Investigating Kernel Performance for Super-Resolution of Point Sources from Multiple Observations

Camille Taltas and Dan Turkel, with Efe Onaran and Tim Kunisky

May 15, 2020

1 Introduction

1.1 Background

Scientists measuring signals are often faced with obstacles arising from the limited resolution of their observations. Diffraction limits imposed by optical systems or bandlimiting created by finite sampling rates result in observations with a lower level of detail than the "true" signal.

This imperfect measurement process can be modeled as the convolution of the signal with a low-pass kernel, resulting in the attenuation or removal of high-frequency content.[1] Super-resolution is the task of recovering that high-frequency (or high-resolution) content [2]. One of the tools in signal estimation is to leverage assumptions about the original signal. In this paper, we are interested in particular in the super-resolution of point sources: signals (or spectra) with sparse, discrete support (though the support points may lie anywhere in a continuous range). These types of signals arise in astronomy, where celestial objects are so distant that light or radio emissions from them are observed as point sources; nuclear stectroscopy, where substances may be identified by the line spectra of their emissions; and fluorescence microscopy, where individual fluorescent probes are located across a series of observations to create a high-resolution image. We will see that the multiple observations inherent to the fluorescent microscopy setting are particularly useful in our problem.

1.2 Prior Work

The theory and foundational results regarding guarantees for exact recovery of point source signals from a low-pass observation were developed in [4] and [2], where a key insight was that recovery was reliant on the *minimum separation* of the supports of the original signal. In particular, exact recovery was shown to be possible via convex optimization when minimum separation was at least $2/f_c$, where f_c is the cutoff frequency of the measurement process.

The result was expanded in [3], reducing the required minimum separation to $1.26/f_c$ and adapting the convex optimization approach to related problems.[2] Both [3] and [2] formulate the problem as a convex minimization of the total variation norm, which can be seen as an extension of the ℓ_1 norm (commonly used to induce sparse coefficients in optimization problems) from a discrete set of coefficients to a continuous space. Conveniently, for a signal consisting of discrete spikes, the TV norm is the sum of the absolute values of the spike amplitudes.

Thus we have the problem formulation

$$\min_{\hat{x}} \|\hat{x}\|_{\mathrm{TV}} \quad \text{such that} \quad \mathcal{F}_n \hat{x} = y$$

where $\mathcal{F}_n$ is the lowest $2n+1$ Fourier coefficients and $y = \mathcal{F}_n x$, the low-pass coefficients of the true signal.

The multiple observation setting is introduced in [3] and elaborated upon in [8]. In this setting, we have *multiple* independent signal observations where each signal consists of superimposed point sources and all signals have the same support points. During optimization, we now seek to impose *group sparsity*, which can be done with what [3] calls the "group total variation" norm and which has been used in prior simultaneous sparse recovery problems [11]. This norm makes it "cheaper" to put a support point for one signal in a time position where other signals already have support. The problem now takes the

[1] In the context of *spectral* super-resolution, the signal in question is itself the spectrum of another signal, but the same principle applies.

[2] While the proof in [3] accounts for separation greater above $1.26/f_c$, experiments in the same paper suggest the actual limit may be $1/f_c$, which corresponds to the Rayleigh limit, the smallest possible separation where recovery could still be guaranteed.

form

$$\min_{\hat{x}} \|\hat{x}\|_{\text{gTV}} \quad \text{such that} \quad \mathcal{F}_i \hat{x}_i = y_i, \text{ for all } i \in [1, m]$$

where m is the number of observations, and y_i is the low-pass spectrum of signal i.

Numerical experiments in [8] suggest that exact recovery from multiple observations is possible up to minimum separation $0.5/f_c$, under some assumptions on the observations: namely, the multiple observations must not be identical. A simplification is to consider m observations such that the signs of each vector form a standard basis of $\mathbb{R}^m$. It remains to be proven theoretically that recovery is guaranteed under these assumptions for minimum separation of $0.5/f_c$. Our work seeks to further the outlined proof by studying properties and behavior of interpolation kernels used to satisfy the dual problem and guarantee recovery.

2 State of the Art

While some of the first papers to approach recovery of point sources from low-pass observations from a theoretical angle—[4], [2], [3], [8]—do so via convex optimization methods, others have since brought different techniques to the problem.

In [9], the TV norm minimization in the single-observation setting is replaced with two non-convex norms: the ℓ_{1-2} norm $\|x\|_1 - \|x\|_2$ and the capped ℓ_1 norm $\sum_j \min(|x_j|, \alpha)$. The authors found that in cases where minimum separation was below the Rayleigh limit, the corresponding minimizers outperformed the TV minimization in [3], and noted in particular the tendency for the recovered solutions to be at least as sparse as the ground truth.

In [6], the authors train a deep neural network on simulated data to solve the line spectra recovery problem. The network forward pass creates an approximate "pseudo-spectrum," from which the point sources are estimated via peak-picking. (An alternative network was also trained to estimate the point sources directly). The authors note that prediction is much less computationally expensive using a neural network than convex optimization, though training the network is still expensive. The result does not come with any theoretical *guarantees* on when exact recovery is possible and at what threshold of minimum separation.

We note also two papers which continue the convex optimization approach to the problem. In [10], atomic norm minimization is used to guarantee MSE bounds for the recovered spectral supports. In [5], atomic norm minimization is also used, including in the multiple observations case. The authors claim to be able to recover the supports and amplitudes of sufficiently well-separated line spectra when the measurements are isotropic and incoherent (the same assumptions in [8]), where more measurements are required if the measurements themselves are more coherent. The catch, however, is that the calculation of the atomic norm (and thus the proposed algorithm) becomes prohibitively expensive as the dimension increases. Consequently, the authors instead restrict their discussion to recovery within a grid imposed on the frequency spectrum, rather than the true continuous axis.

In [1], a slight variant of the problem is explored where the original signal has been convolved with a *known* kernel. Line spectra are then recovered via convex optimization given a similar condition on minimum separation and a new condition wherein the sample locations must be located close to the point sources.

3 Problem Setup

In our setting, we are interested in lowering the minimum separation towards the Rayleigh limit of $0.5/f_c$ given that we have multiple low-resolution observations of our signal. The approach chosen follows the convex optimization approach introduced in [3] and adapted to multiple observations in [8].

Here, we attempt to recover the original signal by applying Lagrangian duality to the gTV minimization problem from Section 1.2. More specifically, we are interested in establishing the existence of a certain subgradient of the gTV norm, also known as the dual certificate, that is orthogonal to the null space of the measurement operator. In order to further understand how this construction works we first introduce the concept of sign patterns.

Definition 3.1 (Sign Pattern) *Let $\mu^0 \subset \mathbb{C}$ denote the complex unit circle. A sign pattern on a set is an assignment of points of μ^0 to each point of the set.*

From here, we are able to introduce the theorem (Lemma 3.6 in [3]) which certifies the exact recov-

ery of point sources under the existence of the dual certificate.

Theorem 3.2 (gTV Norm Minimization Duality) *For a true signal $x^*(t) = \sum_{t_j \in T} a_j \delta_{t_j}(t)$ where δ_{t_j} is a Dirac measure supported on $t_j \in T \subset [0,1]$ and $a_j \in \mathbb{C}$ is the corresponding amplitude of the point source, if a dual certificate for the sign pattern $v_j = a_j / \|a_j\|_{\ell^2}$ exists, then x^* is the unique solution to the gTV minimization problem.*

The construction of this dual certificate relies on the building of a low-pass trigonometric polynomial $q : [0,1] \to \mathbb{C}^m$ interpolating the points (t_j, v_j) while remaining inside the unit sphere in $\mathbb{C}^m$ elsewhere.

Definition 3.3 (Low-Pass Trigonometric Polynomial) *For a sign pattern $v \in (\mu^{m-1})^T$, a low-pass trigonometric polynomial $q : [0,1] \to \mathbb{C}^m$ given by $q(t) = (q_1(t), \ldots, q_m(t))$ and $q_\ell(t) = \sum_{k=-f_c}^{f_c} c_k e^{2\pi i k t}$ for $\ell \in \{1, \ldots, m\}$, is an m-observation dual certificate for v if $q(t_j) = v_j$ for $t_j \in T$ and $\|q_\ell\|_{\ell^2} < 1$ for $t \notin T$.*

In [7], a form for this interpolating polynomial is constructed by introducing a derivative condition which encourages the polynomial to remain within the unit sphere. More specifically we require every $q'(t_j)$ to be orthogonal to its respective v_j. In order to do so, we rewrite $q_\ell = q_\ell^R(t) + i q_\ell^I(t)$ as a combination of its real and complex parts giving us the $2m$-dimensional trigonometric polynomial

$$\tilde{q}(t) = (q_1^R(t), q_1^I(t), \ldots, q_m^R(t), q_1^I(t)).$$

Similarly, we rewrite

$$\tilde{v_j} = (v_{j,1}^R(t), v_{j,2}^I(t), \ldots, v_{j,m}^R(t), v_{j,m}^I(t)).$$

From here, we can formalize the conditions for our interpolant. Notably,

$$\tilde{q}(t_j) = \tilde{v_j}$$

and

$$\langle \tilde{q}'(t_j), \tilde{v_j} \rangle = 0.$$

Now that the above conditions and transformations are set, we can define our low-pass trigonometric polynomial as

$$\tilde{q}_\ell(t) = \sum_{t_j \in T} \alpha_{j,\ell} K(t - t_j) + \sum_{t_j \in T} \beta_{j,\ell} K'(t - t_j)$$

where K is a real-valued low-pass kernel.

Recall that in our case, in order to recover x^* more effectively than in the one observation case, we are assuming that each a_j is being drawn randomly. More specifically, we are presuming that the signs v_j are isotropically random unit vectors. In [7] and our work, we chose to simplify this assumption by approximating these random unit vectors as orthogonal to one another and, more particularly, as the standard basis in $\mathbb{R}^m$.

This assumption along with letting $|T| = m$ allows us to recover the following version of interpolating polynomial defined above

$$\tilde{q}(t) = K_0^{-1} k_0(t) + D_\beta [k_1(t) - K_1 K_0^{-1} k_0(t)]$$

where

$$\beta = \frac{\operatorname{diag}(K_0^{-1} K_1)}{\operatorname{diag}(K_2) - \operatorname{diag}(K_1 K_0^{-1} K_1)},$$

$$(K_r)_{i,j} = K^{(r)}(t_i - t_j)$$

and

$$(k_r)_{i,j} = K^{(r)}(t - t_j)$$

for K a real-valued low-pass kernel and D_β denoting the diagonal matrix with β as its diagonal.

4 Methodology

Our goal is to investigate kernel choices and characterize their behavior as interpolating polynomials. For which choices of kernel does the interpolating polynomial stay within the unit (hyper)sphere between support points, and for which choices of kernel does the derivative term interpolate the kernel derivative well?

Following [3], we chose to study "multi-Dirichlet kernels" of the form

$$K_\gamma(t) = \prod K(\gamma_i f_c, t), \quad \sum \gamma_i = 1$$

$$K(\gamma_i f_c, t) = \frac{1}{2f+1} \sum_{k=-f_c}^{f_c} e^{-2\pi k t}$$

We experimented[3] with γs of length 2 and 3. As an illustrative example, in Figure 1, we show two kernels, one for $\gamma = [0.1, 0.9]$ and one for $\gamma = [0.5, 0.5]$.

[3]Our experiments were all supported by the many relevant functions provided in the codebase of Tim Kunisky and Efe Onaran, found at https://github.com/eonaran/sr_multisignals.

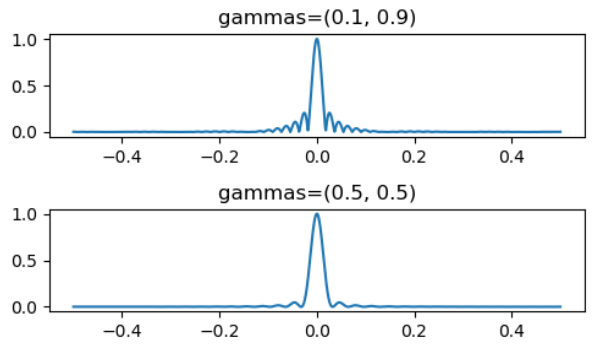

Figure 1: Multi-Dirichlet kernels for two γ choices.

For $\gamma = [0.1, 0.9]$, we have the desirable property that the central spike of the kernel is very sharp, meaning our interpolating polynomials will be "well-behaved" by decaying quickly in the immediate neighborhood of the support points. But the same kernel has the undesirable property of having nontrivial sidelobes which can create interference (especially if minimum separation is low and the kernels are thus close together) that can push our polynomial outside the unit sphere.

On the other hand, for $\gamma = [0.5, 0.5]$, the central spike is not quite as sharp, which means the immediate decay around the supports is not as fast, but the side lobes are more negligible.

These trade-offs become directly apparent when we plot the resultant interpolating polynomials. We fixed $f_c = 25$ and plotted polynomials for minimum separation of $0.9/f_c$ (relatively "wide") and $0.65/f_c$ (relatively narrow, close to the Rayleigh limit). Figure 2 shows the ℓ_2 norm of this polynomial for $\gamma = [0.1, 0.9]$ and $\gamma = [0.5, 0.5]$, and is discussed in the next session.

Lastly, we plotted the two derivative terms of our polynomial formulation, $D_\beta k_1(t)$ and $D_\beta K_1 K_0^{-1} k_0(t)$, to see how well the latter interpolates the former. We expect the two plots to intersect at 0 at each support point, by design, but we want to know how well bounded the difference between them remains between the supports so that this term does not threaten to "break" our certificate by raising the ℓ_2 norm too high. Figure 3 shows the plot for a single coordinate of this polynomial (the 6th coordinate, corresponding to the 6th support point) for $\gamma = [0.1, 0.9]$ and $\gamma = [0.5, 0.5]$, which is also discussed in the next section.

5 Results

Recall that in order to be a sufficient dual certificate, the polynomial must be 1 at the support points and have ℓ_2 norm less than 1 between the supports.

Here, we first look at the ℓ_2 norm of the interpolating polynomials.

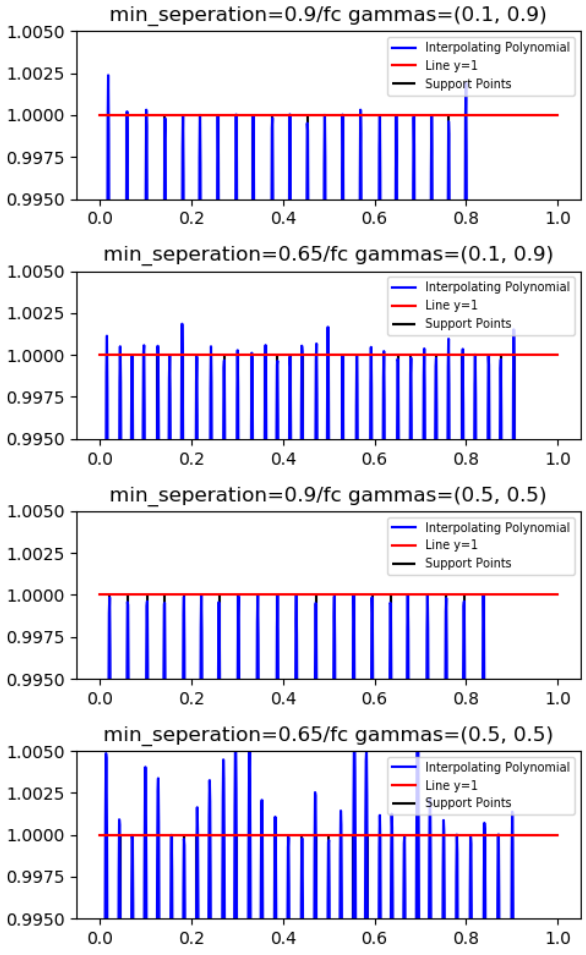

Figure 2: Interpolating for two γ choices and two minimum separations.

From this graph, we can see that for a minimum separation of $0.9/f_c$, the $\gamma = [0.5, 0.5]$ Dirichlet kernel seems to work much better in that all the values are below 1 outside of the supports and exactly one at the support. In this case, it seems that the width of

the spike is not an issue and the fast decay to zero is the more important factor in terms of performance.

On the other hand, for a minimum separation of $0.65/f_c$, this $\gamma = [0.5, 0.5]$ kernel performs significantly worse than the $\gamma = [0.1, 0.9]$ kernel in that its values outside the support points are above 1 and a lot larger than for the $\gamma = [0.1, 0.9]$ kernel. This is where we see that the width of the spike starts causing an issue and making the points near the support reach more than 1. In this case we are better off choosing the more "spiky" kernel even though its decay is still not fast enough.

Thus, it seems that for larger minimum separations, we value a faster decay to zero and for smaller minimum separations we value a thinner spike, which represents well the trade-off we expected to see within these graphs.

We next consider the derivative terms of the polynomial. Recall that in order to be a sufficient dual certificate, the derivatives must not point outside the unit hypersphere.

Here, to the right, we plot the two derivative terms of our interpolating polynomial and look at the 6th coordinate.

First note that we indeed notice that both of these terms are zero at the 6th support point.

Now, for a minimum separation of 0.9 and $\gamma = [0.5, 0.5]$ we notice that the two derivative terms are extremely close to zero and, thus, also close to each other. This means that they interfere very little with the first term of the polynomial which is what we hope to achieve.

On the other hand, as the minimum separation decreases, near the support, the two terms begin to differ a lot which makes the derivative term in the interpolating polynomial more significant relative to the first term and, thus, makes this kernel choice less adapted to our problem.

Lastly, for $\gamma = [0.1, 0.9]$, it seems like changing the minimum separation does not change much to the difference between the two terms in the second part of the polynomial. Overall, near the support they seem to differ a fair amount which makes them interfere with the first term of the polynomial. However, the fact that the difference between these two terms does not change much as the minimum separation decreases, this kernel seems like a more stable choice for our problem.

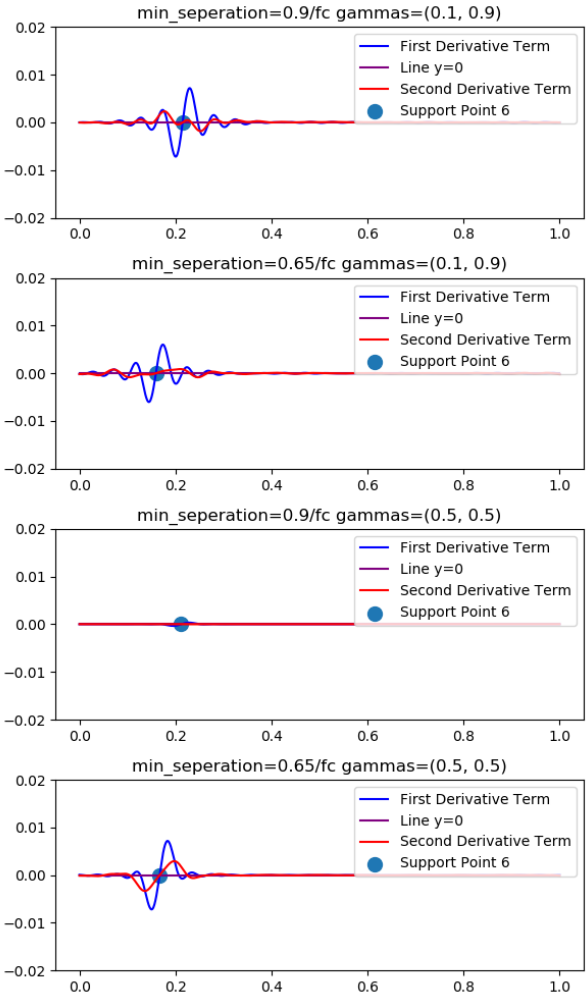

Figure 3: Two derivative terms for two γ choices and two minimum separations.

6 Discussion

Figure 2 suggests that the "fat" peaks of the kernel $\gamma = [0.5, 0.5]$ are simply *too* fat one the separation shrinks close to the limit. In the second and third subplots, we see that the transition from separation of $0.9/f_c$ to $0.5/f_c$ degrades the performance for the kernel $\gamma = [0.1, 0.9]$, but catastrophically reduces the performance of the kernel $\gamma = [0.5, 0.5]$. If a "spikier" kernel has slightly worse baseline performance (as seen in the first versus third subplot), it is more robust to shrinking minimum separation.

Our experiments with "triple" Dirichlet kernels (where γ is of length 3) showed that the best per-

formance occurred when one of the entries of γ was negligible, which suggested to us that it was a less fruitful direction than the "double" Dirichlet kernels. However, we did not take a sample of the full space of potential "triple" kernels (compared to many more "double" kernels that we experimented with) and further exploring these and kernels of higher order remains an area we would like to visit in the future.

Non-Dirichlet kernels also offer a rich avenue of exploration, as the trade-off between "spikiness" on the support and flatness between supports is a characteristic of Dirichlet kernels that may not exist in other kernels. We would like to further research the literature on interpolating kernels and splines in order to find alternative candidates for the kernel.

There is also the option of modifying our entire polynomial construction. We could add more constraints, potentially on higher-order derivatives, to attempt to control the behavior further.

Lastly, we ran most of our experiments with $f_c = 25$ in order to keep them computationally inexpensive, but many real-world applications involve substantially higher cutoff frequencies and the kernel behavior in these "toy" examples may change or degrade as f_c grows substantially higher.

7 Acknowledgments

We would like to thank Efe Onaran, Tim Kunisky, and Carlos Fernandez-Granda for their invaluable insight and patience over Zoom calls and email.

References

[1] B. Bernstein and C. Fernandez-Granda, "Deconvolution of Point Sources: A Sampling Theorem and Robustness Guarantees," *Communications on Pure and Applied Mathematics*, vol. 72, no. 6, pp. 1152–1230, 2019.

[2] 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, Jun. 2014. DOI: 10.1002/cpa.21455.

[3] C. Fernandez-Granda, "Super-Resolution of Point Sources via Convex Programming," *Information and Inference*, vol. 5, no. 3, pp. 251–303, Jul. 2015. DOI: 10.1093/imaiai/iaw005. arXiv: 1507.07034.

[4] ——, "Support detection in super-resolution," in *Proceedings of the 10th International Conference on Sampling Theory and Applications (SampTA)*, 2013, pp. 145–148.

[5] R. Heckel and M. Soltanolkotabi, "Generalized Line Spectral Estimation via Convex Optimization," *IEEE Transactions on Information Theory*, vol. 64, no. 6, pp. 4001–4023, Jun. 2018. DOI: 10.1109/TIT.2017.2757003. arXiv: 1609.08198.

[6] G. Izacard, B. Bernstein, and C. Fernandez-Granda, "A Learning-Based Framework for Line-Spectra Super-resolution," *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, vol. 2019-May, pp. 3632–3636, Nov. 2018. arXiv: 1811.05844.

[7] D. Kunisky and E. Onaran, "Large Minimum Separation," Tech. Rep., 2019, pp. 1–2.

[8] ——, "Super-Resolution of Point Sources Down to the Rayleigh Limit from Multiple Observations," Tech. Rep., 2017.

[9] Y. Lou, P. Yin, and J. Xin, "Point Source Super-resolution Via Non-convex L1 Based Methods," *Journal of Scientific Computing*, vol. 68, no. 3, pp. 1082–1100, Sep. 2016. DOI: 10.1007/s10915-016-0169-x.

[10] 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, Jan. 2015, ISSN: 00189448. DOI: 10.1109/TIT.2014.2368122. arXiv: 1303.4348.

[11] J. A. Tropp, "Algorithms for simultaneous sparse approximation. Part II: Convex relaxation," *Signal Processing*, vol. 86, no. 3, pp. 589–602, 2006. DOI: https://doi.org/10.1016/j.sigpro.2005.05.031.