Super-Resolution from Noisy Data

Carlos Fernandez-Granda

www.stanford.edu/~cfgranda/

SIAM Conference on Imaging Science

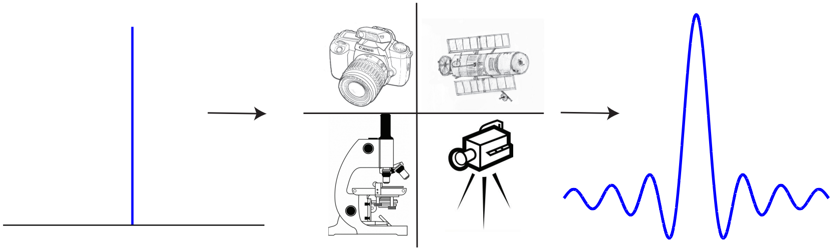
5/14/2014

Acknowledgements

- This work was supported by a Fundación La Caixa Fellowship and a Fundación Caja Madrid Fellowship
- **Collaborator** : Emmanuel Candès (Department of Mathematics and of Statistics, Stanford)

Motivation : Limits of resolution in imaging

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

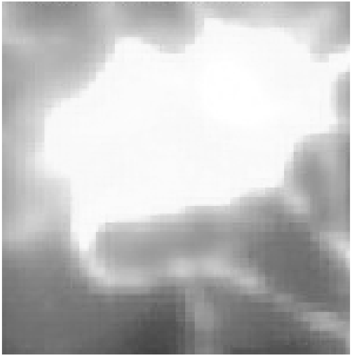


$\delta(t-\tau)$ optical system $h(t-\tau)$

Diffraction imposes a fundamental limit on the resolution of optical systems

Aim

Estimation from data that have limited resolution



- Microscopy
- Astronomy
- Electronic imaging
- Medical imaging
- Signal processing
- Radar
- Spectroscopy
- Geophysics
- ...

Super-resolution

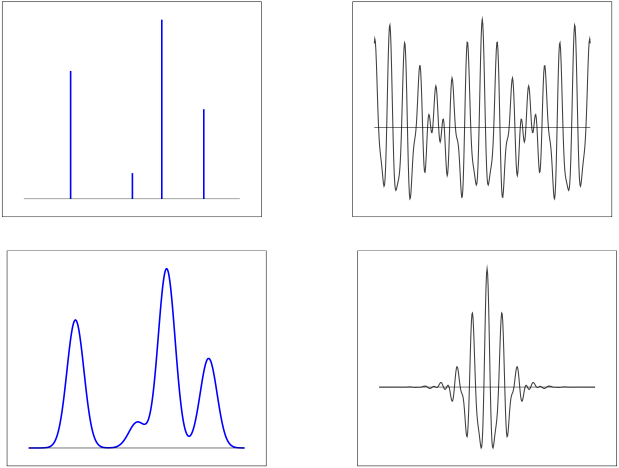
- **Optics** : Data-acquisition techniques to overcome the diffraction limit
- **Image processing** : Methods to upsample images onto a finer grid while preserving edges and hallucinating textures
- **This talk** : Signal estimation from low-pass measurements

Spatial Super-resolution

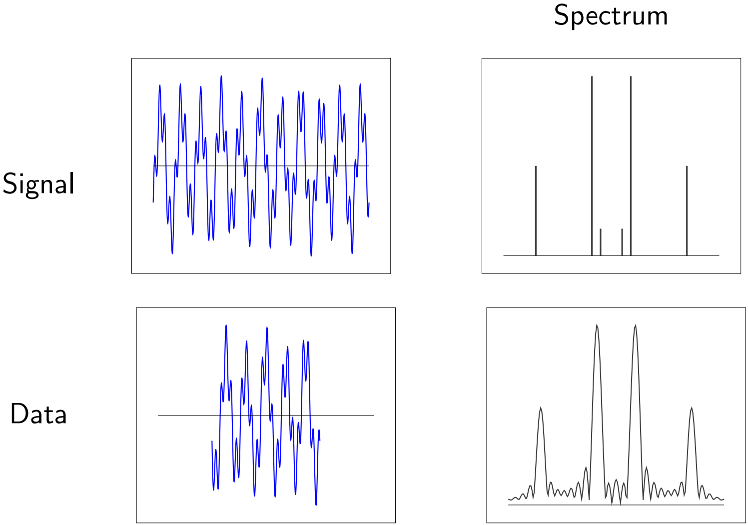
Spectrum

Signal

Data



Spectral Super-resolution



Point sources

- In many applications signals of interest are point sources :
  - Celestial bodies (astronomy)
  - Fluorescent molecules (microscopy)
  - Line spectra (spectroscopy, signal processing)

Point sources

- In many applications signals of interest are point sources :
  - Celestial bodies (astronomy)
  - Fluorescent molecules (microscopy)
  - Line spectra (spectroscopy, signal processing)
- Traditional approaches
  1. Fitting point-spread function to each source (matched filtering)
     - Sensitive to noise and high dynamic ranges
  2. Algorithms based on Prony's method : MUSIC, ESPRIT, ...
     - Parametric (number of sources must be known)
     - Extension to 2D is very computationally intensive
     - Strong assumptions on noise (Gaussian, white), signal and measurement model

Point sources

- In many applications signals of interest are point sources :
  - Celestial bodies (astronomy)
  - Fluorescent molecules (microscopy)
  - Line spectra (spectroscopy, signal processing)
- Traditional approaches
  1. Fitting point-spread function to each source (matched filtering)
     - Sensitive to noise and high dynamic ranges
  2. Algorithms based on Prony's method : MUSIC, ESPRIT, ...
     - Parametric (number of sources must be known)
     - Extension to 2D is very computationally intensive
     - Strong assumptions on noise (Gaussian, white), signal and measurement model
- This talk : Super-resolution via convex programming

Outline of the talk

Basic model

Estimation from noisy data

Basic model

Estimation from noisy data

Mathematical model

- Signal : superposition of Dirac measures with support T

$$x = \sum_j a_j \delta_{t_j} \qquad a_j \in \mathbb{C},\ t_j \in T \subset [0,1]$$

- Data : low-pass Fourier coefficients with cut-off frequency f_c

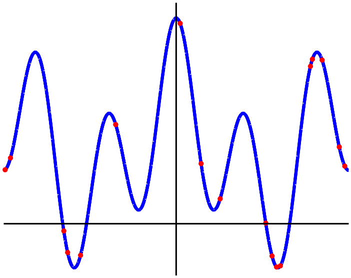
$$y = \mathcal{F}_c\, x$$

$$y(k) = \int_0^1 e^{-i2\pi kt} x\,(\mathrm{d}t) = \sum_j a_j e^{-i2\pi k t_j}, \quad k \in \mathbb{Z},\ |k| \leq f_c$$

Compressed sensing vs super-resolution

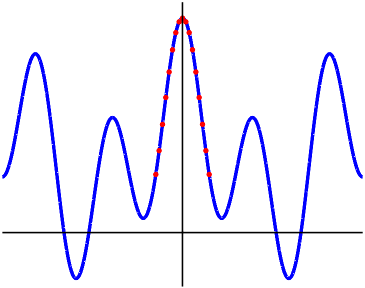
Estimation of sparse signals from undersampled measurements suggests connections to compressed sensing

Compressed sensing



spectrum interpolation

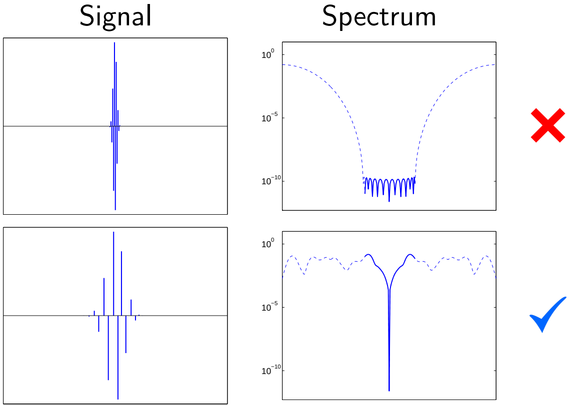
Super-resolution



spectrum extrapolation

Sparsity is not enough

Compressed sensing : measurement operator is well conditioned when acting upon any sparse signal (restricted isometry property)

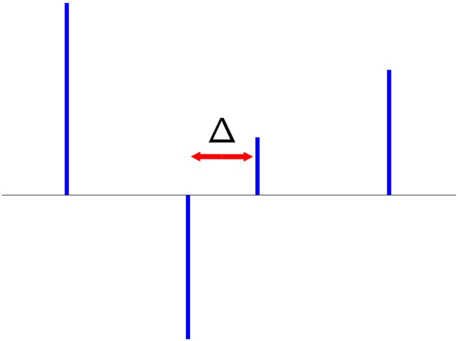


Not the case in super-resolution !

Minimum separation

Definition : The minimum separation Δ of a discrete set T is

$$\Delta = \inf_{(t,t') \in T \,:\, t \neq t'} |t - t'|$$



Total-variation norm

- Continuous counterpart of the ℓ_1 norm
- If $x = \sum_j a_j \delta_{t_j}$ then $||x||_{\text{TV}} = \sum_j |a_j|$
- Not the total variation of a piecewise-constant function

Total-variation norm

- Continuous counterpart of the ℓ_1 norm
- If $x = \sum_j a_j \delta_{t_j}$ then $||x||_{\text{TV}} = \sum_j |a_j|$
- Not the total variation of a piecewise-constant function
- **Formal definition :** For a complex measure ν

$$||\nu||_{\text{TV}} = \sup \sum_{j=1}^{\infty} |\nu(B_j)|,$$

(supremum over all finite partitions B_j of $[0, 1]$)

Estimation via convex programming

In a zero-noise limit, i.e. $y = \mathcal{F}_c\, x$, we solve

$$\min_{\tilde{x}} ||\tilde{x}||_{\mathrm{TV}} \quad \text{subject to} \quad \mathcal{F}_c\, \tilde{x} = y,$$

over all finite complex measures $\tilde{x}$ supported on $[0, 1]$

Estimation via convex programming

In a zero-noise limit, i.e. $y = \mathcal{F}_c \, x$, we solve

$$\min_{\tilde{x}} ||\tilde{x}||_{\text{TV}} \quad \text{subject to} \quad \mathcal{F}_c \, \tilde{x} = y,$$

over all finite complex measures $\tilde{x}$ supported on $[0, 1]$

Theorem [Candès, F. '12]

If the minimum separation of the signal support T obeys

$$\Delta \geq 2 \, / f_c := 2 \, \lambda_c,$$

then recovery is exact in 1D

Nonparametric approach (no previous knowledge of the number of spikes)

Estimation via convex programming

In a zero-noise limit, i.e. $y = \mathcal{F}_c\, x$, we solve

$$\min_{\tilde{x}} ||\tilde{x}||_{\mathrm{TV}} \quad \text{subject to} \quad \mathcal{F}_c\, \tilde{x} = y,$$

over all finite complex measures $\tilde{x}$ supported on $[0,1]$

Theorem [Candès, F. '12]

If the minimum separation of the signal support T obeys

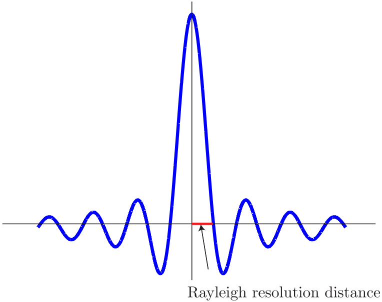
$$\Delta \geq 2.38\,/f_c := 2.38\,\lambda_c,$$

then recovery is exact in 2D

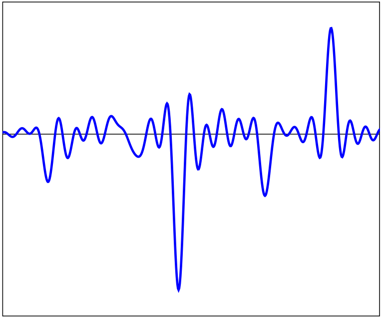
Nonparametric approach (no previous knowledge of the number of spikes)

Minimum separation

Point-spread function



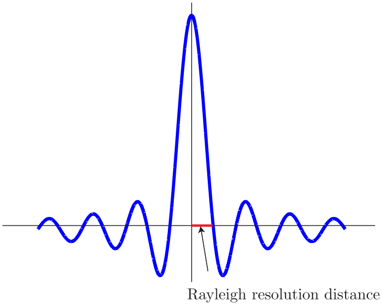
$\Delta = 1.4\,\lambda_c$



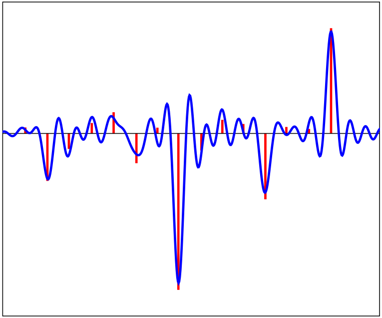
$\lambda_c/2$ is the Rayleigh resolution limit

Minimum separation

Point-spread function



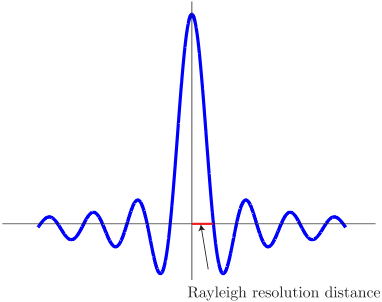
$\Delta = 1.4\,\lambda_c$



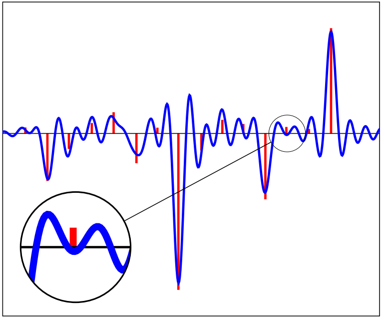
$\lambda_c/2$ is the Rayleigh resolution limit

Minimum separation

Point-spread function



$\Delta = 1.4\,\lambda_c$



$\lambda_c/2$ is the Rayleigh resolution limit

Sketch of proof : Dual certificate

A sufficient condition for

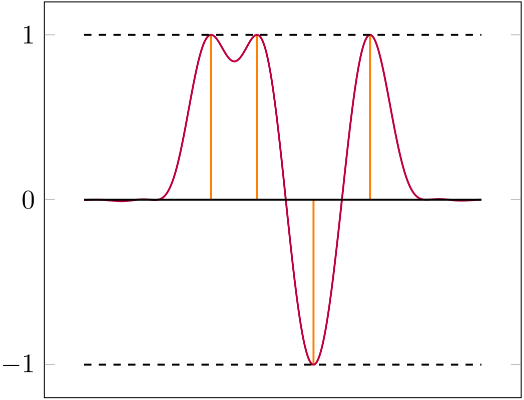
$$x = \sum_{j \in T} a_j \delta_{t_j} = \sum_{j \in T} |a_j| \, e^{i\phi_j} \delta_{t_j}$$

to be the unique solution is that there exists q such that

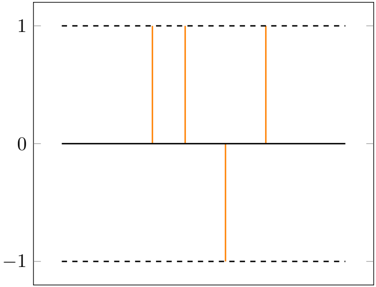
1. $q(t) = \sum_{k=-f_c}^{f_c} b_k e^{i2\pi kt}$ (low pass polynomial)
2. $q(t_j) = e^{i\phi_j}$, $t_j \in T$ (interpolates the sign of the signal on T)
3. $|q(t)| < 1$, $t \in T^c$

q is a subgradient of the TV norm at the signal x that is orthogonal to the null space of the measurement operator

Sketch of proof : Dual certificate



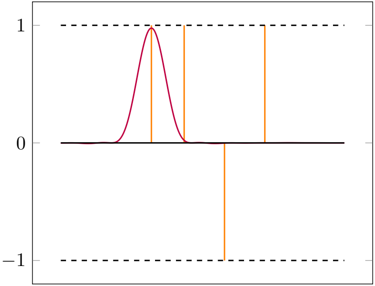
Sketch of proof : Construction by interpolation



1st idea : Interpolation with a low-frequency fast-decaying kernel K

$$q(t) = \sum_{t_j \in T} \alpha_j \, K(t - t_j),$$

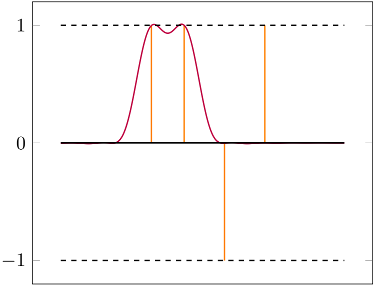
Sketch of proof : Construction by interpolation



1st idea : Interpolation with a low-frequency fast-decaying kernel K

$$q(t) = \sum_{t_j \in T} \alpha_j \, K(t - t_j),$$

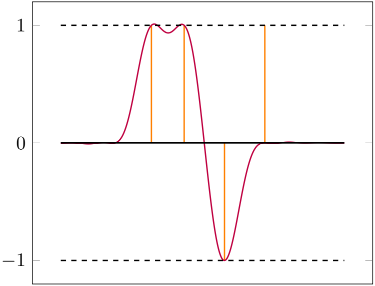
Sketch of proof : Construction by interpolation



1st idea : Interpolation with a low-frequency fast-decaying kernel K

$$q(t) = \sum_{t_j \in T} \alpha_j \, K(t - t_j),$$

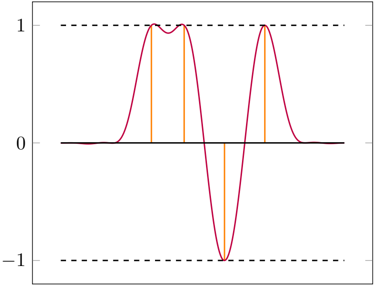
Sketch of proof : Construction by interpolation



1st idea : Interpolation with a low-frequency fast-decaying kernel K

$$q(t) = \sum_{t_j \in T} \alpha_j \, K(t - t_j),$$

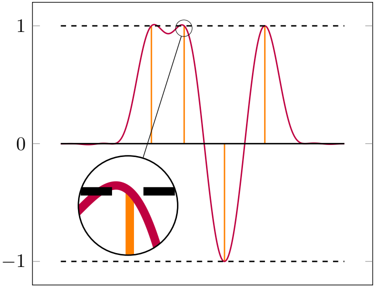
Sketch of proof : Construction by interpolation



1st idea : Interpolation with a low-frequency fast-decaying kernel K

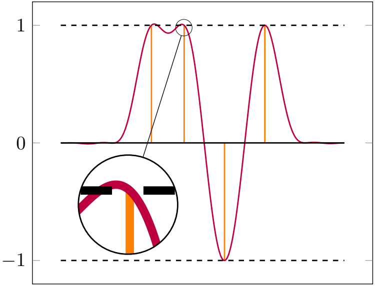
$$q(t) = \sum_{t_j \in T} \alpha_j \, K(t - t_j),$$

Sketch of proof : Construction by interpolation



Problem : Magnitude of polynomial locally exceeds 1

Sketch of proof : Construction by interpolation

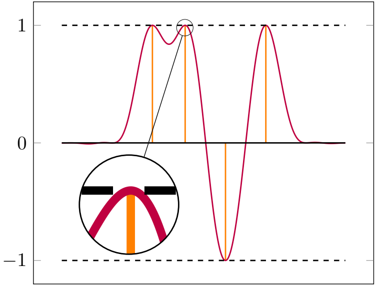


Problem : Magnitude of polynomial locally exceeds 1

Solution : Add correction term and force $q'(t_k) = 0$ for all $t_k \in T$

$$q(t) = \sum_{t_j \in T} \alpha_j K(t - t_j) + \beta_j K'(t - t_j)$$

Sketch of proof : Construction by interpolation



Problem : Magnitude of polynomial locally exceeds 1

Solution : Add correction term and force $q'(t_k) = 0$ for all $t_k \in T$

$$q(t) = \sum_{t_j \in T} \alpha_j K(t - t_j) + \beta_j K'(t - t_j)$$

Basic model

Estimation from noisy data

Estimation from noisy data

We assume additive noise with norm bounded by δ

$$y = \mathcal{F}_c x + z$$

Our estimator is the solution to

$$\min_{\tilde{x}} ||\tilde{x}||_{\mathrm{TV}} \quad \text{subject to} \quad ||\mathcal{F}_c\, \tilde{x} - y||_2 \leq \delta,$$

Estimation from noisy data

We assume additive noise with norm bounded by δ

$$y = \mathcal{F}_c x + z$$

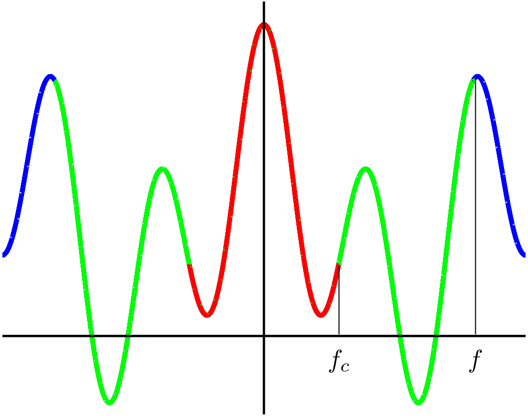
Our estimator is the solution to

$$\min_{\tilde{x}} ||\tilde{x}||_{\mathrm{TV}} \quad \text{subject to} \quad ||\mathcal{F}_c\, \tilde{x} - y||_2 \leq \delta,$$

Metrics to quantify estimation accuracy :

1. Approximation error at a higher resolution
2. Support-detection error

Super-resolution factor : spectral viewpoint

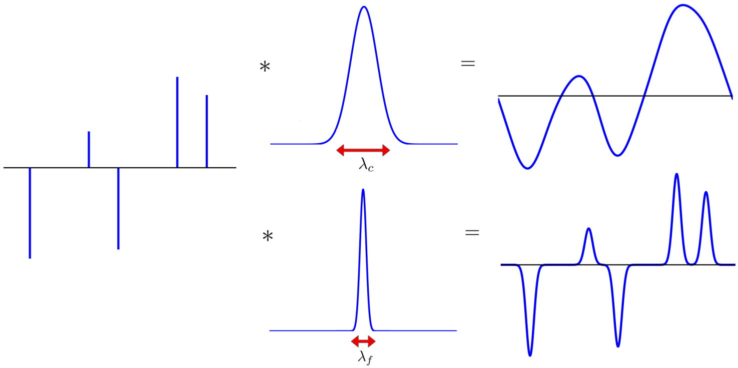


Super-resolution factor

$$\mathsf{SRF} = \frac{f}{f_c}$$

Super-resolution factor : spatial viewpoint

Signal at a resolution λ : convolution with a kernel ϕ_λ of width λ



Super-resolution factor

$$\mathrm{SRF} = \frac{\lambda_c}{\lambda_f}$$

Approximation at a higher resolution

At the resolution of the measurements

$$||\phi_{\lambda_c} * (x_{\text{est}} - x)||_{L_1} \leq \delta$$

How does the estimate degrade at a higher resolution ?

Approximation at a higher resolution

At the resolution of the measurements

$$||\phi_{\lambda_c} * (x_{\text{est}} - x)||_{L_1} \leq \delta$$

How does the estimate degrade at a higher resolution ?

Theorem [Candès, F. 2012]

If $\Delta \geq 2\,/f_c$ then the solution $\hat{x}$ to

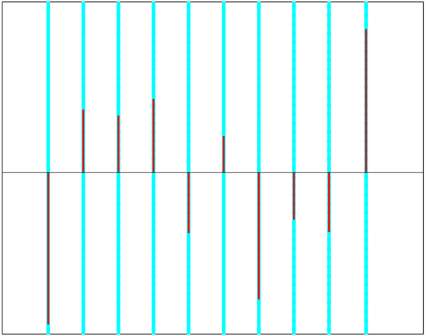
$$\min_{\tilde{x}} ||\tilde{x}||_{\text{TV}} \quad \text{subject to} \quad ||\mathcal{F}_c\, \tilde{x} - y||_2 \leq \delta,$$

$$\text{satisfies} \quad \left|\left|\phi_{\lambda_f} * (\hat{x} - x)\right|\right|_{L_1} \lesssim \text{SRF}^2\, \delta$$

Some comments

- Non-asymptotic results, whereas most theory for Prony-based methods is asymptotic (convergence of sample autocorrelation matrices)
- Usual proof techniques from high-dimensional statistics do not apply
  1. Conditions (restricted-isometry property, restricted-eigenvalue condition, etc.) do not hold
  2. Estimation takes place over a continuous domain
- Proofs are based on generalizations of the dual certificate for the noiseless problem

Sketch of proof



We partition the unit interval into

$$\text{NEAR} := \left\{ t \;:\; \min_{t_j \in T} |t - t_j| \leq 0.1\,\lambda_f \right\} \qquad \text{FAR} := \text{NEAR}^c$$

Sketch of proof

$$e := \hat{x} - x$$

We establish an approximate null-space property to bound

$$||e_{\text{FAR}}||_{\text{TV}} \lesssim \text{SRF}^2\,\delta$$

Controlling $\quad \left|\left|\left(e * \phi_{\lambda_f}\right)_{\text{NEAR}}\right|\right|_{L_1} \lesssim \text{SRF}^2\,\delta \quad$ is more challenging

Sketch of proof : $\left|\left|\left(e * \phi_{\lambda_f}\right)_{\text{NEAR}}\right|\right|_{L_1} \lesssim \text{SRF}^2\,\delta$

We apply a Taylor expansion at each $t_j \in T$

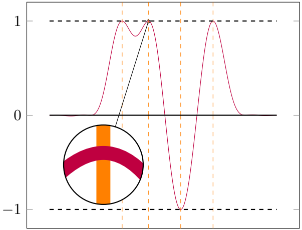
$$e * \phi_{\lambda_f}(t) \approx e * \phi_{\lambda_f}(t_j) + \left(e * \phi_{\lambda_f}\right)'(t_j)\,(t - t_j)$$

This yields the bound

$$\left|\left|\left(e * \phi_{\lambda_f}\right)_{\text{NEAR}}\right|\right|_{L_1} \leq \sum_{t_j \in T} \left|\int_{|t-t_j| \leq 0.1\,\lambda_f} e\,(\mathrm{d}t)\right| + \frac{1}{\lambda_f}\left|\int_{|t-t_j| \leq 0.1\,\lambda_f} (t - t_j)\,e\,(\mathrm{d}t)\right|$$

To complete the proof we show that both quantities $\lesssim \text{SRF}^2\,\delta$

Sketch of proof : $\|(e * \phi_{\lambda_f})_{\text{NEAR}}\|_{L_1} \lesssim \text{SRF}^2\,\delta$



Build low-pass polynomial q almost constant in NEAR so

$$\sum_{t_j \in T} \left| \int_{|t-t_j| \leq 0.1\,\lambda_f} e\,(\mathrm{d}t) \right| \approx \langle q_{\text{NEAR}}, e_{\text{NEAR}} \rangle$$

$$\leq |\langle q, e \rangle| + |\langle q_{\text{FAR}}, e_{\text{FAR}} \rangle|$$

Sketch of proof : $\quad ||(e * \phi_{\lambda_f})_{\text{NEAR}}||_{L_1} \lesssim \text{SRF}^2\, \delta$

Because q is low-pass and both x and $\hat{x}$ are feasible

$$\begin{aligned} |\langle q, e\rangle| &\leq ||q||_2\, ||\mathcal{F}_c e||_2 \\ &\leq ||\mathcal{F}_c x - y||_2 + ||y - \mathcal{F}_c \hat{x}||_2 \\ &\leq 2\delta \end{aligned}$$

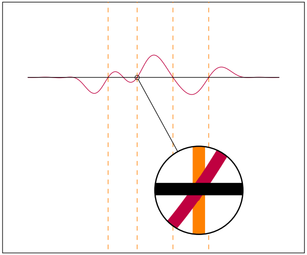
We can show $||q||_\infty \leq 1$ so

$$|\langle q_{\text{FAR}}, e_{\text{FAR}}\rangle| \leq ||q||_\infty\, ||e_{\text{FAR}}||_{\text{TV}} \lesssim \text{SRF}^2\, \delta$$

As a result

$$\begin{aligned} \sum_{t_j \in T} \left| \int_{|t - t_j| \leq 0.1\, \lambda_f} e\,(\mathrm{d}t) \right| &\leq |\langle q, e\rangle| + |\langle q_{\text{FAR}}, e_{\text{FAR}}\rangle| \\ &\lesssim \text{SRF}^2\, \delta \end{aligned}$$

Sketch of proof : $\|(e * \phi_{\lambda_f})_{\text{NEAR}}\|_{L_1} \lesssim \text{SRF}^2\,\delta$



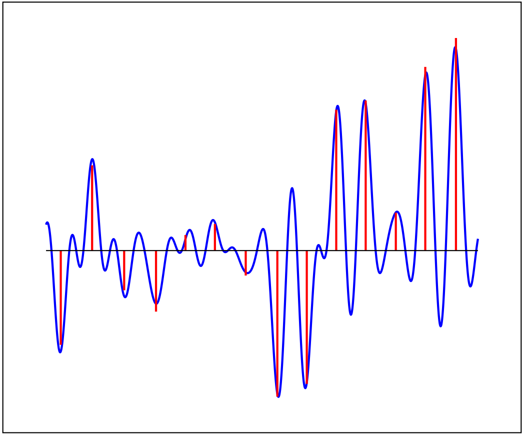
Build low-pass polynomial q almost linear in NEAR and $\|q\|_\infty \lesssim \lambda_c$

$$\sum_{t_j \in T} \frac{1}{\lambda_f} \left| \int_{|t-t_j| \le 0.1\,\lambda_f} (t - t_j)\, e\,(\mathrm{d}t) \right| \approx \frac{1}{\lambda_f} \langle q_{\text{NEAR}}, e_{\text{NEAR}} \rangle$$

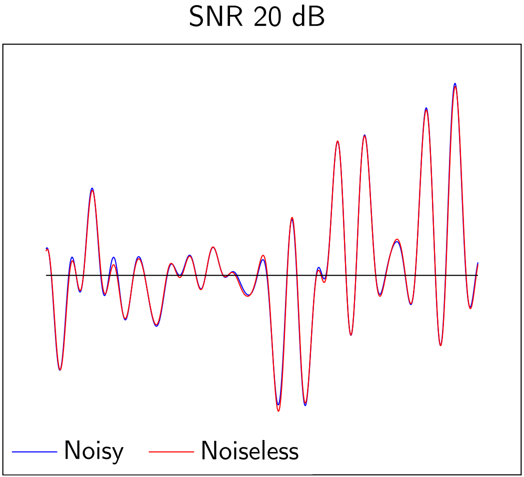
$$\lesssim \text{SRF}^2\,\delta$$

Example

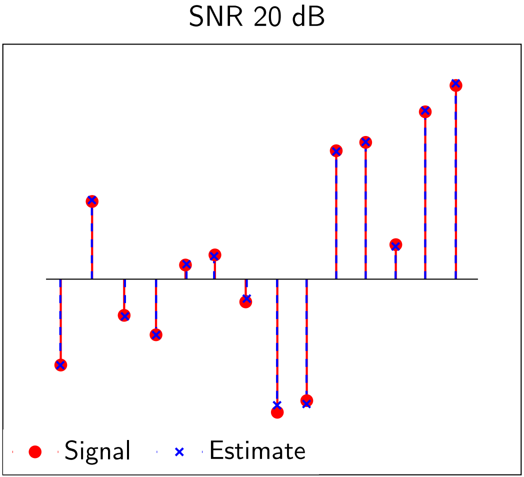
Minimum separation : $1.5\,\lambda_c$



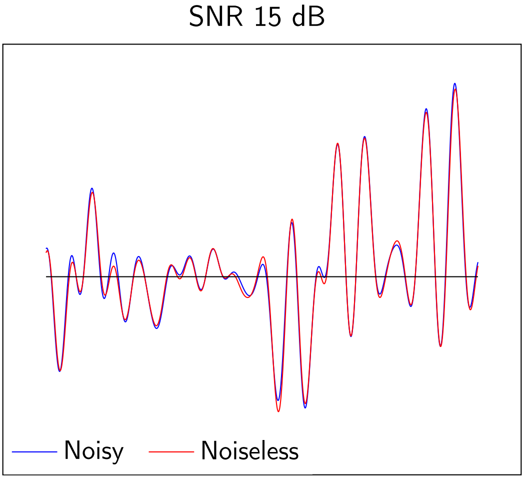
Example



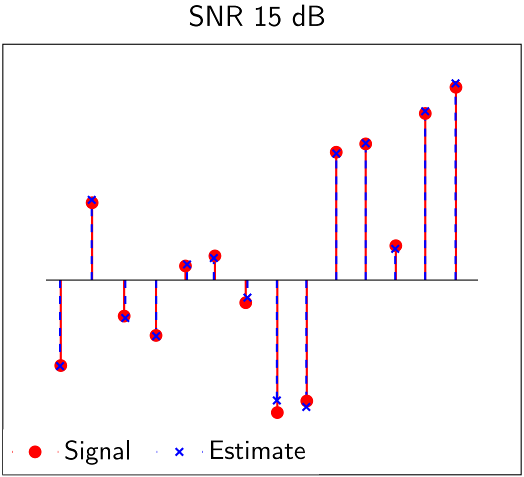
Example



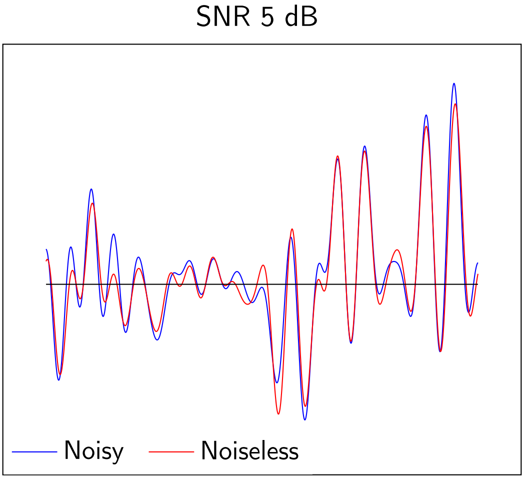
Example



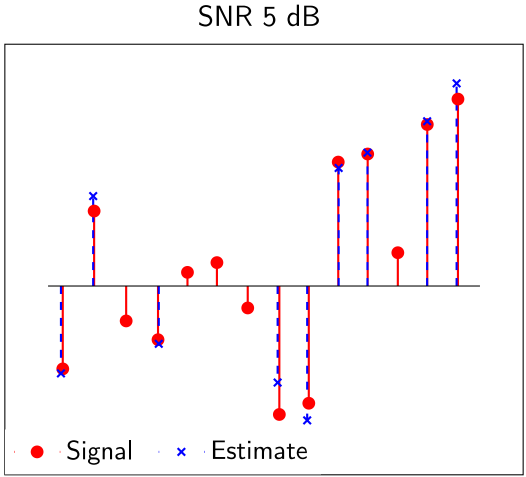
Example



Example



Example



Support-detection accuracy

- Original support : T
- Estimated support : $\widehat{T}$

Theorem [F. 2013]

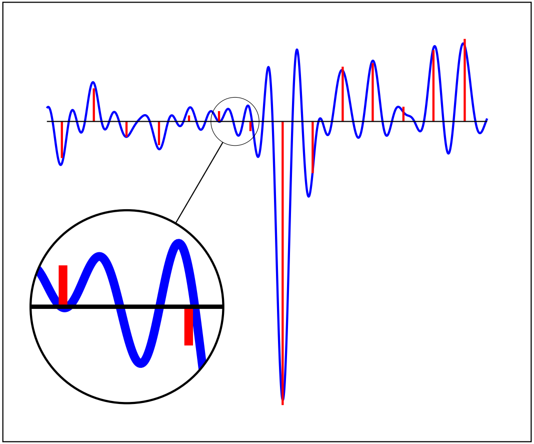
For any $t_i \in T$, if $|a_i| > C_1\delta$ there exists $\hat{t}_i \in \widehat{T}$ such that

$$\left|t_i - \hat{t}_i\right| \leq \frac{1}{f_c}\sqrt{\frac{C_2\delta}{|a_i| - C_1\delta}}$$

No dependence on the amplitude of the signal at other locations

Consequence

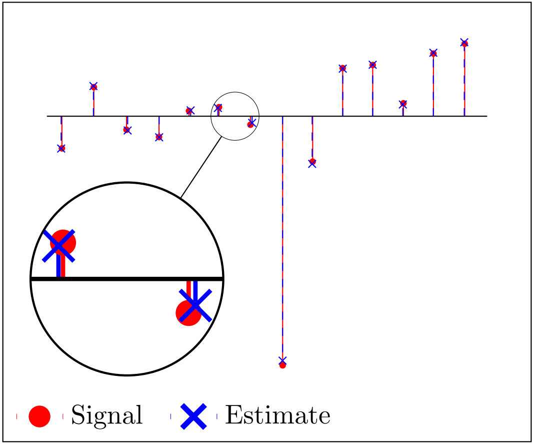
Robustness of the algorithm to high dynamic ranges



SNR 20 dB (15 dB without the large spike)

Consequence

Robustness of the algorithm to high dynamic ranges



SNR 20 dB (15 dB without the large spike)

Conclusion

Convex programming is a powerful tool for estimation from low-res data :

- Precise theoretical analysis
- Non-asymptotic stability guarantees

Conclusion

Convex programming is a powerful tool for estimation from low-res data :

- Precise theoretical analysis
- Non-asymptotic stability guarantees

Things I haven't talked about :

- In 1D, infinite-dimensional problem can be solved *without discretizing*
- Other noise and signal models

Conclusion

Convex programming is a powerful tool for estimation from low-res data :

- Precise theoretical analysis
- Non-asymptotic stability guarantees

Things I haven't talked about :

- In 1D, infinite-dimensional problem can be solved *without discretizing*
- Other noise and signal models

Lots of work to do :

- Poisson noise
- Super-resolution of 2D curves
- *Blind deconvolution* : joint estimation of signal + point-spread function

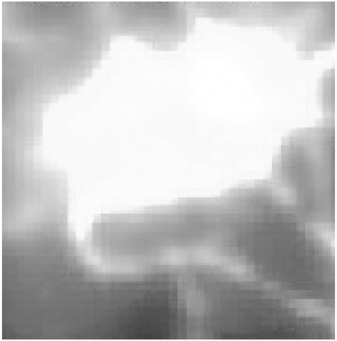
Related work

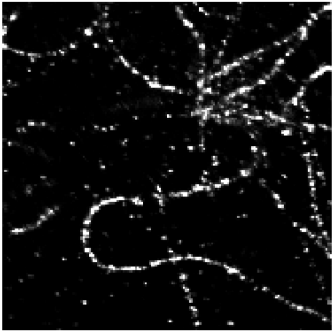
- Deconvolution in seismography [Claerbout, Muir '73], [Levy, Fullagar '81], [Santosa, Symes '86]
- Modulus of continuity of super-resolution [Donoho 1992]
- Line-spectra estimation with missing data [Tang *et al* 2012], denoising via convex optimization [Tang *et al* 2013]
- Other work on super-resolution of spikes via convex programming [Azais *et al* 2012, Duval and Peyré 2013]

For more details

- **Towards a mathematical theory of super-resolution**. E. J. Candès and C. Fernandez-Granda. *Communications on Pure and Applied Math* **67**(6), 906-956.
- **Super-resolution from noisy data**. E. J. Candès and C. Fernandez-Granda. *Journal of Fourier Analysis and Applications* **19** (6), 1229-1254.
- **Support detection in super-resolution**. C. Fernandez-Granda. *Proceedings of SampTA 2013*, 145-148.

Thank you





Figures courtesy of V. Morgenshtern, Stanford

Practical implementation

- **Primal problem** :

$$\min_{\tilde{x}} ||\tilde{x}||_{\mathsf{TV}} \quad \text{subject to} \quad \mathcal{F}_c\,\tilde{x} = y$$

Infinite-dimensional variable $\tilde{x}$ (measure in $[0,1]$)

First option : Discretizing + ℓ_1-norm minimization

Practical implementation

- **Primal problem** :

$$\min_{\tilde{x}} ||\tilde{x}||_{\mathrm{TV}} \quad \text{subject to} \quad \mathcal{F}_c\, \tilde{x} = y$$

Infinite-dimensional variable $\tilde{x}$ (measure in $[0,1]$)

First option : Discretizing + ℓ_1-norm minimization

- **Dual problem** :

$$\max_{\tilde{u} \in \mathbb{C}^n} \mathrm{Re}\,[y^* \tilde{u}] \quad \text{subject to} \quad ||\mathcal{F}_c^*\, \tilde{u}||_\infty \leq 1, \quad n := 2f_c + 1$$

Finite-dimensional variable $\tilde{u}$, but infinite-dimensional constraint

$$\mathcal{F}_c^*\, \tilde{u} = \sum_{k \leq |f_c|} \tilde{u}_k e^{i2\pi kt}$$

Second option : Solving the dual problem

Lemma : Semidefinite representation

The Fejér-Riesz Theorem and the semidefinite representation of polynomial sums of squares imply that

$$||\mathcal{F}_c^* \, \tilde{u}||_\infty \leq 1$$

is equivalent to

There exists a Hermitian matrix $Q \in \mathbb{C}^{n \times n}$ such that

$$\begin{bmatrix} Q & \tilde{u} \\ \tilde{u}^* & 1 \end{bmatrix} \succeq 0, \qquad \sum_{i=1}^{n-j} Q_{i,i+j} = \begin{cases} 1, & j = 0, \\ 0, & j = 1, 2, \ldots, n-1. \end{cases}$$

Consequence : The dual problem is a tractable semidefinite program

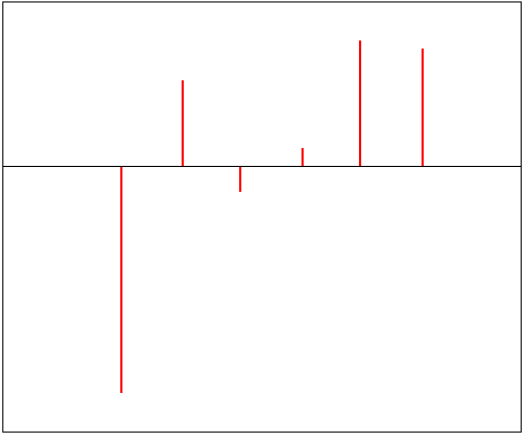
Support-locating polynomial

How do we obtain an estimator from the dual solution ?

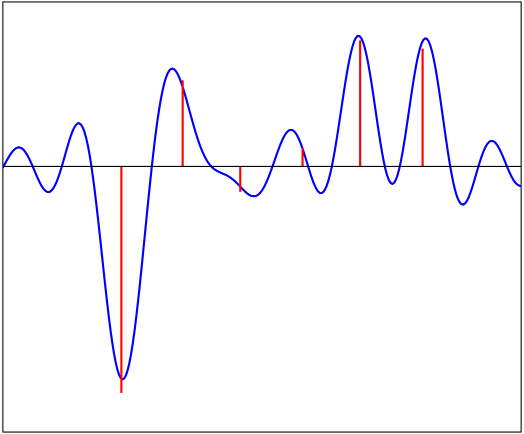
Dual solution vector : Fourier coefficients of low-pass polynomial that interpolates the sign of the primal solution (follows from strong duality)

Idea : Use the polynomial to locate the support of the signal

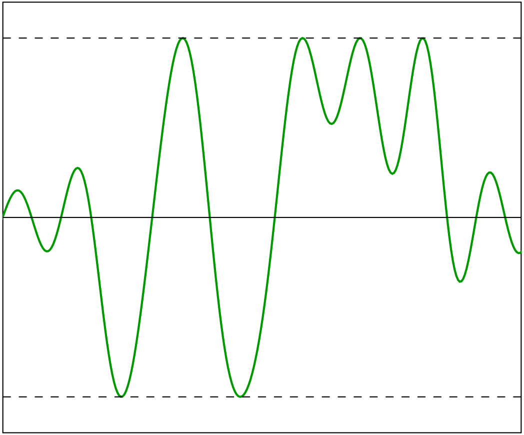
Super-resolution via semidefinite programming



Super-resolution via semidefinite programming

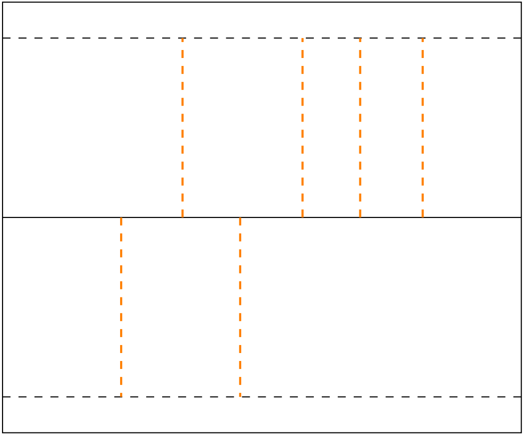


Super-resolution via semidefinite programming



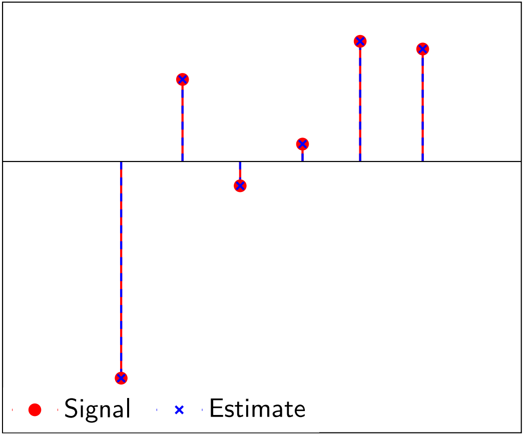
1. Solve semidefinite program to obtain dual solution

Super-resolution via semidefinite programming



2. Locate points at which corresponding polynomial has unit magnitude

Super-resolution via semidefinite programming



3. Estimate amplitudes via least squares