NON-NEGATIVE SUPER-RESOLUTION IS STABLE

Armin Eftekhari[†‡], *Jared Tanner*[†§], *Andrew Thompson*[§], *Bogdan Toader*[§], *Hemant Tyagi*[†‡]

[†]Alan Turing Institute, [‡]University of Edinburgh, [§]University of Oxford

ABSTRACT

We consider the problem of localizing point sources on an interval from possibly noisy measurements. In the absence of noise, we show that measurements from Chebyshev systems are an injective map for non-negative sparse measures, and therefore non-negativity is sufficient to ensure uniqueness for sparse measures. Moreover, we characterize non-negative solutions from inexact measurements and show that any non-negative solution consistent with the measurements is proportionally close to the solution of the system with exact measurements. Our results substantially simplify, extend, and generalize the prior work by De Castro et al. [1] and Schiebinger et al. [2], which relies upon sparsifying penalties, by showing that it is the non-negativity constraint, rather than any particular algorithm, that imposes uniqueness of the sparse non-negative measure, and by extending the results to inexact samples.

Index Terms— Super-resolution, non-negative sparse measures, feasibility programs, Chebyshev systems.

1. PROBLEM SETUP

Consider an unknown number of point sources, with unknown locations and amplitudes. A sensing mechanism provides us with a few (possibly noisy) measurements, from which we wish to estimate the locations and amplitudes of these sources. Because of the finite *resolution* or *bandwidth* of the imaging device, poorly separated sources may be visually indistinguishable from the measurements, but they can be exactly identified by taking the signal model into account. This *super-resolution* problem of localizing point sources finds various applications in astronomy [3], imaging in chemistry, medicine and neuroscience [4, 5, 6, 7, 8, 9, 10, 11], spectral estimation [12, 13], geophysics [14], and system identification [15]. The rich literature of super-resolution which is mostly closely tied to the results herein are briefly reviewed in Section 1.1. In this paper, we study the "grid-free" and non-negative super-resolution in the presence of inexact samples in one dimension.

To be concrete, consider a *non-negative* Borel measure x supported on the interval $I = [0, 1] \subset \mathbb{R}$. While we often consider a general measure x, we typically compare these measures with non-negative discrete measures of finite support such as

$$x = \sum_{i=1}^{k} a_i \cdot \delta_{t_i} \quad \text{with} \;\; a_i > 0 \; \forall i. \tag{1}$$

Consider also real-valued and continuous functions $\{\phi_j\}_{j=1}^m$ and let $\{y_j\}_{j=1}^m$ be the possibly noisy measurements collected from x by convolving against sampling functions $\phi_j(t)$:

$$y_j = \int_I \phi_j(t)x(\mathrm{d}t) + \eta_j, \tag{2}$$

where η_j can represent additive noise, with $\|\eta\|_2 \leqslant \delta$. Herein we characterize non-negative measures consistent with measurements (2) in relation to the discrete measure (1). That is, we consider any non-negative Borel measure z from the Program (3) [1]

$$\text{find } z \geqslant 0 \text{ subject to } \sum_{j=1}^{m} \left| y_j - \int_I \phi_j(t)z(\mathrm{d}t) \right|^2 \leqslant \delta^2, \tag{3}$$

and show that any such z is close to x from (1) in an appropriate metric, see Theorems 2 and 5.

1.1. Comparison with other techniques

Our contribution builds on a growing literature of super-resolution, but differs primarily in that Program (3) does not implicitly impose a sparsifying penalty beyond non-negativity. In particular, the majority of results in this area consider Program (3) augmented to include minimizing $\int_I |z(\mathrm{d}t)|$, the total-variation (TV) of z, see for instance [16, 17, 18, 19, 20, 21, 22, 23] as examples of such approaches. Our focus on x which is non-negative builds on

This publication is based on work supported by the EPSRC Centre For Doctoral Training in Industrially Focused Mathematical Modelling (EP/L015803/1) in collaboration with the National Physical Laboratory and by the Alan Turing Institute under the EPSRC grant EP/N510129/1 and the Turing Seed Funding grant SF019. The authors thank Stephane Chretien, Alistair Forbes and Peter Harris from the National Physical Laboratory for their insightful contributions which motivated this work through applications in maritime sensing.

[1] An equivalent formulation of Program (3) minimizes $\|y - \int_I \Phi(t)z(\mathrm{d}t)\|_2$ over all non-negative measures on I (without any constraints). In this context, however, we find it somewhat more intuitive to work with Program (3), particularly considering the importance of the case $\delta = 0$.

[1, 2] which proved that no minimum separation condition was needed to ensure exact reconstruction in the noiseless case using TV norm minimization.

In contrast to all prior work, we show in this paper that it is possible to dispense with the TV norm altogether in the case of non-negative measures. That is, instead of searching for the sparsest, one can look for *any* solution consistent with the measurements and it must be similarly proportionally close to the noise free discrete sparse measure which generates noise free measurements. A more thorough discussion of prior work is given in [24].

2. STABILITY FOR GAUSSIAN WINDOW

Recall that $I = [0,1]$ and consider a sequence of source and sample locations T and S respectively as given in (4):

$$T = \{t_i\}_{i=1}^k \subset \text{int}(I) \quad \text{and } S = \{s_j\}_{j=1}^m \subseteq I, \tag{4}$$

with arbitrary k-sparse non-negative measure x supported on T, namely x as given in (1), sampled by windows

$$\phi_j(t) = g(t - s_j) = e^{-\frac{(t-s_j)^2}{\sigma^2}}. \tag{5}$$

The Gaussian window (5) can be interpreted as the "point spread function" of the sensing mechanism at location s_j and the set S as the "sampling points" in the sense that

$$y(s_j) = \int_I \phi_j(t)x(\mathrm{d}t) = \int_I g(t - s_j)x(\mathrm{d}t), \tag{6}$$

for all $j \in [m]$, where $[m] = \{1, 2, \dots, m\}$. The stability of Program (3) is determined by the source and sample configuration as given in (4), the relative number of sources to samples k and m, and properties of the window $\phi(t)$. The conditions we impose to ensure stability are as follows:

Conditions 1. (Gaussian window conditions) When the window function is a Gaussian $\phi(t) = e^{-\frac{t^2}{\sigma^2}}$, we require its width σ, the source locations and sampling locations to satisfy the following conditions:

1. Boundary samples: $s_1 = 0$ and $s_m = 1$,
2. Samples near sources: for every $i \in [k]$, there exists a pair of samples $s, s' \subset S$ such that $|s - t_i| \leqslant \eta$ and $s' - s = \eta$, for η small enough; which is quantified in [24].
3. Sources away from the boundary: $\sigma\sqrt{\log(1/\eta^3)} \leqslant t_i, s_j \leqslant 1 - \sigma\sqrt{\log(1/\eta^3)}$ for every $i \in [k]$ and $j \in [2 : m-1]$,
4. Minimum separation of sources: $\sigma \leqslant \sqrt{2}$ and $\Delta(T) > \sigma\sqrt{\log\left(3 + \frac{4}{\sigma^2}\right)}$, where the minimum separation $\Delta(T)$ of the sources is defined in Definition 1.

Definition 1. (Minimum separation) For finite $T \subset I$, let $\Delta(T) > 0$ be the minimum separation between the points in T and the endpoints of I, namely $\Delta(T)$ is the largest number s such that

$$s \leqslant |t_i - t_{i'}|, \qquad i \neq i',\ i, i' \in [k],$$
$$s \leqslant |t_i - 0|, \qquad s \leqslant |t_i - 1|, \qquad i \in [k]. \tag{7}$$

When referring to the minimum separation between points in the support of the true signal, we will use $\Delta(T)$.

The four properties in Conditions 1 can be interpreted as follows: Property 1 imposes that the sources are within the interval defined by the minimum and maximum sample; Property 2 ensures that there are a pair of samples near each source which translates into a sampling density condition in relation to the minimum separation between sources and in particular requires the number of samples $m \geqslant 2k+2$; Property 3 is a technical condition to ensure sources are not overly near the sampling boundary; and Property 4 relates the minimum separation between the sources to the width, σ, of the Gaussian window.

To make Property 2 of Conditions 1 more transparent we introduce the distance $\lambda\Delta(T)$ between each source and its closest sample, namely for $\lambda \in (0, \frac{1}{2})$ and each source location t_i we have

$$|t_i - s_{l(i)}| \leqslant \lambda\Delta(T), \tag{8}$$

where $s_{l(i)}$ is the sample that is the closest to the source t_i. That is, λ, $\Delta(T)$ and η defined in Conditions 1 are related by $\lambda\Delta(T) \leqslant \eta/2$. We can now present stability bounds for Program (3) for the Gaussian window in terms of the error near sources, on $T_{i,\epsilon} = (t_i - \epsilon, t_i + \epsilon)$, and away from the sources, on $T_\epsilon^C = \left(\bigcup_{i=1}^k T_{i,\epsilon}\right)^C$.

Theorem 2. (Stability of Program (3) for Gaussian $\phi(t)$) *Let $I = [0,1]$ and consider a k-sparse non-negative measure x supported on T and sample locations S as given in* (4) *and for positive σ, let $\{\phi_j(t)\}_{j=1}^m$ as defined in* (5)*. If the Conditions 1 hold, then, in the presence of additive noise, Program* (3) *is stable and it holds that, for any solution $\hat{x}$ of Program* (3)*:*

$$\left|\int_{T_{i,\epsilon}} \hat{x}(\mathrm{d}t) - a_i\right| \leqslant (c_1 + F)\cdot\delta + c_2\frac{\|\hat{x}\|_{TV}}{\sigma^2}\cdot\epsilon, \tag{9}$$

$$\left|\int_{T_\epsilon^C} \hat{x}(\mathrm{d}t)\right| \leqslant F \cdot \delta, \tag{10}$$

where the exact expression of $F = F(k, \Delta(T), \frac{1}{\sigma}, \frac{1}{\epsilon})$ is given in [24], provided that λ, $\Delta = \Delta(T)$ and σ satisfy

$$\phi(\lambda\Delta) = \phi(\Delta - \lambda\Delta) + \phi(\Delta + \lambda\Delta) + \frac{1}{\Delta}\int_{\Delta-\lambda\Delta}^{1/2-\lambda\Delta} \phi(x)\,\mathrm{d}x + \frac{1}{\Delta}\int_{\Delta+\lambda\Delta}^{1/2+\lambda\Delta} \phi(x)\,\mathrm{d}x, \tag{11}$$

In particular, for $\sigma < \frac{1}{\sqrt{3}}$, $\Delta(T) > \sigma\sqrt{\log\frac{5}{\sigma^2}}$, and $\lambda < 0.4$, we have:

$$F(k, \Delta(T), \frac{1}{\sigma}, \frac{1}{\epsilon}) < c_3 \frac{k\bar{C}(\frac{1}{\epsilon})}{\sigma^2}\left[\frac{c_4}{\sigma^6(1-3\sigma^2)^2}\right]^k.$$

Above, c_1, c_2, c_3, c_4 are universal constants and $\bar{C}(\frac{1}{\epsilon})$ is given in [24].

Theorem 2 is a particular case of the more general theorem[2] which allows for other sampling windows and source sample configurations. In particular, the more general Theorem 5 is valid for any windows $\phi(t)$ which are Chebyshev systems as defined in Definition 3, of which the Gaussian window in (5) is an example.

Definition 3. (Chebyshev system [25]) *Real-valued and continuous functions $\{\phi_j\}_{j=1}^m$ form a Chebyshev system on the interval I if the $m \times m$ matrix $[\phi_j(\tau_l)]_{l,j=1}^m$ is nonsingular for any increasing sequence $\{\tau_l\}_{l=1}^m \subseteq I$.*

We pose the more general extension of Theorem 2 in terms of minimum separation in Definition 1, rather than the more restrictive Conditions 1.

While the limiting case of δ approaching zero in Theorem 2 is suggestive that the solution z of Program (3) approaches a discrete measure, the proof of this result follows more directly and is so stated separately in Proposition 4.

Proposition 4. (Uniqueness of exactly sampled sparse non-negative measures) *Let x be a non-negative k-sparse discrete measure supported on I, see (1). Let $\delta = 0$ and, with $m \geqslant 2k + 1$, assume that $\{\phi_j\}_{j=1}^m$ form a Chebyshev system on I. Then x is the unique solution of Program (3) with $\delta = 0$.*

Proposition 4 states that Program (3) successfully localizes the k impulses present in x given only $2k + 1$ measurements when $\{\phi_j\}_{j=1}^m$ form a T-system on I.

Lastly, the more general variant of Theorem 2 bounds the error using the norm of the vector of coefficients of the dual polynomial rather than explicitly in terms of the parameters of the problem.

Theorem 5. (Stability of Program (3) **for general** $\phi(t)$**)** *Let $\hat{x}$ be a solution of Program* (3) *and $\Delta(T)$ the minimum separation of T as defined in Definition 1. If for each source t_i there exists a closest sampling location s_i as defined in* (8) *for $\lambda = \lambda_0 \in [0, 1/2)$ which satisfies* (11) *then, for a given $\epsilon \in (0, \Delta(T)/2)$ and for all $i \in [k]$,*

$$\left|\int_{T_{i,\epsilon}} \hat{x}(\mathrm{d}t) - a_i\right| \leqslant [2(1 + \frac{\phi^\infty \|b\|_2}{\bar{f}})\delta + \epsilon L \|\hat{x}\|_{TV}] \sum_{j=1}^k (A^{-1})_{ij}, \tag{12}$$

$$\left|\int_{T_\epsilon^C} \hat{x}(\mathrm{d}t)\right| \leqslant \frac{2\|b\|_2}{\bar{f}}\delta, \tag{13}$$

where:

- $\phi^\infty = \max_{s,t\in I} |\phi(s-t)|$,
- *L is the Lipschitz constant of $\phi(t)$,*
- *$A \in \mathbb{R}^{k\times k}$ is the matrix*

$$A = \begin{bmatrix} |\phi_1(t_1)| & -|\phi_1(t_2)| & \dots & -|\phi_1(t_k)| \\ -|\phi_2(t_1)| & |\phi_2(t_2)| & \dots & -|\phi_2(t_k)| \\ \vdots & \vdots & \ddots & \vdots \\ -|\phi_k(t_1)| & -|\phi_k(t_2)| & \dots & |\phi_k(t_k)| \end{bmatrix}, \tag{14}$$

with $\phi_i(t_i) = \phi(t_i - s_{l(i)})$ is evaluated at $s_{l(i)}$ from (8).

- *and $b \in \mathbb{R}^m$ are the coefficients of a dual polynomial $q(t) = \sum_{i=1}^m b_i \phi_i(t)$ which satisfies*

$$q(t) \geqslant F(t) := \begin{cases} f(t - t_i), & \text{when } t \in [t_i - \epsilon, t_i + \epsilon] \\ \bar{f}, & \text{elsewhere on } I, \end{cases}$$

where the equality holds on T.

Note that $\bar{f}$ typically decreases with ϵ and the ratio $\frac{\|b\|_2}{\bar{f}}$ increases as ϵ approaches zero, see [24]. As such, we can set $\epsilon \sim \delta^{1/p}$ and then the overall bound given by the right hand side of (12) is $O(\delta^{1/p})$ for an appropriately chosen $p > 1$, so the discrepancy goes to zero as the samples become consistent with a sparse measure, e.g. as δ goes to zero. Moreover, while Theorem 5 explicitly states that the location of the closest samples to each source is less than $\lambda_0\Delta(T)$, this is achieved without knowing the locations of the sources by placing the samples uniformly at interval $2\lambda_0\Delta(T)$ which gives a sampling complexity of $m = (2\lambda_0\Delta(T))^{-1}$.

3. SOURCE LOCATIONS FOR GAUSSIAN AND POLYNOMIAL WINDOW FUNCTIONS

The source and sample conditions of Theorems 2 and 5 are determined by condition (11) used to ensure the matrix A in (14) is diagonally dominant. In this section we explore further when the sufficient bound (11) is satisfied for both the Gaussian window ϕ as defined in (5) and the function

$$\hat{\phi}(t) = (1 - |t|)^\gamma, \quad \gamma > 1. \tag{15}$$

[2] All proofs of results herein and numerous substantial extensions are available in [24].

as an example of a function that is non-differentiable, has more rapid initial decay and slower decay far from the origin, and moreover is not known to be a Chebyshev system.

Figure 1 shows the relationships between λ, $\Delta(T)$ and the window localization parameters γ and σ, in the left and right panels respectively, by solving equation (11) numerically; recall that (11) considers the worst sampling locations consistent with bound (8). The first row of Figure 1 (panels (a) and (b)) shows the degree to which samples are needed to become closer, that is λ to decrease, as the window function becomes wider (for small values of γ in (a) and large values of σ in (b)). This also depends on the minimum distance between sources $\Delta(T)$ with λ decaying more quickly for small $\Delta(T)$. The second row of Figure 1, in panels (c) and (d), shows the dependence between the width and $\Delta(T)$. When sources are closer to each other, the window function must be narrow for the same value of λ. In both plots we also show the case when $\lambda = 0$, namely when we have samples at the locations of the sources. Going beyond this curve (bottom left in (c) and top left in (d)) leads to not being able to reconstruct the signal. Approximation to these curves as λ approaches zero by taking leading Taylor series in (11) gives the following relationships between $\Delta(T)$ and the localization parameters of the windows: $\Delta(T) \approx \frac{2-2^{-\gamma}}{1+\gamma}$ for (15) and $\Delta(T) \approx \sqrt{\pi}\sigma \operatorname{erf}\left(\frac{1}{2\sigma}\right)$ for the Gaussian window (5).

Finally, in the bottom row of Figure 1, we fix the parameters γ and σ of the windows and show the dependence between λ and $\Delta(T)$. As expected, when the minimum distance between sources is greater, the distance between sources and samples can also be greater.

We show a few examples of parameters that satisfy (11) in Table 1 and signals with sources and sampling locations that have these parameters in Figure 2. Here we see $k = 5$ sources generated using the window function $\hat{\phi}$ in (a), (c), (e) and using the window function ϕ in (b), (d), (f). We start with the sources placed at $t_1 = 0.1, t_2 = 0.4, t_3 = 0.5, t_4 = 0.66, t_5 = 0.78$ in (a) and (b) so that we have the minimum distance between sources $\Delta(T) = 0.1$, then in (c) and (d) we keep the same source locations and we increase the width of the window functions, and in (e) and (f) we have the same width as in (a) and (b) but we move the sources to $t_1 = 0.1, t_2 = 0.25, t_3 = 0.4, t_4 = 0.63, t_5 = 0.78$ so that we increase the minimum distance to $\Delta(T) = 0.15$. For each of these configurations, we place the samples uniformly at intervals $2\lambda\Delta(T)$, so that the distance between each source and its closest sample is at most $\lambda\Delta(T)$.

γ	$\Delta(T)$	λ	σ	$\Delta(T)$	λ
25	0.1	0.4245	0.07	0.1	0.4292
15	0.1	0.2401	0.085	0.1	0.3386
25	0.15	0.4720	0.07	0.15	0.4833

Table 1. Examples of parameter values that satisfy (11) for $\hat{\phi}$ (left) and ϕ (right). Rows correspond to rows in Figure 2.

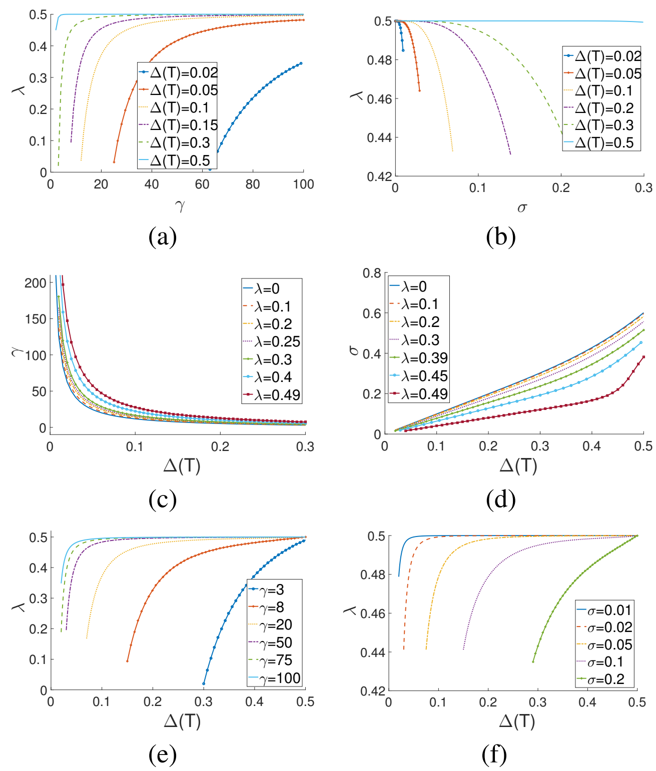

Fig. 1. Dependence of λ on $\Delta(T)$ and the width of the window function as given by (11) for (15) (left), and (5) (right).

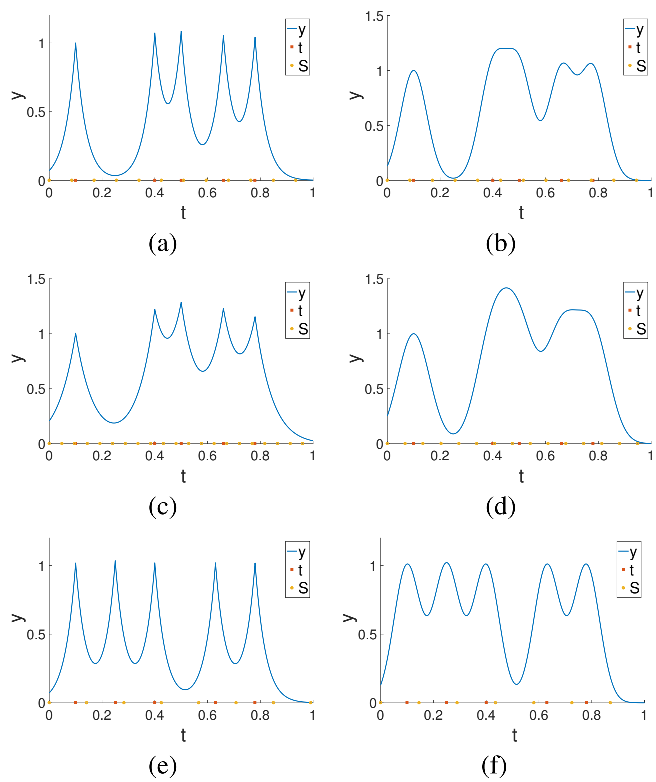

Fig. 2. Examples from Table 1 for (15) (left) and (5) (right), where sampling points are located uniformly at interval $2\lambda\Delta(T)$.

4. REFERENCES

[1] Y. De Castro and F. Gamboa, "Exact reconstruction using beurling minimal extrapolation," *Journal of Mathematical Analysis and applications*, vol. 395, no. 1, pp. 336–354, 2012.

[2] G. Schiebinger, E. Robeva, and B. Recht, "Superresolution without separation," *Information and Inference: a journal of the IMA*, vol. 7, no. 1, pp. 1–30, 2018.

[3] K.G. Puschmann and F. Kneer, "On super-resolution in astronomical imaging," *Astronomy & Astrophysics*, vol. 436, no. 1, pp. 373–378, 2005.

[4] E. Betzig, G.H. Patterson, R. Sougrat, O.W. Lindwasser, S. Olenych, J.S. Bonifacino, M.W. Davidson, J. Lippincott-Schwartz, and H.F. Hess, "Imaging intracellular fluorescent proteins at nanometer resolution," *Science*, vol. 313, no. 5793, pp. 1642–1645, 2006.

[5] S.T. Hess, T.P.K. Girirajan, and M.D. Mason, "Ultra-high resolution imaging by fluorescence photoactivation localization microscopy," *Biophysical journal*, vol. 91, no. 11, pp. 4258–4272, 2006.

[6] M.J. Rust, M. Bates, and X. Zhuang, "Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (storm)," *Nature methods*, vol. 3, no. 10, pp. 793–796, 2006.

[7] C.W. Mccutchen, "Superresolution in microscopy and the abbe resolution limit," *JOSA*, vol. 57, no. 10, pp. 1190–1192, 1967.

[8] H. Greenspan, "Super-resolution in medical imaging," *The Computer Journal*, vol. 52, no. 1, pp. 43–63, 2009.

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

[10] S. Hell, "Primer: fluorescence imaging under the diffraction limit," *Nature methods*, vol. 6, no. 1, pp. 19, 2009.

[11] R. Tur, Y.C. Eldar, and Z. Friedman, "Innovation rate sampling of pulse streams with application to ultrasound imaging," *IEEE Transactions on Signal Processing*, vol. 59, no. 4, pp. 1827–1842, 2011.

[12] D.J. Thomson, "Spectrum estimation and harmonic analysis," *Proceedings of the IEEE*, vol. 70, no. 9, pp. 1055–1096, 1982.

[13] G. Tang, B.N. Bhaskar, and B. Recht, "Near minimax line spectral estimation," *IEEE Transactions on Information Theory*, vol. 61, no. 1, pp. 499–512, 2015.

[14] V. Khaidukov, E. Landa, and T.J. Moser, "Diffraction imaging by focusing-defocusing: An outlook on seismic superresolution," *Geophysics*, vol. 69, no. 6, pp. 1478–1490, 2004.

[15] P. Shah, B.N. Bhaskar, G. Tang, and B. Recht, "Linear system identification via atomic norm regularization," in *Decision and Control (CDC), 2012 IEEE 51st Annual Conference on*. IEEE, 2012, pp. 6265–6270.

[16] E.J. Candès and C. Fernandez-Granda, "Towards a mathematical theory of super-resolution," *Communications on Pure and Applied Mathematics*, vol. 67, no. 6, pp. 906–956, 2014.

[17] G. Tang, B.N. Bhaskar, P. Shah, and B. Recht, "Compressed sensing off the grid," *IEEE Transactions on Information Theory*, vol. 59, no. 11, pp. 7465–7490, 2013.

[18] K. Fyhn, H. Dadkhahi, and M.F. Duarte, "Spectral compressive sensing with polar interpolation," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2013.

[19] L. Demanet, D. Needell, and N. Nguyen, "Super-resolution via superset selection and pruning," *Proc. SampTA Conference, Bremen*, 2013.

[20] V. Duval and G. Peyré, "Exact support recovery for sparse spikes deconvolution," *Foundations of Computational Mathematics*, pp. 1–41, 2015.

[21] Q. Denoyelle, V. Duval, and G. Peyré, "Support recovery for sparse super-resolution of positive measures," *Journal of Fourier Analysis and Applications*, vol. 23, no. 5, pp. 1153–1194, 2017.

[22] J.M. Azais, Y. De Castro, and F. Gamboa, "Spike detection from inaccurate samplings," *Applied and Computational Harmonic Analysis*, vol. 38, no. 2, pp. 177–195, 2015.

[23] T. Bendory, S. Dekel, and A. Feuer, "Robust recovery of stream of pulses using convex optimization," *Journal of Mathematical Analysis and Applications*, vol. 442, no. 2, pp. 511–536, 2016.

[24] A. Eftekhari, J. Tanner, A. Thompson, B. Toader, and H. Tyagi, "Sparse non-negative super-resolution: simplified and stabilized," 2018.

[25] S. Karlin and W.J. Studden, *Tchebycheff systems: with applications in analysis and statistics*, Pure and applied mathematics. Interscience Publishers, 1966.