

# Super-Resolution of Point Sources Down to the Rayleigh Limit from Multiple Observations

Dmitriy Kunisky, Efe Onaran

December 16, 2017

## 1 Introduction

We consider the problem of super-resolution, that of extracting high-frequency information from low-frequency observations, approached through convex optimization. Although its name comes from optics, where spatial resolution is naturally limited due to the relationship between light wavelength and the numerical aperture of the optical system [1], the super-resolution problem has come up in other settings as well, including image processing to produce high-resolution images from low-resolution observations [2], and the recovery of neural activity—which consists of characteristic high-frequency “spiky” behavior—from imperfect measurements [3]. If we interchange the time (or space) and frequency domains, we obtain the spectral super-resolution problem, which is relevant when time domain samples are not of infinite length for technical reasons, creating a low-pass effect in the frequency domain. This formulation of the problem finds applications in, for instance, identifying nuclear spectra [4].

In our setting, we assume the original data are a train of spikes, as shown in Figure 1. Note that recovering the original signal from its low-pass observation without additional assumptions is an ill-posed problem, as the high-frequency component of the signal could be arbitrary. The necessary assumption for recovery turns out to be the intuitive property of a *minimum separation* between spikes. Suppose the cut-off frequency of the observation filter is  $f_c$ . As illustrated in Figure 2 of [5], the existence of a purely high-frequency spike train with minimum separation  $\frac{1}{2}\lambda_c$  implies that the spectrum of the difference of two spike trains with separation  $\lambda_c$  can be made to lie entirely outside of the low-pass band with cutoff  $f_c$ , whereby it is in principle impossible to distinguish those two signals. This phenomenon is known as the Rayleigh diffraction limit [5].

An interesting aspect of super-resolution that has been studied recently is the possibility of solving the recovery problem by convex relaxation techniques. Several works have been devoted to characterizing the accuracy and robustness to noise of different convex relaxations of the super-resolution problem [6, 7, 8, 9, 10]. In particular, TV-norm relaxation (a generalization of the  $\ell^1$  norm to continuous domain) has been proven to successfully super-resolve spikes with minimum separation bounded by  $\approx 1.26\lambda_c$ , although numerical experiments suggest that it in fact achieves the separation  $\lambda_c$  described by the above heuristic reasoning.

In this work, we study convex relaxations of a generalization of the super-resolution problem to multiple observations where it is assumed that multiple independent spike trains are observed that all share the same support. This version of the problem was first described in [5]. We propose a dual certificate that certifies the tightness of convex gTV-norm relaxation for this problem,

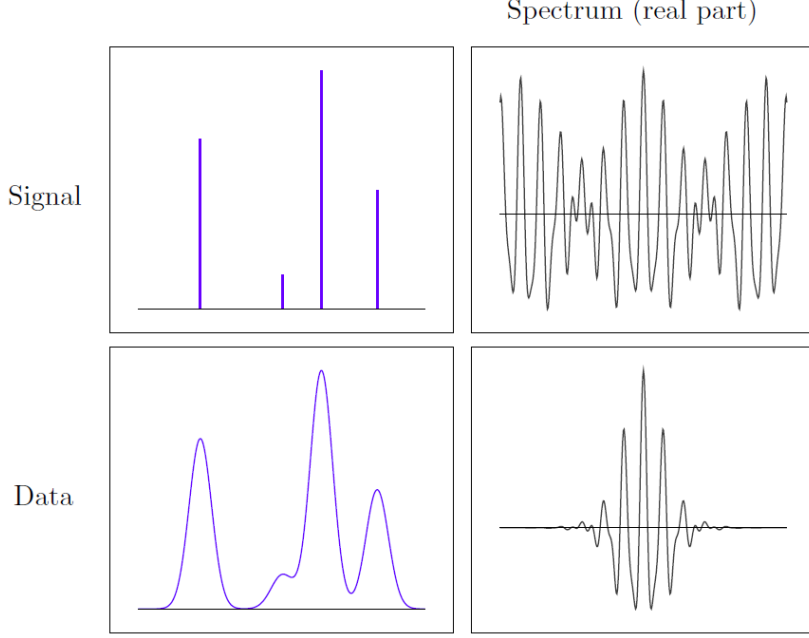


Figure 1: **Schematic of Super-Resolution.** The original signal and observed data for the super-resolution problem with their frequency spectra, as illustrated in [5].

as proposed in [5]. Numerical experiments suggest that our certificate is dual-feasible up to the Rayleigh limit  $\frac{1}{2}\lambda_c$ , provided that the number of observations is large enough. We also discuss a theoretical approach to prove the feasibility of this dual certificate under suitable conditions on the minimum separation and number of observations, although the proof is left for future work.

## 2 Formal Setup

### 2.1 Single Observation Problem Statement

Suppose we observe a true signal whose representation in the physical domain is

$$x^*(t) = \sum_{j=1}^s a_j \delta_{t_j}(t),$$

for  $t \in \mathbb{T}$  where  $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ , which we will usually think of as the unit interval  $[0, 1]$  with its endpoints identified, and  $T = \{t_j\}_{j=1}^s \subset \mathbb{T}$  a discrete support set. In the Fourier domain, this signal takes the form

$$\widehat{x^*}(k) = \sum_{j=1}^s a_j \exp(-2\pi i k t_j)$$

for  $k \in \mathbb{Z}$ .

Now, we are interested in recovering this true signal  $x$  from an observation that suffers from low resolution, which we represent as convolution of  $x(t)$  with a point-spread function (PSF) denoted

by  $\phi(t)$ . We then observe the signal

$$(\phi * x^*)(t) = \sum_{j=1}^s a_j \phi(t - t_j).$$

In the simplest case,  $\phi$  is a Dirichlet kernel with cutoff frequency  $f_c$ , in which case the Fourier transform of the above is simply the truncation of the Fourier transform of  $x^*$ . Denoting by  $y$  the signal we observe in the Fourier domain, we have

$$y(k) = \left( \sum_{j=1}^s a_j \exp(-2\pi i k t_j) \right) \mathbb{1}_{\{|k| \leq f_c\}}.$$

We may then think of the data we observe as simply the  $2f_c + 1$  values  $y(-f_c), \dots, y(f_c)$ . We write  $\mathcal{F}_n$  for the sensing operator mapping  $x^*$  to these  $n$  Fourier coefficients. A popular technique for solving this problem is *total variation minimization*, which attempts to recover  $x^*$  by solving the convex problem

$$\text{minimize } \|x\|_{\text{TV}} \quad \text{subject to } \mathcal{F}_n x = y.$$

## 2.2 Extension to Multiple Observations

We now consider a generalization of the problem presented in the previous section, where we make *several* observations signals sharing the support  $T = \{t_1, \dots, t_s\}$  of  $x^*$ , but having varying amplitudes  $a_j$ . To formalize this, our true signal is now

$$x_\ell^*(t) = \sum_{j=1}^s a_{\ell,j} \delta_{t_j} \text{ for } \ell \in \{1, \dots, m\},$$

and our observations are  $y_{\ell,k} = (\mathcal{F}_n x_\ell^*)_k$  for  $k \in \{-f_c, \dots, f_c\}$  and  $\ell \in \{1, \dots, m\}$ . We think of the  $a_{\ell,j}$  as organized into vectors  $\mathbf{a}_j \in \mathbb{R}^m$  for  $j \in \{1, \dots, s\}$ , and the  $y_{\ell,k}$  as organized into a matrix  $\mathbf{Y} \in \mathbb{C}^{m \times n}$ . The *group total variation minimization* is the natural extension of the previous convex problem to this setting, where we solve

$$\text{minimize } \|x\|_{\text{gTV}} \quad \text{subject to } \begin{bmatrix} \mathcal{F}_n x_1 & \mathcal{F}_n x_2 & \dots & \mathcal{F}_n x_m \end{bmatrix}^\top = \mathbf{Y}.$$

## 2.3 Dual Certificates

In this section, we briefly describe the results of applying Lagrangian duality to TV and gTV norm minimization, the main tool for theoretical analysis of the performance of these algorithms.

**Definition 1** Let  $\mu^0 \subset \mathbb{C}$  denote the complex unit circle. A sign pattern on a set is an assignment of points of  $\mu^0$  to each point of the set.

**Definition 2** For a sign pattern  $v \in (\mu^0)^T$ , a low-pass trigonometric polynomial  $q : \mathbb{T} \rightarrow \mathbb{C}$ ,

$$q(t) = \sum_{k=-f_c}^{f_c} c_k e^{2\pi i k t},$$

is a single-observation dual certificate for  $v$  if  $q(t_j) = v_j$  for  $t_j \in T$  and  $|q(t)| < 1$  for  $t \notin T$ .

**Proposition 1 (TV Norm Minimization Duality)** *If a dual certificate for the sign pattern  $v_j = a_j/|a_j|$  exists, then  $x^*$  is the unique solution of TV minimization for the super-resolution problem.*

Therefore, to prove the effectiveness of TV minimization it suffices to show that a dual certificate exists under some conditions on  $T$ . The main result of [5] establishes that this is true under a minimum separation condition.

**Definition 3** *The minimum separation of a set  $T \subset \mathbb{T}$  is*

$$\Delta(T) = \min_{\substack{s, t \in T \\ s \neq t}} |s - t|,$$

where we measure the wrap-around distance as on a circle.

**Theorem 1 (Proposition 2.3 of [5])** *If  $\Delta(T) \geq 1.26\lambda_c$  and  $f_c \geq 10^3$ , then a dual certificate exists for any sign pattern on  $T$ .*

As also observed in [5], the same ideas extend to the multiple observation case in a straightforward fashion.

**Definition 4** *Let  $\mu^{m-1} \subset \mathbb{C}^m$  denote the  $m$ -dimensional complex unit sphere. A  $m$ -dimensional sign pattern on a set is an assignment of points of  $\mu^{m-1}$  to each point of the set.*

**Definition 5** *For a sign pattern  $\mathbf{v} \in (\mu^{m-1})^T$ , a low-pass trigonometric polynomial  $q : \mathbb{T} \rightarrow \mathbb{C}^m$  given by  $q(t) = (q_1(t), \dots, q_m(t))$  and*

$$q_\ell(t) = \sum_{k=-f_c}^{f_c} c_k \exp(2\pi i k t) \text{ for } \ell \in \{1, \dots, m\}$$

is an  $m$ -observation dual certificate for  $v$  if  $q(t_j) = \mathbf{v}_j$  for  $t_j \in T$ , and  $\|q(t)\|_{\ell^2} < 1$  for  $t \notin T$ .

**Proposition 2 (gTV Norm Minimization Duality)** *If a dual certificate for the sign pattern  $\mathbf{v}_j = \mathbf{a}_j/\|\mathbf{a}_j\|_{\ell^2}$  exists, then  $x^*$  is the unique solution of gTV minimization for the super-resolution problem.*

However, while numerical experiments in [5] suggest that when the amplitude vectors  $\mathbf{a}_j$  are taken randomly then recovery is possible down to a lower critical minimum separation which accumulates at  $\frac{1}{2}\lambda_c$  as  $m \rightarrow \infty$ , the same uniform argument cannot apply—indeed, it is always possible that the amplitudes  $\mathbf{a}_j$  are all identical, in which case clearly it cannot be possible to recover  $x^*$  any more effectively than in the one observation case. Thus, to show that gTV norm minimization substantially improves on TV norm minimization, it is necessary to make some incoherence assumptions about the data generating process of the  $\mathbf{a}_j$ . One simple such assumption that captures the idea of these points being in “general position” is to assume that the signs  $\mathbf{a}_j/\|\mathbf{a}_j\|_{\ell^2}$  are isotropically random unit vectors.

### 3 Dual Certificate Construction

Recall that we are interested in building a low-pass trigonometric polynomial  $q : \mathbb{T} \rightarrow \mathbb{C}^m$  interpolating the points  $(t_j, v_j)$  while remaining strictly inside the unit sphere in  $\mathbb{C}^m$  elsewhere. The idea of the construction of [5] when  $m = 1$  is to build  $q$  via kernel interpolation, and adjust it so that  $q'(t_j) = 0$  for each  $j$ , which one would hope might encourage the polynomial to remain in the unit sphere (in particular, this solves a problem with the naive approach of building an interpolant directly, where the polynomial does not remain inside the sphere near interpolated points but rather very slightly escapes).

The naive extension of this to the  $m$ -dimensional case is to require  $q'(t_j) = 0$  for each  $t_j$ . However, numerical experiments show that the regime where this construction succeeds essentially does not depend on  $m$ , which is not surprising—after all, this is very close to just repeating the one-dimensional construction in each coordinate. In this section, we propose a different dual certificate, which does appear to succeed numerically in certifying reconstruction down to the Rayleigh limit as  $m \rightarrow \infty$ .

#### 3.1 Step 1: Relaxing the Derivative Condition

The first observation is that the condition  $q'(t_j) = 0$  is overkill for encouraging  $q$  to remain in the unit sphere. Intuitively, all that is really necessary is that  $q'(t_j)$  not have a component pointing out of the sphere, i.e. in the direction of  $v_j$ . Thus, a reasonable condition is to require  $q'(t_j)$  and  $v_j$ , viewed as vectors, to be orthogonal. However, it is not possible to express this constraint in complex arithmetic, so we must rewrite the dual certificate in terms of its real and complex parts (each of which is still low-pass).

Thus, we write  $q_\ell(t) = q_\ell^R(t) + iq_\ell^I(t)$  for real-valued low-pass trigonometric polynomials  $q_\ell^R$  and  $q_\ell^I$ . We combine these into a single real-valued  $2m$ -dimensional trigonometric polynomial,

$$\tilde{q}(t) = (q_1^R(t), q_1^I(t), \dots, q_m^R(t), q_m^I(t)) \in \mathbb{R}^{2m}.$$

Likewise writing  $v_{j,\ell} = v_{j,\ell}^R + iv_{j,\ell}^I$  and  $\tilde{v}_j = (v_{j,1}^R, v_{j,1}^I, \dots, v_{j,m}^R, v_{j,m}^I)$ , we can express the interpolation conditions that the dual certificate  $q$  must satisfy in terms of these new variables in essentially the same way:  $\tilde{q}(t_j) = \tilde{v}_j$  for  $j \in \{1, \dots, s\}$ , and  $\|\tilde{q}(t)\|_{\ell^2} < 1$  for all  $t \notin T$ . We can also now formulate the conditions we will require of our interpolant.

$$\text{INT1. } \tilde{q}(t_j) = \tilde{v}_j \text{ for } j \in \{1, \dots, s\}.$$

$$\text{INT2. } \langle \tilde{q}'(t_j), \tilde{v}_j \rangle = 0 \text{ for } j \in \{1, \dots, s\}.$$

#### 3.2 Step 2: Kernel Interpolation

We now proceed just as in [5], viewing the problem constructed in the previous section as a purely real-valued one: fix a real-valued low-pass kernel  $K$ , and define

$$\tilde{q}_\ell(t) = \sum_{j=1}^s \alpha_{j,\ell} K(t - t_j) + \sum_{j=1}^s \beta_{j,\ell} K'(t - t_j) \quad (1)$$

for  $\ell \in [2m]$  and coefficients  $\alpha_{j,\ell}, \beta_{j,\ell} \in \mathbb{R}$  to be determined. We think of these as organized into vectors  $\alpha, \beta \in \mathbb{R}^{2sm}$  where the entries are written iterating first over  $j$ , then over  $\ell$  (in “row major”

order). To concisely write conditions INT1 and INT2 from the previous part, we first make a small definition.

**Definition 6** *For a function  $K : \mathbb{T} \rightarrow \mathbb{R}$  and an ordered point set  $T = \{t_1, \dots, t_s\} \subset \mathbb{T}$ , define  $\mathbf{D}(K, T) \in \mathbb{R}^{s \times s}$  to have entries*

$$(\mathbf{D}(K, T))_{j,k} = K(t_j - t_k).$$

We also define the vector  $\tilde{\mathbf{v}}$  to be the  $\tilde{v}_{j,\ell}$  flattened, also in row-major order. Conditions INT1 and INT2 may then be written as the matrix equation

$$\begin{bmatrix} \mathbf{I}_{2m} \otimes \mathbf{D}(K, T) & \mathbf{I}_{2m} \otimes \mathbf{D}(K', T) \\ (\mathbf{1}_s \tilde{\mathbf{v}}^\top) \odot (\mathbf{1}_{2m}^\top \otimes \mathbf{D}(K', T)) & (\mathbf{1}_s \tilde{\mathbf{v}}^\top) \odot (\mathbf{1}_{2m}^\top \otimes \mathbf{D}(K'', T)) \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{bmatrix} = \begin{bmatrix} \tilde{\mathbf{v}} \\ \mathbf{0} \end{bmatrix}.$$

We set some notation for the matrix on the left-hand side and the blocks corresponding to the set of equality constraints and derivative constraints:

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{\text{eq}} \\ \mathbf{A}_{\text{der}} \end{bmatrix} = \begin{bmatrix} \mathbf{I}_{2m} \otimes \mathbf{D}(K, T) & \mathbf{I}_{2m} \otimes \mathbf{D}(K', T) \\ (\mathbf{1}_s \tilde{\mathbf{v}}^\top) \odot (\mathbf{1}_{2m}^\top \otimes \mathbf{D}(K', T)) & (\mathbf{1}_s \tilde{\mathbf{v}}^\top) \odot (\mathbf{1}_{2m}^\top \otimes \mathbf{D}(K'', T)) \end{bmatrix}.$$

### 3.3 Step 3: Choosing Solution of Underconstrained System

Note that the system described in the previous section has  $4sm$  variables but only  $2sm + s = (2m + 1)s$  constraints. When the sign pattern is purely real, this reduces to  $2s$  variables and  $2s$  constraints, so there is (generically) a unique solution, as in the construction treated in [5]. In our more general setting, however, uniqueness typically fails, and so we must build a criterion for choosing a solution.

This extra freedom lets us further encourage  $q(t)$  to remain inside the unit sphere. In particular, a natural quantity to minimize is the  $L^2$  norm of  $q(t)$ :

$$\|q\|_{L^2(\mathbb{T})}^2 = \int_0^1 \|q(t)\|_{\ell^2}^2 dt.$$

The added benefit of this choice is that it is a quadratic form in terms of the coefficients  $\boldsymbol{\alpha}, \boldsymbol{\beta}$ , so its minimization subject to linear constraints has a convenient closed form. We now develop some notation to write down this quadratic form.

**Definition 7** *For real-valued functions  $K_1, K_2 \in L^2(\mathbb{T})$  and an ordered point set  $T = \{t_1, \dots, t_s\} \subset \mathbb{T}$ , define  $\mathbf{G}(K_1, K_2, T) \in \mathbb{R}^{s \times s}$  to have entries*

$$(\mathbf{G}(K_1, K_2, T))_{j,k} = \int_0^1 K_1(t - t_j) K_2(t - t_k) dt.$$

Now, the  $L^2$  norm of the polynomial  $q(t)$  as defined in (1) is given by

$$\begin{aligned} \|q\|_{L^2(\mathbb{T})}^2 &= \sum_{\ell=1}^{2m} \sum_{j=1}^s \sum_{k=1}^s \left( \alpha_{\ell,j} \alpha_{\ell,k} \langle K(\bullet - t_j), K(\bullet - t_k) \rangle + \beta_{\ell,j} \beta_{\ell,k} \langle K'(\bullet - t_j), K'(\bullet - t_k) \rangle \right. \\ &\quad \left. + \alpha_{\ell,j} \beta_{\ell,k} \langle K(\bullet - t_j), K'(\bullet - t_k) \rangle \right) \\ &= \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{bmatrix}^\top \begin{bmatrix} \mathbf{I}_{2m} \otimes \mathbf{G}(K, K, T) & \mathbf{I}_{2m} \otimes \mathbf{G}(K, K', T) \\ \mathbf{I}_{2m} \otimes \mathbf{G}(K', K, T) & \mathbf{I}_{2m} \otimes \mathbf{G}(K', K', T) \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{bmatrix}. \end{aligned}$$

We set a notation for the inner matrix:

$$\begin{aligned} \mathbf{G}_0 &= \begin{bmatrix} \mathbf{G}(K, K, T) & \mathbf{G}(K, K', T) \\ \mathbf{G}(K', K, T) & \mathbf{G}(K', K', T) \end{bmatrix} \succeq 0, \\ \mathbf{G} &= \begin{bmatrix} \mathbf{I}_{2m} \otimes \mathbf{G}(K, K, T) & \mathbf{I}_{2m} \otimes \mathbf{G}(K, K', T) \\ \mathbf{I}_{2m} \otimes \mathbf{G}(K', K, T) & \mathbf{I}_{2m} \otimes \mathbf{G}(K', K', T) \end{bmatrix}. \end{aligned}$$

We may then concisely write down the full quadratic problem by which we will determine the coefficients  $\boldsymbol{\alpha}, \boldsymbol{\beta}$  as follows:

$$\text{minimize} \quad \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{bmatrix}^\top \mathbf{G} \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{bmatrix} \quad \text{subject to} \quad \mathbf{A} \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{bmatrix} = \begin{bmatrix} \tilde{\mathbf{v}} \\ \mathbf{0} \end{bmatrix}. \quad (2)$$

### 3.4 Explicit Solution

As mentioned previously, (2) has a closed-form solution, which may be derived with a straightforward Lagrange multiplier calculation, similar to the usual derivation of linearly constrained least-squares (corresponding to the case  $\mathbf{G} = \mathbf{I}$ ). In our setting, the coefficients  $\boldsymbol{\alpha}, \boldsymbol{\beta}$  are given by

$$\begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \\ \boldsymbol{\lambda} \end{bmatrix} = \begin{bmatrix} \mathbf{G} & \mathbf{A}^\top \\ \mathbf{A} & \mathbf{0} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{0} \\ \tilde{\mathbf{v}} \\ \mathbf{0} \end{bmatrix}$$

where  $\boldsymbol{\lambda} \in \mathbb{R}^{(2m+1)s}$  are the Lagrange multipliers introduced in the calculation. (The vectors on either side belong to  $\mathbb{R}^{(6m+1)s}$ .) Moreover, we may explicitly invert this block matrix, and dispose of the Lagrange multipliers to obtain the simpler formula

$$\begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{bmatrix} = \mathbf{G}^{-1} \mathbf{A}^\top (\mathbf{A} \mathbf{G}^{-1} \mathbf{A}^\top)^{-1} \begin{bmatrix} \tilde{\mathbf{v}} \\ \mathbf{0} \end{bmatrix}. \quad (3)$$

Formula (3) will be the beginning of our subsequent analysis; the remaining task is to obtain a detailed understanding of the structure of the matrix on the right-hand side.

## 4 Numerical Evaluation

In this section, we present some numerical results illustrating the performance of this dual certificate construction as the parameter  $m$  increases. These should be compared to Figures 3 and 9 of [5] to

observe that the dual certificate is feasible (and therefore certifies recovery) in basically all cases where the  $\mathbf{gTV}$  convex program itself successfully achieves recovery.

First, in Figure 2, we compare the case  $m = 1$  with the case  $m = 2$ , which is the difference between real and complex sign patterns, and see that our dual certificate indeed is feasible more often in the latter case. Second, in Figure 3, we consider higher values of  $m$ , and note that the improvement continues, and the minimum separation at which the dual certificate stops being feasible for large support sets appears to approach the Rayleigh limit of  $\frac{1}{2}\lambda_c$  as  $m \rightarrow \infty$ .

## 5 Proof Outline for Dual Feasibility

The following is a brief sketch of the ideas in our proof-to-be of the feasibility of this dual certificate construction.

1. First, we argue that the  $\alpha$  are close to  $\tilde{\mathbf{v}}$  and the  $\beta$  are small. Based on (3), this follows from arguing that

$$\mathbf{G}^{-1}\mathbf{A}^\top(\mathbf{A}\mathbf{G}^{-1}\mathbf{A}^\top)^{-1} \approx \begin{bmatrix} \mathbf{I}_{2m} & \bullet \\ \mathbf{0} & \bullet \end{bmatrix},$$

where the bullets indicate  $s$  columns whose values do not affect the values of  $\alpha$  and  $\beta$ . This should follow just by properties of the kernel  $K$ : since the kernel is localized near zero,  $\mathbf{A}$  and  $\mathbf{G}$  should both be roughly diagonal, or roughly Toeplitz with very quickly decaying values away from the main diagonal. A numerical example of this structure is presented in Figure 4. (The block  $\mathbf{A}_{\text{der}}$  may cause problems here, especially since it depends upon the random  $\tilde{\mathbf{v}}$ , but it appears to only influence the right-most columns of this matrix, which do not affect the values of  $\alpha$  and  $\beta$ ). Moreover, since the  $L^2$  norm of  $K'$  is much higher than that of  $K$ ,  $\mathbf{G}^{-1}$  should be close to zero on the lower half of its diagonal, which ensures that the lower block of this matrix is close to zero as needed. Many of the relevant bounds for this calculation are already collected in [5].

2. Next, we argue that in a small neighborhood of the support points  $t_j$ ,  $q(t)$  satisfies the necessary bound  $\|q(t)\|_{\ell^2} < 1$ . As explained nicely in [11] for the one-dimensional case, this can be done just by using the derivative condition and showing that the Hessian of  $\|q(t)\|_{\ell^2}^2$  (or something analogous) is negative-definite, so that there is a local maximum of the magnitude and Taylor's theorem gives the result in a sufficiently small neighborhood.
3. Finally, we must bound  $\|q(t)\|_{\ell^2}$  away from all of the support points. Here it is important to use our probabilistic assumption on the  $\mathbf{v}_j$ : since they are distributed uniformly on the unit sphere, they are roughly orthogonal, and so the tails of kernels centered at each  $\mathbf{v}_j$  should not be able to align sufficiently well to create a large value of  $q(t)$  away from the support points. The orthogonality intuition can be formalized in a few ways, depending on what sorts of bounds we end up needing, but the most useful would probably be that the Gram matrix of the  $\mathbf{v}_j$  is distributed like a Wishart matrix normalized by the norms of Gaussian vectors (which concentrate tightly), so associated random matrix results can help bound its off-diagonal entries, which are distributed like the inner products of the  $\mathbf{v}_j$ .



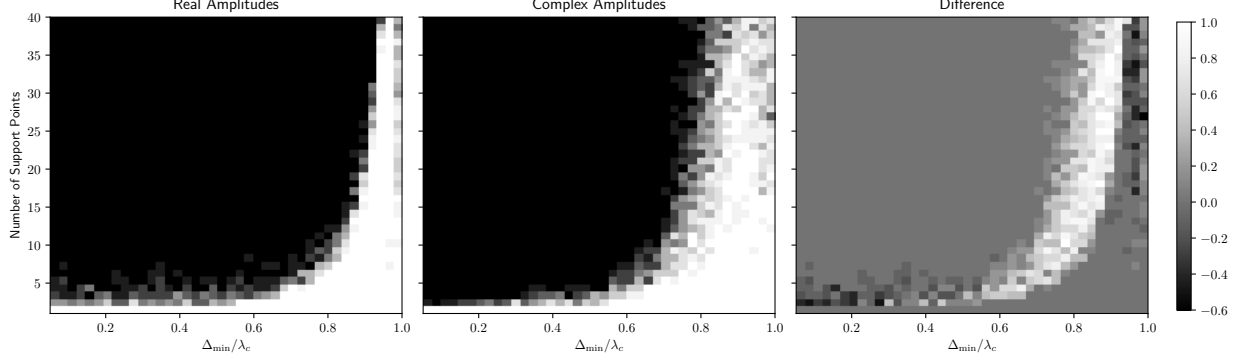


Figure 2: **Dual Certificate Feasibility: Real vs. Complex Amplitudes.** We show that our dual certificate construction performs better on random complex sign patterns than on random real sign patterns, showing the fraction of trials (out of 10) that the dual certificate is feasible in each case for fixed number of signals and minimum separation, and the difference between the fractions in the last panel.

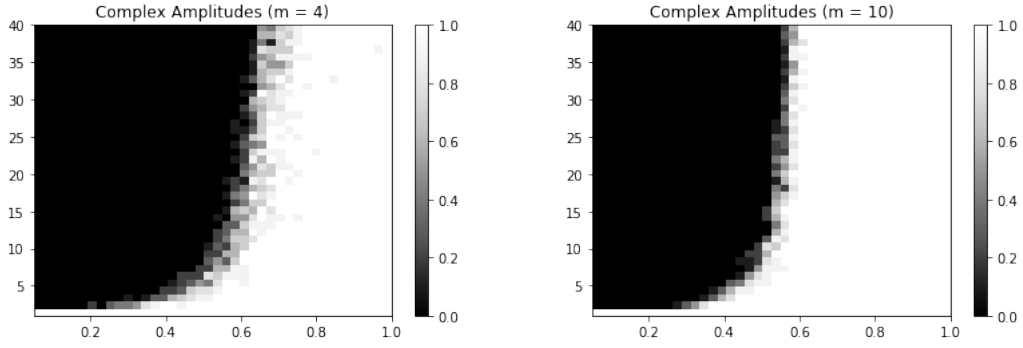


Figure 3: **Dual Certificate Feasibility: Increasing Dimension.** We show that the trend of improving feasibility performance as  $m \rightarrow \infty$  continues for higher values of  $m$ , with the feasibility phase transition appearing to saturate around a minimum separation of  $\frac{1}{2}\lambda_c$ .

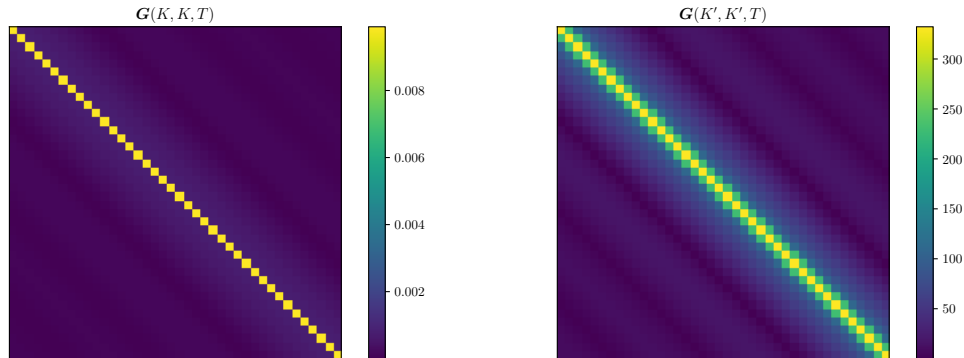


Figure 4: **Shifted Kernel Gram Matrices.** We show the structure of Gram matrices of almost-regular shifts of a Dirichlet kernel, which are approximately Toeplitz with dominant diagonal values. The same for the kernel's derivative exhibits poorer decay and much higher values.

## References

- [1] Jari Lindberg. Mathematical concepts of optical superresolution. *Journal of Optics*, 14(8):083001, 2012.
- [2] Sina Farsiu, Dirk Robinson, Michael Elad, and Peyman Milanfar. Advances and challenges in super-resolution. *International Journal of Imaging Systems and Technology*, 14(2):47–57, 2004.
- [3] Fred Rieke. *Spikes: exploring the neural code*. MIT press, 1999.
- [4] Lindsey F Smith and Lawrence H Aller. On the classification of emission-line spectra of planetary nuclei. *The Astrophysical Journal*, 157:1245, 1969.
- [5] Carlos Fernandez-Granda. Super-resolution of point sources via convex programming. *Information and Inference: A Journal of the IMA*, 5(3):251–303, 2016.
- [6] Emmanuel J Candès and Carlos Fernandez-Granda. Towards a mathematical theory of super-resolution. *Communications on Pure and Applied Mathematics*, 67(6):906–956, 2014.
- [7] Jean-Marc Azais, Yohann De Castro, and Fabrice Gamboa. Spike detection from inaccurate samplings. *Applied and Computational Harmonic Analysis*, 38(2):177–195, 2015.
- [8] Vincent Duval and Gabriel Peyré. Exact support recovery for sparse spikes deconvolution. *Foundations of Computational Mathematics*, 15(5):1315–1355, 2015.
- [9] Carlos Fernandez-Granda. Support detection in super-resolution. *arXiv preprint arXiv:1302.3921*, 2013.
- [10] Gongguo Tang, Badri Narayan Bhaskar, and Benjamin Recht. Near minimax line spectral estimation. *IEEE Transactions on Information Theory*, 61(1):499–512, 2015.
- [11] Gongguo Tang, Badri Narayan Bhaskar, Parikshit Shah, and Benjamin Recht. Compressed sensing off the grid. *IEEE transactions on information theory*, 59(11):7465–7490, 2013.
- [12] Veniamin I Morgenshtern and Emmanuel J Candes. Super-resolution of positive sources: The discrete setup. *SIAM Journal on Imaging Sciences*, 9(1):412–444, 2016.