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

February 28, 2018

1 Formal Setup

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, ϕ 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

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

We may then think of the data we observe as simply the $2f_c + 1$ values $y(-f_c), \ldots, y(f_c)$. 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

$$\begin{array}{ll} \text{minimize} & \|x\|_{\mathsf{TV}} \\ \text{subject to} & \mathcal{F}_n x = y. \end{array}$$

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 $T = \{t_1, \ldots, t_s\}$ of x^*, but having varying amplitudes a_j. To formalize this, our true signal is now

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

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

$$\begin{array}{ll} \text{minimize} & \|x\|_{\mathsf{gTV}} \\ \text{subject to} & \begin{bmatrix} \mathcal{F}_n x_1 & \mathcal{F}_n x_2 \cdots \mathcal{F}_n x_m \end{bmatrix}^\top = \boldsymbol{Y}. \end{array}$$

1.3 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 μ^0 to each point of the set.*

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

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

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 [**?**] 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 [?]) *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 [?], 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 μ^{m-1} to each point of the set.*

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

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

*is an m-*observation dual certificate *for v if $q(t_j) = \boldsymbol{v}_j$ 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 $\boldsymbol{v}_j = \boldsymbol{a}_j / \|\boldsymbol{a}_j\|_{\ell^2}$ exists, then x^* is the unique solution of* gTV *minimization for the super-resolution problem.*

However, while numerical experiments in [?] suggest that when the amplitude vectors $\boldsymbol{a}_j$ are taken randomly then recovery is possible down to a lower critical minimum separation which accumulates at $\frac{1}{2}\lambda_c$ as $m \to \infty$, the same uniform argument cannot apply—indeed, it is always possible that the amplitudes $\boldsymbol{a}_j$ 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 $\boldsymbol{a}_j$. One simple such assumption that captures the idea of these points being in "general position" is to assume that the signs $\boldsymbol{v}_j = \boldsymbol{a}_j / \|\boldsymbol{a}_j\|_{\ell^2}$ are i.i.d. isotropically random unit vectors.

2 Dual Certificate Construction

Recall that we are interested in building a low-pass trigonometric polynomial $q : \mathbb{T} \to \mathbb{C}^m$ interpolating the points $(t_j, \boldsymbol{v}_j)$ while remaining strictly inside the unit sphere in $\mathbb{C}^m$ elsewhere. The general idea of such constructions is usually to choose some *interpolation basis* of polynomials, and build q as a linear combination of them under some constraints. We first describe a simple approach to choosing interpolation bases based on constraints and a variational heuristic, that justifies several existing techniques and motivates our own.

2.1 Equivalence of L^2 Norm Minimization and Interpolation

Minimizing $\|q\|_{L^2([0,1])}$ under some interpolation constraints is equivalent to interpolating with a restricted basis, the basis arising from the nature of the constraints. This general principle can be implemented in specific cases with simple Lagrange multiplier calculations. Consider, for instance, the problem

$$\begin{array}{ll} \text{minimize} & \|q\|_{L^2([0,1])} \\ \text{subject to} & q(t_j) = \boldsymbol{v}_j \quad \forall j \in [s] \end{array}$$

Let us write

$$q_{\boldsymbol{c}}(t) = \sum_{f=-f_c}^{f_c} c_f \exp(2\pi i f t),$$

and define the matrix $\boldsymbol{F}_T \in \mathbb{C}^{s \times (2f_c+1)}$ by $(\boldsymbol{F}_T)_{jf} = \exp(2\pi i f t_j)$, with $-f_c \leq f \leq f_c$. Then, the calculation of the values of $q_{\boldsymbol{c}}$ on T is a matrix-vector multiplication,

$$\begin{bmatrix} q_{\boldsymbol{c}}(t_1) \\ \vdots \\ q_{\boldsymbol{c}}(t_s) \end{bmatrix} = \boldsymbol{F}_T \boldsymbol{c},$$

and $\|q_{\boldsymbol{c}}\|_{L^2([0,1])} = \|\boldsymbol{c}\|_2$. Thus, the above problem is equivalently

$$\begin{array}{ll} \text{minimize} & \|\boldsymbol{c}\|_2 \\ \text{subject to} & \boldsymbol{F}_T \boldsymbol{c} = \boldsymbol{v} \end{array}$$

We write a Lagrangian $L(\boldsymbol{c}, \boldsymbol{\alpha}) = \frac{1}{2}\|\boldsymbol{c}\|_2^2 - \langle \boldsymbol{\alpha}, \boldsymbol{F}_T \boldsymbol{c} - \boldsymbol{v} \rangle$, and upon differentiating and solving the first-order optimality condition obtain $\boldsymbol{c} = \boldsymbol{F}_T^* \boldsymbol{\alpha}$. Generally, we have

$$q_{\boldsymbol{F}_T^* \boldsymbol{\alpha}}(t) = \sum_{f=-f_c}^{f_c} \left(\sum_{j=1}^{s} \exp(-2\pi i f t_j) \alpha_j \right) \exp(2\pi i f t) = \sum_{j=1}^{s} \alpha_j K(t - t_j)$$

where K is the Dirichlet kernel. We therefore have found the formal equivalence:

$$\begin{array}{ll} \text{minimize} & \|q\|_{L^2([0,1])} \\ \text{subject to} & q(t_j) = \boldsymbol{v}_j \quad \forall j \in [s] \end{array} \quad \leftrightarrow \quad \begin{array}{ll} \text{find} & q(t) = \sum_{j=1}^{s} \alpha_j K(t - t_j) \\ \text{subject to} & q(t_j) = \boldsymbol{v}_j \quad \forall j \in [s] \end{array}$$

In the same way, one obtains a justification for the popular trick for allowing one to force the derivative of q to be zero at each t_j by adding terms $\beta_j K'(t - t_j)$ to the sum above: an analogous calculation shows

$$\begin{array}{ll} \text{minimize} & \|q\|_{L^2([0,1])} \\ \text{subject to} & q(t_j) = v_j \quad \forall j \in [s] \\ & q'(t_j) = 0 \quad \forall j \in [s] \end{array} \quad \leftrightarrow \quad \begin{array}{ll} \text{find} & q(t) = \sum_{j=1}^{s} \alpha_j K(t - t_j) + \\ & \qquad \sum_{j=1}^{s} \beta_j K'(t - t_j) \\ \text{subject to} & q(t_j) = v_j \quad \forall j \in [s] \\ & q'(t_j) = 0 \quad \forall j \in [s] \end{array}$$

2.2 Relaxing the Derivative Condition

As mentioned before, in the multidimensional setting, we do not expect setting the derivative to equal zero at each interpolation point will allow us to lower the minimum separation threshold substantially beyond where the analogous one-dimensional dual certificate is feasible. Instead, we propose a new construction where this condition is relaxed. Rather than setting the derivative at each interpolation point to equal zero, we only constrain it to the tangent plane to the complex unit sphere at that interpolation point. [TODO: something about not dividing real from complex for clarity though this helps in numerics]. Following the ideas in the previous section, we propose as a dual certificate the polynomial solving the following optimization problem:

$$\begin{aligned} \text{minimize} \quad & \|q\|_{L^2([0,1])} \\ \text{subject to} \quad & q(t_j) = \boldsymbol{v}_j \quad \forall j \in [s] \\ & \langle q'(t_j), \boldsymbol{v}_j \rangle = 0 \quad \forall j \in [s] \end{aligned}$$

Performing the translation to an interpolation problem as before gives the following:

$$\begin{aligned} \text{find} \quad & q_k(t) = \sum_{j=1}^{s} \alpha_{j,k} K(t - t_j) + \sum_{j=1}^{s} \beta_j V_{j,k} K'(t - t_j) \\ \text{subject to} \quad & q_k(t_j) = V_{j,k} \quad \forall j \in [s], k \in [m] \\ & \langle q'(t_j), \boldsymbol{v}_j \rangle = 0 \quad \forall j \in [s] \end{aligned}$$

where we organize the sign pattern vectors $\boldsymbol{v}_j$ as the rows of the matrix $\boldsymbol{V} \in \mathbb{C}^{s \times m}$. Note that if we instead took the constraint $q'(t_j) = \boldsymbol{0}$, then the second term would involve terms $\beta_{j,k} K'(t - t_j)$ in coordinate k, but here we are only allowed *one* set of coefficients β_j to use across all coordinates. This, along with the randomness of the $V_{j,k}$, essentially ensures that the second term cannot often be large: a single vector $\boldsymbol{\beta}$ cannot align with many weakly dependent random vectors $\boldsymbol{V}_{\bullet,k}$ at once.

2.3 Explicit Solution of Interpolation Problem

Next, we derive concise formulae solving for the coefficients $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ (when this is possible, which will later need some justification). Let $\boldsymbol{K}_r = (\frac{d^r}{dt^r} K(t_i - t_j))_{i,j=1}^{s}$, then we may write

$$\begin{aligned} (q_k(t_j))_{j \in [s], k \in [m]} &= \boldsymbol{K}_0 \boldsymbol{\alpha} + \boldsymbol{K}_1 \mathsf{diag}(\boldsymbol{\beta}) \boldsymbol{V} \\ (q_k'(t_j))_{j \in [s], k \in [m]} &= \boldsymbol{K}_1 \boldsymbol{\alpha} + \boldsymbol{K}_2 \mathsf{diag}(\boldsymbol{\beta}) \boldsymbol{V} \end{aligned}$$

Thus, the interpolation equations may be written

$$\begin{aligned} \boldsymbol{K}_0 \boldsymbol{\alpha} + \boldsymbol{K}_1 \mathsf{diag}(\boldsymbol{\beta}) \boldsymbol{V} &= \boldsymbol{V} \\ \mathsf{diag}((\boldsymbol{K}_1 \boldsymbol{\alpha} + \boldsymbol{K}_2 \mathsf{diag}(\boldsymbol{\beta}) \boldsymbol{V}) \boldsymbol{V}^*) &= \boldsymbol{0} \end{aligned}$$

From the first equation we obtain

$$\boldsymbol{\alpha} = \boldsymbol{K}_0^{-1} (\boldsymbol{I}_s - \boldsymbol{K}_1 \mathsf{diag}(\boldsymbol{\beta})) \boldsymbol{V},$$

and substituting this into the second obtain

$$\boldsymbol{\beta} = \frac{\mathsf{diag}(\boldsymbol{V}\boldsymbol{V}^*\boldsymbol{K}_0^{-1}\boldsymbol{K}_1)}{\mathsf{diag}(\boldsymbol{V}\boldsymbol{V}^*(\boldsymbol{K}_2 - \boldsymbol{K}_1\boldsymbol{K}_0^{-1}\boldsymbol{K}_1))},$$

where division in the latter is coordinatewise, and these formulae hold provided $\boldsymbol{K}_0$ is invertible and $\mathsf{diag}(\boldsymbol{V}\boldsymbol{V}^*(\boldsymbol{K}_2 - \boldsymbol{K}_1\boldsymbol{K}_0^{-1}\boldsymbol{K}_1))$ is nowhere zero.