

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

December 10, 2017

## 1 Introduction

### 1.1 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

$$\operatorname{argmin}_x \|x\|_{\text{TV}} \text{ subject to } \mathcal{F}_n x = y.$$

## 1.2 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

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

## 1.3 Dual Certificates

In this section, we lay out the Lagrangian duality theory for 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 [1] establishes that this is true under a minimum separation condition.

**Theorem 1 (Proposition 2.3 of [1])** 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 [1], the same ideas extend to the multiple observation case in a straightforward fashion.

**Definition 3** 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 4** 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  $\mathbf{v}$  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 [1] 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 signs  $\mathbf{a}_j$  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 probabilistic assumptions about the data generating process of the  $\mathbf{a}_j$ .

## 2 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  $(t_j, \mathbf{v}_j)$  while remaining strictly inside the unit sphere in  $\mathbb{C}^m$  elsewhere. The idea of the construction of [1] 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. The naive extension of this to the  $m$ -dimensional is to require  $\nabla q(t_j) = 0$  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 repeating the one-dimensional construction in each coordinate. We propose a different dual certificate, which does succeed numerically in certifying reconstruction down to the Rayleigh limit as  $m \rightarrow \infty$ .

### 2.1 Step 1: Relaxing Gradient Condition

The first observation is that the condition  $\nabla q(t_j) = 0$  is overkill for encouraging  $q$  to remain in the unit sphere. Intuitively, all that is really necessary is that  $\nabla q(t_j)$  not have a component pointing out of the sphere, i.e. in the direction of  $\mathbf{v}_j$ . Thus, a reasonable condition is to require  $\nabla q(t_j)$  and  $\mathbf{v}_j$ , 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  $q_\ell(t) = q_\ell^R(t) + i q_\ell^I(t)$  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  $\tilde{\mathbf{v}}_j = (v_{j,1}^R, v_{j,1}^I, \dots, v_{j,m}^R, v_{j,m}^I)$ , we can express the interpolation conditions that the dual certificate  $q$  must satisfy in terms of these new variables in essentially the same way:  $\tilde{q}(t_j) = \tilde{\mathbf{v}}_j$  for  $j \in \{1, \dots, s\}$ , and  $\|\tilde{q}(t)\|_{\ell^2} < 1$  for all  $t \notin T$ . We can now formulate the conditions we will require of our interpolant.

- I1.  $\tilde{q}(t_j) = \tilde{\mathbf{v}}_j$  for  $j \in \{1, \dots, s\}$ .
- I2.  $\langle \nabla \tilde{q}(t_j), \tilde{\mathbf{v}}_j \rangle = 0$  for  $j \in \{1, \dots, s\}$ .

## 2.2 Step 2: Kernel Interpolation

We now proceed just as in [1]: fix a real-valued low-pass kernel  $K$ , and define

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

$$q_\ell^I(t) = \sum_{j=1}^s \alpha_{j,\ell}^I K(t - t_j) + \sum_{j=1}^s \beta_{j,\ell}^I K'(t - t_j), \quad (2)$$

for coefficients  $\alpha_{j,\ell}^R, \alpha_{j,\ell}^I, \beta_{j,\ell}^R, \beta_{j,\ell}^I \in \mathbb{R}$  to be determined. We think of these as organized into four matrices,  $\boldsymbol{\alpha}^R, \boldsymbol{\alpha}^I, \boldsymbol{\beta}^R, \boldsymbol{\beta}^I \in \mathbb{R}^{s \times m}$ . We also organize the real and imaginary parts of the sign patterns  $v_{j,\ell}^R$  and  $v_{j,\ell}^I$  into matrices  $\mathbf{V}^R, \mathbf{V}^I \in \mathbb{R}^{s \times m}$  of the same shape. Conditions I1 and I2 from the previous part may now be expressed as a system of linear constraints on these coefficients, which we introduce some notation to write down concisely with some matrix algebra.

**Definition 5** For a function  $K : \mathbb{T} \rightarrow \mathbb{R}$  and an ordered point set  $T = \{t_1, \dots, t_s\}$ , define  $\mathbf{D}(K, T) \in \mathbb{R}^{s \times s}$  to have entries

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

The constraint I1 may then be written as the two linear matrix equations

$$\mathbf{D}(K, T) \boldsymbol{\alpha}^R + \mathbf{D}(K', T) \boldsymbol{\beta}^R = \mathbf{V}^R, \quad (3)$$

$$\mathbf{D}(K, T) \boldsymbol{\alpha}^I + \mathbf{D}(K', T) \boldsymbol{\beta}^I = \mathbf{V}^I, \quad (4)$$

and the constraint I2 as the one equation

$$\text{diag} \left( \mathbf{D}(K', T) \boldsymbol{\alpha}^R \mathbf{V}^{R^\top} + \mathbf{D}(K'', T) \boldsymbol{\beta}^R \mathbf{V}^{R^\top} + \mathbf{D}(K', T) \boldsymbol{\alpha}^I \mathbf{V}^{I^\top} + \mathbf{D}(K'', T) \boldsymbol{\beta}^I \mathbf{V}^{I^\top} \right) = \mathbf{0}. \quad (5)$$

For numerical calculations, it is straightforward to unfold these constraints into a vectorized linear system for the coefficients.

## 2.3 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 [1]. 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  $\alpha, \beta$ , and so its minimization subject to linear constraints is well-understood and has a convenient closed form. We now develop some notation to write down this quadratic form.

**Definition 6** For real-valued functions  $K_1, K_2 \in L^2(\mathbb{T})$  and an ordered point set  $T = \{t_1, \dots, t_s\}$ , define  $\mathbf{G}(K_1, K_2, T) \in \mathbb{R}^{s \times s}$  to have entries

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

In the special case  $K_1 = K_2 =: K$ , we simply write  $\mathbf{G}(K, T)$  for the same matrix, and observe that it is the Gram matrix of the shifted functions  $f_j(t) = K(t - t_j)$  in the Hilbert space  $L^2(\mathbb{T})$ .

Now, the  $L^2$  norm of the polynomial  $q(t)$  with real and imaginary parts as defined in (1) and (2) respectively is given by

$$\begin{aligned} \|q\|_{L^2(\mathbb{T})}^2 &= \sum_{X \in \{R, I\}} \sum_{\ell=1}^m \sum_{j=1}^s \sum_{k=1}^s \left( \alpha_{\ell,j}^X \alpha_{\ell,k}^X \langle K(\bullet - t_j), K(\bullet - t_k) \rangle + \beta_{\ell,j}^X \beta_{\ell,k}^X \langle K'(\bullet - t_j), K'(\bullet - t_k) \rangle \right. \\ &\quad \left. + \alpha_{\ell,j}^X \beta_{\ell,k}^X \langle K'(\bullet - t_j), K(\bullet - t_k) \rangle \right) \\ &= \sum_{X \in \{R, I\}} \left( \text{Tr}(\alpha^{X^\top} \mathbf{G}(K, T) \alpha^X) + \text{Tr}(\beta^{X^\top} \mathbf{G}(K', T) \beta^X) + \text{Tr}(\beta^{X^\top} \mathbf{G}(K', K, T) \alpha^X) \right. \\ &\quad \left. + \text{Tr}(\alpha^{X^\top} \mathbf{G}(K, K', T) \beta^X) \right) \\ &= \text{Tr} \left( \begin{bmatrix} \alpha^R \\ \beta^R \\ \alpha^I \\ \beta^I \end{bmatrix}^\top \left( \mathbf{I}_2 \otimes \begin{bmatrix} \mathbf{G}(K, T) & \mathbf{G}(K, K', T) \\ \mathbf{G}(K', K, T) & \mathbf{G}(K', T) \end{bmatrix} \right) \begin{bmatrix} \alpha^R \\ \beta^R \\ \alpha^I \\ \beta^I \end{bmatrix} \right), \end{aligned} \quad (6)$$

where in the last expression each block of the inner matrix is a Gram matrix of the  $2s$  functions  $K(\bullet - t_j)$  and  $K'(\bullet - t_j)$  in  $L^2(\mathbb{T})$ .

## 2.4 Numerical Evaluation

## 3 Proof

We start by writing the Lagrangian of the optimization problem where we minimize (6) with constraints I1 and I2 at (3), (4) and (5).

$$\begin{aligned}
L(\boldsymbol{\alpha}^{\{R,I\}}, \boldsymbol{\beta}^{\{R,I\}}, \boldsymbol{\lambda}^{\{1,2,3\}}) = & \sum_{X \in \{R,I\}} \left( \text{Tr}(\boldsymbol{\alpha}^{X^\top} \mathbf{G}(K, T) \boldsymbol{\alpha}^X) + \text{Tr}(\boldsymbol{\beta}^{X^\top} \mathbf{G}(K', T) \boldsymbol{\beta}^X) \right. \\
& + \text{Tr}(\boldsymbol{\beta}^{X^\top} \mathbf{G}(K', K, T) \boldsymbol{\alpha}^X) + \text{Tr}(\boldsymbol{\alpha}^{X^\top} \mathbf{G}(K, K', T) \boldsymbol{\beta}^X) \Big) \\
& - \text{Tr}((\boldsymbol{\lambda}^1)^T (\mathbf{D}(K, T) \boldsymbol{\alpha}^R + \mathbf{D}(K', T) \boldsymbol{\beta}^R - \mathbf{V}^R)) \\
& - \text{Tr}((\boldsymbol{\lambda}^2)^T (\mathbf{D}(K, T) \boldsymbol{\alpha}^I + \mathbf{D}(K', T) \boldsymbol{\beta}^I - \mathbf{V}^I)) \\
& - \text{Tr}\left((\boldsymbol{\lambda}^3)^T \left( \mathbf{D}(K', T) \boldsymbol{\alpha}^R \mathbf{V}^{R^\top} + \mathbf{D}(K'', T) \boldsymbol{\beta}^R \mathbf{V}^{R^\top} + \mathbf{D}(K', T) \boldsymbol{\alpha}^I \mathbf{V}^{I^\top} + \mathbf{D}(K'', T) \boldsymbol{\beta}^I \mathbf{V}^{I^\top} \right)\right)
\end{aligned}$$

where  $\boldsymbol{\lambda}^1, \boldsymbol{\lambda}^2 \in \mathbb{R}^{s \times m}$  and  $\boldsymbol{\lambda}^3 \in \mathbb{R}^{s \times s}$  is a diagonal matrix.

The first order optimality condition for  $L(\boldsymbol{\alpha}^{\{R,I\}}, \boldsymbol{\beta}^{\{R,I\}}, \boldsymbol{\lambda}^{\{1,2,3\}})$  reads as

$$\begin{aligned}
\nabla_{\boldsymbol{\alpha}^R} L(\boldsymbol{\alpha}^{\{R,I\}}, \boldsymbol{\beta}^{\{R,I\}}, \boldsymbol{\lambda}^{\{1,2,3\}}) &= \mathbf{0} \quad \Leftrightarrow \\
(G(K, T) + G(K, T)^T) \boldsymbol{\alpha}^R + (G(K, K', T)^T + G(K', K, T)^T) \boldsymbol{\beta}^R - D(K, T)^T \boldsymbol{\lambda}^1 - D(K', T)^T \boldsymbol{\lambda}^3 \mathbf{V}^R &= \mathbf{0} \\
\nabla_{\boldsymbol{\beta}^R} L(\boldsymbol{\alpha}^{\{R,I\}}, \boldsymbol{\beta}^{\{R,I\}}, \boldsymbol{\lambda}^{\{1,2,3\}}) &= \mathbf{0} \quad \Leftrightarrow \\
(G(K', T) + G(K', T)^T) \boldsymbol{\beta}^R + (G(K', K, T)^T + G(K, K', T)^T) \boldsymbol{\alpha}^R - D(K', T)^T \boldsymbol{\lambda}^1 - D(K'', T)^T \boldsymbol{\lambda}^3 \mathbf{V}^R &= \mathbf{0} \\
\nabla_{\boldsymbol{\alpha}^I} L(\boldsymbol{\alpha}^{\{R,I\}}, \boldsymbol{\beta}^{\{R,I\}}, \boldsymbol{\lambda}^{\{1,2,3\}}) &= \mathbf{0} \quad \Leftrightarrow \\
(G(K, T) + G(K, T)^T) \boldsymbol{\alpha}^I + (G(K, K', T)^T + G(K', K, T)^T) \boldsymbol{\beta}^I - D(K, T)^T \boldsymbol{\lambda}^2 - D(K', T)^T \boldsymbol{\lambda}^3 \mathbf{V}^I &= \mathbf{0} \\
\nabla_{\boldsymbol{\beta}^I} L(\boldsymbol{\alpha}^{\{R,I\}}, \boldsymbol{\beta}^{\{R,I\}}, \boldsymbol{\lambda}^{\{1,2,3\}}) &= \mathbf{0} \quad \Leftrightarrow \\
(G(K', T) + G(K', T)^T) \boldsymbol{\beta}^I + (G(K', K, T)^T + G(K, K', T)^T) \boldsymbol{\alpha}^I - D(K', T)^T \boldsymbol{\lambda}^2 - D(K'', T)^T \boldsymbol{\lambda}^3 \mathbf{V}^I &= \mathbf{0}
\end{aligned}$$

Define matrix  $\tilde{G} \in \mathbb{R}^{2s \times 2s}$  as

$$\tilde{G} = \begin{bmatrix} G(K, T) + G(K, T)^T & G(K, K', T) + G(K', K, T)^T \\ G(K, K', T)^T + G(K', K, T) & G(K', T) + G(K', T)^T \end{bmatrix}$$

and  $\tilde{D} \in \mathbb{R}^{2s \times 2s}$  as

$$\tilde{D} = \begin{bmatrix} D(K, T) & D(K', T) \\ D(K', T) & D(K'', T) \end{bmatrix}$$

Then first order conditions in compact form,

$$\tilde{G} \begin{bmatrix} \boldsymbol{\alpha}^R \\ \boldsymbol{\beta}^R \end{bmatrix} = \tilde{D}^T \begin{bmatrix} \boldsymbol{\lambda}^1 \\ \boldsymbol{\lambda}^3 \mathbf{V}^R \end{bmatrix}$$

and

$$\tilde{G} \begin{bmatrix} \boldsymbol{\alpha}^I \\ \boldsymbol{\beta}^I \end{bmatrix} = \tilde{D}^T \begin{bmatrix} \boldsymbol{\lambda}^2 \\ \boldsymbol{\lambda}^3 \mathbf{V}^I \end{bmatrix}$$

## References

- [1] Carlos Fernandez-Granda. Super-resolution of point sources via convex programming. *Information and Inference: A Journal of the IMA*, 5(3):251–303, 2016.