Non-Negative Super-Resolution: Simplified and Stabilized

Jared Tanner

joint with Eftekhari, Thompson, Toader, and Tyagi

GDO 2019, Cluj, 10^{th} April 2019

University of Oxford & Alan Turing Institute[1]

[1]Supported by: EPSRC, the National Physical Laboratory, and The Turing Institute

The team:

Armin
Eftekhari
EPFL

Andrew
Thompson
NPL

Bogdan
Toader
Oxford

Hemant
Tyagi
Inria Lille

Some simplified models for data:

- Compressed Sensing: $\min_{x,s-sparse} \|A\Phi x - y\|$
  Relies on knowledge of data being well represented by Φx
  For $x \in \mathbb{R}^n$ need A to have $m \sim s\log(n/m)$ rows
  unless more structure is imposed.

- Matrix Completion: $\min_{X,rank(X)=r} \|\mathcal{A}(X) - Y\|$
  Relies on matrices of interest being approximately low rank
  For $X \in \mathbb{R}^{m\times n}$ need $\mathcal{A}(X)$ to map to $p \sim r(m+n-r)$
  Representation of X via singular vec. which vary continuously.

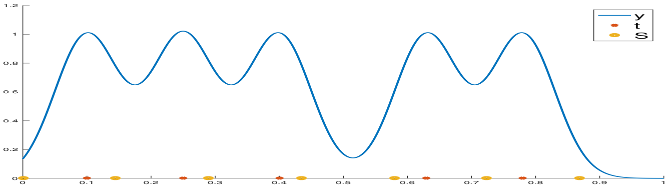
- Super-resolution (grid free CS): $\min_{x,|\text{supp}(x)|=k} \|\phi * x - y\|$
  Uses knowledge of $\phi(s,t)$ and relies on $x(t) = \sum_{i=1}^{k} a_i \delta_{t_i}$
  Similar to CS, but locations of t_i vary continuously.
  Discretizing t gives CS with A convolutional, highly coherent.

Super-resolution model:

Objects of interest essentially point sources as compared to the width of the measurement point-spread response function, $\phi(s,t)$. Have access to data of the form:

$$y(s) = \sum_{i=1}^{k} a_i \phi(s, t_i)$$

where t_i denotes locations of point source with magnitude a_i. Motivated by microscopy let $\phi(s,t_i) = \exp\left(-\frac{|s-t_i|^2}{\sigma^2}\right)$, then:



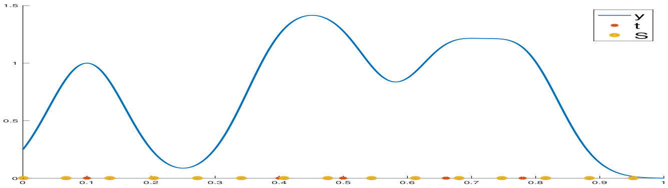
Seek to recover t_i and a_i for $i = 1, \ldots, k$ from samples of $y(s)$.

Super-resolution model:

Objects of interest essentially point sources as compared to the width of the measurement point-spread response function, $\phi(s,t)$. Have access to data of the form:

$$y(s) = \sum_{i=1}^{k} a_i \phi(s, t_i)$$

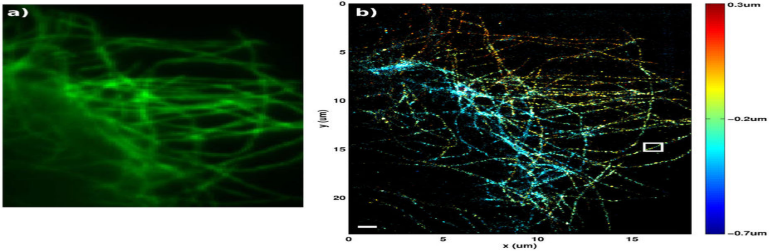
where t_i denotes locations of point source with magnitude a_i. Motivated by microscopy let $\phi(s,t_i) = \exp\left(-\frac{|s-t_i|^2}{\sigma^2}\right)$, then:



Seek to recover t_i and a_i for $i = 1, \ldots, k$ from samples of $y(s)$.

Fluorescence image [Barsic, Grover, Piestun; Sci. Rep. 14']

Three-dimensional super-resolution and localization of dense clusters of single molecules. Used Gaussian blurring model.



"A standard fluorescence image is shown in (a). The 3D super-resolution image (b) of labeled tubulin in PtK1 cells demonstrates that the method can be applied to localization-based super-resolution imaging with a wide field of view."

The Nobel Prize in Chemistry 2014 was awarded jointly to Eric Betzig, Stefan W. Hell and William E. Moerner "for the development of super-resolved fluorescence microscopy".

Error metric: locality different from normal comp. sensing:

Consider discrete measure $x(t) = \sum_{i=1}^{k} a_i \delta_{t_i} \geq 0$,
$y(s) = \phi(s,t) * x(t) + e(s)$, and general measure $\hat{x} \geq 0$.

- Distance between measures via Wasserstein distance:

$$d_W(x, \hat{x}) = \inf_{\gamma} \int_{I \times I} |\tau_1 - \tau_2| \cdot \gamma(d\tau_1, d\tau_2)$$

  where $\hat{x}(t) = \int_I \gamma(d\tau_1, \tau_2)$
  and $x(t) = \int_I \gamma(\tau_1, d\tau_2)$
- Energy conservation depends on source sample locations, requires minimum separation for signed discrete measures:

$$\sum_{j=1}^{m} y(s_j)^2 \geq Const. \sum_{i=1}^{k} a_i^2$$

  for universal $Const.$ requires $\Delta = \min_{i \neq l} |t_i - t_l|$ lower bounded and for local $\phi(s,t)$ requires s_j sufficiently near t_i.

Super-resolution without separation: TV-minimization

Theorem (Schiebinger, Robeva, Recht 2018)

Suppose $x(t) = \sum_{i=1}^{k} a_i \delta_{t_i}$ *with:* $a_i > 0$ *and* $t_i \in [0,1]$ *for* $i = 1, \ldots, k$ *and* $\phi(s,t)$ *is a Tchebysheff system (e.g.* $\phi(s,t) = \exp\left(-\frac{|s-t|^2}{\sigma^2}\right)$*) and given* $y(s_j) = \phi(s_j, t) * x$ *for* $j = 1, \ldots, m$ *for* $m > 2k$ *and for* $z(t) \geq 0$ *let*

$$\hat{x} = argmin_z \int dz(t) \quad subject\ to \quad y(s_j) = \phi(s_j, t) * dz(t),$$

then $\hat{x} = x$.

Main innovation: no need for minimum separation $\Delta = \min_{i \neq j} |t_i - t_j|$ which must be bounded for x signed; typically $m \sim \Delta^{-1}$ which is analogous to uniform sampling near sources.

Super-resolution without separation: uniqueness

Theorem (Eft. Tan. Tho. Toa. Tya. 2018)

Suppose $x(t) = \sum_{i=1}^{k} a_i \delta_{t_i}$ *with:* $a_i > 0$ *and* $t_i \in [0, 1]$ *for* $i = 1, \ldots, k$ *and* $\phi(s, t)$ *is a Tchebysheff system (e.g.* $\phi(s, t) = \exp\left(-\frac{|s-t|^2}{\sigma^2}\right)$*) and given* $y(s_j) = \phi(s_j, t) * x(t)$ *for* $j = 1, \ldots, m$ *with* $m > 2k$ *and for* $\hat{x} \geq 0$ *let*

$$y(s_j) = \phi(s_j, t) * d\hat{x}(t),$$

then there is a unique k*-sparse solution, e.g.* $\hat{x} = x$*.*

Super-resolution without separation: uniqueness

Theorem (Eft. Tan. Tho. Toa. Tya. 2018)

Suppose $x(t) = \sum_{i=1}^{k} a_i \delta_{t_i}$ *with:* $a_i > 0$ *and* $t_i \in [0, 1]$ *for* $i = 1, \ldots, k$ *and* $\phi(s, t)$ *is a Tchebysheff system (e.g.* $\phi(s, t) = \exp\left(-\frac{|s-t|^2}{\sigma^2}\right)$*) and given* $y(s_j) = \phi(s_j, t) * x(t)$ *for* $j = 1, \ldots, m$ *with* $m > 2k$ *and for* $\hat{x} \geq 0$ *let*

$$y(s_j) = \phi(s_j, t) * d\hat{x}(t),$$

then there is a unique k*-sparse solution, e.g.* $\hat{x} = x$*.*

Result by Schiebinger, Robeva, and Recht is true for all non-negative measures matching the measurements; no need for TV. Essentially a corollary by Karlin from the 60s.

Super-resolution without separation: uniqueness

Theorem (Eft. Tan. Tho. Toa. Tya. 2018)

Suppose $x(t) = \sum_{i=1}^{k} a_i \delta_{t_i}$ *with:* $a_i > 0$ *and* $t_i \in [0,1]$ *for* $i = 1, \ldots, k$ *and* $\phi(s,t)$ *is a Tchebysheff system (e.g.* $\phi(s,t) = \exp\left(-\frac{|s-t|^2}{\sigma^2}\right)$*) and given* $y(s_j) = \phi(s_j, t) * x(t)$ *for* $j = 1, \ldots, m$ *with* $m > 2k$ *and for* $\hat{x} \geq 0$ *let*

$$y(s_j) = \phi(s_j, t) * d\hat{x}(t),$$

then there is a unique k*-sparse solution, e.g.* $\hat{x} = x$*.*

Result by Schiebinger, Robeva, and Recht is true for all non-negative measures matching the measurements; no need for TV. Essentially a corollary by Karlin from the 60s.

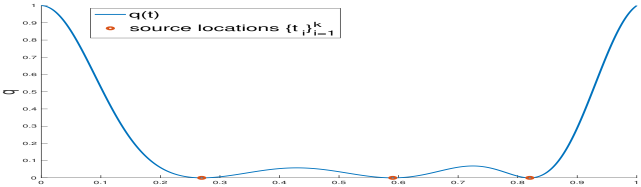
Similar result from compressed sensing of vectors [Donoho, Tanner 2010], roughly $y = Ax$ for $x \geq 0$ and s-sparse, then x unique if $y \in \mathbb{R}^m$ for m the same as if one solved ℓ^1 minimization, e.g. for $m \geq 2s \log(n/m)$ for $m/n \ll 1$.

Sketch of the proof: dual polynomial 1

Lemma

Let x be a nonnegative k-sparse atomic measure supported on $T = \{t_i\}_{i=1}^k \in I$, and $y(s) = \phi(s,t) * x$, then, x is the unique solution to $y(s) = \phi(s,t) * z$ over $z \geq 0$ (including non-atomic measures) if

- *the $k \times m$ matrix $[\phi(s_j, t_i)]_{i=1,j=1}^{i=k,j=m}$ is full rank (Tchebysheff systems satisfy this for all sample and source sequences), and*
- *there exist real coefficients $\{b_j\}_{j=1}^m$ and dual polynomial $q(t) = \sum_{j=1}^m b_j \phi(s_j, t)$ such that q is nonnegative on I and vanishes only on the set of sources $T = \{t_i\}_{i=1}^k$.*



Sketch of the proof: dual polynomial 2

Proof:

Let $x \geq 0$ be a nonnegative k-sparse atomic measure supported on $T = \{t_i\}_{i=1}^{k} \in I$ and both x and $\hat{x} \geq 0$ satisfy $y(s) = \phi(s,t) * x(t) = \phi(s,t) * \hat{x}(t)$ for $s \in \{s_j\}_{j=1}^{m}$. Let $q(t) = \sum_{j=1}^{m} b_j \phi(s_j, t) \geq 0$ satisfy $q(t_i) = 0$ for $t_i \in T$; then let $h(t) = \hat{x}(t) - x(t)$ and note $\int_I h(t)\phi(s_j,t) = 0$ for $j = 1, \cdot m$.

- $\text{supp}(\hat{x}) = \text{supp}(x)$: Consider

$$0 = \sum_{j=1}^{m} b_j \int_I \phi(s_j,t) dh(t) = \int_I q(t) dh(t) = \int_{I/T} q(t) dh(t) \geq 0,$$

  as $q(t)h(t) > 0$ for $t \in I/T$ requires $h(t) = 0$ to $t \in I/T$.
- $\hat{x} = x$: given $h(t) = \sum_{i=1}^{k} c_i \delta_{t_i}$ the m measurements equalities can be expressed as a linear system $\Phi c = 0$ where $\Phi_{i,j} = \phi(s_i, t_j)$, if invertible has only the trivial solution $c = 0$.

Existence of $q(t)$ for $\phi(s,t)$ Chebyshev System; e.g. Gaussian.

Non-negative super-resolution stability: Gaussian, local

Theorem (Eft. Tan. Tho. Toa. Tya. 2018)

Let $I = [0,1]$ and consider a k-sparse nonnegative measure x supported on $T \subset int(I)$. Consider also an arbitrary increasing sequence $\{s_j\}_{j=1}^m \subset \mathbb{R}$ and any $\hat{x} \geq 0$ satisfy $\sum_{j=1}^m \left| y_j - \int_I \exp\left(-\frac{|s_j - t|^2}{\sigma^2}\right) \hat{x}(dt)\right|^2 \leq \delta^2$. If technical conditions are satisfied (stated in a few slides) and $\epsilon = \frac{\sigma^2}{2}\delta$ then:

$$\left| \int_{t_i-\epsilon}^{t_i+\epsilon} \hat{x}(dt) - a_i \right| \leq \left[(c_1 + F_1) \cdot \delta + c_2 \frac{\|\hat{x}\|_{TV}}{\sigma^2} \cdot \epsilon \right] F_2$$

Theorem (Eft. Tan. Tho. Toa. Tya. 2018)

Let $I = [0,1]$ *and consider a k-sparse nonnegative measure* x *supported on* $T \subset int(I)$. *Consider also an arbitrary increasing sequence* $\{s_j\}_{j=1}^m \subset \mathbb{R}$ *and any* $\hat{x} \geq 0$ *satisfy* $\sum_{j=1}^m \left| y_j - \int_I \exp\left(-\frac{|s_j - t|^2}{\sigma^2}\right) \hat{x}(dt) \right|^2 \leq \delta^2$. *If technical conditions are satisfied (stated in a few slides) and* $\epsilon = \frac{\sigma^2}{2}\delta$ *then:*

$$\left| \int_{t_i-\epsilon}^{t_i+\epsilon} \hat{x}(dt) - a_i \right| \leq \left[(c_1 + F_1) \cdot \delta + c_2 \frac{\|\hat{x}\|_{TV}}{\sigma^2} \cdot \epsilon \right] F_2$$

where, for precise formulae are available for $F_1(k, \Delta(T), \sigma, \epsilon)$ *and* $F_2(\Delta(T), \sigma, \lambda)$, *and in some settings the overall bound can be simplified to be proportional to* $\delta^{1/6}$.

Non-negative super-resolution stability: Wasserstein

Theorem (Eft. Tan. Tho. Toa. Tya. 2018)

Let $I = [0,1]$ and consider a k-sparse nonnegative measure x supported on $T \subset int(I)$. Consider also an arbitrary increasing sequence $\{s_j\}_{j=1}^m \subset \mathbb{R}$ and any $\hat{x} \geq 0$ satisfy

$$\sum_{j=1}^{m} \left| y_j - \int_I \exp\left(-\frac{|s_j - t|^2}{\sigma^2}\right) \hat{x}(dt) \right|^2 \leq \delta^2. \qquad (1)$$

If technical conditions are satisfied and $\epsilon \leq \Delta/2$, then

$$d_{GW}(x, \hat{x}) \leq F(k, \Delta, \sigma) \cdot \delta + \|x\|_{TV} \cdot \epsilon,$$

where d_{GW} is the generalized Wasserstein distance. If $\sigma < 3^{-1/2}$ and $\Delta > \sigma\sqrt{\log(5/\sigma^2)}$ then $d_{GW}(x, \hat{x}) \leq F_3(k, \Delta, \sigma) \cdot \delta^{1/7}$.

Sketch of the proof: dual polynomial with $\delta > 0$

Lemma

Let $\widehat{x}$ be a solution of Program (1) and set $h = \widehat{x} - x$ to be the error. Consider a bounded function $f : \mathbb{R} \to \mathbb{R}_+$ such that $f(0) = 0$ and also a positive scalar $\bar{f}$. Suppose that there exist a positive $\epsilon \le \min_{i,j} |t_i - t_j|$, let $T_{i,\epsilon} = [t_i - \epsilon, t_i + \epsilon]$, real coefficients $\{b_j\}_{j=1}^m$, and a polynomial $q(t) = \sum_{j=1}^m b_j \phi(s_j, t)$ such that

$$q(t) \ge F(t) := \begin{cases} f(t - t_i), & \text{for } i \in [k] \text{ with } t \in T_{i,\epsilon}, \\ \bar{f}, & \text{elsewhere on } I, \end{cases}$$

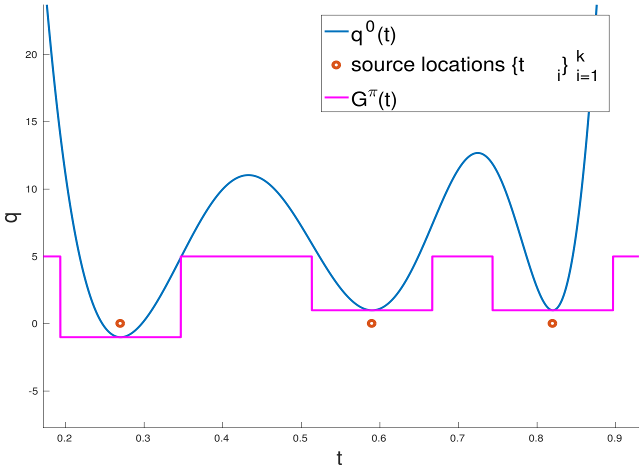
where the equality holds on T. Then we have that

$$\bar{f} \int_{T_\epsilon^C} h(dt) + \sum_{i=1}^k \int_{T_{i,\epsilon}} f(t - t_i)\, h(dt) \le 2\|b\|_2 \delta,$$

where $b \in \mathbb{R}^m$ is the vector formed by the coefficients $\{b_j\}_{j=1}^m$.

Sketch of the proof: dual polynomial with $\delta > 0$

Quality of bound on $h = \hat{x} - x$ determined by dual polynomial, controlling size of $\bar{f}$, how quickly $f(t)$ grows away from 0, and $\|b\|_2$



Those technical conditions in the theorems:

When the window function is a Gaussian $\phi(s,t) = e^{-\frac{|s-t|^2}{\sigma^2}}$, 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 $i \in [k]$ and $\eta = O(\sigma^2)$, there exists a pair of samples $s, s' \subset S$ with $s' - s = \eta$ such that $|s - t_i| \leq \eta$ and for η small enough (quantified in the paper),
3. Sources away from the boundary: $\sigma\sqrt{\log(1/\eta)} \ll t_i, s_j \ll 1 - \sigma\sqrt{\log(1/\eta)}$ for every $i \in [k]$ and $j \in [2 : m-1]$,
4. Minimum separation of sources: $\sigma \leq \sqrt{2}$ and $\Delta > \sigma\sqrt{\log\left(3 + \frac{4}{\sigma^2}\right)}$.

Roughly this translates to $m \sim \Delta^{-1}$ and σ can't be to large.

Stability implied for windows that are T^*-systems

Definition of $T*$-systems

For an even integer m, real-valued functions $\{\phi_j\}_{j=0}^m$ form a T*-system on $I = [0, 1]$ if the following holds for every $T = \{t_1, t_2, \ldots, t_k\} \subset I$ when $\rho > 0$ is sufficiently small. For any increasing sequence $\tau = \{\tau_l\}_{l=0}^m \subset I$ such that

- $\tau_0 = 0$, $\tau_m = 1$,
- except exactly three points, namely τ_0, τ_m, and say $\tau_{\underline{l}} \in \text{int}(I)$, the other points belong to $T_\rho = \cup_{i=1}^k [t_i - \rho, t_i + \rho]$,
- every $T_{i,\rho} = [t_i - \rho, t_i + \rho]$ contains an even number of points,

we have that

1. the determinant of the $(m+1) \times (m+1)$ matrix $M_\rho := [\phi_j(\tau_l)]_{l,j=0}^m$ is positive, and
2. the magnitudes of all minors of M_ρ along the row containing $\tau_{\underline{l}}$ approach zero at the same rate when $\rho \to 0$.

Some extensions:

- The stability results can be trivially generalized to $x \geq 0$ which is not atomic, but can be well approximated by a k-atomic measure with a prescribed separation between the atoms.
- The can be adapted to the setting where there are some unresolvable sources so that the sample complexity can be relaxed; results show the average of any consistent solution is close to the sum of the unresolved sources.
- The results can be extended to higher dimensional settings for samples on a cartesian grid by appropriately combining the one dimensional dual polynomials, but requires the number of measurements to be quadratic in the number of measurements (not as trivial as it sounds).
- In limited settings we have extended some results to the setting of continuous paths in higher dimensions; preliminary.

Conclusions:

- Non-negative super-resolution is robust to model misfit and additive noise
- All solutions whose measurements are within δ in ℓ^2 are within δ of noise free case in Wasserstein distance.
- Robust even for sources within δ, or high noise regime; let $\tilde{T}_{i,\epsilon}$ be the set of overlapping $T_{i,\epsilon}$, then

$$\left| \int_{\tilde{T}_{i,\epsilon}} \hat{x}(dt) - \sum_{r=1}^{p} a_{i(r)} \right| \sim \delta^{1/7}$$

- When considering algorithms for non-negative super-resolution, seek fast methods for feasibility problem, not necessarily TV; justifies use of conditional gradient and non-linear heuristics.

Literature: very incomplete

- Towards a mathematical theory of super-resolution, by Candes and Fernandez-Granda; Communications on Pure and Applied Mathematics (2014)
- Support Recovery for Sparse Deconvolution of Positive Measures, by Denoyelle, Duval, and Peyre; J. of Fourier Analysis and Applications (2017).
- Super-resolution without separation, by Schiebinger, Robeva, and Recht; Information and Inference (2018).
- Demixing sines and spikes: robust spectral super-resolution in the presence of outliers, by Fernandez-Granda, Tang, Wang, and Zheng; Information and Inference (2018)
- Non-negative super-resolution: simplified and stabilized, by Eftekhari, Tanner, Thompson, Toader, and Tyagi; preprint (2018).