NYU COURANT INSTITUTE OF MATHEMATICAL SCIENCES

Overview

Optimization-Based Data Analysis

http://www.cims.nyu.edu/~cfgranda/pages/OBDA_spring16

Carlos Fernandez-Granda

1/25/2016

Sparsity

- Denoising
- Regression
- Inverse problems

Low-rank models

- Matrix completion
- Low rank + sparse model
- Nonnegative matrix factorization

Sparsity

- Denoising
- Regression
- Inverse problems

Low-rank models

- Matrix completion
- Low rank + sparse model
- Nonnegative matrix factorization

Denoising

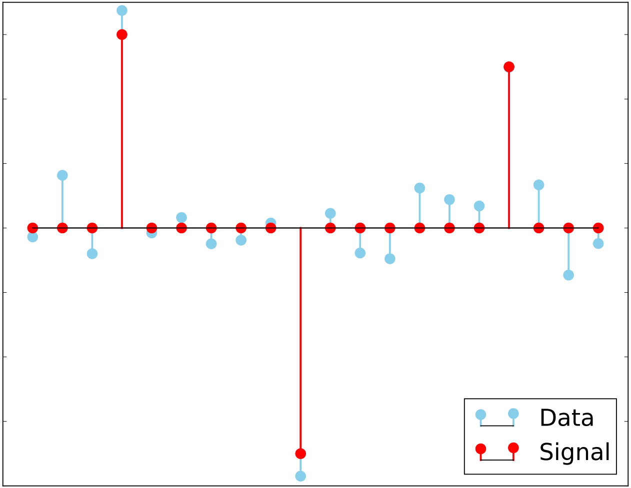
Additive noise model

$$\text{data} = \text{signal} + \text{noise}$$

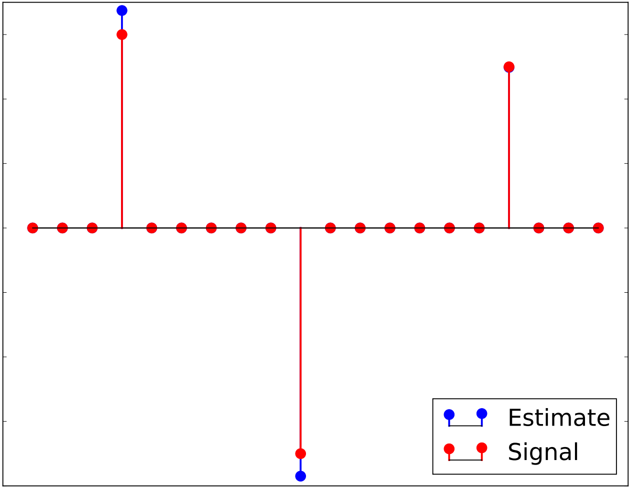
Hard thresholding

$$\mathcal{H}_\eta(x)_i := \begin{cases} x_i & \text{if } |x_i| > \eta, \\ 0 & \text{otherwise} \end{cases}$$

Denoising via thresholding



Denoising via thresholding

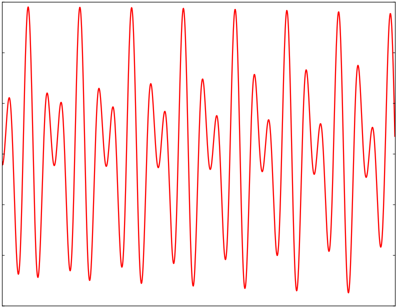


Sparsifying transforms

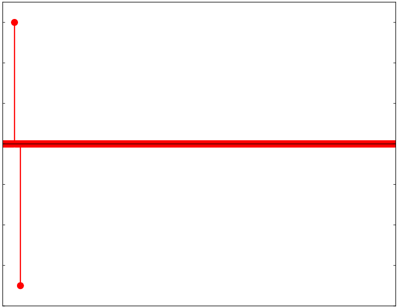
$$x = Dc = \sum_{i=1}^{n} D_i \, c_i,$$

Discrete cosine transform

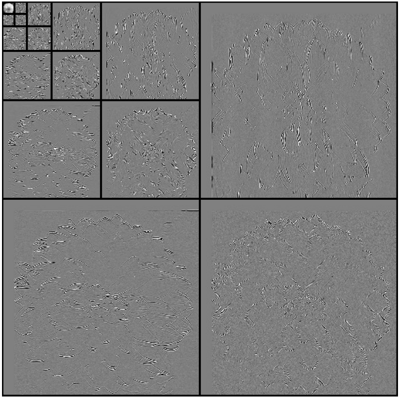
Signal



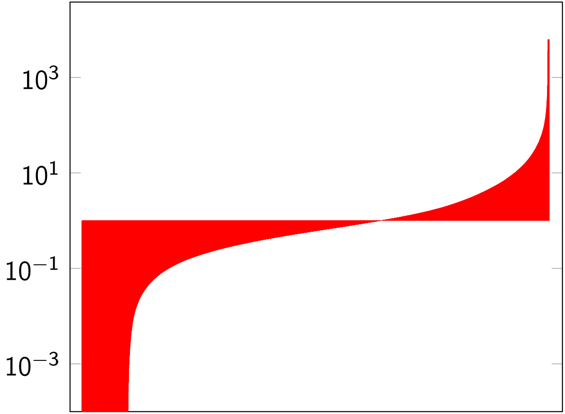
DCT coefficients



Wavelets



Sorted wavelet coefficients



Denoising via thresholding in a basis

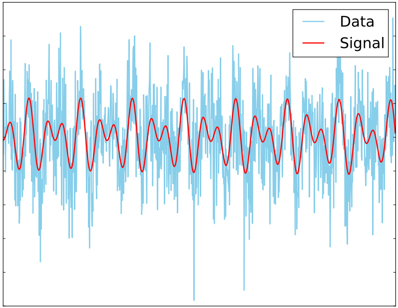
$$x = Bc$$

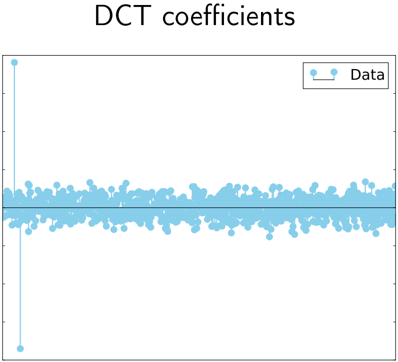
$$y = Bc + z$$

$$\hat{c} = \mathcal{H}_\eta \left(B^{-1} y \right)$$

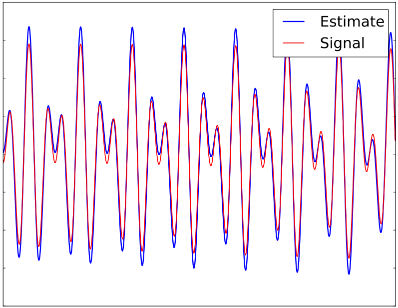
$$\hat{y} = B\hat{c}$$

Denoising via thresholding in a DCT basis

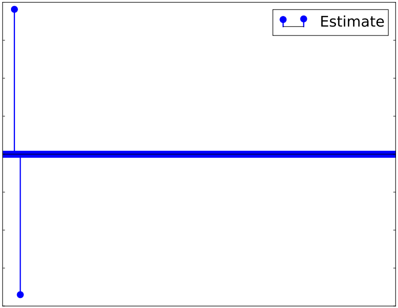




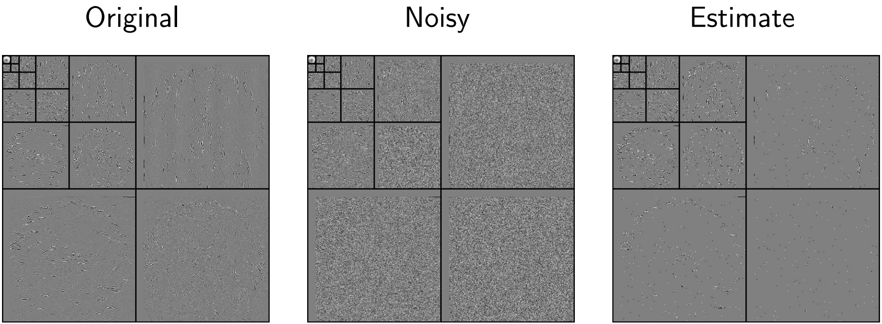
Denoising via thresholding in a DCT basis



DCT coefficients



Denoising via thresholding in a wavelet basis



Denoising via thresholding in a wavelet basis

Original

Noisy

Estimate

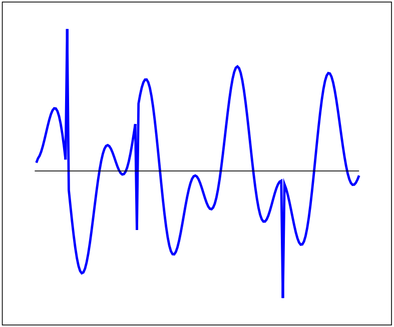
Denoising via thresholding in a wavelet basis

Original

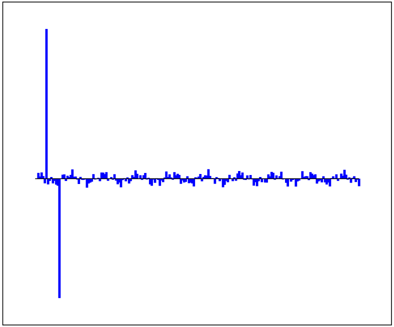
Noisy

Estimate

Overcomplete dictionary



DCT coefficients



Overcomplete dictionary

$$x = Dc = \begin{bmatrix} A & B \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = Aa + Bb$$

Sparsity estimation

First idea:

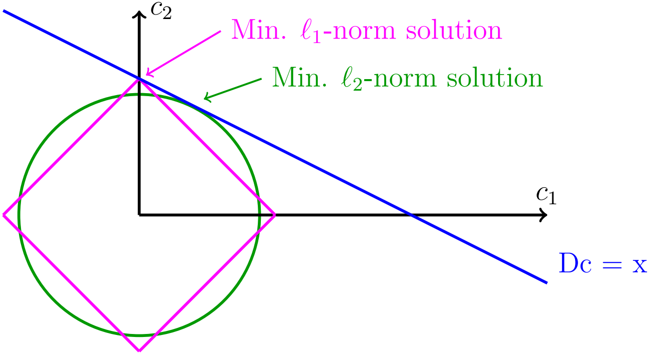
$$\min_{\tilde{c} \in \mathbb{R}^m} ||\tilde{c}||_0 \quad \text{such that } x = Dc$$

Computationally intractable!

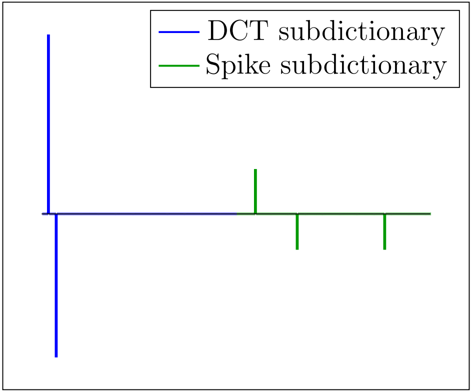
Tractable alternative:

$$\min_{\tilde{c} \in \mathbb{R}^m} ||\tilde{c}||_1 \quad \text{such that } x = Dc$$

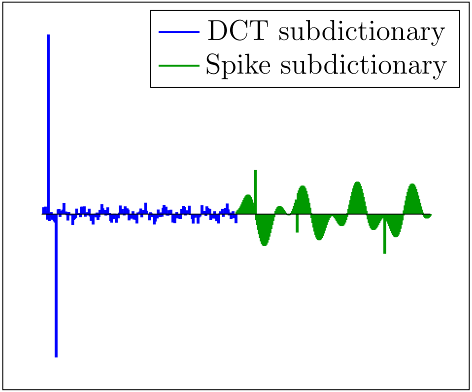
Geometric intuition



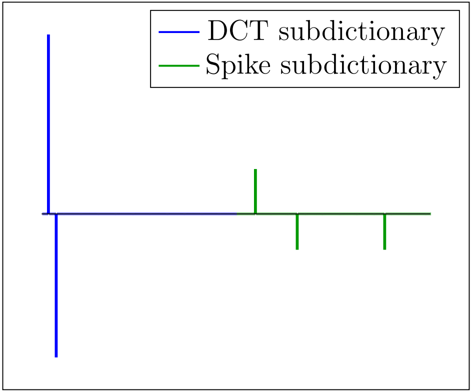
Sparsity estimation



Minimum ℓ_2-norm coefficients



Minimum ℓ_1-norm coefficients

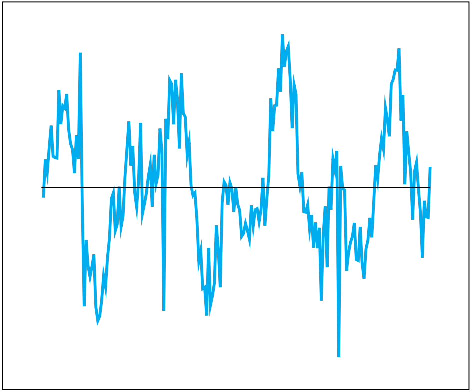


Denoising via ℓ_1-norm regularized least squares

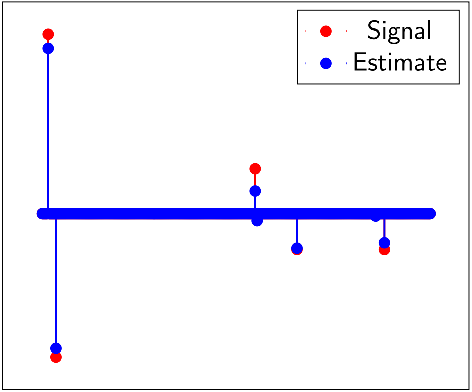
$$\hat{c} = \arg\min_{\tilde{c} \in \mathbb{R}^m} ||x - D\tilde{c}||_2^2 + \lambda ||\tilde{c}||_1$$

$$\hat{x} = D\hat{c}$$

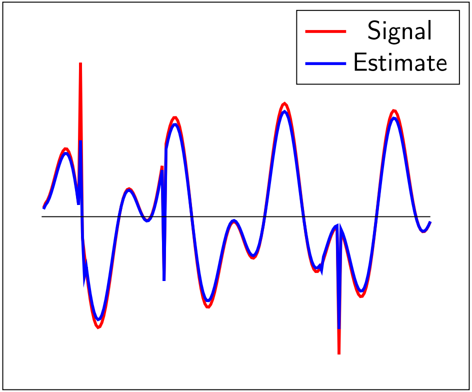
Denoising via ℓ_1-norm regularized least squares



Denoising via ℓ_1-norm regularized least squares



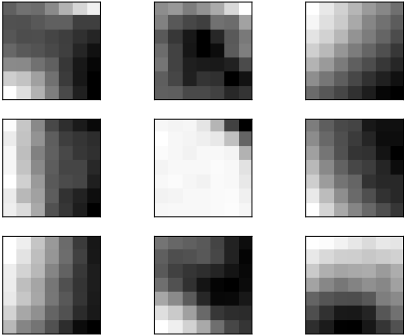
Denoising via ℓ_1-norm regularized least squares



Learning the dictionary

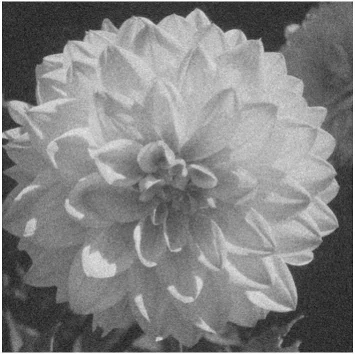
$$\min_{\widetilde{C} \in \mathbb{R}^{m \times k}} \left|\left| X - \widetilde{D}\widetilde{C} \right|\right|_{\mathrm{F}}^{2} + \lambda \left|\left| \widetilde{c} \right|\right|_{1} \quad \text{such that} \quad \left|\left| \widetilde{D}_{i} \right|\right|_{2} = 1, \quad 1 \leq i \leq m$$

Dictionary learning



Denoising via dictionary learning

Denoising via dictionary learning



Denoising via dictionary learning

Sparsity

Denoising

Regression

Inverse problems

Low-rank models

Matrix completion

Low rank + sparse model

Nonnegative matrix factorization

Linear regression

$$y_i \approx \sum_{j=1}^{p} \theta_j X_{ij}, \quad 1 \leq i \leq n$$

$$y \approx X\theta$$

Sparse regression

Least squares

$$\hat{\theta}_{\text{ls}} := \arg\min_{\tilde{\theta} \in \mathbb{R}^n} \left|\left| y - X\tilde{\theta} \right|\right|_2$$

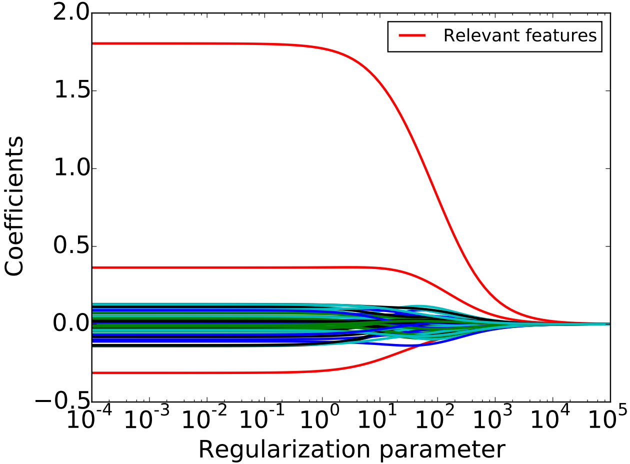
Ridge regression (aka Tikhonov regularization)

$$\hat{\theta}_{\text{ridge}} := \arg\min_{\tilde{\theta} \in \mathbb{R}^n} \left|\left| y - X\tilde{\theta} \right|\right|_2^2 + \lambda \left|\left| \tilde{\theta} \right|\right|_2^2$$

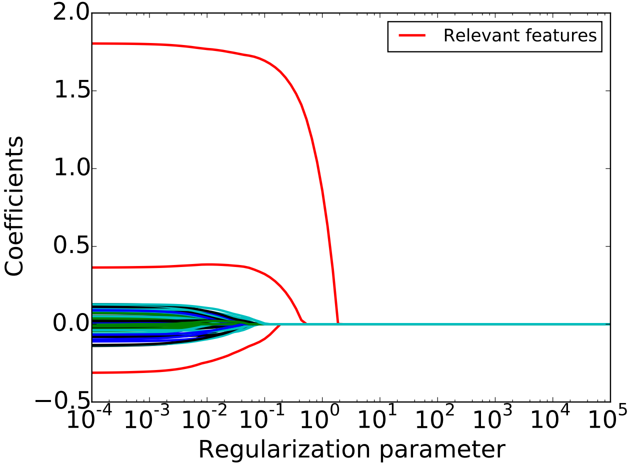
The lasso

$$\hat{\theta}_{\text{lasso}} := \arg\min_{\tilde{\theta} \in \mathbb{R}^n} \left|\left| y - X\tilde{\theta} \right|\right|_2^2 + \lambda \left|\left| \tilde{\theta} \right|\right|_1$$

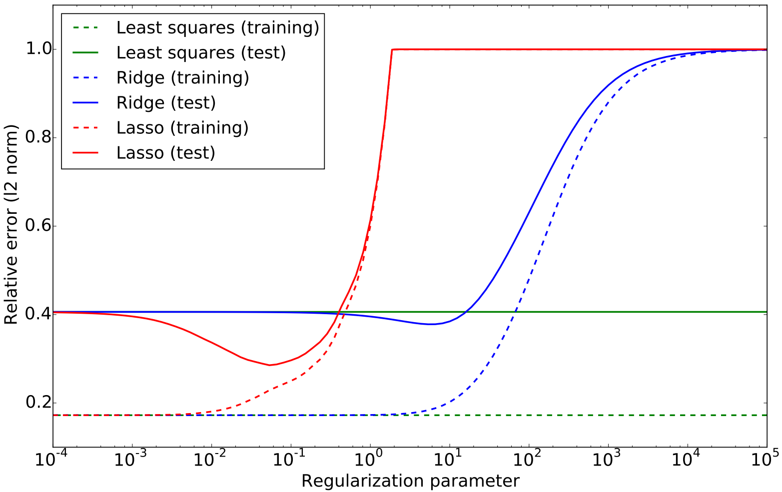
Ridge-regression coefficients



Lasso coefficients



Results


Least squares (training)
Least squares (test)
Ridge (training)
Ridge (test)
Lasso (training)
Lasso (test)
Relative error (l2 norm)
Regularization parameter
1.0
0.8
0.6
0.4
0.2
10^{-4}
10^{-3}
10^{-2}
10^{-1}
10^{0}
10^{1}
10^{2}
10^{3}
10^{4}
10^{5}

Sparsity

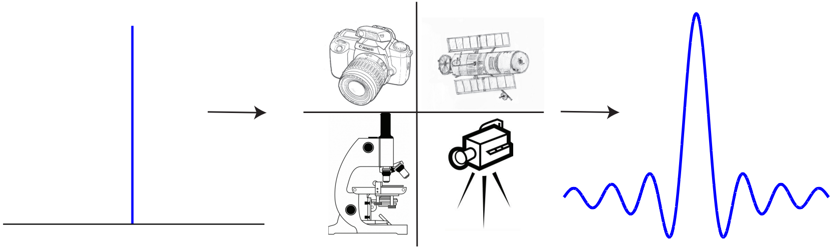
- Denoising
- Regression
- Inverse problems

Low-rank models

- Matrix completion
- Low rank + sparse model
- Nonnegative matrix factorization

Limits of resolution in imaging

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

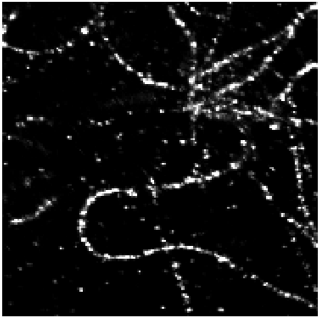


$\delta(t-\tau)$ optical system $h(t-\tau)$

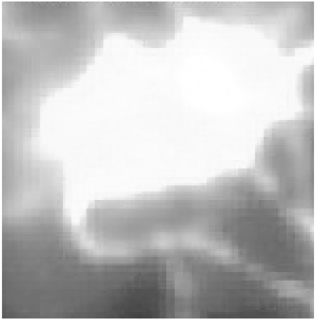
Diffraction imposes a fundamental limit on the resolution of optical systems

Fluorescence microscopy

Data



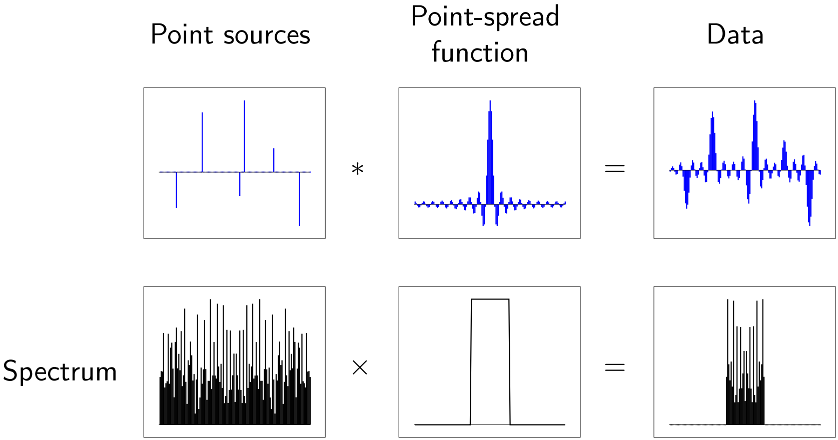
Point sources



Low-pass blur

(Figures courtesy of V. Morgenshtern)

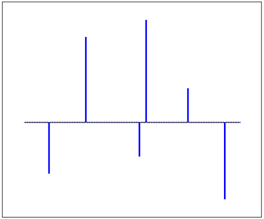
Sensing model for super-resolution



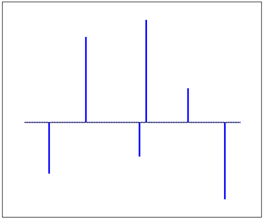
Minimum ℓ_1-norm estimate

$$\begin{array}{ll} \textit{minimize} & ||\text{estimate}||_1 \\ \textit{subject to} & \text{estimate} * \text{psf} = \text{data} \end{array}$$

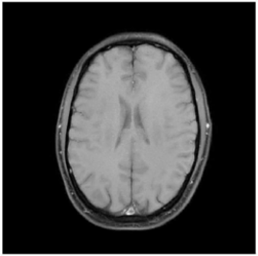
Point sources

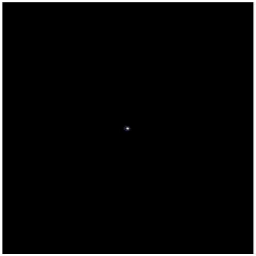


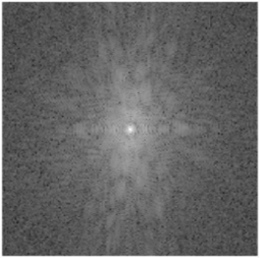
Estimate



Magnetic resonance imaging







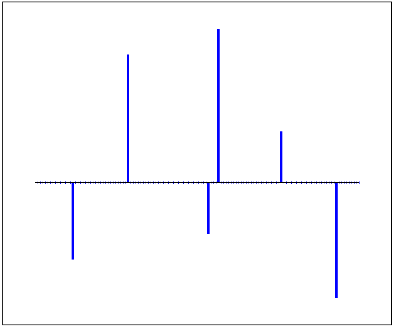
Magnetic resonance imaging

- Data: Samples from spectrum
- Problem: Sampling is time consuming (patients get annoyed, kids move during data acquisition)
- Images are compressible ($\approx$ sparse)
- Can we recover compressible signals from less data?

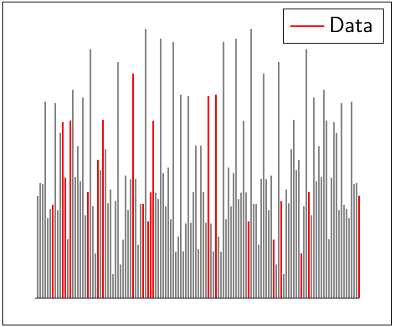
Compressed sensing

1. Undersample the spectrum randomly

Signal



Spectrum

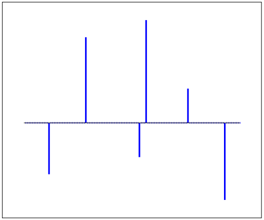


Compressed sensing

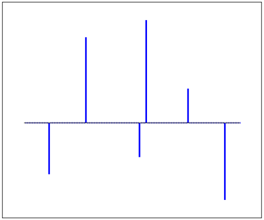
2. Solve the optimization problem

$$\begin{array}{ll} \textit{minimize} & ||\text{estimate}||_1 \\ \textit{subject to} & \text{frequency samples of estimate} = \text{data} \end{array}$$

Signal

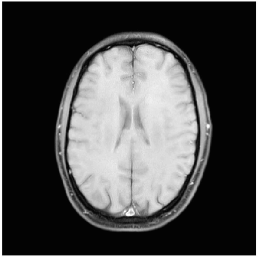


Estimate

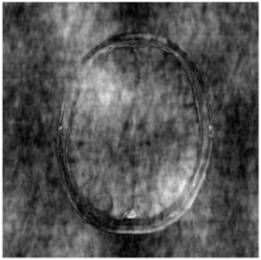


Compressed sensing in MRI

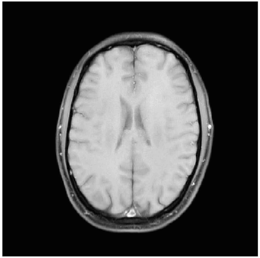
Original



Min. ℓ_2-norm estimate



Min. ℓ_1-norm estimate



Sparsity

- Denoising
- Regression
- Inverse problems

Low-rank models

- Matrix completion
- Low rank + sparse model
- Nonnegative matrix factorization

Sparsity

Denoising

Regression

Inverse problems

Low-rank models

Matrix completion

Low rank + sparse model

Nonnegative matrix factorization

Netflix Prize

Collaborative filtering

$$A := \begin{pmatrix} 1 & 1 & 5 & 4 \\ 2 & 1 & 4 & 5 \\ 4 & 5 & 2 & 1 \\ 5 & 4 & 2 & 1 \\ 4 & 5 & 1 & 2 \\ 1 & 2 & 5 & 5 \end{pmatrix}$$

	Bob	Molly	Mary	Larry
The Dark Knight	1	1	5	4
Spiderman 3	2	1	4	5
Love Actually	4	5	2	1
Bridget Jones's Diary	5	4	2	1
Pretty Woman	4	5	1	2
Superman 2	1	2	5	5

Centering

$$\mu := \frac{1}{n} \sum_{i=1}^{m} \sum_{j=1}^{n} A_{ij},$$

$$\bar{A} := \begin{bmatrix} \mu & \mu & \cdots & \mu \\ \mu & \mu & \cdots & \mu \\ \cdots & \cdots & \cdots & \cdots \\ \mu & \mu & \cdots & \mu \end{bmatrix}$$

SVD

$$A - \bar{A} = USV^T = U \begin{bmatrix} 7.79 & 0 & 0 & 0 \\ 0 & 1.62 & 0 & 0 \\ 0 & 0 & 1.55 & 0 \\ 0 & 0 & 0 & 0.62 \end{bmatrix} V^T$$

First left singular vector

$$U_1 = \begin{pmatrix} \text{D. Knight} & \text{Sp. 3} & \text{Love Act.} & \text{B.J.'s Diary} & \text{P. Woman} & \text{Sup. 2} \\ -0.45 & -0.39 & 0.39 & 0.39 & 0.39 & -0.45 \end{pmatrix}$$

First right singular vector

	Bob	Molly	Mary	Larry
$V_1 =$	(0.48	0.52	−0.48	−0.52)

Rank 1 model

$$\bar{A} + \sigma_1 U_1 V_1^T = \begin{pmatrix} 1.34\,(1) & 1.19\,(1) & 4.66\,(5) & 4.81\,(4) \\ 1.55\,(2) & 1.42\,(1) & 4.45\,(4) & 4.58\,(5) \\ 4.45\,(4) & 4.58\,(5) & 1.55\,(2) & 1.42\,(1) \\ 4.43\,(5) & 4.56\,(4) & 1.57\,(2) & 1.44\,(1) \\ 4.43\,(4) & 4.56\,(5) & 1.57\,(1) & 1.44\,(2) \\ 1.34\,(1) & 1.19\,(2) & 4.66\,(5) & 4.81\,(5) \end{pmatrix}$$

	Bob	Molly	Mary	Larry
The Dark Knight	1.34 (1)	1.19 (1)	4.66 (5)	4.81 (4)
Spiderman 3	1.55 (2)	1.42 (1)	4.45 (4)	4.58 (5)
Love Actually	4.45 (4)	4.58 (5)	1.55 (2)	1.42 (1)
B.J.'s Diary	4.43 (5)	4.56 (4)	1.57 (2)	1.44 (1)
Pretty Woman	4.43 (4)	4.56 (5)	1.57 (1)	1.44 (2)
Superman 2	1.34 (1)	1.19 (2)	4.66 (5)	4.81 (5)

Matrix completion

Bob	Molly	Mary	Larry	
1	?	5	4	The Dark Knight
?	1	4	5	Spiderman 3
4	5	2	?	Love Actually
5	4	2	1	Bridget Jones's Diary
4	5	1	2	Pretty Woman
1	2	?	5	Superman 2

Low-rank matrix estimation

First idea:

$$\min_{\widetilde{X} \in \mathbb{R}^{m \times n}} \operatorname{rank}\left(\widetilde{X}\right) \quad \text{such that } \widetilde{X}_{\Omega} \approx y$$

Computationally intractable because of missing entries

Tractable alternative:

$$\min_{\widetilde{X} \in \mathbb{R}^{m \times n}} \left|\left|\widetilde{X}_{\Omega} - y\right|\right|_2^2 + \lambda \left|\left|\widetilde{X}\right|\right|_*$$

Matrix completion via nuclear-norm minimization

Bob	Molly	Mary	Larry	
1	2 (1)	5	4	The Dark Knight
2 (2)	1	4	5	Spiderman 3
4	5	2	2 (1)	Love Actually
5	4	2	1	Bridget Jones's Diary
4	5	1	2	Pretty Woman
1	2	5 (5)	5	Superman 2

Sparsity

- Denoising
- Regression
- Inverse problems

Low-rank models

- Matrix completion
- Low rank + sparse model
- Nonnegative matrix factorization

Background subtraction

Low rank + sparse model

$$\min_{\widetilde{L},\widetilde{S}\in\mathbb{R}^{m\times n}} \left|\left|\widetilde{L}\right|\right|_* + \lambda \left|\left|\widetilde{S}\right|\right|_1 \quad \text{such that } \widetilde{L} + \widetilde{S} = Y$$

Frame 17

Low-rank component

Sparse component

Frame 42

Low-rank component

Sparse component

Frame 75

Low-rank component

Sparse component

Sparsity

- Denoising
- Regression
- Inverse problems

Low-rank models

- Matrix completion
- Low rank + sparse model
- Nonnegative matrix factorization

Topic modeling

$A :=$

singer	GDP	senate	election	vote	stock	bass	market	band	Articles
6	1	1	0	0	1	9	0	8	a
1	0	9	5	8	1	0	1	0	b
8	1	0	1	0	0	9	1	7	c
0	7	1	0	0	9	1	7	0	d
0	5	6	7	5	6	0	7	2	e
1	0	8	5	9	2	0	0	1	f

SVD

$$A - \bar{A} = USV^T = U \begin{bmatrix} 19.32 & 0 & 0 & 0 & & \\ 0 & 14.46 & 0 & 0 & 0 & 0 \\ 0 & 0 & 4.99 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2.77 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1.67 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.93 \end{bmatrix} V^T$$

Left singular vectors

$$\begin{array}{lcllllll} & & \text{a} & \text{b} & \text{c} & \text{d} & \text{e} & \text{f} \\ U_1 & = & (-0.51 & -0.40 & -0.54 & -0.11 & -0.38 & -0.38) \\ U_2 & = & (\ 0.19 & -0.45 & -0.19 & -0.69 & -0.2 & -0.46) \\ U_3 & = & (\ 0.14 & -0.27 & -0.09 & -0.58 & -0.69 & -0.29) \end{array}$$

Right singular vectors

	singer	GDP	senate	election	vote	stock	bass	market	band
$V_1 =$	-0.38	0.05	0.40	0.27	0.40	0.17	-0.52	0.14	-0.38
$V_2 =$	0.16	-0.46	0.33	0.15	0.38	-0.49	0.10	-0.47	0.12
$V_3 =$	-0.18	-0.18	-0.04	-0.74	-0.05	0.11	-0.10	-0.43	-0.43

Nonnegative matrix factorization

$$
\begin{aligned}
M \approx WH, \quad W_{i,j} &\geq 0, \qquad 1 \leq i \leq m,\ 1 \leq j \leq r, \\
H_{i,j} &\geq 0, \qquad 1 \leq i \leq r,\ 1 \leq i \leq n,
\end{aligned}
$$

Right nonnegative factors

	singer	GDP	senate	election	vote	stock	bass	market	band
H_1 =	(0.34	0	3.73	2.54	3.67	0.52	0	0.35	0.35)
H_2 =	(0	2.21	0.21	0.45	0	2.64	0.21	2.43	0.22)
H_3 =	(3.22	0.37	0.19	0.2	0	0.12	4.13	0.13	3.43)

Left nonnegative factors

		a	b	c	d	e	f
W_1	=	(0.03	2.23	0	0	1.59	2.24)
W_2	=	(0.1	0	0.08	3.13	2.32	0)
W_3	=	(2.13	0	2.22	0	0	0.03)