta, \theta) = e^{-i\frac{\theta}{2}}\delta\left(t + \frac{\Delta}{2}\right) + e^{i\frac{\theta}{2}}\delta\left(t - \frac{\Delta}{2}\right). \tag{25}$$

We can now restrict our analysis to canonical measures of the form $\mu_\star = \mu_\star(\Delta, \theta)$ without loosing any generality, and exploit the Hermitian symmetry of $\mu_\star$ to simplify the ensuing calculations.

B. The Minimal Vanishing Derivative Polynomial

The minimal norm solution $\widehat{\boldsymbol{p}}_{\min}$ (defined in (15)) associated to the measure $\mu_\star = \mu_\star(\Delta, \theta)$ can be equivalently interpreted as a projection onto the spectrahedra of bounded trigonometric polynomials. Such projection is difficult to derive analytically, limiting the ability to establish the non-degenerate source condition (c.f. Def. 4) through a direct analysis of $\widehat{\boldsymbol{p}}_{\min}$. To bypass this problem, it is proposed in [6] to study instead the behaviors of a surrogate vector $\boldsymbol{p}_V$, defined in the present context as the unique solution to the following quadratic problem:

$$\boldsymbol{p}_V = \underset{\boldsymbol{p} \in \mathbb{C}^N}{\arg\min}\, \|\boldsymbol{p}\|_2$$

$$\text{subject to} \quad \boldsymbol{a}\left(\frac{\Delta}{2}\right)^* \mathrm{diag}(\bar{\boldsymbol{g}})\boldsymbol{p} = e^{i\theta/2},$$

$$\boldsymbol{a}\left(-\frac{\Delta}{2}\right)^* \mathrm{diag}(\bar{\boldsymbol{g}})\boldsymbol{p} = e^{-i\theta/2},$$

$$\dot{\boldsymbol{a}}\left(\frac{\Delta}{2}\right)^* \mathrm{diag}(\bar{\boldsymbol{g}})\boldsymbol{p} = 0,$$

$$\dot{\boldsymbol{a}}\left(-\frac{\Delta}{2}\right)^* \mathrm{diag}(\bar{\boldsymbol{g}})\boldsymbol{p} = 0. \tag{26}$$

A key property of $\boldsymbol{p}_V$, recalled in the following lemma, is its equivalence with the minimal norm solution $\widehat{\boldsymbol{p}}_{\min}$ under additional assumptions.

Lemma 8 (Equivalence of the Minimal Vanishing Derivative Solution [6]): If the solution $\boldsymbol{p}_V$ of (26) satisfies $\left\|\Phi_g^*(\boldsymbol{p}_V)\right\|_\infty \le 1$, then $\boldsymbol{p}_V = \widehat{\boldsymbol{p}}_{\min}$ is equal to the minimum norm solution for the measure $\mu_\star$.

Let $\boldsymbol{q}_V = \mathrm{diag}\,(\bar{\boldsymbol{g}})\,\boldsymbol{p}_V \in \mathbb{C}^N$, and the associated polynomial $Q_V(t) = \boldsymbol{a}(t)^*\boldsymbol{q}_V$ be the minimal vanishing derivative

polynomial. Note that by construction, $Q_V(t)$ verifies the interpolation constraints (13a) and (13b). Additionally, if $\boldsymbol{q}_V$ also verifies the extremal constraint (13c), $\left\|\Phi_g^*(\boldsymbol{p}_V)\right\|_\infty = \sup_{t\in\mathbb{T}} |Q_V(t)| \leq 1$, then $\boldsymbol{p}_V = \widehat{\boldsymbol{p}}_{\min}$ by Lemma 8. Therefore, it is enough to show that $Q_V(t)$ verifies (13c) and (16) to conclude on the support stability of $\mu_\star$ for the sampling operator Φ_g. To ensure that those conditions can be met, we start by studying the asymptotic of $Q_V(t)$ when $N \to \infty$.

C. Asymptotic Analysis of $Q_V(t)$

We start the asymptotic analysis of the trigonometric polynomial $Q_V(t) = \boldsymbol{a}(t)^* \boldsymbol{q}_V$ by defining the discrete auto-correlation function $K(t)$ of the PSF $g(\tau)$ as

$$K(t) \triangleq \sum_{k=-n}^{n} |g_k|^2 \, e^{i2\pi kt} = \sum_{k=-n}^{n} \left| G\left(\frac{Bk}{N}\right) \right|^2 e^{i2\pi kt}, \quad (27)$$

for any $t \in \mathbb{R}$, which is a real and even trigonometric polynomial. The following lemma, whose proof is delayed to Appendix B, gives an explicit expression of the polynomial $Q_V(t)$ in terms of the parameters Δ, θ and the polynomial $K(t)$.

Lemma 9: Suppose that $\boldsymbol{g}$ has at least four non-zero coefficients, then for all $\Delta \in (0, \frac{1}{2}]$ and all $\theta \in [0, \pi]$, Q_V can be decomposed as

$$Q_V(t) = \cos\left(\frac{\theta}{2}\right) R_\Delta(t) + i \sin\left(\frac{\theta}{2}\right) S_\Delta(t), \quad \forall t \in \mathbb{T}, \tag{28}$$

where R_Δ and S_Δ are respectively the even and odd real trigonometric polynomials, independent of θ, given by

$$C_R(\Delta)R_\Delta(t) = (-K''(0) + K''(\Delta))\left(K(t - \frac{\Delta}{2}) + K(t + \frac{\Delta}{2})\right) + K'(\Delta)\left(K'(t - \frac{\Delta}{2}) - K'(t + \frac{\Delta}{2})\right), \tag{29a}$$

$$C_S(\Delta)S_\Delta(t) = (-K''(0) - K''(\Delta))\left(K(t - \frac{\Delta}{2}) - K(t + \frac{\Delta}{2})\right) - K'(\Delta)\left(K'(t - \frac{\Delta}{2}) + K'(t + \frac{\Delta}{2})\right), \tag{29b}$$

where the quantities $C_R(\Delta) > 0$ and $C_S(\Delta) > 0$ are positive and for all $\Delta \in (0, \frac{1}{2}]$ and given as

$$C_R(\Delta) = (-K''(0) + K''(\Delta))(K(0) + K(\Delta)) - K'(\Delta)^2 \tag{30a}$$

$$C_S(\Delta) = (-K''(0) - K''(\Delta))(K(0) - K(\Delta)) - K'(\Delta)^2. \tag{30b}$$

Next, we demonstrate in Lemma 10 the uniform convergence of K towards the auto-correlation function κ.

Lemma 10 (Uniform Convergence of K): Under the hypothesis of Theorem 6, the equality

$$\lim_{N\to\infty} \sup_{t\in\mathbb{T}} \left| \left(\frac{B}{N}\right)^{\ell+1} K^{(\ell)}(t) - \kappa^{(\ell)}\left(\frac{Nt}{B}\right) \right| = 0, \tag{31}$$

holds for $\ell = 0, 1, 2, 3$.

The proof of the above is presented in Appendix C. By (H1) and (H2), the Fourier transform G of the PSF g is non-zero on a non-empty open interval $I \subset \mathbb{R}$. By (H2), G is band-limited within B, and one must have $I \subset (-B/2, B/2)$. It comes that if $N \geq \left\lceil \frac{4}{|I|} \right\rceil$, at least four elements of the form $\{kB/N\}_{|k|\leq n}$ will fall into I, and the vector $\boldsymbol{g}$ has at least four non-zero coefficients. Lemma 10 can be applied to the expression of Q_V (28) to get the existence of a function $\mathcal{Q}_V$ verifying the convergence

$$\lim_{N\to\infty} \sup_{t\in\mathbb{T}} \left| \left(\frac{B}{N}\right)^{\ell} Q_V^{(\ell)}(t) - \mathcal{Q}_V^{(\ell)}\left(\frac{Nt}{B}\right) \right| = 0. \tag{32}$$

for $\ell = 0, 1, 2$. Moreover, defining $\beta = N\Delta/B$, the limit function $\mathcal{Q}_V$ writes as

$$\mathcal{Q}_V(\tau) = \cos\left(\frac{\theta}{2}\right) r_\beta(\tau) + i \sin\left(\frac{\theta}{2}\right) s_\beta(\tau), \quad \forall \tau \in \mathbb{R}. \tag{33}$$

The intermediate real functions r_β, s_β in the above expression are even and respectively, given by the limits of $R_{B\beta/N}(B\tau/N)$ and $S_{B\beta/N}(B\tau/N)$, so that

$$C_r(\beta)r_\beta(\tau) = (-\kappa''(0) + \kappa''(\beta)) u_\beta(\tau) + \kappa'(\beta)v_\beta'(\tau), \tag{34a}$$

$$C_s(\beta)s_\beta(\tau) = (-\kappa''(0) - \kappa''(\beta)) v_\beta(\tau) - \kappa'(\beta)u_\beta'(\tau), \tag{34b}$$

where the function u_β, v_β are defined in (18), and $C_r(\beta), C_s(\beta)$ are positive quantities given by

$$C_r(\beta) = (-\kappa''(0) + \kappa''(\beta))(\kappa(0) + \kappa(\beta)) - \kappa'(\beta)^2 > 0, \tag{35a}$$

$$C_s(\beta) = (-\kappa''(0) - \kappa''(\beta))(\kappa(0) - \kappa(\beta)) - \kappa'(\beta)^2 > 0. \tag{35b}$$

We highlight that the compositions (28) and (33) indicate that both $Q_V(t)$ and $\mathcal{Q}_V(\tau)$ are Hermitian: $\overline{Q_V(t)} = Q_V(-t)$ for all $t \in \mathbb{T}$ and $\overline{\mathcal{Q}_V(\tau)} = \mathcal{Q}_V(-\tau)$ for all $\tau \in \mathbb{R}$. Suppose that the limit function $\mathcal{Q}_V$ verifies the conditions

$$\frac{d^2 |\mathcal{Q}_V|}{d\tau^2}\left(\frac{\beta}{2}\right) < 0, \tag{36a}$$

$$|\mathcal{Q}_V(\tau)| < 1, \quad \forall \tau \in \mathbb{R}^+ \setminus \left\{\frac{\beta}{2}\right\}. \tag{36b}$$

It follows from the uniform convergence (32) of Q_V towards $\mathcal{Q}(Nt/B)$, and the Hermitian symmetry of Q_V and $\mathcal{Q}_V$ that there must exist an integer $\widetilde{N}_0$ so that for all $N \geq \widetilde{N}_0$, Q_V verifies (13c) and (16).

Therefore, using the results of Section IV-B, if $\mathcal{Q}_V$ verifies (65), then $\mu_\star(\Delta, \theta)$ verifies the non-degenerate source condition with respect to the measurement operator Φ_g whenever $N \geq N_0 \triangleq \max\left\{\left\lceil \frac{4}{|I|} \right\rceil, \widetilde{N}_0\right\}$.

D. Non-Degeneracy of $\mu_\star(\Delta, \theta)$

It remains to derive sufficient conditions under which the limit function $\mathcal{Q}_V$ satisfies (65) to conclude on the desired result. First of all, it comes immediately from (28) that

$$|\mathcal{Q}_V(\tau)|^2 = \cos^2\left(\frac{\theta}{2}\right)|r_\beta(\tau)|^2 + \sin^2\left(\frac{\theta}{2}\right)|s_\beta(\tau)|^2, \quad (37)$$

holds for any $\tau \in \mathbb{R}$.

We start by discussing the condition (36b). From (37), it is clear that (36b) is verified for every θ in $[0, \pi]$ if and only if $|r_\beta(\tau)| < 1$ and $|s_\beta(\tau)| < 1$ for all τ in $\mathbb{R}^+\backslash\{\beta/2\}$. Those last inequalities can be equivalently interpreted in terms of the functions $\widetilde{r}_\beta(\tau)$, $\widetilde{s}_\beta(\tau)$ introduced in (21) from the identity $\widetilde{r}_\beta(\tau) = C_r(\beta)r_\beta(\tau)$ and $\widetilde{s}_\beta(\tau) = C_s(\tau)s_\beta(\tau)$ for all $\tau \in \mathbb{R}$, as we have that

$$|r_\beta(\tau)| < 1 \Longleftrightarrow \frac{|\widetilde{r}_\beta(\tau)|}{C_r(\beta)} < 1,$$

$$|s_\beta(\tau)| < 1 \Longleftrightarrow \frac{|\widetilde{s}_\beta(\tau)|}{C_s(\beta)} < 1.$$

Hence, the condition (36b) is equivalent to

$$C_r(\beta) - |\widetilde{r}_\beta(\tau)| > 0, \quad \forall \tau \in \mathbb{R}^+\backslash\left\{\frac{\beta}{2}\right\} \quad (38a)$$

$$C_s(\beta) - |\widetilde{s}_\beta(\tau)| > 0, \quad \forall \tau \in \mathbb{R}^+\backslash\left\{\frac{\beta}{2}\right\}. \quad (38b)$$

From the definitions (21) and (35), one has $C_r(\beta) = \widetilde{r}_\beta(\frac{\beta}{2})$ and $C_s(\beta) = \widetilde{s}_\beta(\frac{\beta}{2})$. Leveraging the hypothesis $\beta = N\Delta/B > \max\{\gamma_1^\star, \gamma_2^\star\}/B$, the functions $C_r(\beta) - |\widetilde{r}_\beta(\tau)|$ and $C_s(\beta) - |\widetilde{s}_\beta(\tau)|$ are non-negative and can reach 0 only for $\tau = \beta/2$. Hence (38) and subsequently (36b) are both verified under the hypothesis of Theorem 6.

It remains to show that $\mathcal{Q}_V$ satisfies (36a). We recall that the second derivative of the modulus of a function $q = q_R + iq_I$, where q_R, q_I are its real and imaginary part, respectively, reads (see e.g. [36])

$$\frac{\mathrm{d}^2|q|}{\mathrm{d}\tau^2}(\tau) = -\frac{(q_R(\tau)q_R'(\tau) + q_I(\tau)q_I'(\tau))^2}{|q(\tau)|^3}$$
$$+ \frac{|q'(\tau)|^2 + q_R(\tau)q_R''(\tau) + q_I(\tau)q_I''(\tau)}{|q(\tau)|}. \quad (39)$$

Evaluating the above at $\tau = \frac{\beta}{2}$ for the function $\mathcal{Q}_V$ leads to

$$\frac{\mathrm{d}^2|\mathcal{Q}_V|}{\mathrm{d}\tau^2}(\beta/2)$$
$$= \cos(\theta/2)\, r_\beta''(\beta/2) + \sin(\theta/2)\, s_\beta''(\beta/2)$$
$$= \left(-\kappa''(0)^2 + \kappa''(\beta)^2 - \kappa'(\beta)\kappa'''(\beta)\right)$$
$$\left(C_r(\beta)^{-1}\cos(\theta/2) + C_s(\beta)^{-1}\sin(\theta/2)\right), \quad (40)$$

which is strictly negative for every $\theta \in [0, \pi]$ by the hypothesis $\beta = N\Delta/B > \gamma_3^\star/B$.

As a result, under the hypothesis of Theorem 6, if $\Delta = B\beta/N \geq \gamma^\star/N$ with $\gamma^\star = \max\{\gamma_1^\star, \gamma_2^\star, \gamma_3^\star\}$, there exists $N_0 \in \mathbb{N}$ such that for every $N \geq N_0$ the canonical measure $\mu_\star(\Delta, \theta)$ will verify the non-degenerate source condition as in Def. 4 for the sampling operator Φ_g for every $\theta \in [0, \pi]$.

We conclude on the desired statement by an application of Theorem 5. ∎

V. EXTENSION TO MULTIPLE POINT SOURCES

The stable resolution limit presented in Section III only applies to the simple case when only two point sources are present. In this section, we show that the same resolution limit continues to govern the support stability in a multi-source setting including two close-located sources and other well-separated sources.

We consider an extension of the normalized measure $\mu_\star$ in (3) composing S point sources at locations $\mathcal{T} = \{t_s\}_{s=1}^S \subset \mathbb{T}$, given as

$$\mu_\star(t) = \sum_{s=1}^S c_s\delta(t - t_s), \quad (41)$$

where $\{c_s \in \mathbb{C}\}_{s=1}^S$ are non-zero complex amplitudes. Similarly, we can extend the notion of support stability, defined in Def. 1 to include more spikes. Without loss of generality, we assume that t_1 and t_2 are closely located, while the other sources are well-separated from each other. Theorem 11 establishes that the support stability of the Beurling-LASSO estimator can be ensured under the same separation condition between the two close sources as that of Theorem 6, depending only on the characteristics of the PSF.

Theorem 11 (Extension to Multiple Point Sources): Suppose that the PSF g satisfies the regularity conditions (H1)-(H4) of Theorem 6. Let $S \geq 3$ be a constant that does not grow with N, and the support set $\mathcal{T} = \{t_s\}_{s=1}^S$ of $\mu_\star$ verifies

$$\Delta := |t_2 - t_1|_\mathbb{T} > \gamma^\star/N, \quad (42a)$$
$$|t_{s'} - t_s|_\mathbb{T} = \omega(1/N), \text{ for } s \neq s' \text{ and } \{s, s'\} \neq \{1, 2\}, \quad (42b)$$

where $\gamma^\star$ is the constant defined in the statement of Theorem 6. Then there exists $N_0 \in \mathbb{N}$ such that, for every $N \geq N_0$ and every measure $\mu_\star$ of the form (41) satisfying (42), there exists $\alpha > 0$ such that the Beurling-LASSO estimator $\widehat{\mu}_\lambda(\Phi_g(\mu_\star) + \boldsymbol{w})$ with the regularization parameter $\lambda = \alpha^{-1}\|\boldsymbol{w}\|_2$ is support stable.

The proof of this theorem is given in Appendix D. Similar to the proof of Theorem 6 outlined in Section IV, it relies on the characterization of the non-degenerate source condition presented in [6]. More precisely, we show that, under the assumptions of Theorem 11, the minimal vanishing derivative polynomial associated to the measure $\mu_\star$ with S sources as in (41) has the same asymptotic behavior as the sum of the polynomial associated with the measure with two close spikes, whose properties were studied in Section IV, and the polynomials associated with each well-separated single spike.

We compare in Fig. 4 the theoretical threshold $\gamma^\star$ provided by Theorem 11 with the empirical success rate of the Beurling-LASSO estimator to output a measure $\widehat{\mu}$ with S spikes given a ground truth $\mu_\star$ of the form (41), for different values of the separation parameter Δ and for different PSF $g(\tau)$. In the experiments, we set $|t_{s'} - t_s|_\mathbb{T} = 5/N$ as a separation for the well-separated spikes, where $s \neq s'$ and $\{s, s'\} \neq \{1, 2\}$. The amplitude of the colliding spikes c_1, c_2 are chosen with
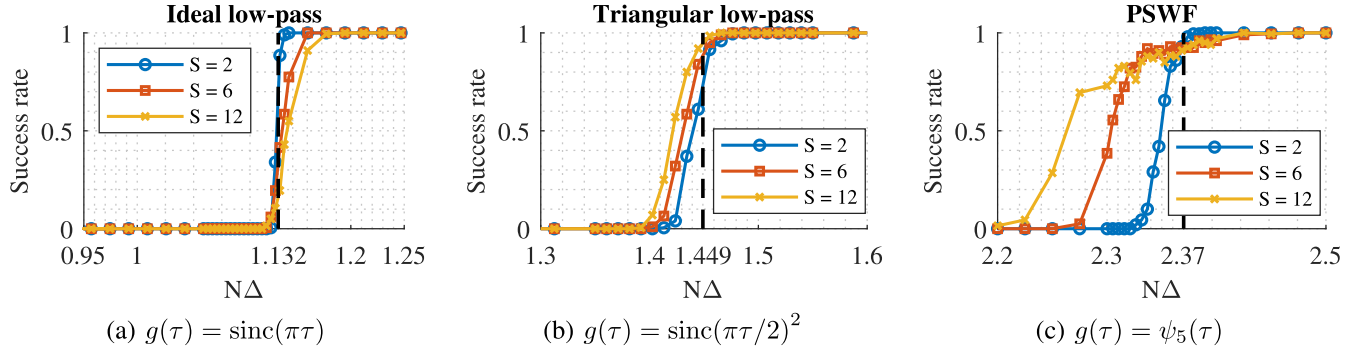
Fig. 4. Empirical success rate for the Beurling-LASSO estimator to return a measure with S point sources, for three different point spread functions, under additive white Gaussian noise, as a function of the separation parameter $N\Delta$. The support stability threshold $\gamma^\star$ predicted by Theorem 11 is shown in a black dashed line. Here, we set $N = 101$, SNR = 60dB and $|t_{s'} - t_s|_{\mathbb{T}} = 5/N$ as a separation for the well separated spikes $s \neq s'$ and $\{s, s'\} \neq \{1, 2\}$. The results are averaged over 200 trials.

opposite signs $c_1 = 1$, $c_2 = -1$, while the amplitudes of the well-separated ones $\{c_s\}_{s=3}^S$ are drawn uniformly at random over the complex unit circle.

VI. CONCLUSIONS

This paper studies the support stability of the Beurling-LASSO estimator for estimating two closely located point sources, possibly in the presence of a finite number of well-separated point sources, and characterizes the resolution limit as a function of the PSF, above which the Beurling-LASSO estimator is support stable. Our result highlights and quantifies the role of PSF in noisy super resolution, which is not evident in the study of the noiseless setting. Our analysis is based on verifying the non-degenerate source conditions put forth in [6]. We believe it is possible to develop similar stable resolution limits in higher dimensions [2], [37], by carefully generalizing the arguments in our paper, which is left for the future work. Finally, it will be interesting to study how the availability of multiple snapshots impact the stable resolution limit [38].

APPENDIX A
PROOF OF LEMMA 7

We restrict the proof to showing the translation invariance of the non-degenerate source condition. The scaling and time-reversal invariances can be demonstrated by following an analogous reasoning.

Consider a measure $\mu_1(t) = c_1\delta(t - t_{1,1}) + c_2\delta(t - t_{1,2})$ for all $t \in \mathbb{T}$, and let

$$\mu_2(t) = \mu_1(t - t_0) = c_1\delta(t - t_{2,1}) + c_2\delta(t - t_{2,2}),$$

for some $t_0 \in \mathbb{T}$ with $t_{2,1} = t_{1,1} - t_0$ and $t_{2,2} = t_{1,2} - t_0$. Moreover, let $\boldsymbol{x}_j = \Phi_g(\mu_j)$ for $j = 1, 2$, and consider an element $\boldsymbol{p}_1$ within the feasible set of (14). For any $t_0 \in \mathbb{T}$, we let $\boldsymbol{p}_2 = \boldsymbol{a}(-t_0) \odot \boldsymbol{p}_1$. Since $\boldsymbol{p}_1 \in \mathcal{D}_0(\boldsymbol{x}_1)$ by hypothesis, we have that

$$\left\|\Phi_g^*(\boldsymbol{p}_2)\right\|_\infty = \sup_{t \in \mathbb{T}} |\boldsymbol{a}(t)^* \operatorname{diag}(\bar{\boldsymbol{g}})\boldsymbol{p}_2|$$
$$= \sup_{t \in \mathbb{T}} |\boldsymbol{a}(t)^* \operatorname{diag}(\bar{\boldsymbol{g}}) (\boldsymbol{a}(-t_0) \odot \boldsymbol{p}_1)|$$
$$= \sup_{t \in \mathbb{T}} |\boldsymbol{a}(t + t_0)^* \operatorname{diag}(\bar{\boldsymbol{g}})\boldsymbol{p}_1|$$

$$= \sup_{t \in \mathbb{T}} |\boldsymbol{a}(t)^* \operatorname{diag}(\bar{\boldsymbol{g}})\boldsymbol{p}_1|$$
$$= \left\|\Phi_g^*(\boldsymbol{p}_1)\right\|_\infty \leq 1. \quad (43)$$

It comes that $\boldsymbol{p}_2$ also lies in the feasible set of (14). Moreover, from the definition (5) of the sampling operator Φ_g, we have that

$$\boldsymbol{x}_2 = \Phi_g(\mu_2) = \Phi_g(\mu_1(\cdot - t_0)) = \boldsymbol{a}(-t_0) \odot \boldsymbol{x}_1.$$

Evaluating the cost function of (14) for $\boldsymbol{p}_2$ yields

$$\langle \boldsymbol{x}_2, \boldsymbol{p}_2 \rangle_{\mathbb{R}} = \langle \boldsymbol{a}(-t_0) \odot \boldsymbol{x}_1, \boldsymbol{a}(-t_0) \odot \boldsymbol{p}_1 \rangle_{\mathbb{R}} = \langle \boldsymbol{x}_1, \boldsymbol{p}_1 \rangle_{\mathbb{R}}.$$

This implies that $\boldsymbol{p}_2 \in \mathcal{D}_0(\boldsymbol{x}_2)$ is a solution of (14) with an input $\boldsymbol{x}_2$ if and only if $\boldsymbol{p}_2 \in \mathcal{D}_0(\boldsymbol{x}_1)$ is a solution of (14) with an input $\boldsymbol{x}_1$. Further noticing that $\|\boldsymbol{p}_2\|_2 = \|\boldsymbol{p}_1\|_2$, we conclude that the minimal norm elements $\widehat{\boldsymbol{p}}_{1,\min}$, $\widehat{\boldsymbol{p}}_{2,\min}$ of the sets $\mathcal{D}_0(\boldsymbol{x}_1)$, $\mathcal{D}_0(\boldsymbol{x}_2)$, respectively, are linked by the relation

$$\widehat{\boldsymbol{p}}_{2,\min} = \boldsymbol{a}(-t_0) \odot \widehat{\boldsymbol{p}}_{1,\min}.$$

Now, suppose that μ_1 verifies the non-degenerate condition with respect to Φ_g, and define the trigonometric polynomials $\widehat{Q}_{j,\min}(t) = \boldsymbol{a}(t)^* \operatorname{diag}(\bar{\boldsymbol{g}})\widehat{\boldsymbol{p}}_{j,\min}$, for $j = 1, 2$. By Def. 4, we have that for any $t \in \mathbb{T}$

$$\widehat{Q}_{2,\min}(t - t_0) = \boldsymbol{a}(t - t_0)^* \operatorname{diag}(\bar{\boldsymbol{g}})\widehat{\boldsymbol{p}}_{2,\min}$$
$$= \boldsymbol{a}(t - t_0)^* \operatorname{diag}(\bar{\boldsymbol{g}}) (\boldsymbol{a}(-t_0) \odot \widehat{\boldsymbol{p}}_{1,\min})$$
$$= \boldsymbol{a}(t)^* \operatorname{diag}(\bar{\boldsymbol{g}})\widehat{\boldsymbol{p}}_{1,\min} = \widehat{Q}_{1,\min}(t). \quad (44)$$

Therefore $\widehat{Q}_{2,\min}(t_{2,s}) = \widehat{Q}_{2,\min}(t_{1,s}) = \operatorname{sgn}(c_s)$ for $s = 1, 2$, and $|\widehat{Q}_{2,\min}(t)| < 1$ for all $t \in \mathbb{T}\backslash \{t_{2,1}, t_{2,2}\}$. Consequently $\widehat{Q}_{2,\min}(t)$ verifies conditions (13). A similar reasoning also yields

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}\left|\widehat{Q}_{2,\min}(t_{2,s})\right| = \frac{\mathrm{d}^2}{\mathrm{d}t^2}\left|\widehat{Q}_{1,\min}(t_{1,s})\right| < 0, \quad s = 1, 2,$$

and $\widehat{Q}_{2,\min}(t)$ verifies (16). We conclude that μ_2 verifies the non-degenerate source condition of the sampling operator Φ_g, completing the proof of the translation invariance. $\blacksquare$

APPENDIX B
PROOF OF LEMMA 9

We derive the expression of the polynomial $Q_V(t) = a(t)^* \operatorname{diag}(\bar{g}) p_V$. By the orthogonal projection theorem, the solution p_V of the quadratic program (26) belongs to the subspace

$$p_V \in \operatorname{span}\left\{ \operatorname{diag}(g)a\left(\frac{\Delta}{2}\right), \operatorname{diag}(g)a\left(-\frac{\Delta}{2}\right),\right.$$
$$\left. \operatorname{diag}(g)\dot{a}\left(\frac{\Delta}{2}\right), \operatorname{diag}(g)\dot{a}\left(-\frac{\Delta}{2}\right) \right\}.$$

Recalling that $q_V = \operatorname{diag}(\bar{g})p_V$, it yields

$$q_V \in \operatorname{span}\left\{ \operatorname{diag}(|g|^2)a\left(\frac{\Delta}{2}\right), \operatorname{diag}(|g|^2)a\left(-\frac{\Delta}{2}\right),\right.$$
$$\left. \operatorname{diag}(|g|^2)\dot{a}\left(\frac{\Delta}{2}\right), \operatorname{diag}(|g|^2)\dot{a}\left(-\frac{\Delta}{2}\right) \right\},$$

where, with a slight abuse of notation, the operator $|\cdot|^2$ is interpreted element-wise. Remarkably, denoting by $K(t) = \sum_{k=-n}^{n} |g_k|^2 \, e^{i2\pi kt}$, we have that, for every $t, t' \in \mathbb{T}$.

$$a(t)^* \operatorname{diag}(|g|^2)a(t') = \sum_{k=-n}^{n} |g_k|^2 \, e^{2\pi ik(t-t')} = K(t-t')$$

$$a(t)^* \operatorname{diag}(|g|^2)\dot{a}(t') = -2\pi i \sum_{k=-n}^{n} k |g_k|^2 \, e^{2\pi ik(t-t')}$$
$$= -K'(t-t').$$

Therefore the polynomial $Q_V(t) = a(t)^* q_V$ lies in the span of the translations of the auto-correlation $K(t)$ and its derivative $K'(t)$ at the spike locations $\{-\Delta/2, \Delta/2\}$. Introducing a change of basis for the purpose of convenience, the polynomial $Q_V(t)$ writes for all $t \in \mathbb{T}$ as

$$\begin{aligned} Q_V(t) = \quad & \alpha_R\left(K(t-\frac{\Delta}{2}) + K(t+\frac{\Delta}{2})\right) \\ & - \beta_R\left(K'(t-\frac{\Delta}{2}) - K'(t+\frac{\Delta}{2})\right) \\ & + i\alpha_S\left(K(t-\frac{\Delta}{2}) - K(t+\frac{\Delta}{2})\right) \\ & - i\beta_S\left(K'(t-\frac{\Delta}{2}) + K'(t+\frac{\Delta}{2})\right), \end{aligned} \quad (45)$$

for some coefficients $\{\alpha_R, \beta_R, \alpha_S, \beta_S\} \subset \mathbb{C}$, which are fully determined by the four interpolation constraints of (26), that can be reinterpreted as

$$\frac{1}{2}\left(Q_V(\frac{\Delta}{2}) + Q_V(-\frac{\Delta}{2})\right) = \cos\left(\frac{\theta}{2}\right), \quad (46a)$$

$$\frac{1}{2}\left(Q'_V(\frac{\Delta}{2}) - Q'_V(-\frac{\Delta}{2})\right) = 0, \quad (46b)$$

$$\frac{1}{2i}\left(Q_V(\frac{\Delta}{2}) - Q_V(-\frac{\Delta}{2})\right) = \sin\left(\frac{\theta}{2}\right), \quad (46c)$$

$$\frac{1}{2i}\left(Q'_V(\frac{\Delta}{2}) + Q'_V(-\frac{\Delta}{2})\right) = 0. \quad (46d)$$

The linear system (46) can be reformulated in terms of the 4×4 block diagonal system

$$\begin{bmatrix} M_R & 0 \\ 0 & M_S \end{bmatrix} \begin{bmatrix} \alpha_R \\ \beta_R \\ \alpha_S \\ \beta_S \end{bmatrix} = \begin{bmatrix} \cos\left(\frac{\theta}{2}\right) \\ 0 \\ \sin\left(\frac{\theta}{2}\right) \\ 0 \end{bmatrix}, \quad (47)$$

where each of the diagonal blocks can be decomposed as

$$\begin{aligned} M_R &= \begin{bmatrix} K(0) + K(\Delta) & K'(\Delta) \\ K'(\Delta) & -K''(0) + K''(\Delta) \end{bmatrix} \\ &= A_R^* \operatorname{diag}(g)^* \operatorname{diag}(g) A_R, \end{aligned} \quad (48a)$$

$$\begin{aligned} M_S &= \begin{bmatrix} K(0) - K(\Delta) & -K'(\Delta) \\ -K'(\Delta) & -K''(0) - K''(\Delta) \end{bmatrix} \\ &= A_S^* \operatorname{diag}(g)^* \operatorname{diag}(g) A_S. \end{aligned} \quad (48b)$$

Here, the intermediate matrices $A_R, A_S \in \mathbb{C}^{N \times 2}$ are given by

$$A_R = 2^{-1/2}\left[a(\tfrac{\Delta}{2}) + a(-\tfrac{\Delta}{2}) \quad \dot{a}(\tfrac{\Delta}{2}) - \dot{a}(-\tfrac{\Delta}{2})\right],$$
$$A_S = 2^{-1/2}\left[a(\tfrac{\Delta}{2}) - a(-\tfrac{\Delta}{2}) \quad \dot{a}(\tfrac{\Delta}{2}) + \dot{a}(-\tfrac{\Delta}{2})\right].$$

Using the invertibility properties of Vandermonde matrices, it can be verified that the matrix $[A_R, A_S]$ and all its submatrices of size 4×4 are of the maximal rank whenever $0 < \Delta \le \frac{1}{2}$. Consequently, if g has four or more non-zero entries, the matrices $\operatorname{diag}(g)A_R$ and $\operatorname{diag}(g)A_S$ are also of the maximal rank whenever $0 < \Delta \le \frac{1}{2}$. We conclude, using the decomposition on the right-hand-side of (48), that the matrices M_R, M_S are positive definite Hermitian matrices, and that the linear system (47) has a unique solution. Solving this system leads to the desired expression of Q_V, where we let $C_R = \det(M_R) > 0$ and $C_S = \det(M_S) > 0$. ∎

APPENDIX C
PROOF OF LEMMA 10

Let by $\kappa = \mathcal{K}(g) \in L_2$ the autocorrelation of g. We start by recalling that, by the Wiener-Khinchin theorem, $\mathcal{F}(\kappa) = |G(\cdot)|^2 \in L_2$. It comes by (H2) and (H3) that the function $f \mapsto (i2\pi f)^\ell \mathcal{F}(\kappa)(f)$ is also band-limited within B and bounded for any $\ell \in \mathbb{N}$. Therefore it is absolutely integrable over $\mathbb{R}$ for any $\ell \in \mathbb{N}$. We conclude that κ possesses derivatives of all orders and that

$$\mathcal{F}(\kappa^{(\ell)}): \, f \mapsto (i2\pi f)^\ell \mathcal{F}(\kappa)(f), \quad \forall \ell \in \mathbb{N}.$$

Fix an odd integer $N = 2n + 1$. Using the definition (27) of K, we have for $\ell = 0, 1, 2, 3$ that

$$\begin{aligned} \left(\frac{B}{N}\right)^{\ell+1} K^{(\ell)}(t) &= \frac{B}{N} \sum_{k=-n}^{n} \left(i2\pi \frac{Bk}{N}\right)^\ell \left|G\left(\frac{Bk}{N}\right)\right|^2 e^{i2\pi kt} \\ &= \frac{B}{N} \sum_{k \in \mathbb{Z}} \left(i2\pi \frac{Bk}{N}\right)^\ell \left|G\left(\frac{Bk}{N}\right)\right|^2 e^{i2\pi kt} \\ &= \sum_{k \in \mathbb{Z}} \mathcal{F}\left(\kappa^{(\ell)}\left(\frac{N}{B}(t+\cdot)\right)\right)(k), \quad (49) \end{aligned}$$

where we used the fact that $G(\frac{Bk}{N}) = 0$ for all $|k| \ge n + 1$ by the hypothesis (H2) in the second equality, and identified

the expression with the Fourier transform of the derivatives of κ in the last equality. Hence, the series in the last equality converges absolutely and uniformly in t as it has finitely many non-zero terms. Moreover the series

$$\sum_{k \in \mathbb{Z}} \kappa^{(\ell)} \left(\frac{N}{B} (t+k) \right) = \kappa^{(\ell)} \left(\frac{N}{B} t \right) + \sum_{\substack{k \in \mathbb{Z} \\ k \neq 0}} \kappa^{(\ell)} \left(\frac{N}{B} (t+k) \right)$$

$$(50)$$

converges uniformly for $\ell = 0, 1, 2, 3$ as $\kappa^{(\ell)}$ is bounded and by the hypothesis (H4). Applying the Poisson summation formula between the right-hand-side of (49) and the left-hand-side of (50) yields

$$\left(\frac{B}{N} \right)^{\ell+1} K^{(\ell)}(t) = \sum_{k \in \mathbb{Z}} \mathcal{F} \left(\kappa^{(\ell)} \left(\frac{N}{B} (t+\cdot) \right) \right) (k)$$

$$= \sum_{k \in \mathbb{Z}} \kappa^{(\ell)} \left(\frac{N}{B} (t+k) \right). \qquad (51)$$

Substracting the term of index 0 in (51), taking the absolute value and the supremum over t on both sides of the equality yields

$$\sup_{t \in \mathbb{T}} \left| \left(\frac{B}{N} \right)^{\ell+1} K^{(\ell)}(t) - \kappa^{(\ell)} \left(\frac{Nt}{B} \right) \right|$$

$$= \sup_{t \in \mathbb{T}} \left| \sum_{\substack{k \in \mathbb{Z} \\ k \neq 0}} \kappa^{(\ell)} \left(\frac{N}{B} (t+k) \right) \right|.$$

We conclude on the desired statement by setting the limit $N \to \infty$ and using Assumption (H4). ∎

APPENDIX D
PROOF OF THEOREM 11

Similarly to the proof given in Section IV, we leverage the invariance of the non-degenerate source condition (proved in Lemma 7) defined in [6] and may restrict, without loss of generality, the analysis to measures $\mu_\star$ with support $\mathcal{T} = \{t_1, \ldots, t_S\}$ of the form

$$\mu_\star = e^{-i\frac{\theta}{2}} \delta \left(t - \Delta/2 \right) + e^{i\frac{\theta}{2}} \delta \left(t + \Delta/2 \right) + \sum_{s \geq 3} c_s \delta(t - t_s),$$

$$(52)$$

where we impose $t_1 = \Delta/2$ and $t_2 = -\Delta/2$ for some $\Delta > 0$, and $c_1 = e^{-i\frac{\theta}{2}}$ and $c_2 = e^{i\frac{\theta}{2}}$ for some $\theta \in [0, \pi]$. The minimal norm vanishing derivative polynomial Q_V associated to the measure $\mu_\star$ in (52) is given by $Q_V(t) = \boldsymbol{a}(t)^* \operatorname{diag}(\overline{\boldsymbol{g}}) \boldsymbol{p}_V$, where $\boldsymbol{p}_V \in \mathbb{C}^N$ is defined as the solution of the quadratic program

$$\boldsymbol{p}_V = \arg \min_{\boldsymbol{p} \in \mathbb{C}^N} \|\boldsymbol{p}\|_2$$

$$\text{subject to} \quad \boldsymbol{a}(t_s)^* \operatorname{diag}(\bar{\boldsymbol{g}}) \boldsymbol{p} = \operatorname{sgn}(c_s),$$

$$\dot{\boldsymbol{a}}(t_s)^* \operatorname{diag}(\bar{\boldsymbol{g}}) \boldsymbol{p} = 0, \quad s = 1, \ldots, S. \quad (53)$$

Invoking from [6] a generalization of Lemma 8 to the case of measures with an arbitrary number S of point sources, it is sufficient to show that $Q_V(t)$ verifies the conditions

$$|Q_V(t)| < 1, \quad \forall t \notin \mathcal{T}, \qquad (54a)$$

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2} |Q_V(t_s)| < 0, \quad s = 1, \ldots, S, \qquad (54b)$$

for a large enough Δ for any $N \geq N_0$ to conclude on the support stability of the Beurling-LASSO estimator for any measure $\mu_\star$ satisfying (42).

To proceed, we study, as in the proof of Theorem 6, the asymptotic behavior of the polynomial $Q_V(t)$ when $N \to \infty$. Define for any $s = 1, \ldots, S$, the subspace $E_s \subset \mathbb{C}^N$ as

$$E_s = \operatorname{span} \{ \operatorname{diag}(\boldsymbol{g}) \boldsymbol{a}(t_s), \operatorname{diag}(\boldsymbol{g}) \dot{\boldsymbol{a}}(t_s) \}, \qquad (55)$$

and denote by $F = E_1 + E_2$ the sum of the two subspaces associated with the two close point sources t_1 and t_2. It is clear, from an orthogonality projection argument, that $\boldsymbol{p}_V \in F + \sum_{s=3}^S E_s$. Let $\boldsymbol{p}_F \in F$ be the solution of the quadratic program (26) associated with the two close point sources t_1 and t_2. Similarly, for any $s \geq 3$ denote by $\boldsymbol{p}_{E_s} \in E_s$ the unique solution of the quadratic program

$$\boldsymbol{p}_{E_S} = \arg \min_{\boldsymbol{p} \in \mathbb{C}^N} \|\boldsymbol{p}\|_2$$

$$\text{subject to} \quad \boldsymbol{a}(t_s)^* \operatorname{diag}(\bar{\boldsymbol{g}}) \boldsymbol{p} = \operatorname{sgn}(c_s),$$

$$\dot{\boldsymbol{a}}(t_s)^* \operatorname{diag}(\bar{\boldsymbol{g}}) \boldsymbol{p} = 0, \qquad (56)$$

which is equal to

$$\boldsymbol{p}_{E_s} = \frac{\operatorname{sgn}(c_s)}{K(0)} \operatorname{diag}(\boldsymbol{g}) \boldsymbol{a}(t_s) \qquad (57)$$

for all $s = 3, \ldots, S$. We wish to show that $\boldsymbol{p}_V$ converges to $\boldsymbol{p}_F + \sum_{s=3}^S \boldsymbol{p}_{E_s}$ in the limit of N to get the asymptotic properties of $Q_V(t)$. To that end, we define the matrix $\boldsymbol{A}_{E_s} \in \mathbb{C}^{N \times 2}$ for $s = 1, \ldots, S$ as

$$\boldsymbol{A}_{E_s} = \left[\operatorname{diag}(\boldsymbol{g}) \boldsymbol{a}(t_s) / \sqrt{K(0)}, \operatorname{diag}(\boldsymbol{g}) \dot{\boldsymbol{a}}(t_s) / \sqrt{|K''(0)|} \right].$$

It is immediate that the columns of $\boldsymbol{A}_{E_s}$ form an orthonormal basis of E_s. Let $\boldsymbol{A}_F \in \mathbb{C}^{N \times 4}$ and $\boldsymbol{A} \in \mathbb{C}^{N \times 2S}$ be the concatenations

$$\boldsymbol{A}_F = [\boldsymbol{A}_{E_1}, \boldsymbol{A}_{E_2}],$$

$$\boldsymbol{A} = [\boldsymbol{A}_F, \boldsymbol{A}_{E_3}, \ldots, \boldsymbol{A}_{E_s}] = [\boldsymbol{A}_{E_1}, \boldsymbol{A}_{E_2}, \boldsymbol{A}_{E_3}, \ldots, \boldsymbol{A}_{E_s}],$$

and $\boldsymbol{w} = [\operatorname{sgn}(c_1), 0, \operatorname{sgn}(c_2), 0, \ldots \operatorname{sgn}(c_s), 0]^\top \in \mathbb{C}^{2S}$. By the orthogonal projection theorem, we have that

$$\boldsymbol{p}_F = K(0)^{-\frac{1}{2}} \boldsymbol{A}_F \left(\boldsymbol{A}_F^* \boldsymbol{A}_F \right)^{-1} [\operatorname{sgn}(c_1), 0, \operatorname{sgn}(c_2), 0]^\top,$$

$$(58a)$$

$$\boldsymbol{p}_{E_s} = K(0)^{-\frac{1}{2}} \boldsymbol{A}_{E_s} \left(\boldsymbol{A}_{E_s}^* \boldsymbol{A}_{E_s} \right)^{-1} [\operatorname{sgn}(c_s), 0]^\top, \quad s = 3, \ldots S,$$

$$(58b)$$

$$\boldsymbol{p}_V = K(0)^{-\frac{1}{2}} \boldsymbol{A} \left(\boldsymbol{A}^* \boldsymbol{A} \right)^{-1} \boldsymbol{w}. \qquad (58c)$$

Moreover, denote by $\boldsymbol{G} \in \mathbb{C}^{2S \times 2S}$ the matrix

$$\boldsymbol{G} = \begin{bmatrix} \boldsymbol{A}_F^* \boldsymbol{A}_F & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{A}_{E_3}^* \boldsymbol{A}_{E_3} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{A}_{E_s}^* \boldsymbol{A}_{E_s} \end{bmatrix},$$

then (58) implies that

$$
\boldsymbol{p}_V - \boldsymbol{p}_F - \sum_{s=3}^{S} \boldsymbol{p}_{E_s} = K(0)^{-1/2} \boldsymbol{A} \left((\boldsymbol{A}^* \boldsymbol{A})^{-1} - \boldsymbol{G}^{-1} \right) \boldsymbol{w}.
$$
(59)

Next, we show that the difference $(\boldsymbol{A}^* \boldsymbol{A})^{-1} - \boldsymbol{G}^{-1}$ is small in norm when $N \to \infty$. To proceed, we start by noticing that

$$
\boldsymbol{A}^* \boldsymbol{A} - \boldsymbol{G} = \begin{bmatrix} \boldsymbol{0} & \boldsymbol{A}_F^* \boldsymbol{A}_{E_3} & \cdots & \boldsymbol{A}_F^* \boldsymbol{A}_{E_s} \\ \boldsymbol{A}_{E_3}^* \boldsymbol{A}_F & \boldsymbol{0} & \cdots & \boldsymbol{A}_{E_3}^* \boldsymbol{A}_{E_s} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{A}_{E_s}^* \boldsymbol{A}_F & \boldsymbol{A}_{E_s}^* \boldsymbol{A}_{E_3} & \cdots & \boldsymbol{0} \end{bmatrix}.
$$

Denote by $\|\boldsymbol{M}\|_{\max} = \max_{i,j} |m_{i,j}|$ the max-norm of $\boldsymbol{M}$, which is the maximum of the absolute values of the entries $m_{i,j}$ of $\boldsymbol{M}$. A direct calculation establishes that each off-diagonal block of $\boldsymbol{A}^* \boldsymbol{A} - \boldsymbol{G}$ verifies for any $s, s' \geq 3$ with $s \neq s'$

$$
\|\boldsymbol{A}_F^* \boldsymbol{A}_{E_s}\|_{\max} = \max \left\{ \left| \frac{K(t_s \pm \Delta/2)}{K(0)} \right|, \frac{|K'(t_s \pm \Delta/2)|}{\sqrt{|K(0)K''(0)|}}, \right.
$$
$$
\left. \left| \frac{K''(t_s \pm \Delta/2)}{K''(0)} \right| \right\},
$$
(60a)

$$
\left\| \boldsymbol{A}_{E_{s'}}^* \boldsymbol{A}_{E_s} \right\|_{\max} = \max \left\{ \left| \frac{K(t_s - t_{s'})}{K(0)} \right|, \frac{|K'(t_s - t_{s'})|}{\sqrt{|K(0)K''(0)|}}, \right.
$$
$$
\left. \left| \frac{K''(t_s - t_{s'})}{K''(0)} \right| \right\}.
$$
(60b)

Taking the limit $N \to \infty$ in (60a) yields

$$
\lim_{N \to \infty} \|\boldsymbol{A}_F^* \boldsymbol{A}_{E_s}\|_{\max}
$$
$$
= \lim_{N \to \infty} \max \left\{ \left| \frac{\kappa(Nt_s/B \pm \gamma/2)}{\kappa(0)} \right|, \right.
$$
$$
\left. \frac{|\kappa'(Nt_s/B \pm \gamma/2)|}{\sqrt{|\kappa(0)\kappa''(0)|}}, \left| \frac{\kappa''(Nt_s/B \pm \gamma/2)}{\kappa''(0)} \right| \right\}
$$
$$
= \max \left\{ \lim_{|\tau| \to \infty} \left| \frac{\kappa(\tau)}{\kappa(0)} \right|, \right.
$$
$$
\left. \lim_{|\tau| \to \infty} \left| \frac{\kappa'(\tau)}{\sqrt{|\kappa(0)\kappa''(0)|}} \right|, \lim_{|\tau| \to \infty} \left| \frac{\kappa''(\tau)}{\kappa''(0)} \right| \right\}
$$
$$
= 0,
$$

where the first equality comes from Lemma 10, the second equality from the assumption (42) which implies $|Nt_s/B \pm \gamma/2| \to \infty$, $s = 3, \ldots, S$, when $N \to \infty$, and the last equality holds under assumptions (H2) and (H3) by the application of the Riemann-Lebesgue lemma on the integrable functions $\mathcal{F}\{\kappa^{(\ell)}\}(f) = (i2\pi f)^\ell |G(f)|^2$, $\ell = 0, 1, 2$. An analogous reasoning on (60b) leads to $\lim_{N \to \infty} \left\| \boldsymbol{A}_{E_{s'}}^* \boldsymbol{A}_{E_s} \right\|_{\max} = 0$, for all $s, s' \geq 3$ and $s \neq s'$. We conclude that

$$
\lim_{N \to \infty} \|\boldsymbol{A}^* \boldsymbol{A} - \boldsymbol{G}\|_{\max} = 0.
$$
(61)

Next, by Lemma 10, the matrix $\boldsymbol{G}$ admits an invertible limit, and we denote $\overline{\boldsymbol{G}} = \lim_{N \to \infty} \boldsymbol{G}$. By continuity of the inversion $\boldsymbol{M} \mapsto \boldsymbol{M}^{-1}$ at $\overline{\boldsymbol{G}}$, we have $\lim_{N \to \infty} \boldsymbol{G}^{-1} = \overline{\boldsymbol{G}}^{-1}$. Moreover, from (61) we have

$$
\lim_{N \to \infty} \left\| \boldsymbol{A}^* \boldsymbol{A} - \overline{\boldsymbol{G}} \right\|_{\max}
$$
$$
\leq \lim_{N \to \infty} \left(\|\boldsymbol{A}^* \boldsymbol{A} - \boldsymbol{G}\|_{\max} + \|\boldsymbol{G} - \overline{\boldsymbol{G}}\|_{\max} \right) = 0,
$$

implying by continuity that $\lim_{N \to \infty} (\boldsymbol{A}^* \boldsymbol{A})^{-1} = \overline{\boldsymbol{G}}^{-1}$. Let by $\|\cdot\|_2$ and $\|\cdot\|_F$ the spectral norm and the Frobenius norm of a matrix, respectively. We have by the previous arguments that

$$
\lim_{N \to \infty} \left\| (\boldsymbol{A}^* \boldsymbol{A})^{-1} - \boldsymbol{G}^{-1} \right\|_2
$$
$$
\leq \lim_{N \to \infty} \left(\left\| (\boldsymbol{A}^* \boldsymbol{A})^{-1} - \overline{\boldsymbol{G}}^{-1} \right\|_2 + \left\| \overline{\boldsymbol{G}}^{-1} - \boldsymbol{G}^{-1} \right\|_2 \right)
$$
$$
= 0,
$$

since S is finite and does not grow with N. This, together with (59), yields

$$
\left\| \boldsymbol{p}_V - \boldsymbol{p}_F - \sum_{s=3}^{S} \boldsymbol{p}_{E_s} \right\|_2
$$
$$
= K(0)^{-1/2} \left\| \boldsymbol{A} \left((\boldsymbol{A}^* \boldsymbol{A})^{-1} - \boldsymbol{G}^{-1} \right) \boldsymbol{w} \right\|_2
$$
$$
\leq K(0)^{-1/2} \|\boldsymbol{A}\|_F \left\| (\boldsymbol{A}^* \boldsymbol{A})^{-1} - \boldsymbol{G}^{-1} \right\|_2 \|\boldsymbol{w}\|_2
$$
$$
= \sqrt{2} S K(0)^{-1/2} \left\| (\boldsymbol{A}^* \boldsymbol{A})^{-1} - \boldsymbol{G}^{-1} \right\|_2
$$
$$
\leq \sqrt{2} S \left(\frac{N\kappa(0)}{B} + o(N) \right)^{-1/2} \left\| (\boldsymbol{A}^* \boldsymbol{A})^{-1} - \boldsymbol{G}^{-1} \right\|_2
$$
$$
\leq o(N^{-1/2}), \quad N \to \infty,
$$
(62)

where the third line follows from $\|\boldsymbol{w}\|_2 = \sqrt{S}$ and $\|\boldsymbol{A}\|_F = \sqrt{2S}$, and the fourth line follows from Lemma 10.

Denote by $Q_F(t)$ and $Q_{E_s}(t)$ for any $s \geq 3$ the trigonometric polynomials

$$
Q_F(t) = \boldsymbol{a}(t)^* \operatorname{diag}(\overline{\boldsymbol{g}}) \boldsymbol{p}_F,
$$
$$
Q_{E_s}(t) = \boldsymbol{a}(t)^* \operatorname{diag}(\overline{\boldsymbol{g}}) \boldsymbol{p}_{E_s} = \frac{\operatorname{sgn}(c_s)}{K(0)} K(t - t_s),
$$

for any $t \in \mathbb{T}$ and $s = 3, \ldots, S$. We are now ready to derive the asymptotic behavior of $Q_V(t)$. First of all, for $\ell = 0, 1, 2$, we have that

$$
\sup_{t \in \mathbb{T}} \left| Q_V^{(\ell)}(t) - Q_F^{(\ell)}(t) - \sum_{s=3}^{S} Q_{E_s}^{(\ell)}(t) \right|
$$
$$
= \sup_{t \in \mathbb{T}} \left| \boldsymbol{a}^{(\ell)}(t)^* \operatorname{diag}(\overline{\boldsymbol{g}}) \left(\boldsymbol{p}_V - \boldsymbol{p}_F - \sum_{s=3}^{S} \boldsymbol{p}_{E_s} \right) \right|
$$
$$
\leq \sup_{t \in \mathbb{T}} \left\| \operatorname{diag}(\boldsymbol{g}) \boldsymbol{a}^{(\ell)}(t)^* \right\|_2 \left\| \boldsymbol{p}_V - \boldsymbol{p}_F - \sum_{s=3}^{S} \boldsymbol{p}_{E_s} \right\|_2
$$
$$
\leq \|\boldsymbol{g}\|_\infty \sup_{t \in \mathbb{T}} \left\| \boldsymbol{a}^{(\ell)}(t) \right\|_2 \left\| \boldsymbol{p}_V - \boldsymbol{p}_F - \sum_{s=3}^{S} \boldsymbol{p}_{E_s} \right\|_2
$$
$$
= \|\boldsymbol{g}\|_\infty (2\pi)^\ell \sqrt{\sum_{k=-n}^{n} k^{2\ell}} \left\| \boldsymbol{p}_V - \boldsymbol{p}_F - \sum_{s=3}^{S} \boldsymbol{p}_{E_s} \right\|_2
$$

$$\leq \|g\|_\infty \, \mathcal{O}\left(N^{\ell+\frac{1}{2}}\right) o\left(N^{-\frac{1}{2}}\right)$$
$$\leq o\left(N^\ell\right), \qquad N \to \infty, \tag{63}$$

where we used in the last inequality assumption (H3) to bound

$$\|g\|_\infty = \max\left\{\left|G\left(\frac{kB}{N}\right)\right|\right\}_{k=-n}^n \leq \sup_{f\in[-\frac{B}{2},\frac{B}{2}]} |G(f)| < \infty.$$

Next, we denote by $\mathcal{Q}_V(\tau)$ the function

$$\mathcal{Q}_V(\tau) = \mathcal{Q}_F(\tau) + \sum_{s=3}^S \mathcal{Q}_{E_s}(\tau), \quad \forall \tau \in \mathbb{R},$$

where $\mathcal{Q}_F(\tau)$ is the function given in (33) and $\mathcal{Q}_{E_s}(\tau) = \frac{\mathrm{sgn}(c_k)}{\kappa(0)}\kappa(\tau - \frac{Nt_s}{B})$ for any $s \geq 3$. It follows that

$$\sup_{t\in\mathbb{T}}\left|\left(\frac{B}{N}\right)^\ell Q_V^{(\ell)}(t) - \mathcal{Q}_V^{(\ell)}\left(\frac{Nt}{B}\right)\right|$$

$$= \sup_{t\in\mathbb{T}}\left|\left(\frac{B}{N}\right)^\ell Q_V^{(\ell)}(t) - \mathcal{Q}_F^{(\ell)}\left(\frac{Nt}{B}\right) - \sum_{s=3}^S \mathcal{Q}_{E_s}^{(\ell)}\left(\frac{Nt}{B}\right)\right|$$

$$\leq \left(\frac{B}{N}\right)^\ell \sup_{t\in\mathbb{T}}\left|Q_V^{(\ell)}(t) - Q_F^{(\ell)}(t) - \sum_{s=3}^S Q_{E_s}^{(\ell)}(t)\right|$$

$$+ \sup_{t\in\mathbb{T}}\left|\left(\frac{B}{N}\right)^\ell Q_F^{(\ell)}(t) - \mathcal{Q}_F^{(\ell)}\left(\frac{Nt}{B}\right)\right|$$

$$+ \sum_{s=3}^S \sup_{t\in\mathbb{T}}\left|\left(\frac{B}{N}\right)^\ell Q_{E_s}^{(\ell)}(t) - \mathcal{Q}_{E_s}^{(\ell)}\left(\frac{Nt}{B}\right)\right|$$

$$\leq o(1), \qquad N \to \infty, \tag{64}$$

where the first inequality follows from the triangle inequality, and the second inequality follows from the fact that each term in the line above is controlled by (63) and (32) and Lemma 10, respectively.

With (64) in place, similarly to the proof presented in Section IV, we argue that the non-degenerate source condition is verified for any $N \geq N_0$ for N_0 sufficiently large if the limit function $\mathcal{Q}_V(\tau)$ meets the conditions

$$\lim_{N\to\infty} \frac{\mathrm{d}^2 |\mathcal{Q}_V|}{\mathrm{d}\tau^2}\left(\frac{Nt_s}{B}\right) < 0, \quad s = 1, \ldots, S, \tag{65a}$$

$$\lim_{N\to\infty}\left|\mathcal{Q}_V\left(\frac{Nt}{B}\right)\right| < 1, \quad \forall t \notin \{t_s\}. \tag{65b}$$

In order to verify (65), we start by noticing that for any $\tau \in \mathbb{R}$ and $\ell = 0,1,2$,

$$\lim_{N\to\infty}\left|\mathcal{Q}_V^{(\ell)}(\tau)\right| \leq \lim_{N\to\infty}\left|\mathcal{Q}_F^{(\ell)}(\tau)\right| + \sum_{s=3}^S \lim_{N\to\infty}\left|\mathcal{Q}_{E_s}^{(\ell)}(\tau)\right|. \tag{66}$$

The condition (65a) can be verified by picking $t_{s_\star} \in \{t_1, \ldots, t_S\}$ and distinguishing two cases:
- If $s_\star = 1, 2$, and we have that for $\ell = 0, 1, 2$ and for $s = 3, \ldots S$ that

$$\lim_{N\to\infty}\left|\mathcal{Q}_{E_s}^{(\ell)}\left(\frac{Nt_{s_\star}}{B}\right)\right| = \lim_{N\to\infty}\left|\frac{\kappa^{(\ell)}\left(\frac{N}{B}(t_{s_\star} - t_s)\right)}{\kappa(0)}\right|$$

$$= \lim_{|\tau|\to\infty}\left|\frac{\kappa^{(\ell)}(\tau)}{\kappa(0)}\right| = 0, \tag{67}$$

where we used the separation assumption and the Riemann-Lebesgue lemma on the integrable functions $\mathcal{F}^{(\ell)}(\kappa)$, $\ell = 0,1,2$, in (67). This leads to

$$\lim_{N\to\infty}\frac{\mathrm{d}^2 |\mathcal{Q}_V|}{\mathrm{d}\tau^2}\left(\frac{Nt_{s_\star}}{B}\right) = \frac{\mathrm{d}^2 |\mathcal{Q}_F|}{\mathrm{d}\tau^2}\left(\frac{Nt_{s_\star}}{B}\right) < 0$$

by (66), and by going to the limit $N \to \infty$ in the expression (39), and invoking the derivations on $\frac{\mathrm{d}^2 |\mathcal{Q}_V|}{\mathrm{d}\tau^2}$ presented in Section IV-B under the separation condition (42).
- If $s_\star = 3, \ldots, S$, a similar reasoning yields that for any $\ell = 0,1,2$ and any $s = 3, \ldots, S$

$$\lim_{N\to\infty}\left|\mathcal{Q}_{E_s}^{(\ell)}\left(\frac{Nt_{s_\star}}{B}\right)\right| = \lim_{N\to\infty}\left|\frac{\kappa^{(\ell)}\left(\frac{N}{B}(t_{s_\star} - t_s)\right)}{\kappa(0)}\right|$$

$$= \lim_{|\tau|\to\infty}\left|\frac{\kappa^{(\ell)}(\tau)}{\kappa(0)}\right| = 0,$$

$$\lim_{N\to\infty}\left|\mathcal{Q}_F^{(\ell)}\left(\frac{Nt_{s_\star}}{B}\right)\right| = \lim_{|\tau|\to\infty}\left|\mathcal{Q}_F^{(\ell)}(\tau)\right| = 0.$$

We analogously obtain that

$$\lim_{N\to\infty}\frac{\mathrm{d}^2 |\mathcal{Q}_V|}{\mathrm{d}\tau^2}\left(\frac{Nt_{s_\star}}{B}\right) = \frac{\mathrm{d}^2 |\mathcal{Q}_{E_{s_\star}}|}{\mathrm{d}\tau^2}\left(\frac{Nt_{s_\star}}{B}\right)$$

$$= \frac{\kappa''(0)}{\kappa(0)} < 0.$$

Finally, to show that (65b) holds, for any $t \in \mathbb{T}\backslash\{t_s\}$, we denote by $s_\star$ the index of the closest point source in the sense of the wrap-around distance

$$t_{s_\star} = \min_{s=1,\ldots,S}|t - t_s|_\mathbb{T}.$$

Here again, we distinguish two cases:
- If $s_\star = 1, 2$, we have for any $s = 3, \ldots, S$ that

$$\lim_{N\to\infty}\left|\mathcal{Q}_{E_s}\left(\frac{Nt}{B}\right)\right| = \lim_{N\to\infty}\left|\frac{\kappa\left(\frac{N}{B}(t - t_s)\right)}{\kappa(0)}\right|$$

$$= \lim_{|\tau|\to\infty}\left|\frac{\kappa(\tau)}{\kappa(0)}\right| = 0, \tag{68a}$$

$$\lim_{N\to\infty}\left|\mathcal{Q}_F\left(\frac{Nt}{B}\right)\right| < 1, \tag{68b}$$

where we used the Riemann-Lebesgue Lemma on $\mathcal{F}(\kappa)$ in (68a), and by invoking the derivations on $\mathcal{Q}_F$ of Section IV-B under the separation condition (42) in (68b).
- If $s_\star = 3, \ldots S$, we have for any $s = 3, \ldots, S$ and $s \neq s^\star$

$$\lim_{N\to\infty}\left|\mathcal{Q}_{E_s}\left(\frac{Nt}{B}\right)\right| = \lim_{N\to\infty}\left|\frac{\kappa\left(\frac{N}{B}(t - t_s)\right)}{\kappa(0)}\right|$$

$$= \lim_{|\tau|\to\infty}\left|\frac{\kappa(\tau)}{\kappa(0)}\right| = 0, \tag{69a}$$

$$\lim_{N\to\infty}\left|\mathcal{Q}_{E_{s_\star}}\left(\frac{Nt}{B}\right)\right| = \lim_{N\to\infty}\left|\frac{\kappa\left(\frac{N}{B}(t - t_{s_\star})\right)}{\kappa(0)}\right| < 1, \tag{69b}$$

$$\lim_{N \to \infty} \left| \mathcal{Q}_F \left(\frac{Nt}{B} \right) \right| = \lim_{|\tau| \to \infty} |\mathcal{Q}_F(\tau)| = 0. \quad (69c)$$

Summing all the terms in both cases, and invoking (66) concludes on (65b).

In conclusion, under the conditions of Theorem 11, there must exists $N_0 \in \mathbb{N}$ such that the polynomial $\mathcal{Q}_V(\tau)$ verifies the conditions (54). We conclude that the measure $\mu_\star$ verifies the non-degenerate source condition, implying the support stability of the Beurling-LASSO estimator. ∎

References

[1] J. Lindberg, "Mathematical concepts of optical superresolution," *J. Opt.*, vol. 14, no. 8, Aug. 2012, Art. no. 083001.

[2] E. J. Candés and C. Fernandez-Granda, "Towards a mathematical theory of super-resolution," *Commun. Pure Appl. Math.*, vol. 67, no. 6, pp. 906–956, Jun. 2014.

[3] Y. Chi, A. Pezeshki, R. Calderbank, and S. Howard, "Range sidelobe suppression in a desired Doppler interval," in *Proc. Int. Waveform Diversity Design Conf.*, Orlando, FL, USA, Feb. 2009, pp. 258–262.

[4] S.-H. Lee, J. Y. Shin, A. Lee, and C. Bustamante, "Counting single photoactivatable fluorescent molecules by photoactivated localization microscopy (PALM)," *Proc. Nat. Acad. Sci. USA*, vol. 109, no. 43, pp. 17436–17441, Oct. 2012.

[5] C. Fernandez-Granda, "Super-resolution of point sources via convex programming," *Inf. Inference*, vol. 5, no. 3, pp. 251–303, Sep. 2016.

[6] V. Duval and G. Peyré, "Exact support recovery for sparse spikes deconvolution," *Found. Comput. Math.*, vol. 15, no. 5, pp. 1315–1355, Oct. 2015.

[7] Y. Chi, L. L. Scharf, A. Pezeshki, and A. R. Calderbank, "Sensitivity to basis mismatch in compressed sensing," *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 2182–2195, May 2011.

[8] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via 1 minimization," *Proc. Nat. Acad. Sci. USA*, vol. 100, no. 5, pp. 2197–2202, Mar. 2003.

[9] V. Chandrasekaran, B. Recht, P. A. Parrilo, and A. S. Willsky, "The convex geometry of linear inverse problems," *Found. Comput. Math.*, vol. 12, no. 6, pp. 805–849, Dec. 2012.

[10] Y. Chi and M. F. Da Costa, "Harnessing sparsity over the continuum: Atomic norm minimization for superresolution," *IEEE Signal Process. Mag.*, vol. 37, no. 2, pp. 39–57, Mar. 2020.

[11] B. N. Bhaskar, G. Tang, and B. Recht, "Atomic norm denoising with applications to line spectral estimation," *IEEE Trans. Signal Process.*, vol. 61, no. 23, pp. 5987–5999, Dec. 2013.

[12] Y. de Castro and F. Gamboa, "Exact reconstruction using Beurling minimal extrapolation," *J. Math. Anal. Appl.*, vol. 395, no. 1, pp. 336–354, Nov. 2012.

[13] M. Born and E. Wolf, *Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light*. Cambridge, U.K.: Cambridge Univ. Press, 1999.

[14] A. Moitra, "Super-resolution, extremal functions and the condition number of vandermonde matrices," in *Proc. 47th Annu. ACM Symp. Theory Comput. (STOC)*, 2015, pp. 821–830.

[15] C. Aubel, D. Stotz, and H. Bölcskei, "A theory of super-resolution from short-time Fourier transform measurements," *J. Fourier Anal. Appl.*, vol. 24, no. 1, pp. 45–107, Feb. 2018.

[16] D. Batenkov, L. Demanet, G. Goldman, and Y. Yomdin, "Stability of partial Fourier matrices with clustered nodes," 2018, *arXiv:1809.00658*. [Online]. Available: http://arxiv.org/abs/1809.00658

[17] B. Diederichs, "Well-posedness of sparse frequency estimation," 2019, *arXiv:1905.08005*. [Online]. Available: http://arxiv.org/abs/1905.08005

[18] D. Slepian and H. O. Pollak, "Prolate spheroidal wave functions, Fourier analysis and uncertainty—I," *Bell Syst. Tech. J.*, vol. 40, no. 1, pp. 43–63, 1961.

[19] I. C. Moore and M. Cada, "Prolate spheroidal wave functions, an introduction to the slepian series and its properties," *Appl. Comput. Harmon. Anal.*, vol. 16, no. 3, pp. 208–230, May 2004.

[20] E. J. Candès and C. Fernandez-Granda, "Super-resolution from noisy data," *J. Fourier Anal. Appl.*, vol. 19, no. 6, pp. 1229–1254, Dec. 2013.

[21] M. F. Da Costa and W. Dai, "A tight converse to the spectral resolution limit via convex programming," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2018, pp. 901–905.

[22] D. L. Donoho, "Superresolution via sparsity constraints," *SIAM J. Math. Anal.*, vol. 23, no. 5, pp. 1309–1331, Sep. 1992.

[23] M. Shahram and P. Milanfar, "Statistical and information-theoretic analysis of resolution in imaging," *IEEE Trans. Inf. Theory*, vol. 52, no. 8, pp. 3411–3437, Aug. 2006.

[24] D. Batenkov, G. Goldman, and Y. Yomdin, "Super-resolution of near-colliding point sources," 2019, *arXiv:1904.09186*. [Online]. Available: http://arxiv.org/abs/1904.09186

[25] L. Demanet and N. Nguyen, "The recoverability limit for super-resolution via sparsity," 2015, *arXiv:1502.01385*. [Online]. Available: http://arxiv.org/abs/1502.01385

[26] S. T. Smith, "Statistical resolution limits and the complexified Cramér-Rao bound," *IEEE Trans. Signal Process.*, vol. 53, no. 5, pp. 1597–1609, May 2005.

[27] W. Liao and A. Fannjiang, "MUSIC for single-snapshot spectral estimation: Stability and super-resolution," *Appl. Comput. Harmon. Anal.*, vol. 40, no. 1, pp. 33–67, Jan. 2016.

[28] C. Fernandez-Granda, G. Tang, X. Wang, and L. Zheng, "Demixing sines and spikes: Robust spectral super-resolution in the presence of outliers," *Inf. Inference, J. IMA*, vol. 7, no. 1, pp. 105–168, Mar. 2018.

[29] V. Duval, "A characterization of the non-degenerate source condition in super-resolution," *Inf. Inference, J. IMA*, vol. 9, no. 1, pp. 235–269, Mar. 2019.

[30] Q. Denoyelle, V. Duval, and G. Peyré, "Support recovery for sparse super-resolution of positive measures," *J. Fourier Anal. Appl.*, vol. 23, no. 5, pp. 1153–1194, Oct. 2017.

[31] C. Fernandez-Granda, "Support detection in super-resolution," in *Proc. 10th Int. Conf. Sampling Theory Appl. (SampTA)*, 2013, pp. 145–148.

[32] G. Tang, B. N. Bhaskar, and B. Recht, "Near minimax line spectral estimation," *IEEE Trans. Inf. Theory*, vol. 61, no. 1, pp. 499–512, Jan. 2015.

[33] V. I. Morgenshtern and E. J. Candès, "Super-resolution of positive sources: The discrete setup," *SIAM J. Imag. Sci.*, vol. 9, no. 1, pp. 412–444, Jan. 2016.

[34] Q. Li and G. Tang, "Approximate support recovery of atomic line spectral estimation: A tale of resolution and precision," *Appl. Comput. Harmon. Anal.*, vol. 48, no. 3, pp. 891–948, May 2020.

[35] T. Bendory, S. Dekel, and A. Feuer, "Robust recovery of stream of pulses using convex optimization," *J. Math. Anal. Appl.*, vol. 442, no. 2, pp. 511–536, Oct. 2016.

[36] G. Tang, B. N. Bhaskar, P. Shah, and B. Recht, "Compressed sensing off the grid," *IEEE Trans. Inf. Theory*, vol. 59, no. 11, pp. 7465–7490, Nov. 2013.

[37] Y. Chi and Y. Chen, "Compressive two-dimensional harmonic retrieval via atomic norm minimization," *IEEE Trans. Signal Process.*, vol. 63, no. 4, pp. 1030–1042, Feb. 2015.

[38] Y. Li and Y. Chi, "Off-the-grid line spectrum denoising and estimation with multiple measurement vectors," *IEEE Trans. Signal Process.*, vol. 64, no. 5, pp. 1257–1269, Mar. 2016.

Maxime Ferreira Da Costa (Member, IEEE) received the M.Sc. degree in signal processing from Imperial College London, U.K., the Diplôme d'Ingénieur degree from Supélec, France, in 2012, and the Ph.D. degree in electrical and electronic engineering from Imperial College London in 2018. He is currently a Research Associate with the Department of Electrical and Computer Engineering, Carnegie Mellon University, Pennsylvania. His research focuses on mathematical signal processing, inverse problems, and optimization. In 2018, he was shortlisted among the finalists for the Jack Keil Wolf Student Paper Award at the IEEE International Symposium on Information Theory.

Yuejie Chi (Senior Member, IEEE) received the B.E. degree (Hon.) in electrical engineering from Tsinghua University, Beijing, China, in 2007, and the Ph.D. degree in electrical engineering from Princeton University in 2012. She was with The Ohio State University from 2012 to 2017. Since 2018, she has been an Associate Professor with the Department of Electrical and Computer Engineering, Carnegie Mellon University, where she holds the Robert E. Doherty Early Career Development Professorship. Her research interests include signal processing, statistical inference, machine learning, large-scale optimization, and their applications in data science, inverse problems, imaging, and sensing systems. She has served as an Elected Member of the SPTM, SAM, and MLSP Technical Committees of the IEEE Signal Processing Society. She was a recipient of the PECASE Award, the NSF CAREER Award, the AFOSR and ONR Young Investigator Program Awards, the Ralph E. Powe Junior Faculty Enhancement Award, and the Google Faculty Research Award. She received the IEEE SPS Early Career Technical Achievement Award and the Young Author Best Paper Award from IEEE Signal Processing Society, and the Best Paper Award at the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP). She serves as an Associate Editor of the IEEE Transactions on Signal Processing.