Solving Underdetermined Linear Systems with Highly Correlated Columns

Veniamin I. Morgenshtern

Statistics Department, Stanford

"Big data" world

Problem: Estimate a **high-dimensional** object from (relatively) **few** corrupted data points.

"Big data" world

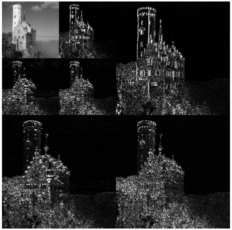
Problem: Estimate a **high-dimensional** object from (relatively) **few** corrupted data points.

Assumption: The object is **simple**: low-dimensional structure.

"Big data" world

Problem: Estimate a **high-dimensional** object from (relatively) **few** corrupted data points.

Assumption: The object is **simple**: low-dimensional structure.

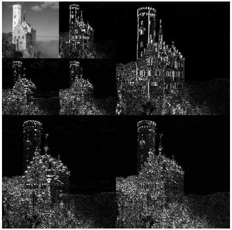


signal processing

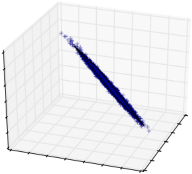
“Big data” world

Problem: Estimate a **high-dimensional** object from (relatively) **few** corrupted data points.

Assumption: The object is **simple**: low-dimensional structure.



signal processing

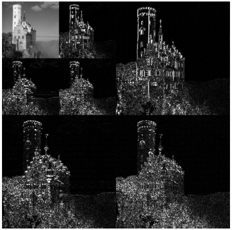


statistics

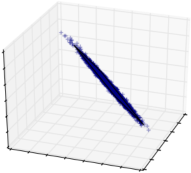
“Big data” world

Problem: Estimate a **high-dimensional** object from (relatively) **few** corrupted data points.

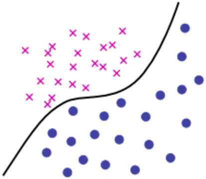
Assumption: The object is **simple**: low-dimensional structure.



signal processing

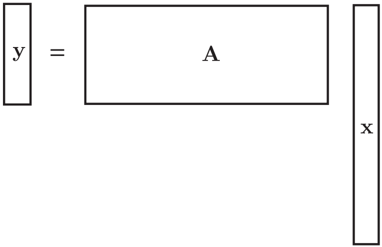


statistics



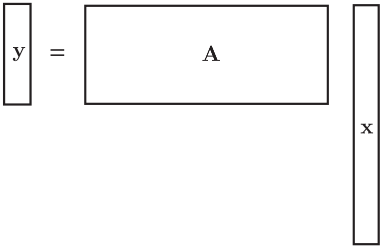
machine learning

Underdetermined linear system



Few linear measurements about a high-dimensional object

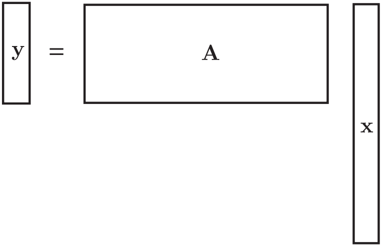
Underdetermined linear system



Few linear measurements about a high-dimensional object

How can we possibly recover the object?

Underdetermined linear system

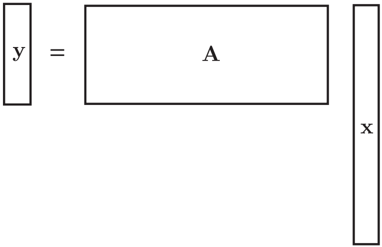


Few linear measurements about a high-dimensional object

How can we possibly recover the object?

The object has low-dimensional representation (sparsity)

Underdetermined linear system



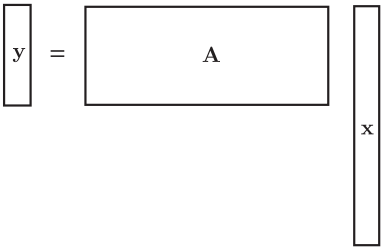
Few linear measurements about a high-dimensional object

How can we possibly recover the object?

The object has low-dimensional representation (sparsity)

Many problems are naturally of this form.

Underdetermined linear system



Few linear measurements about a high-dimensional object

How can we possibly recover the object?

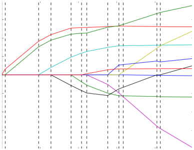
The object has low-dimensional representation (sparsity)

Many problems are naturally of this form.

Even more problems can be forced into this form!

Huge range of **“big data”** applications

compressed sensing
MRI

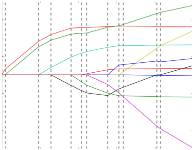


model selection via
LASSO

low rank matrix
recovery
(Netflix Prize)

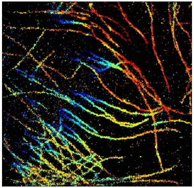
Huge range of **"big data"** applications

compressed sensing MRI



model selection via LASSO

low rank matrix recovery (Netflix Prize)

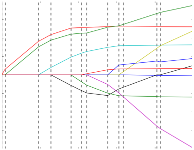


super-resolution microscopy

radar imaging

Huge range of **“big data”** applications

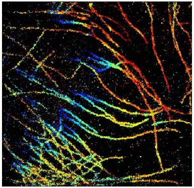
compressed sensing MRI



model selection via LASSO

low rank matrix recovery (Netflix Prize)

Beautiful mathematical core!



super-resolution microscopy

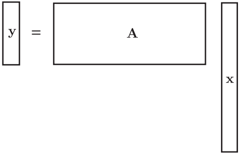
radar imaging

Structure of this talk

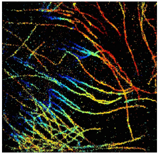
radar imaging
(main novelty)

Structure of this talk

basic sparse signal recovery



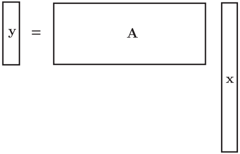
super-resolution microscopy



radar imaging
(main novelty)

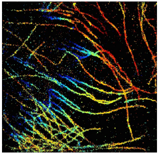
Structure of this talk

basic sparse signal recovery



$\mathbf{A}$ has random uncorrelated columns

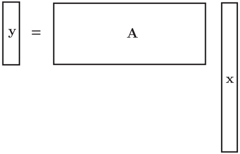
super-resolution microscopy



radar imaging
(main novelty)

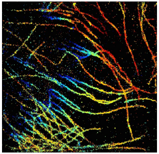
Structure of this talk

basic sparse signal recovery



$\mathbf{A}$ has random uncorrelated columns

super-resolution microscopy

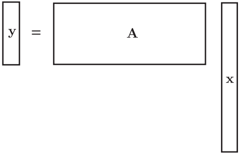


$\mathbf{A}$ has deterministic highly correlated columns

radar imaging
(main novelty)

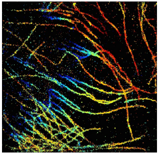
Structure of this talk

basic sparse signal recovery



A has random
uncorrelated columns

super-resolution microscopy



A has deterministic
highly correlated columns

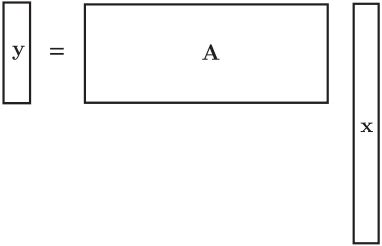
radar imaging
(main novelty)

A has random
highly correlated
columns

Basic theory of sparse signal recovery

E. Candès
D. Donoho
J. Romberg
T. Tao
...

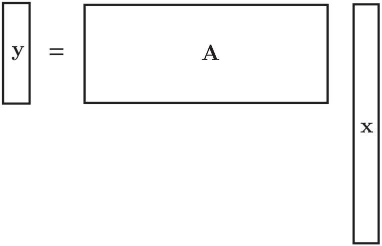
Prototypical model



Assumptions:

(1) $\mathbf{y}$ is M dimenstional

Prototypical model

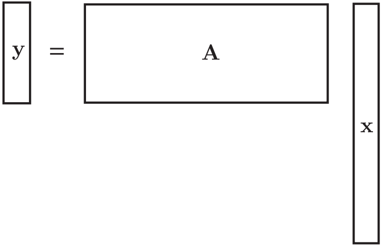


Assumptions:

(1) $\mathbf{y}$ is M dimenstional

(2) Sparsity: $\mathbf{x}$ has at most S nonzero entries ($S < M$)

Prototypical model



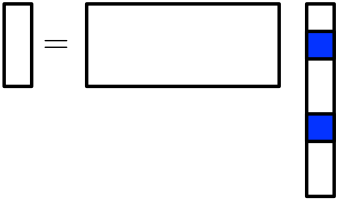
Assumptions:

(1) $\mathbf{y}$ is M dimenstional

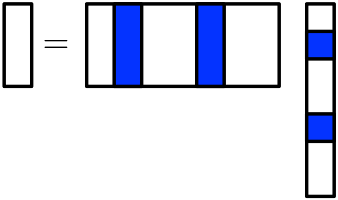
(2) Sparsity: $\mathbf{x}$ has at most S nonzero entries ($S < M$)

(3) Properties of $\mathbf{A}$: $A_{ij} \sim \mathcal{N}(0, 1)$

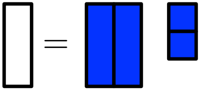
If we knew where nonzeros are ...



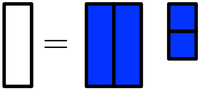
If we knew where nonzeros are ...



If we knew where nonzeros are ...

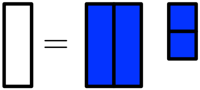


If we knew where nonzeros are ...



The search for nonzeros is combinatorial in nature!

If we knew where nonzeros are ...



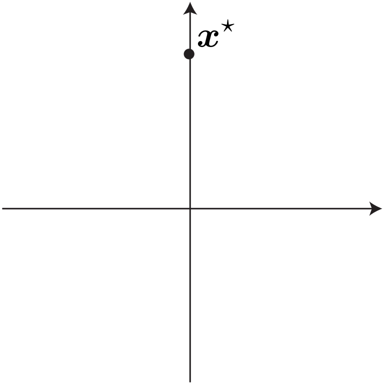
The search for nonzeros is combinatorial in nature!

Recovery by convex programming (relaxation):

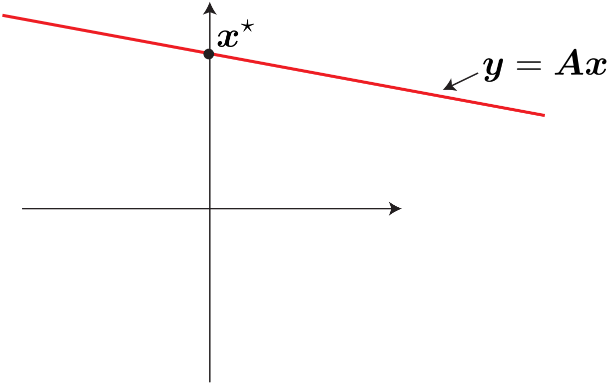
$$\text{minimize} \quad \underbrace{\sum_i |x_i|}_{\text{l1-norm } \|\mathbf{x}\|_1} \quad \text{subject to} \quad \mathbf{y} = \mathbf{A}\mathbf{x}$$

Min norm problem is a convex program and computationally tractable

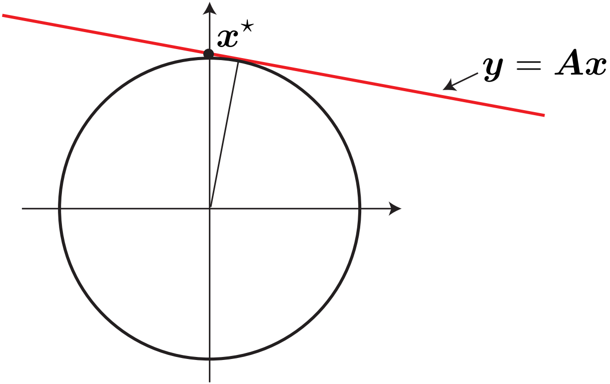
Why does ℓ_1 work?



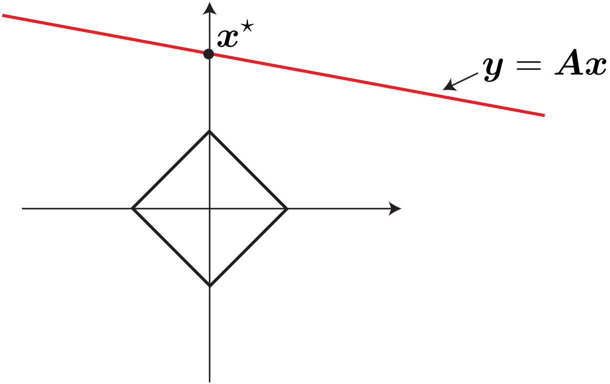
Why does ℓ_1 work?



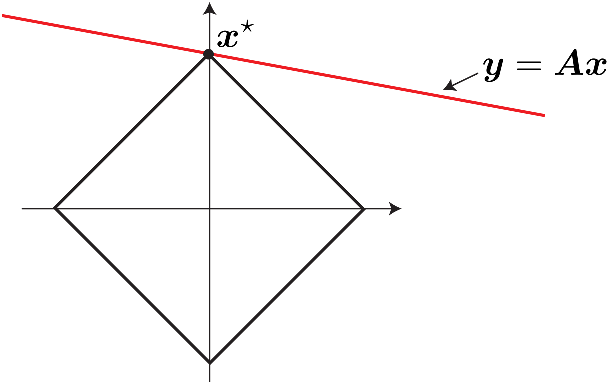
Why does ℓ_1 work?



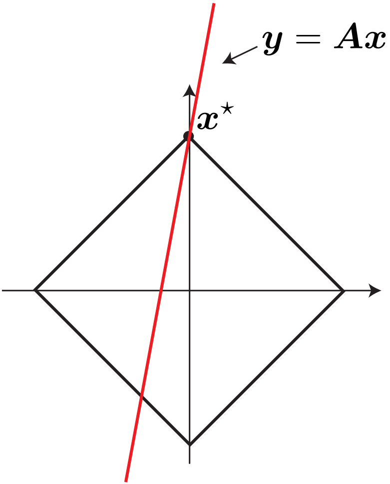
Why does ℓ_1 work?



Why does ℓ_1 work?



Why ℓ_1 may not always work

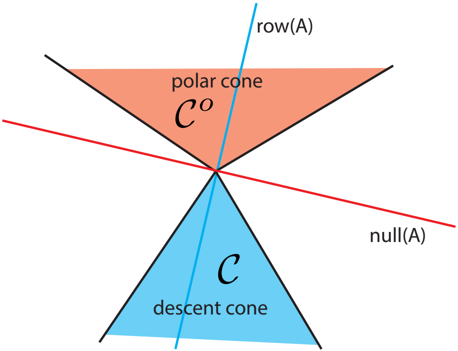


Dual certificates

$$\text{minimize} \quad \|\mathbf{x}\|_1 \quad \text{such that} \quad \mathbf{y} = \mathbf{A}\mathbf{x}$$

$\mathbf{x}^*$ solution iff there exists

$\mathbf{v} \perp \text{null}(\mathbf{A})$ and $\mathbf{v} \in \mathcal{C}^o \Leftrightarrow \mathbf{v} \in \partial\|\mathbf{x}^*\|$

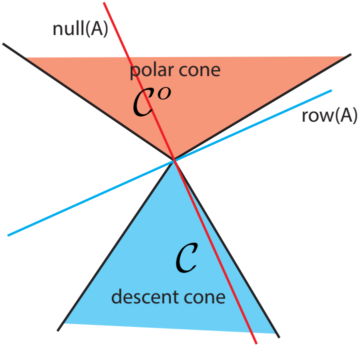


Dual certificates

$$\text{minimize} \quad \|\mathbf{x}\|_1 \quad \text{such that} \quad \mathbf{y} = \mathbf{A}\mathbf{x}$$

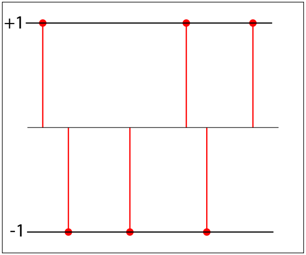
$\mathbf{x}^*$ solution iff there exists

$\mathbf{v} \perp \text{null}(\mathbf{A})$ and $\mathbf{v} \in \mathcal{C}^o \Leftrightarrow \mathbf{v} \in \partial\|\mathbf{x}^*\|$



Construction of dual certificate

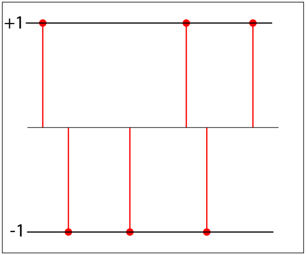
dual certificate $\quad \mathbf{v} \in \mathrm{row}(\mathbf{A}) \quad$ and $\quad \begin{cases} v_i = \mathrm{sgn}(x_i^*) & x_i^* \neq 0 \\ |v_i| < 1 & x_i^* = 0 \end{cases}$



$\mathrm{sgn}(\mathbf{x}^*)$

Construction of dual certificate

dual certificate $\quad \mathbf{v} \in \text{row}(\mathbf{A}) \quad$ and $\quad \begin{cases} v_i = \text{sgn}(x_i^*) & x_i^* \neq 0 \\ |v_i| < 1 & x_i^* = 0 \end{cases}$

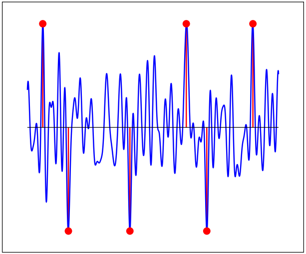


$\text{sgn}(\mathbf{x}^*)$

Least-squares solution to $\mathbf{v}_S = \text{sgn}(\mathbf{x})$:

Construction of dual certificate

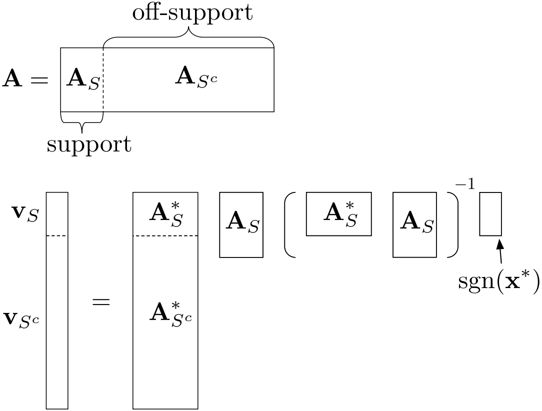
dual certificate $\quad \mathbf{v} \in \text{row}(\mathbf{A}) \quad$ and $\quad \begin{cases} v_i = \text{sgn}(x_i^*) & x_i^* \neq 0 \\ |v_i| < 1 & x_i^* = 0 \end{cases}$



$\text{sgn}(\mathbf{x}^*)$

$\mathbb{E}\, \mathbf{v}_{S^c} = \mathbf{0}$

Least-squares solution to $\mathbf{v}_S = \text{sgn}(\mathbf{x})$:



Sparse recovery guarantee

$$\text{minimize } \|\mathbf{x}\|_1 \ \text{ subject to } \ \mathbf{y} = \mathbf{A}\mathbf{x}$$

Assume:

- $\mathbf{x}$ is arbitrary N-dimensional S-sparse vector
- data vector $\mathbf{y}$ is M-dimensional with

$$M \geq S \log(N)$$

- $A_{ij} \sim \mathcal{N}(0, 1)$

Then, with high probability, l1 solution is exact!

Sparse recovery guarantee

$$\text{minimize } \|\mathbf{x}\|_1 \text{ subject to } \mathbf{y} = \mathbf{A}\mathbf{x}$$

Assume:

- $\mathbf{x}$ is arbitrary N-dimensional S-sparse vector
- data vector $\mathbf{y}$ is M-dimensional with

$$M \geq S \log(N)$$

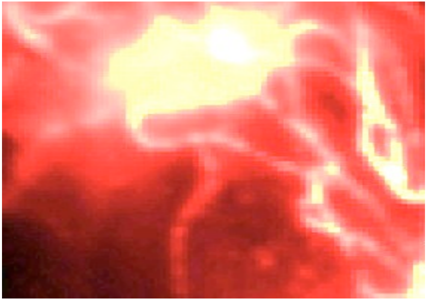
- $A_{ij} \sim \mathcal{N}(0, 1)$

Then, with high probability, l1 solution is exact!

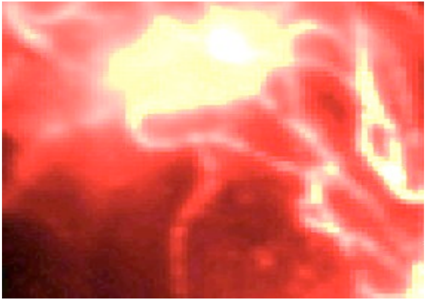
The **log** is needed to bound deviations of $\mathbf{v}_{S^c}$ around $\mathbb{E}\,\mathbf{v}_{S^c} = \mathbf{0}$.

Super-resolution microscopy

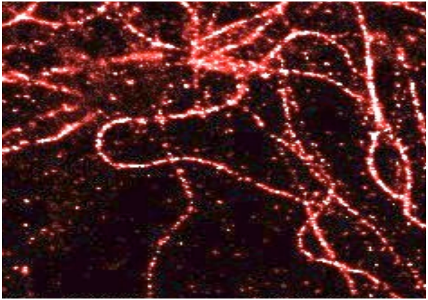
Abbe's diffraction limit for microscopy



Nobel Prize in Chemistry 2014

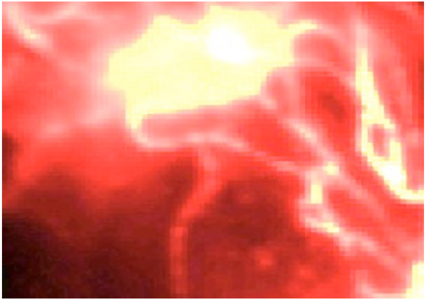


conventional microscopy

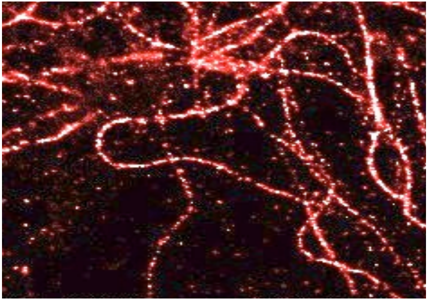


single-molecule microscopy

Nobel Prize in Chemistry 2014



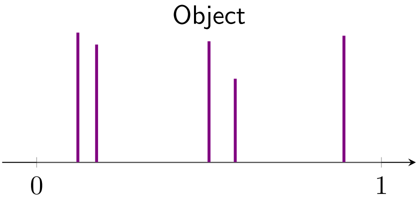
conventional microscopy

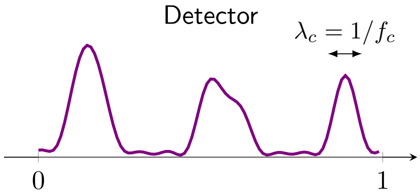


single-molecule microscopy

To make imaging faster, need a powerfull algorithm for sparse signal recovery problem!

Mathematical model



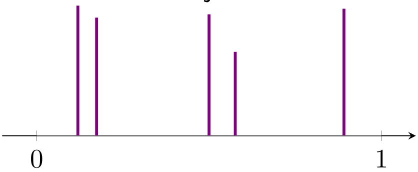


$$x(t) = \sum_s x_s \delta(t - t_s)$$

$$s(t) = (f_{\text{low}} \star x)(t) = \sum_s x_s f_{\text{low}}(t - t_s)$$

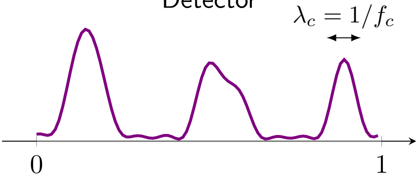
Mathematical model

Object



$$x(t) = \sum_s x_s \delta(t - t_s)$$

Detector



$$s(t) = (f_{\text{low}} \star x)(t) = \sum_s x_s f_{\text{low}}(t - t_s)$$

$$\mathbf{x} = [x_0 \cdots x_{N-1}]^{\mathsf{T}}$$

$\mathbf{x}$ is **sparse**

$$\mathbf{y} = \mathbf{A}\mathbf{x}$$

$\mathbf{A}$ ... $2f_c \times N$ low-frequency DFT

$$A_{kt} = e^{-i2\pi kt/N}, \ |k| \leq f_c$$

Columns of $\mathbf{A}$ are highly correlated

Solve:

minimize $\|\mathbf{x}\|_1$ subject to $\mathbf{y} = \mathbf{A}\mathbf{x}$

Question:

When does l1 work?

Columns of $\mathbf{A}$ are highly correlated

Solve:

$$\text{minimize } \|\mathbf{x}\|_1 \text{ subject to } \mathbf{y} = \mathbf{A}\mathbf{x}$$

Question:

When does l1 work?

First observation:

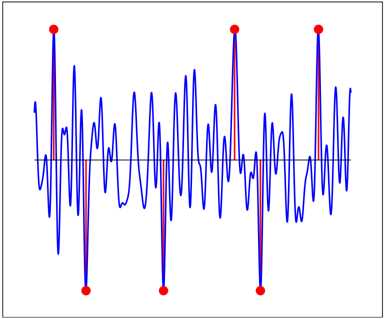
$$\mathbf{A} = [\mathbf{a}_1, \ldots \mathbf{a}_N] \quad 2f_c \times N$$

A_{kt} ... Gaussian:
$\langle \mathbf{a}_l, \mathbf{a}_{l+1} \rangle \approx \frac{1}{\sqrt{2f_c}}$

A_{kt} ... $e^{i2\pi kt}, |k| < f_c$
$\langle \mathbf{a}_l, \mathbf{a}_{l+1} \rangle \approx 1$

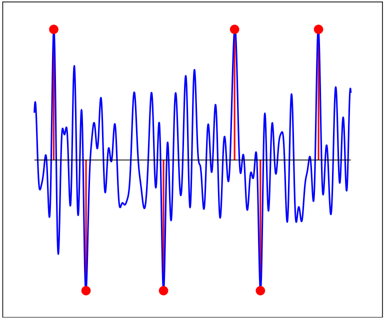
The dual polynomial for super-resolution

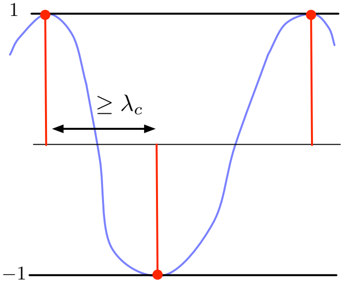
$$A_{kt} = e^{-i2\pi kt/N}, |k| \leq f_c \;\Rightarrow\; v(t) = \sum_{m=-f_c}^{f_c} \hat{v}_m e^{i2\pi mt}$$



The dual polynomial for super-resolution

$$A_{kt} = e^{-i2\pi kt/N}, |k| \leq f_c \;\Rightarrow\; v(t) = \sum_{m=-f_c}^{f_c} \hat{v}_m e^{i2\pi mt}$$



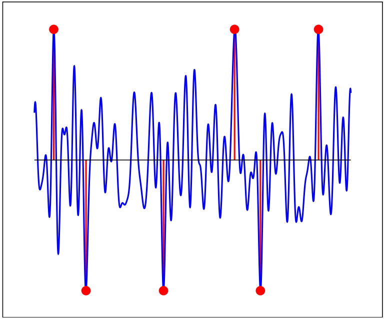


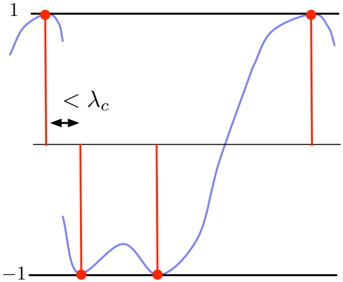
E. Candès and C. Fernandez-Granda '14

L1 works if spikes are further than $2\lambda_c$

The dual polynomial for super-resolution

$$A_{kt} = e^{-i2\pi kt/N}, |k| \leq f_c \;\Rightarrow\; v(t) = \sum_{m=-f_c}^{f_c} \hat{v}_m e^{i2\pi mt}$$

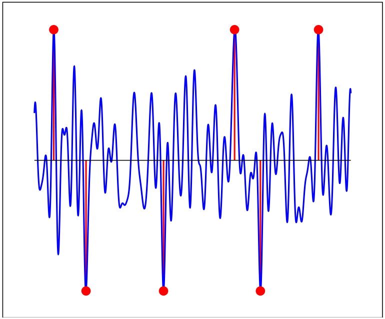


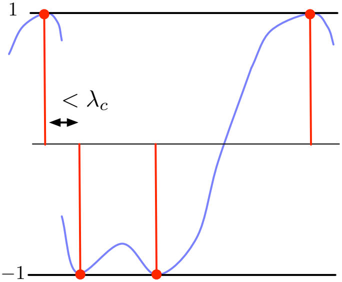


If spikes are closer than λ_c $\Rightarrow$ L1 breaks

The dual polynomial for super-resolution

$$A_{kt} = e^{-i2\pi kt/N}, |k| \le f_c \;\Rightarrow\; v(t) = \sum_{m=-f_c}^{f_c} \hat{v}_m e^{i2\pi mt}$$





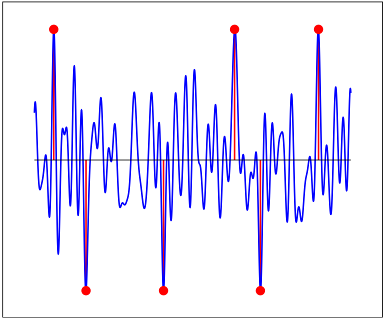
Bernstein theorem:

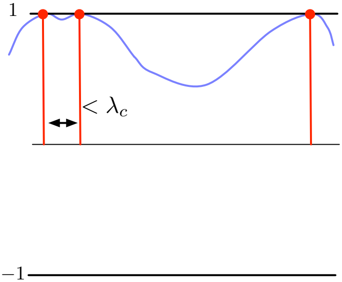
Consider: $v(t) = \sum_{k=-f_c}^{f_c} \hat{v}_k e^{-i2\pi kt}$ with $\left|v(t)\right| \le 1$ for all t

Then: $\left|v'(t)\right| \le 2f_c$ for all t.

The dual polynomial for super-resolution

$$A_{kt} = e^{-i2\pi kt/N}, |k| \leq f_c \;\Rightarrow\; v(t) = \sum_{m=-f_c}^{f_c} \hat{v}_m e^{i2\pi mt}$$





Donoho '92:

$\mathbf{x} \geq \mathbf{0} \;\Rightarrow$ L1 works if the number of spikes is less than $f_c + 1$

Super-resolution in the presence of noise

Model:

$$\mathbf{s} = f_{\text{low}} \star \hat{\mathbf{x}} + \mathbf{z}, \quad \|\mathbf{z}\|_1 \leq \delta$$

Solve:

$$\text{minimize } \|\mathbf{s} - f_{\text{low}} \star \hat{\mathbf{x}}\|_1 \quad \text{subject to} \quad \hat{\mathbf{x}} \geq 0$$

Theorem: [V. Morgenshtern and E. Candès, 2015]

Assume $\mathbf{x} \geq 0$, $\mathbf{x}$ is r-regular. Then,

$$\|\hat{\mathbf{x}} - \mathbf{x}\|_1 \leq \delta \left(\frac{N}{2f_c} \right)^{2r}.$$

Super-resolution in the presence of noise

Model:

$$\mathbf{s} = f_{\text{low}} \star \hat{\mathbf{x}} + \mathbf{z}, \quad \|\mathbf{z}\|_1 \leq \delta$$

Solve:

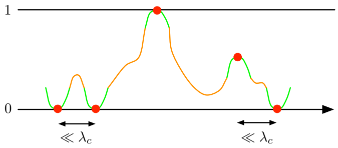
$$\text{minimize } \|\mathbf{s} - f_{\text{low}} \star \hat{\mathbf{x}}\|_1 \quad \text{subject to} \quad \hat{\mathbf{x}} \geq 0$$

Theorem: [V. Morgenshtern and E. Candès, 2015]

Assume $\mathbf{x} \geq 0$, $\mathbf{x}$ is r-regular. Then,

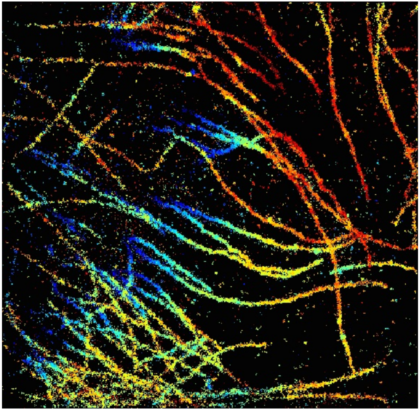
$$\|\hat{\mathbf{x}} - \mathbf{x}\|_1 \leq \delta \left(\frac{N}{2 f_c} \right)^{2r}.$$

Key novelty: a set of new tools in Fourier analysis



Reconstruction of 3D signals from 2D data

Preliminary result: **4 times faster than state-of-the-art**



10000 CVX problems solved
TFOCS first order solver
millions of variables

Radar imaging

Radar imaging

Recap:

- Dual certificate is a tool to analyse success of l1.
- Structure of $\mathbf{A}$ determines when certificate exists/does not exist.
- When A_{kl} are i.i.d. Gaussian, dual certificate is random. It exists if there are sufficiently many measurements.
- In the super-resolution problem, the certificate is deterministic low frequency trigonometric polynomial. It exists if the spikes are sufficiently separated.

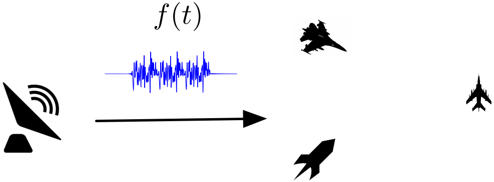
Radar imaging

Recap:

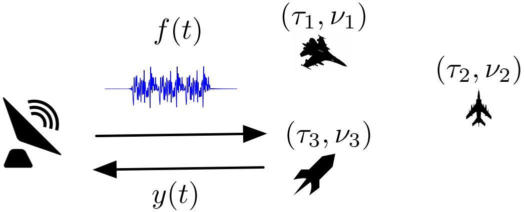
- Dual certificate is a tool to analyse success of l1.
- Structure of $\mathbf{A}$ determines when certificate exists/does not exist.
- When A_{kl} are i.i.d. Gaussian, dual certificate is random. It exists if there are sufficiently many measurements.
- In the super-resolution problem, the certificate is deterministic low frequency trigonometric polynomial. It exists if the spikes are sufficiently separated.

We will see: in radar, the certificate is random **and** it approximates a low frequency trigonometric polynomial.

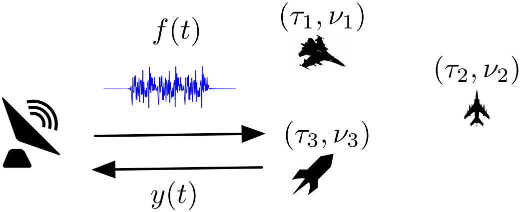
Mathematical model



Mathematical model

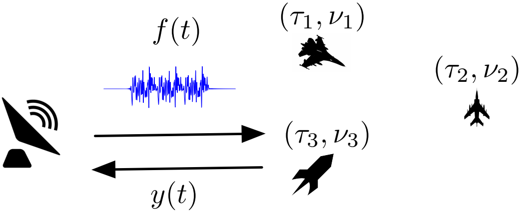


Mathematical model



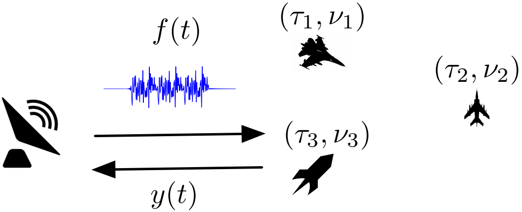
$$y(t) = \sum_{s=1}^{S} x_s (\mathcal{T}_{\tau_s} \mathcal{F}_{\nu_s} f)(t)$$

Mathematical model



$$
\begin{aligned}
y(t) &= \sum_{s=1}^{S} x_s (\mathcal{T}_{\tau_s} \mathcal{F}_{\nu_s} f)(t) \\
&= \sum_{s=1}^{S} x_s f(t - \tau_s) e^{i2\pi \nu_s t}
\end{aligned}
$$

Mathematical model



$$y(t) = \sum_{s=1}^{S} x_s (\mathcal{T}_{\tau_s} \mathcal{F}_{\nu_s} f)(t)$$
$$= \sum_{s=1}^{S} x_s f(t - \tau_s) e^{i2\pi\nu_s t}$$

Goal: recover (x_s, τ_s, ν_s)

Time and bandwidth limitations

In practice:

- $f(t)$ is bandlimited to $B\,\mathrm{Hz}$
- $y(t)$ is observed over $T\,\mathrm{sec}$
- $\Rightarrow y(t)$ is BT-dimensional

Time and bandwidth limitations

In practice:

- $f(t)$ is bandlimited to $B\,\mathrm{Hz}$
- $y(t)$ is observed over $T\,\mathrm{sec}$
- $\Rightarrow y(t)$ is BT-dimensional

Goal: estimate τ and ν with precision higher than $1/B$ and $1/T$

$$y(t) = \sum_{s=1}^{S} x_s f(t-\tau_s) e^{i2\pi\nu_s t}$$

Time and bandwidth limitations

In practice:

- $f(t)$ is bandlimited to $B\,\mathrm{Hz}$
- $y(t)$ is observed over $T\,\mathrm{sec}$
- $\Rightarrow y(t)$ is BT-dimensional

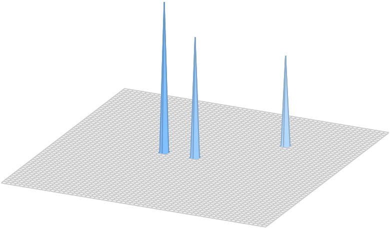
Goal: estimate τ and ν with precision higher than $1/B$ and $1/T$

$$y(t) = \sum_{s=1}^{S} x_s f(t - \tau_s) \qquad \text{(super-resolution)}$$

Blurring of time and frequency shifts

input

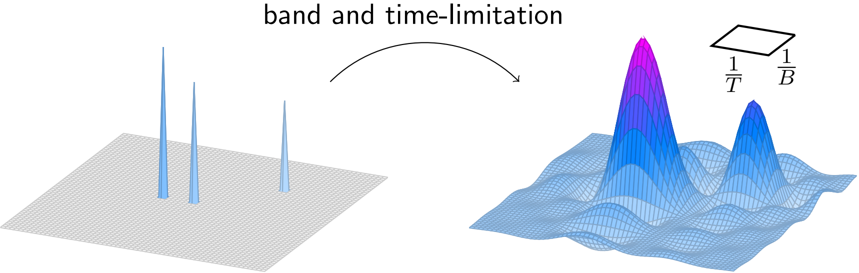
$$h(\tau, \nu) = \sum_s b_s \delta(\tau - \tau_s) \delta(\nu - \nu_s)$$



Blurring of time and frequency shifts

input

$$h(\tau,\nu) = \sum_s b_s \delta(\tau - \tau_s)\delta(\nu - \nu_s)$$



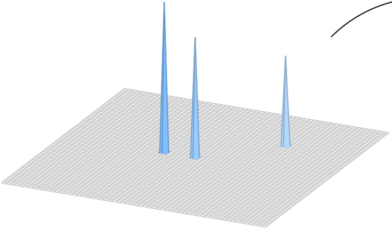
$$\operatorname{sinc}(\tau B)\operatorname{sinc}(\nu T) * h(\tau,\nu)$$

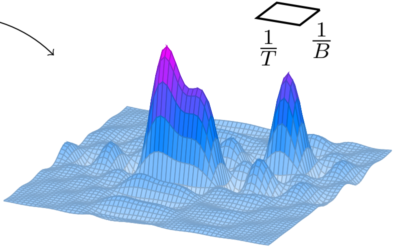
Blurring of time and frequency shifts

input

$$h(\tau,\nu) = \sum_s b_s \delta(\tau - \tau_s)\delta(\nu - \nu_s)$$

band and time-limitation





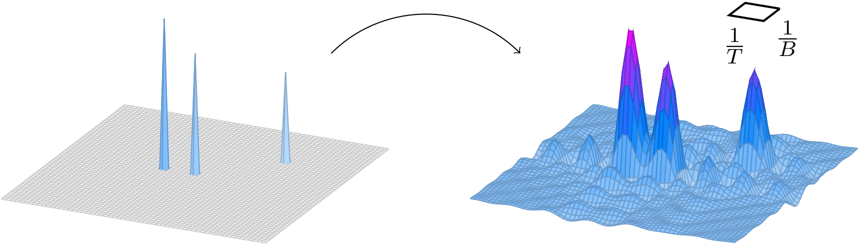
$$\mathrm{sinc}(\tau B)\,\mathrm{sinc}(\nu T) * h(\tau,\nu)$$

Blurring of time and frequency shifts

input

$$h(\tau,\nu) = \sum_s b_s \delta(\tau - \tau_s)\delta(\nu - \nu_s)$$

band and time-limitation



$$\operatorname{sinc}(\tau B)\operatorname{sinc}(\nu T) * h(\tau,\nu)$$

Blurring of time and frequency shifts

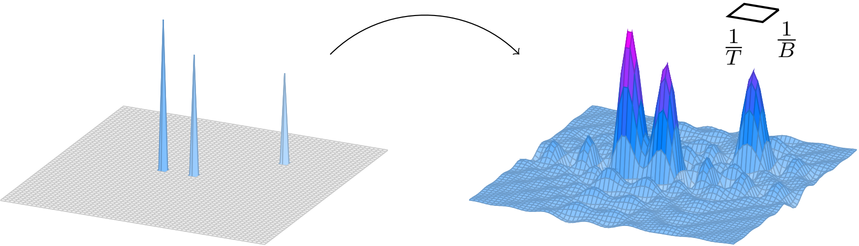
input

$$h(\tau,\nu) = \sum_s b_s \delta(\tau - \tau_s)\delta(\nu - \nu_s)$$

output

$$y(t) = \iint h(\tau,\nu) f(t-\tau) e^{i2\pi\nu t} d\tau d\nu$$

band and time-limitation



$$\mathrm{sinc}(\tau B)\,\mathrm{sinc}(\nu T) * h(\tau,\nu)$$

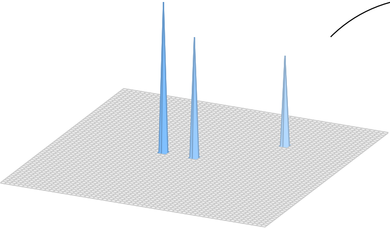
Blurring of time and frequency shifts

input

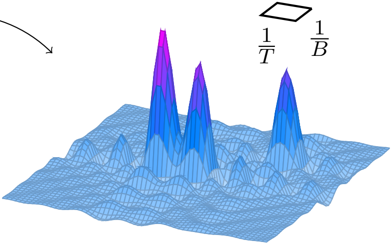
$$h(\tau,\nu) = \sum_s b_s \delta(\tau - \tau_s)\delta(\nu - \nu_s)$$

output

$$y(t) = \iint h(\tau,\nu) f(t-\tau) e^{i2\pi\nu t} d\tau d\nu$$



band and time-limitation



$$\mathrm{sinc}(\tau B)\,\mathrm{sinc}(\nu T) * h(\tau,\nu)$$

Resolution achieved by classic Radar via matched filtering is $\left(\frac{1}{B}, \frac{1}{T}\right)$!

Main result

Notation:

- $\mathbf{y}$, $\mathbf{f}$ contain samples of $y(t)$, $f(t)$ at the rate $1/B$
- Columns of $\mathbf{A}$ are $\mathcal{T}_\tau \mathcal{F}_\nu \mathbf{f}$ (indexed by τ and ν):

$$\mathbf{A} = \boxed{\mathcal{F}_\nu \mathcal{T}_\tau \mathbf{x}} \updownarrow BT$$

$$\longleftrightarrow$$

$$\gg (BT)^2$$

- Random probing signal: f_ℓ i.i.d. $\mathcal{N}(0,1)$
- $\mathbf{x}$ contains x_s at location indexed by τ and ν

Solve:

$$\text{minimize } \|\mathbf{x}\|_1 \quad \text{subject to} \quad \mathbf{y} = \mathbf{A}\mathbf{x}$$

Main result

Solve:

$$\text{minimize } \|\mathbf{x}\|_1 \quad \text{subject to} \quad \mathbf{y} = \mathbf{A}\mathbf{x}$$

Theorem: [Heckel, Morgenshtern, Soltanolkotabi '15]

Assume:

$$|\tau_s - \tau_r| \geq \frac{5}{B} \text{ or } |\nu_s - \nu_r| \geq \frac{5}{T}, \text{ for all } s \neq r$$

and

$$S \lesssim BT \log^{-3}(BT).$$

Then: with high probability, l1 minimization recovers $\mathbf{x}$ exactly.
Hence, (τ_s, ν_s, x_s) are recovered perfectly.

Key novelty: dual polynomial for radar

Recall:

$$\mathbf{A} = \boxed{\mathcal{F}_\nu \mathcal{T}_\tau \mathbf{x}} \updownarrow BT$$

$$\longleftrightarrow \gg (BT)^2$$

Need:

$$v(\tau, \nu) = [\mathcal{F}_\nu \mathcal{T}_\tau \mathbf{x}]^H \hat{\mathbf{v}}$$

Key novelty: dual polynomial for radar

Recall:

$$\mathbf{A} = \boxed{\mathcal{F}_\nu \mathcal{T}_\tau \mathbf{x}} \updownarrow BT$$

$$\longleftrightarrow \gg (BT)^2$$

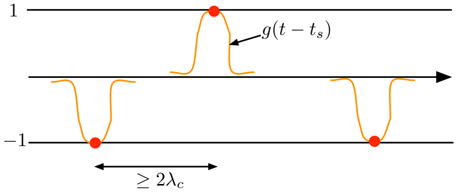
Need:

$$v(\tau, \nu) = [\mathcal{F}_\nu \mathcal{T}_\tau \mathbf{x}]^H \hat{\mathbf{v}}$$

Ingredient: dual certificate for super-resolution (with separation) [Candes and Fernandez-Granda '14]

$$v(t) = \sum_s c_s g(t - t_s) + \text{corrections}$$

Low pass and concentrated kernel: $g(t) = \sum_{k=-f_c}^{f_c} \hat{g}_k e^{i2\pi kt}$



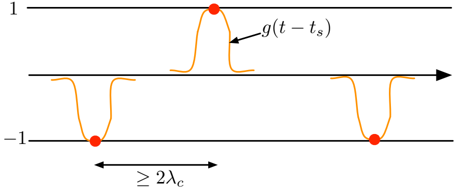
Key novelty: dual polynomial for radar

Need: $v(\tau,\nu) = [\mathcal{F}_\nu \mathcal{T}_\tau \mathbf{x}]^{\mathsf{H}} \hat{\mathbf{v}}$

Ingredient:

$$v(t) = \sum_s c_s g(t - t_s) + \text{corrections}$$

low pass and concentrated kernel: $g(t) = \sum_{k=-f_c}^{f_c} \hat{g}_k e^{i2\pi kt}$



For radar:

$$v(\tau,\nu) = \sum_s c_s g_{\tau_s,\nu_s}(\tau,\nu) + \text{corrections}$$

$$g_{\tau_s,\nu_s}(\tau,\nu) \ \textbf{“resembles”} \ g(\cdot - \tau_s) \times g(\cdot - \nu_s)$$

Key novelty: construction of $g_{\tau_s,\nu_s}(\cdot,\cdot)$

Kernel:

$$g_{\tau_s,\nu_s}(\tau,\nu) = \left[\mathcal{F}_\nu \mathcal{T}_\tau \mathbf{x}\right]^H \hat{\mathbf{v}}$$

Key novelty: construction of $g_{\tau_s,\nu_s}(\cdot,\cdot)$

Kernel:

$$g_{\tau_s,\nu_s}(\tau,\nu) = [\mathcal{F}_\nu \mathcal{T}_\tau \mathbf{x}]^H \, \hat{\mathbf{v}}$$

We can write:

$$(\mathcal{F}_\nu \mathcal{T}_\tau \mathbf{x})^{\mathsf{H}} = \left[\cdots e^{i2\pi(\tau r + \nu q)} \cdots\right] \mathbf{F}\mathbf{G}^{\mathsf{H}}$$

$\mathbf{F}$ is 2D DFT matrix

$$\mathbf{G} = \boxed{\mathcal{F}_{\frac{r}{BT}} \mathcal{T}_{\frac{q}{BT}} \mathbf{x}} \updownarrow BT$$

$$\longleftrightarrow$$
$$(BT)^2$$

Key novelty: construction of $g_{\tau_s,\nu_s}(\cdot,\cdot)$

Kernel:

$$g_{\tau_s,\nu_s}(\tau,\nu) = [\mathcal{F}_\nu \mathcal{T}_\tau \mathbf{x}]^H \, \hat{\mathbf{v}}$$

We can write:

$$(\mathcal{F}_\nu \mathcal{T}_\tau \mathbf{x})^{\mathsf{H}} = \left[\cdots e^{i2\pi(\tau r + \nu q)} \cdots \right] \mathbf{F}\mathbf{G}^{\mathsf{H}}$$

$\mathbf{F}$ is 2D DFT matrix

$$\mathbf{G} = \boxed{\mathcal{F}_{\frac{r}{BT}} \mathcal{T}_{\frac{q}{BT}} \mathbf{x}} \updownarrow BT, \quad \longleftrightarrow (BT)^2$$

Choose coefficients:

$$\hat{\mathbf{v}} = \mathbf{G}\mathbf{F}^{\mathsf{H}} \left[\cdots \hat{g}_r \hat{g}_q e^{-i2\pi(\tau r + \nu q)} \cdots \right]^{\mathsf{T}}$$

Key novelty: construction of $g_{\tau_s,\nu_s}(\cdot,\cdot)$

Kernel:

$$g_{\tau_s,\nu_s}(\tau,\nu) = [\mathcal{F}_\nu \mathcal{T}_\tau \mathbf{x}]^H \, \hat{\mathbf{v}}$$

We can write:

$$(\mathcal{F}_\nu \mathcal{T}_\tau \mathbf{x})^{\mathsf{H}} = \left[\cdots e^{i2\pi(\tau r+\nu q)} \cdots\right] \mathbf{F}\mathbf{G}^{\mathsf{H}}$$

$\mathbf{F}$ is 2D DFT matrix

$$\mathbf{G} = \boxed{\mathcal{F}_{\frac{r}{BT}} \mathcal{T}_{\frac{q}{BT}} \mathbf{x}} \updownarrow BT \qquad \longleftrightarrow (BT)^2$$

Choose coefficients:

$$\hat{\mathbf{v}} = \mathbf{G}\mathbf{F}^{\mathsf{H}} \left[\cdots \hat{g}_r \hat{g}_q e^{-i2\pi(\tau r+\nu q)} \cdots\right]^{\mathsf{T}}$$

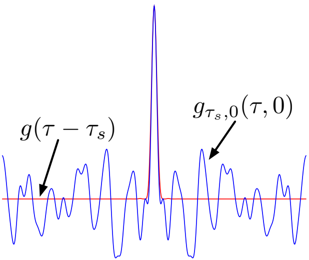
Observe: $\mathbb{E}\left[\mathbf{F}\mathbf{G}^{\mathsf{H}}\mathbf{G}\mathbf{F}^{\mathsf{H}}\right] = \mathbf{F}\,\mathbb{E}\left[\mathbf{G}^{\mathsf{H}}\mathbf{G}\right]\mathbf{F}^{\mathsf{H}} = \mathbf{I}$

Therefore: $\mathbb{E}\left[g_{\tau_s,0}(\tau,0)\right] = \sum_{k=-T}^{T} \hat{g}_k e^{i2\pi k(\tau-\tau_j)} = g(\tau-\tau_j)$

Random kernel approximates deterministic kernel

Observe: $\mathbb{E}\left[\mathbf{F}\mathbf{G}^{\mathsf{H}}\mathbf{G}\mathbf{F}^{\mathsf{H}} = \mathbf{I}\right]$

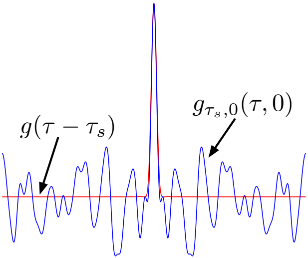
Therefore: $\mathbb{E}\left[g_{\tau_s,0}(\tau,0)\right] = \sum_{k=-T}^{T} \hat{g}_k e^{i2\pi k(\tau-\tau_j)} = g(\tau-\tau_j)$

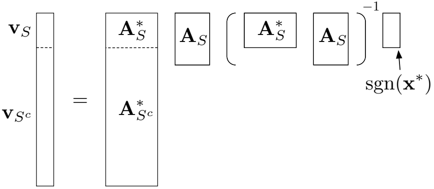


Random kernel approximates deterministic kernel

Observe: $\mathbb{E}\left[\mathbf{F}\mathbf{G}^{\mathsf{H}}\mathbf{G}\mathbf{F}^{\mathsf{H}} = \mathbf{I}\right]$

Therefore: $\mathbb{E}\left[g_{\tau_s,0}(\tau,0)\right] = \sum_{k=-T}^{T} \hat{g}_k e^{i2\pi k(\tau-\tau_j)} = g(\tau-\tau_j)$



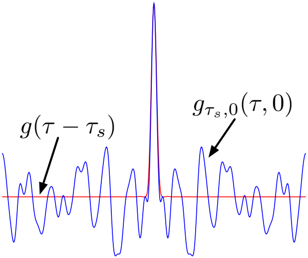


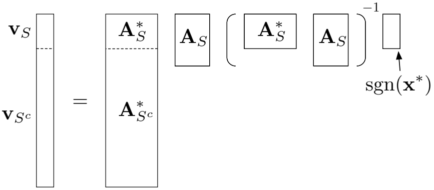
$$\mathbb{E}\,\mathbf{v}_{S^c} = \mathbf{0}$$

Random kernel approximates deterministic kernel

Observe: $\mathbb{E}\left[\mathbf{F}\mathbf{G}^{\mathsf{H}}\mathbf{G}\mathbf{F}^{\mathsf{H}} = \mathbf{I}\right]$

Therefore: $\mathbb{E}\left[g_{\tau_s,0}(\tau,0)\right] = \sum_{k=-T}^{T} \hat{g}_k e^{i2\pi k(\tau-\tau_j)} = g(\tau-\tau_j)$



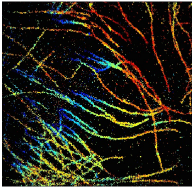


$$\mathbb{E}\,\mathbf{v}_{S^c} = \mathbf{0}$$

Now we can use:

$$v(\tau,\nu) = \sum_s c_s g_{\tau_s,\nu_s}(\tau,\nu) + \text{corrections}$$

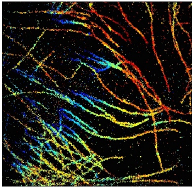
Mathematics of information



super-resolution microscopy

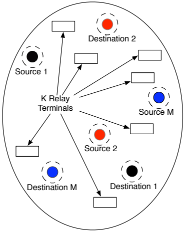
radar imaging

Mathematics of information

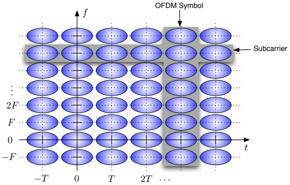


super-resolution microscopy

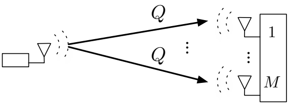
radar imaging



large wireless
networks

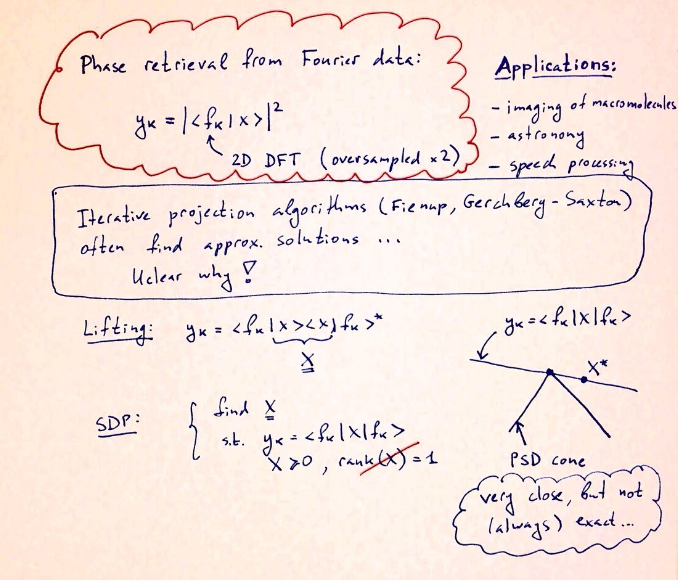


communication
under
channel uncertainty



systems
with
multiple receive antennas

Related open problems



Thank you