On robustness of ℓ_1-regularization methods for spectral estimation

Johan Karlsson
Department of Mathematics
KTH Royal Institute of Technology

ROYAL INSTITUTE OF TECHNOLOGY

Lipeng Ning
Brigham & Women's Hospital
Harvard Medical School

HARVARD MEDICAL SCHOOL

Conference on Decision and Control
Los Angeles, December 2014

On robustness of ℓ_1-regularization methods for spectral estimation

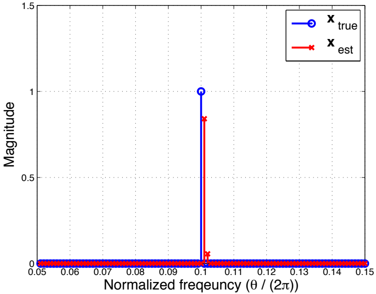
$$\begin{aligned} \text{Data:} \quad & \mathbf{y} && = \mathbf{A}\mathbf{x}_{\text{true}} + \text{noise} \\ \text{Recovery:} \quad & \mathbf{x}_{\text{est}} && = \arg\min \|\mathbf{x}\|_1 \\ &&& \quad \text{subject to } \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2 \leq \text{error} \end{aligned}$$

Distance between $\mathbf{x}_{\text{true}}$ and $\mathbf{x}_{\text{est}}$?

A incoherent: Robust recovery (ℓ_2) by Candès, Donoho, Tao, Tropp, etal.

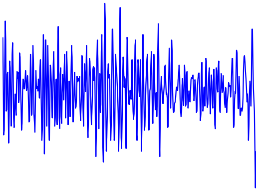
A coherent:

- Typical in spectral estimation.
- Robust recovery (ℓ_2) impossible.
- Robust recovery in Transportation distance for sparse separated $\mathbf{x}_{\text{true}}$.

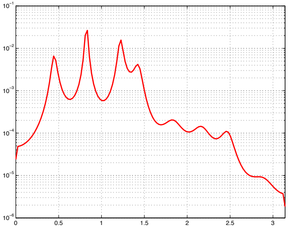


Outline

Spectral estimation



$\Rightarrow$



$\ldots y_0, y_1, \ldots$

$$y_k = \int_{\mathbb{T}} e^{i\theta k} dX(\theta)$$

$$d\mu(\theta) = E(|dX(\theta)|^2)$$

Applications:

- Speech analysis
- Medical diagnostics
- Radar/Sonar
- Communications
- System identification

Estimation Methods:

- Periodogram, Correlogram
- Burg, THREE (Maximum entropy)
- Capon
- MUSIC, ESPRIT
- **Sparse methods**

Problem setting

Discrete-time signal $y_n \in \mathbb{C}$, $n \in \mathbb{Z}$. Sinusoids in noise:

$$y_n = \sum_{\ell=1}^{L} x_\ell e^{in\lambda_\ell} + w_n, \quad \text{for } n = 0, 1 \ldots, N-1$$

$\lambda_\ell \in [-\pi, \pi]$	frequency
$x_\ell \in \mathbb{C}$	magnitude and phase
$w_n \in \mathbb{C}$	error/noise.

Typically the frequency grid is discretized resulting in the linear system

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{w} \tag{1}$$

columns of $\mathbf{A} \in \mathbb{C}^{N \times K}$ form a normalized overcomplete Fourier basis

$$\mathbf{a}(\theta_k) = \frac{1}{\sqrt{N}} \left(1, e^{i\theta_k}, \ldots, e^{i(N-1)\theta_k}\right)^T$$

for $\Omega = \{\theta_1, \theta_2, \ldots, \theta_K\}$. Here $\mathbf{x} \in \mathbb{C}^K$, $\mathbf{w} \in \mathbb{C}^N$, $N < K$.

(1) is underdetermined and to single out a unique solution, the use ℓ_1 regularization has recently been very popular (Chen, Donoho, 1998).

Sparse methods

- Find the most sparse solution: a combinatorial problem.
- Use ℓ_1-norm as surrogate for the cardinality:

$$\arg\min_{\mathbf{x}\in\mathbb{C}^K} \|\mathbf{x}\|_1 \quad \text{subject to } \|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2 \leq \epsilon. \tag{2}$$

  Empirically: good recovery of the true $\mathbf{x}_{\rm true}$.

Theorem 1: (Candes, Wakin, 2008) Assume that for $\delta_{2s} < \sqrt{2}-1$, the inequality

$$(1-\delta_{2s})\|\mathbf{x}\|_2^2 \leq \|\mathbf{A}\mathbf{x}\|_2^2 \leq (1+\delta_{2s})\|\mathbf{x}\|_2^2 \qquad \text{(RIP)}$$

holds for all $2s-$sparse x. Let $\mathbf{y} = \mathbf{A}\mathbf{x}_{\rm true} + \mathbf{w}$ where $\mathbf{x}_{\rm true}$ is s-sparse and $\|\mathbf{w}\|_2 \leq \epsilon$, then the minimizer $\mathbf{x}_{\rm est}$ of (2) satisfy

$$\|\mathbf{x}_{\rm true} - \mathbf{x}_{\rm est}\|_2 \leq C(\delta_{2s})\epsilon.$$

□

- In high resolution spectral estimation, typically $\delta_2 \geq 0.95 > \sqrt{2}-1 \approx 0.41$.
- **A** highly coherent $\Rightarrow$ standard sparse results not applicable.
- What can we say in this case?

Sparse methods

Example:

- One sinusoid $\mathbf{y} := \mathbf{a}(0.1) + \mathbf{w}$
- $\|\mathbf{w}\|_2 = \epsilon = 10\%$
- Results in relative ℓ_2-error of 130%.

A is highly coherent: robust recovery in the sense of ℓ_2-norm is not possible.

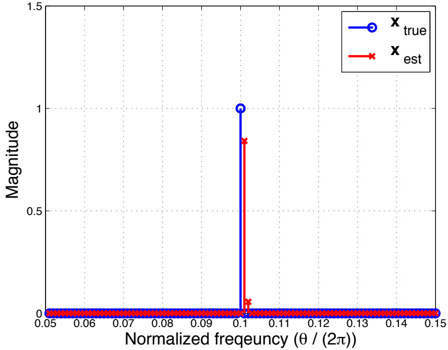


Figure: Here SNR:= $\|\mathbf{Ax}\|_2/\|\mathbf{w}\|_2 = 10$, $N = 100$, and $K = 1000$.

Motivating Example

Consider the example where $\mathbf{x}_{\text{true}} = \mathbf{e}(\lambda)$ has support in only one point λ. The signal is given by

$$\mathbf{y} = \mathbf{A}\mathbf{x}_{\text{true}} + \mathbf{w} = \mathbf{a}(\lambda) + \mathbf{w}, \quad \text{where } \|\mathbf{w}\|_2 \le \epsilon,$$

and let $\mathbf{x}_{\text{est}}$ be the recovered solution from (2):

$$\mathbf{x}_{\text{est}} = \arg\min \|\mathbf{x}\|_1 \text{ subject to } \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2 \le \epsilon.$$

Note that

$$\begin{aligned}
\epsilon &\ge \|\mathbf{A}\mathbf{x}_{\text{est}} - \mathbf{y}\|_2 \ge |\mathbf{a}(\lambda)^*(\mathbf{A}\mathbf{x}_{\text{est}} - \mathbf{y})| \\
&= \Big|\underbrace{\mathbf{a}(\lambda)^*\mathbf{a}(\lambda)}_{=1=\|\mathbf{x}_{\text{true}}\|_1} - \sum_{\theta\in\Omega} \mathbf{x}_{\text{est}}(\theta)\mathbf{a}(\lambda)^*\mathbf{a}(\theta) + \underbrace{\mathbf{a}(\lambda)^*\mathbf{w}}_{|\cdot|\le\epsilon}\Big| \\
&\ge \|\mathbf{x}_{\text{true}}\|_1 - \sum_{\theta\in\Omega} |\mathbf{x}_{\text{est}}(\theta)||\mathbf{a}(\lambda)^*\mathbf{a}(\theta)| - \epsilon \\
&= \sum_{\theta\in\Omega} \underbrace{|\mathbf{x}_{\text{est}}(\theta)|(1 - |\mathbf{a}(\lambda)^*\mathbf{a}(\theta)|)}_{\text{"mass" transport}} + \underbrace{\|\mathbf{x}_{\text{true}}\|_1 - \|\mathbf{x}_{\text{est}}\|_1}_{\text{"mass" deviation}} - \epsilon.
\end{aligned}$$

Motivating Example

Consider the example where $\mathbf{x}_{\text{true}} = \mathbf{e}(\lambda)$ has support in only one point λ. The signal is given by

$$\mathbf{y} = \mathbf{A}\mathbf{x}_{\text{true}} + \mathbf{w} = \mathbf{a}(\lambda) + \mathbf{w}, \quad \text{where } \|\mathbf{w}\|_2 \leq \epsilon,$$

and let $\mathbf{x}_{\text{est}}$ be the recovered solution from (2):

$$\mathbf{x}_{\text{est}} = \arg\min \|\mathbf{x}\|_1 \text{ subject to } \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2 \leq \epsilon.$$

Note that

$$\begin{aligned}
\epsilon &\geq \|\mathbf{A}\mathbf{x}_{\text{est}} - \mathbf{y}\|_2 \geq |\mathbf{a}(\lambda)^*(\mathbf{A}\mathbf{x}_{\text{est}} - \mathbf{y})| \\
&= \Big| \underbrace{\mathbf{a}(\lambda)^*\mathbf{a}(\lambda)}_{=1=\|\mathbf{x}_{\text{true}}\|_1} - \sum_{\theta\in\Omega} \mathbf{x}_{\text{est}}(\theta)\mathbf{a}(\lambda)^*\mathbf{a}(\theta) + \underbrace{\mathbf{a}(\lambda)^*\mathbf{w}}_{|\cdot|\leq\epsilon} \Big| \\
&\geq \|\mathbf{x}_{\text{true}}\|_1 - \sum_{\theta\in\Omega} |\mathbf{x}_{\text{est}}(\theta)||\mathbf{a}(\lambda)^*\mathbf{a}(\theta)| - \epsilon \\
&= \sum_{\theta\in\Omega} \underbrace{|\mathbf{x}_{\text{est}}(\theta)|(1 - |\mathbf{a}(\lambda)^*\mathbf{a}(\theta)|)}_{\text{“mass” transport}} + \underbrace{\|\mathbf{x}_{\text{true}}\|_1 - \|\mathbf{x}_{\text{est}}\|_1}_{\text{“mass” deviation}} - \epsilon.
\end{aligned}$$

Motivating Example

$$2\epsilon \geq \sum_{\theta\in\Omega} \underbrace{|\mathbf{x}_{\text{est}}(\theta)|(1-|\mathbf{a}(\lambda)^*\mathbf{a}(\theta)|)}_{\text{“mass” transport}} + \underbrace{\|\mathbf{x}_{\text{true}}\|_1 - \|\mathbf{x}_{\text{est}}\|_1}_{\text{“mass” deviation}}.$$

- Term 1: Transportation cost where the cost of transporting from θ to λ is

$$c(\lambda,\theta) = 1 - |\mathbf{a}(\lambda)^*\mathbf{a}(\theta)|.$$

- Term 2: Deviation in the total mass. By optimality $\|\mathbf{x}_{\text{true}}\|_1 - \|\mathbf{x}_{\text{est}}\|_1 \geq 0$.

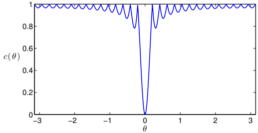
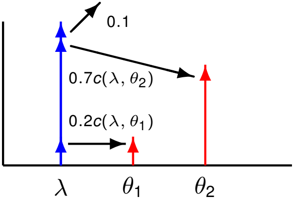


Figure: $c(0,\theta)$ for $N=30$.

Example:

$\mathbf{x}_{\text{true}} = \mathbf{e}(\lambda)$ and $\mathbf{x}_{\text{est}} = 0.2\mathbf{e}(\theta_1) + 0.7\mathbf{e}(\theta_2)$

$$\Rightarrow 2\epsilon \geq 0.2c(\lambda,\theta_1) + 0.7c(\lambda,\theta_2) + 0.1.$$



Transportation distances

Transportation cost (Monge 1781, Kantorovich 1942):

$$
\begin{aligned}
T_N(\rho_0,\rho_1) := \min \quad & \sum_{\theta\in\Omega}\sum_{\omega\in\Omega} c_N(\theta,\omega)M(\theta,\omega) \\
\text{subject to} \quad & \sum_{\omega\in\Omega} M(\theta,\omega)=\rho_0(\theta) \;\text{ and }\; \sum_{\theta\in\Omega} M(\theta,\omega)=\rho_1(\omega) \\
& M(\theta,\omega)\geq 0, \quad \theta,\omega\in\Omega
\end{aligned}
$$

Relaxed transportation cost allowing for different masses
(K., Georgiou, Takyar, 2009):

$$
\tilde{T}_N(\mathbf{x}_0,\mathbf{x}_1) \quad := \inf_{\|\rho_0\|_1=\|\rho_1\|_1}\left(T_N(\rho_0,\rho_1) + \sum_{j=0}^{1} \||\mathbf{x}_j| - \rho_j\|_1 \right)
$$

where $|\mathbf{x}| := (|\mathbf{x}(j)|)_{j=1}^N$ denotes element-wise absolute value.

Worst case bounds

General case

$$\mathbf{x}_{\text{true}} = \sum_{\lambda \in \Lambda} \alpha_\lambda \mathbf{e}(\lambda), \quad \text{supp}(\mathbf{x}_{\text{true}}) = \Lambda \subset \Omega$$

Definition 2: Let **A** be a dictionary with index set Ω and let $\Lambda \subset \Omega$. Then we define

$$\mu_\Lambda \quad := \quad \max_{\theta \in \Omega} \left(\sum_{\lambda \in \Lambda} |\mathbf{a}(\lambda)^* \mathbf{a}(\theta)| - \max_{\lambda \in \Lambda} |\mathbf{a}(\lambda)^* \mathbf{a}(\theta)| \right),$$

which we denote by the *cumulative intercoherence*. □

μ_Λ: quantifies sparsity and separateness of the support Λ .

Proposition 3:

$$\mu_\Lambda \leq \frac{(|\Lambda| - 1)\pi}{N \Delta(\Lambda)}$$

where

$$\Delta(\Lambda) = \min_{\lambda_0, \lambda_1 \in \Lambda, \lambda_0 \neq \lambda_1} |\lambda_0 - \lambda_1|.$$

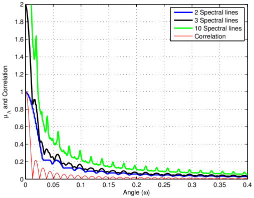


Figure: Intercoherence for $L \in \{2, 3, 10\}$ and Λ consisting of L equispaced frequencies with distance θ. Here $N = 100$.

Worst case bounds

We will derive the bound based on the following properties

$$\begin{array}{ll} 3a) & \mathbf{x}_{\text{true}} \text{ has support } \Lambda, \\ 3b) & \|\mathbf{x}_{\text{est}}\|_1 \leq \|\mathbf{x}_{\text{true}}\|_1, \\ 3c) & \|\mathbf{A}\mathbf{x}_{\text{est}} - \mathbf{A}\mathbf{x}_{\text{true}}\|_2 \leq 2\epsilon. \end{array} \tag{3}$$

They hold for $\mathbf{x}_{\text{est}}$ obtained from

$$\arg\min_{\mathbf{x}\in\mathbb{C}^K} \|\mathbf{x}\|_1 \quad \text{subject to } \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 \leq \epsilon,$$

Theorem 4: Let $\mathbf{x}_{\text{true}} \in \mathbb{C}^{K\times 1}$ be a vector with support Λ, and let $\mathbf{w} \in \mathbb{C}^{K\times 1}$ with $\|\mathbf{w}\|_2 \leq \epsilon$. Then the optimal solution $\mathbf{x}_{\text{est}}$ satisfies

$$\tilde{T}(\mathbf{x}_{\text{est}}, \mathbf{x}_{\text{true}}) \leq 6\left(\epsilon\sqrt{L(1+\mu_\Lambda)} + \mu_\Lambda\|\mathbf{x}_{\text{true}}\|_1\right),$$

where μ_Λ is given by Definition 2 and $L = |\Lambda|$. □

Error bounds with given confidence level

If noise $\geq$ signal, worst case bounds useless.

If $\mathbf{w}$ nearly orthogonal to the columns of $\mathbf{A}$, i.e., $\delta \ll 1$ and

$$|\mathbf{a}(\theta)^*\mathbf{w}| \leq \delta\|\mathbf{w}\|_2 \quad \text{for all} \quad \theta \in [-\pi, \pi] = \mathbb{T}. \tag{4}$$

In addition assume $\|\mathbf{w}\|_2 = \epsilon$ and

$$\|\mathbf{A}\mathbf{x}_{\rm est} - \mathbf{A}\mathbf{x}_{\rm true} - \mathbf{w}\|_2 \leq \epsilon.$$

Then it follows that

$$\begin{aligned} \|\mathbf{A}\mathbf{x}_{\rm est} - \mathbf{A}\mathbf{x}_{\rm true}\|_2^2 &\leq 2|\mathbf{w}^*(A\mathbf{x}_{\rm est} - A\mathbf{x}_{\rm true})| \\ &\leq 2\delta\epsilon(\|\mathbf{x}_{\rm est}\|_1 + \|\mathbf{x}_{\rm true}\|_1) \\ &\leq 4\delta\epsilon\|\mathbf{x}_{\rm true}\|_1. \end{aligned}$$

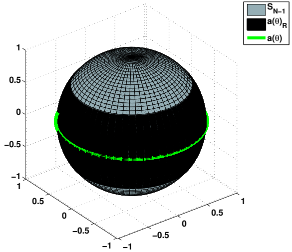
Theorem 5: Let $\mathbf{x}_{\rm true} \in \mathbb{C}^{K\times 1}$ be a vector with support Λ, and let $\mathbf{w} \in \mathbb{C}^{K\times 1}$ with $\|\mathbf{w}\|_2 = \epsilon$ and which satisfies (4). Then the optimal solution $\mathbf{x}_{\rm est}$ satisfies

$$\tilde{T}(\mathbf{x}_{\rm est}, \mathbf{x}_{\rm true}) \leq 6\left(\sqrt{2\delta\epsilon\|\mathbf{x}_{\rm true}\|_1 L(1+\mu_\Lambda)} + \mu_\Lambda\|\mathbf{x}_{\rm true}\|_1\right).$$

Error bounds with given confidence level

When is the near orthogonality assumption justified?

$$|\mathbf{a}(\theta)^*\mathbf{w}| \le \delta\|\mathbf{w}\|_2 \quad \text{for all } \theta$$
$$\Leftrightarrow$$
$$\frac{\mathbf{w}}{\|\mathbf{w}\|_2} \notin B_{\sqrt{(1-\delta)/2}}(\mathbf{a}(\theta)e^{i\phi} : \theta, \phi \in \mathbb{T}) \cap S_{N-1}$$



Volume of tube can be bounded based on Neiman's inequality (Johnstone, Siegmund, 1989).

Proposition 6: Let $\delta \in (0,1)$ be given and let $\mathbf{w} \in \mathbb{C}^N$ be a random vector that is complex Gaussian with zero mean and unit variance. Then

$$\operatorname{Prob}\left(\max_{0\le\theta\le 2\pi} |\mathbf{a}(\theta)^*\mathbf{w}| \ge \delta\|\mathbf{w}\|_2\right) \le (1-\delta)^{N-3/2}\left(1+\delta\frac{e^2}{\sqrt{6}}N^{3/2}\right).$$

□

$\Rightarrow \max_{0\le\theta\le 2\pi} \frac{|\mathbf{a}(\theta)^*\mathbf{w}|}{\|\mathbf{w}\|_2} \to 0$ as $N \to \infty$ in probability.

Spectral estimation based on ℓ_1 regularization

Let $N \in \mathbb{N}$ and let the signal be a sum of sinusoids in noise

$$y_n = \sum_{\lambda \in \Lambda}^{L} e^{in\lambda} x(\lambda) + w_n, \quad \text{for } n = 0, \ldots, N-1$$

where $x(\lambda) \in \mathbb{C}$ for $\lambda \in \Lambda = \{\lambda_1, \ldots, \lambda_L\} \subset \Omega_{K(N)}$, and let $w_n \in CN(0, \sigma^2)$ be white circular complex-valued Gaussian noise.

Let

$$\mathbf{x}_{\text{est}}^N = \arg \min_{\mathbf{x}_N \in \mathbb{C}^K} \|\mathbf{x}_N\|_1 \quad \text{subject to } \|\mathbf{y}_N - \mathbf{A}_N \mathbf{x}_N\|_2 \leq \epsilon = \|\mathbf{w}_N\|,$$

$\mathbf{A}_N \in \mathbb{C}^{K(N) \times N}, \mathbf{x}_N \in \mathbb{C}^{K(N)}, \mathbf{y}_N \in \mathbb{C}^N, \mathbf{w}_N \in \mathbb{C}^N$ are defines as before.

Then

$$\tilde{T}_N(\mathbf{x}_{\text{est}}^N, \mathbf{x}_{\text{true}}) \to 0 \quad \text{in probability, as } N \to \infty.$$

Related work

- Candès and Fernandez-Grada (2012) consider a similar problem:
  If $\Delta(\Lambda) \geq \frac{8\pi}{N}$ then their bound is

$$\|\mathbf{x}_{\text{true}} - \mathbf{x}_{\text{est}}\|_1 \leq C \left(\frac{K}{N}\right)^2 \epsilon.$$

  (+) Transparant condition on Λ.
  (+) Exact recovery in the noiseless case.
  (-) Error bound in noisy case grows quickly as K/N increase.

- Tang, Bhaskar, Recht (2013) use atomic norm minimization:
  (+) Convergence in probability in the weak topology.
  (-) No explicit bounds for finite N.

Conclusions

Conclusions

- Standard ℓ_2 distance is not appropriate for quantifying uncertainty in high-resolution spectral estimation based on ℓ_1 regularization.
- Framework for spectral estimation with uncertainty bounds based on transportation distance.
- Uncertainty bounds, both worst-case and bounds that hold with a guaranteed probability defined.
- The latter bound go to 0 as number of data go to ∞ (in probability).

Thank you for your attention!

Conclusions

Conclusions

- Standard ℓ_2 distance is not appropriate for quantifying uncertainty in high-resolution spectral estimation based on ℓ_1 regularization.
- Framework for spectral estimation with uncertainty bounds based on transportation distance.
- Uncertainty bounds, both worst-case and bounds that hold with a guaranteed probability defined.
- The latter bound go to 0 as number of data go to ∞ (in probability).

Thank you for your attention!