Robust Super-resolution via Convex Programming

Carlos Fernandez-Granda

www.stanford.edu/~cfgranda/

International Conference on Continuous Optimization

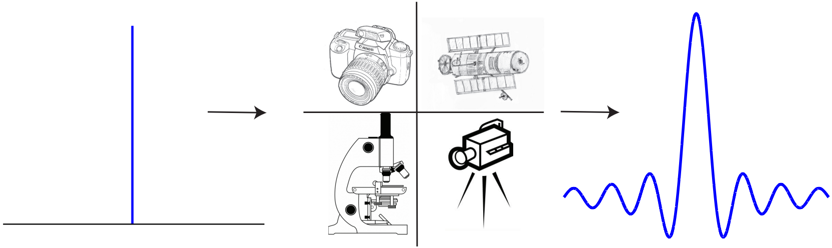
7/31/2013

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**)

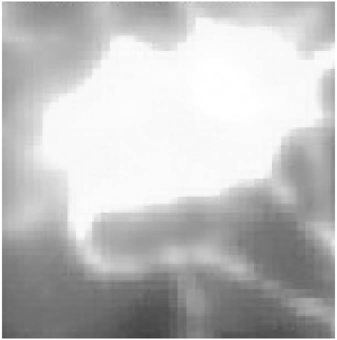


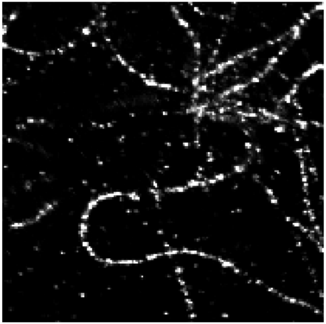
$\delta(t-\tau)$ optical system $h(t-\tau)$

Diffraction imposes a fundamental limit on the resolution of optical systems

Motivation

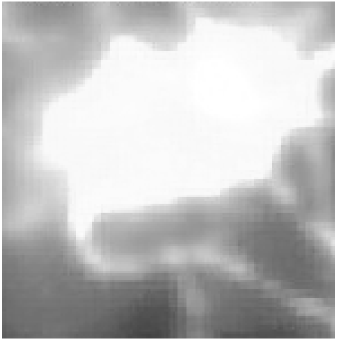
Similar problems arise in electronic imaging, signal processing, radar, spectroscopy, medical imaging, astronomy, geophysics, etc.

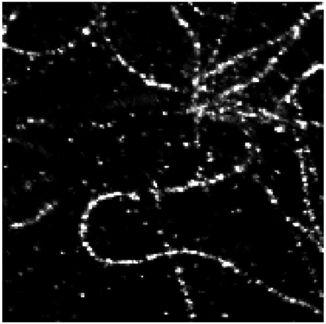




Motivation

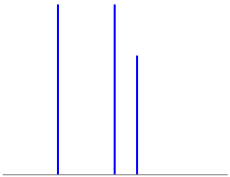
Similar problems arise in electronic imaging, signal processing, radar, spectroscopy, medical imaging, astronomy, geophysics, etc.

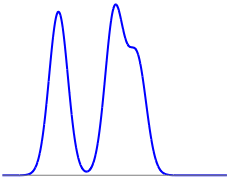




Signals of interest are often point sources : celestial bodies (astronomy), line spectra (signal processing), molecules (fluorescence microscopy), ...

Super-resolution





Aim : estimating fine-scale structure from low-resolution data

Super-resolution

Aim : estimating fine-scale structure from low-resolution data

Equivalently, extrapolating the high end of the spectrum

Mathematical model

- Signal : superposition of Dirac measures with support T

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

Mathematical model

- Signal : superposition of Dirac measures with support T

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

- Measurements : low-pass filtering with cut-off frequency f_c

$$y = \mathcal{F}_c\, x \quad \text{(vector of low-pass Fourier coefficients)}$$

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

Equivalent problem : line-spectra estimation

Swapping time and frequency

- Signal : superposition of sinusoids

$$x(t) = \sum_j a_j e^{i2\pi\omega_j t} \qquad a_j \in \mathbb{C},\ \omega_j \in T \subset [0,1]$$

Equivalent problem : line-spectra estimation

Swapping time and frequency

- Signal : superposition of sinusoids

$$x(t) = \sum_j a_j e^{i2\pi\omega_j t} \qquad a_j \in \mathbb{C},\ \omega_j \in T \subset [0,1]$$

- Measurements : equispaced samples

$$x(1), x(2), x(3), \ldots x(n)$$

Equivalent problem : line-spectra estimation

Swapping time and frequency

- Signal : superposition of sinusoids

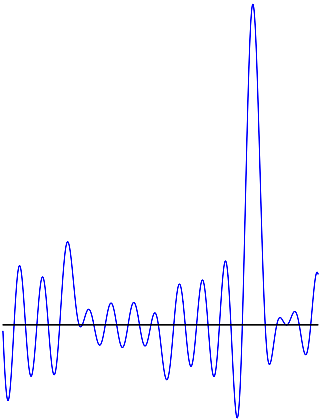
$$x(t) = \sum_j a_j e^{i2\pi\omega_j t} \qquad a_j \in \mathbb{C},\ \omega_j \in T \subset [0,1]$$

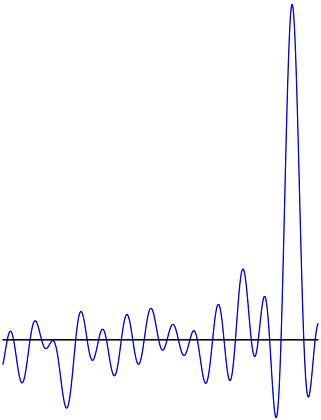
- Measurements : equispaced samples

$$x(1), x(2), x(3), \dots x(n)$$

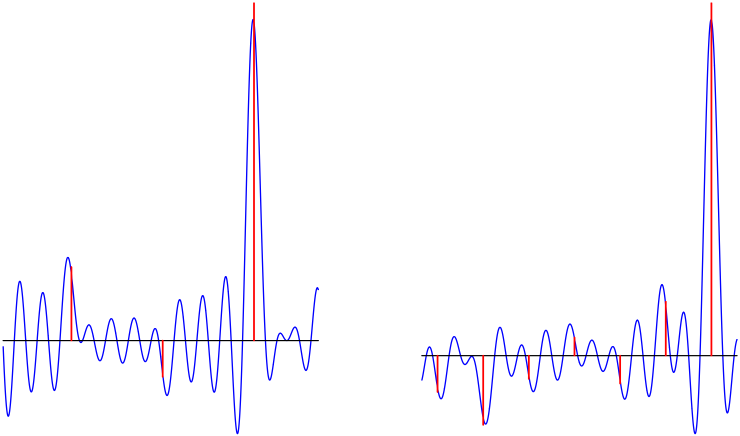
- Classical problem in signal processing

Can you find the spikes?



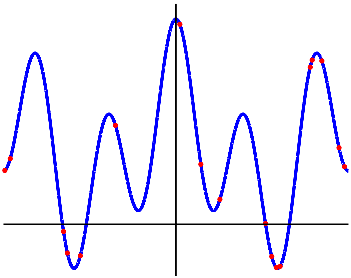


Can you find the spikes?



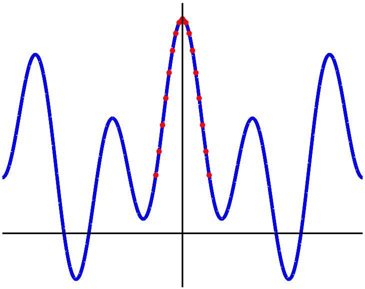
Compressed sensing vs super-resolution

Compressed sensing



spectrum interpolation

Super-resolution



spectrum extrapolation

Outline of the talk

Theory

Proof (sketch)

Robustness to noise

Algorithms

Theory

Proof (sketch)

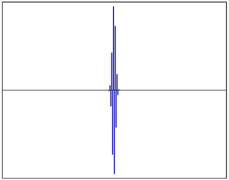
Robustness to noise

Algorithms

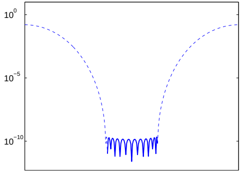
Sparsity is not enough

Sparse signals can be almost completely destroyed by low-pass filtering (shown rigorously by Slepian in the 1970s)

Signal



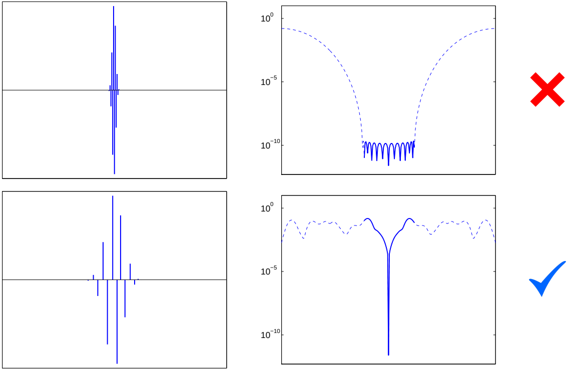
Spectrum



Sparsity is not enough

Sparse signals can be almost completely destroyed by low-pass filtering (shown rigorously by Slepian in the 1970s)

Signal Spectrum

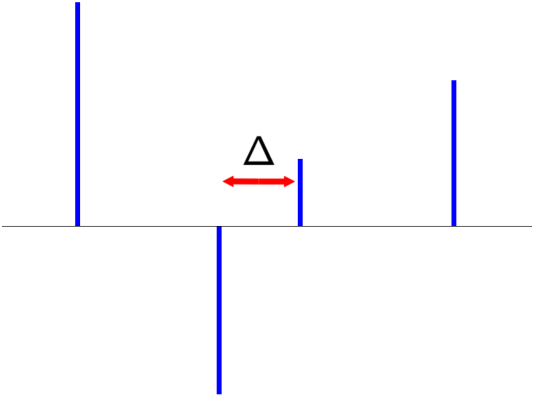


Stable reconstruction is only possible for signals with non-clustered supports

Minimum separation

To exclude highly-clustered signals from our model, we control the minimum separation Δ of the support T

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



Total-variation norm

- Continuous counterpart of the ℓ_1 norm

Total-variation norm

- Continuous counterpart of the ℓ_1 norm
- If $x = \sum_j a_j \delta_{t_j}$ then $||x||_{\text{TV}} = \sum_j |a_j|$

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 $[0, 1]$)

Recovery via convex programming

In the absence of noise, i.e. if $y = \mathcal{F}_c\, x$, 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 $[0, 1]$

Recovery via convex programming

In the absence of noise, i.e. if $y = \mathcal{F}_c\, x$, 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. 2012]

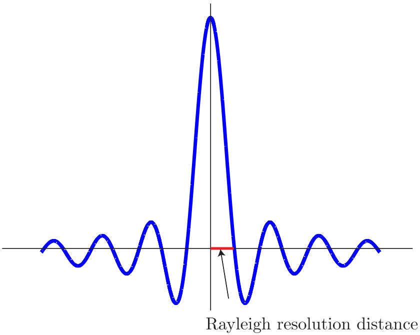
If the minimum separation of the signal support T obeys

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

then recovery is exact

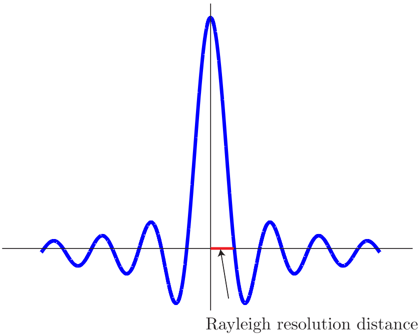
Minimum-distance condition

- $\lambda_c/2$ is the Rayleigh resolution limit (half-width of measurement filter)



Minimum-distance condition

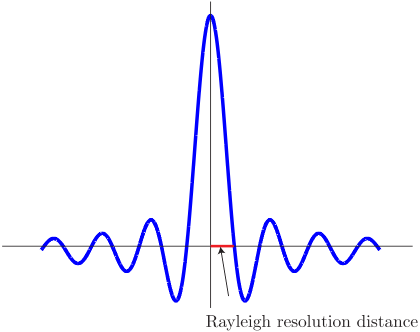
- $\lambda_c/2$ is the Rayleigh resolution limit (half-width of measurement filter)



- Numerical simulations show that TV-norm minimization fails if $\Delta < \lambda_c$

Minimum-distance condition

- $\lambda_c/2$ is the Rayleigh resolution limit (half-width of measurement filter)



- Numerical simulations show that TV-norm minimization fails if $\Delta < \lambda_c$
- If $\Delta < \lambda_c/2$ some signals are *almost* in the nullspace of the measurement operator (no method can achieve stable estimation)

Higher dimensions

- Signal : superposition of point sources (delta measures) in 2D
- Measurements : low-pass 2D Fourier coefficients

Higher dimensions

- Signal : superposition of point sources (delta measures) in 2D
- Measurements : low-pass 2D Fourier coefficients

Theorem [Candès, F. 2012]

TV-norm minimization yields exact recovery if

$$\Delta \geq 2.38\, \lambda_c$$

Higher dimensions

- Signal : superposition of point sources (delta measures) in 2D
- Measurements : low-pass 2D Fourier coefficients

Theorem [Candès, F. 2012]

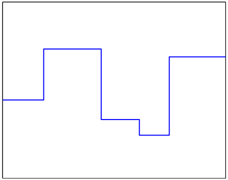
TV-norm minimization yields exact recovery if

$$\Delta \geq 2.38 \, \lambda_c$$

In dimension d, $\Delta \geq C_d \, \lambda_c$, where C_d only depends on d

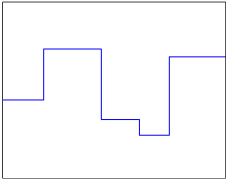
Extensions

- Signal : piecewise-constant function
- Measurements : low-pass Fourier coefficients



Extensions

- Signal : piecewise-constant function
- Measurements : low-pass Fourier coefficients



Corollary

$$\text{Solving } \min \|\tilde{x}^{(1)}\|_{\text{TV}} \quad \text{subject to} \quad \mathcal{F}_c \tilde{x} = y$$

$$\text{yields exact recovery if } \Delta \geq 2\,\lambda_c$$

Similar result for cont. differentiable piecewise-smooth functions

Theory

Proof (sketch)

Robustness to noise

Algorithms

Certificate of optimality

Proposition

For any support $T \subset [0,1]$ satisfying $\Delta \geq 2\,\lambda_c$ and every vector $v \in \mathbb{C}^{|T|}$ such that

$$|v_j| = 1 \quad \text{for all } 1 \leq j \leq |T|$$

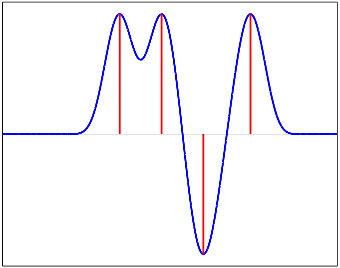
there exists a low-frequency trigonometric polynomial

$$q(t) = \sum_{k=-f_c}^{f_c} c_k e^{i2\pi kt}$$

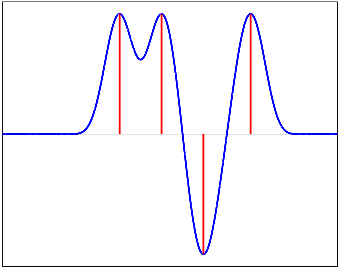
obeying

$$\begin{cases} q(t_j) = v_j, & t_j \in T, \\ |q(t)| < 1, & t \in [0,1] \setminus T. \end{cases}$$

Certificate of optimality



Certificate of optimality



Lemma

The proposition implies that $x = \sum_{t_j \in T} a_j \delta_{t_j}$ is the unique solution to

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

Certificate of optimality : Proof

Take any feasible signal $x' = x + h$ and decompose $h = h_T + h_{T^c}$, where

$$h_T = \sum_{t_j \in T} b_j \delta_{t_j} = \sum_{t_j \in T} e^{i\phi_j} |b_j| \delta_{t_j} \quad \text{(proof is easily generalized)}$$

(1) $v_j = e^{-i\phi_j}$ yields q such that $\langle q_T, h_T \rangle = \sum_{t_j \in T} |b_j| = ||h_T||_{\text{TV}}$

(2) x' is feasible so $\langle q, h \rangle = \langle q_T, h_T \rangle + \langle q_{T^c}, h_{T^c} \rangle = 0$

$$\begin{aligned} ||h_{T^c}||_{\text{TV}} &> |\langle q_{T^c}, h_{T^c} \rangle| && \text{by Hölder's inequality and } ||q_{T^c}||_{\infty} < 1 \\ &= |\langle q_T, h_T \rangle| && \text{by (2)} \\ &= ||h_T||_{\text{TV}} && \text{by (1)} \end{aligned}$$

Certificate of optimality : Proof

Take any feasible signal $x' = x + h$ and decompose $h = h_T + h_{T^c}$, where

$$h_T = \sum_{t_j \in T} b_j \delta_{t_j} = \sum_{t_j \in T} e^{i\phi_j} |b_j| \delta_{t_j} \quad \text{(proof is easily generalized)}$$

(1) $v_j = e^{-i\phi_j}$ yields q such that $\langle q_T, h_T \rangle = \sum_{t_j \in T} |b_j| = ||h_T||_{\text{TV}}$

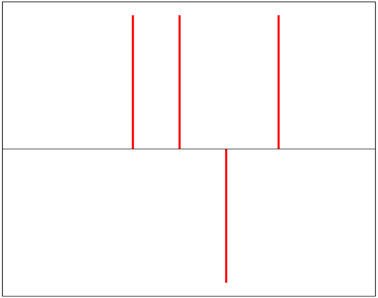
(2) x' is feasible so $\langle q, h \rangle = \langle q_T, h_T \rangle + \langle q_{T^c}, h_{T^c} \rangle = 0$

$$\begin{aligned} ||h_{T^c}||_{\text{TV}} &> |\langle q_{T^c}, h_{T^c} \rangle| && \text{by Hölder's inequality and } ||q_{T^c}||_\infty < 1 \\ &= |\langle q_T, h_T \rangle| && \text{by (2)} \\ &= ||h_T||_{\text{TV}} && \text{by (1)} \end{aligned}$$

By this null-space condition and the fact that the TV norm is separable,

$$\begin{aligned} \left|\left|x'\right|\right|_{\text{TV}} &= ||x + h_T||_{\text{TV}} + ||h_{T^c}||_{\text{TV}} \geq ||x||_{\text{TV}} + ||h_{T^c}||_{\text{TV}} - ||h_T||_{\text{TV}} \\ &> ||x||_{\text{TV}} \end{aligned}$$

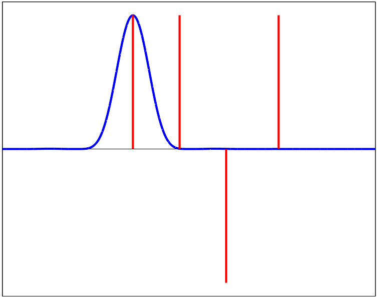
Construction of the certificate



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

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

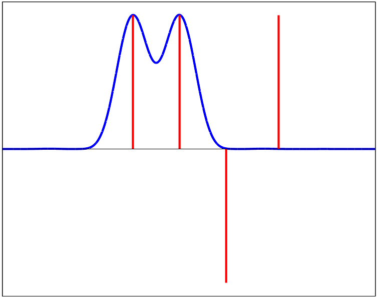
Construction of the certificate



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

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

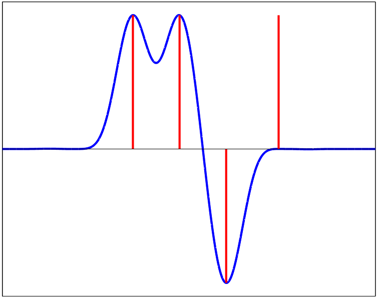
Construction of the certificate



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

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

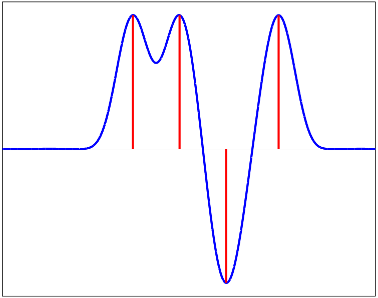
Construction of the certificate



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

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

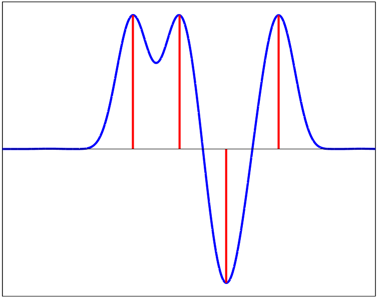
Construction of the certificate



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

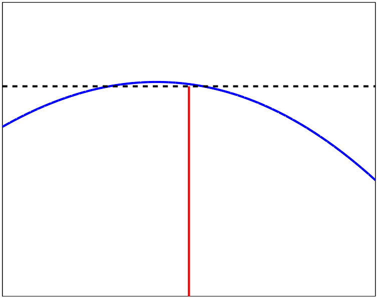
$$q(t) = \sum_{t_j \in T} \alpha_j K(t - t_j),$$

Construction of the certificate



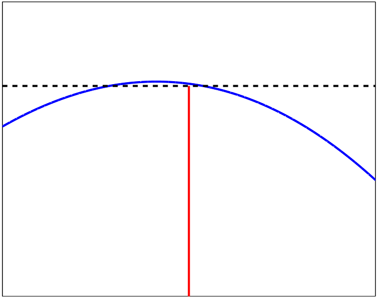
Problem : magnitude of polynomial locally exceeds 1

Construction of the certificate



Problem : magnitude of polynomial locally exceeds 1

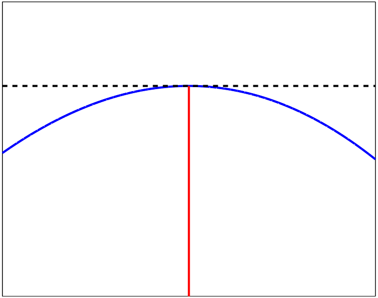
Construction of the certificate



Problem : magnitude of polynomial locally exceeds 1
Solution : add correction term and force $q'(t_k) = 0$ for all $t_k \in T$

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

Construction of the certificate

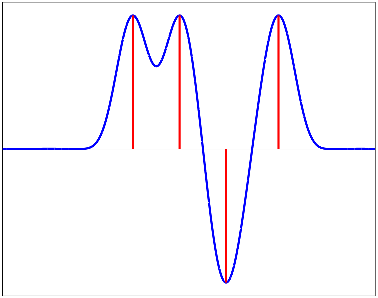


Problem : magnitude of polynomial locally exceeds 1

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

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

Construction of the certificate



Problem : magnitude of polynomial locally exceeds 1
Solution : add correction term and force $q'(t_k) = 0$ for all $t_k \in T$

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

Theory

Proof (sketch)

Robustness to noise

Algorithms

Approximation at a higher resolution

Without noise, we achieve perfect precision, i.e. infinite resolution

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

Approximation at a higher resolution

Without noise, we achieve perfect precision, i.e. infinite resolution

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

This is impossible if the measurements are noisy

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

Approximation at a higher resolution

Without noise, we achieve perfect precision, i.e. infinite resolution

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

This is impossible if the measurements are noisy

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

Resolution at scale λ is quantified by convolution with kernel ϕ_λ of width λ

Approximation at a higher resolution

Without noise, we achieve perfect precision, i.e. infinite resolution

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

This is impossible if the measurements are noisy

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

Resolution at scale λ is quantified by convolution with kernel ϕ_λ of width λ

If the noise z has norm δ, a trivial estimate x_{est} achieves

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

Approximation at a higher resolution

Without noise, we achieve perfect precision, i.e. infinite resolution

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

This is impossible if the measurements are noisy

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

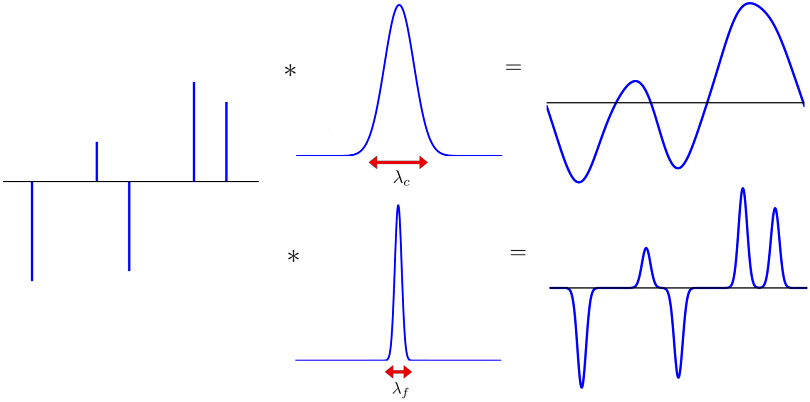
Resolution at scale λ is quantified by convolution with kernel ϕ_λ of width λ

If the noise z has norm δ, a trivial estimate x_{est} achieves

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

How does the estimate degrade at a higher resolution ?

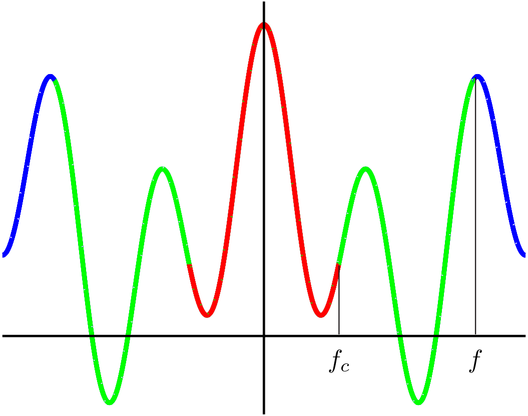
Super-resolution factor : spatial viewpoint



Super-resolution factor

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

Super-resolution factor : spectral viewpoint



Super-resolution factor

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

Approximation at a higher resolution

At the resolution of the measurements

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

Approximation at a higher resolution

At the resolution of the measurements

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

At a higher resolution

Theorem [Candès, F. 2012]

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

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

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

Approximation at a higher resolution

At the resolution of the measurements

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

At a higher resolution

Theorem [Candès, F. 2012]

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

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

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

Another metric for stability is the accuracy of support detection :
Support detection in super-resolution. C. Fernandez-Granda. SampTA 2013

Theory

Proof (sketch)

Robustness to noise

Algorithms

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 x (measure in $[0, 1]$)

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 x (measure in $[0,1]$)

- **Dual problem** :

$$\max_{u \in \mathbb{C}^n} \mathrm{Re}\left[y^* u\right] \quad \text{subject to} \quad ||\mathcal{F}_c^*\, u||_\infty \leq 1,$$

$$n := 2f_c + 1$$

Finite-dimensional variable u, but infinite-dimensional constraint

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

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 x (measure in $[0, 1]$)

- **Dual problem** :

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

$$n := 2f_c + 1$$

  Finite-dimensional variable u, but infinite-dimensional constraint

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

- Similar for relaxed versions that account for noise

Lemma : Semidefinite representation

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

is equivalent to

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

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

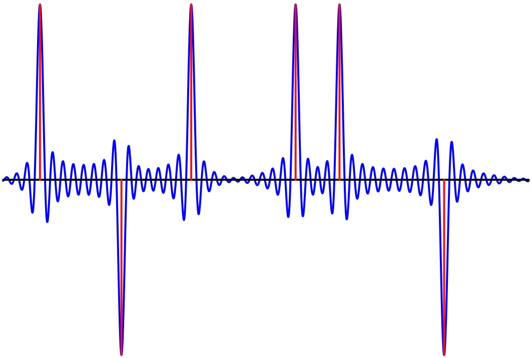
We can solve the dual problem, but how do we retrieve a primal solution ?

Implementation

Dual solution vector : Fourier coefficients of low-pass polynomial that interpolates the sign of the primal solution

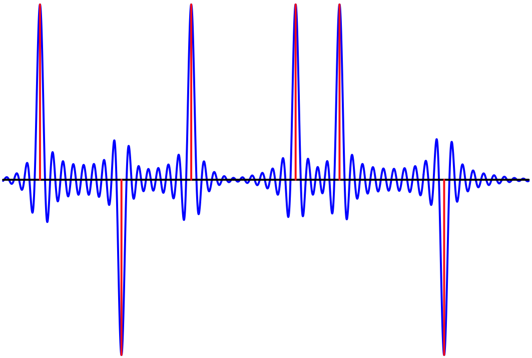
Implementation

Dual solution vector : Fourier coefficients of low-pass polynomial that interpolates the sign of the primal solution



Implementation

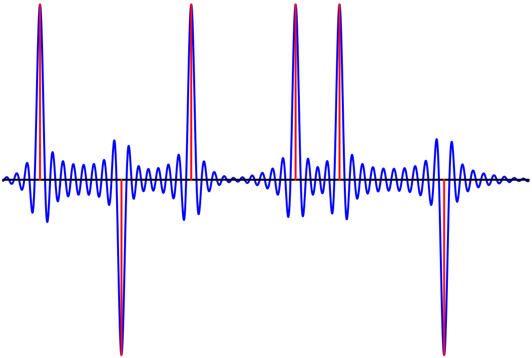
Dual solution vector : Fourier coefficients of low-pass polynomial that
interpolates the sign of the primal solution



To estimate the support we

Implementation

Dual solution vector : Fourier coefficients of low-pass polynomial that
interpolates the sign of the primal solution

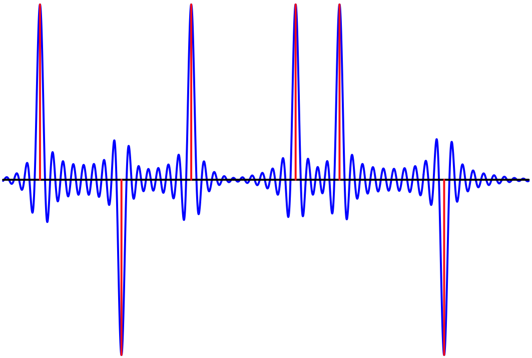


To estimate the support we

1. solve the sdp

Implementation

Dual solution vector : Fourier coefficients of low-pass polynomial that interpolates the sign of the primal solution

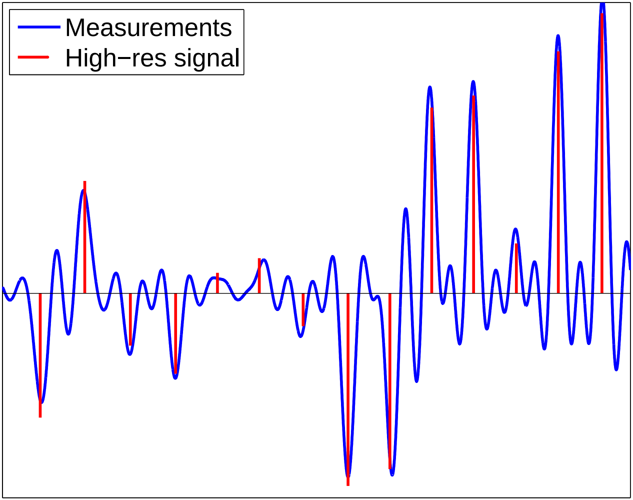


To estimate the support we

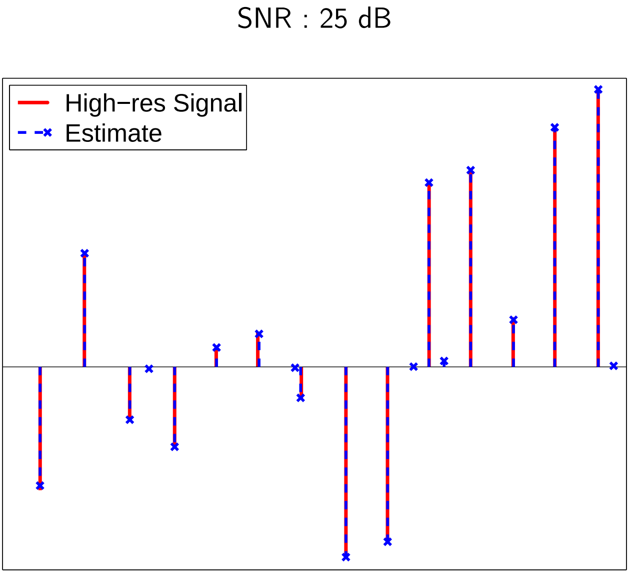
1. solve the sdp
2. determine where the magnitude of the polynomial equals 1

Example

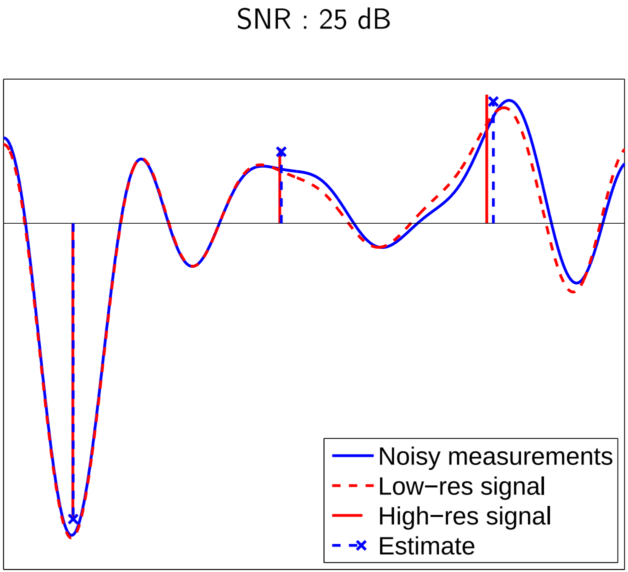
SNR : 25 dB



Example



Example


SNR : 25 dB
Noisy measurements
Low−res signal
High−res signal
Estimate

Conclusion and future work

- To obtain theoretical guarantees for super-resolution in realistic settings, we need conditions that avoid clustered supports

Conclusion and future work

- To obtain theoretical guarantees for super-resolution in realistic settings, we need conditions that avoid clustered supports
- Under a minimum-separation condition, convex programming achieves exact recovery

Conclusion and future work

- To obtain theoretical guarantees for super-resolution in realistic settings, we need conditions that avoid clustered supports
- Under a minimum-separation condition, convex programming achieves exact recovery
- The method is provably robust to noise

Conclusion and future work

- To obtain theoretical guarantees for super-resolution in realistic settings, we need conditions that avoid clustered supports
- Under a minimum-separation condition, convex programming achieves exact recovery
- The method is provably robust to noise
- The optimization problem can be recast as a tractable semidefinite program

Conclusion and future work

- To obtain theoretical guarantees for super-resolution in realistic settings, we need conditions that avoid clustered supports
- Under a minimum-separation condition, convex programming achieves exact recovery
- The method is provably robust to noise
- The optimization problem can be recast as a tractable semidefinite program
- Research directions :
  - Super-resolution of images with sharp edges
  - Developing fast solvers to solve sdp formulation

For more details

- **Towards a mathematical theory of super-resolution**. E. J. Candès and C. Fernandez-Granda. *Comm. on Pure and Applied Math.*
- **Super-resolution from noisy data**. E. J. Candès and C. Fernandez-Granda. *Journal of Fourier Analysis and Applications*
- **Support detection in super-resolution**. C. Fernandez-Granda. *Proceedings of SampTA 2013*