Sparse Recovery Beyond Compressed Sensing

Carlos Fernandez-Granda
www.cims.nyu.edu/~cfgranda

Applied Math Colloquium, MIT

4/30/2018
Acknowledgements

Project funded by NSF award DMS-1616340
Separable Nonlinear Inverse Problems

Sparse Recovery for SNL Problems

Super-resolution

Deconvolution

SNL Problems with Correlation Decay
Separable Nonlinear Inverse Problems

Sparse Recovery for SNL Problems

Super-resolution

Deconvolution

SNL Problems with Correlation Decay
Separable nonlinear inverse (SNL) problems

Aim: estimate parameters $\theta_1, \ldots, \theta_s \in \mathbb{R}^d$ from data $y \in \mathbb{R}^n$

Relation between data and each $\theta_j$ governed by nonlinear function $\phi$

Contributions of $\theta_1, \ldots, \theta_s$ combine linearly with unknown coeffs $c \in \mathbb{R}^s$

$$y = \sum_{j=1}^{s} c(j) \phi(\theta_j) = \begin{bmatrix} \phi(\theta_1) & \phi(\theta_2) & \cdots & \phi(\theta_s) \end{bmatrix} c$$

$n > s$, easy if we know $\theta_1, \ldots, \theta_s$
SNL problems

- Super-resolution
- Deconvolution
- Source localization in EEG
- Direction of arrival in radar / sonar
- Magnetic-resonance fingerprinting
Magnetic-resonance fingerprinting (Ma et al, 2013)

Goal: Estimate magnetic relaxation-time constants of tissues in a voxel
Separable Nonlinear Inverse Problems

Sparse Recovery for SNL Problems

Super-resolution

Deconvolution

SNL Problems with Correlation Decay
Methods to tackle SNL problems

- Nonlinear least-squares solved by descent methods
  *Drawback*: local minima

- Prony-based / Finite-rate of innovation
  *Drawback*: challenging to apply beyond super-resolution

- Reformulate as sparse-recovery problem
  (drawbacks discussed at the end)
Linearization

Linearize problem by lifting to a higher-dimensional space

True parameters: \( \theta_{T_1}, \ldots, \theta_{T_s} \)

Grid of parameters: \( \theta_1, \ldots, \theta_N, \; N \gg n \)

\[
y = \begin{bmatrix} \phi(\theta_1) & \cdots & \phi(\theta_{T_1}) & \cdots & \phi(\theta_{T_s}) & \cdots & \phi(\theta_N) \end{bmatrix} \begin{bmatrix} 0 \\ \vdots \\ c(1) \\ \vdots \\ c(s) \\ 0 \end{bmatrix} = \sum_{j=1}^{s} c(j) \phi(\theta_{T_j})
\]
Sparse Recovery for SNL Problems

Find a sparse $\tilde{c}$ such that

$$y = \Phi_{\text{grid}} \tilde{c}$$

Underdetermined linear inverse problem with sparsity prior
Popular approach: $\ell_1$-norm minimization

\[
\begin{align*}
\text{minimize} & \quad ||\tilde{c}||_1 \\
\text{subject to} & \quad \Phi_{\text{grid}}\tilde{c} = y
\end{align*}
\]
Popular approach: \( \ell_1 \)-norm minimization

- Deconvolution:
  
  *Deconvolution with the \( \ell_1 \) norm*, Taylor et al (1979)

- EEG:
  

- Direction-of-arrival in radar / sonar:
  
  *A sparse signal reconstruction perspective for source localization with sensor arrays*, Malioutov et al (2005)

- and many, many others...
Magnetic-resonance fingerprinting

Multicompartment magnetic resonance fingerprinting
(Tang, F., Lannuzel, Bernstein, Lattanzi, Cloos, Knoll, Asslaender 2018)
Main question

Under what conditions can SNL problems be solved by \( \ell_1 \)-norm minimization?
Continuous dictionary

Analysis should apply to arbitrarily fine grids

We model the coefficients / parameter values as an atomic measure

\[
x := \sum_{j=1}^{s} c(j) \delta_{\theta T_j}
\]

\[
y = \sum_{j=1}^{s} c(j) \phi(\theta T_j)
\]

\[
= \int \phi(\theta) x(d\theta) = \Phi x
\]

Intuitively, \(\Phi\) is a continuous dictionary with \(n\) rows
Find a sparse $\tilde{x}$ such that

$$y = \int \phi(\theta) \tilde{x}(d\theta)$$

(Extremely) underdetermined linear inverse problem with sparsity prior
Total-variation norm

Continuous counterpart of the $\ell_1$ norm

Not the total variation of a piecewise-constant function

$$\|c\|_1 = \sup_{\|\vec{v}\|_\infty \leq 1} \langle v, c \rangle$$

$$\|x\|_{TV} = \sup_{f \in C[0,1], \|f\|_\infty \leq 1} \int_{[0,1]} f(t)x(t)\,dt$$

If $x = \sum_j c_j \delta_{\theta_j}$ then $\|x\|_{TV} = \|c\|_1$
Main question

For an SNL problem, when does

\[
\begin{align*}
\text{minimize} & \quad \|\tilde{x}\|_{TV} \\
\text{subject to} & \quad \int \phi(\theta) \tilde{x}(d\theta) = y
\end{align*}
\]

achieve exact recovery?

*Wait, isn’t this just compressed sensing?*
Compressed sensing

Recover $s$-sparse vector $x$ of dimension $m$ from $n < m$ measurements $y = Ax$

Key assumption: $A$ is random, and hence satisfies restricted-isometry properties with high probability
Restricted isometry property (Candès, Tao 2006)

An \( m \times n \) matrix \( A \) satisfies the restricted isometry property (RIP) if there exists \( 0 < \kappa < 1 \) such that for any \( s \)-sparse vector \( x \)

\[
(1 - \kappa) \| x \|_2 \leq \| Ax \|_2 \leq (1 + \kappa) \| x \|_2
\]

2s-RIP implies that for any \( s \)-sparse signals \( x_1, x_2 \)

\[
\| Ax_2 - Ax_1 \|_2 \geq (1 - \kappa) \| x_2 - x_1 \|_2
\]
Restricted isometry property (Candès, Tao 2006)

An $m \times n$ matrix $A$ satisfies the restricted isometry property (RIP) if there exists $0 < \kappa < 1$ such that for any $s$-sparse vector $x$

$$(1 - \kappa) \|x\|_2 \leq \|Ax\|_2 \leq (1 + \kappa) \|x\|_2$$

$2s$-RIP implies that for any $s$-sparse signals $x_1, x_2$

$$\|Ax_2 - Ax_1\|_2 = \|A(x_2 - x_1)\|_2$$
Restricted isometry property (Candès, Tao 2006)

An $m \times n$ matrix $A$ satisfies the restricted isometry property (RIP) if there exists $0 < \kappa < 1$ such that for any $s$-sparse vector $x$

$$(1 - \kappa) \|x\|_2 \leq \|Ax\|_2 \leq (1 + \kappa) \|x\|_2$$

2s-RIP implies that for any $s$-sparse signals $x_1, x_2$

$$\|Ax_2 - Ax_1\|_2 = \|A(x_2 - x_1)\|_2 \geq (1 - \kappa) \|x_2 - x_1\|_2$$
Separable nonlinear problems

If $\phi$ is smooth, nearby columns in

$$\Phi_{\text{grid}} := [\phi(\theta_1) \ \phi(\theta_2) \ \cdots \ \phi(\theta_N)]$$

are highly correlated so RIP does not hold!

There are $x_1, x_2$ such that $Ax_1 \approx Ax_2$

Sparsity is not enough, we need additional restrictions!
Separable Nonlinear Inverse Problems

Sparse Recovery for SNL Problems

Super-resolution

Deconvolution

SNL Problems with Correlation Decay
Super-resolution

Joint work with Emmanuel Candès (Stanford)
Limits of resolution in imaging

The resolving power of lenses, however perfect, is limited (Lord Rayleigh)

\[ \delta(t - \tau) \quad \text{optical system} \quad h(t - \tau) \]

Diffraction imposes a fundamental limit on the resolution of optical systems
Fluorescence microscopy

Data

Point sources

Low-pass blur

(Figures courtesy of V. Morgenshtern)
Sensing model for super-resolution

Point sources \( \times \) Point-spread function = Data

Spectrum \( \times \) =
Super-resolution

$s$ sources with locations $\theta_1, \ldots, \theta_s$, modeled as superposition of spikes

$$x = \sum_j c(j) \delta_{\theta_j} \quad c_j \in \mathbb{C}, \theta_j \in T \subset [0, 1]$$

We observe Fourier coefficients up to cut-off frequency $f_c$

$$y(k) = \int_0^1 \exp(-i2\pi kt) x(\text{d}t)$$

$$= \sum_{j=1}^s c(j) \exp(-i2\pi k\theta_j)$$

SNL problem where

$$\phi(\theta_j) = \begin{bmatrix} \exp(-i2\pi \theta_j(-f_c)) \\ \ldots \\ \exp(-i2\pi \theta_j f_c) \end{bmatrix}$$
Fundamental questions

1. Is the problem well posed?

2. Does $TV$-norm minimization work?
Is the problem well posed?

Measurement operator = low-pass samples with cut-off frequency $f_c$
Is the problem well posed?

Effect of measurement operator on sparse vectors?
Is the problem well posed?

Submatrix can be very ill conditioned!
Is the problem well posed?

If support is spread out there is hope
Minimum separation

The minimum separation $\Delta$ of the support of $x$ is

$$\Delta = \inf_{(\theta, \theta') \in \text{support}(x): \theta \neq \theta'} |\theta - \theta'|$$
Conditioning of submatrix with respect to $\Delta$

- If $\Delta < 1/f_c$ the problem is **ill posed**
- If $\Delta > 1/f_c$ the problem becomes **well posed**
- Proved asymptotically by Slepian and non-asymptotically by Moitra

$1/f_c$ is the diameter of the main lobe of the point-spread function (twice the Rayleigh distance)
Example: 25 spikes, $f_c = 10^3$, $\Delta = 0.8/f_c$
Example: 25 spikes, $f_c = 10^3$, $\Delta = 0.8/f_c$
Example: 25 spikes, $f_c = 10^3$, $\Delta = 0.8/f_c$

The difference is almost in the null space of the measurement operator
Theoretical questions

1. Is the problem well posed?

2. Does $TV$-norm minimization work?
Super-resolution via TV-norm minimization

\[
\begin{align*}
\text{minimize} & \quad \| \tilde{x} \|_{TV} \\
\text{subject to} & \quad \int \phi(\theta) \tilde{x}(\theta) \, d\theta = y
\end{align*}
\]
Dual certificate for TV-norm minimization

\( v \in \mathbb{R}^n \) is a **dual certificate** associated to

\[
x = \sum_j c_j \delta_{\theta_j} \quad c_j \in \mathbb{R}, \theta_j \in T
\]

if

\[
Q(\theta) := v^T \phi(\theta)
\]

\[
Q(\theta_j) = \text{sign}(c_j) \quad \text{if } \theta_j \in T
\]

\[
|Q(\theta)| < 1 \quad \text{if } \theta \notin T
\]

Dual variable guaranteeing that \( \|x\|_{TV} \) is optimal
Dual certificate

For any $x + h$ such that $\int \phi(\theta) h(\theta) \, d\theta = 0$

$$\|x + h\|_{TV} = \sup_{\|f\|_{\infty} \leq 1} \int_{[0,1]} f(\theta) x(\theta) \, d\theta + \int_{[0,1]} f(\theta) h(\theta) \, d\theta$$
Dual certificate

For any $x + h$ such that $\int \phi(\theta) h(\,d\theta) = 0$

$$\|x + h\|_{TV} = \sup_{\|f\|_\infty \leq 1} \left( \int_{[0,1]} f(\theta) x(\,d\theta) + \int_{[0,1]} f(\theta) h(\,d\theta) \right)$$

$$\geq \int_{[0,1]} Q(\theta) x(\,d\theta) + \int_{[0,1]} Q(\theta) h(\,d\theta)$$
Dual certificate

For any $x + h$ such that $\int \phi(\theta) h(\,d\theta) = 0$

$$\|x + h\|_{TV} = \sup_{\|f\|_{\infty} \leq 1} \int_{[0,1]} f(\theta) x(\,d\theta) + \int_{[0,1]} f(\theta) h(\,d\theta)$$

$$\geq \int_{[0,1]} Q(\theta) x(\,d\theta) + \int_{[0,1]} Q(\theta) h(\,d\theta)$$

$$\geq \sum_{\theta_j \in T} \int_{[0,1]} Q(\theta_j) c_j \delta_{\theta_j}(\,d\theta) + \nu^T \int_{[0,1]} \phi(\theta) h(\,d\theta)$$
Dual certificate

For any $x + h$ such that $\int \phi(\theta) h(\,d\theta) = 0$

$$\|x + h\|_{\text{TV}} = \sup_{\|f\|_{\infty} \leq 1} \int_{[0,1]} f(\theta) x(\,d\theta) + \int_{[0,1]} f(\theta) h(\,d\theta)$$

$$\geq \int_{[0,1]} Q(\theta) x(\,d\theta) + \int_{[0,1]} Q(\theta) h(\,d\theta)$$

$$\geq \sum_{\theta_j \in T} \int_{[0,1]} Q(\theta_j) c_j \delta_{\theta_j}(\,d\theta) + \nu^T \int_{[0,1]} \phi(\theta) h(\,d\theta)$$

$$= \|x\|_{\text{TV}}$$
Dual certificate

For any $x + h$ such that $\int \phi(\theta) h(\,d\theta) = 0$

$$\|x + h\|_{TV} = \sup_{\|f\|_\infty \leq 1} \left( \int_{[0,1]} f(\theta) x(\,d\theta) + \int_{[0,1]} f(\theta) h(\,d\theta) \right)$$

$$\geq \int_{[0,1]} Q(\theta) x(\,d\theta) + \int_{[0,1]} Q(\theta) h(\,d\theta)$$

$$\geq \sum_{\theta_j \in T} \int_{[0,1]} Q(\theta_j) c_j \delta_{\theta_j}(\,d\theta) + v^T \int_{[0,1]} \phi(\theta) h(\,d\theta)$$

$$= \|x\|_{TV}$$

Existence of $Q$ for any sign pattern implies that $x$ is the unique solution.
Dual certificate for super-resolution

\( \nu \in \mathbb{C}^n \) is a dual certificate associated to

\[
x = \sum_j c_j \delta_{\theta_j} \quad c_j \in \mathbb{C}, \; \theta_j \in T
\]

if

\[
Q(\theta) := \nu^* \phi(\theta) = \sum_{k=-f_c}^{f_c} \overline{\nu_k} \exp(i2\pi k\theta)
\]

\[
Q(\theta_j) = \text{sign}(c_j) \quad \text{if } \theta_j \in T
\]

\[
|Q(\theta)| < 1 \quad \text{if } \theta \notin T
\]

Linear combination of low pass sinusoids
Certificate for super-resolution

Aim: Interpolate sign pattern
Certificate for super-resolution

Interpolation with a low-frequency fast-decaying kernel $F$

$$q(t) = \sum_{\theta_j \in T} \alpha_j F(t - \theta_j)$$
Certificate for super-resolution

Interpolation with a low-frequency fast-decaying kernel $F$

$$q(t) = \sum_{\theta_j \in T} \alpha_j F(t - \theta_j)$$
Certificate for super-resolution

Interpolation with a low-frequency fast-decaying kernel $F$

$$q(t) = \sum_{\theta_j \in T} \alpha_j F(t - \theta_j)$$
Certificate for super-resolution

Interpolation with a low-frequency fast-decaying kernel $F$

$$q(t) = \sum_{\theta_j \in T} \alpha_j F(t - \theta_j)$$
Certificate for super-resolution

Interpolation with a low-frequency fast-decaying kernel $F$

$$q(t) = \sum_{\theta_j \in T} \alpha_j F(t - \theta_j)$$
Certificate for super-resolution

Technical detail: Magnitude of certificate locally exceeds 1

\[
\text{Solution: Add correction term and force derivative to vanish on support}
\]

\[
Q(\theta) = \sum_{\theta_j \in T} \alpha_j F(\theta - \theta_j) + \beta_j F'(\theta - \theta_j)
\]
Certificate for super-resolution

Technical detail: Magnitude of certificate locally exceeds 1

Solution: Add correction term and force derivative to vanish on support

\[ Q(\theta) = \sum_{\theta_j \in T} \alpha_j F(\theta - \theta_j) + \beta_j F'(\theta - \theta_j) \]
Certificate for super-resolution

Technical detail: Magnitude of certificate locally exceeds 1

Solution: Add correction term and force derivative to vanish on support

\[ Q(\theta) = \sum_{\theta_j \in T} \alpha_j F(\theta - \theta_j) + \beta_j F'(\theta - \theta_j) \]
Guarantees for super-resolution

Theorem [Candès, F. 2012]

If the minimum separation of the signal support obeys

\[ \Delta \geq 2 / f_c \]

then recovery via convex programming is exact

Theorem [Candès, F. 2012]

In 2D convex programming super-resolves point sources with a minimum separation of

\[ \Delta \geq 2.38 / f_c \]

where \( f_c \) is the cut-off frequency of the low-pass kernel
Guarantees for super-resolution

Theorem [F. 2016]

If the minimum separation of the signal support obeys

$$\Delta \geq 1.26 / f_c,$$

then recovery via convex programming is exact.

Theorem [Candès, F. 2012]

In 2D convex programming super-resolves point sources with a minimum separation of

$$\Delta \geq 2.38 / f_c$$

where $f_c$ is the cut-off frequency of the low-pass kernel.
Numerical evaluation of minimum separation

\[ f_c = 30 \]

\[ f_c = 40 \]

\[ f_c = 50 \]

Numerically TV-norm minimization succeeds if \( \Delta \geq \frac{1}{f_c} \)
Separable Nonlinear Inverse Problems

Sparse Recovery for SNL Problems

Super-resolution

Deconvolution

SNL Problems with Correlation Decay
Deconvolution

Joint work with Brett Bernstein (Courant)
Reflection seismology

Geological section | Acoustic impedance | Reflection coefficients
Reflection seismology

Data $\approx$ convolution of pulse and reflection coefficients
Model for the pulse: Ricker wavelet
Toy model for reflection seismology
Toy model for reflection seismology
Toy model for reflection seismology
Toy model for reflection seismology
Toy model for reflection seismology
Toy model for reflection seismology
Toy model for reflection seismology
Toy model for reflection seismology
Toy model for reflection seismology
Deconvolution

$s$ sources with locations $\theta_1, \ldots, \theta_s$, modeled as superposition of spikes

$$x = \sum_j c(j) \delta_{\theta_j} \quad c_j \in \mathbb{R}, \theta_j \in T \subset [0, 1]$$

We observe samples of convolution with kernel $K$

$$y(k) = (K \ast x)(s_k)$$

$$= \sum_{j=1}^s c(j) K(s_k - \theta_j)$$

SNL problem where

$$\phi(\theta_j) = \begin{bmatrix} K(s_1 - \theta_j) \\ \vdots \\ K(s_n - \theta_j) \end{bmatrix}$$
Theoretical questions

1. Is the problem well posed?

2. Does $TV$-norm minimization work?
Minimum separation

Kernels are approximately low-pass

The support cannot be too clustered
Sampling proximity

We need two samples per spike

Convolution kernel decays: at least two samples close to each spike

Samples $S$ and support $T$ have sample proximity $\gamma$ if for every $\theta_i \in T$ there exist $s_i, s'_i \in S$ such that

$$|\theta_i - s_i| \leq \gamma \quad \text{and} \quad |\theta_i - s'_i| \leq \gamma$$

We consider arbitrary non-uniform sampling patterns with fixed $\gamma$
Sampling proximity
Theoretical questions

1. Is the problem well posed?

2. Does $TV$-norm minimization work?
Deconvolution via TV-norm minimization

minimize \|\tilde{x}\|_{TV}

subject to \int \phi(\theta) \tilde{x}(d\theta) = y
Dual certificate for SNL problems

$v \in \mathbb{R}^n$ is a dual certificate associated to

$$x = \sum_j c_j \delta_{\theta_j} \quad c_j \in \mathbb{R}, \theta_j \in T$$

if

$$Q(\theta) := v^T \phi(\theta) = \sum_{k=1}^n v_k K(s_k - \theta)$$

$$Q(\theta_j) = \text{sign}(c_j) \quad \text{if } \theta_j \in T$$

$$|Q(\theta)| < 1 \quad \text{if } \theta \notin T$$

Linear combination of shifted copies of $K$ fixed at the samples
Certificate for deconvolution

\[ \theta_1 \quad \theta_2 \quad \theta_3 \]
Certificate construction

Only use subset $\tilde{S}$ containing 2 samples close to each spike

$$Q(\theta) = \sum_{s_j \in \tilde{S}} v_j K(s_j - \theta)$$

Fit $v$ so that for all $\theta_i \in T$

$$Q(\theta_i) = \text{sign}(c_i)$$
$$Q'(\theta_i) = 0$$
It works!
It works!
Certificate construction

**Problem:** The construction is difficult to analyze (coefficients vary)

**Solution:** Reparametrization into *bumps* and *waves*

\[ Q(\theta) = \sum_{s_j \in \tilde{S}} v_j K(s_j - \theta) \]
\[ = \sum_{\theta_i \in T} \alpha_i B_{\theta_i}(\theta, \tilde{s}_{i,1}, \tilde{s}_{i,2}) + \beta_i W_{\theta_i}(\theta, \tilde{s}_{i,1}, \tilde{s}_{i,2}), \]
Bump function (Gaussian kernel)

\[ B_{\theta_i}(\theta, \tilde{s}_i, 1, \tilde{s}_i, 2) := b_{i,1}K(\tilde{s}_i, 1 - \theta) + b_{i,2}K(\tilde{s}_i, 2 - \theta) \]

\[ B_{\theta_i}(\theta_i, \tilde{s}_i, 1, \tilde{s}_i, 2) = 1 \]
\[ \frac{\partial}{\partial \theta} B_{\theta_i}(\theta_i, \tilde{s}_i, 1, \tilde{s}_i, 2) = 0 \]
Wave function (Gaussian kernel)

\[ W_{\theta_i}(\theta, \tilde{s}_{i,1}, \tilde{s}_{i,2}) = w_{i,1} K(\tilde{s}_{i,1} - \theta) + w_{i,2} K(\tilde{s}_{i,2} - \theta) \]

\[ W_{\theta_i}(\theta_i, \tilde{s}_{i,1}, \tilde{s}_{i,2}) = 0 \]

\[ \frac{\partial}{\partial \theta} W_{\theta_i}(\theta_i, \tilde{s}_{i,1}, \tilde{s}_{i,2}) = 1 \]
Bump function (Ricker wavelet)

\[
B_{\theta_i}(\theta, \tilde{s}_{i,1}, \tilde{s}_{i,2}) := b_{i,1} K(\tilde{s}_{i,1} - \theta) + b_{i,2} K(\tilde{s}_{i,2} - \theta)
\]

\[
B_{\theta_i}(\theta_i, \tilde{s}_{i,1}, \tilde{s}_{i,2}) = 1
\]

\[
\frac{\partial}{\partial \theta} B_{\theta_i}(\theta_i, \tilde{s}_{i,1}, \tilde{s}_{i,2}) = 0
\]
Wave function (Ricker wavelet)

\[ W_{\theta_i}(\theta, \tilde{s}_{i,1}, \tilde{s}_{i,2}) = w_{i,1} K(\tilde{s}_{i,1} - \theta) + w_{i,2} K(\tilde{s}_{i,2} - \theta) \]

\[ W_{\theta_i}(\theta_i, \tilde{s}_{i,1}, \tilde{s}_{i,2}) = 0 \]

\[ \frac{\partial}{\partial \theta} W_{\theta_i}(\theta_i, \tilde{s}_{i,1}, \tilde{s}_{i,2}) = 1 \]
Certificate construction

Reparametrization decouples the coefficients

\[ Q(\theta) = \sum_{s_j \in \tilde{S}} v_j K(s_j - \theta) \]

\[ = \sum_{\theta_i \in T} \alpha_i B_{\theta_i}(\theta, \tilde{s}_{i,1}, \tilde{s}_{i,2}) + \beta_i W_{\theta_i}(\theta, \tilde{s}_{i,1}, \tilde{s}_{i,2}) \]

\[ \approx \sum_{\theta_i \in T} \text{sign}(c_i) B_{\theta_i}(\theta, \tilde{s}_{i,1}, \tilde{s}_{i,2}) \]
Certificate for deconvolution (Gaussian kernel)
Certificate for deconvolution (Gaussian kernel)
Certificate for deconvolution (Ricker wavelet)
Certificate for deconvolution (Ricker wavelet)
Exact recovery guarantees [Bernstein, F. 2017]

Sample Proximity
Gaussian Kernel
Numerical Recovery
Exact Recovery

Spike Separation
0.2
0.4
0.6
0.8
1
1.2
Sample Proximity
Gaussian Kernel
Numerical Recovery
Exact Recovery
Exact recovery guarantees [Bernstein, F. 2017]
Separable Nonlinear Inverse Problems

Sparse Recovery for SNL Problems

Super-resolution

Deconvolution

SNL Problems with Correlation Decay
SNL problems

Joint work with Brett Bernstein (Courant) and Sheng Liu (CDS, NYU)
General SNL problems

The function $\phi$ may not be available explicitly but can often be computed numerically by solving a differential equation.

- Source localization in EEG
- Direction of arrival in radar / sonar
- Magnetic-resonance fingerprinting
Mathematical model

- **Signal**: superposition of Dirac measures with support $T$

  $x = \sum_{j} c_j \delta_{\theta_j} \quad c_j \in \mathbb{R}, \; \theta_j \in T \subset [0, 1]$

- **Data**: $n$ measurements following SNL model

  $y = \int \phi(\theta) \times (d\theta)$
Diffusion on a rod with varying conductivity

\[ \theta_1 \]
Diffusion on a rod with varying conductivity
Diffusion on a rod with varying conductivity

\[ \theta_1 \quad \theta_2 \quad \theta_3 \]
Diffusion on a rod with varying conductivity
Diffusion on a rod with varying conductivity

$\phi(\theta)$ can be computed by solving differential equation
Time-frequency pulses

$\theta_1$
Time-frequency pulses
Time-frequency pulses
Time-frequency pulses
Time-frequency pulses

Gabor wavelets in 1D

\[ \phi(\theta)_k = \exp\left(-\frac{(s_k - \theta)^2}{2\sigma}\right) \sin(150 s_k (s_k - \theta)) \quad \theta \in [0, 1] \]
Sparse estimation for general SNL problems

**Problem:** Sparse recovery requires RIP-like properties that do not hold for SNL problems with smooth $\phi$ (even if we discretize)

We cannot hope to recover all sparse signals

How about signals such that $\phi(\theta_i)^T \phi(\theta_j)$ is small for all $\theta_i \neq \theta_j$ in $T$?

**Challenge:** Prove guarantees for general SNL problems that only depend on correlation structure
Aim: Guarantees for signals under separation conditions with respect to support-centered correlations $\rho_1, \ldots, \rho_s$

$$\rho_i(\theta) := \phi(\theta_i)^T \phi(\theta)$$
Diffusion on a rod with varying conductivity
Support-centered correlations
\[ \phi(\theta)_k = \exp \left( -\frac{(s_k - \theta)^2}{2\sigma} \right) \sin(150 s_k (s_k - \theta)) \quad \theta \in [0, 1] \]
Support-centered correlations

\[ \theta_0 \quad \theta_1 \quad \theta_2 \]
Correlation decay

Parametrized by $F_i^- < N_i^- < N_i^+ < F_i^+$ and $\sigma_i$

Parameters can be different at each $\theta_i$
Correlation decay

- $\rho_i$ is concave in $[N_i^-, N_i^+]$: $\rho''_i(\theta) < -\gamma_0$

- $\rho_i$ is bounded outside $[N_i^-, N_i^+]$: $|\rho_i(\theta)| < \gamma_1$

- $\rho_i$ decays for $\theta < F_i^-$ and $\theta > F_i^+$:
  
  $|\rho_i(\theta)| < \gamma_2 e^{-(|\theta - F_i^-|)/\sigma_i}$ for $\theta < F_i^-$
  
  $|\rho_i(\theta)| < \gamma_2 e^{-(|\theta - F_i^+|)/\sigma_i}$ for $\theta > F_i^+$

Choice of exponential decay is arbitrary
Correlation decay

Additional condition on correlation derivatives

\[ \rho_i^{(q,r)}(\theta) := \phi(q)(\theta_i)^T \phi(r)(\theta) \]

\( \rho_i^{(q,r)} \) decays for \( \theta < F_i^- \) and \( \theta > F_i^+ \), for \( q = 0, 1, r = 0, 1, 2 \):

\[ \left| \rho_i^{(q,r)}(\theta) \right| < \gamma_2 e^{-|\theta - F_i^-|/\sigma_i} \text{ for } \theta < F_i^- \]
\[ \left| \rho_i^{(q,r)}(\theta) \right| < \gamma_2 e^{-|\theta - F_i^+|/\sigma_i} \text{ for } \theta > F_i^+ \]

Open question: Is this necessary?
Normalized distance

Normalized distance from $\theta$ to $\theta_j > \theta$

$$d_j(\theta) := \frac{F_j^- - \theta}{\sigma_j}$$

If $d_j(\theta)$ is large, $\phi(\theta)$ and $\phi(\theta_j)$ are not very correlated
Minimum separation conditions

Normalized distance between spikes is equal to $\Delta$

\[
\theta_1 \quad F_1^+ \quad \geq \Delta \sigma_2 \quad F_2^- \quad \theta_2 \quad F_3^- \quad \theta_3
\]

\[
\geq 2\Delta \sigma_3
\]

\[
\theta_1 \quad \theta_{1-2} \quad F_2^- \quad \theta_2 \quad F_3^- \quad \theta_3
\]

\[
\geq \frac{\Delta \sigma_2}{2}
\]

\[
\geq \frac{3\Delta \sigma_3}{2}
\]
Guarantees for SNL problems with decaying correlation

Theorem [Bernstein, F. 2018]

For any SNL problem with decaying correlation TV-norm minimization achieves exact recovery under the separation conditions if

\[ \Delta > C \]

for a fixed constant \( C \) depending on the decay bounds \( \gamma_0, \gamma_1, \gamma_2 \)
Dual certificate for SNL problems

$v \in \mathbb{R}^n$ is a dual certificate associated to

$$x = \sum_j c_j \delta_{\theta_j} \quad c_j \in \mathbb{R}, \, \theta_j \in T$$

if

$$Q(\theta) := v^T \phi(\theta)$$

$$Q(\theta_j) = \text{sign}(c_j) \quad \text{if } \theta_j \in T$$

$$|Q(\theta)| < 1 \quad \text{if } \theta \notin T$$
Dual certificate construction

Use support-centered correlations to interpolate sign pattern

\[ Q(\theta) := \sum_{i=1}^{s} \alpha_i \rho_i(\theta) \]
Dual certificate construction

Use support-centered correlations to interpolate sign pattern

\[ Q(\theta) := \sum_{i=1}^{s} \alpha_i \rho_i(\theta) \]
Dual certificate construction

Use support-centered correlations to interpolate sign pattern

\[
Q(\theta) := \sum_{i=1}^{s} \alpha_i \rho_i(\theta)
\]
Dual certificate construction

Use support-centered correlations to interpolate sign pattern

\[ Q(\theta) := \sum_{i=1}^{s} \alpha_i \rho_i(\theta) \]
Dual certificate construction

Use support-centered correlations to interpolate sign pattern

\[ Q(\theta) := \sum_{i=1}^{s} \alpha_i \rho_i(\theta) \]

\[ = \sum_{i=1}^{s} \alpha_i \phi(\theta_i)^T \phi(\theta) \]
Dual certificate construction

Use support-centered correlations to interpolate sign pattern

\[ Q(\theta) := \sum_{i=1}^{s} \alpha_i \rho_i(\theta) \]

\[ = \sum_{i=1}^{s} \alpha_i \phi(\theta_i) ^T \phi(\theta) \]

\[ = v ^T \phi(\theta) \quad v := \sum_{i=1}^{s} \alpha_i \phi(\theta_i) \]
Dual certificate construction

Use support-centered correlations to interpolate sign pattern

\[ Q(\theta) := \sum_{i=1}^{s} \alpha_i \rho_i(\theta) \]

\[ = \sum_{i=1}^{s} \alpha_i \phi(\theta_i)^T \phi(\theta) \]

\[ = v^T \phi(\theta) \quad v := \sum_{i=1}^{s} \alpha_i \phi(\theta_i) \]

Technical detail: Correction term to ensure derivative vanishes
Robustness to noise / outliers

Variations of dual certificates establish robustness at small noise levels (Candès, F. 2013), (F. 2013), (Bernstein, F. 2017)

Exact recovery with constant number of outliers (up to log factors) (F., Tang, Wang, Zheng 2017), (Bernstein, F. 2017)

Open questions: Analysis of higher-noise levels and discretization error, robustness for positive amplitudes
Drawbacks

Solving convex program is computationally expensive

Approach doesn’t scale well at high dimensions

In practice, reweighting is need to obtain sparse solutions for noisy data

Open question: Analysis of other techniques (reweighting methods, descent methods on nonconvex cost functions)
Conclusion

Previous works focus mostly on random operators

For deterministic problems sparsity is not enough!

Under separation conditions:

1. Sharp guarantees for super-resolution and deconvolution

2. General guarantees for SNL problems with correlation decay
References


- *The recoverability limit for superresolution via sparsity*. L. Demanet, N. Nguyen


References


References


▶ Deconvolution of point sources: A sampling theorem and robustness guarantees. B. Bernstein, C. Fernandez-Granda

▶ Sparse recovery beyond compressed sensing: Separable nonlinear inverse problems. B. Bernstein, C. Fernandez-Granda, S. Liu