Section 5
Convex Optimisation 1

Convex Combination

Definition 5.1

A *convex combination* is a linear combination of points where all coefficients are non-negative and sum to 1.
More specifically, let $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_\ell \in \mathbb{R}^n$. A convex combination of these points is of the form

$$\sum_{i=1}^{\ell} \lambda_i \boldsymbol{x}_i,$$

where the real coefficients λ_i satisfy $\lambda_i \geq 0$ and $\sum_{i=1}^{n} \lambda_i = 1$.

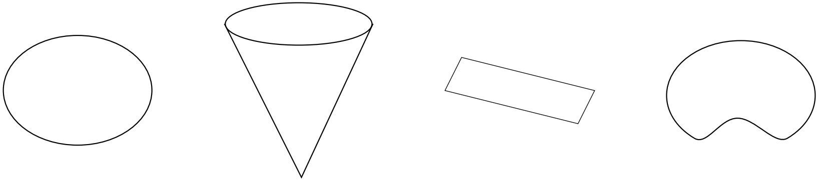
Convex Sets

Definition 5.2

A set $\mathcal{X}$ is a *convex set* if and only if the convex combination of any two points in the set belongs to the set.
That is,

$$\mathcal{X} \subseteq \mathbb{R}^n \text{ is convex } \Leftrightarrow \forall \boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathcal{X},\ \lambda \boldsymbol{x}_1 + (1-\lambda)\, \boldsymbol{x}_2 \in \mathcal{X},\ \forall \lambda \in [0,1]\,.$$

Examples



Example of convex sets:

- A *hyperplane* $\mathcal{H} = \left\{ \boldsymbol{x} : \ \boldsymbol{a}^T \boldsymbol{x} = b \right\}$, where $\boldsymbol{a} \in \mathbb{R}^n$, $\boldsymbol{a} \neq \mathbf{0}$, and $b \in \mathbb{R}$.
- A *halfspace* $\mathcal{H}_+ = \left\{ \boldsymbol{x} : \ \boldsymbol{a}^T \boldsymbol{x} \leq b \right\}$, where $\boldsymbol{a} \in \mathbb{R}^n$, $\boldsymbol{a} \neq 0$, and $b \in \mathbb{R}$.
- A *polyhedron*
  $\mathcal{P} = \left\{ \boldsymbol{x} : \ \boldsymbol{a}_j^T \boldsymbol{x} \leq b_j, \ j = 1, \cdots, m, \ \boldsymbol{c}_j^T \boldsymbol{x} = d_j, \ j = 1, \cdots, p \right\}$.
- Intersections of convex sets are convex.

Convex Functions

Definition 5.3

The *domain* of a function $f : \mathbb{R}^n \to \mathbb{R}$ is defined as the set of the points where the function f is finite, i.e.,

$$\mathrm{dom} f = \left\{ \boldsymbol{x} \in \mathbb{R}^n : \left| f\left(\boldsymbol{x}\right) \right| < \infty \right\}.$$

Example: $\mathrm{dom} \log x = \mathbb{R}^+$.

Definition 5.4 (Convex functions)

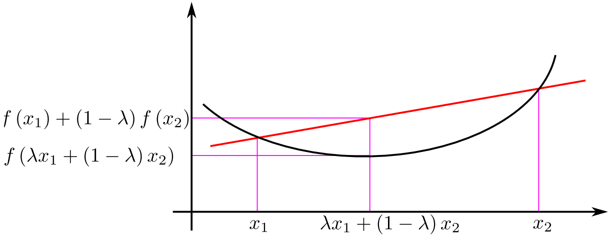
A function $f : \mathbb{R}^n \to \mathbb{R}$ is *convex* if for any $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathrm{dom} f \subseteq \mathbb{R}^n$, $\lambda \in [0, 1]$, it holds

$$\lambda f\left(\boldsymbol{x}_1\right) + \left(1 - \lambda\right) f\left(\boldsymbol{x}_2\right) \geq f\left(\lambda \boldsymbol{x}_1 + \left(1 - \lambda\right) \boldsymbol{x}_2\right).$$

This definition implies that $\mathrm{dom} f$ is convex. However, in this lecture notes, we usually assume $\mathrm{dom}\, f = \mathbb{R}^n$ for simplicity.

A function f is *strictly convex* if strict inequality holds whenever $\boldsymbol{x} \neq \boldsymbol{y}$ and $\lambda \in (0, 1)$.

A Convex Function

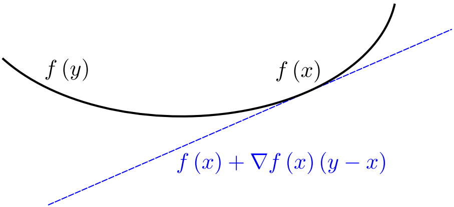


First-Order Condition of Convexity

Theorem 5.5

Suppose a function $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable. Then it is convex if and only if for all $\boldsymbol{x}, \boldsymbol{y} \in \operatorname{dom} f$, it holds

$$f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^T (\boldsymbol{y} - \boldsymbol{x}). \quad (12)$$



Necessity

Assume first that f is convex and $\boldsymbol{x}, \boldsymbol{y} \in \operatorname{dom}(f)$. Since $\operatorname{dom}(f)$ is convex, $\boldsymbol{x} + t(\boldsymbol{y} - \boldsymbol{x}) \in \operatorname{dom}(f)$ for all $0 < t \leq 1$. By convexity of f,

$$f(\boldsymbol{x} + t(\boldsymbol{y} - \boldsymbol{x})) \leq (1 - t) f(\boldsymbol{x}) + t f(\boldsymbol{y}).$$

Divide both sides by t. It holds

$$f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + \frac{f(\boldsymbol{x} + t(\boldsymbol{y} - \boldsymbol{x})) - f(\boldsymbol{x})}{t}.$$

Take the limit as $t \to 0$ yields (12).

Sufficiency

To show the other direction (sufficiency), assume that (12) holds. Choose any $\boldsymbol{x} \neq \boldsymbol{y}$ and $\lambda \in [0, 1]$. Let $\boldsymbol{z} = \lambda \boldsymbol{x} + (1 - \lambda)\, \boldsymbol{y}$. Applying (12) twice yields

$$f(\boldsymbol{x}) - f(\boldsymbol{z}) \geq \nabla f(\boldsymbol{z})^T (\boldsymbol{x} - \boldsymbol{z}),$$
$$f(\boldsymbol{y}) - f(\boldsymbol{z}) \geq \nabla f(\boldsymbol{z})^T (\boldsymbol{y} - \boldsymbol{z}).$$

Multiply the first inequality by λ and the second by $1 - \lambda$, and then add them together. It holds

$$\lambda f(\boldsymbol{x}) + (1 - \lambda)\, f(\boldsymbol{y}) - f(\boldsymbol{z}) \geq \nabla f(\boldsymbol{z})^T (\lambda \boldsymbol{x} + (1 - \lambda)\, \boldsymbol{y} - \boldsymbol{z}).$$

By the definition of $\boldsymbol{z}$, the left side of the inequality is zero. Hence,

$$\lambda f(\boldsymbol{x}) + (1 - \lambda)\, f(\boldsymbol{y}) \geq f(\boldsymbol{z}),$$

which proves that f is convex.

Sublevel Sets

Definition 5.6 (Sublevel Sets, a.k.a. Lower Contour Sets)

The α-*sublevel set* of a function $f:\ \mathbb{R}^n \to \mathbb{R}$ is defined as

$$\mathcal{C}_\alpha = \{\boldsymbol{x} \in \mathrm{dom}\,(f) :\ f\,(\boldsymbol{x}) \leq \alpha\}\,.$$

Sublevel Sets of Convex Functions

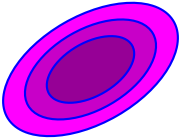
Lemma 5.7

Sublevel sets of a convex function f are convex.

Proof: We shall show that for all $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{C}_\alpha$, it holds $\lambda \boldsymbol{x} + (1-\lambda)\, \boldsymbol{y} \in \mathcal{C}_\alpha$ for all $\lambda \in [0,1]$. By the definition of $\mathcal{C}_\alpha$, $f(\boldsymbol{x}) \le \alpha$ and $f(\boldsymbol{y}) \le \alpha$. By the convexity of f,

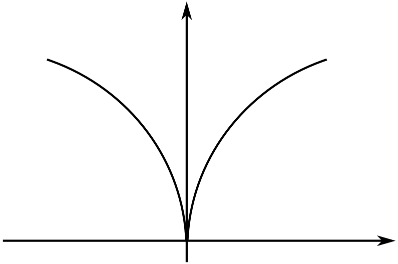
$$f(\lambda \boldsymbol{x} + (1-\lambda)\, \boldsymbol{y}) \le \lambda f(\boldsymbol{x}) + (1-\lambda)\, f(\boldsymbol{y}) \le \alpha,$$

which proves this proposition.



Sublevel Sets

The converse of Lemma 5.7 is not true.
That sublevel sets of a function f are convex does not imply that f is convex.



Norm

We've seen ℓ_p-norm in Definition 4.7.

Definition 5.8

Given a vector space $\mathcal{V}$ over the field $\mathbb{F}$ of complex (real) numbers, a norm on $\mathcal{V}$ is a function $p: \mathcal{V} \to \mathbb{R}$ with the following properties:
For all $a \in \mathbb{F}$ and all $\boldsymbol{u}, \boldsymbol{v} \in \mathcal{V}$,

1. $p(a\boldsymbol{v}) = |a| \, p(\boldsymbol{v})$, (absolute scalability)
2. $p(\boldsymbol{u} + \boldsymbol{v}) \leq p(\boldsymbol{u}) + p(\boldsymbol{v})$, (triangle inequality)
3. if $p(\boldsymbol{v}) = 0$ then $\boldsymbol{v}$ is the zero vector. (separates points)

Positivity follows: By the first axiom, $p(\boldsymbol{0}) = 0$ and $p(-\boldsymbol{v}) = p(\boldsymbol{v})$.
Then by triangle inequality,

$$0 \leq p(\boldsymbol{v}) + p(-\boldsymbol{v}) = 2p(\boldsymbol{v}) \;\Rightarrow\; 0 \leq p(\boldsymbol{v}).$$

Convexity of a Norm

Lemma 5.9

A norm is a convex function.

Proof: For any given $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$ and $\lambda \in [0, 1]$, it holds that

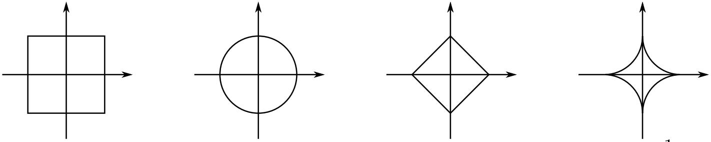
$$\begin{aligned} \|\lambda \boldsymbol{u} + (1-\lambda)\,\boldsymbol{v}\| &\le \|\lambda \boldsymbol{u}\| + \|(1-\lambda)\,\boldsymbol{v}\| \\ &= \lambda\,\|\boldsymbol{u}\| + (1-\lambda)\,\|\boldsymbol{v}\|, \end{aligned}$$

where we have used the triangle inequality and the absolute scalability. This establishes the convexity of the norm.

ℓ_p-Norm

In Definition 4.7, it mentioned that ℓ_p-norm is a proper norm iff $p \geq 1$.

Can be verified by using sub-level argument.



$p = \infty$ $\quad$ $p = 2$ $\quad$ $p = 1$ $\quad$ $p = \frac{1}{2}$

Constrained Convex Optimization Problems

A constrained optimization problem of the form

$$\begin{aligned} &\min_{\boldsymbol{x}} f(\boldsymbol{x}) \\ \text{subject to } & h_i(\boldsymbol{x}) \le 0,\ i = 1, \cdots, m, \\ & \ell_i(\boldsymbol{x}) = 0,\ i = 1, \cdots, r, \end{aligned}$$

is convex if

- the objective function f_0 is convex, and
- the feasible set is convex.
  - h_i's are convex (consequence of Lemma 5.7).
  - ℓ_i's are affine, i.e., in the form of $\boldsymbol{a}_i^T \boldsymbol{x} + b_i = 0$.
    $\ell_i(\boldsymbol{x}) = 0 \Leftrightarrow \ell_i(\boldsymbol{x}) \le 0$ and $-\ell_i(\boldsymbol{x}) \le 0$.
    Both ℓ_i and $-\ell_i$ need to be convex $\Rightarrow \ell_i$ is affine.

Local Optimality and Global Optimality

Theorem 5.10

Suppose that a feasible point $\boldsymbol{x}$ is locally optimal for a convex optimization problem. Then it is also globally optimal.

Proof: Suppose that $\boldsymbol{x}$ is locally optimal but not globally optimal, i.e., there exists a feasible $\boldsymbol{y} \neq \boldsymbol{x}$ such that $f(\boldsymbol{y}) < f(\boldsymbol{x})$. Consider a point $\boldsymbol{z}$ on the line segment between $\boldsymbol{x}$ and $\boldsymbol{y}$, i.e.,

$$\boldsymbol{z} = (1-\lambda)\,\boldsymbol{x} + \lambda \boldsymbol{y},\ \lambda \in (0,1).$$

Then it is clear that

$$\begin{aligned} & f(\boldsymbol{z}) \leq (1-\lambda) f(\boldsymbol{x}) + \lambda f(\boldsymbol{y}) < f(\boldsymbol{x}), \\ & h_i(\boldsymbol{z}) \leq (1-\lambda) h_i(\boldsymbol{x}) + \lambda h_i(\boldsymbol{y}) \leq 0,\ i = 0, 1, \cdots, m, \\ & \boldsymbol{a}_i^T \boldsymbol{z} = (1-\lambda)\,\boldsymbol{a}_i^T \boldsymbol{x} + \lambda \boldsymbol{a}_i^T \boldsymbol{y} = b_i,\ i = 1, \cdots, r, \end{aligned}$$

where the inequalities follow from the convexity of the functions f and h_i's. Hence, the point $\boldsymbol{z}$ is feasible and $f(\boldsymbol{z}) < f(\boldsymbol{x})$ for all $\lambda \in (0,1)$. This contradicts with that $\boldsymbol{x}$ is locally optimal and proves the global optimality of $\boldsymbol{x}$.

A Global Optimality Criterion

Theorem 5.11

Suppose that the objective f_0 in a convex optimization problem is differentiable, i.e.,

$$f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^T (\boldsymbol{y} - \boldsymbol{x}), \ \forall \boldsymbol{x}, \boldsymbol{y} \in \mathrm{dom}(f).$$

Let $\mathcal{X}$ denote the feasible set

$$\mathcal{X} = \left\{ \boldsymbol{x} : \ h_i(\boldsymbol{x}) \leq 0, \ i = 1, \cdots, m, \ \boldsymbol{a}_i^T \boldsymbol{x} = b_i, \ i = 1, \cdots, r \right\}.$$

Then an $\boldsymbol{x} \in \mathcal{X}$ is optimal if and only if

$$\nabla f(\boldsymbol{x})^T (\boldsymbol{y} - \boldsymbol{x}) \geq 0, \ \forall \boldsymbol{y} \in \mathcal{X}.$$

Consequence of Theorem 5.11

- For an unconstrained convex optimization problem, the sufficient and necessary condition for a globally optimal point $\boldsymbol{x}$ is given by

$$\nabla f(\boldsymbol{x}) = \mathbf{0}.$$

- In a constrained convex optimization problem, it may happen that

$$\nabla f(\boldsymbol{x}) \neq \mathbf{0}.$$

  This implies that $\boldsymbol{x}$ is at the boundary of the feasible set. (This is actually linked to KKT conditions and will be discussed later.)

Proof

The proof of sufficiency is straightforward. Suppose the inequality holds. Then for all $\boldsymbol{y} \in \mathcal{X}$,

$$f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^T (\boldsymbol{y} - \boldsymbol{x}) \geq f(\boldsymbol{x}).$$

Hence, the point $\boldsymbol{x}$ is globally optimal.

Conversely, suppose $\boldsymbol{x}$ is optimal, but the inequality does not hold, i.e., for some $\boldsymbol{y} \in \mathcal{X}$ we have

$$\nabla f(\boldsymbol{x})^T (\boldsymbol{y} - \boldsymbol{x}) < 0.$$

Consider the point $\boldsymbol{z}(t) = t\boldsymbol{y} + (1-t)\boldsymbol{x},\ t \in [0,1]$. Clearly, $\boldsymbol{z}(t)$ is feasible. Now

$$\begin{aligned} \tfrac{d}{dt} f(\boldsymbol{z}(t))\big|_{t=0} &= \nabla f(\boldsymbol{z}(0)) \cdot \tfrac{d}{dt} \boldsymbol{z}(t)\big|_{t=0} \\ &= \nabla f(\boldsymbol{x}) \cdot (\boldsymbol{y} - \boldsymbol{x}) < 0, \end{aligned}$$

where the inequality comes from the assumption. It implies that for small positive t, we have $f(\boldsymbol{z}(t)) < f(\boldsymbol{x})$, which contradicts the optimality of $\boldsymbol{x}$. The necessity is therefore proved.

Non-differentiable Functions: Subgradient

Definition 5.12

If $f : \mathcal{U} \to \mathbb{R}$ is a convex function defined on a convex open set $\mathcal{U} \subset \mathbb{R}^n$, a vector $\boldsymbol{v} \in \mathbb{R}^n$ is called a subgradient at a point $\boldsymbol{x} \in \mathcal{U}$ if

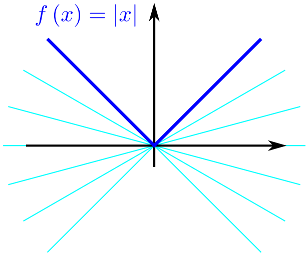
$$f(\boldsymbol{y}) - f(\boldsymbol{x}) \geq \boldsymbol{v}^T (\boldsymbol{y} - \boldsymbol{x}), \ \forall \boldsymbol{y} \in \mathcal{U}.$$

The set of all subgradients at $\boldsymbol{x}$ is called the subdifferential at $\boldsymbol{x}$ and is denoted $\partial f(\boldsymbol{x})$.

Remark: If f is convex and its subdifferential at $\boldsymbol{x}$ contains exactly one subgradient, then f is differentiable at $\boldsymbol{x}$.

Example

$$f\left(x\right)=\left|x\right|\;\Rightarrow\;\partial f=\begin{cases}1 & \text{if } x>0,\\ \left[-1,1\right] & \text{if } x=0,\\ -1 & \text{if } x<0.\end{cases}$$



Section 6
ℓ_1-Minimization

Three Algorithms

- Cyclic Coordinate Descent (CCD)
- Iterative Shrinkage Thresholding (IST)
- Least Angle Regression (LAR)

ℓ_1-Minimization

Want to solve the sparse linear inverse problem:

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{e}.$$

Constrained optimization problem: if we know $\|\boldsymbol{e}\| \le \epsilon$,

$$\min_{\boldsymbol{x}} \|\boldsymbol{x}\|_1 \text{ subject to } \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_2 \le \epsilon.$$

Unconstrained optimization problem: LASSO

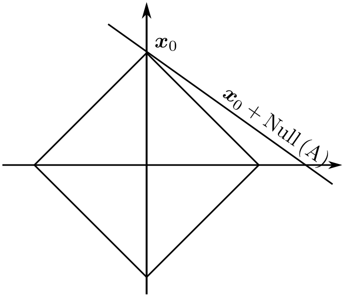
$$\min_{\boldsymbol{x}} \frac{1}{2} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \lambda \|\boldsymbol{x}\|_1 .$$

$\exists$ a one-to-one correspondence between ϵ and λ.

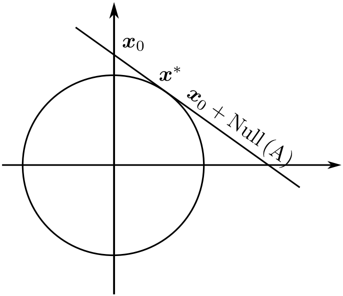
- $\lambda \to 0$ implies $\epsilon \to 0$.
- $\lambda \to \infty$ implies $\epsilon \to \infty$.

Why ℓ_1-Minimization

A geometric intuition:



ℓ_1 tends to give sparse solutions



ℓ_2 tends to give non-sparse solutions

Feasible solution for $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$: $\boldsymbol{x} \in \mathcal{X} = \boldsymbol{x}_0 + \mathcal{N}ull\,(\boldsymbol{A})$.

Scalar Lasso Problem

$$\min_x \underbrace{\frac{1}{2}(x-y)^2 + \lambda |x|}_{f(x)}.$$

The minimum of $f(x)$ is achieved at $x^{\#}$ s.t. $\frac{d}{dx} f\left(x^{\#}\right) = 0$:

$$x^{\#} - y + \lambda\, \partial_x |x||_{x^{\#}} = 0,$$

where

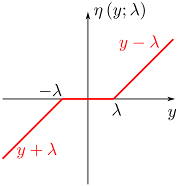
$$\partial_x |x| = \begin{cases} 1 & \text{if } x > 0, \\ [-1, 1] & \text{if } x = 0, \\ -1 & \text{if } x < 0. \end{cases}$$

Scalar Lasso Problem: The Solution

$x^{\#}$ is given by the soft thresholding function:

$$x^{\#} = \begin{cases} y - \lambda & \text{if } y > \lambda, \\ 0 & \text{if } |y| \leq \lambda, \\ y + \lambda & \text{if } y < -\lambda. \end{cases}$$
$$= \eta(y; \lambda) = \eta_\lambda(y) = \mathrm{sign}(y)(|y| - \lambda)_+ ,$$

where $(z)_+ = \max(z, 0)$.



Lasso Problem: Scalar Input Vector Observation

Assume that $\|\boldsymbol{a}\|_2^2 = 1$. Consider the problem

$$\min_x \frac{1}{2} \|\boldsymbol{y} - \boldsymbol{a}x\|_2^2 + \lambda |x|$$

It optimal solution $x^{\#}$ is given by

$$x^{\#} = \begin{cases} \langle \boldsymbol{y}, \boldsymbol{a} \rangle - \lambda & \text{if } \langle \boldsymbol{y}, \boldsymbol{a} \rangle > \lambda, \\ 0 & \text{if } |\langle \boldsymbol{y}, \boldsymbol{a} \rangle| \leq \lambda, \\ \langle \boldsymbol{y}, \boldsymbol{a} \rangle + \lambda & \text{if } \langle \boldsymbol{y}, \boldsymbol{a} \rangle < -\lambda. \end{cases}$$
$$= \eta_\lambda \left(\langle \boldsymbol{y}, \boldsymbol{a} \rangle \right).$$

Solving General Lasso: Cyclic Coordinate Descent

$$\min_{\boldsymbol{x}} \frac{1}{2} \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2 + \lambda \|\boldsymbol{x}\|_1$$

Objective function with respect to x_i:

- $\frac{1}{2} \left\|\boldsymbol{y} - \sum_{j\neq i} \boldsymbol{a}_j x_j - \boldsymbol{a}_i x_i\right\|_2^2 + \lambda \sum_{j\neq i} |x_j| + \lambda |x_i|$
- $\frac{1}{2} \|\boldsymbol{r}_i - \boldsymbol{a}_i x_i\|_2^2 + c + \lambda |x_i|$

Optimal solution for x_i is given by

$$\begin{aligned} x_i^{\#} &= \eta_\lambda \left(\langle \boldsymbol{a}_i, \boldsymbol{r}_i \rangle\right) = \eta_\lambda \left(\left\langle \boldsymbol{a}_i, \boldsymbol{y} - \sum_{j\neq i} \boldsymbol{a}_j \hat{x}_j \right\rangle \right) \\ &= \eta_\lambda \left(\hat{x}_i + \left\langle \boldsymbol{a}_i, \boldsymbol{y} - \sum_{j} \boldsymbol{a}_j \hat{x}_j \right\rangle \right) \end{aligned}$$

Three Algorithms

- Cyclic Coordinate Descent (CCD)
- Iterative Shrinkage Thresholding (IST)
- Least Angle Regression (LAR)

The Gradient Descent Method

Gradient descent method: To solve $\min_{\boldsymbol{x}} \; f(\boldsymbol{x})$,
one iteratively updates

$$\boldsymbol{x}^k = \boldsymbol{x}^{k-1} - t_k \nabla f\left(\boldsymbol{x}^{k-1}\right),$$

where $t_k > 0$ is a suitable stepsize.

For Lasso problem $f(x) = \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2 + \lambda \|\boldsymbol{x}\|_1$ which is non-smooth.
Its gradient is given by (see details on page 6-35)

$$-\boldsymbol{A}^T(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}) + \partial \|\boldsymbol{x}\|_1 .$$

Gradient descent converges very slow.

Gradient Descent Method: Another View

In gradient descent method:

$$\boldsymbol{x}^k = \boldsymbol{x}^{k-1} - t_k \nabla f\left(\boldsymbol{x}^{k-1}\right).$$

This is equivalent to minimize $\tilde{f}$,

$$\boldsymbol{x}_k = \arg\min_{\boldsymbol{x}} \tilde{f}\left(\boldsymbol{x}\right)$$

where

$$\begin{aligned}\tilde{f}\left(\boldsymbol{x}\right) &:= f\left(\boldsymbol{x}^{k-1}\right) + \left\langle \boldsymbol{x} - \boldsymbol{x}^{k-1}, \nabla f\left(\boldsymbol{x}^{k-1}\right)\right\rangle + \frac{1}{2t_k}\left\|\boldsymbol{x} - \boldsymbol{x}^{k-1}\right\|_2^2 \\ &= \frac{1}{2t_k}\left\|\boldsymbol{x} - \left(\boldsymbol{x}^{k-1} - t_k \nabla f\left(\boldsymbol{x}^{k-1}\right)\right)\right\|_2^2 + c.\end{aligned}$$

Iterative Shrinkage Thresholding (IST)

To solve $\min_{\boldsymbol{x}} f(\boldsymbol{x}) + \lambda \|\boldsymbol{x}\|_1$, we apply the proximal regularization:

$$\boldsymbol{x}^k = \arg \min_{\boldsymbol{x}} \tilde{f}(\boldsymbol{x}) + \lambda \|\boldsymbol{x}\|_1$$

where

$$\begin{aligned}
&\tilde{f}(\boldsymbol{x}) + \lambda \|\boldsymbol{x}\|_1 \\
&:= f\left(\boldsymbol{x}^{k-1}\right) + \left\langle \boldsymbol{x} - \boldsymbol{x}^{k-1}, \nabla f\left(\boldsymbol{x}^{k-1}\right)\right\rangle + \frac{1}{2t_k}\left\|\boldsymbol{x} - \boldsymbol{x}^{k-1}\right\|_2^2 + \lambda \|\boldsymbol{x}\|_1 \\
&= \frac{1}{2t_k}\left\|\boldsymbol{x} - \left(\boldsymbol{x}^{k-1} - t_k \nabla f\left(\boldsymbol{x}^{k-1}\right)\right)\right\|_2^2 + \lambda \|\boldsymbol{x}\|_1 + c \\
&= \sum_i \left[\frac{1}{2t_k}(x_i - z_i)^2 + \lambda |x_i|\right] + c.
\end{aligned}$$

Therefore,

$$\boldsymbol{x}^k = \eta\left(\boldsymbol{x}^{k-1} + t_k \boldsymbol{A}^T\left(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}^{k-1}\right); \lambda t_k\right).$$

Three Algorithms

- Cyclic Coordinate Descent (CCD)
- Iterative Shrinkage Thresholding (IST)
- Least Angle Regression (LAR)

Lasso and Sparsity

$$\min_{\boldsymbol{x}} \frac{1}{2} \left\| \boldsymbol{y} - \boldsymbol{A}\boldsymbol{x} \right\|_2^2 + \lambda \left\| \boldsymbol{x} \right\|_1$$

$\lambda = 0$: $\boldsymbol{x}$ is not sparse.
$\lambda \to \infty$: $\boldsymbol{x} = \mathbf{0}$.
$\lambda \in [0, \infty)$: one-to-one correspondence between λ and $\|\boldsymbol{x}_\lambda\|_0$, where

$$\boldsymbol{x}_\lambda := \arg \min_{\boldsymbol{x}} \frac{1}{2} \left\| \boldsymbol{y} - \boldsymbol{A}\boldsymbol{x} \right\|_2^2 + \lambda \left\| \boldsymbol{x} \right\|_1 .$$

Least angle regression: Find a λ to give an x with a specific sparsity.

Piecewise Linearity

Theorem 6.1

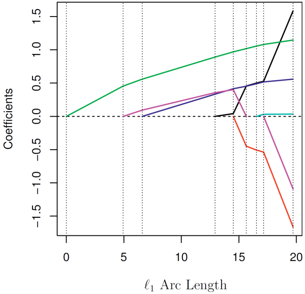
$\boldsymbol{x}_\lambda$ is a piecewise linear function of λ.

Proof: $\boldsymbol{x}_\lambda$ is an optimal solution to the Lasso problem if and only if

$$\boldsymbol{A}^T \left(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}_\lambda\right) = \lambda \text{sign}\left(\boldsymbol{x}_\lambda\right).$$

Let $0 < \lambda_1 < \lambda_2$ be two sufficiently close values of λ, so that going from the solution $\boldsymbol{x}_{\lambda_1}$ to $\boldsymbol{x}_{\lambda_2}$ does not require any coordinate of $\boldsymbol{x}_\lambda$ to change its sign. Then it is easy to see that, for all $\lambda = \alpha\lambda_1 + (1-\alpha)\lambda_2$, $\alpha \in [0,1]$, $\hat{\boldsymbol{x}} = \alpha\boldsymbol{x}_{\lambda_1} + (1-\alpha)\boldsymbol{x}_{\lambda_2}$ satisfies the optimality condition. Therefore, $\boldsymbol{x}_\lambda = \hat{\boldsymbol{x}} = \alpha\boldsymbol{x}_{\lambda_1} + (1-\alpha)\boldsymbol{x}_{\lambda_2}$.

Typical Behavior of $\boldsymbol{x}_\lambda$



T. Hastie, et al., Statistical learning with sparsity: the lasso and generalizations. Chapman and Hall/CRC, 2015: page 120.

Find the knots, i.e., $\lambda_s > 0$ such that

$$\lim_{\epsilon \to 0^+} \operatorname{sign}(\boldsymbol{x}_{\lambda_s+\epsilon}) \neq \lim_{\epsilon \to 0^+} \operatorname{sign}(\boldsymbol{x}_{\lambda_s-\epsilon})$$

Finding λ_0

Goal: to find λ_0 such that

$$0 = \lim_{\epsilon \to 0^+} \|\boldsymbol{x}_{\lambda_0+\epsilon}\|_0 \neq \lim_{\epsilon \to 0^+} \|\boldsymbol{x}_{\lambda_s-\epsilon}\|_0 = 1.$$

When $\|\boldsymbol{x}_\lambda\|_0 = 1$, let $\mathrm{supp}(\boldsymbol{x}_\lambda) = \{i\}$.
Then the Lasso problem is reduced to $\frac{1}{2}\|\boldsymbol{y} - \boldsymbol{a}_i x_i\|_2^2 + \lambda |x_i|$, and

$$x_i^{\#} = (|\boldsymbol{a}_i^T \boldsymbol{y}| - \lambda)\mathrm{sign}(\boldsymbol{a}_i^T \boldsymbol{y}).$$

This implies that

$$\lambda_0 = \mathrm{max}_j \ |\boldsymbol{a}_j^T \boldsymbol{y}|.$$

- For a $\lambda > \lambda_0$, $\|\boldsymbol{x}_\lambda\|_0 = 0$
- For a sufficiently small $\epsilon > 0$ and $\lambda \in (\lambda_0 - \epsilon, \lambda_0)$, $\|\boldsymbol{x}_\lambda\|_0 = 1$

Finding the Next Knot (1)

Starting from λ_{s-1}, want to find the next knot $\lambda_s < \lambda_{s-1}$.

- For a $\lambda \in [\lambda_s, \lambda_{s-1}]$, let $\mathcal{I} = \mathrm{supp}\,(\boldsymbol{x}_\lambda)$ and $\boldsymbol{\delta}_\lambda = \boldsymbol{x}_\lambda - \boldsymbol{x}_{s-1}$:

$$
\begin{aligned}
\lambda \mathrm{sign}(\boldsymbol{x}_{\lambda,\mathcal{I}}) &= \boldsymbol{A}_{\mathcal{I}}^T(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}_\lambda) \\
&= \boldsymbol{A}_{\mathcal{I}}^T(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}_{s-1} - \boldsymbol{A}\boldsymbol{\delta}_\lambda) \\
&= \boldsymbol{A}_{\mathcal{I}}^T\boldsymbol{r}_{s-1} - \boldsymbol{A}_{\mathcal{I}}^T\boldsymbol{A}_{\mathcal{I}}\boldsymbol{\delta}_{\lambda,\mathcal{I}}
\end{aligned}
$$

- But $\boldsymbol{A}_{\mathcal{I}}^T\boldsymbol{r}_{s-1} = \lambda_{s-1}\mathrm{sign}(\boldsymbol{x}_{\lambda_{s-1},\mathcal{I}}) = \lambda_{s-1}\mathrm{sign}(\boldsymbol{x}_{\lambda,\mathcal{I}})$.

$$
\begin{aligned}
\boldsymbol{A}_{\mathcal{I}}^T\boldsymbol{A}_{\mathcal{I}}\boldsymbol{\delta}_{\lambda,\mathcal{I}} &= \boldsymbol{A}_{\mathcal{I}}^T\boldsymbol{r}_{s-1} - \lambda\mathrm{sign}(\boldsymbol{x}_{\lambda,\mathcal{I}}) \\
&= (\lambda_{s-1} - \lambda)\,\mathrm{sign}(\boldsymbol{x}_{\lambda,\mathcal{I}}).
\end{aligned}
$$

  Hence $\boldsymbol{x}_\lambda = \boldsymbol{x}_{s-1} + \boldsymbol{\delta}_\lambda$ where

$$
\boldsymbol{\delta}_{\lambda,\mathcal{I}} = \frac{\lambda_{s-1} - \lambda}{\lambda_{s-1}} \left(\boldsymbol{A}_{\mathcal{I}}^T\boldsymbol{A}_{\mathcal{I}}\right)^{-1} \boldsymbol{A}_{\mathcal{I}}^T\boldsymbol{r}_{s-1}.
$$

Finding the Next Knot (2)

- Keep track $\boldsymbol{x}_\lambda$ and $|\langle \boldsymbol{a}_j, \boldsymbol{r}_\lambda \rangle|$ until
  - Either
    $$\max_{j \notin \mathcal{I}} |\langle \boldsymbol{a}_j, \boldsymbol{r}_\lambda \rangle| = \lambda.$$
    Define $i = \arg \max_{j \notin \mathcal{I}} |\langle \boldsymbol{a}_j, \boldsymbol{r}_\lambda \rangle|$ and set $\mathcal{I} = \mathcal{I} \bigcup \{i\}$.
  - Or for some $i \in \mathcal{I}$,
    $$(\boldsymbol{x}_\lambda)_i = 0.$$
    Set $\mathcal{I} = \mathcal{I} \backslash \{i\}$.

  Set λ_s accordingly.

Least Angle Regression

1. $\boldsymbol{r}_0 = \boldsymbol{y}$ and $\boldsymbol{x} = \boldsymbol{0}$.
2. Let $\lambda_0 = \max_j \ |\langle \boldsymbol{a}_j, \boldsymbol{r}_0 \rangle|$, $i = \arg\ \max_j \ |\langle \boldsymbol{a}_j, \boldsymbol{r}_0 \rangle|$ and $\mathcal{I} = \{i\}$.
3. For $s = 1, 2, \cdots$, do
   - 3.1 Find the next knot λ_s.
   - 3.2 Set $\boldsymbol{x}_s = \boldsymbol{x}_{s-1} + \boldsymbol{\delta}_{\lambda_s}$ and $\boldsymbol{r}_s = \boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}_s$.

Return the sequence $\{\lambda_s, \boldsymbol{x}_s\}$, $s = 0, 1, 2, \cdots$

Stable Recovery of Exact Sparse Signals

Theorem 6.2

Let S be such that $\delta_{4S} \le \frac{1}{2}$. Then for any signal $\boldsymbol{x}_0$ supported on $\mathcal{T}_0$ with $|\mathcal{T}_0| \le S$ and any perturbation $\boldsymbol{e}$ with $\|\boldsymbol{e}\|_2 \le \epsilon$, the solution $\boldsymbol{x}^{\#}$ obeys

$$\left\| \boldsymbol{x}^{\#} - \boldsymbol{x}_0 \right\|_2 \le C_S \cdot \epsilon,$$

where the constant C_S depends only on δ_{4S}.

Typical value of C_S

$$C_S \approx \begin{cases} 8.82 & \text{for } \delta_{4S} = \frac{1}{5}, \\ 10.47 & \text{for } \delta_{4S} = \frac{1}{4}. \end{cases}$$

Stable Recovery of Approximately Sparse Signals

Theorem 6.3

Suppose that $\boldsymbol{x}_0$ *is an an arbitrary vector in* $\mathbb{R}^n$ *and let* $\boldsymbol{x}_{0,S}$ *be the truncated vector corresponding to the* S *largest values of* $\boldsymbol{x}_0$ *(in absolute value). When the matrix* $\boldsymbol{A}$ *satisfies RIP, the solution* $\boldsymbol{x}^{\#}$ *obeys*

$$\left\|\boldsymbol{x}^{\#} - \boldsymbol{x}_0\right\|_2 \leq C_{1,S} \cdot \epsilon + C_{2,S} \cdot \frac{\|\boldsymbol{x}_0 - \boldsymbol{x}_{0,S}\|_1}{\sqrt{S}}.$$

No algorithm performs fundamentally better than ℓ_1-min.

Typical values

$C_{1,S} \approx 12.04$ and $C_{2,S} \approx 8.77$ for $\delta_{4S} = \frac{1}{5}$.

Analysis for Exact Sparse Signals (1)

Assume that $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{w}$, $\|\boldsymbol{x}\|_0 \leq S$, and $\|\boldsymbol{w}\|_2 \leq \epsilon$.
Cast the recovery problem as

$$\min_{\boldsymbol{x}} \|\boldsymbol{x}\|_1 \text{ subject to } \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_2 \leq \epsilon.$$

Tube constraint:

$$\|\boldsymbol{A}\boldsymbol{h}\|_2 = \left\|\boldsymbol{A}\boldsymbol{x}^{\#} - \boldsymbol{A}\boldsymbol{x}_0\right\|_2 \leq \left\|\boldsymbol{A}\boldsymbol{x}^{\#} - \boldsymbol{y}\right\|_2 + \|\boldsymbol{A}\boldsymbol{x}_0 - \boldsymbol{y}\|_2 \leq 2\epsilon.$$

Analysis for Exact Sparse Signals (2)

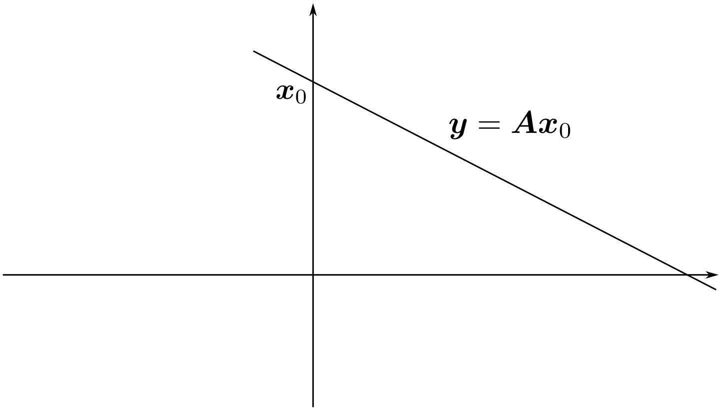
Cone constraint: Let $\boldsymbol{x}^{\#} = \boldsymbol{x}_0 + \boldsymbol{h}$. Then

$$\left\| \boldsymbol{h}_{\mathcal{T}_0^c} \right\|_1 \leq \left\| \boldsymbol{h}_{\mathcal{T}_0} \right\|_1 .$$

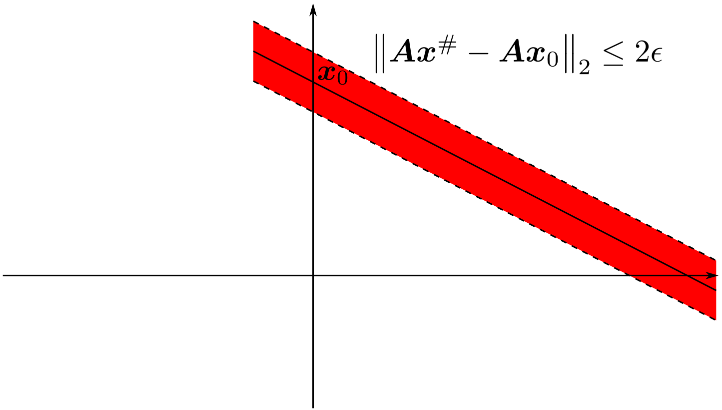
Proof:

$$\begin{aligned} \left\| \boldsymbol{x}_0 \right\|_1 &\geq \left\| \boldsymbol{x}^{\#} \right\|_1 = \left\| \boldsymbol{x}_0 + \boldsymbol{h} \right\|_1 \\ &= \left\| (\boldsymbol{x}_0 + \boldsymbol{h})_{\mathcal{T}_0} \right\|_1 + \left\| \boldsymbol{h}_{\mathcal{T}_0^c} \right\|_1 \\ &\geq \left\| \boldsymbol{x}_0 \right\|_1 - \left\| \boldsymbol{h}_{\mathcal{T}_0} \right\|_1 + \left\| \boldsymbol{h}_{\mathcal{T}_0^c} \right\|_1 . \end{aligned}$$

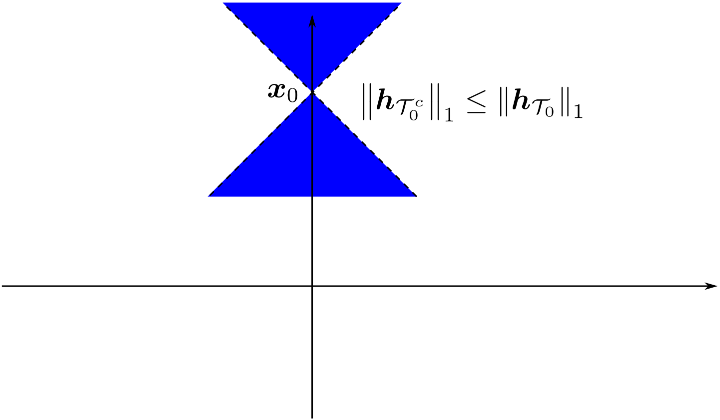
Geometric Interpretation



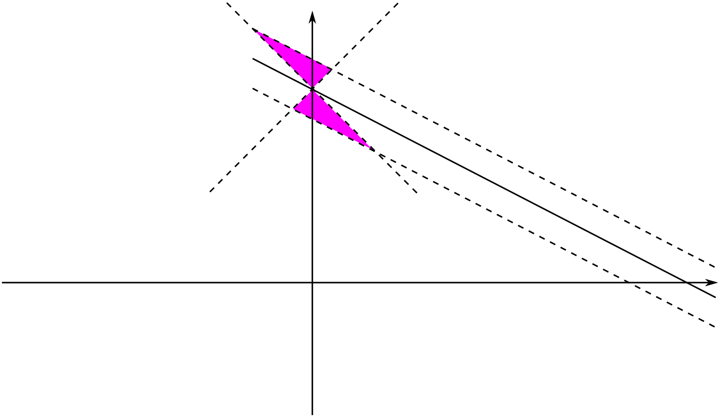
Geometric Interpretation



Geometric Interpretation



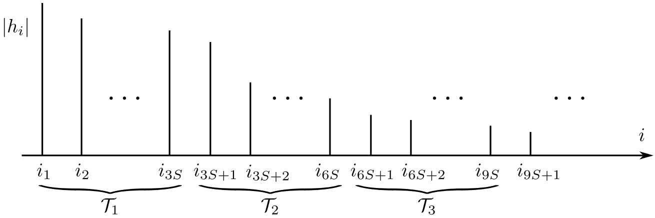
Geometric Interpretation



Proof

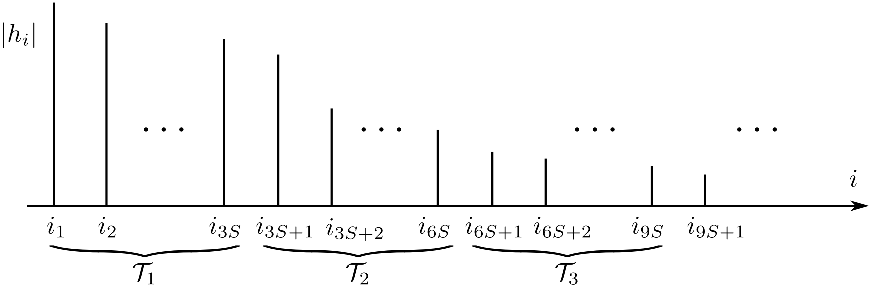
Since $\|\boldsymbol{A}\boldsymbol{h}\|_2 \le 2\epsilon$, want to show $\|\boldsymbol{h}\|_2 \approx \|\boldsymbol{A}\boldsymbol{h}\|_2$.
(This is not true in general. For example $\boldsymbol{A}\boldsymbol{h} = \boldsymbol{0}$ but $\|\boldsymbol{h}\|_2$ can be ∞)

Divide $\mathcal{T}_0^c$ into subsets of size M ($M = 3\,|\mathcal{T}_0|$).
List the entries in $\mathcal{T}_0^c$ as $n_1, \cdots, n_{N-|\mathcal{T}_0|}$ in decreasing order of their magnitudes.
Set $\mathcal{T}_j = \{n_\ell,\ (j-1)\,M + 1 \le \ell \le jM\}$.
Hence $\mathcal{T}_1$ contains the indices of the M largest entries (in magnitude) of $\boldsymbol{h}_{\mathcal{T}_0^c}$, $\mathcal{T}_2$ contains the indices of the next M largest entries (in magnitude) of $\boldsymbol{h}_{\mathcal{T}_0^c}$.



Define $\rho = |\mathcal{T}_0|\,/M$ ($\rho = 1/3$ when $M = 3\,|\mathcal{T}_0|$).

Some Observations



- The k^{th}-largest value of $\boldsymbol{h}_{\mathcal{T}_0^c}$ obeys

$$\left|\boldsymbol{h}_{\mathcal{T}_0^c}(k)\right| \leq \frac{\sum_{\ell=1}^{k}\left|\boldsymbol{h}_{\mathcal{T}_0^c}(\ell)\right|}{k} \leq \left\|\boldsymbol{h}_{\mathcal{T}_0^c}\right\|_1 / k.$$

-

$$\left|\boldsymbol{h}_{\mathcal{T}_{j+1}}(k)\right| \leq \frac{\left\|\boldsymbol{h}_{\mathcal{T}_j}\right\|_1}{M}.$$

Proof: Step 1

The ℓ_2-norm of $\boldsymbol{h}$ concentrates on $\mathcal{T}_{01} = \mathcal{T}_0 \bigcup \mathcal{T}_1$.

$$\|\boldsymbol{h}\|_2^2 = \|\boldsymbol{h}_{\mathcal{T}_{01}}\|_2^2 + \left\|\boldsymbol{h}_{\mathcal{T}_{01}^c}\right\|_2^2 \leq (1+\rho) \|\boldsymbol{h}_{\mathcal{T}_{01}}\|_2^2 .$$

Proof: From $\left|\boldsymbol{h}_{\mathcal{T}_0^c}\right|_{(k)} \leq \left\|\boldsymbol{h}_{\mathcal{T}_0^c}\right\|_1 / k$, it holds

$$\begin{aligned}
\left\|\boldsymbol{h}_{\mathcal{T}_{01}^c}\right\|_2^2 &\leq \left\|\boldsymbol{h}_{\mathcal{T}_0^c}\right\|_1^2 \sum_{k=M+1}^{N} \frac{1}{k^2} \\
&\overset{(a)}{\leq} \left\|\boldsymbol{h}_{\mathcal{T}_0^c}\right\|_1^2 / M \overset{(b)}{\leq} \frac{\|\boldsymbol{h}_{\mathcal{T}_0}\|_1^2}{M} \\
&\overset{(c)}{\leq} \frac{\|\boldsymbol{h}_{\mathcal{T}_0}\|_2^2 \cdot |\mathcal{T}_0|}{M} \leq \rho \|\boldsymbol{h}_{\mathcal{T}_{01}}\|_2^2 ,
\end{aligned}$$

where (a) holds as $\sum_{k=M+1}^{N} 1/k^2 \leq 1/M$, (b) is from the ℓ_1-cone constraint, and (c) comes from the Cauchy-Schwartz inequality.

Proof: Step 2 - A Technical Result

$$\sum_{j\geq 2}\left\|\boldsymbol{h}_{\mathcal{T}_j}\right\|_2 \leq \sqrt{\rho}\cdot\left\|\boldsymbol{h}_{\mathcal{T}_0}\right\|_2.$$

Proof: By construction $\left|\boldsymbol{h}_{\mathcal{T}_{j+1}}(k)\right| \leq \left\|\boldsymbol{h}_{\mathcal{T}_j}\right\|_1/M$. Then

$$\left\|\boldsymbol{h}_{\mathcal{T}_{j+1}}\right\|_2^2 = \sum_{k\in\mathcal{T}_{j+1}}\left|\boldsymbol{h}_{\mathcal{T}_{j+1}}(k)\right|^2 \leq M\cdot\frac{\left\|\boldsymbol{h}_{\mathcal{T}_j}\right\|_1^2}{M^2} = \frac{\left\|\boldsymbol{h}_{\mathcal{T}_j}\right\|_1^2}{M}.$$

Hence,

$$\sum_{j\geq 2}\left\|\boldsymbol{h}_{\mathcal{T}_j}\right\|_2 \leq \sum_{j\geq 2}\left\|\boldsymbol{h}_{\mathcal{T}_{j-1}}\right\|_1/\sqrt{M} \overset{(a)}{=} \sum_{j\geq 1}\left\|\boldsymbol{h}_{\mathcal{T}_j}\right\|_1/\sqrt{M} = \left\|\boldsymbol{h}_{\mathcal{T}_0^c}\right\|_1/\sqrt{M}$$
$$\overset{(b)}{\leq}\left\|\boldsymbol{h}_{\mathcal{T}_0}\right\|_1/\sqrt{M} \overset{(c)}{\leq} \sqrt{\frac{|\mathcal{T}_0|}{M}}\left\|\boldsymbol{h}_{\mathcal{T}_0}\right\|_2 = \sqrt{\rho}\left\|\boldsymbol{h}_{\mathcal{T}_0}\right\|_2,$$

where (a) uses the variable change $j' = j-1$, (b) and (c) follow from the cone constraint and the Cauchy-Schwartz inequality respectively.

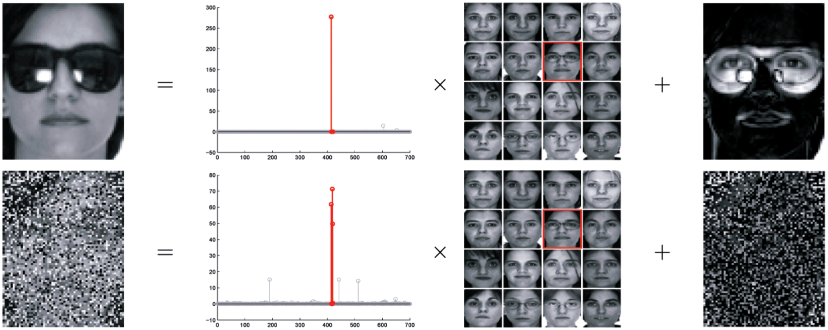
Proof: Step 3

$$
\begin{aligned}
\|\boldsymbol{A}\boldsymbol{h}\|_2 &= \left\| \boldsymbol{A}_{\mathcal{T}_{01}}\boldsymbol{h}_{\mathcal{T}_{01}} + \sum_{j\geq 2} \boldsymbol{A}_{\mathcal{T}_j}\boldsymbol{h}_{\mathcal{T}_j} \right\|_2 \geq \|\boldsymbol{A}_{\mathcal{T}_{01}}\boldsymbol{h}_{\mathcal{T}_{01}}\|_2 - \left\| \sum_{j\geq 2} \boldsymbol{A}_{\mathcal{T}_j}\boldsymbol{h}_{\mathcal{T}_j} \right\|_2 \\
&\geq \|\boldsymbol{A}_{\mathcal{T}_{01}}\boldsymbol{h}_{\mathcal{T}_{01}}\|_2 - \sum_{j\geq 2} \left\| \boldsymbol{A}_{\mathcal{T}_j}\boldsymbol{h}_{\mathcal{T}_j} \right\|_2 \\
&\geq \sqrt{1-\delta_{|\mathcal{T}_0|+M}}\, \|\boldsymbol{h}_{\mathcal{T}_{01}}\|_2 - \sqrt{1+\delta_M} \sum_{j\geq 2} \left\| \boldsymbol{h}_{\mathcal{T}_j} \right\|_2 \\
&\geq \underbrace{\left(\sqrt{1-\delta_{4S}} - \sqrt{\rho}\sqrt{1+\delta_{4S}} \right)}_{C_{4S}} \|\boldsymbol{h}_{\mathcal{T}_{01}}\|_2 .
\end{aligned}
$$

Hence,

$$
\|\boldsymbol{h}\|_2 \leq \sqrt{1+\rho}\, \|\boldsymbol{h}_{\mathcal{T}_{01}}\|_2 \leq \frac{\sqrt{1+\rho}}{C_{4S}} \|\boldsymbol{A}\boldsymbol{h}\|_2 \leq \frac{\sqrt{1+\rho}}{C_{4S}} \cdot 2\epsilon.
$$

Face Recognition with Block Occlusion [Wright et al., 2009]



The Setup

- A set of training samples $\{\boldsymbol{\phi}_i, l_i\}$
  - $\boldsymbol{\phi}_i \in \mathbb{R}^m$ is the vector representation of the images.
  - $l_i \in \{1, 2, \cdots, C\}$ label for the C subjects.
- Test sample $\boldsymbol{y}$

Assumption:

- For simplicity, assume a good face alignment.

Face Recognition via Sparse Linear Regression

Sufficiently many images of the same subject i form a low-dimensional linear subspace in $\mathbb{R}^m$.

$$\boldsymbol{y} \approx \sum_{\{j|l_j=i\}} \boldsymbol{\phi}_j c_j =: \boldsymbol{\Phi}_i \boldsymbol{c}_i.$$

Or equivalently

$$\boldsymbol{y} \approx [\boldsymbol{\Phi}_1, \boldsymbol{\Phi}_2, \cdots, \boldsymbol{\Phi}_C]\, \boldsymbol{c} = \boldsymbol{\Phi}\boldsymbol{c} \in \mathbb{R}^m$$
$$\text{where } \boldsymbol{c} = \left[\cdots, \boldsymbol{0}^T, \boldsymbol{c}_i^T, \boldsymbol{0}^T, \cdots\right]^T.$$

The ℓ_1-minimisation formulation for face recognition:

$$\min \ \|\boldsymbol{c}\|_1 \quad \text{s.t.} \ \|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{c}\|_2 \le \epsilon.$$

Robust Face Recognition

When we have corruption and occlusion $\boldsymbol{y} \not\approx \boldsymbol{\Phi x}$. Instead

$$\boldsymbol{y} \approx \boldsymbol{\Phi c} + \boldsymbol{e},$$

where $\boldsymbol{e}$ is an unknown error vector whose entries can be very large.

Assumption: only a fraction of pixels is corrupted ($\geq 70\%$ in some cases).

Robust face recognition formulation:

$$\min \; \|\boldsymbol{c}\|_1 + \|\boldsymbol{e}\|_1 \quad \text{s.t. } \boldsymbol{y} = \boldsymbol{\Phi c} + \boldsymbol{e}.$$

Or

$$\min \; \|\boldsymbol{w}\|_1 \quad \text{s.t. } \boldsymbol{y} = [\boldsymbol{\Phi}, \boldsymbol{I}]\, \boldsymbol{w}.$$

Gradient Computation

Definition 6.4 (Gradient)

$$\nabla f(\boldsymbol{x}) := \left[\frac{d}{dx_1} f, \cdots, \frac{d}{dx_n} f\right]^T.$$

Example 6.5

Let $f(\boldsymbol{x}) = \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2$. Then $\nabla f = -\boldsymbol{A}^T(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x})$.

- $$\frac{d}{d\boldsymbol{x}} \boldsymbol{a}^T \boldsymbol{x} = \frac{d}{d\boldsymbol{x}} \boldsymbol{x}^T \boldsymbol{a} = \boldsymbol{a}.$$
- $$\frac{d}{d\boldsymbol{x}} \boldsymbol{x}^T \boldsymbol{A}^T \boldsymbol{A}\boldsymbol{x} = 2\boldsymbol{A}^T \boldsymbol{A}\boldsymbol{x}.$$
- $f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T \boldsymbol{A}^T \boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}^T \boldsymbol{A}\boldsymbol{x} + \frac{1}{2}\boldsymbol{y}^T \boldsymbol{y}$,

$$\frac{d}{d\boldsymbol{x}} f = \boldsymbol{A}^T \boldsymbol{A}\boldsymbol{x} - \boldsymbol{A}^T \boldsymbol{y} = -\boldsymbol{A}^T(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}).$$

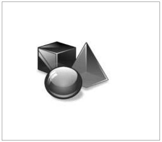
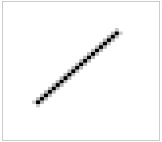
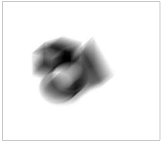
Section 7
Low Rank Matrix Recovery

Netflix Problem

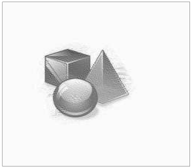
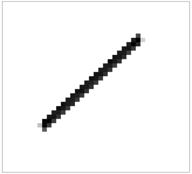
	Black Swan	Titanic	True Grit	The King's Speech
J. Cameron		★★★★★	★★★☆☆	
C. Eastwood	★★★★☆		★★★★★	
P. Jackson		★★★☆☆		★★★★☆
Roman Polanski	★★★★★			★★★★★

Blind Deconvolution [Ahmed, Recht, and Romberg, 2013]

$\boldsymbol{y} = \boldsymbol{s} \star \boldsymbol{h}: \; y[n] = \sum_{\ell=0}^{L} s[n-\ell]\, h[\ell].$

After deblurring:

Low Rank Matrices and Approximations

Consider a matrix $\boldsymbol{X}_0 \in \mathbb{R}^{m \times n}$ with its SVD

$$\boldsymbol{X}_0 = \sum_{k=1}^{\min(m,n)} \sigma_k \boldsymbol{u}_k \boldsymbol{v}_k^T,$$

where $K = \min(m, n)$ and $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_K \geq 0$.

Theorem 7.1 (The Eckart-Young Theorem)

The best low-rank approximation *of $\boldsymbol{X}_0$, i.e.,*

$$\min_{\boldsymbol{X}} \ \|\boldsymbol{X} - \boldsymbol{X}_0\|_F^2 \quad \text{s.t. rank}\,(\boldsymbol{X}) = R,$$

is given by simply truncating the SVD

$$\hat{\boldsymbol{X}} = \sum_{k=1}^{R} \sigma_k \boldsymbol{u}_k \boldsymbol{v}_k^T.$$

Remark: $\|\boldsymbol{X}\|_F^2 = \sum_{i,j} X_{i,j}^2 = \|\text{vec}\,(\boldsymbol{X})\|_2^2$.

Low Rank Matrix Recovery

Let $\mathcal{A}: \mathbb{R}^{m\times n} \to \mathbb{R}^L$ is a linear measurement operator that takes L inner products with predefined matrices $\boldsymbol{A}_1, \cdots, \boldsymbol{A}_L$:

$$\mathcal{A}: \quad \mathbb{R}^{m\times n} \to \mathbb{R}^L$$

$$\boldsymbol{X}_0 \mapsto y_l = \langle \boldsymbol{X}_0, \boldsymbol{A}_l \rangle = \operatorname{trace}\left(\boldsymbol{A}_l^T \boldsymbol{X}_0\right) = \sum_{i=1}^{m} \sum_{j=1}^{n} X_0[i,j] \, A_l[i,j].$$

The low-rank matrix recovery problem is given by

$$\min_{\boldsymbol{X}} \ \|\boldsymbol{y} - \mathcal{A}(\boldsymbol{X})\|_2^2 \quad \text{s.t. } \operatorname{rank}(\boldsymbol{X}) \le R.$$

Example 7.2

In the Netflix problem, $\boldsymbol{A}_l[i,j] = 1$ and $\boldsymbol{A}_l[s,t] = 0$ for all $[s,t] \neq [i,j]$.

Another Look at the Linear Operator $\mathcal{A}$

$$\begin{aligned} \mathcal{A}: \quad & \mathbb{R}^{m\times n} \to \mathbb{R}^{L} \\ & \boldsymbol{X} \mapsto \boldsymbol{y} = \boldsymbol{A}\text{vect}\left(\boldsymbol{X}\right), \end{aligned}$$

where $\boldsymbol{A} \in \mathbb{R}^{L\times(m\cdot n)}$.

$$\begin{bmatrix} 1 & & & & & & \\ & 1 & & & & & \\ & & 1 & & & & \\ & & & \ddots & & & \\ & & & & 1 & & \\ & & & & & 1 & \\ & & & & & & 1 \end{bmatrix}$$

Alternating Projection

To solve

$$\min_{\boldsymbol{X}} \ \|\boldsymbol{y} - \mathcal{A}(\boldsymbol{X})\|_2^2 \quad \text{s.t. rank}(\boldsymbol{X}) \leq R$$

is the same as to look for an $\boldsymbol{L} \in \mathbb{R}^{m \times R}$ and a $\boldsymbol{R} \in \mathbb{R}^{n \times R}$ s.t.

$$\min_{\boldsymbol{L},\boldsymbol{R}} \ \left\|\boldsymbol{y} - \mathcal{A}\left(\boldsymbol{L}\boldsymbol{R}^T\right)\right\|_2^2.$$

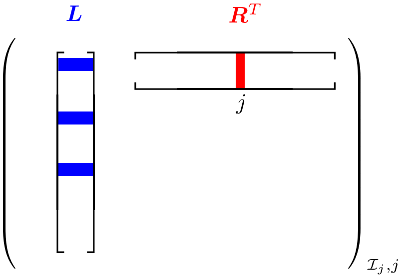
Alternating projection:

$$\boldsymbol{R}_{k+1} = \arg\min_{\boldsymbol{R}} \ \left\|\boldsymbol{y} - \mathcal{A}\left(\boldsymbol{L}_k\boldsymbol{R}^T\right)\right\|_2^2,$$
$$\boldsymbol{L}_{k+1} = \arg\min_{\boldsymbol{L}} \ \left\|\boldsymbol{y} - \mathcal{A}\left(\boldsymbol{L}\boldsymbol{R}_{k+1}^T\right)\right\|_2^2.$$

Alternating Projection (2)

Details on fixing $\boldsymbol{L}$ and updating $\boldsymbol{R}$:

$$\begin{array}{c} \begin{array}{ccc} & j & \end{array} \\ \begin{bmatrix} & 1 & \\ & ? & \\ & 3 & \\ \cdots & ? & \cdots \\ & 5 & \\ & ? & \\ & \vdots & \end{bmatrix}_{\mathcal{I}_j, j} \end{array} =$$



$$\begin{bmatrix} 1 \\ 3 \\ 5 \\ \vdots \end{bmatrix} = \boldsymbol{X}_0 \left[\mathcal{I}_j, j \right] = \boldsymbol{L}_{\mathcal{I}_j,:} \boldsymbol{R}_{j,:}^T$$

Nuclear Norm Minimization

Define the nuclear norm

$$\|\boldsymbol{X}\|_* = \sum_{k=1}^{\min(m,n)} \sigma_i,$$

which is the ℓ_1-norm of the singular value vector.

Constrained optimization problem:

$$\min_{\boldsymbol{X}} \ \|\boldsymbol{X}\|_* \quad \text{s.t. } \|\boldsymbol{y} - \mathcal{A}(\boldsymbol{X})\|_2^2 \leq \epsilon.$$

Unconstrained optimization problem:

$$\min_{\boldsymbol{X}} \ \frac{1}{2} \|\boldsymbol{y} - \mathcal{A}(\boldsymbol{X})\|_2^2 + \lambda \|\boldsymbol{X}\|_* .$$

ℓ_1-norm and Nuclear Norm

ℓ_1-norm

Write $\boldsymbol{x} = \sum_{i=1}^n x_i \boldsymbol{e}_i$ where $\boldsymbol{e}_i$ is the i^{th} natural basis vector.
$\|\boldsymbol{x}\|_1 = \sum_{i=1}^n |x_i|$.

$$\partial \|\boldsymbol{x}\|_1 = \sum_{i=1}^n \text{sign}\left(x_i\right) \boldsymbol{e}_i = \left\{\boldsymbol{v} :\ v_i = \text{sign}\left(x_i\right)\right\}.$$

Nuclear norm

$\boldsymbol{X} = \sum_{i=1}^{\min(m,n)} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^T$ and $\|\boldsymbol{X}\|_* = \sum_{i=1}^{\min(m,n)} \sigma_i$.

$$\begin{aligned} \partial \|\boldsymbol{X}\|_* &= \sum_{i=1}^{\min(m,n)} \text{sign}\left(\sigma_i\right) \boldsymbol{u}_i \boldsymbol{v}_i^T \\ &= \left\{\boldsymbol{U}_r \boldsymbol{V}_r^T + \boldsymbol{U}_{m-r} \boldsymbol{T} \boldsymbol{V}_{n-r}^T :\ \boldsymbol{T} \in \mathbb{R}^{(m-r)\times(n-r)},\ \sigma\left(\boldsymbol{T}\right) \le 1\right\}. \end{aligned}$$

Soft Thresholding Function

ℓ_1-norm minimization with given $\boldsymbol{z} \in \mathbb{R}^n$

Let $\hat{\boldsymbol{x}} = \arg\ \min_{\boldsymbol{x}}\ \frac{1}{2} \left\| \boldsymbol{x} - \boldsymbol{z} \right\|_2^2 + \lambda \left\| \boldsymbol{x} \right\|_1$. Then

$$\hat{\boldsymbol{x}} = \sum_i \eta\left(z_i; \lambda\right) \boldsymbol{e}_i \quad \text{where } \eta\left(z_i; \lambda\right) = \operatorname{sign}\left(z_i\right) \max\left(0, \left|z_i\right| - \lambda\right).$$

Nuclear norm minimization with given $\boldsymbol{Z} \in \mathbb{R}^{m \times n}$

Let $\hat{\boldsymbol{X}} = \arg\ \min_{\boldsymbol{X}}\ \frac{1}{2} \left\| \boldsymbol{X} - \boldsymbol{Z} \right\|_F^2 + \lambda \left\| \boldsymbol{X} \right\|_*$. Then

$$\hat{\boldsymbol{X}} = \sum_{i=1}^{\min(m,n)} \eta\left(\sigma_i; \lambda\right) \boldsymbol{u}_i \boldsymbol{v}_i^T \quad \text{where } \eta\left(\sigma_i; \lambda\right) = \operatorname{sign}\left(\sigma_i\right) \max\left(0, \left|\sigma_i\right| - \lambda\right).$$

ISTA

$\min \; \frac{1}{2}\left\|\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}\right\|_2^2+\lambda\left\|\boldsymbol{x}\right\|_1$

- $\frac{\partial}{\partial \boldsymbol{x}}\frac{1}{2}\left\|\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}\right\|_2^2=-\boldsymbol{A}^T\left(\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}\right)$.
- $f=\frac{1}{2}\left\|\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}\right\|_2^2 \Rightarrow \frac{1}{2t_k}\left\|\boldsymbol{x}-\left(\boldsymbol{x}^{k-1}-t_k\nabla f\right)\right\|_2^2$.
- $$\boldsymbol{x}^k=\eta\left(\boldsymbol{x}^{k-1}+t_k\boldsymbol{A}^T\left(\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}^{k-1}\right);\lambda t_k\right).$$

$\min \; \frac{1}{2}\left\|\boldsymbol{y}-\mathcal{A}\left(\boldsymbol{X}\right)\right\|_2^2+\lambda\left\|\boldsymbol{X}\right\|_*$

- $\frac{\partial}{\partial \boldsymbol{X}}\frac{1}{2}\left\|\boldsymbol{y}-\mathcal{A}\left(\boldsymbol{X}\right)\right\|_2^2=-\mathcal{A}^*\left(\boldsymbol{y}-\mathcal{A}\left(\boldsymbol{x}\right)\right)$.
- $f=\frac{1}{2}\left\|\boldsymbol{y}-\mathcal{A}\left(\boldsymbol{X}\right)\right\|_2^2 \Rightarrow \frac{1}{2t_k}\left\|\boldsymbol{X}-\left(\boldsymbol{X}^{k-1}-t_k\nabla f\right)\right\|_F^2$.
- $$\boldsymbol{X}^k=\eta_{\boldsymbol{\sigma}}\left(\boldsymbol{X}^{k-1}+t_k\mathcal{A}^*\left(\boldsymbol{y}-\mathcal{A}\left(\boldsymbol{X}^{k-1}\right)\right);\lambda t_k\right).$$

Iterative Hard Thresholding Algorithm

$$\min \; \frac{1}{2} \left\| \boldsymbol{y} - \boldsymbol{A}\boldsymbol{x} \right\|_2^2 \quad \text{s.t.} \; \left\| \boldsymbol{x} \right\|_0 \leq S$$

$$\boldsymbol{x}^k = H_S\left(\boldsymbol{x}^{k-1} + \mu_k \boldsymbol{A}^T \left(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}^{k-1}\right)\right).$$

$$\min \; \frac{1}{2} \left\| \boldsymbol{y} - \mathcal{A}\left(\boldsymbol{X}\right) \right\|_2^2 \quad \text{s.t. rank}\left(\boldsymbol{X}\right) \leq R$$

$$\boldsymbol{X}^k = H_{R,\boldsymbol{\sigma}}\left(\boldsymbol{X}^{k-1} + t_k \mathcal{A}^* \left(\boldsymbol{y} - \mathcal{A}\left(\boldsymbol{X}^{k-1}\right)\right)\right).$$

Comments on Performance Guarantees

- When $\mathcal{A}(\cdot)$ is a Gaussian random 'projection', RIP condition will hold with high probability:

$$1-\delta \leq \|\mathcal{A}(\boldsymbol{X})\|_2^2 \leq 1+\delta, \quad \forall \boldsymbol{X} \text{ s.t. } \operatorname{rank}(\boldsymbol{X}) \leq R.$$

- For matrix completion: difficult when $\boldsymbol{X}$ is low-rank and sparse.

$$\begin{bmatrix} \bullet \\ \bullet \\ \\ \\ \end{bmatrix} \begin{bmatrix} \bullet & & \bullet \end{bmatrix} = \begin{bmatrix} \bullet & & \bullet \\ \bullet & & \bullet \\ & & \\ & & \end{bmatrix}$$

  - Want coherence constant small:

$$\mu(\boldsymbol{U}) := \frac{N}{R} \max_{1\leq i\leq N} \|\mathcal{P}_{\boldsymbol{U}} \boldsymbol{e}_i\|_2^2 = O(1).$$

Blind Deconvolution: The Problem

Given a convolution of two signals

$$y[n] = \sum_{\ell=0}^{L} s[n-\ell]\, h[\ell],$$

what are $x[n]$ and $h[n]$?

This bilinear problem is difficult to solve.

- Scaling ambiguity.

Blind Deconvolution: The Idea

$$
\boldsymbol{s}\boldsymbol{h}^T = \begin{array}{c} \\ \\ y[0] \\ y[1] \\ y[2] \\ y[3] \\ y[4] \\ y[5] \\ \vdots \\ \\ \end{array} \begin{bmatrix} s[-2]\,h[0] & s[-2]\,h[1] & s[-2]\,h[2] \\ s[-1]\,h[0] & s[-1]\,h[1] & s[-1]\,h[2] \\ s[0]\,h[0] & s[0]\,h[1] & s[0]\,h[2] \\ s[1]\,h[0] & s[1]\,h[1] & s[1]\,h[2] \\ s[2]\,h[0] & s[2]\,h[1] & s[2]\,h[2] \\ s[3]\,h[0] & s[3]\,h[1] & s[3]\,h[2] \\ s[4]\,h[0] & s[4]\,h[1] & s[4]\,h[2] \\ s[5]\,h[0] & s[5]\,h[1] & s[5]\,h[2] \\ s[6]\,h[0] & s[6]\,h[1] & s[6]\,h[2] \\ \vdots & \vdots & \vdots \end{bmatrix}
$$

Each entries of $\boldsymbol{y} = \boldsymbol{x} \star \boldsymbol{h}$ is a sum along a skew diagonal of the rank-1 matrix $\boldsymbol{x}\boldsymbol{h}^T$:

$$\min \ \|\boldsymbol{X}\|_* \ \text{s.t.} \ \boldsymbol{y} = \mathcal{A}(\boldsymbol{X}).$$