A Sampling Theorem for Deconvolution in Two Dimensions

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

Mathematics in Imaging, Imaging and Applied Optics Congress

Acknowledgements

Joint work with Brett Bernstein and Joey McDonald

Project funded by NSF award DMS-1616340

Deconvolution as sparse recovery

Certifying optimality

A sampling theorem for deconvolution

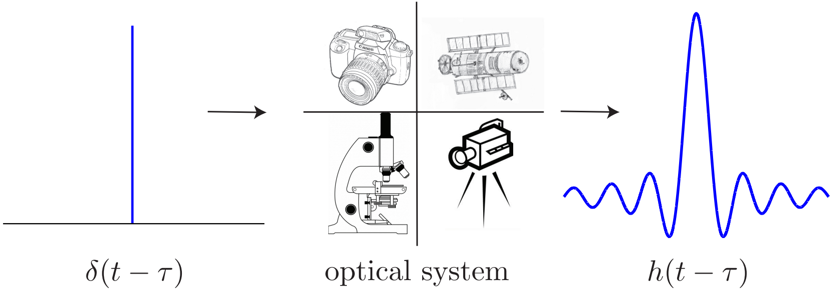
Deconvolution as sparse recovery

Certifying optimality

A sampling theorem for deconvolution

Limits of resolution in imaging

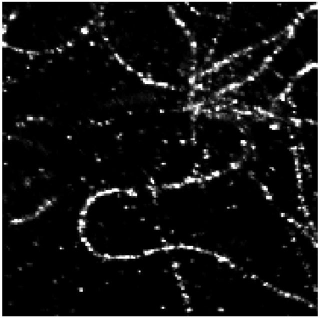
The resolving power of lenses, however perfect, is limited (**Lord Rayleigh**)



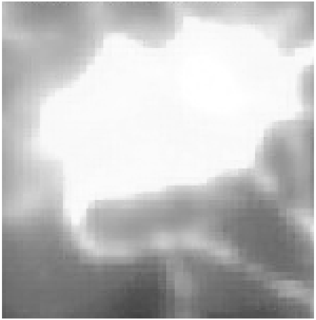
Diffraction imposes a fundamental limit on the resolution of optical systems

Fluorescence microscopy

Data



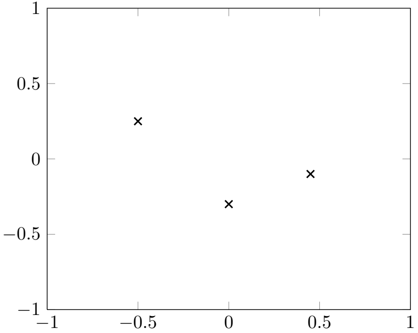
Point sources



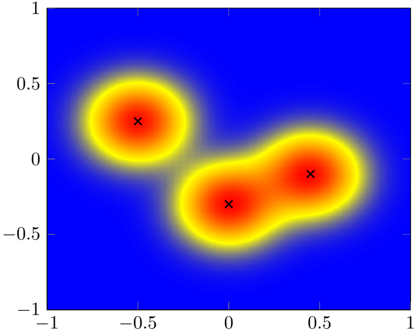
Low-pass blur

(Figures courtesy of V. Morgenshtern)

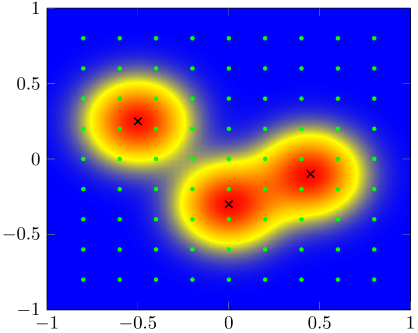
Sensing model for diffraction-limited imaging



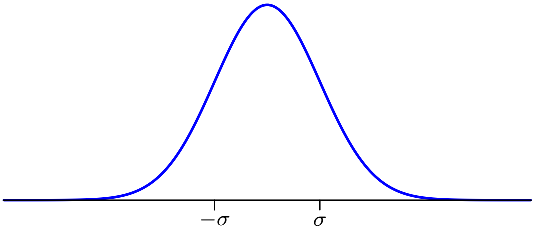
Sensing model for diffraction-limited imaging



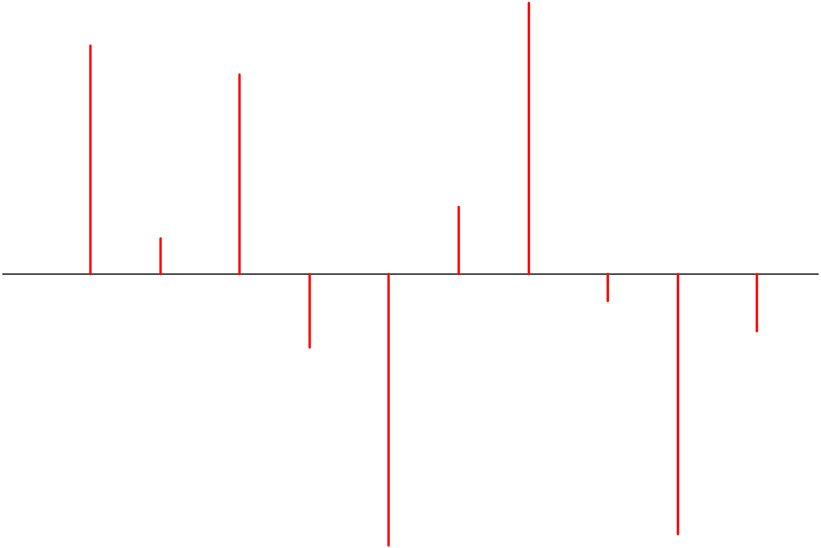
Sensing model for diffraction-limited imaging



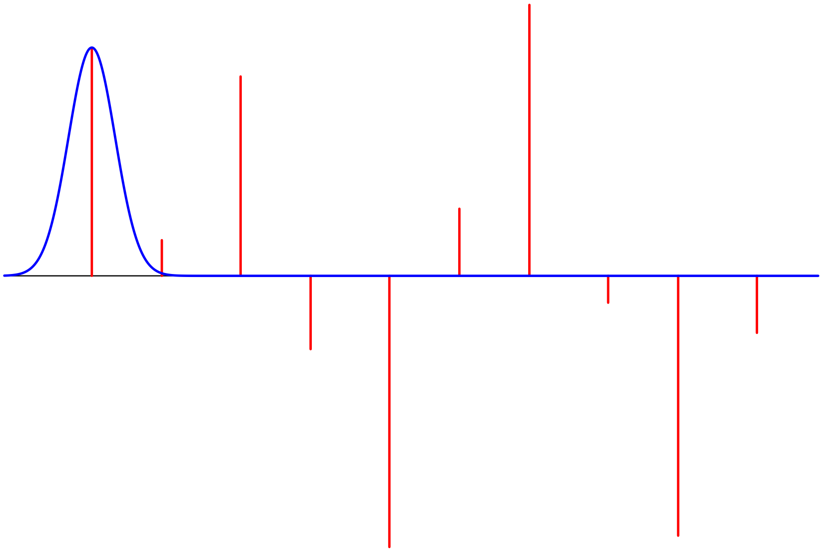
Model for the point-spread function: Gaussian kernel



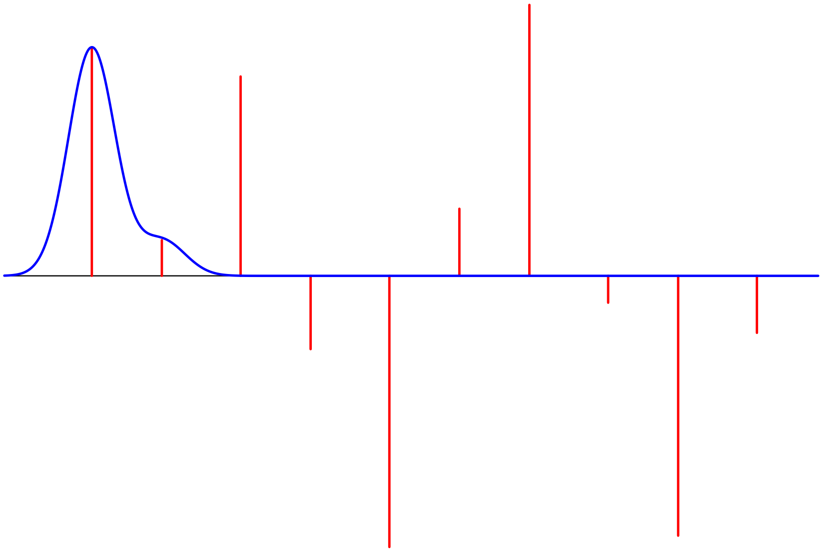
Sensing model for diffraction-limited imaging



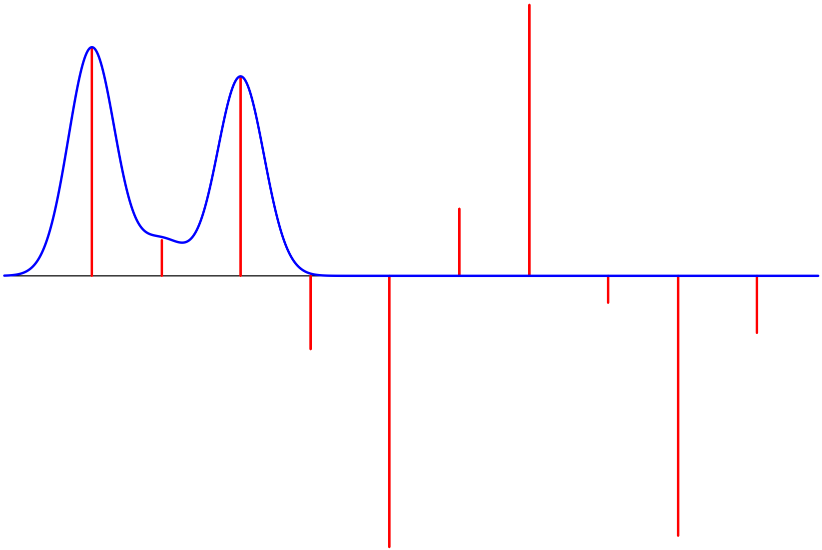
Sensing model for diffraction-limited imaging



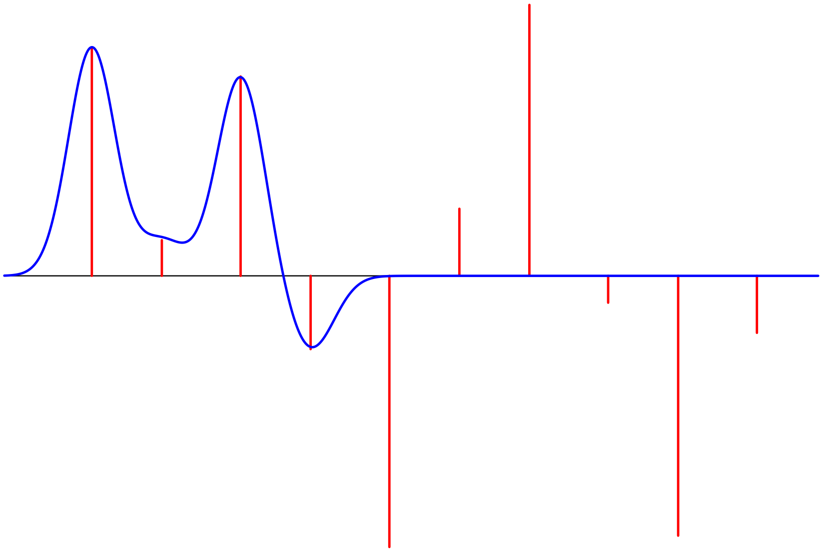
Sensing model for diffraction-limited imaging



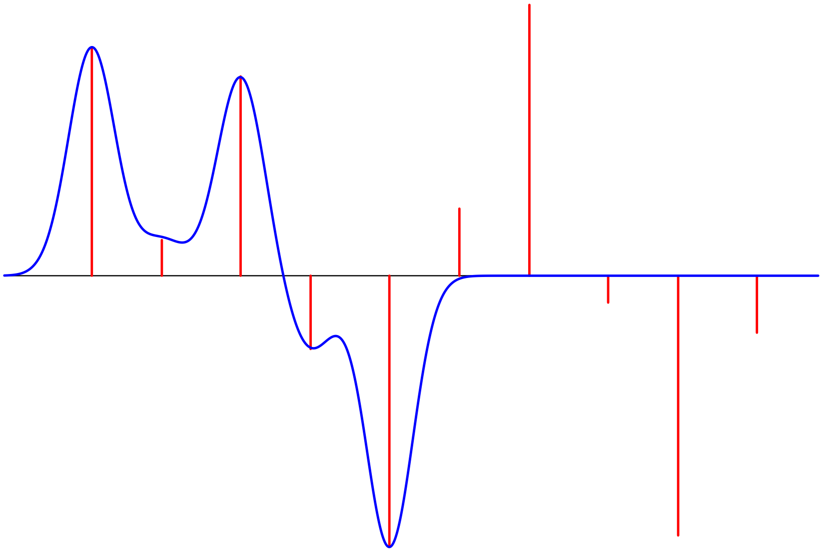
Sensing model for diffraction-limited imaging



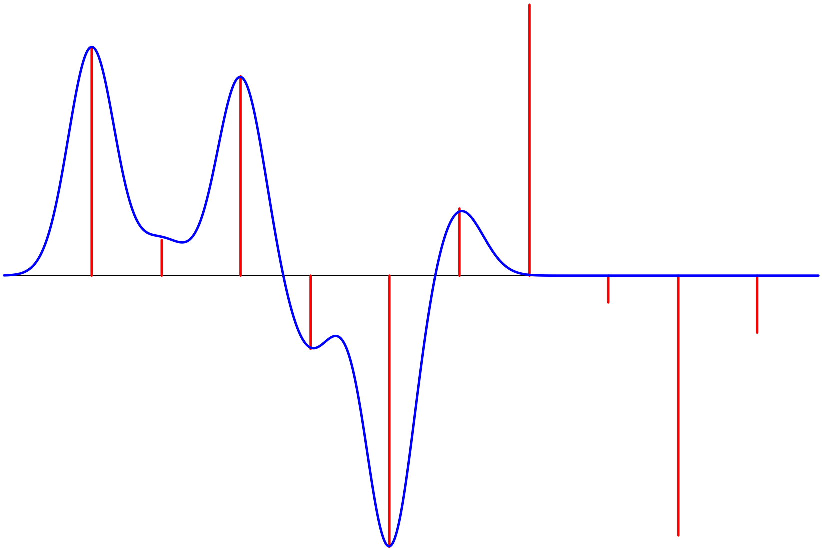
Sensing model for diffraction-limited imaging



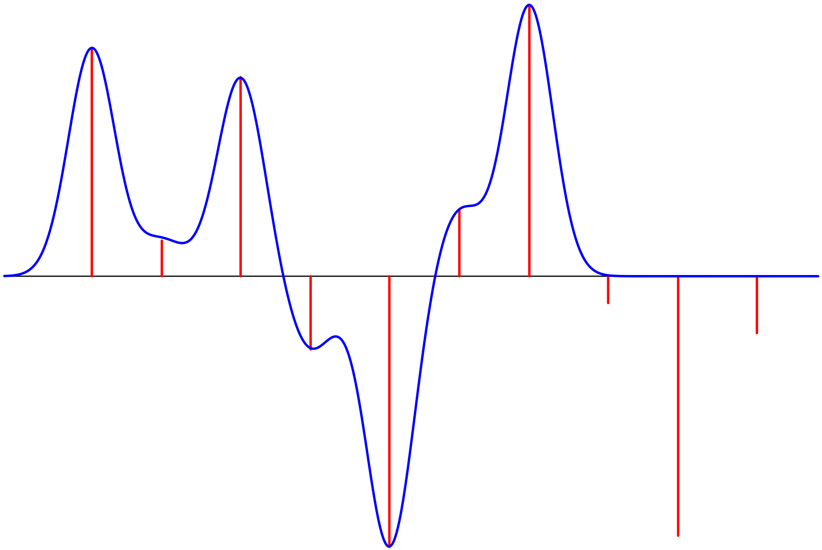
Sensing model for diffraction-limited imaging



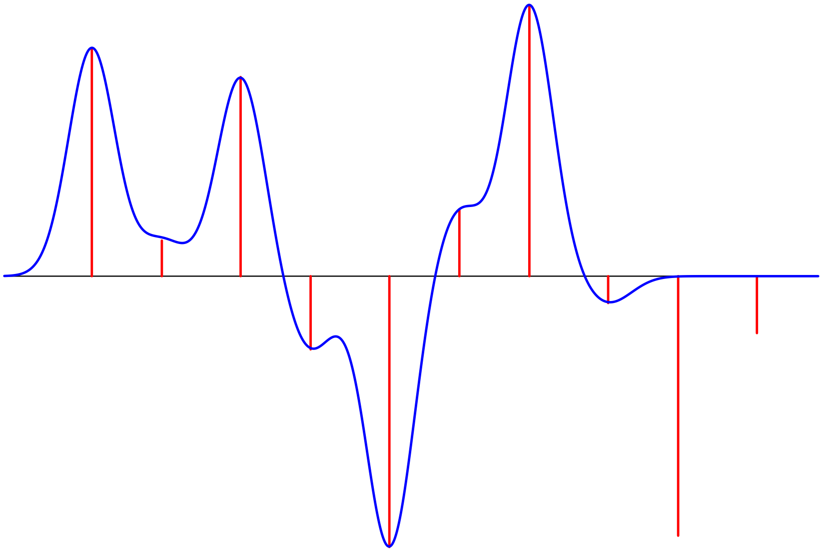
Sensing model for diffraction-limited imaging



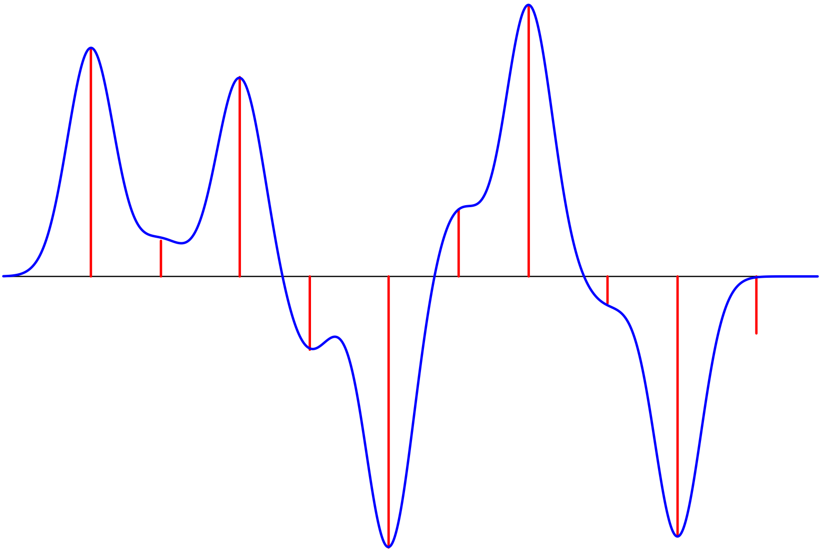
Sensing model for diffraction-limited imaging



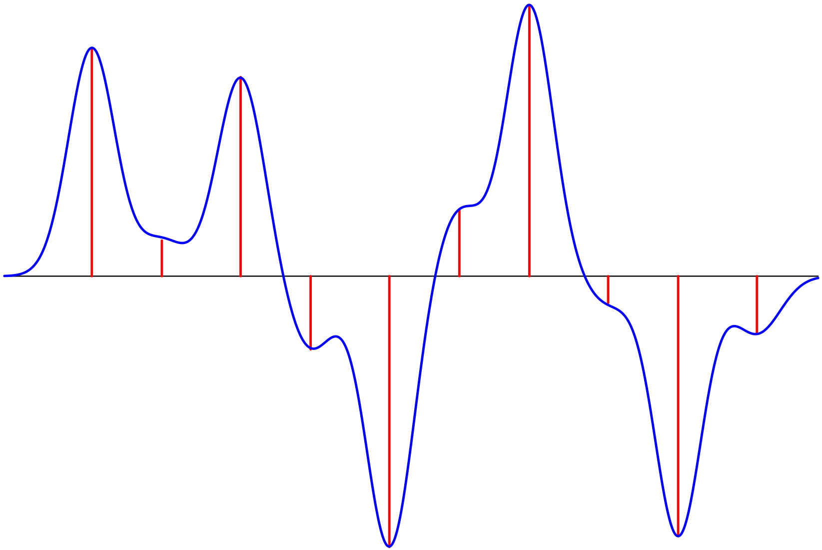
Sensing model for diffraction-limited imaging



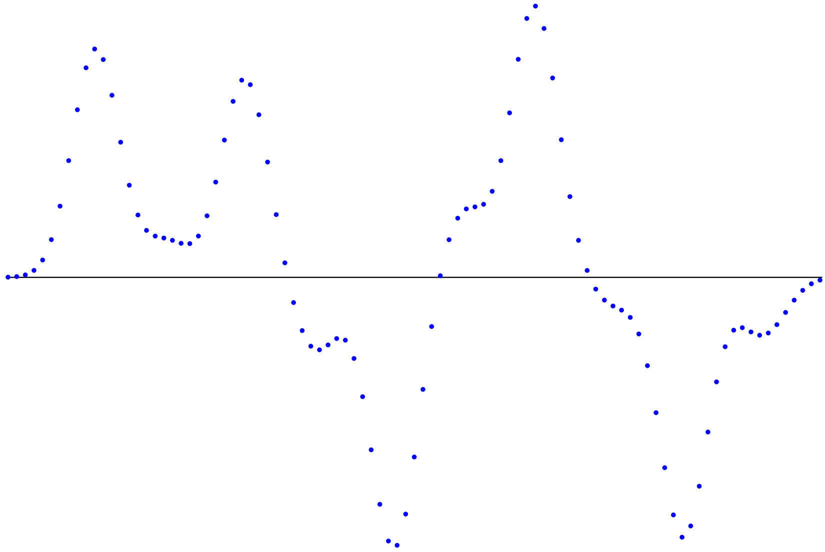
Sensing model for diffraction-limited imaging



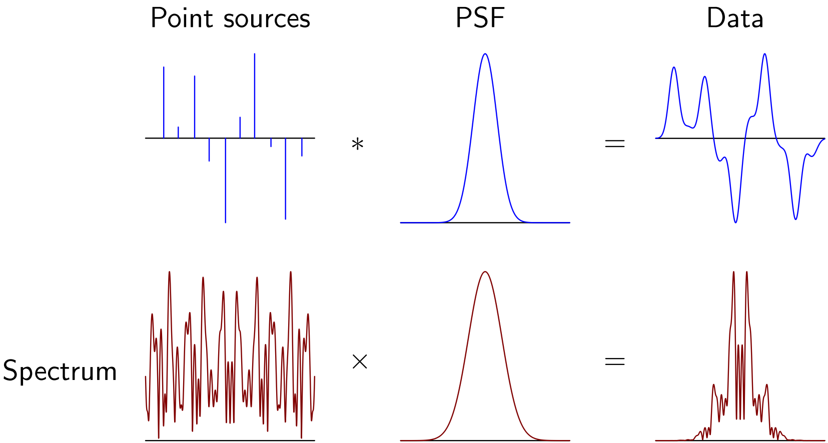
Sensing model for diffraction-limited imaging



Sensing model for diffraction-limited imaging



Sensing model for diffraction-limited imaging



Convolution in time = Pointwise multiplication in frequency

Ill-posed problem! How do we choose between signals consistent with data?

Mathematical model

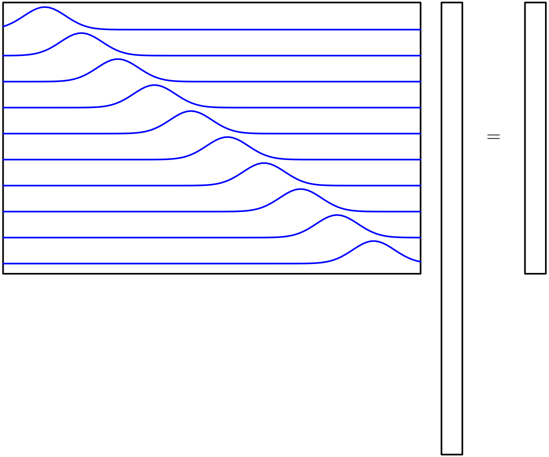
- Signal: superposition of k Dirac measures

$$x = \sum_{j=1}^{k} a_j \delta_{t_j} \qquad a_j \in \mathbb{R},\ t_j \in \mathbb{R}^d$$

- Data: n samples of convolution with PSF kernel K

$$\begin{aligned} y &:= \mathcal{K}\, x \\ y_i &:= (K * x)(s_i) \\ &= \int K(s_i - t)\, \mathrm{d}x, \quad i = 1, 2, \ldots, n \end{aligned}$$

In 1D



(Extremely) underdetermined linear inverse problem!

Sparse recovery for deconvolution

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

$$y := \mathcal{K}\,\tilde{x}$$

We need a tractable method to promote sparsity

Minimize ℓ_1 norm

Faster STORM using compressed sensing. *Nature methods*
Zhu, L., Zhang, W., Elnatan, D., Huang, B. (2012), 9(7), 721

Minimize ℓ_1 norm

Faster STORM using compressed sensing. *Nature methods*
Zhu, L., Zhang, W., Elnatan, D., Huang, B. (2012), 9(7), 721

Approach originally pioneered by geophysicists

Deconvolution with the ℓ_1 norm
Howard L. Taylor,* Stephen C. Banks,‡ and John F. McCoy§
GEOPHYSICS, VOL. 44, NO. 1 (JANUARY 1979)

LINEAR INVERSION OF BAND-LIMITED REFLECTION SEISMOGRAMS*
FADIL SANTOSA† AND WILLIAM W. SYMES‡
SIAM J. SCI. STAT. COMPUT.
Vol. 7, No. 4, October 1986

Reconstruction of a sparse spike train from a portion of its spectrum and application to high-resolution deconvolution
Shlomo Levy* and Peter K. Fullagar‡
GEOPHYSICS, VOL. 46, NO. 9 (SEPTEMBER 1981)

ROBUST MODELING WITH ERRATIC DATA†
JON F. CLAERBOUT* AND FRANCIS MUIR‡
GEOPHYSICS, VOL. 38, NO. 5 (OCTOBER 1973)

ℓ_1-norm minimization

$$\begin{aligned} &\textit{minimize} && ||\text{estimate}||_1 \\ &\textit{subject to} && \text{samples of convolution with kernel} = \text{data} \end{aligned}$$

Total-variation norm

Aim: Analysis for arbitrarily fine grids

Continuous counterpart of the ℓ_1 norm

Not the total variation of a piecewise-constant function

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

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

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

Goal of this talk

Analysis of ℓ_1-norm/TV-norm minimization for spike deconvolution

Goal of this talk

Analysis of ℓ_1-norm/TV-norm minimization for spike deconvolution

But wait, isn't this just compressed sensing?

Compressed sensing

Recover k-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 $\mathbf{x}$

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

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 $\mathbf{x}$

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

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

$$||Ax_2 - Ax_1||_2$$

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 $\mathbf{x}$

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

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

$$\|Ax_2 - Ax_1\|_2 = \|A (x_2 - x_1)\|_2$$

Restricted isometry property (Candès, Tao 2006)

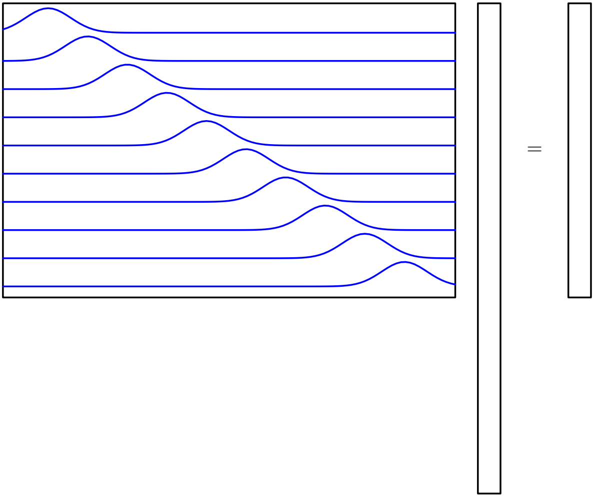
An $m \times n$ matrix A satisfies the restricted isometry property (RIP) if there exists $0 < \kappa < 1$ such that for any s-sparse vector $\mathbf{x}$

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

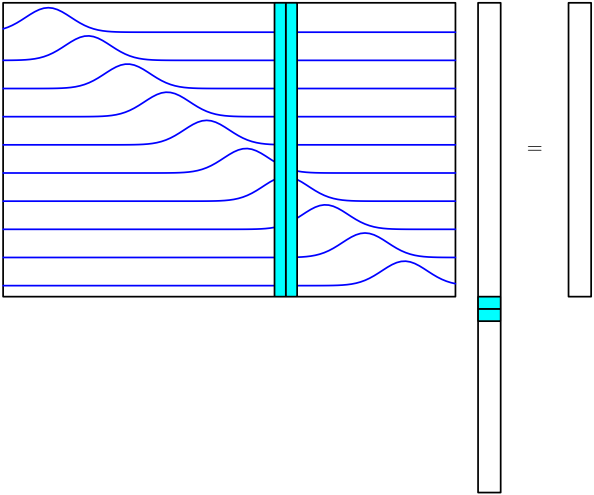
$2k$-RIP implies that for any k-sparse signals x_1, x_2

$$\begin{aligned} ||Ax_2 - Ax_1||_2 &= ||A\,(x_2 - x_1)||_2 \\ &\geq (1-\kappa)\,||x_2 - x_1||_2 \end{aligned}$$

Does the RIP hold for deconvolution?

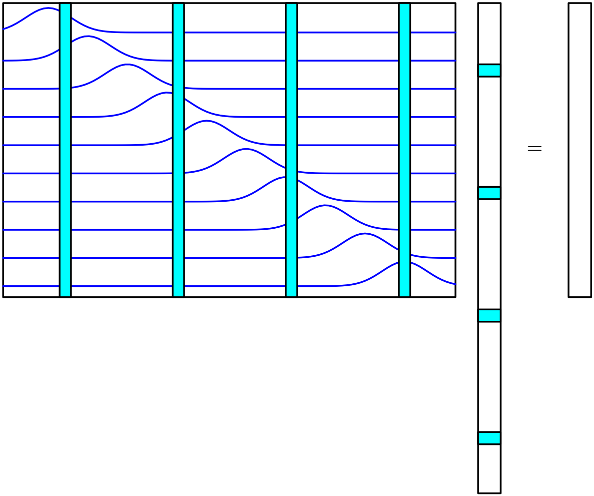


Does the RIP hold for deconvolution?



No!

Does the RIP hold for deconvolution?

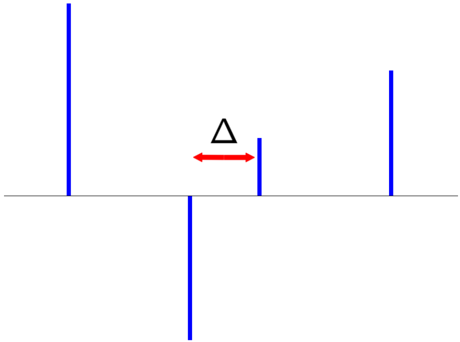


In deconvolution, sparsity is not enough...

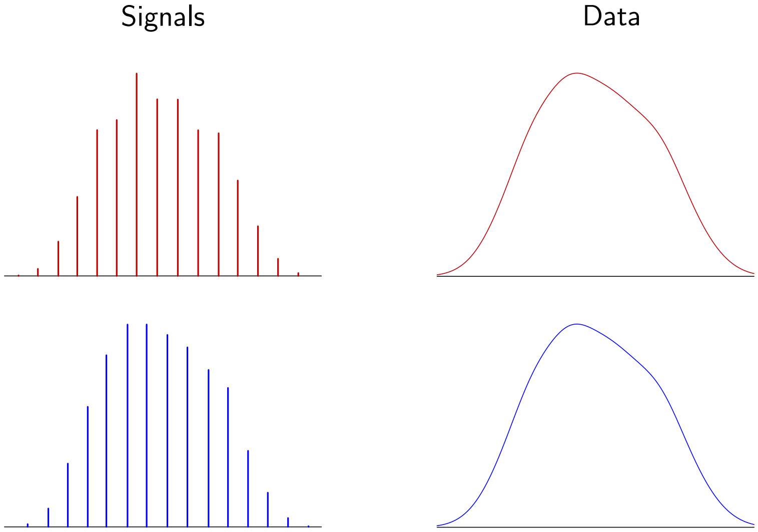
Minimum separation

The minimum separation Δ of the support of x is

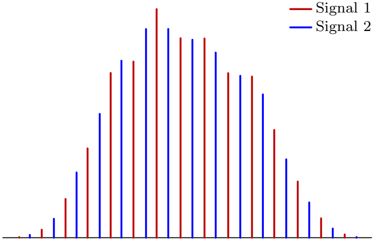
$$\Delta = \inf_{(t,t') \in \text{support}(x) \,:\, t \neq t'} |t - t'|$$



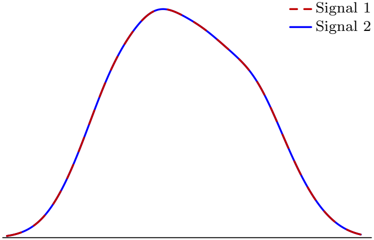
Example: 15 spikes, $\Delta = \sigma$



Example: 15 spikes, $\Delta = \sigma$



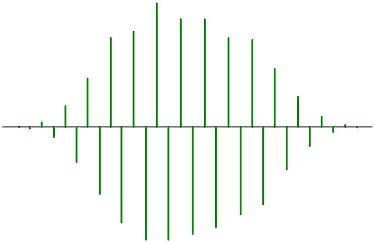
Signals



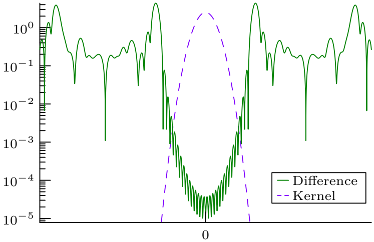
Data

Example: 15 spikes, $\Delta = \sigma$

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

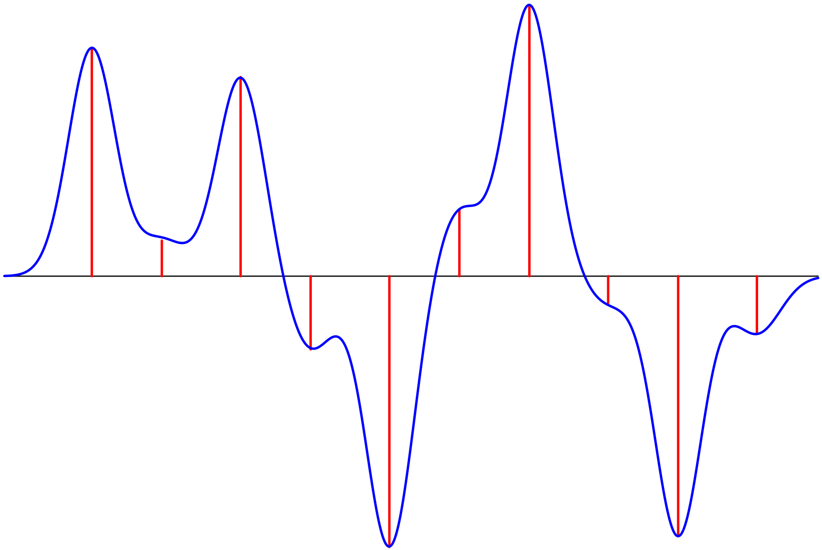


Difference

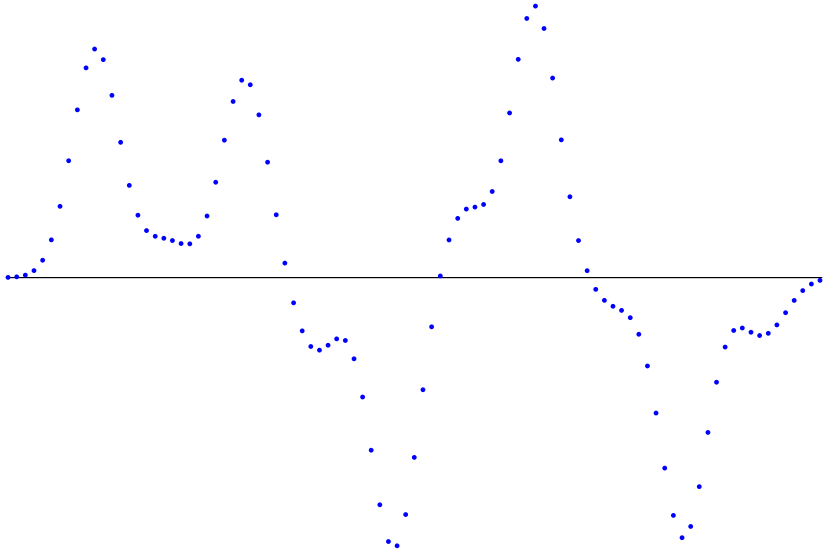


Spectrum

Sampling



Sampling



Sampling proximity

We need two samples per spike

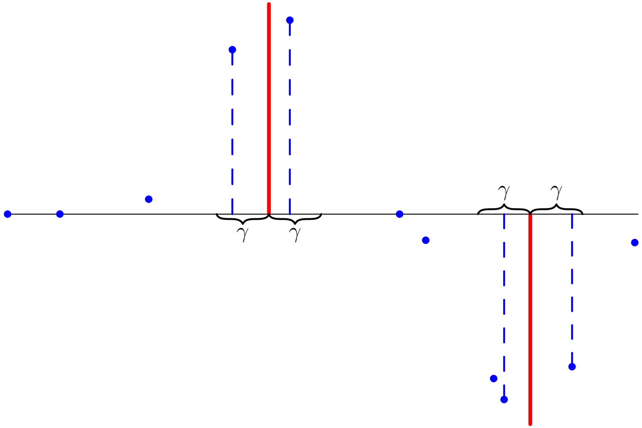
Convolution kernel decays: at least two samples close to each spike

Samples S and support T have sample proximity γ if for every $t_i \in T$ there exist $s, s' \in S$ such that

$$|t_i - s| \leq \gamma \quad \text{and} \quad \left|t_j - s'\right| \leq \gamma$$

We consider arbitrary non-uniform sampling patterns with fixed γ

Sampling proximity

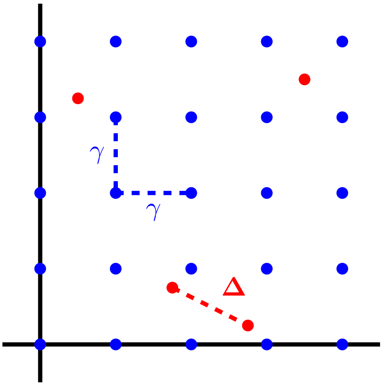


Aim

Prove exact recovery under two assumptions:

1. Signal support has *minimum separation*
2. Measurements satisfy *sampling-proximity* condition with respect to signal support

In 2D, regular grid



Aim (2D)

Prove exact recovery under two assumptions:

1. Signal support has *minimum separation*
2. Measurements are on a grid with a certain width

Deconvolution as sparse recovery

Certifying optimality

A sampling theorem for deconvolution

Analysis of ℓ_1-norm minimization

- Aim: Prove that any sparse x such that $Ax = y$ is the unique solution of

$$\begin{array}{ll} \textit{minimize} & \left|\left|x'\right|\right|_1 \\ \textit{subject to} & A\,x' = y \end{array}$$

Analysis of ℓ_1-norm minimization

- Aim: Prove that any sparse x such that $Ax = y$ is the unique solution of

$$\begin{aligned} &\textit{minimize} \quad && \left|\left|x'\right|\right|_1 \\ &\textit{subject to} \quad && A\,x' = y \end{aligned}$$

- Strategy: Build dual certificate associated to each sparse x

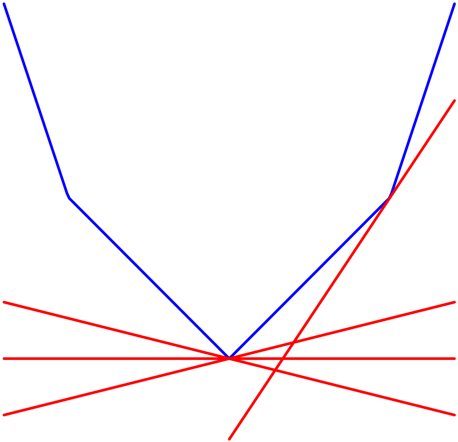
Subgradient

The subgradient of $f : \mathbb{R}^n \rightarrow \mathbb{R}$ at $x \in \mathbb{R}^n$ is a vector $g \in \mathbb{R}^n$ such that

$$f(y) \geq f(x) + g^T (y - x), \quad \text{for all } y \in \mathbb{R}^n$$

The set of all subgradients at x is called the subdifferential

Subgradients



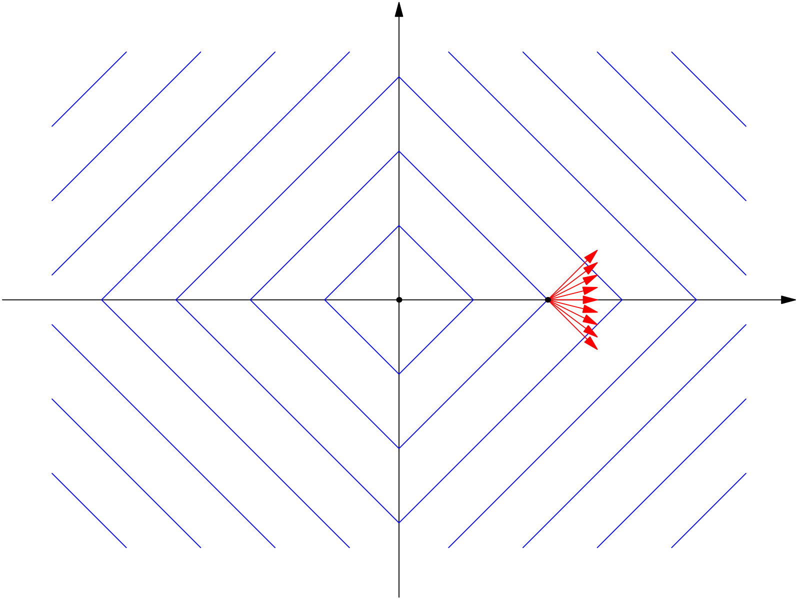
Subdifferential of ℓ_1 norm

g is a subgradient of the ℓ_1 norm at $x \in \mathbb{R}^n$ if and only if

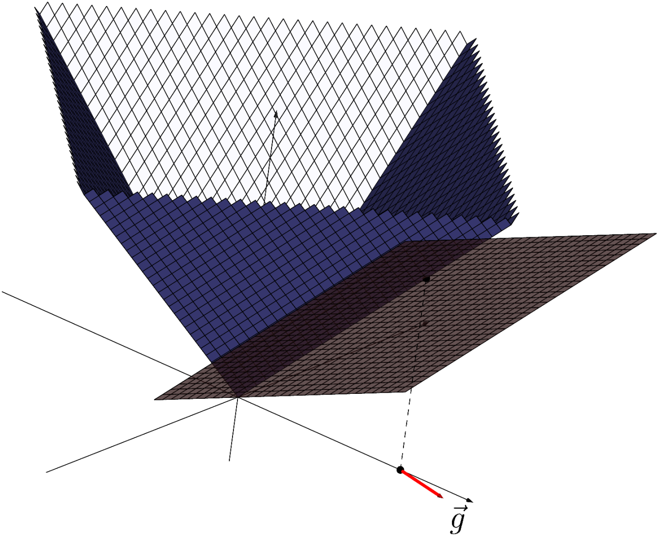
$$g[i] = \text{sign}\left(x[i]\right) \qquad \text{if } x[i] \neq 0$$

$$|g[i]| \leq 1 \qquad \text{if } x[i] = 0$$

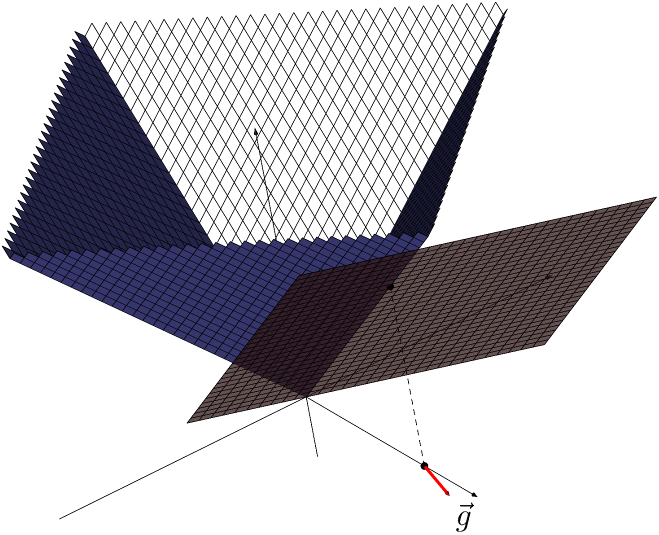
Subdifferential of ℓ_1 norm



Subdifferential of ℓ_1 norm



Subdifferential of ℓ_1 norm



Dual certificate

$v \in \mathbb{R}^m$ is a dual certificate associated to x if

$$q := A^T v$$

satisfies

$$\begin{aligned} q_i &= \operatorname{sign}(x_i) && \text{if } x_i \neq 0 \\ |q_i| &< 1 && \text{if } x_i = 0 \end{aligned}$$

Dual certificate

$v \in \mathbb{R}^m$ is a dual certificate associated to x if

$$q := A^T v$$

satisfies

$$\begin{aligned} q_i &= \operatorname{sign}(x_i) && \text{if } x_i \neq 0 \\ |q_i| &< 1 && \text{if } x_i = 0 \end{aligned}$$

q is a subgradient of the ℓ_1 norm at x

For any vector u

$$||x + u||_1 \geq ||x||_1 + q^T u$$

Dual certificate

For any $x + h$ such that $Ah = 0$

$$||x + h||_1 \geq ||x||_1 + q^T h \qquad (q \text{ is a subgradient})$$

Dual certificate

For any $x + h$ such that $Ah = 0$

$$\begin{aligned} \|x+h\|_1 &\geq \|x\|_1 + q^T h && (q \text{ is a subgradient}) \\ &= \|x\|_1 + v^T A h && (q = A^T v) \end{aligned}$$

Dual certificate

For any $x + h$ such that $Ah = 0$

$$\begin{aligned} \|x+h\|_1 &\geq \|x\|_1 + q^T h && (q \text{ is a subgradient}) \\ &= \|x\|_1 + v^T A h && (q = A^T v) \\ &= \|x\|_1 \end{aligned}$$

Dual certificate

For any $x + h$ such that $Ah = 0$

$$\begin{aligned} \|x + h\|_1 &\geq \|x\|_1 + q^T h && (q \text{ is a subgradient}) \\ &= \|x\|_1 + v^T A h && (q = A^T v) \\ &= \|x\|_1 \end{aligned}$$

If A_T (where T is the support of x) is injective, x is the unique solution

Dual certificate

A dual certificate of the TV norm at

$$x = \sum_i a_i \delta_{t_i} \qquad a_i \in \mathbb{R},\ t_i \in T$$

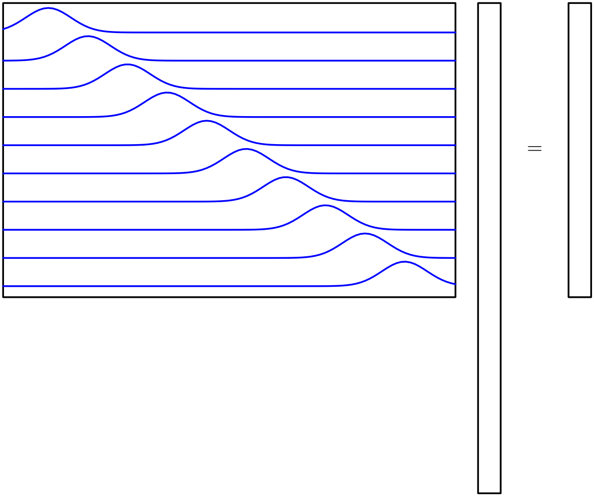
guarantees that x is the unique solution if

$$q(t) := \left(\mathcal{K}^T v\right)(t) = \sum_{j=1}^{n} v_j K(s_j - t)$$

$$q(t_i) = \operatorname{sign}(a_i) \qquad \text{if } t_i \in T$$

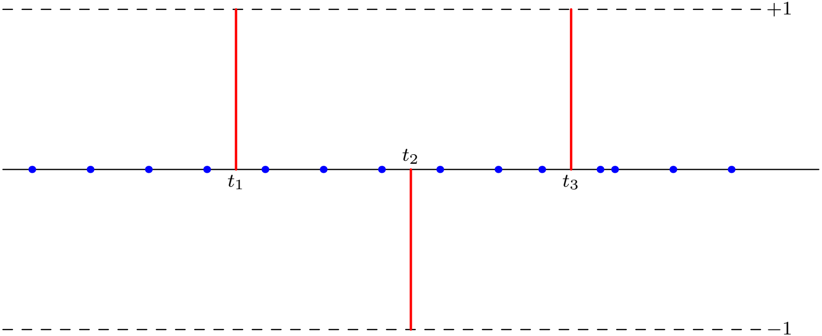
$$|q(t)| < 1 \qquad \text{if } t \notin T$$

In 1D



Range of $\mathcal{K}^T$ is spanned by shifted copies of K fixed at the samples

Certificate for deconvolution



Aim: Interpolate using PSFs centered at samples

Deconvolution as sparse recovery

Certifying optimality

A sampling theorem for deconvolution

Aim (1D)

Build certificate for arbitrary signals/measurements assuming:

1. Signal support has *minimum separation*
2. Measurements satisfy *sampling-proximity* condition with respect to signal support

Certificate construction

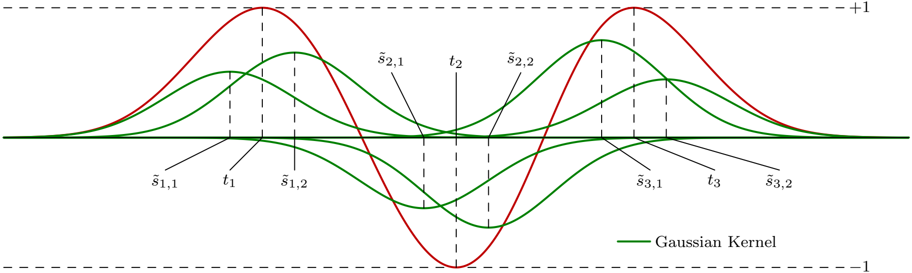
Idea: Only use subset of data $\widetilde{S}$ containing 2 samples close to each spike

$$q(t) = \sum_{s_j \in \widetilde{S}} v_j K\left(s_j - t\right)$$

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

$$\begin{aligned} q\left(t_i\right) &= \operatorname{sign}\left(a_i\right) \\ q'\left(t_i\right) &= 0 \end{aligned}$$

It works!


+1
−1
$\tilde{s}_{2,1}$
t_2
$\tilde{s}_{2,2}$
$\tilde{s}_{1,1}$
t_1
$\tilde{s}_{1,2}$
$\tilde{s}_{3,1}$
t_3
$\tilde{s}_{3,2}$
Gaussian Kernel

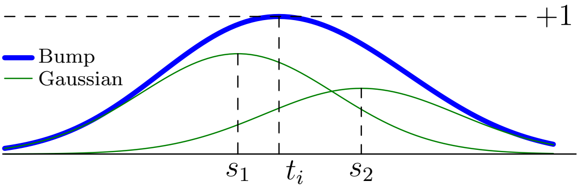
Certificate construction

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

Solution: Reparametrization into *bumps* and *waves*

$$
\begin{aligned}
q(t) &= \sum_{s_j \in \widetilde{S}} v_j K\left(s_j - t\right) \\
&= \sum_{t_i \in T} \alpha_i B_{t_i}(t, \tilde{s}_{i,1}, \tilde{s}_{i,2}) + \beta_i W_{t_i}(t, \tilde{s}_{i,1}, \tilde{s}_{i,2}),
\end{aligned}
$$

Bump function

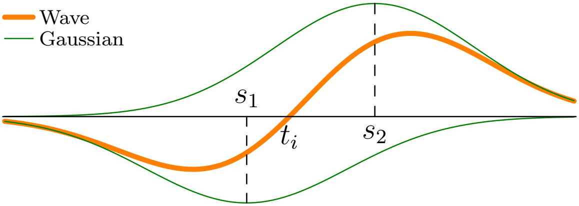


$$B_{t_i}(t, \tilde{s}_{i,1}, \tilde{s}_{i,2}) := b_{i,1} K(\tilde{s}_{i,1} - t) + b_{i,2} K(\tilde{s}_{i,2} - t)$$

$$B_{t_i}(t_i, \tilde{s}_{i,1}, \tilde{s}_{i,2}) = 1$$

$$\frac{\partial}{\partial t} B_{t_i}(t_i, \tilde{s}_{i,1}, \tilde{s}_{i,2}) = 0$$

Wave function



$$W_{t_i}(t, \tilde{s}_{i,1}, \tilde{s}_{i,2}) = w_{i,1} K(\tilde{s}_{i,1} - t) + w_{i,2} K(\tilde{s}_{i,2} - t)$$

$$W_{t_i}(t_i, \tilde{s}_{i,1}, \tilde{s}_{i,2}) = 0$$

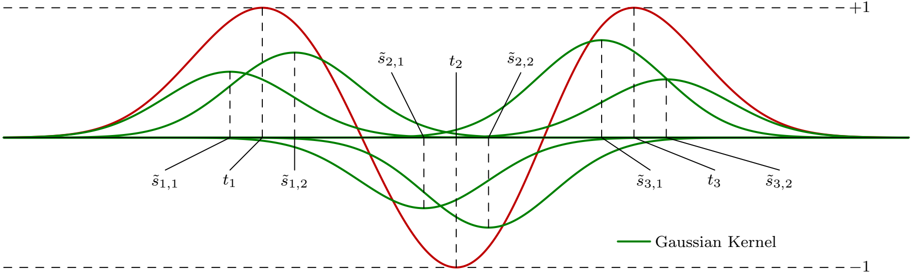
$$\frac{\partial}{\partial t} W_{t_i}(t_i, \tilde{s}_{i,1}, \tilde{s}_{i,2}) = 1$$

Certificate construction

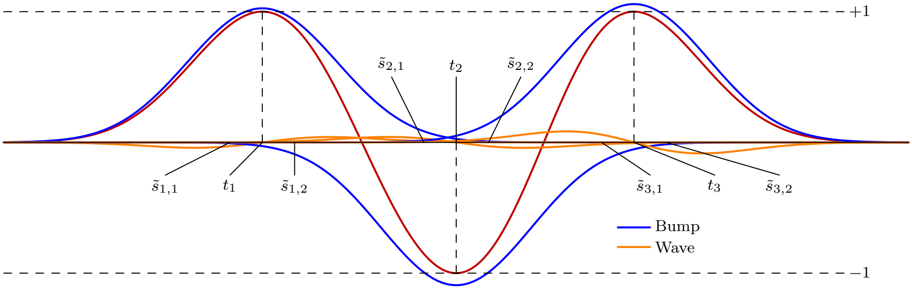
Reparametrization decouples the coefficients

$$
\begin{aligned}
q(t) &= \sum_{s_j \in \widetilde{S}} v_j K\left(s_j - t\right) \\
&= \sum_{t_i \in T} \alpha_i B_{t_i}(t, \tilde{s}_{i,1}, \tilde{s}_{i,2}) + \beta_i W_{t_i}(t, \tilde{s}_{i,1}, \tilde{s}_{i,2}) \\
&\approx \sum_{t_i \in T} \operatorname{sign}\left(a_i\right) B_{t_i}(t, \tilde{s}_{i,1}, \tilde{s}_{i,2})
\end{aligned}
$$

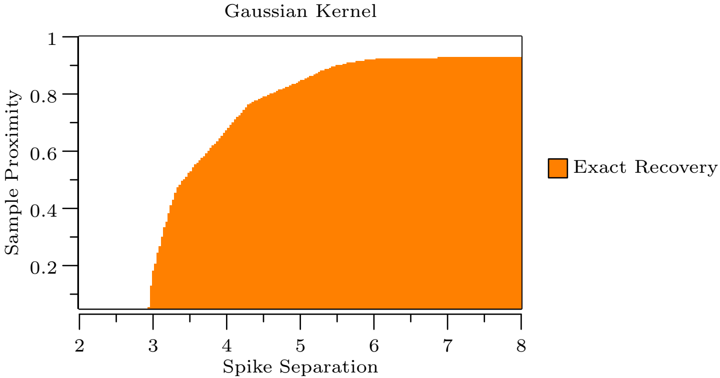
Certificate for deconvolution



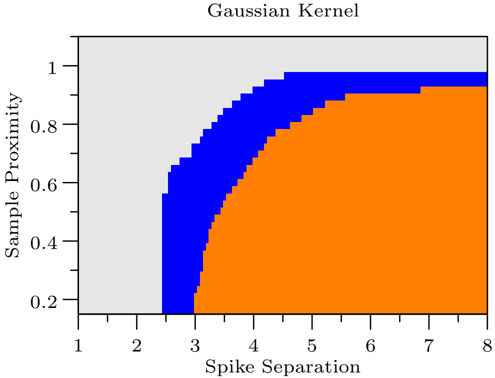
Certificate for deconvolution



Exact recovery guarantees in 1D [Bernstein, F. 2017]



Guarantees vs numerical experiments (1D)

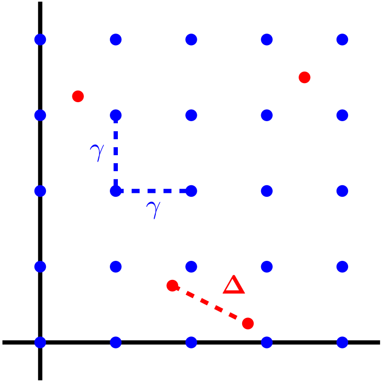


Aim (2D)

Build certificate for arbitrary signals/measurements assuming:

1. Signal support has *minimum separation*
2. Measurements are on a grid with a certain *width*

In 2D, regular grid



2D certificate construction

Same idea: Use subset of data $\widetilde{S}$ containing 3 samples close to each spike

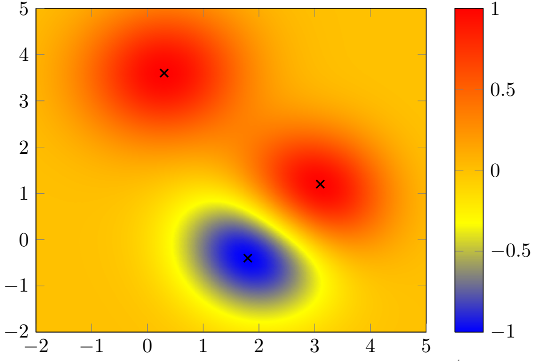
$$q(t) = \sum_{s_j \in \widetilde{S}} v_j K\left(s_j - t\right)$$

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

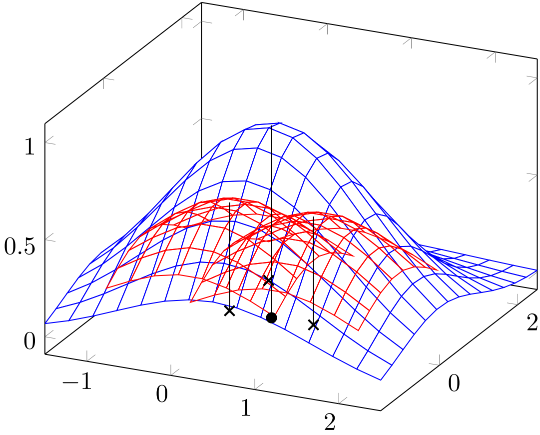
$$\begin{aligned} q\left(t_i\right) &= \operatorname{sign}\left(a_i\right) \\ \nabla q\left(t_i\right) &= 0 \end{aligned}$$

Reparametrize into bumps and waves to make analysis tractable

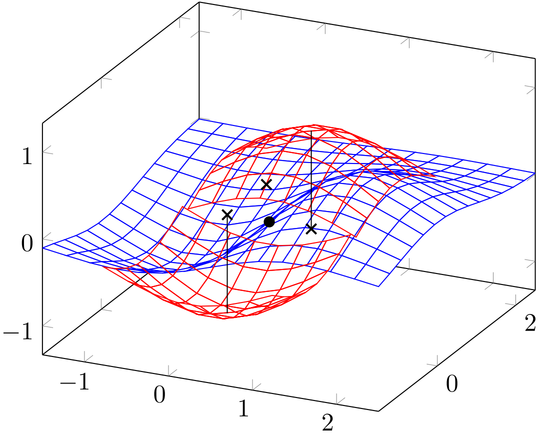
2D certificate



2D bump function

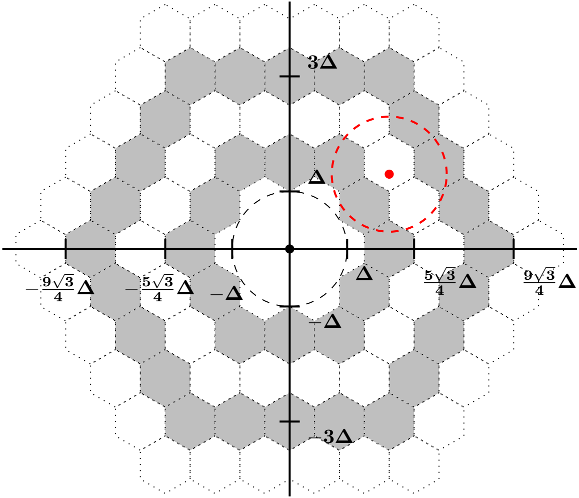


2D wave function

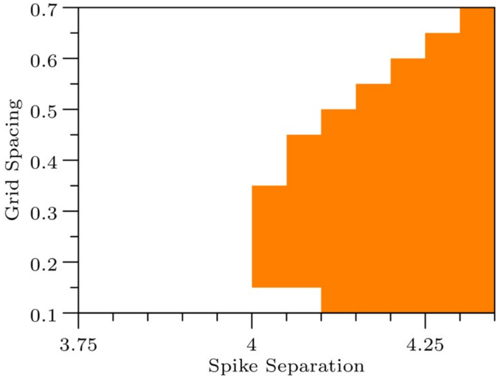


Challenge

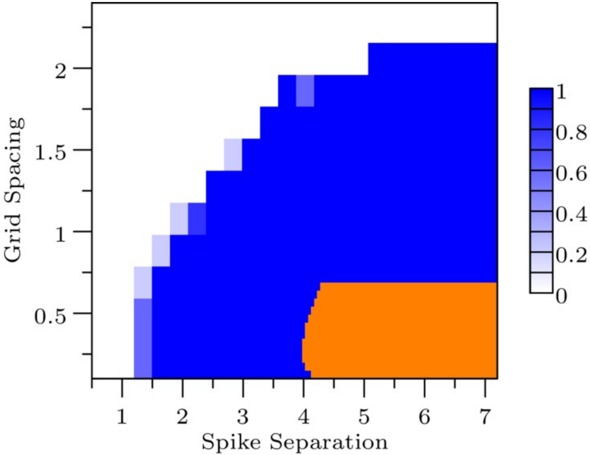
Controlling the magnitude of the certificate requires geometric arguments



Exact recovery guarantees in 2D [McDonald, Bernstein, F. 2019]



Guarantees vs numerical experiments (2D)



Conclusion

Compressed-sensing intuition / tools for randomized measurements do not apply to deconvolution

Conditions beyond sparsity are necessary to make the problem well posed

Under such conditions the method achieves exact recovery

Proofs rely on novel dual-certificate construction (*bumps* and *waves*)

References

- *A sampling theorem for deconvolution in two dimensions*. J. McDonald, B. Bernstein, C. Fernandez-Granda 2019
- *Deconvolution of point sources: A sampling theorem and robustness guarantees*. B. Bernstein, C. Fernandez-Granda *Communications on Pure and Applied Math.*, 2019