Section 4
Greedy Algorithms

Greedy Algorithms: the Approach

Recall: $\|\boldsymbol{x}\|_0 =$ number of nonzero entries in $\boldsymbol{x}$.

- When we roughly know the sparsity $\|\boldsymbol{x}\|_0$,

$$\min_{\boldsymbol{x}} \ \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2 \ \text{s.t.} \ \|\boldsymbol{x}\|_0 \leq S.$$

- Otherwise if we roughly know the noise energy,

$$\min_{\boldsymbol{x}} \|\boldsymbol{x}\|_0 \ \text{s.t.} \ \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2 \leq \epsilon.$$

Major Greedy Algorithms

- Orthogonal matching pursuit (OMP)
- Subspace pursuit (SP)
- Compressive sampling matching pursuit (CoSaMP)
- Iterative hard thresholding (IHT)

Intuition: When $S = 1$

When $S = 1$: The location of the nonzero entries is given by

$$i^* = \arg \min_i \left(\min_{x_i} \|\boldsymbol{y} - \boldsymbol{a}_i x_i\|_2^2 \right)$$
$$= \arg \min_i \left\| \boldsymbol{y} - \boldsymbol{a}_i \left(\boldsymbol{a}_i^\dagger \boldsymbol{y} \right) \right\|_2^2$$

Once i^* is found,

$$x_{i^*} = \boldsymbol{a}_i^\dagger \boldsymbol{y}, \quad x_j = 0, \ \forall j \neq i^*.$$

Intuition: A Simplification

In practice, we often normalize the columns of $\boldsymbol{A}$, i.e. $\|\boldsymbol{a}_i\|_2 = 1$, such that $\boldsymbol{a}_i^\dagger = \boldsymbol{a}_i^T$.

$$\begin{aligned}
&\left\|\boldsymbol{y} - \boldsymbol{a}_i \left(\boldsymbol{a}_i^T \boldsymbol{y}\right)\right\|_2^2 \\
&= \boldsymbol{y}^T \boldsymbol{y} - 2\boldsymbol{y}^T \boldsymbol{a}_i \boldsymbol{a}_i^T \boldsymbol{y} + \boldsymbol{y}^T \boldsymbol{a}_i \boldsymbol{a}_i^T \boldsymbol{a}_i \boldsymbol{a}_i^T \boldsymbol{y} \\
&= \|\boldsymbol{y}\|_2^2 - \left|\langle \boldsymbol{a}_i, \boldsymbol{y} \rangle\right|^2 .
\end{aligned}$$

Hence

$$\begin{aligned}
i^* &= \arg\ \min_i \ \left\|\boldsymbol{y} - \boldsymbol{a}_i \left(\boldsymbol{a}_i^\dagger \boldsymbol{y}\right)\right\|_2^2 \\
&= \arg\ \max_i \ \left|\langle \boldsymbol{a}_i, \boldsymbol{y} \rangle\right| .
\end{aligned}$$

Henceforth, we assume that $\|\boldsymbol{a}_i\|_2 = 1, \ \forall i$.

Intuition: $S = 2$

Suppose that we knew $S = 2$ and the location of one nonzero entry, i.e. the support set $\mathcal{I} = \{i_1, ?\}$.

- Cancel the effect from i_1:

$$\boldsymbol{y}_r := \boldsymbol{y} - \boldsymbol{a}_{i_1} \boldsymbol{a}_{i_1}^{\dagger} \boldsymbol{y} = \boldsymbol{y} - \boldsymbol{a}_{i_1} \boldsymbol{a}_{i_1}^{T} \boldsymbol{y}.$$

- Choose i_2 via

$$i_2 = \arg \max_i \left| \langle \boldsymbol{a}_i, \boldsymbol{y}_r \rangle \right|.$$

Remark: It holds that $i_2 \neq i_1$. We get two locations indeed.
Proof: Clearly $\boldsymbol{y}_r$ is orthogonal to $\boldsymbol{a}_{i_1}$, i.e. $\langle \boldsymbol{y}_r, \boldsymbol{a}_{i_1} \rangle = 0$.

Intuition: $S = 3$

Suppose that we knew $S = 3$ and the locations of two nonzero entries, i.e. the support set $\mathcal{I} = \{i_1, i_2, ?\}$.

- Cancel the effect from i_1 and i_2: Let $\mathcal{I}_2 = \{i_1, i_2\}$.

$$\boldsymbol{y}_r := \boldsymbol{y} - \boldsymbol{A}_{\mathcal{I}_2} \boldsymbol{A}_{\mathcal{I}_2}^{\dagger} \boldsymbol{y}.$$

- Choose i_2 via

$$i_3 = \arg\max_i \left| \langle \boldsymbol{a}_i, \boldsymbol{y}_r \rangle \right|.$$

Remark: It holds that $i_3 \notin \mathcal{I}_2$. We get three locations.

The Orthogonal Matching Pursuit (OMP) Algorithm

Input: S, $\boldsymbol{A}$, $\boldsymbol{y}$.
Initialization:
$\boldsymbol{x} = \boldsymbol{0}$, $\mathcal{T}^{\ell} = \phi$, and $\boldsymbol{y}_r = \boldsymbol{y}$.
Iteration: $\ell = 1, 2, \cdots, S$

1. Let $i_\ell = \arg\max_j |\langle \boldsymbol{a}_j, \boldsymbol{y}_r \rangle|$
2. $\mathcal{T}^{\ell} = \mathcal{T}^{\ell-1} \bigcup \{i_\ell\}$. (Add one index)
3. $\boldsymbol{x}_{\mathcal{T}^\ell} = \boldsymbol{A}^{\dagger}_{\mathcal{T}^\ell} \boldsymbol{y}$. (Estimate ℓ-sparse signal)
4. $\boldsymbol{y}_r = \boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}$. (Compute estimation error)

Performance?

Suppose that

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}_0 + \boldsymbol{w},$$

where the signal $\boldsymbol{x}_0$ is S-sparse and the noise satisfies $\|\boldsymbol{w}\|_2 \leq \epsilon$.
The question is

$$\|\hat{\boldsymbol{x}} - \boldsymbol{x}_0\|_2 \leq ?.$$

- Noise free case $(\epsilon = 0)$: when $\hat{\boldsymbol{x}} = \boldsymbol{x}_0$?
- Noisy case $(\epsilon > 0)$:
  - How the recovery error $\|\hat{\boldsymbol{x}} - \boldsymbol{x}_0\|_2$ behaves with ϵ.
- Approximately sparse case:
  - Let $\boldsymbol{x}_{0,S}$ be the best S-term approximation of $\boldsymbol{x}_0$.
  - How the recovery error $\|\hat{\boldsymbol{x}} - \boldsymbol{x}_0\|_2$ behaves with
    - ϵ, and
    - $\|\boldsymbol{x}_0 - \boldsymbol{x}_{0,S}\|_2$.

Performance Guarantee of OMP: Mutual Coherence

Definition 4.1 (Mutual coherence)

The mutual coherence of a matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$, denoted by $\mu(\boldsymbol{A})$, is the maximal correlation (in magnitude) between two (normalized) columns.

$$\mu(\boldsymbol{A}) = \max_{i \neq j} \frac{|\langle \boldsymbol{a}_i, \boldsymbol{a}_j \rangle|}{\|\boldsymbol{a}_i\|_2 \|\boldsymbol{a}_j\|_2}.$$

When $\|\boldsymbol{a}_i\|_2 = 1, \ \forall i \in [n]$, then $\mu(\boldsymbol{A}) = \max_{i \neq j} |\langle \boldsymbol{a}_i, \boldsymbol{a}_j \rangle|$.

Performance Guarantee of OMP

Theorem 4.2

Suppose that $\boldsymbol{A}$ satisfies that

$$\mu < \frac{1}{2S}.$$

Then the OMP algorithm is guaranteed to exactly recover all S-sparse $\boldsymbol{x}$ from $\boldsymbol{y}$.

The key for the proof: To show $\hat{\boldsymbol{x}} = \boldsymbol{x}_0$:

- Want to show that $\operatorname{supp}(\hat{\boldsymbol{x}}) = \operatorname{supp}(\boldsymbol{x}_0)$.
- Or show that at the ℓ-th iteration of OMP, the chosen index $i_\ell \in \mathcal{T}_0 := \operatorname{supp}(\boldsymbol{x}_0)$.

The proof needs Cauchy–Schwartz Inequality in Theorem 4.9 in Appendix.

The First Iteration of OMP (1)

Want to show that $i_1 := \arg\max_i |\langle \boldsymbol{a}_i, \boldsymbol{y} \rangle| \in \mathcal{T}_0$.

- $\forall i$, $|\langle \boldsymbol{a}_i, \boldsymbol{y} \rangle| = \left| \left\langle \boldsymbol{a}_i, \sum_{j \in \mathcal{T}_0} \boldsymbol{a}_j x_{0,j} \right\rangle \right| = \left| \sum_{j \in \mathcal{T}_0} x_{0,j} \langle \boldsymbol{a}_i, \boldsymbol{a}_j \rangle \right|$.
- For all $i \notin \mathcal{T}_0$:

$$|\langle \boldsymbol{a}_i, \boldsymbol{y} \rangle| = \left| \sum_{j \in \mathcal{T}_0} x_{0,j} \langle \boldsymbol{a}_i, \boldsymbol{a}_j \rangle \right| \le \sum_{j \in \mathcal{T}_0} |x_{0,j}| \, |\langle \boldsymbol{a}_i, \boldsymbol{a}_j \rangle|$$
$$\le \mu \sum_{j \in \mathcal{T}_0} |x_{0,j}| \overset{(a)}{\le} \mu \sqrt{S} \, \|\boldsymbol{x}\|_2$$

  where (a) follows from Cauchy–Schwartz Inequality (Theorem 4.9).
- Hence,

$$\max_{i \notin \mathcal{T}_0} |\langle \boldsymbol{a}_i, \boldsymbol{y} \rangle| \le \mu \sqrt{S} \, \|\boldsymbol{x}\|_2 . \quad (6)$$

The First Iteration of OMP (2)

- For all $i \in \mathcal{T}_0$:

$$\left|\langle \boldsymbol{a}_i, \boldsymbol{y}\rangle\right| = \left|\sum_{j\in\mathcal{T}_0} x_{0,j} \langle \boldsymbol{a}_i, \boldsymbol{a}_j\rangle\right| \geq \left|x_{0,i} \langle \boldsymbol{a}_i, \boldsymbol{a}_i\rangle\right| - \left|\sum_{j\neq i} x_{0,j} \langle \boldsymbol{a}_i, \boldsymbol{a}_j\rangle\right|$$

$$\geq |x_{0,i}| - \mu \sum_{j\neq i} |x_{0,j}| \overset{(a)}{\geq} |x_{0,i}| - \mu\sqrt{S}\,\|\boldsymbol{x}\|_2\,,$$

where (a) follows from Cauchy–Schwartz Inequality.

- $$\max_{i\in\mathcal{T}_0} \left|\langle \boldsymbol{a}_i, \boldsymbol{y}\rangle\right| \geq \frac{1}{\sqrt{S}}\|\boldsymbol{x}\|_2 - \mu\sqrt{S}\,\|\boldsymbol{x}\|_2\,,$$

where we have used the fact that

$$\frac{1}{\sqrt{S}}\|\boldsymbol{x}\|_2 = \frac{\left(\sum x_i^2\right)^{\frac{1}{2}}}{\sqrt{S}} \leq \frac{\left(\sum \left(\max_i \ |x_i|\right)^2\right)^{\frac{1}{2}}}{\sqrt{S}} = \max_{i\in\mathcal{T}_0} |x_i|\,. \tag{7}$$

The First Iteration of OMP (3)

- Now suppose that $\mu < \frac{1}{2S}$ (the assumption in Theorem 4.2). Then

$$\frac{1}{\sqrt{S}} \left\| \boldsymbol{x} \right\|_2 > 2\mu\sqrt{S} \left\| \boldsymbol{x} \right\|_2 ,$$

- Or equivalently,

$$\max_{i \in \mathcal{T}_0} \left| \langle \boldsymbol{a}_i, \boldsymbol{y} \rangle \right| \geq \frac{1}{\sqrt{S}} \left\| \boldsymbol{x} \right\|_2 - \mu\sqrt{S} \left\| \boldsymbol{x} \right\|_2 > \mu\sqrt{S} \left\| \boldsymbol{x} \right\|_2 \geq \max_{i \notin \mathcal{T}_0} \left| \langle \boldsymbol{a}_i, \boldsymbol{y} \rangle \right|.$$

- One conclues that

$$i_1 \in \mathcal{T}_0.$$

The ℓ^{th} Iteration: Mathematical Induction

- Let $i_1, \cdots, i_{\ell-1}$ be the indices chosen in the first $\ell - 1$ iterations. Let $\mathcal{T}^{\ell-1} = \{i_1, \cdots, i_{\ell-1}\}$. Assume that $\mathcal{T}^{\ell-1} \subset \mathcal{T}_0$.
- Then

$$\boldsymbol{y}_r = \boldsymbol{y} - \boldsymbol{A}_{\mathcal{T}^{\ell-1}} \boldsymbol{A}^{\dagger}_{\mathcal{T}^{\ell-1}} \boldsymbol{y} = \boldsymbol{y} - \boldsymbol{A}_{\mathcal{T}^{\ell-1}} \tilde{\boldsymbol{y}}_{\ell-1} \in \operatorname{span}\left(\boldsymbol{A}_{\mathcal{T}_0}\right).$$

  Or

$$\boldsymbol{y}_r = \boldsymbol{A}_{\mathcal{T}_0} \tilde{\boldsymbol{v}}_{\mathcal{T}_0}.$$

  for some $\tilde{\boldsymbol{v}}_{\mathcal{T}_0}$.
- Use the same arguments as before, $i_\ell \in \mathcal{T}_0$.
  At the same time, $\boldsymbol{A}^T_{\mathcal{T}^{\ell-1}} \boldsymbol{y}_r = \boldsymbol{0}$ and hence $i_\ell \notin \mathcal{T}^{\ell-1}$.
  $\left|\mathcal{T}^{\ell}\right| = \ell$.
- OMP algorithm needs S iterations to recover S-sparse signals.

Hard Thresholding Function

Hard thresholding function $H_S(\boldsymbol{a})$:

Set all but the largest (in magnitude) S elements of $\boldsymbol{a}$ to zero.

Example:

$$\boldsymbol{a} = [3, -4, 1] \Rightarrow H_1(\boldsymbol{a}) = [0, -4, 0] \ \& \ H_2(\boldsymbol{a}) = [3, -4, 0].$$

$\mathrm{supp}(\boldsymbol{a})$: Index set of nonzero entries in $\boldsymbol{a}$.

$\mathrm{supp}(H_1(\boldsymbol{a})) = \arg\max_i |a_i|$.

$\mathrm{supp}(H_S(\boldsymbol{a})) = \{S \text{ indices of the largest magnitude entries in } \boldsymbol{a}\}$.

In the following greedy algorithms:

$\mathrm{supp}\left(H_1\left(\boldsymbol{A}^T\boldsymbol{y}\right)\right) = \arg\max_j |\langle \boldsymbol{y}, \boldsymbol{a}_j \rangle|$.

$\mathrm{supp}\left(H_S\left(\boldsymbol{A}^T\boldsymbol{y}\right)\right) = \{S \text{ indices corr. to the } S \text{ largest } |\langle \boldsymbol{y}, \boldsymbol{a}_j \rangle|\}$.

The Subspace Pursuit (SP) Algorithm

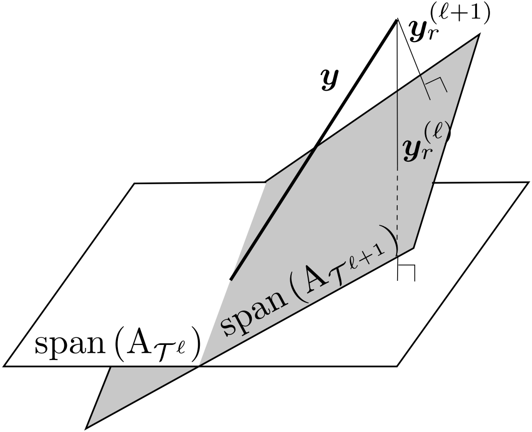
Input: S, $\boldsymbol{A}$, $\boldsymbol{y}$.
Initialization:

1. $\mathcal{T}^0 = \text{supp}\left(H_S\left(\boldsymbol{A}^T\boldsymbol{y}\right)\right)$.
2. $\boldsymbol{y}_r = \text{resid}\left(\boldsymbol{y}, \boldsymbol{A}_{\mathcal{T}^0}\right)$.

Iteration: $\ell = 1, 2, \cdots$ until exit criteria are true.

1. $\tilde{\mathcal{T}}^\ell = \mathcal{T}^{\ell-1} \bigcup \text{supp}\left(H_S\left(\boldsymbol{A}^T\boldsymbol{y}_r\right)\right)$. (Expand support)
2. Let $\boldsymbol{b}_{\tilde{\mathcal{T}}^\ell} = \boldsymbol{A}^\dagger_{\tilde{\mathcal{T}}^\ell}\boldsymbol{y}$ and $\boldsymbol{b}_{\left(\tilde{\mathcal{T}}^\ell\right)^c} = \boldsymbol{0}$. (Estimate $2S$-sparse signal)
3. Set $\mathcal{T}^\ell = \text{supp}\left(H_S\left(\boldsymbol{b}\right)\right)$. (Shrink support)
4. Let $\boldsymbol{x}^\ell_{\mathcal{T}^\ell} = \boldsymbol{A}^\dagger_{\mathcal{T}^\ell}\boldsymbol{y}$ and $\boldsymbol{x}^\ell_{\left(\mathcal{T}^\ell\right)^c} = \boldsymbol{0}$. (Estimate S-sparse signal)
5. Let $\boldsymbol{y}_r = \boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}^\ell$. (Compute estimation error)

Geometric Interpretation



The Compressive Sampling Matching Pursuit (CoSaMP) Algorithm

Input: S, $\boldsymbol{A}$, $\boldsymbol{y}$.
Initialization:
$\boldsymbol{x}^0 = \boldsymbol{0}$, and $\boldsymbol{y}_r = \boldsymbol{y}$.
Iteration: $\ell = 1, 2, \cdots$ until exit criterion true.

1. $\tilde{\mathcal{T}}^\ell = \mathcal{T}^{\ell-1} \bigcup \operatorname{supp}\left(H_{2S}\left(\boldsymbol{A}^T \boldsymbol{y}_r\right)\right)$. (Expand support)
2. Let $\boldsymbol{b}_{\tilde{\mathcal{T}}^\ell} = \boldsymbol{A}^\dagger_{\tilde{\mathcal{T}}^\ell} \boldsymbol{y}$ and $\boldsymbol{b}_{\left(\tilde{\mathcal{T}}^\ell\right)^c} = \boldsymbol{0}$. (Estimate $3S$-sparse signal)
3. $\boldsymbol{x}^\ell = H_S(\boldsymbol{b})$. ($\mathcal{T}^\ell = \operatorname{supp}(H_S(\boldsymbol{b}))$.) (Shrink support)
4. $\boldsymbol{y}_r = \boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}^\ell$. (Update estimation error)

The Iterative Hard Thresholding (IHT) Algorithm

Input: S, $\boldsymbol{A}$, $\boldsymbol{y}$.
Initialization:
$\boldsymbol{x}^0 = \boldsymbol{0}$.
Iteration: $\ell = 1, 2, \cdots$ until exit criterion true.

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

A more general form: for some $\mu > 0$.

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

Comments

History

- MP: Friedman and Stuetzle, 1981; Mallat and Zhang, 1993; Qian and Chen, 1994.
- OMP: Chen, et al., 1989; Pati, et al., 1993; Davis, et al., 1994. Analysed by Tropp, 2004.
- SP: Dai and Milenkovic, 2009. (Online available 06/03/2008)
  CoSaMP: Needell and Tropp, 2009. (Online available 17/03/2008)
  IHT: Blumensath and Davies, 2009. (Online available 05/05/2008)

Comparison:

	# of measurements	# of iterations
Exhaustive Search	$2S+1$	$\binom{n}{S} = O\left(n^S\right)$
OMP	$O\left(S^2 \log n\right)$	S
SP, CoSaMP, IHT	$O\left(S \cdot \log \frac{n}{S}\right)$	Typically $O\left(\log S\right)$, at most S

# of measurements is based on random Gaussian matrices.

Restricted Isometry Property (RIP)

SOS

Definition 4.3 (Restricted isometry property (RIP) and restricted isometry constant (RIC))

A matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ is said to satisfy the RIP with parameters (K, δ), if for all $\mathcal{T} \subset [n]$ such that $|\mathcal{T}| \leq K$ and for all $\boldsymbol{q} \in \mathbb{R}^{|\mathcal{T}|}$, it holds that

$[n] = \{1, 2, \ldots, n\}$,

$$(1-\delta)\|\boldsymbol{q}\|_2^2 \leq \|\boldsymbol{A}_{\mathcal{T}}\boldsymbol{q}\|_2^2 \leq (1+\delta)\|\boldsymbol{q}\|_2^2 .$$

The RIC δ_K is defined as the smallest constant δ for which the K-RIP holds, i.e.,

inferior

$$\delta_K = \inf\left\{\delta : \ (1-\delta)\|\boldsymbol{q}\|_2^2 \leq \|\boldsymbol{A}_{\mathcal{T}}\boldsymbol{q}\|_2^2 \leq (1+\delta)\|\boldsymbol{q}\|_2^2 . \right.$$
$$\left. \forall |\mathcal{T}| \leq K, \ \forall \boldsymbol{q} \in \mathbb{R}^{|\mathcal{T}|}\right\} .$$

RIP, Eigenvalues and Singular Values

Let $\boldsymbol{B} \in \mathbb{R}^{m \times K}$ be a tall matrix, i.e. $m \geq K$. Then the following statements are equivalent.

- For all $\boldsymbol{q} \in \mathbb{R}^K$,
$$(1-\delta)\left\|\boldsymbol{q}\right\|_2^2 \leq \left\|\boldsymbol{B}\boldsymbol{q}\right\|_2^2 \leq (1+\delta)\left\|\boldsymbol{q}\right\|_2^2.$$
- $$1-\delta_K \leq \lambda_{\min}\left(\boldsymbol{B}^T\boldsymbol{B}\right) \leq \lambda_{\max}\left(\boldsymbol{B}^T\boldsymbol{B}\right) \leq 1+\delta_K.$$
- $$\sqrt{1-\delta_K} \leq \sigma_{\min}\left(\boldsymbol{B}\right) \leq \sigma_{\max}\left(\boldsymbol{B}\right) \leq \sqrt{1+\delta_K}.$$

RIP, Eigenvalues and Singular Values: Proof

- Let $\boldsymbol{B} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$ be the compact SVD.
- $$\begin{aligned} \|\boldsymbol{B}\boldsymbol{q}\|_2^2 = \left\|\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T\boldsymbol{q}\right\|_2^2 &= \boldsymbol{q}^T\boldsymbol{V}\boldsymbol{\Sigma}\boldsymbol{U}^T\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T\boldsymbol{q} \\ &= \boldsymbol{q}^T\boldsymbol{V}\boldsymbol{\Sigma}^2\boldsymbol{V}^T\boldsymbol{q} \\ &= \sum_{i=1}^{K} \sigma_i^2 c_i^2, \end{aligned}$$

  where $c_i := \boldsymbol{v}_i^T\boldsymbol{q}$.
  - $\sum_{i=1}^K c_i^2 = \|\boldsymbol{q}\|_2^2$. This follows from $\left\|\boldsymbol{V}^T\boldsymbol{q}\right\|_2^2 = \|\boldsymbol{q}\|_2^2$.
- $$\sum_{i=1}^{K} \sigma_i^2 c_i^2 \le \sigma_{\max}^2 \sum_{i=1}^{K} c_i^2 = \sigma_{\max}^2 \|\boldsymbol{q}\|_2^2,$$
  $$\sum_{i=1}^{K} \sigma_i^2 c_i^2 \ge \sigma_{\min}^2 \sum_{i=1}^{K} c_i^2 = \sigma_{\min}^2 \|\boldsymbol{q}\|_2^2.$$

Monotonicity of RIC

Theorem 4.4

$\delta_1 \le \delta_2 \le \delta_3 \le \cdots$ *($\delta_K \le \delta_{K'}$ for all $K \le K'$).*

Proof: Let $\mathcal{Q}_K = \{\boldsymbol{q} \in \mathbb{R}^n : \|\boldsymbol{q}\|_0 \le K, \|\boldsymbol{q}\|_2 \le 1\}$. It is clear that $\mathcal{Q}_K \subset \mathcal{Q}_{K'}$ if $K \le K'$.
Then it holds that

$$\delta_K := \sup_{\boldsymbol{q} \in \mathcal{Q}_K} \left(\|\boldsymbol{A}\boldsymbol{q}\|_2^2 - 1 \right) \le \sup_{\boldsymbol{q} \in \mathcal{Q}_{K'}} \left(\|\boldsymbol{A}\boldsymbol{q}\|_2^2 - 1 \right) =: \delta_{K'}.$$

Near Orthogonality of the Columns

Theorem 4.5

Let $\mathcal{I}, \mathcal{J} \subset [n]$ *be two disjoint sets, i.e.,* $\mathcal{I} \bigcap \mathcal{J} = \phi$*. For all* $\boldsymbol{a} \in \mathbb{R}^{|\mathcal{I}|}$ *and* $\boldsymbol{b} \in \mathbb{R}^{|\mathcal{J}|}$,

$$\left|\langle \boldsymbol{A}_{\mathcal{I}}\boldsymbol{a}, \boldsymbol{A}_{\mathcal{J}}\boldsymbol{b}\rangle\right| \leq \delta_{|\mathcal{I}|+|\mathcal{J}|} \left\|\boldsymbol{a}\right\|_2 \left\|\boldsymbol{b}\right\|_2, \tag{8}$$

and

$$\left\|\boldsymbol{A}_{\mathcal{I}}^T \boldsymbol{A}_{\mathcal{J}}\boldsymbol{b}\right\|_2 \leq \delta_{|\mathcal{I}|+|\mathcal{J}|} \left\|\boldsymbol{b}\right\|_2. \tag{9}$$

Proof: From (8) to (9):

$$\begin{aligned} \left\|\boldsymbol{A}_{\mathcal{I}}^* \boldsymbol{A}_{\mathcal{J}}\boldsymbol{b}\right\|_2 &= \max_{\boldsymbol{q}:\ \|\boldsymbol{q}\|_2=1} \left|\left\langle \boldsymbol{q}, \boldsymbol{A}_{\mathcal{I}}^T \boldsymbol{A}_{\mathcal{J}}\boldsymbol{b}\right\rangle\right| = \max_{\boldsymbol{q}:\ \|\boldsymbol{q}\|_2=1} \left|\boldsymbol{q}^T \boldsymbol{A}_{\mathcal{I}}^T \boldsymbol{A}_{\mathcal{J}}\boldsymbol{b}\right| \\ &\leq \max_{\boldsymbol{q}:\ \|\boldsymbol{q}\|_2=1} \delta_{|\mathcal{I}|+|\mathcal{J}|} \left\|\boldsymbol{q}\right\|_2 \left\|\boldsymbol{b}\right\|_2 \\ &= \delta_{|\mathcal{I}|+|\mathcal{J}|} \left\|\boldsymbol{b}\right\|_2 \end{aligned}$$

Proof of (8)

(8) obviously holds when either $\boldsymbol{a}$ or $\boldsymbol{b}$ is zero. Assume $\boldsymbol{a} \neq \mathbf{0}$ and $\boldsymbol{b} \neq \mathbf{0}$. Define

$$\begin{array}{ll} \boldsymbol{a}' = \boldsymbol{a} / \left\| \boldsymbol{a} \right\|_2, & \boldsymbol{b}' = \boldsymbol{b} / \left\| \boldsymbol{b} \right\|_2, \\ \boldsymbol{x}' = \boldsymbol{A}_{\mathcal{I}} \boldsymbol{a}', & \boldsymbol{y}' = \boldsymbol{A}_{\mathcal{J}} \boldsymbol{b}'. \end{array}$$

Then RIP implies that

$$2\left(1 - \delta_{|\mathcal{I}|+|\mathcal{J}|}\right) \leq \left\| \boldsymbol{x}' + \boldsymbol{y}' \right\|_2^2 = \left\| \left[\boldsymbol{A}_{\mathcal{I}} \boldsymbol{A}_{\mathcal{J}} \right] \left[\begin{array}{c} \boldsymbol{a}' \\ \boldsymbol{b}' \end{array} \right] \right\|_2^2 \leq 2\left(1 + \delta_{|\mathcal{I}|+|\mathcal{J}|}\right),$$

$$2\left(1 - \delta_{|\mathcal{I}|+|\mathcal{J}|}\right) \leq \left\| \boldsymbol{x}' - \boldsymbol{y}' \right\|_2^2 = \left\| \left[\boldsymbol{A}_{\mathcal{I}} \boldsymbol{A}_{\mathcal{J}} \right] \left[\begin{array}{c} \boldsymbol{a}' \\ -\boldsymbol{b}' \end{array} \right] \right\|_2^2 \leq 2\left(1 + \delta_{|\mathcal{I}|+|\mathcal{J}|}\right).$$

Thus

$$\begin{aligned} \langle \boldsymbol{x}', \boldsymbol{y}' \rangle &= \frac{\|\boldsymbol{x}'+\boldsymbol{y}'\|_2^2 - \|\boldsymbol{x}'-\boldsymbol{y}'\|_2^2}{4} \leq \delta_{|\mathcal{I}|+|\mathcal{J}|} \\ -\langle \boldsymbol{x}', \boldsymbol{y}' \rangle &= \frac{\|\boldsymbol{x}'-\boldsymbol{y}'\|_2^2 - \|\boldsymbol{x}'+\boldsymbol{y}'\|_2^2}{4} \leq \delta_{|\mathcal{I}|+|\mathcal{J}|} \end{aligned}$$

Therefore,

$$\frac{|\langle \boldsymbol{A}_{\mathcal{I}} \boldsymbol{a}, \boldsymbol{A}_{\mathcal{J}} \boldsymbol{b} \rangle|}{\|\boldsymbol{a}\|_2 \|\boldsymbol{b}\|_2} = |\langle \boldsymbol{x}', \boldsymbol{y}' \rangle| \leq \delta_{|\mathcal{I}|+|\mathcal{J}|}.$$

Why RIP

In OMP, we need near-orthogonality between columns.

- $|\langle \boldsymbol{a}_i, \boldsymbol{a}_j \rangle|$ is small.

In other greedy algorithms, we need near-orthogonality between submatrices.

- $\left\| \boldsymbol{A}_{\mathcal{I}}^T \boldsymbol{A}_{\mathcal{J}} \boldsymbol{b} \right\|_2 \leq \delta_{|\mathcal{I}|+|\mathcal{J}|} \left\| \boldsymbol{b} \right\|_2$ means $\sigma_{\max}\left(\boldsymbol{A}_{\mathcal{I}}^T \boldsymbol{A}_{\mathcal{J}} \right)$ is small.

Example: near-orthogonality of columns does not mean near-orthogonality of submatrices.

Suppose that $\boldsymbol{A}_{\mathcal{I}}^T \boldsymbol{A}_{\mathcal{J}} = \begin{bmatrix} \frac{1}{\ell} & \frac{1}{\ell} & \cdots & \frac{1}{\ell} \\ \frac{1}{\ell} & \frac{1}{\ell} & \cdots & \frac{1}{\ell} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{\ell} & \frac{1}{\ell} & \cdots & \frac{1}{\ell} \end{bmatrix} \in \mathbb{R}^{\ell \times \ell}$.

Then $\sigma\left(\boldsymbol{A}_{\mathcal{I}}^T \boldsymbol{A}_{\mathcal{J}} \right) = 1, 0, \cdots, 0$.

IHT Performance: A Sufficient Condition

Theorem 4.6

Suppose that $\boldsymbol{A}$ satisfies the RIP with $\delta_{3S} < 1/\sqrt{32}$, then the k^{th} iteration of IHT obeys

$$\left\| \boldsymbol{x}_0 - \boldsymbol{x}^k \right\|_2 \leq 2^{-k} \left\| \boldsymbol{x}_0 \right\|_2 + 5 \left\| \boldsymbol{w} \right\|_2 .$$

Consequence: IHT estimates $\boldsymbol{x}$ with accuracy

$$\left\| \boldsymbol{x}_0 - \boldsymbol{x}^k \right\|_2 \leq 6 \left\| \boldsymbol{w} \right\|_2 , \quad \text{if } k > k^* = \left\lceil \log_2 \left(\frac{\left\| \boldsymbol{x}_0 \right\|_2}{\left\| \boldsymbol{w} \right\|_2} \right) \right\rceil .$$

Optimality

Claim: No recovery method can perform fundamentally better.

Suppose that an oracle tells us the support $\mathcal{T}_0$ of $\boldsymbol{x}_0$. Then

$$\hat{\boldsymbol{x}} = \begin{cases} \left(\boldsymbol{A}_{\mathcal{T}_0}^T \boldsymbol{A}_{\mathcal{T}_0}\right)^{-1} \boldsymbol{A}_{\mathcal{T}_0}^T \boldsymbol{y} & \text{on } \mathcal{T}_0, \\ \mathbf{0} & \text{elsewhere.} \end{cases}$$

Thus, $\hat{\boldsymbol{x}} - \boldsymbol{x}_0 = \mathbf{0}$ on $\mathcal{T}_0^c$, while on $\mathcal{T}_0$

$$\hat{\boldsymbol{x}} - \boldsymbol{x}_0 = \left(\boldsymbol{A}_{\mathcal{T}_0}^T \boldsymbol{A}_{\mathcal{T}_0}\right)^{-1} \boldsymbol{A}_{\mathcal{T}_0}^T \boldsymbol{w}.$$

By the RIP property,

$$\frac{1}{\sqrt{1+\delta_S}} \left\|\boldsymbol{w}\right\|_2 \le \left\|\hat{\boldsymbol{x}} - \boldsymbol{x}_0\right\|_2 \le \frac{1}{\sqrt{1-\delta_S}} \left\|\boldsymbol{w}\right\|_2 .$$

Proof Idea

Let $\boldsymbol{r}^k := \boldsymbol{x}_0 - \boldsymbol{x}^k$ ($\boldsymbol{r}^0 = \boldsymbol{x}_0$). The key is to show that

$$\left\|\boldsymbol{r}^{k+1}\right\|_2 \le \sqrt{8}\delta_{3S}\left\|\boldsymbol{r}^k\right\|_2 + 2\sqrt{1+\delta_S}\left\|\boldsymbol{w}\right\|_2.$$

In particular, if $\delta_{3S} < 1/\sqrt{32}$,

$$\left\|\boldsymbol{r}^{k+1}\right\|_2 \le 0.5\left\|\boldsymbol{r}^k\right\|_2 + 2.17\left\|\boldsymbol{w}\right\|_2.$$

Back to the main result:

$$\begin{aligned}\left\|\boldsymbol{r}^k\right\|_2 &\le \frac{1}{2}\left\|\boldsymbol{r}^{k-1}\right\|_2 + 2.17\left\|\boldsymbol{w}\right\|_2 \\ &\le \frac{1}{4}\left\|\boldsymbol{r}^{k-2}\right\|_2 + 2.17\left(1+\frac{1}{2}\right)\left\|\boldsymbol{w}\right\|_2 \\ &\cdots < \frac{1}{2^k}\left\|\boldsymbol{r}^0\right\|_2 + 4.34\left\|\boldsymbol{w}\right\|_2.\end{aligned}$$

Detailed Proof

Recall that

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

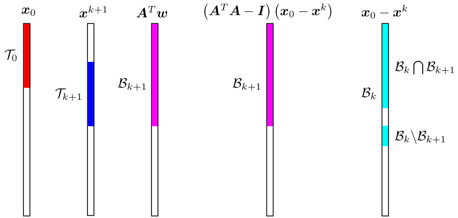
Define

$$\begin{aligned}\boldsymbol{a}^{k+1} &:= \boldsymbol{x}^k + \boldsymbol{A}^T\left(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}^k\right) \\ &= \boldsymbol{x}_0 - \boldsymbol{x}_0 + \boldsymbol{x}^k + \boldsymbol{A}^T\left(\boldsymbol{A}\boldsymbol{x}_0 + \boldsymbol{w} - \boldsymbol{A}\boldsymbol{x}^k\right) \\ &= \boldsymbol{x}_0 + \left(\boldsymbol{A}^T\boldsymbol{A} - \boldsymbol{I}\right)\left(\boldsymbol{x}_0 - \boldsymbol{x}^k\right) + \boldsymbol{A}^T\boldsymbol{w} \\ &= \boldsymbol{x}_0 + \left(\boldsymbol{A}^T\boldsymbol{A} - \boldsymbol{I}\right)\boldsymbol{r}^k + \boldsymbol{A}^T\boldsymbol{w}. \qquad (10)\end{aligned}$$

Then

$$\boldsymbol{x}^{k+1} = H_S\left(\boldsymbol{x}_0 + \left(\boldsymbol{A}^T\boldsymbol{A} - \boldsymbol{I}\right)\boldsymbol{r}^k + \boldsymbol{A}^T\boldsymbol{w}\right).$$

Detailed Proof (Continued)



$$\boldsymbol{x}^{k+1} = H_S\left(\boldsymbol{x}_0 + \left(\boldsymbol{A}^T\boldsymbol{A} - \boldsymbol{I}\right)\boldsymbol{r}^k + \boldsymbol{A}^T\boldsymbol{w}\right).$$

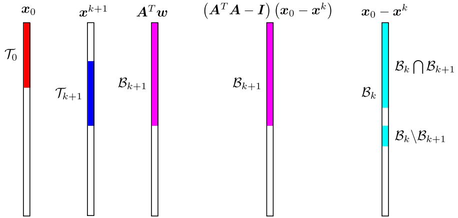
Let $\mathcal{T}_0 = \operatorname{supp}(\boldsymbol{x}_0)$, $\mathcal{T}^k = \operatorname{supp}\left(\boldsymbol{x}^k\right)$, and $\mathcal{B}^k = \mathcal{T}_0 \bigcup \mathcal{T}^k$. → at most 2S non zeros

- $\boldsymbol{r}^{k+1} = \boldsymbol{x}_0 - \boldsymbol{x}^{k+1}$ is supported on $\mathcal{B}^{k+1}$
- $\boldsymbol{r}^k = \boldsymbol{x}_0 - \boldsymbol{x}^k$ is supported on $\mathcal{B}^k$.

Want to show that $\left\|\boldsymbol{r}^{k+1}\right\|_2$ is small.

- Both $\left(\boldsymbol{A}^T\boldsymbol{A} - \boldsymbol{I}\right)\boldsymbol{r}^k$ and $\boldsymbol{A}^T\boldsymbol{w}$ are small.

Detailed Proof (Continued)



Focus on the set $\mathcal{B}^{k+1}$:

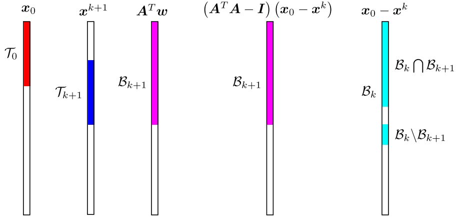
$$\begin{aligned}\left\|\boldsymbol{r}^{k+1}\right\|_2 &= \left\|\boldsymbol{x}_{0,\mathcal{B}^{k+1}} - \boldsymbol{x}^{k+1}_{\mathcal{B}^{k+1}}\right\|_2 \\ &= \left\|\boldsymbol{x}_{0,\mathcal{B}^{k+1}} - \boldsymbol{a}^{k+1}_{\mathcal{B}^{k+1}} + \boldsymbol{a}^{k+1}_{\mathcal{B}^{k+1}} - \boldsymbol{x}^{k+1}_{\mathcal{B}^{k+1}}\right\|_2 \\ &\overset{(a)}{\leq} \left\|\boldsymbol{x}_{0,\mathcal{B}^{k+1}} - \boldsymbol{a}^{k+1}_{\mathcal{B}^{k+1}}\right\|_2 + \left\|\boldsymbol{a}^{k+1}_{\mathcal{B}^{k+1}} - \boldsymbol{x}^{k+1}_{\mathcal{B}^{k+1}}\right\|_2 \\ &\overset{(b)}{\leq} 2\left\|\boldsymbol{x}_{0,\mathcal{B}^{k+1}} - \boldsymbol{a}^{k+1}_{\mathcal{B}^{k+1}}\right\|_2, \end{aligned} \tag{11}$$

where

(a) has used triangle inequality, and

(b) follows from that $\boldsymbol{x}^{k+1}_{\mathcal{B}^{k+1}}$ is the best s-term approximation to $\boldsymbol{a}^{k+1}_{\mathcal{B}^{k+1}}$.

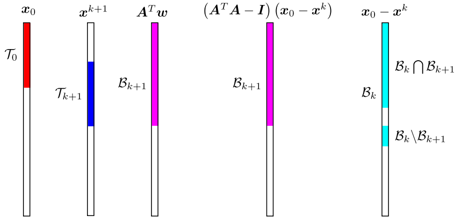
Detailed Proof (Continued)



The noise term: $\boldsymbol{A}^T\boldsymbol{w}$.

$$\left\|\left(\boldsymbol{A}^T\boldsymbol{w}\right)_{\mathcal{B}^{k+1}}\right\|_2 = \left\|\boldsymbol{A}_{\mathcal{B}^{k+1}}^T\boldsymbol{w}\right\|_2 \le \sqrt{1+\delta_{2S}}\,\|\boldsymbol{w}\|_2 \,.$$

Detailed Proof (Continued)



$$\left(\left(\boldsymbol{I}-\boldsymbol{A}^T\boldsymbol{A}\right)\boldsymbol{r}^k\right)_{\mathcal{B}^{k+1}}=\boldsymbol{r}^k_{\mathcal{B}^{k+1}}-\boldsymbol{A}^T_{\mathcal{B}^{k+1}}\boldsymbol{A}\boldsymbol{r}^k$$
$$=\boldsymbol{r}^k_{\mathcal{B}^{k+1}}-\boldsymbol{A}^T_{\mathcal{B}^{k+1}}\boldsymbol{A}_{\mathcal{B}^{k+1}}\cdot\boldsymbol{r}^k_{\mathcal{B}^{k+1}}-\boldsymbol{A}^T_{\mathcal{B}^{k+1}}\boldsymbol{A}_{\mathcal{B}^k\backslash\mathcal{B}^{k+1}}\cdot\boldsymbol{r}^k_{\mathcal{B}^k\backslash\mathcal{B}^{k+1}}$$
$$=\left(\boldsymbol{I}-\boldsymbol{A}^T_{\mathcal{B}^{k+1}}\boldsymbol{A}_{\mathcal{B}^{k+1}}\right)\boldsymbol{r}^k_{\mathcal{B}^{k+1}}-\boldsymbol{A}^T_{\mathcal{B}^{k+1}}\boldsymbol{A}_{\mathcal{B}^k\backslash\mathcal{B}^{k+1}}\cdot\boldsymbol{r}^k_{\mathcal{B}^k\backslash\mathcal{B}^{k+1}}.$$

Hence

$$\left\|\cdots\right\|_2\leq\delta_{2S}\left\|\boldsymbol{r}^k_{\mathcal{B}^{k+1}}\right\|_2+\delta_{3S}\left\|\boldsymbol{r}^k_{\mathcal{B}^k\backslash\mathcal{B}^{k+1}}\right\|_2\leq\sqrt{2}\delta_{3S}\left\|\boldsymbol{r}^k\right\|_2,$$

Detailed Proof (Continued)

where

- The 1st term follows from $\left|\mathcal{B}^{k+1}\right| \leq 2S$ and RIP.
- The 2nd term follows from Theorem 4.5.
- The last term uses $\delta_{2S} \leq \delta_{3S}$ (Theorem 4.4) and Cauchy-Schwartz Inequality

$$\begin{aligned}
&\left\|\boldsymbol{r}^k_{\mathcal{B}^{k+1}}\right\|_2 + \left\|\boldsymbol{r}^k_{\mathcal{B}^k \setminus \mathcal{B}^{k+1}}\right\|_2 \\
&\leq \sqrt{2}\left(\left\|\boldsymbol{r}^k_{\mathcal{B}^{k+1}}\right\|_2^2 + \left\|\boldsymbol{r}^k_{\mathcal{B}^k \setminus \mathcal{B}^{k+1}}\right\|_2^2\right)^{1/2} \\
&= \sqrt{2}\left\|\boldsymbol{r}^k_{\mathcal{B}^k \bigcup \mathcal{B}^{k+1}}\right\|_2 = \sqrt{2}\left\|\boldsymbol{r}^k\right\|_2 .
\end{aligned}$$

Finally,

$$\left\|\boldsymbol{r}^{k+1}\right\|_2 \leq 2\left\|\boldsymbol{x}_{0,\mathcal{B}^{k+1}} - \boldsymbol{a}^{k+1}_{\mathcal{B}^{k+1}}\right\|_2 \leq \sqrt{8}\delta_{3S}\left\|\boldsymbol{r}^k\right\|_2 + \sqrt{1+\delta_{3S}}\left\|\boldsymbol{w}\right\|_2 .$$

ℓ_p-Norm

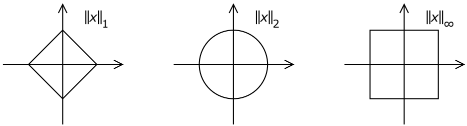
Definition 4.7 (ℓ_p-norm)

For a real number $p \geq 1$, the ℓ_p-norm of $\boldsymbol{x} \in \mathbb{R}^n$ is given by

$$\|\boldsymbol{x}\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{\frac{1}{p}}.$$

Examples

- ℓ_1-norm (Manhattan distance): $\|\boldsymbol{x}\|_1 = \sum |x_i|$.
- ℓ_2-norm (Euclidean norm): $\|\boldsymbol{x}\| = \sqrt{\sum x_i^2}$.
- ℓ_∞-norm: $\|\boldsymbol{x}\|_\infty = \max(|x_1|, \cdots, |x_n|)$.



The Hölder's Inequality

Theorem 4.8 (The Hölder's inequality)

Let $p, q \in [1, \infty]$ *with* $1/p + 1/q = 1$*.*
For all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$*, it holds that*

$$\sum_{i=1}^{n} |x_i \cdot y_i| \le \|\boldsymbol{x}\|_p \|\boldsymbol{y}\|_q = \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p} \cdot \left(\sum_{i=1}^{n} |y_i|^q\right)^{1/q}.$$

The equality holds iff $|x|^p$ *and* $|y|^q$ *are linear dependent, i.e.,* $\alpha |x_i|^p = \beta |y_i|^q$, $\forall i$*.*

(Proof is omitted.)

The Cauchy–Schwartz Inequality

Theorem 4.9 (The Cauchy-Schwartz Inequality)

A special case of the Hölder's inequality is when $p = q = 2$.

$$\sum_{i=1}^{n} |x_i \cdot y_i| \leq \|\boldsymbol{x}\|_2 \cdot \|\boldsymbol{y}\|_2 .$$

In particular, for all $\boldsymbol{x} \in \mathbb{R}^n$,

$$\|\boldsymbol{x}\|_1 = \sum_{i=1}^{n} |x_i| \leq \sqrt{n}\, \|\boldsymbol{x}\|_2 ,$$

where the equality holds iff $|x_i| = |x_j|$.

Section 5
Convex Optimisation 1

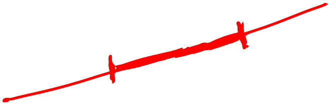
Convex Combination

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 $\boxed{\lambda_i \geq 0}$ and $\boxed{\sum_{i=1}^{n} \lambda_i = 1}$.



linear combination $\to$ line

Convex combination $\to$ section of line

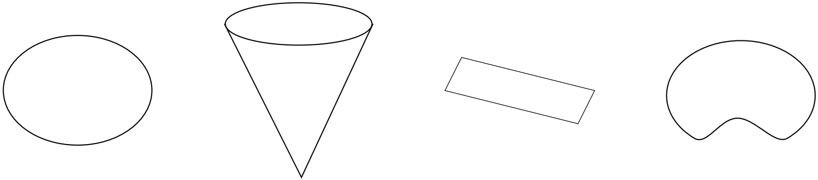
Convex Sets

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.,

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

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

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

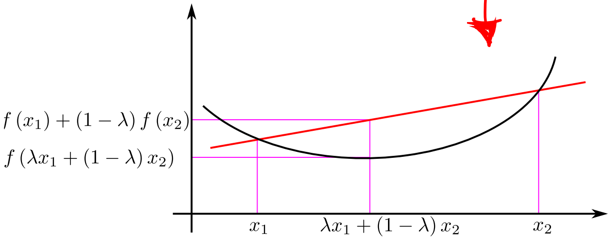
This definition implies that $\mathrm{dom} f$ is convex. However, in this lecture notes, we usually assume $\mathrm{dom}\ f = \mathbb{R}^n$ for simplicity.

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

x_1 x_2

A Convex Function

strictly convex.



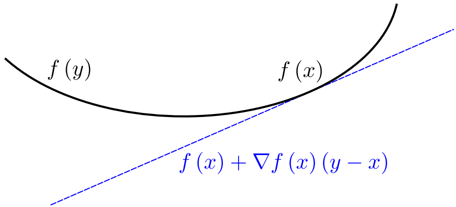
First-Order Condition of Convexity

very important.

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

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



Necessity

Assume first that f is convex and $\boldsymbol{x}, \boldsymbol{y} \in \mathrm{dom}\,(f)$. Since $\mathrm{dom}\,(f)$ is convex, $\boldsymbol{x} + t\,(\boldsymbol{y} - \boldsymbol{x}) \in \mathrm{dom}\,(f)$ 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

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

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

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

Now note that $\boldsymbol{x} - \boldsymbol{z} = (1-\lambda)(\boldsymbol{x} - \boldsymbol{y})$ and $\boldsymbol{y} - \boldsymbol{z} = -\lambda (\boldsymbol{x} - \boldsymbol{y})$. One obtains

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

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

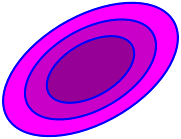
Lemma 5.7

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,

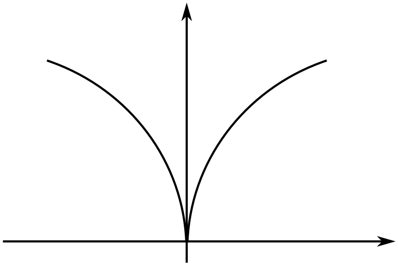
$$f(\lambda \boldsymbol{x} + (1 - \lambda) \boldsymbol{y}) \le \lambda f(\boldsymbol{x}) + (1 - \lambda) f(\boldsymbol{y}) \le \alpha,$$

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, $p(\mathbf{0}) = 0$ 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

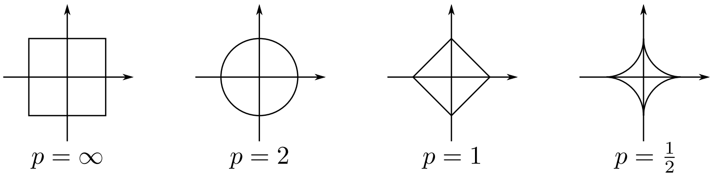
$$\begin{aligned} \|\lambda \boldsymbol{u} + (1-\lambda)\,\boldsymbol{v}\| &\leq \|\lambda \boldsymbol{u}\| + \|(1-\lambda)\,\boldsymbol{v}\| \\ &= \lambda\,\|\boldsymbol{u}\| + (1-\lambda)\,\|\boldsymbol{v}\|\,, \end{aligned}$$

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.



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 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

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

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 not 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

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

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

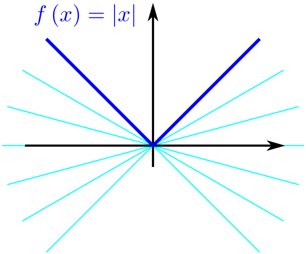
$$f(\boldsymbol{y}) - f(\boldsymbol{x}) \geq \boldsymbol{v}^T (\boldsymbol{y} - \boldsymbol{x}), \ \forall \boldsymbol{y} \in \mathcal{U}.$$

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

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



Section 6
ℓ_1-Minimization

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 $\|\boldsymbol{e}\| \le \epsilon$,

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

ℓ_1-norm convex

ℓ_2-norm [illegible]

Unconstrained optimization problem: LASSO

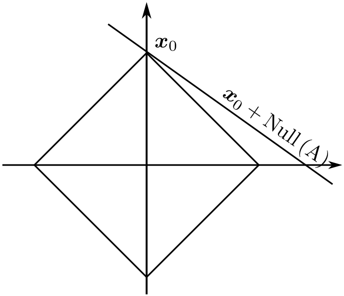
$$\min_{\boldsymbol{x}} \tfrac{1}{2} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \lambda \|\boldsymbol{x}\|_1 .$$

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

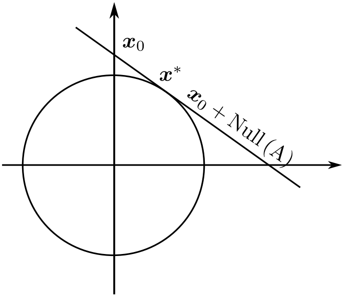
$$\boxed{\begin{array}{l} \blacktriangleright\ \lambda \to 0 \text{ implies } \epsilon \to 0. \\ \blacktriangleright\ \lambda \to \infty \text{ implies } \epsilon \to \infty. \end{array}}$$

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

except of 2 choises null(A) ⇒ 45° or -45°.

That's why we take l-1 norm.

Solve the Lasso Problem: Scalar Case

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

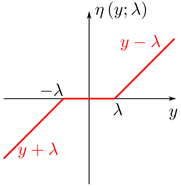
The minimum of $f(x)$ is achieved at $x^\#$ s.t. $\frac{d}{dx} f\left(x^\#\right) = 0$, i.e.,

$$x^\# - y + \lambda \left.\frac{d|x|}{dx}\right|_{x^\#} = 0, \text{ where } \frac{d|x|}{dx} = \begin{cases} 1 & \text{if } x > 0, \\ [-1, 1] & \text{if } x = 0, \\ -1 & \text{if } x < 0. \end{cases}$$

interval?

Or equivalently, $x^\#$ is given by the soft thresholding function

$$x^\# = \eta(y;\lambda) = \begin{cases} y - \lambda & \text{if } y \geq \lambda, \\ 0 & \text{if } -\lambda < y < \lambda, \\ y + \lambda & \text{if } y \leq -\lambda. \end{cases}$$



Vector Case: the Gradient Descent Method

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

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

where $t_k > 0$ is a suitable stepsize.

For Lasso problem $\frac{1}{2}\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2 + \lambda\|\boldsymbol{x}\|_1$, the gradient is given by (see details on page 6-22)

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

Gradient is not unique! Which one should one choose?

- Optimal gradient may depend on t_k.

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 $\tilde{f}$,

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

where

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

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,

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

Stable Recovery of Exact Sparse Signals

Theorem 6.1

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 6.2

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*

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

Typical values

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

Interpretation

Compressible signals: the entries obey a power law

$$|\boldsymbol{x}_0|_{(k)} \leq c \cdot k^{-r},$$

where $|\boldsymbol{x}_0|_{(k)}$ is the k^{th} largest value of $\boldsymbol{x}_0$, $r > 1$.

Consider the noiseless case. Suppose that a gene tells us the true signal $\boldsymbol{x}_0$. The best S-term approximation $\boldsymbol{x}_{0,S}$ gives a distortion

$$\|\boldsymbol{x}_0 - \boldsymbol{x}_{0,S}\|_2 \leq c' \cdot S^{-r+1/2} = c'' \frac{\|\boldsymbol{x}_0 - \boldsymbol{x}_{0,S}\|_1}{\sqrt{S}}.$$

(Computations details are given in Appendix on page 6-23.)

Compare this result with Theorem 6.2. There is no algorithm performing fundamentally better than ℓ_1-min.

Proof for Exact Sparse Signals

$\min \|x\|_1 \text{ s.t. } \|y - Ax\|_2 \le \epsilon,$

Tube constraint:

$$\|\boldsymbol{Ah}\|_2 = \left\|\boldsymbol{Ax}^{\#} - \boldsymbol{Ax}_0\right\|_2 \le \left\|\boldsymbol{Ax}^{\#} - \boldsymbol{y}\right\|_2 + \|\boldsymbol{Ax}_0 - \boldsymbol{y}\|_2 \le 2\epsilon.$$

$h := x^{\#} - x^0.$

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

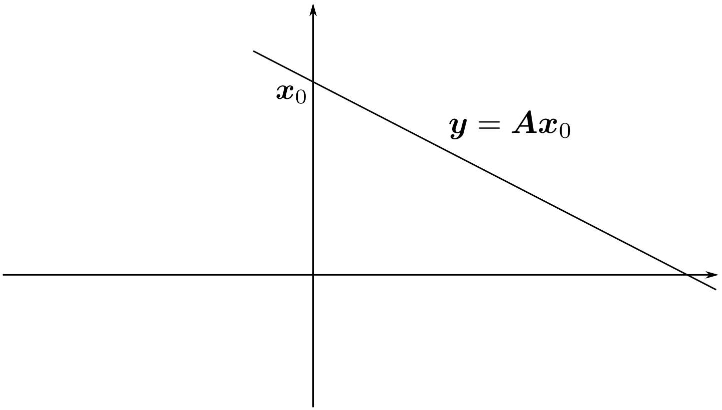
$$\boxed{\left\|\boldsymbol{h}_{\mathcal{T}_0^c}\right\|_1 \le \|\boldsymbol{h}_{\mathcal{T}_0}\|_1 .}$$

Proof:

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

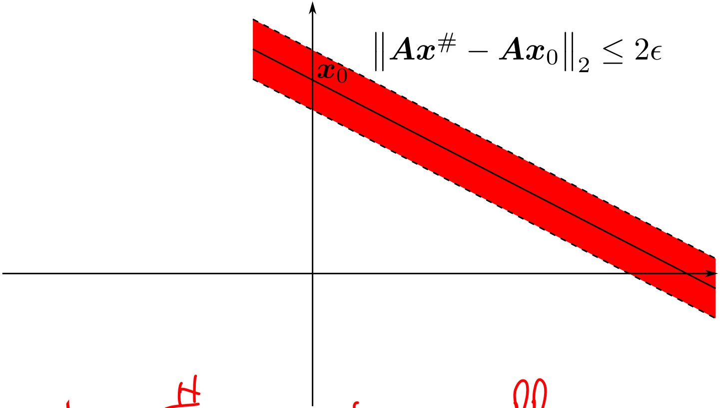
Geometric Interpretation

(without noise)



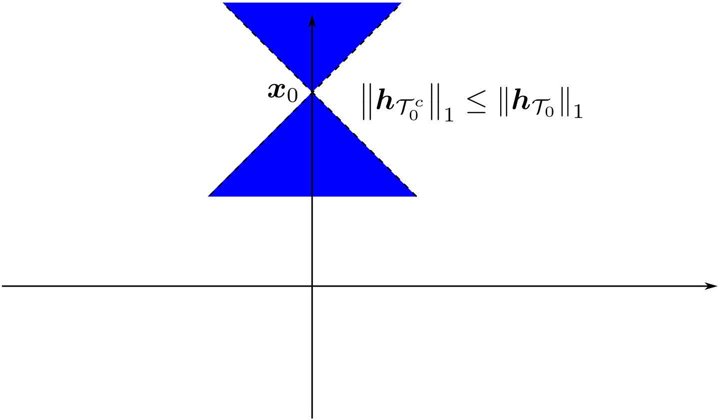
Geometric Interpretation

(with noise)

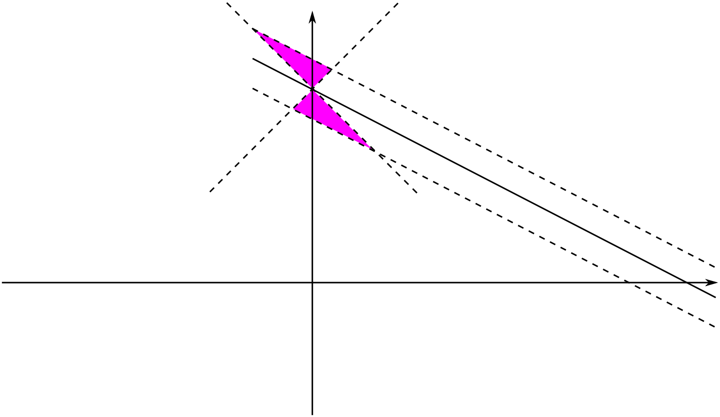


$h := x^{\#} - x_0$ is small.

Geometric Interpretation



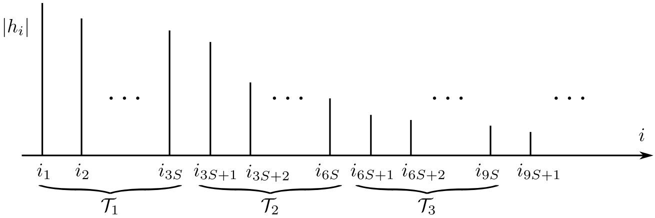
Geometric Interpretation



Proof

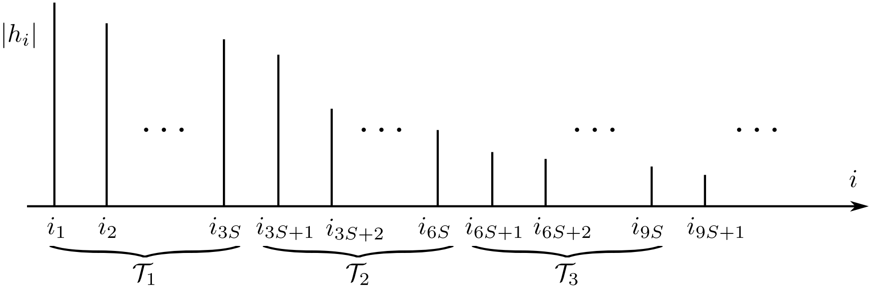
Since $\|\boldsymbol{Ah}\|_2 \leq 2\epsilon$, want to show $\|\boldsymbol{h}\|_2 \approx \|\boldsymbol{Ah}\|_2$.
(This is not true in general. For example $\boldsymbol{Ah} = \boldsymbol{0}$ but $\|\boldsymbol{h}\|_2$ can be ∞)

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 $\mathcal{T}_j = \{n_\ell,\ (j-1)\,M + 1 \leq \ell \leq jM\}$.
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

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

-

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

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.

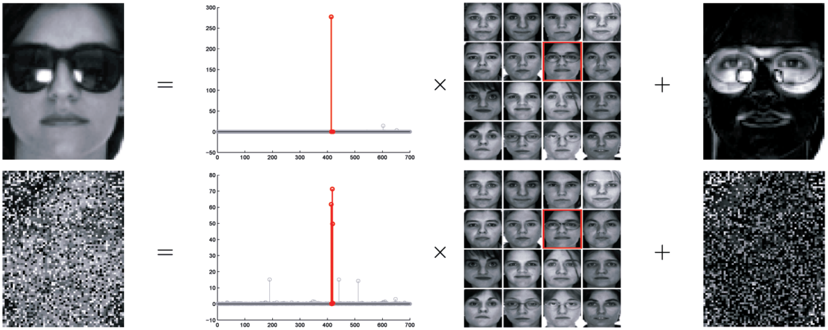
Proof: Step 3

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

Hence,

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

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 $\mathbb{R}^m$.

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

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:

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

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 6.3 (Gradient)

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

Example 6.4

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

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

Sparse Approximation Error for Compressible Signals

Let $|\boldsymbol{x}_0|_{(k)} \le c \cdot k^{-r}$. Then

$$|\boldsymbol{x}_0 - \boldsymbol{x}_{0,S}|_{(k)} \le \begin{cases} 0 & k \le S, \\ c \cdot k^{-r} & k > S. \end{cases}$$

- $$\begin{aligned} \|\boldsymbol{x}_0 - \boldsymbol{x}_{0,S}\|_1 &\le \sum_{k=S+1}^{n} ck^{-r} \le \sum_{k=S+1}^{\infty} ck^{-r} \\ &\le \int_S^{\infty} cx^{-r} dx = c' S^{-r+1}. \end{aligned}$$
- $\|\boldsymbol{x}_0 - \boldsymbol{x}_{0,S}\|_2^2 \le \sum_{k=S+1}^{\infty} ck^{-2r} \le c'' S^{-2r+1}$. Hence

$$\|\boldsymbol{x}_0 - \boldsymbol{x}_{0,S}\|_2 \le c''' S^{-r+\frac{1}{2}}.$$

Section 7
Low Rank Matrix Recovery

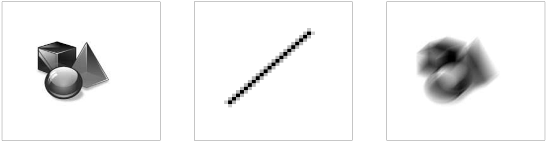
Netflix Problem

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

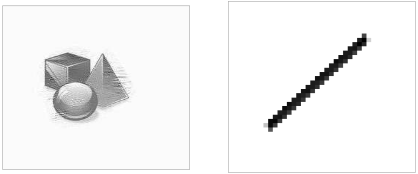
Blind Deconvolution [Ahmed, Recht, and Romberg, 2013]

using low rank matrix recovery approach.

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



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.,

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

is given by simply truncating the SVD

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

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

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 $\boldsymbol{A}_1, \cdots, \boldsymbol{A}_L$:

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

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

The low-rank matrix recovery problem is given by

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

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 $\mathcal{A}$

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

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

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

identity matrix

L rows will be left.

Alternating Projection

To solve

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

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.

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

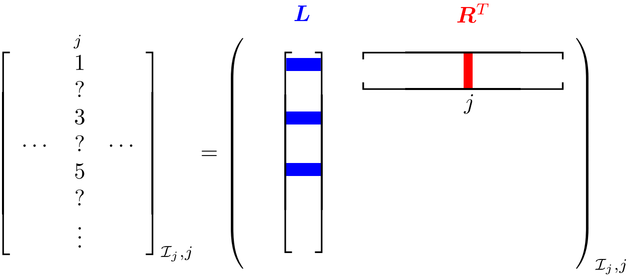
Alternating projection:

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

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

Alternating Projection (2)

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



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

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:

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

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.
$\|\boldsymbol{x}\|_1 = \sum_{i=1}^{n} |x_i|$.

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

Nuclear norm

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

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

Soft Thresholding Function

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

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

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

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

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

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

ISTA

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

ℓ_1-norm minimization

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

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

nuclear-norm minimization

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

transpose.

Iterative Hard Thresholding Algorithm

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

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

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

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

All greedy algorithms can be adapted to low rank matrix recovery

Comments on Performance Guarantees

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

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

- 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:

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

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.

$\hat{S} = S_0 \cdot C.$

$\hat{h} = h_0 \cdot \frac{1}{C}.$

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 $\boldsymbol{x}\boldsymbol{h}^T$:

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