Super-resolution of Positive Sources

Veniamin I. Morgenshtern

Statistics Department, Stanford

Joint work with Emmanuel Candès
Diffraction limits resolution:

\[ \lambda = \frac{\lambda_{\text{LIGHT}}}{2 \cdot \sin(\theta)} \]

Ernst Abbe
Diffraction limits resolution: \( \lambda_c = \frac{\lambda_{\text{LIGHT}}}{2n \sin(\theta)} \)
Eric Betzig, Stefan W. Hell and William E. Moerner are awarded the Nobel Prize in Chemistry 2014 for having bypassed a presumed scientific limitation stipulating that an optical microscope can never yield a resolution better than 0.2 micrometres. Using the fluorescence of molecules, scientists can now monitor the interplay between individual molecules inside cells; they can observe disease-related proteins aggregate and they can track cell division at the nanolevel.

Red blood cells, bacteria, yeast cells and spermatozoids. When scientists in the 17th century for the first time studied living organisms under an optical microscope, a new world opened up before their eyes. This was the birth of microbiology, and ever since, the optical microscope has been one of the most important tools in the life-sciences toolbox. Other microscopy methods, such as electron microscopy, require preparatory measures that eventually kill the cell.

Glowing molecules surpassing a physical limitation

For a long time, however, optical microscopy was held back by a physical restriction as to what size of structures are possible to resolve. In 1873, the microscopist Ernst Abbe published an equation demonstrating how microscope resolution is limited by, among other things, the wavelength of the light. For the greater part of the 20th century this led scientists to believe that, in optical microscopes, they would never be able to observe things smaller than roughly half the wavelength of light, i.e., 0.2 micrometres (figure 1). The contours of some of the cells’ organelles, such as the powerhouse mitochondria, were visible. But it was impossible to discern smaller objects and, for instance, to follow the interaction between individual protein molecules in the cell. It is somewhat akin to being able to see the buildings of a city without being able to discern how citizens live and go about their lives. In order to fully understand how a cell functions, you need to be able to track the work of individual molecules.

Abbe’s equation still holds but has been bypassed just the same. Eric Betzig, Stefan W. Hell and William E. Moerner are awarded the Nobel Prize in Chemistry 2014 for having taken optical microscopy into a new dimension using fluorescent molecules. Theoretically there is no longer any structure too small to be studied. As a result, microscopy has become nanoscopy.

[Abbe’s Diffraction Limit (0.2 µm)]

[picture from nobelprize.org]
Looking inside the cell: conventional microscopy

microtubule
Nobel Prize in Chemistry 2014

Eric Betzig  Stefan W. Hell  W.E. Moerner

Invention of single-molecule microscopy
Looking inside the cell

conventional microscopy

single-molecule microscopy
Single molecule microscopy
(basics)
Controlled photoactivation

Green fluorescent protein (GFP)

Energy states [Dickson et al. '97]
Green fluorescent protein (GFP)  

State $A$ is excited to $A^*$ and returns to $A$ upon photon emission
Controlled photoactivation

Green fluorescent protein (GFP)

- State $A$ is excited to $A^*$ and returns to $A$ upon photon emission.
- When $I$ is reached from $A$ there is no fluorescence until $I$ spontaneously moves to $A$ (blinking).

Energy states [Dickson et al. ’97]

- When Betzig returned to academic science after his post-field exile in private industry, he learnt about Lippincott-Schwartz’ mutant and realized that it could possibly solve the problem of finding an optimal way to combine sparse sets of fluorophores with distinct spectral properties to a dense total set of fluorophores. The simple solution would be to activate a very small and thus sparse, random subset of GFP mutant molecules in a biological structure by low-level irradiation at 413 nm. Subsequent irradiation at 488 nm would then be used to determine the positions of the members of the sparse subset at super-resolution, according to Eq. 5 above.

When the first subset had been irreversibly inactivated by bleaching, a second small subset could be activated and the positions of its members determined at high resolution, and so on until all subsets had been sampled and used to determine the structure under authentic super-resolution conditions.

This fulfilled both the condition of only a sparse subset being observed at a time, and the condition of high-frequency (dense) spatial sampling in order to fulfill the Nyqvist and Shannon theorems, as illustrated in Fig. 8.
Controlled photoactivation

Green fluorescent protein (GFP)  Energy states [Dickson et.al. ’97]

- State $A$ is excited to $A^*$ and returns to $A$ upon photon emission
- When $I$ is reached from $A$ there is no fluorescence until $I$ spontaneously moves to $A$ (blinking)
- When $I$ moves to $N$ there is no fluorescence until $N$ is activated by 405nm light and GFP returns to $A$
Photoactivated localization microscopy (PALM) Setup

[picture from ZEISS]
PALM Process

Step 1

Step 2

Step 3. Algorithm needed.

Step 4
Antibodies: attach fluorescent molecules to the structure

All off
Antibodies: attach fluorescent molecules to the structure

All off

All on

Detector

Cannot resolve the structure!
Antibodies: attach fluorescent molecules to the structure

All off

All on

Detector

Cannot resolve the structure!
“Blinking” molecules: sparsity

Frame 1
“Blinking” molecules: sparsity

Frame 1
“Blinking” molecules: sparsity

Frame 1

Locate centers of “Gaussian” blobs (parametric estimation)
“Blinking” molecules: sparsity

Frame 1

Locate centers of “Gaussian” blobs (parametric estimation)
“Blinking” molecules: sparsity

- Locate centers of “Gaussian” blobs (parametric estimation)
- Combine ~ 10000 frames.
“Blinking” molecules: sparsity

Locate centers of “Gaussian” blobs (parametric estimation)

Combine $\sim 10000$ frames.

The structure is now resolved!
Next Frontier: image dynamical processes

Imaging $\sim$ 10000 frames is slow
Next Frontier: image dynamical processes

- Imaging \( \sim 10000 \) frames is **slow**

- Can we make data acquisition **faster**?
Next Frontier: image dynamical processes

Imaging $\sim 10000$ frames is **slow**

Can we make data acquisition **faster**?

Image $\sim 2500$ frames with 4 times more molecules per frame?

- parametric estimation works
- 4 times more active molecules $\Rightarrow$ parametric estimation **does not** work
Next Frontier: image dynamical processes

Imaging $\sim 10000$ frames is **slow**

Can we make data acquisition **faster**?

Image $\sim 2500$ frames with 4 times more molecules per frame?

- parametric estimation works
- 4 times more active molecules \implies parametric estimation **does not** work

Need powerful super-resolution algorithm!
Theory
Theory

Which algorithm?
Performance guarantees?
Fundamental limits?
Mathematical model (discrete 1D setup for simplicity)

$$x(t) = \sum_{i} x_i \delta(t - t_i), \ x_i \geq 0$$

Detector

$$s(t) = \int f_{\text{low}}(t - t')x(t')dt'$$

$$f_{\text{low}}(t) = \frac{1}{2f_c} \left( \frac{\sin(2\pi f_c t)}{\pi t} \right)^2$$

$$\lambda_c = 1/f_c$$
Mathematical model (discrete 1D setup for simplicity)

Object

\[ x(t) = \sum_{i} x_i \delta(t - t_i), \quad x_i \geq 0 \]

Detector

\[ s(t) = \int f_{\text{low}}(t - t') x(t') dt' \]

\[ s = Px + z \]

\[ P = P_{\text{tri}} \text{ is circulant} \]

Triangular spectrum
Mathematical model (discrete 1D setup for simplicity)

\[ x(t) = \sum_{i} x_i \delta(t - t_i), \quad x_i \geq 0 \]

\[ s(t) = \int f_{\text{low}}(t - t') x(t') dt' \]

\[ x = [x_0 \cdots x_{N-1}]^T \geq 0 \]

\[ s = Px + z \]

\[ P = P_{\text{flat}} \text{ is circulant} \]

Flat spectrum
Super-resolution factor and stability

\[ x = [x_0 \cdots x_{N-1}]^T \]

Triangular spectrum

Flat spectrum

\[ s = Px + z \]

\[ \text{SRF} \triangleq \frac{N}{2f_c} \]
Super-resolution factor and stability

\[ \mathbf{x} = [x_0 \cdots x_{N-1}]^T \]

Triangular spectrum

Flat spectrum

\[ \mathbf{s} = \mathbf{P} \mathbf{x} + \mathbf{z} \]

\[ \text{SRF} \triangleq N/(2f_c) \]

Stability: \[ \| \mathbf{x} - \hat{\mathbf{x}} \| \overset{?}{\leq} \| \mathbf{z} \| \cdot (\text{amplification factor}) \]
Classical resolution criteria: separation is about $\lambda_c$
Rayleigh-regularity: $x \in \mathcal{R}(d, r)$

$x$ has fewer than $r$ spikes in every $\lambda_c d$ interval $[\lambda_c \triangleq 1/f_c]$
Rayleigh-regularity: \( x \in \mathcal{R}(d, r) \)

\( x \) has fewer than \( r \) spikes in every \( \lambda_c d \) interval \([\lambda_c \triangleq 1/f_c]\)

Separation: \( \mathcal{R}(2, 1) \)
Rayleigh-regularity: \( \mathbf{x} \in \mathcal{R}(d, r) \)

\( \mathbf{x} \) has fewer than \( r \) spikes in every \( \lambda_c d \) interval \( [\lambda_c \triangleq 1/f_c] \)

Separation: \( \mathcal{R}(2, 1) \)

\[ \geq 2\lambda_c \]

\( \mathcal{R}(4, 2) \)

\[ \lambda_c \quad \geq 4\lambda_c \]
Rayleigh-regularity: \( x \in \mathcal{R}(d, r) \)

\( x \) has fewer than \( r \) spikes in every \( \lambda_c d \) interval \([\lambda_c \triangleq 1/f_c]\)

**Separation:** \( \mathcal{R}(2, 1) \)

\[
\lambda_c \geq 2\lambda_c
\]

\( \mathcal{R}(4, 2) \)

\[
\lambda_c \geq 4\lambda_c
\]

\( \mathcal{R}(6, 3) \)

\[
\lambda_c \geq 6\lambda_c
\]
Key contribution

[ Prony’1795 ]
\[ x \in \mathbb{C}^N \]
no stability
–
efficient
### Key contribution

<table>
<thead>
<tr>
<th></th>
<th>[Prony’1795]</th>
<th>MUSIC, ESPRIT</th>
</tr>
</thead>
</table>
| $x \in \mathbb{C}^N$ | no stability

|               | efficient
|---------------|-----------------------------------------------------------------------------|
|               | stability not understood

- Rayleigh-regularity

- separation

- convex

|               | efficient
|---------------|-----------------------------------------------------------------------------|
|               | efficient

$x \geq 0$

$x \in \mathbb{C}^N$
### Key contribution

<table>
<thead>
<tr>
<th>[Prony’1795]</th>
<th>MUSIC, ESPRIT</th>
<th>[Donoho’92]</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x \in \mathbb{C}^N$</td>
<td>$x \in \mathbb{C}^N$</td>
<td>$x \in \mathbb{C}^N$</td>
</tr>
<tr>
<td>no stability</td>
<td>stability not understood</td>
<td>stability</td>
</tr>
<tr>
<td>efficient</td>
<td>efficient</td>
<td>Rayleigh-regularity</td>
</tr>
<tr>
<td></td>
<td></td>
<td>combinatorial</td>
</tr>
</tbody>
</table>

$\mathcal{R}(2r, r)$
Key contribution

**[Prony’1795]**

\[ x \in \mathbb{C}^N \]

- no stability
- efficient

**MUSIC, ESPRIT**

\[ x \in \mathbb{C}^N \]

- stability not understood
- efficient

**[Donoho’92]**

\[ x \in \mathbb{C}^N \]

- stability
- Rayleigh-regularity
- combinatorial

**[Donoho et al.’90]**

\[ x \geq 0 \]

- no stability
- convex

\[ x \geq 0 \]
Key contribution

**[Prony’1795]**
\[ x \in \mathbb{C}^N \]
no stability
- efficient

**MUSIC, ESPRIT**
\[ x \in \mathbb{C}^N \]
stability not understood
- efficient

**[Donoho’92]**
\[ x \in \mathbb{C}^N \]
stability
Rayleigh-regularity
combinatorial

**[Donoho et al.’90]**
\[ x \geq 0 \]
no stability
- convex

**[Candès & F.-Granda’12]**
\[ x \in \mathbb{C}^N \]
stability
separation
convex

Works:

\[ \mathcal{R}(2, 1) \]
\[ \geq 2\lambda_c \]
### Key contribution

<table>
<thead>
<tr>
<th>[Prony’1795]</th>
<th>MUSIC, ESPRIT</th>
<th>[Donoho’92]</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mathbf{x} \in \mathbb{C}^N$</td>
<td>$\mathbf{x} \in \mathbb{C}^N$</td>
<td>$\mathbf{x} \in \mathbb{C}^N$</td>
</tr>
<tr>
<td>no stability</td>
<td>stability not understood</td>
<td>stability</td>
</tr>
<tr>
<td>efficient</td>
<td>efficient</td>
<td>Rayleigh-regularity</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>[Donoho et al.’90]</th>
<th>[Candès &amp; F.-Granda’12]</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mathbf{x} \geq 0$</td>
<td>$\mathbf{x} \in \mathbb{C}^N$</td>
</tr>
<tr>
<td>no stability</td>
<td>stability</td>
</tr>
<tr>
<td>–</td>
<td>separation</td>
</tr>
<tr>
<td>efficient</td>
<td>convex</td>
</tr>
</tbody>
</table>

**Breaks:**

$$\geq 4\lambda_c \quad \mathcal{R}(4, 2)$$
Key contribution

- **[Prony’1795]**
  \[ x \in \mathbb{C}^N \]
  no stability
  -
  efficient

- **[Donoho’92]**
  \[ x \in \mathbb{C}^N \]
  stability
  Rayleigh-regularity
  combinatorial

- **[Donoho et al.’90]**
  \[ x \geq 0 \]
  no stability
  -
  convex

- **[Candès & F.-Granda’12]**
  \[ x \in \mathbb{C}^N \]
  stability
  separation
  convex

- **This work**
  \[ x \geq 0 \]
  stability
  Rayleigh-regularity
  convex

\[ \mathcal{R}(2r, r), x \geq 0 \]
Main results

Recall:

\[ s = Px + z \]

Solve:

\[ \begin{align*}
\text{minimize} & \quad \| s - P\hat{x} \|_1 \\
\text{subject to} & \quad \hat{x} \geq 0
\end{align*} \]

Theorem: [V. Morgenshtern and E. Candès, 2014]

Take \( P = P_{\text{tri}} \) or \( P = P_{\text{flat}} \). Assume \( x \geq 0 \), \( x \in \mathcal{R}(2r, r) \). Then,

\[ \| \hat{x} - x \|_1 \leq c \cdot \| z \|_1 \cdot \left( \frac{N}{2f_c} \right)^{2r} \cdot \left( 2f_c \right)^{-1} \]
Main results

Recall:

\[ s = Px + z \]

Solve:

\[
\begin{align*}
\text{minimize} & \quad \|s - \hat{P}\hat{x}\|_1 \\
\text{subject to} & \quad \hat{x} \geq 0
\end{align*}
\]

Theorem: [V. Morgenshtern and E. Candès, 2014]

Take \( P = P_{\text{tri}} \) or \( P = P_{\text{flat}} \). Assume \( x \geq 0, x \in \mathcal{R}(2r, r) \). Then,

\[
\|\hat{x} - x\|_1 \leq c \cdot \|z\|_1 \cdot \left( \frac{N}{2f_c} \right)^{2r}. 
\]

Converse: [V. Morgenshtern and E. Candès, 2014]

For \( P = P_{\text{tri}} \), no algorithm can do better than \( c \cdot \|z\|_1 \cdot \left( \frac{N}{2f_c} \right)^{2r-1} \).
Key ideas

→ **Duality theory:** to prove stability we need a low-frequency trigonometric polynomial that is “curvy”
  - [Dohono, et al.’92] construct trigonometric polynomial that is not “curvy”
  - [Candès and Fernandez-Granda’12] construct trigonometric polynomial that is “curvy”, but construction needs *separation*
  - **New construction:** multiply “curvy” trigonometric polynomials
    - “curvy”
    - construction needs *no separation*
Dual certificate (noiseless case, $z = 0$)

- $\mathcal{T}$ is the support of $x$
Dual certificate (noiseless case, $z = 0$)

- $\mathcal{T}$ is the support of $x$
- Suppose, we can construct a **low-frequency trig. polynomial**:

$$q(t) = \sum_{k=-f_c}^{f_c} \hat{q}_k e^{-i2\pi kt}, \quad 0 \leq q(t) \leq 1, \quad q(t_i) = 0 \text{ for all } t_i \in \mathcal{T}.$$
Dual certificate (noiseless case, \( z = 0 \))

- \( \mathcal{T} \) is the support of \( x \)

- Suppose, we can construct a **low-frequency trig. polynomial**:

\[
q(t) = \sum_{k=-f_c}^{f_c} \hat{q}_k e^{-i2\pi kt}, \quad 0 \leq q(t) \leq 1, \quad q(t_i) = 0 \text{ for all } t_i \in \mathcal{T}.
\]

- Then, \( \hat{x} = x \).
Connection to LASSO (x can be negative here)

\[
\text{minimize} \quad \|\hat{x}\|_1 \quad \text{subject to} \quad s = P\hat{x}
\]
Connection to LASSO (x can be negative here)

\[
\begin{align*}
\text{minimize} & \quad \|\hat{x}\|_1 & \text{subject to} & \quad s = P\hat{x}
\end{align*}
\]
Connection to LASSO (x can be negative here)

\[
\text{minimize} \quad \|\hat{x}\|_1 \quad \text{subject to} \quad s = P\hat{x}
\]

\[\hat{x} = x \text{ iff there exists } q \perp \text{null}(P) \text{ and } q \in \partial\|x\|_1\]
Connection to LASSO (x can be negative here)

\[
\text{minimize } \|\hat{x}\|_1 \quad \text{subject to} \quad s = P\hat{x}
\]

\[\hat{x} = x \text{ iff there exists } q \perp \text{null}(P) \text{ and } q \in \partial \|x\|_1\]

\[P \text{ is orthogonal projection onto the set of low-freq. trig. polynomials: } \]
\[q \perp \text{null}(P) \iff \]
\[q(t) = \sum_{k=-f_c}^{f_c} \hat{q}_k e^{-i2\pi kt}\]
Connection to LASSO (x can be negative here)

\[
\begin{align*}
\text{minimize} & \quad \| \hat{x} \|_1 \quad \text{subject to} \quad s = P \hat{x} \\
\hat{x} & = x \text{ iff there exists } q \perp \text{null}(P) \text{ and } q \in \partial \|x\|_1 \\
P & \text{is orthogonal projection onto the set of low-freq. trig. polynomials:} \\
q \perp \text{null}(P) & \iff q(t) = \sum_{k=-f_c}^{f_c} \hat{q}_k e^{-i2\pi kt} \\
q \in \partial \|x\|_1 & \iff \begin{cases} 
q(t_i) = \text{sign}(x_i) & x_i \neq 0 \\
|q(t_i)| \leq 1 & x_i = 0
\end{cases}
\end{align*}
\]
Dual certificate (noisy case)

- $\mathcal{T}$ is the support of $x$

- Suppose, we can construct a **low-frequency trig. polynomial**:

$$q(t) = \sum_{k=-f_c}^{f_c} \hat{q}_k e^{-i2\pi kt}, \quad 0 \leq q(t) \leq 1, \quad q(t_i) = 0 \text{ for all } t_i \in \mathcal{T}.$$
\[ q(t) = \sum_{k=-f_c}^{f_c} \hat{q}_k e^{-i2\pi kt}, \quad 0 \leq q(t) \leq 1, \quad q(t_i) = 0 \text{ for all } t_i \in \mathcal{T}. \]

Then, \[ \|\hat{x} - x\|_1 \leq 4\|z\|_1 / \rho. \]
Key ideas

- **Duality theory**: to prove stability we need a low-frequency trigonometric polynomial that is “curvy”
  → [Dohono, et al.’92] construct trigonometric polynomial that is not “curvy”

- [Candès and Fernandez-Granda’12] construct trigonometric polynomial that is “curvy”, but construction needs separation

- **New construction**: multiply “curvy” trigonometric polynomials
  - “curvy”
  - construction needs no separation
[Dohono, et al.'92]: “Classical” $q(t)$

$$q(t) = \prod_{t_0 \in \mathcal{T}} \frac{1}{2} \left[ \cos(2\pi(t + 1/2 - t_0)) + 1 \right].$$

No separation required

Low curvature!

$$q(t - t_0) \approx (t - t_0)^2 \Rightarrow \|x - \hat{x}\|_1 \leq \|z\|_1 \cdot N^2.$$
Key ideas

- **Duality theory**: to prove stability we need a low-frequency trigonometric polynomial that is “curvy”

- [Dohono, et al.’92] construct trigonometric polynomial that is not “curvy”

→ [Candès and Fernandez-Granda’12] construct trigonometric polynomial that is “curvy”, but construction needs separation

- **New construction**: multiply “curvy” trigonometric polynomials
  - “curvy”
  - construction needs no separation
[Candès, Fernandez-Granda’12]: “Curvy” $q(t)$

$$q(t) = \sum_{t_j \in \mathcal{T}} a_j K(t - t_j) + \text{corrections},$$

$$K(t) \ldots \text{low-frequency and “curvy”}$$

Separation between zeros required: $\mathcal{T} \in \mathcal{R}(2, 1)$

High curvature!

$$q(t - t_i) \approx f_c^2 (t - t_i)^2 \Rightarrow \|x - \hat{x}\|_1 \leq c \cdot \|z\|_1 \cdot \left(\frac{N}{2f_c}\right)^2$$

$\rho$ —

$\geq 2\lambda_c$
Comparison of Trigonometric Polynomials

\[
\begin{align*}
\text{“classical”} & \quad q(t) \approx t^2 \\
\text{“curvy”} & \quad q(t) \approx f_c^2 t^2 \\
(0, 1) &
\end{align*}
\]
Key ideas

- **Duality theory:** to prove stability we need a low-frequency trigonometric polynomial that is “curvy”
- [Dohono, et al.’92] construct trigonometric polynomial that is not “curvy”
- [Candès and Fernandez-Granda’12] construct trigonometric polynomial that is “curvy”, but construction needs separation
  → **New construction:** multiply “curvy” trigonometric polynomials
    - “curvy”
    - construction needs no separation
New construction: curvature without separation

Partition support: $\mathcal{T} = \mathcal{T}_1 \cup \mathcal{T}_2, \quad r = 2$

Regularity: $\mathcal{T} \in \mathcal{R}(2 \cdot 2, 2) \Rightarrow \mathcal{T}_i \in \mathcal{R}(4, 1)$

$q(t; f_c) = q_1(t; f_c/2) \times q_2(t; f_c/2)$
New construction: curvature without separation

Partition support: $\mathcal{T} = \mathcal{T}_1 \cup \mathcal{T}_2, \quad r = 2$

Regularity: $\mathcal{T} \in \mathcal{R}(2 \cdot 2, 2) \Rightarrow \mathcal{T}_i \in \mathcal{R}(4, 1)$

$$q(t; f_c) = q_1(t; f_c / 2) \times q_2(t; f_c / 2)$$

High curvature!

$$q(t - t_i) \approx \frac{f_c^{2r}}{r^{2r}} (t - t_i)^{2r} \Rightarrow \|x - \hat{x}\|_1 \leq c \cdot \|z\|_1 \cdot \left(\frac{N}{2f_c}\right)^{2r}$$
Summation vs. multiplication

Remember: $q(t)$ must be frequency-limited to $f_c$!
Summation vs. multiplication

Remember: $q(t)$ must be frequency-limited to $f_c$!

[Donoho, et.al.]:

$$q(t) = \prod_{t_j \in T} \frac{1}{2} \left[ \cos(2\pi(t + 1/2 - t_j)) + 1 \right]$$

This work:

$$q(t) = r \prod_{k=1}^{32/50} \sum_{t_{jk} \in T_k} a_{jk} K(t - t_{jk})$$

frequency one
Summation vs. multiplication

Remember: $q(t)$ must be frequency-limited to $f_c$!

[Donoho, et.al.]:

$$q(t) = \prod_{t_j \in \mathcal{T}} \frac{1}{2} \left[ \cos(2\pi(t + 1/2 - t_j)) + 1 \right]$$

[Candès, Fernandez-Granda]:

$$q(t) = \sum_{t_j \in \mathcal{T}} a_j K(t - t_j)$$
**Summation vs. multiplication**

Remember: $q(t)$ must be frequency-limited to $f_c$!

**[Donoho, et.al.]:**

$$q(t) = \prod_{t_j \in \mathcal{T}} \frac{1}{2} \left[ \cos(2\pi(t + 1/2 - t_j)) + 1 \right]$$

**[Candès, Fernandez-Granda]:**

$$q(t) = \sum_{t_j \in \mathcal{T}} a_j K(t - t_j)$$

**This work:**

$$q(t) = \prod_{k=1}^{r} \sum_{t_{jk} \in \mathcal{T}_k} a_{jk} K(t - t_{jk})$$
Complex vs. positive signals

Why do we need $x \geq 0$?

$x \geq 0$

Interpolate zero on supp. of $x$

$x \in \mathbb{C}^N$

Interpolate $\text{sign}(x)$ on supp. of $x$

$\rho$

$0$

$t_1$

$t_2$

$t_3$

$t_4$

$\geq 2\lambda_c$

$1/N$

$1$

$-1$

$\rho$

Does not exist! (Bernstein Th.)
Continuous setup
$f_c$ fixed, $N \to \infty \Rightarrow \text{SRF}_{\text{OLD}} \to \infty$

Is the problem hopeless?

$x(t)$ and $\hat{x}(t)$
$f_c$ fixed, $N \to \infty \Rightarrow SRF_{OLD} \to \infty$

Is the problem hopeless?

No: we need to be less ambitions!
$f_c$ fixed, $N \rightarrow \infty \Rightarrow \text{SRF}_{\text{OLD}} \rightarrow \infty$

$x(t)$ and $\hat{x}(t)$

Is the problem hopeless?

No: we need to be less ambitions!

$s(t) = (f_{\text{low}} \ast x)(t)$

$\hat{x}(t) = (f_{\text{hi}} \ast x)(t)$
$f_c$ fixed, $N \to \infty \Rightarrow SRF_{OLD} \to \infty$

Is the problem hopeless?

No: we need to be less ambitions!

$s(t) = (f_{low} \star x)(t)$

Error$=\|f_{hi} \star (x - \hat{x})\|_1$

$SRF_{NEW} = \lambda_c / \lambda_{hi}$
Theorem: [V. Morgenshtern and E. Candès, 2014]

Assume $x(t) \geq 0, x(t) \in \mathcal{R}(2r, r)$. Then,

$$
\| f_{hi} \ast (x - \hat{x}) \|_1 \leq c \cdot \left( \frac{\lambda_c}{\lambda_{hi}} \right)^{2r} \cdot \| z(t) \|_1.
$$
Need new tools

**Theorem: [V. Morgenshtern and E. Candès, 2014]**

Assume $x(t) \geq 0$, $x(t) \in \mathcal{R}(2r, r)$. Then,

$$\|f_{hi} \ast (x - \hat{x})\|_1 \leq c \cdot \left(\frac{\lambda_c}{\lambda_{hi}}\right)^{2r} \cdot \|z(t)\|_1.$$ 

**Can do:** all zeros

**Need:** arbitrary pattern $\{0, +\rho\}$
2D Super-resolution

Theorem: [V. Morgenshtern and E. Candès, 2014]

Take $\mathbf{P} = \mathbf{P}_{\text{tri,2D}}$ or $\mathbf{P} = \mathbf{P}_{\text{flat,2D}}$. Assume $\mathbf{x} \geq 0$, $\mathbf{x} \in \mathcal{R}(2.38r, r)$. Then,

$$\|\hat{\mathbf{x}} - \mathbf{x}\|_1 \leq c \cdot \left( \frac{N}{2f_c} \right)^{2r} \delta.$$ 

New: number of spikes is linear in the number of observations
Improving microscopes

Collaboration with Moerner Lab, C.A. Sing-Long, E. Candès
Reconstruction of 3D signals from 2D data

Double-helix PSF

Normal PSF

picture from [Pavani and Piston'08]

2D double-helix data
Reconstruction of 3D signals from 2D data

Double-helix PSF

Normal PSF

picture from [Pavani and Piston'08]

2D double-helix data

\[
\begin{align*}
\text{minimize} & \quad \frac{1}{2} \| s - P\hat{x} \|_2^2 + \lambda \sigma \| \text{diag}(w)\hat{x} \|_1 \\
\text{subject to} & \quad \hat{x} \geq 0
\end{align*}
\]
Preliminary result: 4 times faster than state-of-the-art

10000 CVX problems solved
TFOCS first order solver
millions of variables
minimize \[ \frac{1}{2} \| s - P(\hat{x} + b) \|_2^2 + \lambda \sigma \| \hat{x} \|_1 \]
subject to \[ \hat{x} \geq 0 \]
\[ b \text{ low freq. trig. polynomial (background)} \]
Comparison of super-resolution algorithms

Work in progress:

- Need to carefully compare super-resolution algorithms in practice
  - Naive matched-filters
  - Algebraic methods: MUSIC, ESPRIT, ...
  - Convex-optimization algorithms with different regularizers

- Realistic physical model
  - Noise: quantum noise and out-of-focus background
  - Point-spread function (\( P \)) uncertainty and variation
  - Rotation of single molecules
  - ...

- Test images (phantoms):
  - Different densities of sources
  - Different spacial distributions
  - **Correct answer should always be known**

- Create a database of test cases for quick algorithm assessment
Conclusion

Convex optimization is a near-optimal method for super-resolution of positive sources

- Flexibility and good practical performance
- Non-asymptotic precise stability bounds
- Rayleigh-regularity is fundamental: separation between spikes is only one part of the picture
Lots of questions remain

- What is the best regularizer in the presence of stochastic noise?
- Fast parallel solver exploiting the structure of the problem
- Theory for Double-Helix reconstruction: 3D signal from 2D observations
- Tractable near-optimal algorithm for complex-valued signals?
Lots of questions remain

- What is the best regularizer in the presence of stochastic noise?
- Fast parallel solver exploiting the structure of the problem
- Theory for Double-Helix reconstruction: 3D signal from 2D observations
- Tractable near-optimal algorithm for complex-valued signals?
- **Sparse regression where the design matrix has highly correlated columns:**
  - \( \mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_N] \)
  - Shift-invariance: \( \langle \mathbf{a}_k, \mathbf{a}_l \rangle = \langle \mathbf{a}_{k+r}, \mathbf{a}_{l+r} \rangle \)
  - \( \langle \mathbf{a}_k, \mathbf{a}_{k+r} \rangle \) is large for small \( r \)
  - \( \langle \mathbf{a}_k, \mathbf{a}_{k+r} \rangle \) decays quickly with \( r \)
  - Minimum separation likely needed in general
  - If all elements in \( \mathbf{A} \) are nonnegative and the signal is nonnegative, regularity might be enough
Acknowledgements

- Swiss National Science Foundation Fellowship
- Simons Foundation
- Theory: collaboration with E. Candès
- Motivation and applications: collaboration with W.E. Moerner, C.A. Sing-Long, M.D. Lew, A. Backer, S.J. Sahl
- Helpful discussions and related work: C. Fernandez-Granda, M. Soltanolkotabi, R. Heckel
Thank you
Backup slides
Proof of Lemma

Set: \( h = \hat{x} - x, \quad \mathcal{T} = \{l/N : h_l < 0\} \)
Proof of Lemma

Set: \( h = \hat{x} - x \), \( \mathcal{T} = \{ l/N : h_l < 0 \} \subset \text{supp}(x) \).
Proof of Lemma

- Set: \( h = \hat{x} - x \), \( \mathcal{T} = \{ l/N : h_l < 0 \} \subset \text{supp}(x) \).
- Dual vector \( q_l = q(l/N) \) satisfies:

\[
P_{\text{flat}} q = q, \quad \|q\|_\infty = 1, \quad \text{and} \quad \begin{cases} q_l = 0, & l/N \in \mathcal{T} \\ q_l > \rho, & \text{otherwise} \end{cases}
\]
Proof of Lemma

- Set: \( h = \hat{x} - x, \quad \mathcal{T} = \{l/N : h_l < 0\} \subset \text{supp}(x) \).
- Dual vector \( q_l = q(l/N) \) satisfies:

\[
P_{\text{flat}} q = q, \quad \|q\|_\infty = 1, \quad \text{and} \quad \begin{cases} q_l = 0, & l/N \in \mathcal{T} \\ q_l > \rho, & \text{otherwise} \end{cases}
\]

- On the one hand:

\[
|\langle q - \rho/2, h \rangle| = |\langle P(q - \rho/2), h \rangle| = |\langle q - \rho/2, Ph \rangle| \\
\leq \|q - \rho/2\|_\infty \|Ph\|_1 \leq \|Px - s + s - P\hat{x}\|_1 \\
\leq \|Px - s\|_1 + \|s - P\hat{x}\|_1 \\
\leq 2\|Px - s\|_1 \leq 2\|z\|_1.
\]
Proof of Lemma

Set: \( h = \hat{x} - x \), \( \mathcal{T} = \{ l/N : h_l < 0 \} \subset \text{supp}(x) \).

Dual vector \( q_l = q(l/N) \) satisfies:

\[
P_{\text{flat}}q = q, \quad \|q\|_\infty = 1, \quad \text{and} \quad \begin{cases} q_l = 0, & l/N \in \mathcal{T} \\ q_l > \rho, & \text{otherwise.} \end{cases}
\]

On the one hand:

\[
\left| \langle q - \rho/2, h \rangle \right| = \left| \langle P(q - \rho/2), h \rangle \right| = \left| \langle q - \rho/2, Ph \rangle \right| \\
\leq \|q - \rho/2\|_\infty \|Ph\|_1 \leq \|Px - s + s - P\hat{x}\|_1 \\
\leq \|Px - s\|_1 + \|s - P\hat{x}\|_1 \\
\leq 2\|Px - s\|_1 \leq 2\|z\|_1.
\]

On the other hand:

\[
\left| \langle q - \rho/2, h \rangle \right| = \left| \sum_{l=0}^{N-1} (q_l - \rho/2)h_l \right| = \sum_{l=0}^{N-1} (q_l - \rho/2)h_l \geq \rho\|h\|_1/2.
\]
Proof of Lemma

- Set: \( h = \hat{x} - x, \quad \mathcal{T} = \{ l/N : h_l < 0 \} \subset \text{supp}(x). \)
- Dual vector \( q_l = q(l/N) \) satisfies:

\[
P_{\text{flat}} q = q, \quad \| q \|_\infty = 1, \quad \text{and} \quad \begin{cases} q_l = 0, & l/N \in \mathcal{T} \\ q_l > \rho, & \text{otherwise.} \end{cases}
\]

- On the one hand:

\[
|\langle q - \rho/2, h \rangle| = |\langle P(q - \rho/2), h \rangle| = |\langle q - \rho/2, Ph \rangle| \\
\leq \| q - \rho/2 \|_\infty \| Ph \|_1 \leq \| Px - s + s - P\hat{x} \|_1 \\
\leq \| Px - s \|_1 + \| s - P\hat{x} \|_1 \\
\leq 2 \|Px - s\|_1 \leq 2 \|z\|_1.
\]

- On the other hand:

\[
|\langle q - \rho/2, h \rangle| = \left| \sum_{l=0}^{N-1} (q_l - \rho/2) h_l \right| = \sum_{l=0}^{N-1} (q_l - \rho/2) h_l \geq \rho \| h \|_1 / 2.
\]

- Combining: \( \| h \|_1 \leq 4 \| z \|_1 / \rho. \)
Consider: $q(t) = \sum_{k=-f_c}^{f_c} \hat{q}_k e^{-i2\pi kt}$ with $\|q\|_\infty \leq 1$

Then: $\|q'\|_\infty \leq 2f_c$
Connection to Bernstein theorem

Consider: $q(t) = \sum_{k=-f_c}^{f_c} \hat{q}_k e^{-i2\pi kt}$ with $\|q\|_{\infty} \leq 1$

Then: $\|q'\|_{\infty} \leq 2f_c$

“Curvy” $q(t)$ has best possible curvature!
Consider: \[ q(t) = \sum_{k=-f_c}^{f_c} \hat{q}_k e^{-i2\pi kt} \] with \( \|q\|_\infty \leq 1 \)

Then: \( \|q'\|_\infty \leq 2f_c \)

"Curvy" \( q(t) \) has best possible curvature!

Since

\[
q(t_i) = 0 \\
q'(t_i) = 0 \\
\|q\|_\infty \leq 1
\]

We conclude:

\[
\|q'\|_\infty \leq 2f_c \Rightarrow \|q''\|_\infty \leq (2f_c)^2 \\
\Rightarrow q(t - t_i) \leq (2f_c)^2(t - t_i)^2 \\
\Rightarrow q(t_i + 1/N) \leq \frac{(2f_c)^2}{N^2} = \frac{1}{SRF^2}
\]
new tools

1. Control behavior on separated set

2. Multiply

\[ q(t) = q_1(t) \times q_2(t) \]

\[ 0 = q'(t_3) = q'_1(t_3)q_2(t_3) + q_1(t_3)q'_2(t_3) \]
New tools

1. Control behavior on separated set
2. Multiply

\[ q(t) = q_1(t) \times q_2(t) \]

\[ 0 = q'(t_3) = q'_1(t_3)q_2(t_3) + q_1(t_3)q'_2(t_3) \]

3. Sum

\[ q(t) = \sum_r \prod_{k=1}^{r} \sum_{t_{jk} \in T_k} a_{jk}K(t - t_{jk}) \]

Frequency \( f_c/r \)