Sparse Optimization

Wing-Kin (Ken) Ma

Department of Electronic Engineering,
The Chinese University Hong Kong, Hong Kong

ELEG5481, Lecture 13

Acknowledgment: Thank Jiaxian Pan and Xiao Fu for helping prepare the slides.

Introduction

- In signal processing, many problems involve finding a sparse solution.
  - compressive sensing
  - signal separation
  - recommendation system
  - direction of arrival estimation
  - robust face recognition
  - background extraction
  - text mining
  - hyperspectral imaging
  - MRI
  - ...

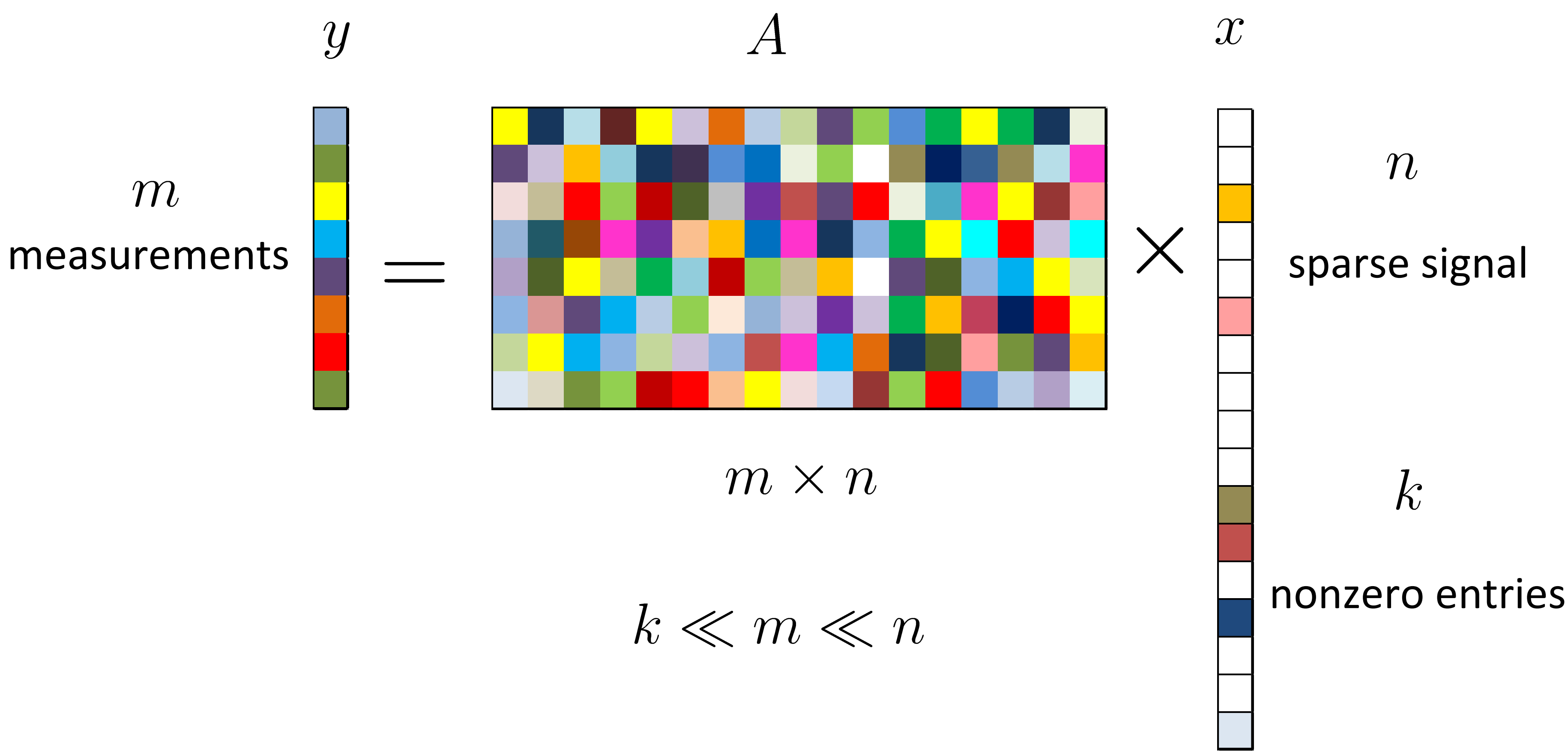
Single Measurement Vector Problem

- **Sparse single measurement vector (SMV) problem:** Given an observation vector $y \in \mathbf{R}^m$ and a matrix $A \in \mathbf{R}^{m \times n}$, find $x \in \mathbf{R}^n$ such that

$$y = Ax,$$

  and x is sparsest, i.e., x has the fewest number of nonzero entries.

- We assume that $m \ll n$, i.e., the number of observations is much smaller than the dimension of the source signal.


y
A
x
m
measurements
n
sparse signal
$m \times n$
k
nonzero entries
$k \ll m \ll n$

Sparse solution recovery

- We try to recover the sparsest solution by solving

$$\min_x \|x\|_0 \quad \text{s.t. } y = Ax$$

where $\|x\|_0$ is the number of nonzero entries of x.

- In the literature, $\|x\|_0$ is commonly called the "ℓ_0-norm", though it is not a norm.

Solving the SMV Problem

- The SMV problem is NP-hard in general.
- An exhaustive search method:
  - Fix the support of $x \in \mathbf{R}^n$, i.e., determine which entry of x is zero or non-zero.
  - Check if the corresponding x has a solution for $y = Ax$.
  - By solving all 2^n equations, an optimal solution can be found.
- A better way is to use the branch and bound method. But it is still very time-consuming.
- It is natural to seek approximate solutions.

Greedy Pursuit

- Greedy pursuit generates an approximate solution to SMV by recursively building an estimate $\hat{x}$.

- Greedy pursuit at each iterations follows two essential operations
  - Element selection: determine the support I of $\hat{x}$ (i.e. which elements are nonzero.)
  - Coefficient update: Update the coefficient $\hat{x}_i$ for $i \in I$.

Orthogonal Matching Pursuit (OMP)

- One of the oldest and simplest greedy pursuit algorithm is the orthogonal matching pursuit (OMP).
- First, initialize the support $I^{(0)} = \emptyset$ and estimate $\hat{x}^{(0)} = 0$.
- For $k = 1, 2, \ldots$ do
  - Element selection: determine an index $j^\star$ and add it to $I^{(k-1)}$.

$$r^{(k)} = y - A\hat{x}^{(k-1)} \qquad \text{(Compute residue } r^{(k)}\text{).}$$

$$j^\star = \arg \min_{\substack{j=1,\ldots,n \\ x}} \|r^{(k)} - a_j x\|_2 \qquad \text{(Find the column that reduces residue most)}$$

$$I^{(k)} = I^{(k-1)} \cup \{j^\star\} \qquad \text{(Add } j^\star \text{ to } I^{(k)}\text{)}$$

  - Coefficient update: with support $I^{(k)}$, minimize the estimation residue,

$$\hat{x}^{(k)} = \arg \max_{x : x_i = 0,\ i \notin I^{(k)}} \|y - Ax\|_2.$$

ℓ_1-norm heuristics

- Another method is to approximate the nonconvex $\|x\|_0$ by a convex function.

$$\min_x \ \|x\|_1 \quad \text{s.t. } y = Ax.$$

- The above problem is also known as basis pursuit in the literature.

- This problem is convex (an LP actually).

Interpretation as convex relaxation

- Let us start with the original formulation (with a bound on x)

$$\min_x \|x\|_0 \quad \text{s.t. } y = Ax, \quad \|x\|_\infty \leq R.$$

- The above problem can be rewritten as a mixed Boolean convex problem

$$\begin{aligned} \min_{x,z} \ & \mathbf{1}^T z \\ \text{s.t. } & y = Ax, \\ & |x_i| \leq R z_i, \quad i = 1, \dots, n \\ & z_i \in \{0, 1\}, \quad i = 1, \dots, n. \end{aligned}$$

- Relax $z_i \in \{0,1\}$ to $z_i \in [0,1]$ to obtain

$$\begin{aligned} \min_{x,z} \ & \mathbf{1}^T z \\ \text{s.t. } & y = Ax, \\ & |x_i| \leq R z_i, \quad i = 1, \ldots, n \\ & 0 \leq z_i \leq 1, \quad i = 1, \ldots, n. \end{aligned}$$

- Observing that $z_i = |x_i|/R$ at optimum , the problem above is equivalent to

$$\begin{aligned} \min_{x,z} \ & \|x\|_1 / R \\ \text{s.t. } & y = Ax, \end{aligned}$$

which is the ℓ_1-norm heuristic.

- The optimal value of the above problem is a lower bound on that of the original problem.

Interpretation via convex envelope

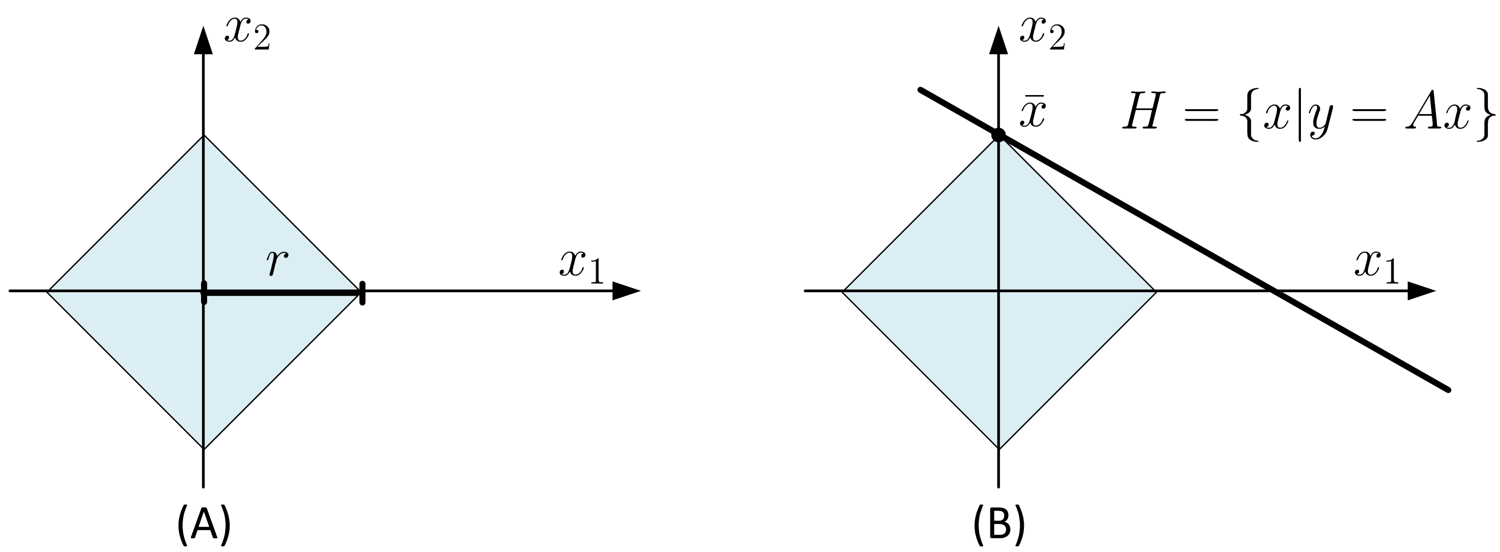
- Given a function f with domain $\mathcal{C}$, the convex envelope f^{env} is the largest possible convex underestimation of f over $\mathcal{C}$, i.e.,

$$f^{\text{env}}(x) = \sup\{g(x) \mid g(x') \leq f(x'), \forall\, x' \in \mathcal{C},\ g(x) \text{ convex}\}.$$

- When x is a scalar, $|x|$ is the convex envelope of $\|x\|_0$ on $[-1, 1]$.

- When x is a vector, $\|x\|_1/R$ is convex envelope of $\|x\|_0$ on $\mathcal{C} = \{x \mid \|x\|_\infty \leq R\}$.

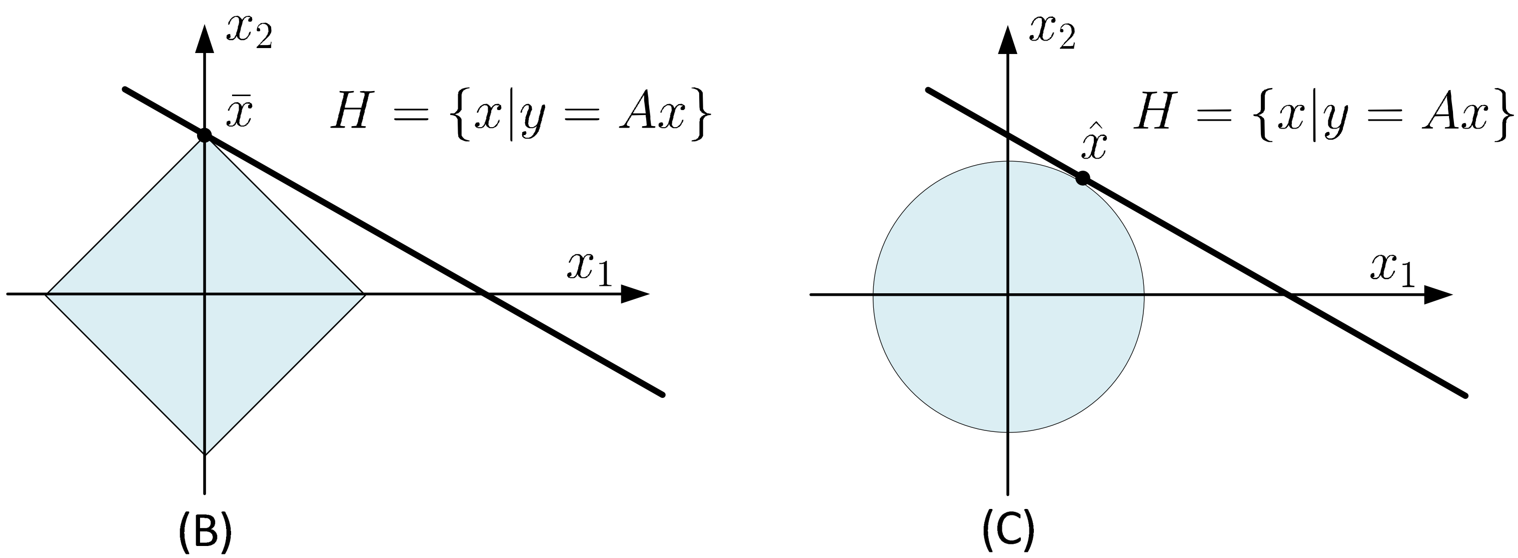
ℓ_1-norm geometry



- Fig. A shows the ℓ_1 ball of some radius r in $\mathbf{R}^2$. Note that the ℓ_1 ball is "pointy" along the axes.

- Fig. B shows the ℓ_1 recovery problem. The point $\bar{x}$ is a "sparse" vector; the line H is the set of x that shares the same measurement y.

ℓ_1-norm geometry



- The ℓ_1 recovery problem is to pick out a point in H that has the minimum ℓ_1 norm. We can see that $\bar{x}$ is such a point.

- Fig. C shows the geometry when ℓ_2 norm is used instead of ℓ_1 norm. We can see that the solution $\hat{x}$ may not be sparse.

Recovery guarantee of ℓ_1-norm minimization

- When ℓ_1-norm minimization is equivalent to ℓ_0-norm minimization?
- Sufficient conditions are provided by characterizing the structure of A and the sparsity of the desirable x.
  - Example: Let $\mu(A) = \max_{i \neq j} \frac{|a_i^T a_j|}{\|a_i\|_2 \|a_j\|_2}$ which is called the mutual coherence. If there exists an x such that $y = Ax$ and

$$\mu(A) \leq \frac{1}{2\|x\|_0 - 1},$$

    then x is the unique solution of ℓ_1-norm minimization. It is also the solution of the corresponding ℓ_0-norm minimization.
  - Such mutual coherence condition means that sparser x and "more orthonormal" A provide better chance of perfect recovery by ℓ_1-norm minimization.
- Other conditions: restricted isometry property (R.I.P.) condition, null space property, ...

Recovery guarantee of ℓ_1-norm minimization

There are several other variations.

- Basis pursuit denoising

$$\begin{aligned} \min\ & \|x\|_1 \\ \text{s.t. } & \|y - Ax\|_2 \le \epsilon. \end{aligned}$$

- Penalized least squares

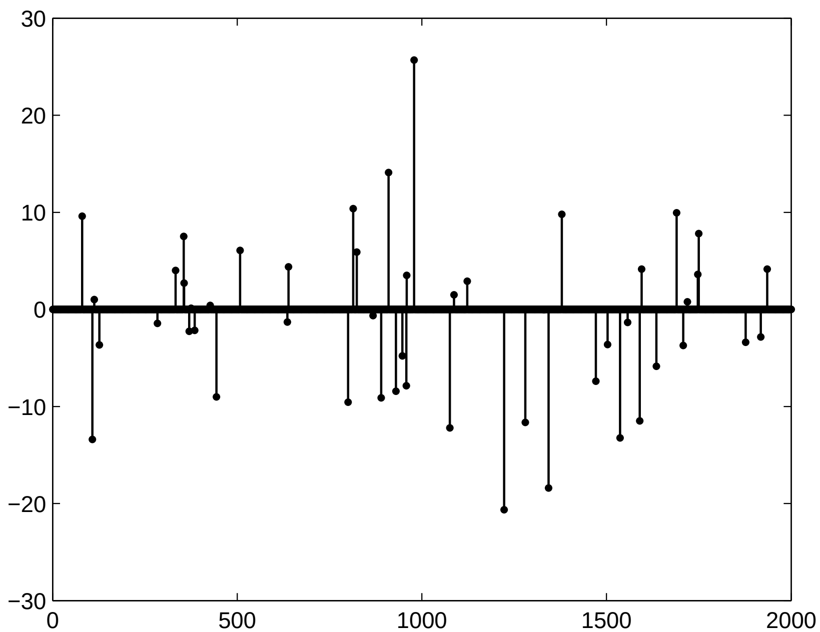
$$\min\ \|Ax - b\|_2^2 + \lambda\|x\|_1.$$

- Lasso Problem

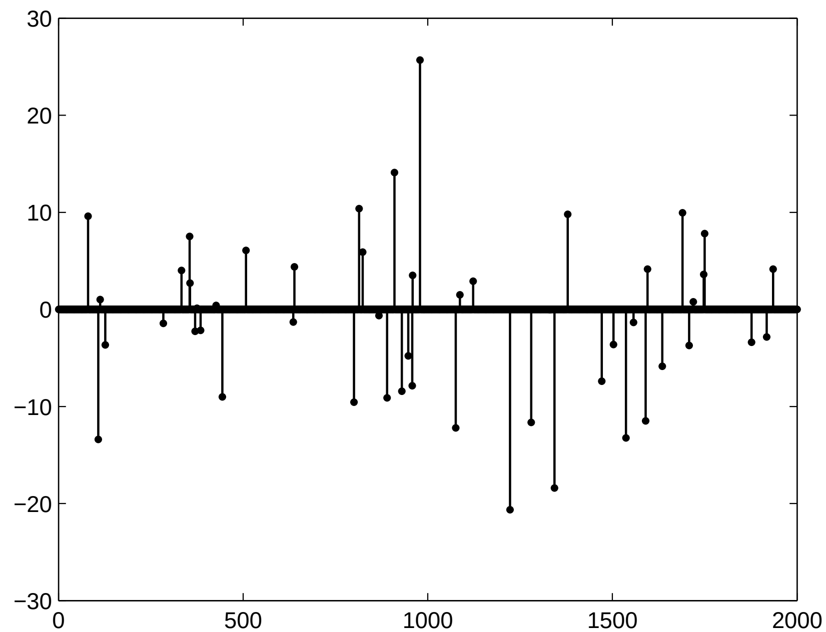
$$\begin{aligned} \min\ & \|Ax - b\|_2^2 \\ \text{s.t. } & \|x\|_1 \le \tau. \end{aligned}$$

Application: Sparse signal reconstruction

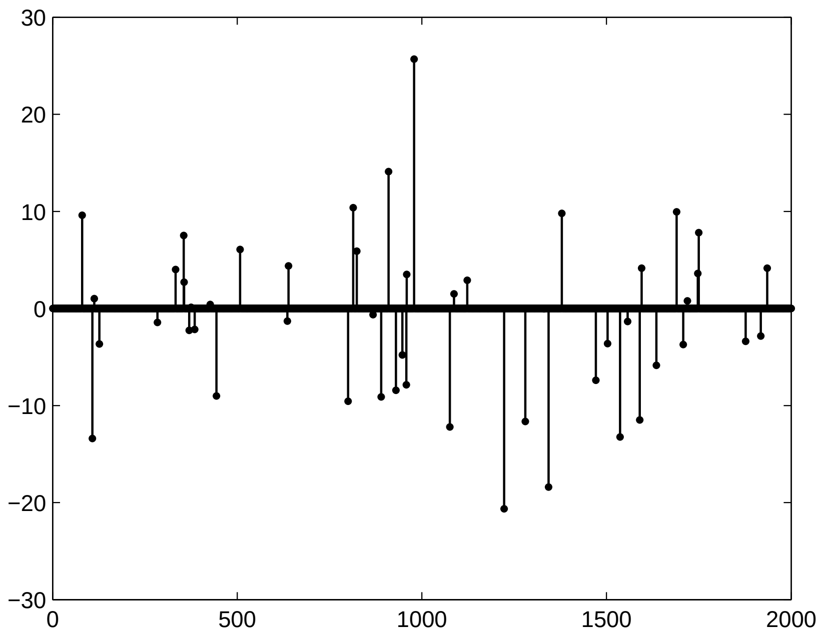
- Sparse signal $x \in \mathbf{R}^n$ with $n = 2000$ and $\|x\|_0 = 50$.
- $m = 400$ noise-free observations of $y = Ax$, where $A_{ij} \sim \mathcal{N}(0, 1)$.



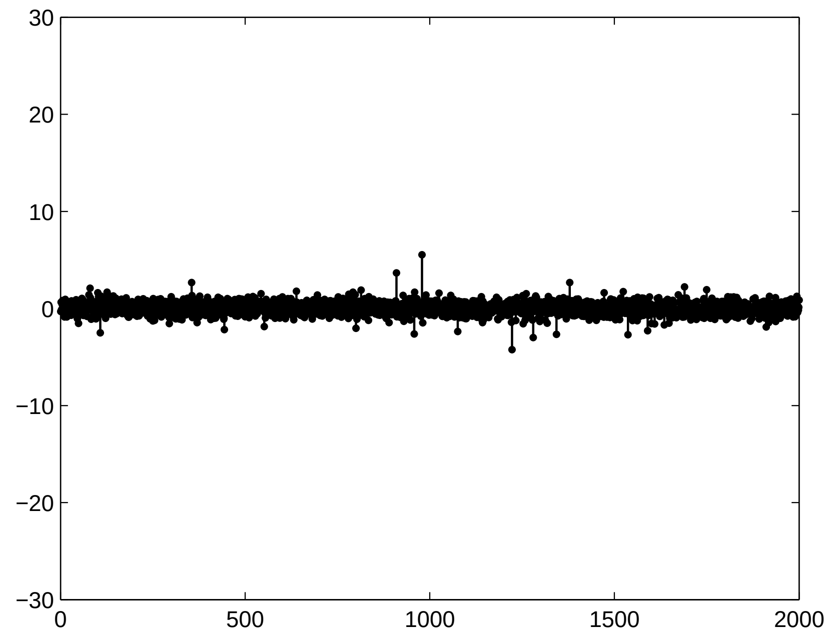
Sparse source signal



Perfect recovery by ℓ_1-norm minimization



Sparse source signal



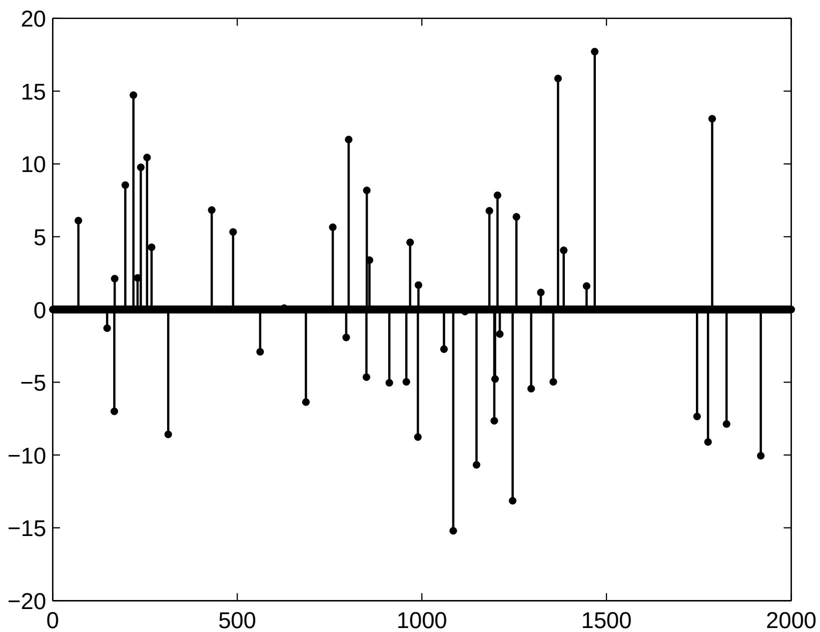
Estimated by ℓ_2-norm minimization

- Sparse signal $x \in \mathbf{R}^n$ with $n = 2000$ and $\|x\|_0 = 50$.

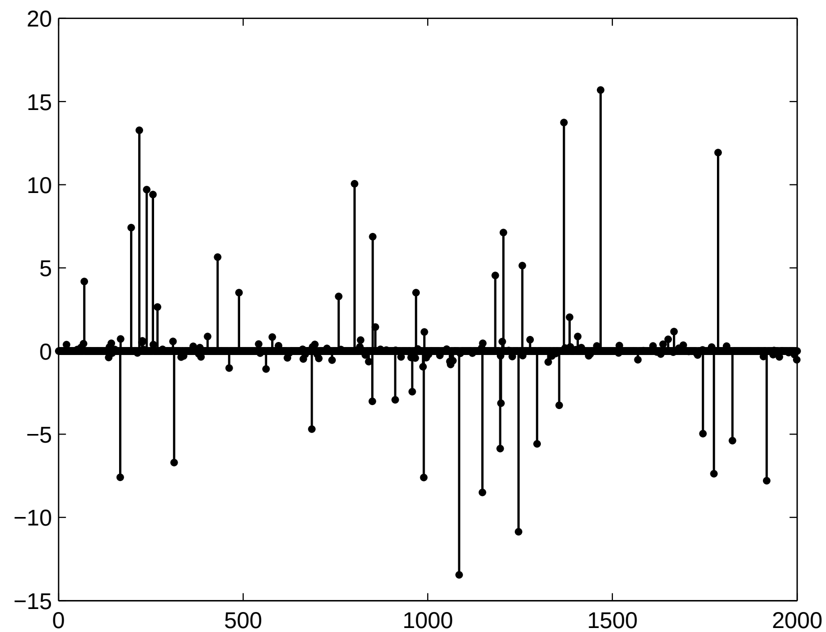
- $m = 400$ noisy observations of $y = Ax + \nu$, where $A_{ij} \sim \mathcal{N}(0,1)$ and $\nu_i \sim \mathcal{N}(0, \delta^2)$.

- Basis pursuit denoising is used.

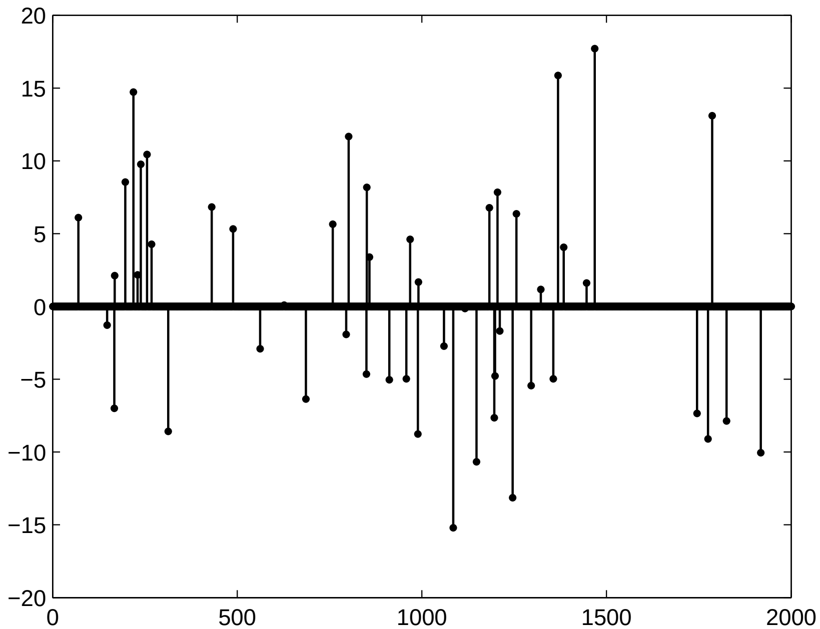
- $\delta^2 = 100$ and $\epsilon = \sqrt{m\delta^2}$.



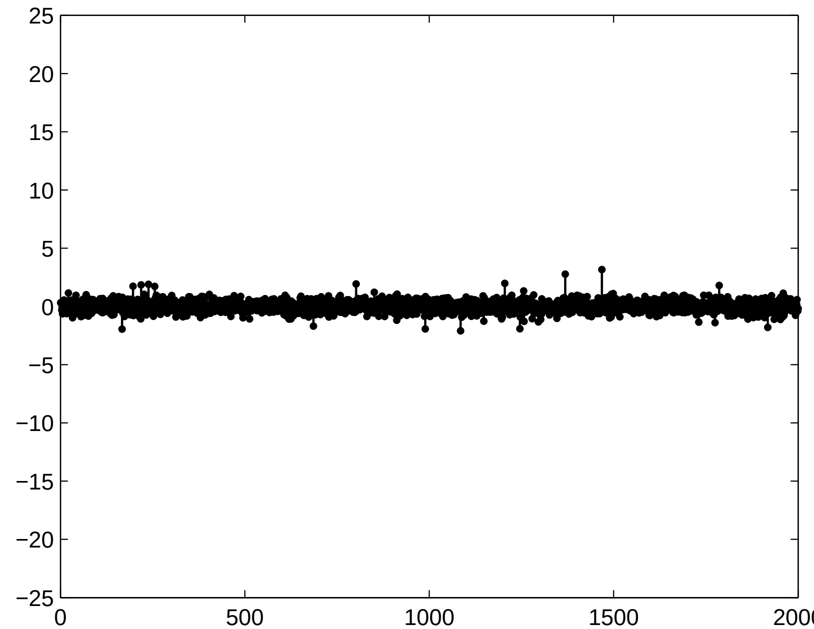
Sparse source signal



Estimated by ℓ_1-norm minimization



Sparse source signal



Estimated by ℓ_2-norm minimization

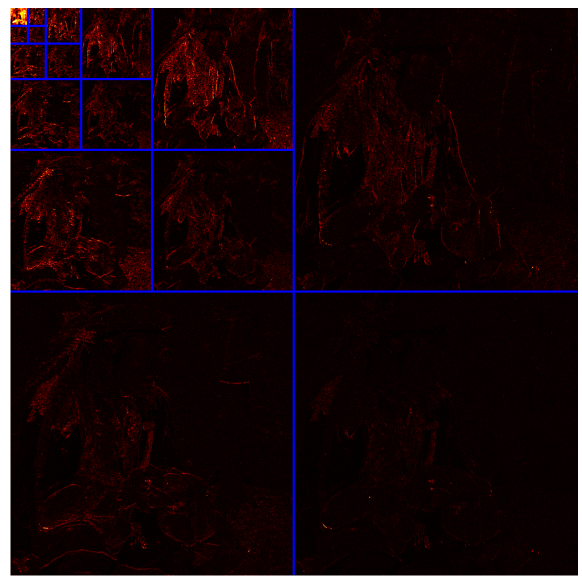
Application: Compressive sensing (CS)

- Consider a signal $\tilde{x} \in \mathbf{R}^n$ that has a sparse representation $x \in \mathbf{R}^n$ in the domain of $\Psi \in \mathbf{R}^{n \times n}$ (e.g. FFT and wavelet), i.e.,

$$\tilde{x} = \Psi x.$$

where x is sparse.

The pirate image $\tilde{x}$



The wavelet transform x

- To acquire information of the signal x, we use a sensing matrix $\Phi \in \mathbf{R}^{m\times n}$ to observe x

$$y = \Phi\tilde{x} = \Phi\Psi x.$$

  Here, we have $m \ll n$, i.e., we only obtain very few observations compared to the dimension of x.

- Such a y will be good for compression, transmission and storage.

- $\tilde{x}$ is recovered by recovering x:

$$\begin{aligned} &\min \ \|x\|_0 \\ &\text{s.t. } y = Ax, \end{aligned}$$

  where $A = \Phi\Psi$.

Application: Total Variation-based Denoising

- Scenario:
  - We want to estimate $x \in \mathbf{R}^n$ from a noisy measurement $x_{\text{cor}} = x + n$.
  - x is known to be piecewise linear, i.e., for most i we have

$$x_i - x_{i-1} = x_{i+1} - x_i \iff -x_{i+1} + 2x_i - x_{i+1} = 0.$$

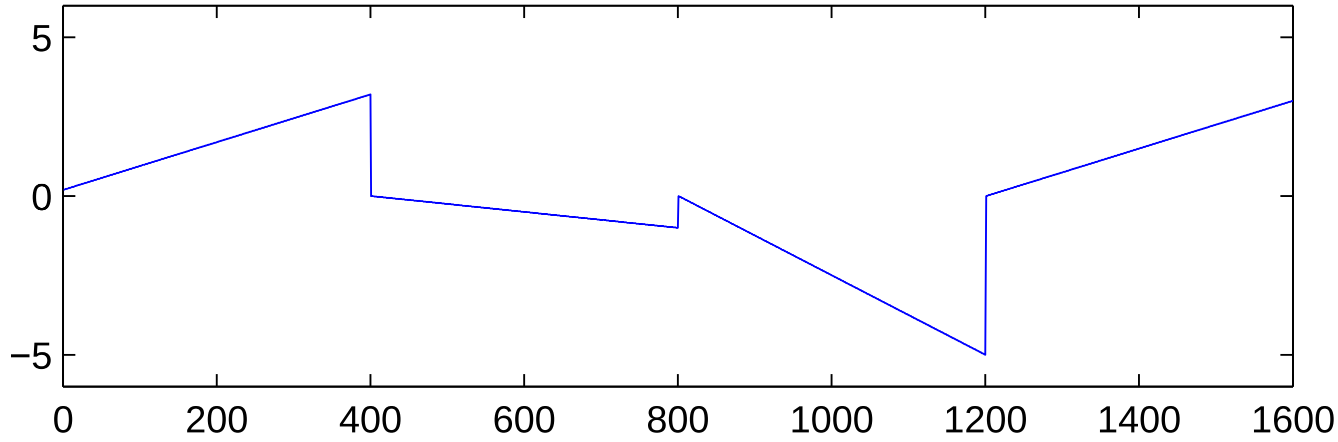
  - Equivalently, Dx is sparse, where

$$D = \begin{bmatrix} -1 & 2 & 1 & 0 & \dots \\ 0 & -1 & 2 & 1 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \dots & \dots & -1 & 2 & 1 \end{bmatrix}.$$

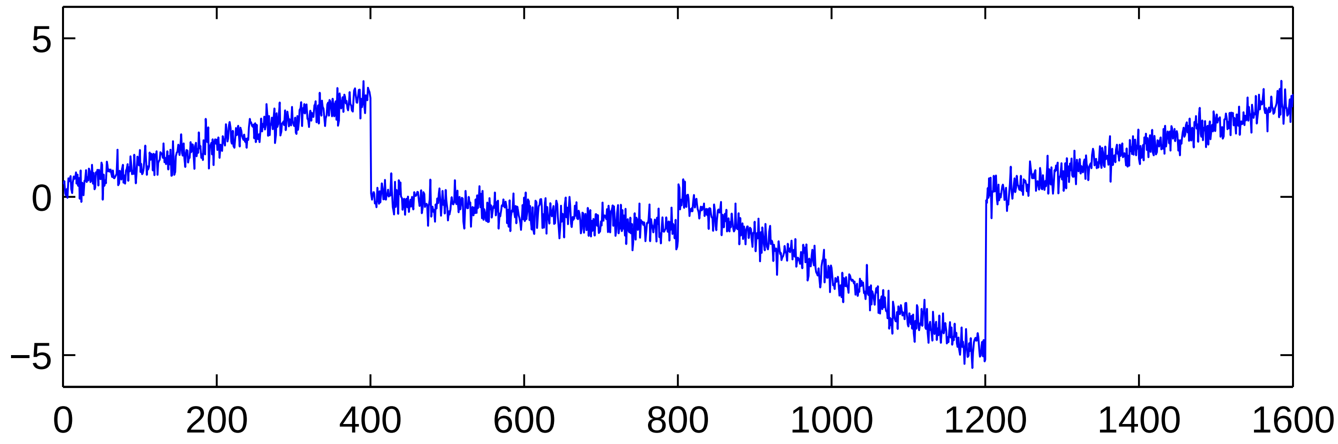
- Problem formulation: $\hat{x} = \arg\min_x \; \|x_{\text{cor}} - x\|_2 + \lambda \|Dx\|_0$.

- Heuristic: change $\|Dx\|_0$ to $\|Dx\|_1$.

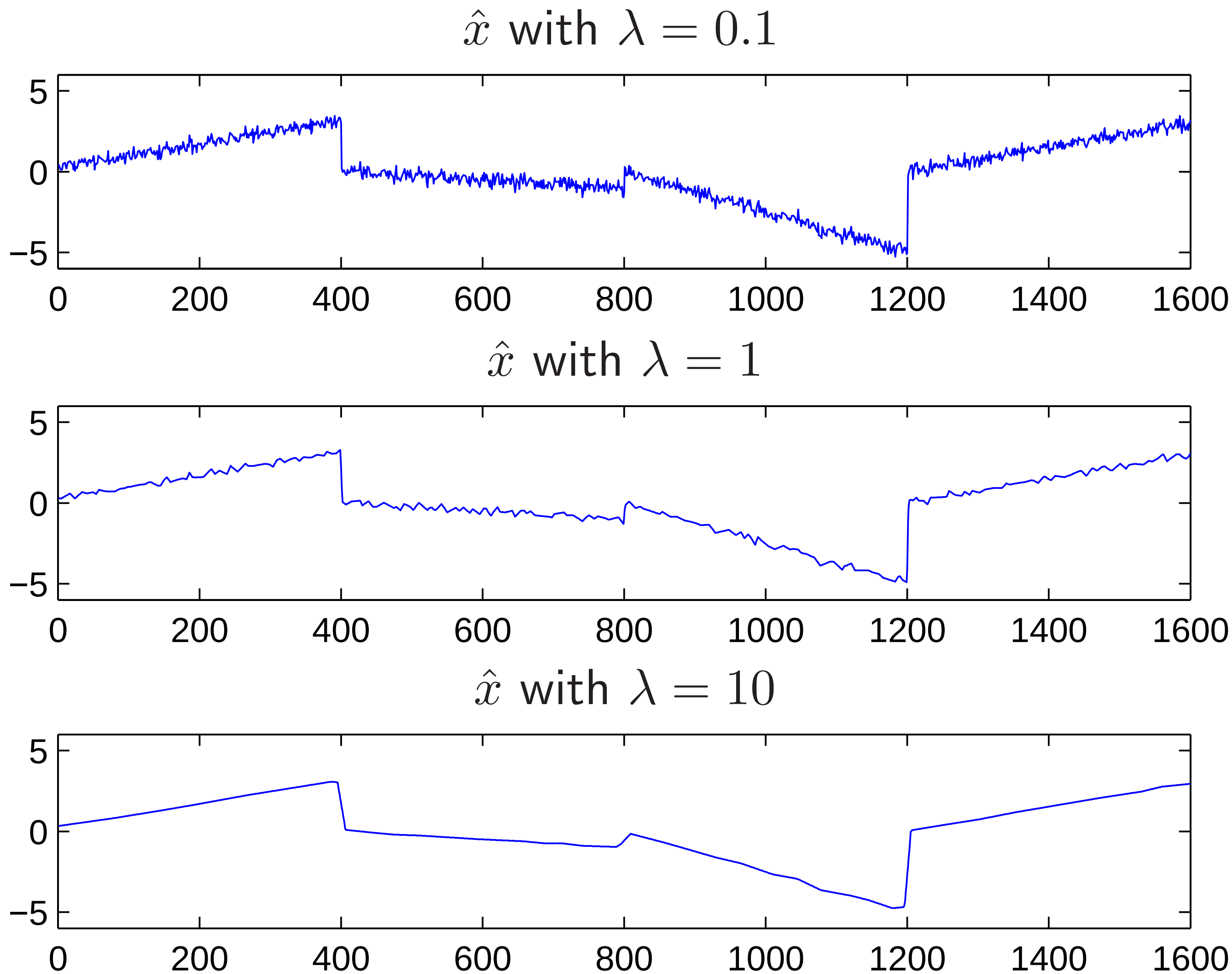
Source



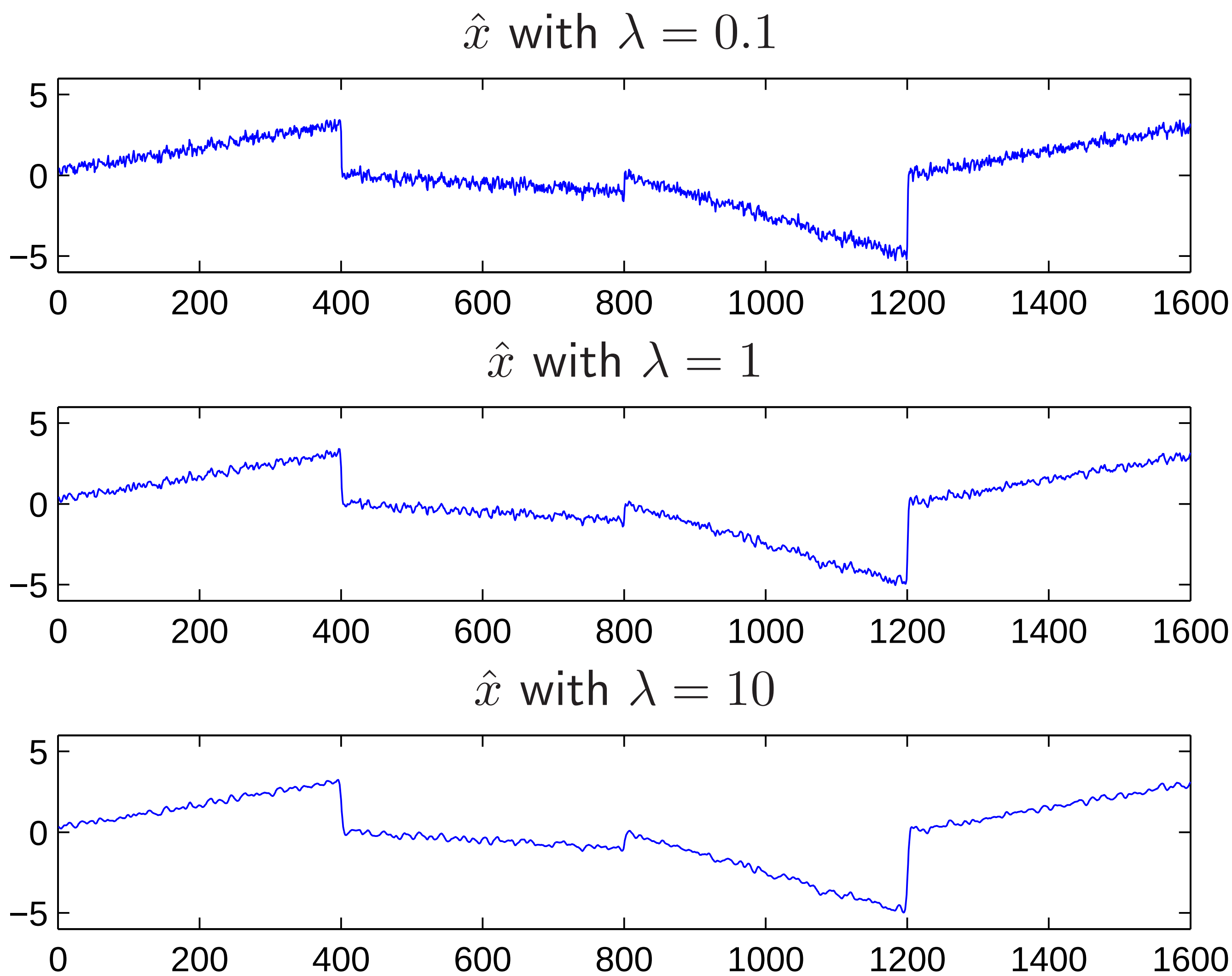
Corrupted by noise



Original x and corrupted x_{cor}



Denoised signals with different λ's and by $\hat{x} = \arg\min_x \; \|x_{\mathrm{cor}} - x\|_2 + \lambda \|Dx\|_1$.



Denoised signals with different λ's and by $\hat{x} = \arg\min_x \ \|x_{\rm cor} - x\|_2 + \lambda \|Dx\|_2$.

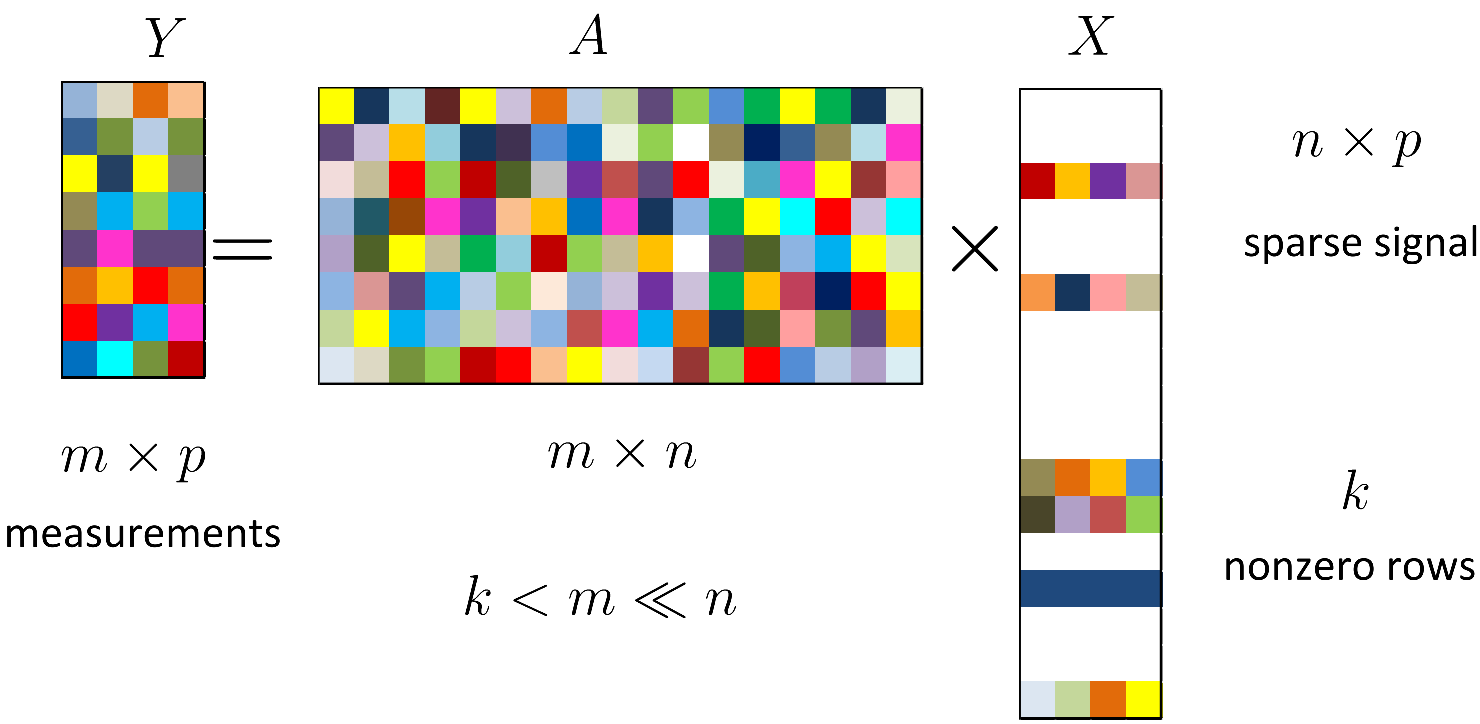
Matrix Sparsity

The notion of sparsity for a matrix X may refer to several different meanings.

- Element-wise sparsity: $\|\mathrm{vec}(X)\|_0$ is small.
- Row sparsity: X only has a few nonzero rows.
- Rank sparsity: $\mathrm{rank}(X)$ is small.

Row sparsity

- Let $X = [x_1, \ldots, x_p]$. Row sparsity means that each x_i shares the same support.



Row sparsity

- Multiple measurement vector (MMV) problem

$$\min_X \|X\|_{\text{row-0}}$$
$$\text{s.t. } Y = AX,$$

where $\|X\|_{\text{row-0}}$ denote the number of nonzero rows.

- Empirically, MMV works (much) better than SMV in many applications.

- Mixed-norm relaxation approach:

$$\begin{aligned} \min_{X} \; & \|X\|_{q,p}^{p} \\ \text{s.t. } & Y = AX, \end{aligned}$$

where $\|X\|_{q,p} = (\sum_{i=1}^{m} \|x^i\|_q^p)^{(1/p)}$ and x^i denotes the ith row in X.

- For $q \in [1, \infty]$ and $p = 1$, this is a convex problem.
- For $(p, q) = (1, 2)$, this problem can be formulated as an SOCP

$$\begin{aligned} \min_{t,X} \; & \sum_{i=1}^{m} t_i \\ \text{s.t. } & Y = AX \\ & \|x^i\|_2 \leq t_i, \; i = 1, \ldots, m. \end{aligned}$$

- Some variations:

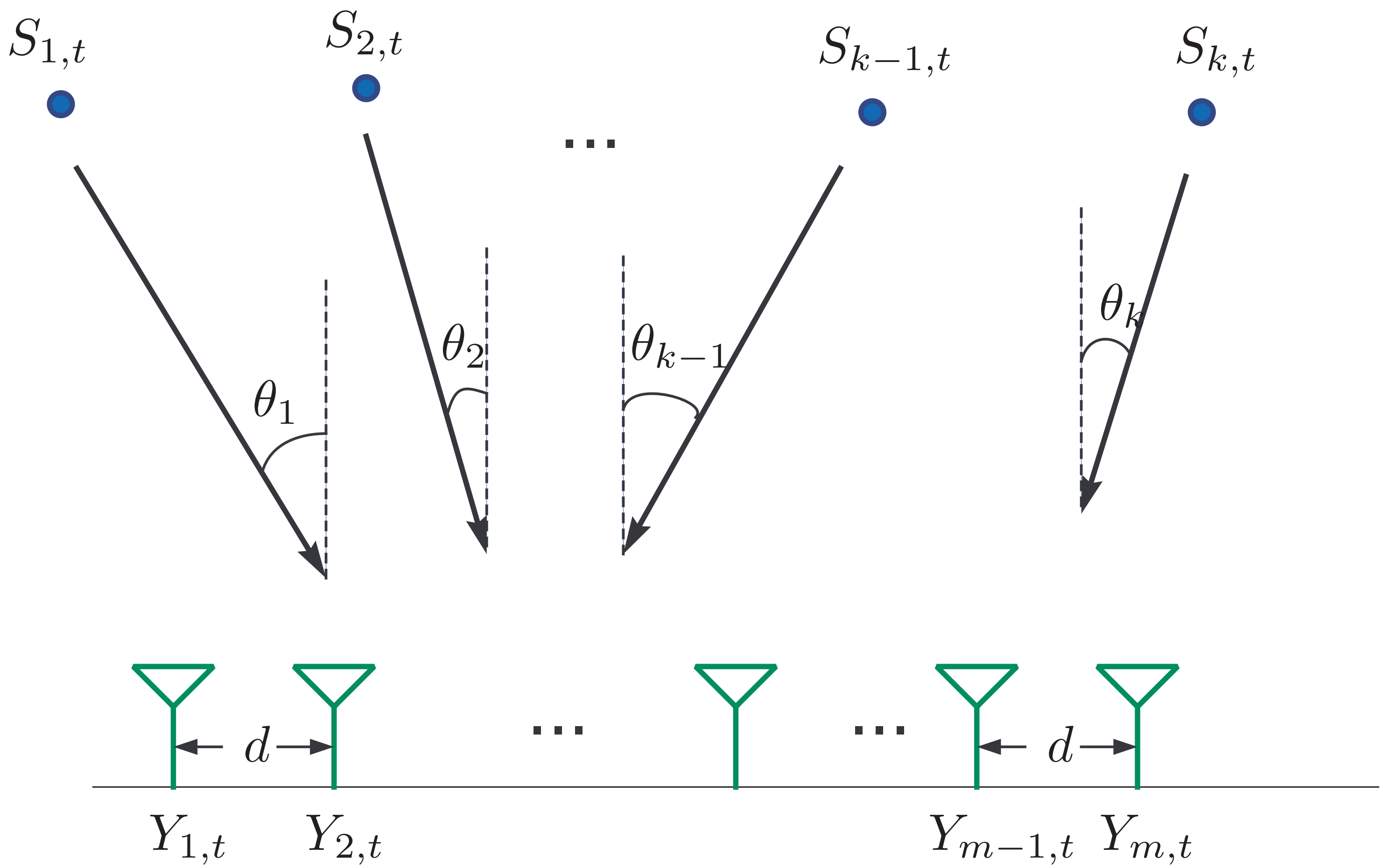
$$\begin{aligned} &\min \|X\|_{2,1} \\ &\text{s.t. } \|Y - AX\|_F \le \epsilon. \end{aligned}$$

$$\min \|AX - Y\|_F^2 + \lambda \|X\|_{2,1}$$

$$\begin{aligned} &\min \|AX - b\|_F^2 \\ &\text{s.t. } \|X\|_{2,1} \le \tau. \end{aligned}$$

- Other algorithms: Simultaneously Orthogonal Matching Pursuit (SOMP), Compressive Multiple Signal Classification (Compressive MUSIC), Nonconvex mixed-norm approach (by choosing $0 < p < 1$), ...

Application: Direction-of-Arrival (DOA) estimation



Application: Direction-of-Arrival (DOA) estimation

- Considering $t = 1, \ldots, p$, the signal model is

$$Y = A(\theta)S + N,$$

where

$$A(\theta) = \begin{bmatrix} 1 & \ldots & 1 \\ e^{-\frac{j2\pi d}{\gamma}\sin(\theta_1)} & \ldots & e^{-\frac{j2\pi d}{\gamma}\sin(\theta_m)} \\ \vdots & \vdots & \vdots \\ e^{-\frac{j2\pi d}{\gamma}(n-1)\sin(\theta_1)} & \ldots & e^{-\frac{j2\pi d}{\gamma}(n-1)\sin(\theta_m)} \end{bmatrix}$$

$Y \in \mathbf{R}^{m \times p}$ are received signals, $S \in \mathbf{R}^{k \times p}$ sources, $N \in \mathbf{R}^{m \times p}$ noise, m and k number of receivers and sources, and γ is the wavelength.

- Objective: estimate $\theta = [\theta_1, \ldots, \theta_k]^T$, where $\theta_i \in [-90^\circ, 90^\circ]$ for $i = 1, \ldots, k$.

- Construct

$$A = [\ a(-90^\circ), a(-89^\circ), a(-88^\circ), \ldots, a(88^\circ), a(89^\circ), a(90^\circ)\ ],$$

where $a(\theta) = [\ 1, e^{-\frac{j2\pi d}{\gamma}\sin(\theta)}, \ldots, e^{-\frac{j2\pi d}{\gamma}\sin(\theta)}\ ]^T$.

- By such construction, we have

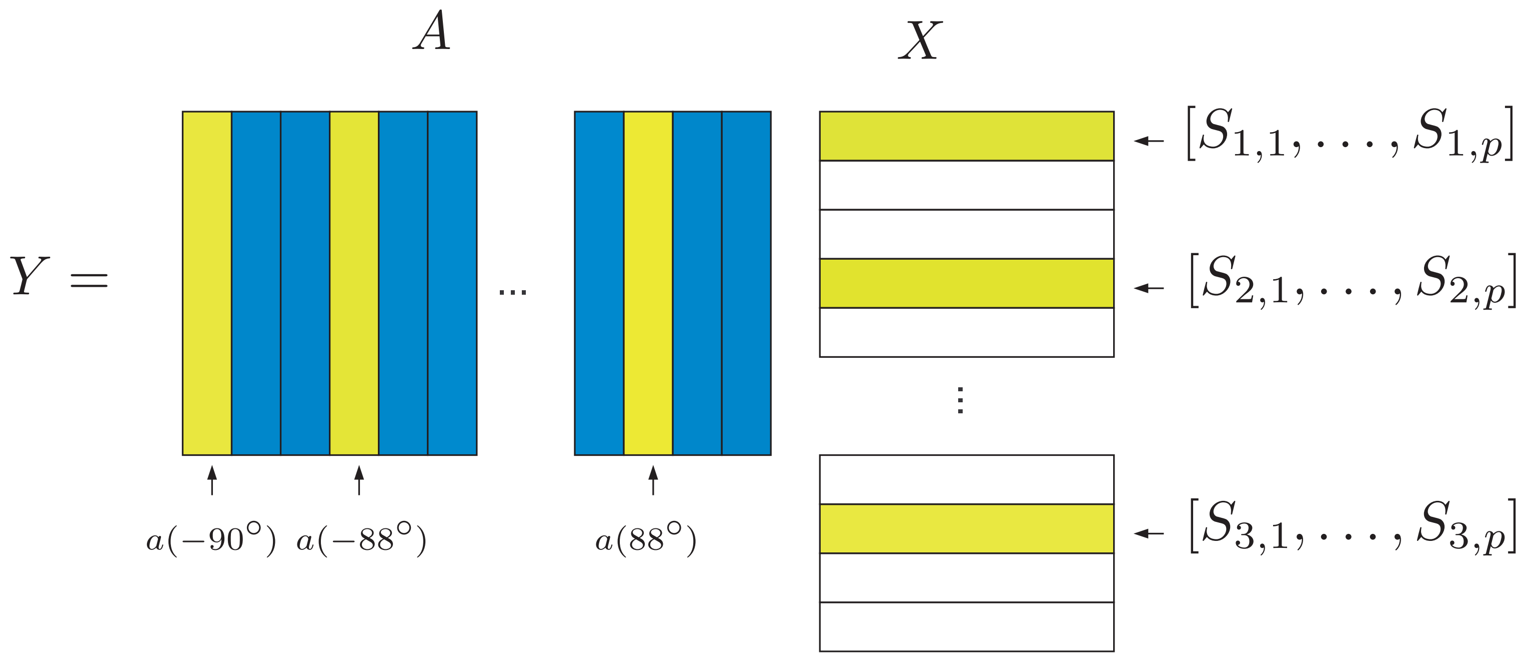
$$A(\theta) = [\ a(\theta_1), \ldots, a(\theta_k)\ ],$$

is approximately a submatrix of A.

- DOA estimation is approximately equivalent to finding the columns of $A(\theta)$ in A.

- Discretizing $[-90^\circ, 90^\circ]$ to more dense grids may increase the estimation accuracy while require more computation resources.

- Example: $k = 3$, $\theta = [-90°, -88°, 88°]$.

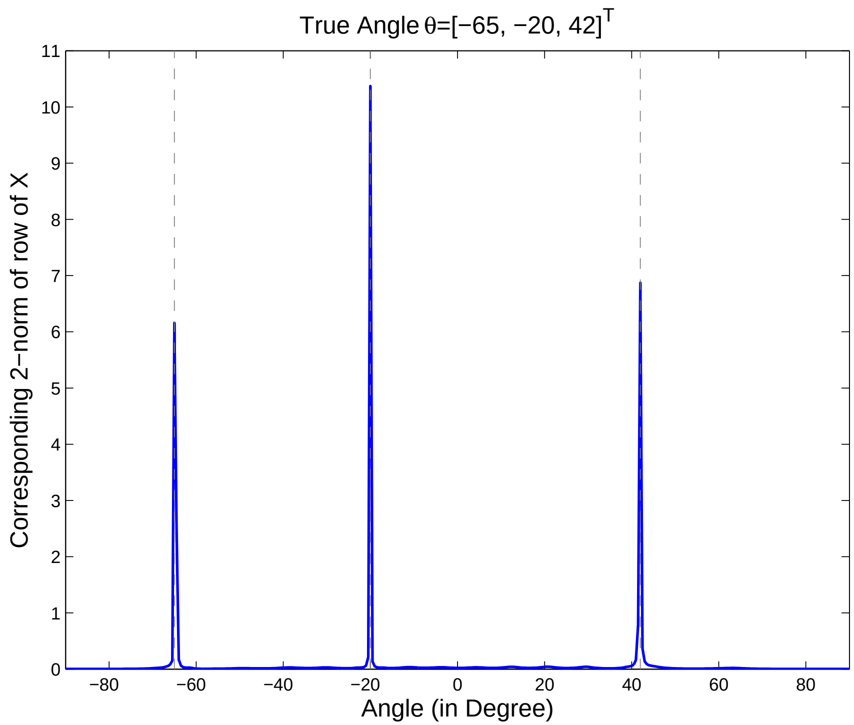


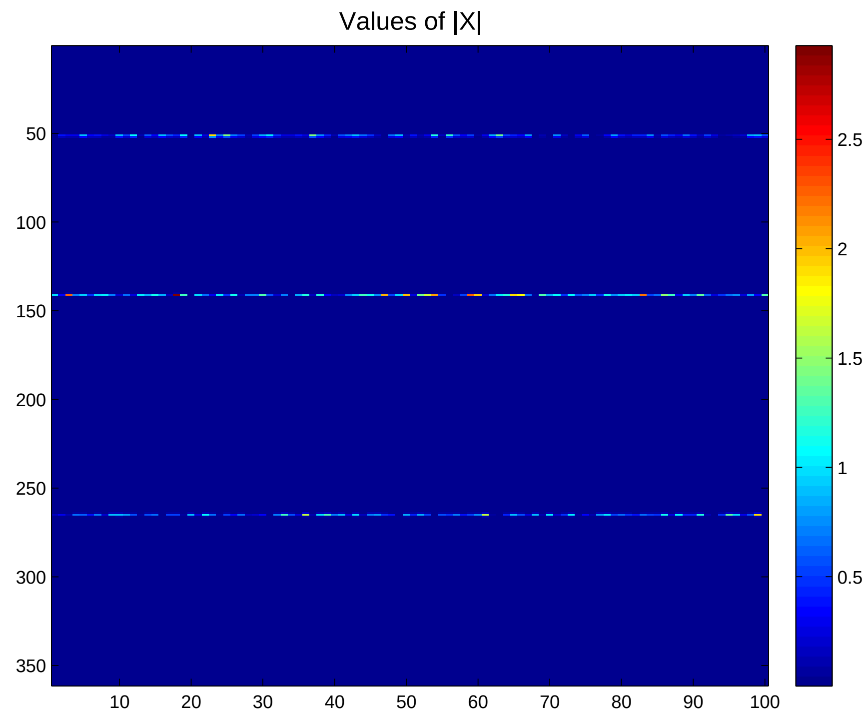
- To locate the "active columns" in A is equivalent to find a row-sparse X.

- Problem formulation:

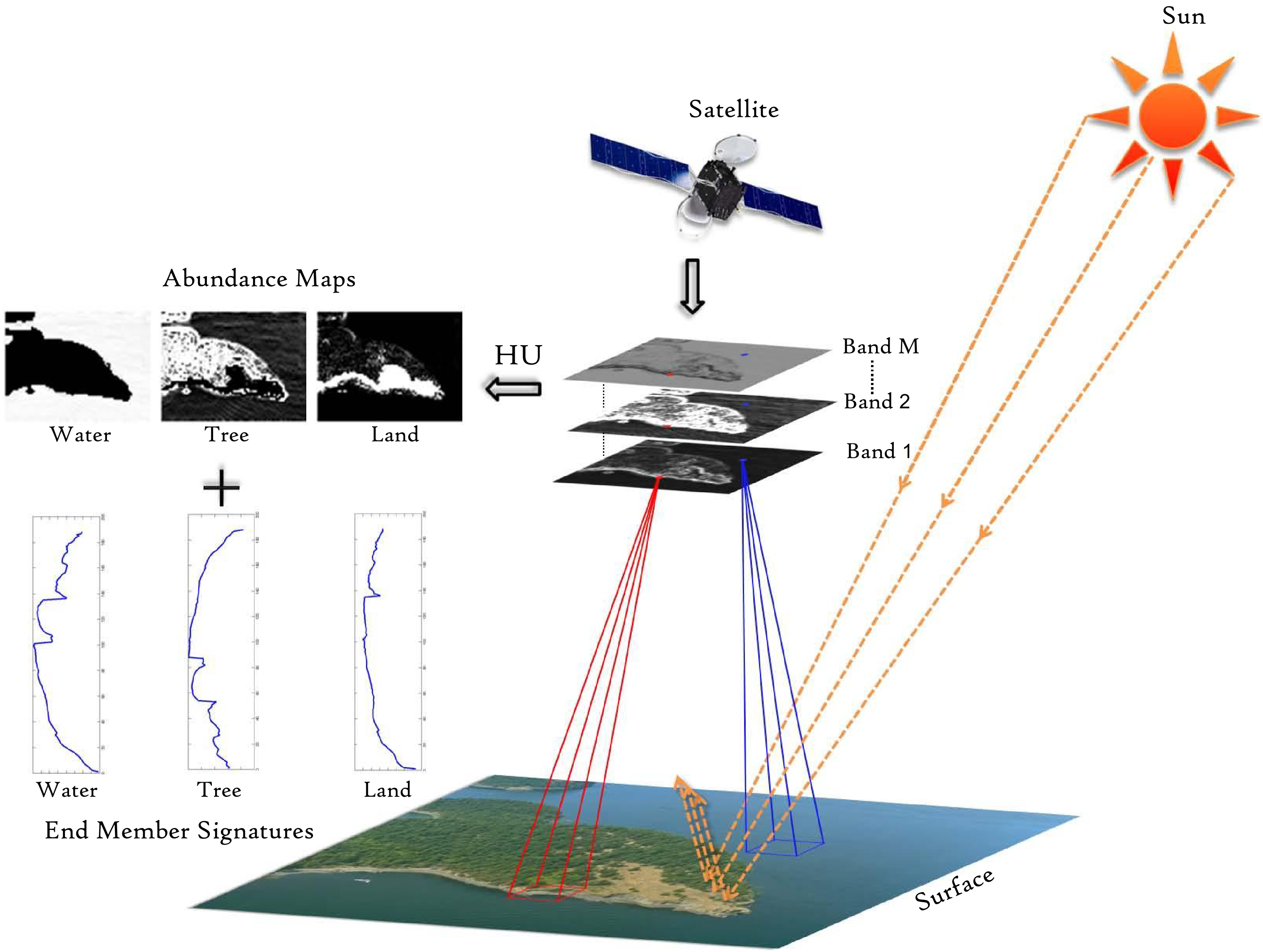
$$\min_X \; \|Y - AX\|_F^2 + \lambda \|X\|_{2,1}.$$

Simulation: $k = 3$, $p = 100$, $n = 8$ and SNR$= 30$dB; three sources come from $-65°$, $-20°$ and $42°$, respectively. $A = [\ a(-90°), a(-89.5°), \ldots, a(90°)\ ] \in \mathbf{R}^{m \times 381}$.

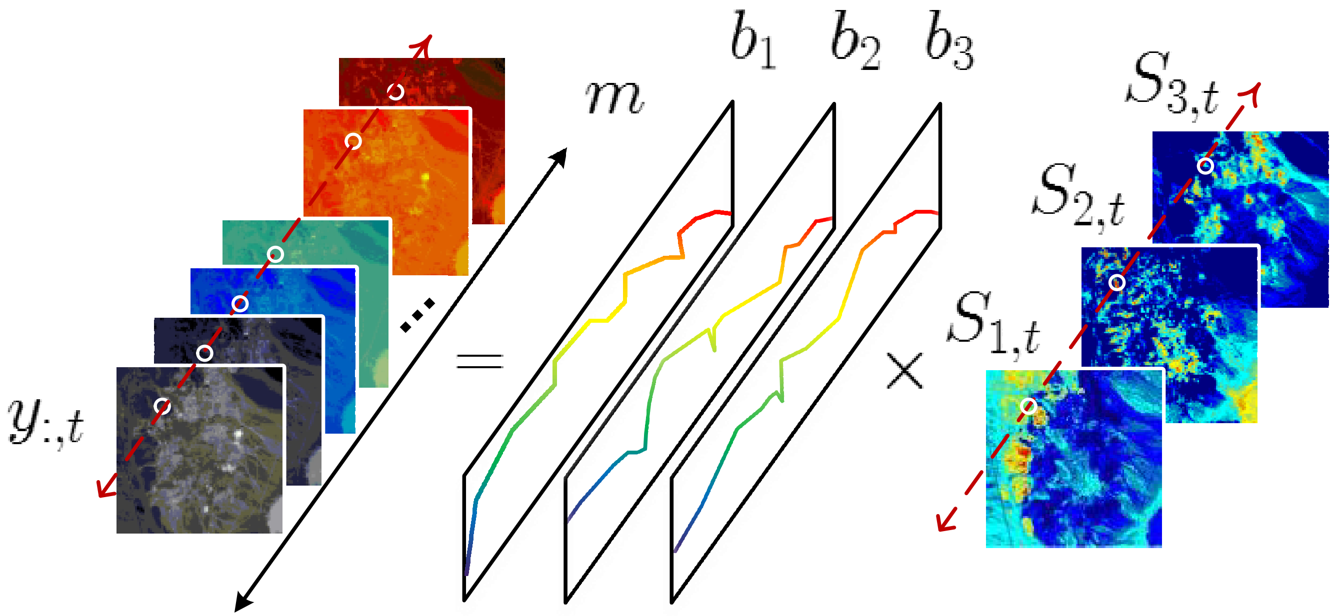




Application: Library-based Hyperspectral Image Separation



- Consider a hyperspectral image (HSI) captured by a remote sensor (satellite, aircraft, etc.).
- Each pixel of HSI is an m-dimensional vector, corresponding to spectral info. of m bands.
- The spectral shape can be used for classifying materials on the ground.
- During the process of image capture, the spectra of different materials might be mixed in pixels.



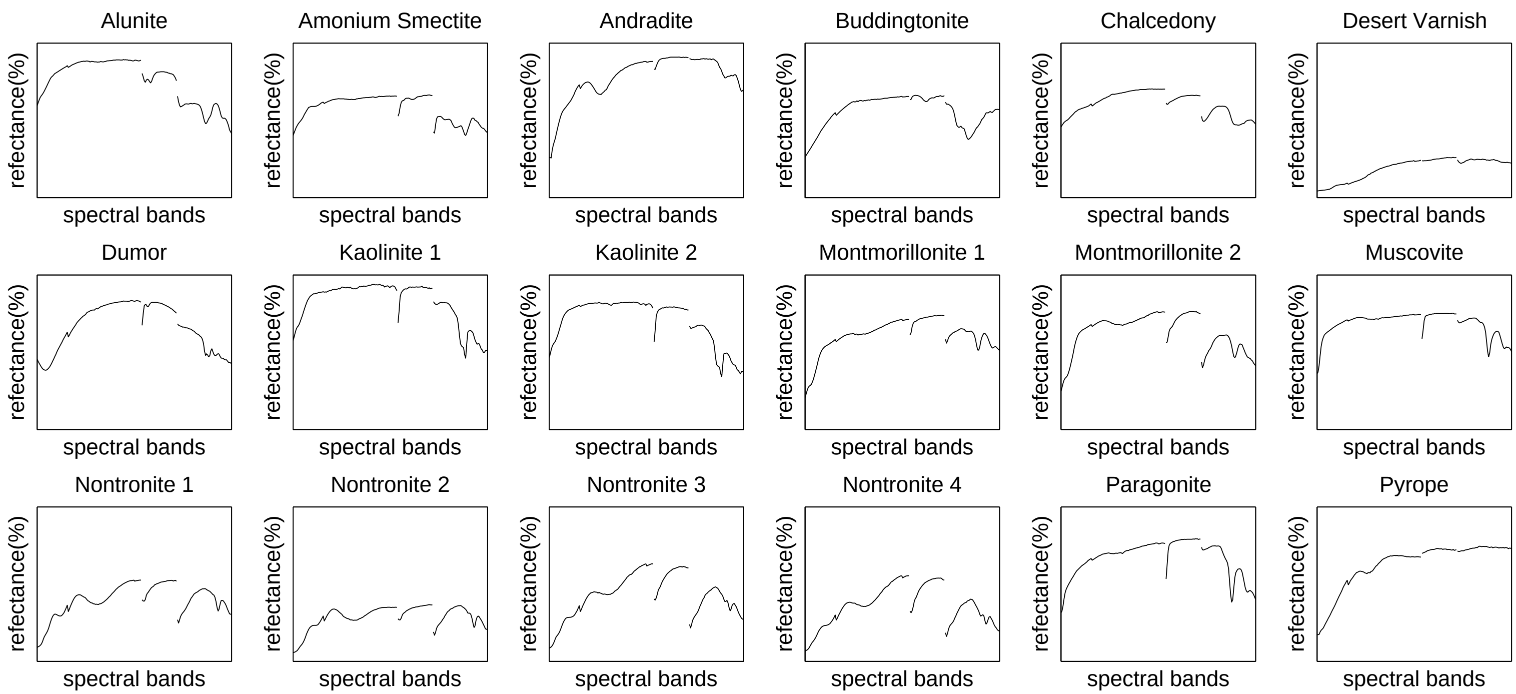
- Signal Model:

$$Y = BS + N,$$

where $Y \in \mathbf{R}^{m \times p}$ is HSI with p pixels, $B = [b_1, \ldots, b_k] \in \mathbf{R}^{m \times k}$ are spectra of materials, $S \in \mathbf{R}_+^{k \times p}$, $S_{i,j}$ represents the amount of material i in pixel j, and N is the noise.

- To know what materials are in pixels, we need to estimate B and S.

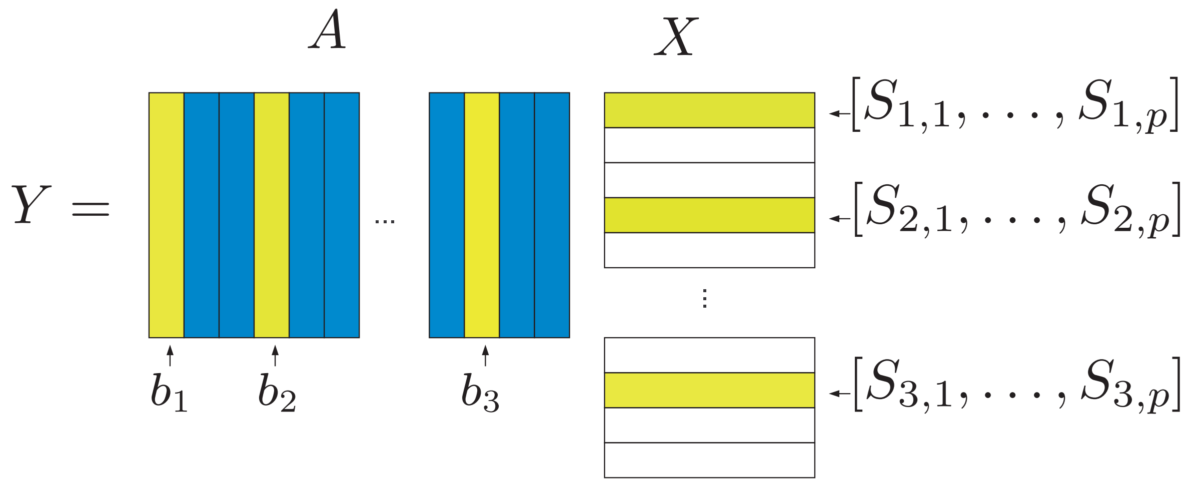
- There are spectral libraries providing spectra of more than a thousand materials.



Some recorded spectra of minerals in U.S.G.S library.

- In many cases, an HSI pixel can be considered as a mixture of 3 to 5 spectra in a known library, which records hundreds of spectra.

- Example: Suppose that $B = [b_1, b_2, b_3]$ is a submatrix of a known library A. Again, we have



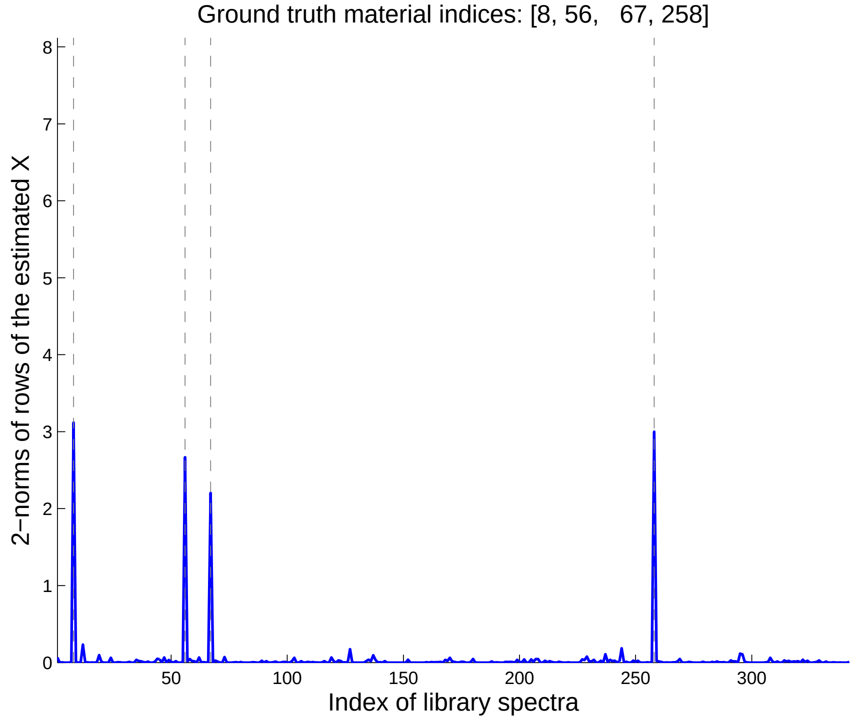
- Estimation of B and S can be done via finding the row-sparse X.

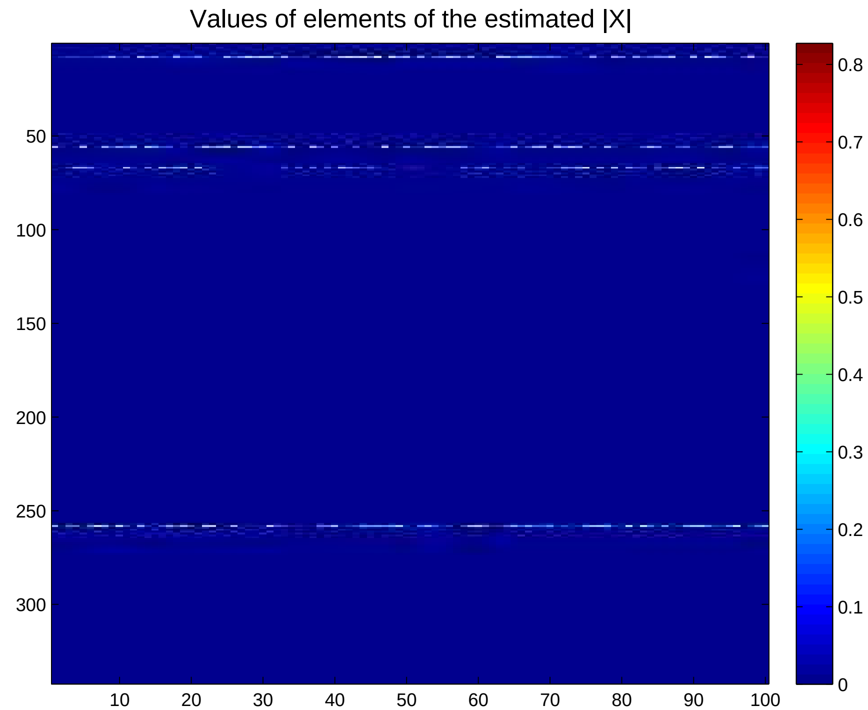
- Problem formulation:

$$\min_{X \geq 0} \|Y - AX\|_F^2 + \lambda \|X\|_{2,1},$$

where the non-negativity of X is added for physical consistency (since elements of S represent amounts of materials in a pixel.)

Simulation: we employ the pruned U.S. Geological Survey (U.S.G.S.) library with $n = 342$ spectra vectors; each spectra vector has $m = 224$ elements; the synthetic HSI consists of $k = 4$ selected materials from the same library; number of pixels $p = 1000$; SNR=40dB.





Rank sparsity

- Rank minimization problem

$$\min_{X} \operatorname{rank}(X)$$
$$\text{s.t. } \mathcal{A}(X) = Y,$$

where $\mathcal{A}$ is a linear operator (i.e., $A \times \operatorname{vec}(X) = \operatorname{vec}(Y)$ for some matrix A).

- When X is restricted to be diagonal, $\operatorname{rank}(X) = \|\operatorname{diag}(X)\|_0$ and the rank minimization problem reduces to the SMV problem.

- Therefore, the rank minimization problem is more general (and more difficult) than the SMV problem.

- The nuclear norm $\|X\|_*$ is defined as the sum of singular values, i.e.

$$\|X\|_* = \sum_{i=1}^{r} \sigma_i.$$

- The nuclear norm is the convex envelope of the rank function on the convex set $\{X \mid \|X\|_2 \leq 1\}$.

- This motivates us to use nuclear norm to approximate the rank function.

$$\min_X \ \|X\|_* \quad \text{s.t. } \mathcal{A}(X) = Y.$$

- Perfect recovery is guaranteed if certain properties hold for $\mathcal{A}$.

- It can be shown that the nuclear norm $\|X\|_*$ can be computed by an SDP

$$\begin{aligned} \|X\|_* = \min_{Z_1, Z_2} \ & \tfrac{1}{2}\mathrm{tr}(Z_1 + Z_2) \\ \text{s.t.} \ & \begin{bmatrix} Z_1 & X \\ X^T & Z_2 \end{bmatrix} \succeq 0. \end{aligned}$$

- Therefore, the nuclear norm approximation can be turned to an SDP

$$\begin{aligned} \min_{X, Z_1, Z_2} \ & \tfrac{1}{2}\mathrm{tr}(Z_1 + Z_2) \\ \text{s.t.} \ & Y = \mathcal{A}(X) \\ & \begin{bmatrix} Z_1 & X \\ X^T & Z_2 \end{bmatrix} \succeq 0. \end{aligned}$$

Application: Matrix Completion Problem

- Recommendation system: recommend new movies to users based on their previous preference.
- Consider a preference matrix Y with y_{ij} representing how user i likes movie j.
- But some y_{ij} are unknown since no one watches all movies
- We would like to predict how users like new movies.
- Y is assumed to be of low rank, as researches show that only a few factors affect users' preferences.

$$\text{movies}$$

$$Y = \begin{bmatrix} 2 & 3 & 1 & ? & ? & 5 & 5 \\ 1 & ? & 4 & 2 & ? & ? & ? \\ ? & 3 & 1 & ? & 2 & 2 & 2 \\ ? & ? & ? & 3 & ? & 1 & 5 \\ 2 & ? & 4 & ? & ? & 5 & 3 \end{bmatrix} \quad \text{users}$$

- Low rank matrix completion

$$\begin{aligned} &\min \operatorname{rank}(X) \\ &\text{ s.t. } x_{ij} = y_{ij}, \text{ for } (i,j) \in \Omega, \end{aligned}$$

where Ω is the set of observed entries.

- Nuclear norm approximation

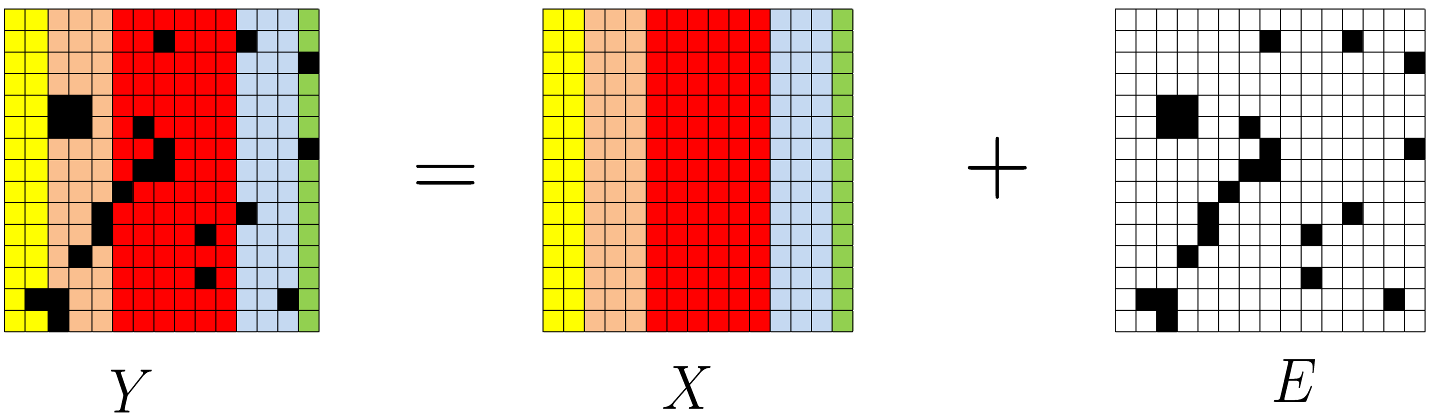
$$\begin{aligned} &\min \|X\|_* \\ &\text{ s.t. } x_{ij} = y_{ij}, \text{ for } (i,j) \in \Omega. \end{aligned}$$

Low-rank matrix + element-wise sparse corruption

- Consider the signal model

$$Y = X + E$$

where Y is the observation, X a low-rank signal, and E some sparse corruption with $|E_{ij}|$ arbitrarily large.



- The objective is to separate X from E via Y.

- Simultaneous rank and element-wise sparse recovery

$$\min \operatorname{rank}(X) + \gamma \|\operatorname{vec}(E)\|_0$$
$$\text{s.t. } Y = X + E,$$

where $\gamma \geq 0$ is used for balancing rank sparsity and element-wise sparsity.

- Replacing $\operatorname{rank}(X)$ by $\|X\|_*$ and $\|\operatorname{vec}(E)\|_0$ by $\|\operatorname{vec}(E)\|_1$, we have a convex problem:

$$\min \|X\|_* + \gamma \|\operatorname{vec}(E)\|_1$$
$$\text{s.t. } Y = X + E.$$

- A theoretical result indicates that when X is of low-rank and E is sparse enough, exact recovery happens with very high probability.

Application: Background extraction

- Suppose that we are given video sequences $F_i, i = 1, \ldots, p$.

- Our objective is to exact the background in the video sequences.

- The background is of low-rank, as the background is static within a short period of time.

- The foreground is sparse, as activities in the foreground only occupy a small fraction of space.

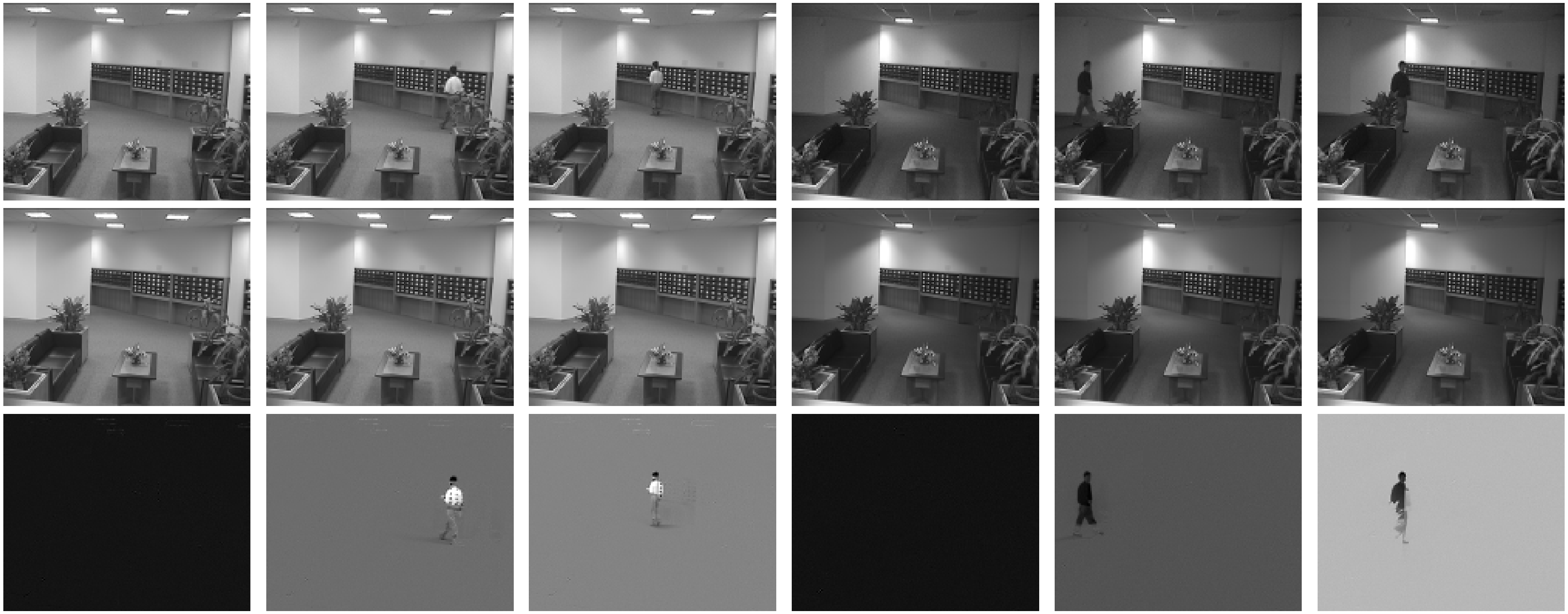
- Stacking the video sequences $Y = [\text{vec}(F_1), \ldots, \text{vec}(F_p)]$, we have

$$Y = X + E,$$

  where X represents the low-rank background, and E the sparse foreground.

- Nuclear norm and ℓ_1-norm approximation:

$$\begin{aligned} &\min \; \|X\|_* + \gamma \|\text{vec}(E)\|_1 \\ &\text{s.t. } Y = X + E. \end{aligned}$$



- 500 images, image size 160×128, $\gamma = 1/\sqrt{160 \times 128}$.
- Row 1: the original video sequences.
- Row 2: the extracted low-rank background.
- Row 3: the extracted sparse foreground.

Low-rank matrix + sparse corruption + dense noise

- A more general model

$$Y = \mathcal{A}(X + E) + V,$$

where X is low-rank, E sparse corruption, V dense but small noise, and $\mathcal{A}(\cdot)$ a linear operator.

- Simultaneous rank and element-wise sparse recovery with denoising

$$\min_{X,E,V} \ \text{rank}(X) + \gamma\|\text{vec}(E)\|_0 + \lambda\|V\|_F$$
$$\text{s.t. } Y = \mathcal{A}(X + E) + V.$$

- Convex approximation

$$\min_{X,E,V} \ \|X\|_* + \gamma\|\text{vec}(E)\|_1 + \lambda\|V\|_F$$
$$\text{s.t. } Y = \mathcal{A}(X + E) + V.$$

- A final remark: In sparse optimization, problem dimension is usually very large. You probably need fast custom-made algorithms instead of relying on CVX.

Reference

S. Boyd, "ℓ_1-norm methods for convex-cardinality problems", lecture note of EE364b, Standford university.

J Romberg, "Imaging via compressive sensing", *Signal Processing Magazine*, 2008.

Y. Ma, A. Yang, and J. Wright, "Sparse Representation and low-rank representation", tutorial note in European conference on computer vision, 2012

Y. C. Eldar and G. Kutyniok, "Compressed Sensing: Theory and Applications", Cambridge University Press, 2012.

J. Chen and X. Huo, "Theoretical Results on Sparse Representations of Multiple-Measurement Vectors", *IEEE Trans. Signal Process.*, 2006

D. Malioutov, M. Cetin, and A. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays", *IEEE Trans. Signal Process.*, 2005.

M.-D. Iordache, J. Bioucas-Dias, and A. Plaza, "Collaborative Sparse Regression for Hyperspectral Unmixing", *accepted for publication in IEEE Trans. Geosci. Remote Sens.*, 2013.