Constrained optimization

DS-GA 1013 / MATH-GA 2824 Optimization-based Data Analysis

http://www.cims.nyu.edu/~cfgranda/pages/OBDA_fall17/index.html

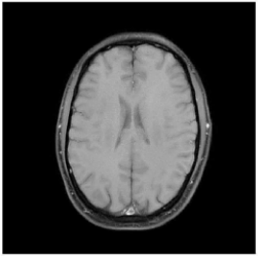
Carlos Fernandez-Granda

Compressed sensing

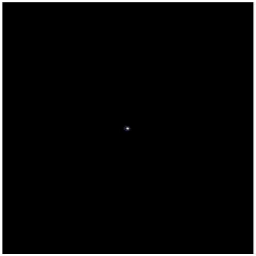
Convex constrained problems

Analyzing optimization-based methods

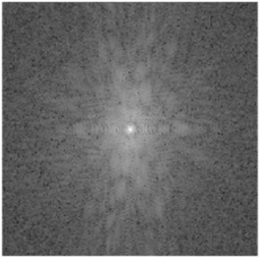
Magnetic resonance imaging



2D DFT (magnitude)



2D DFT (log. of magnitude)



Magnetic resonance imaging

Data: Samples from spectrum

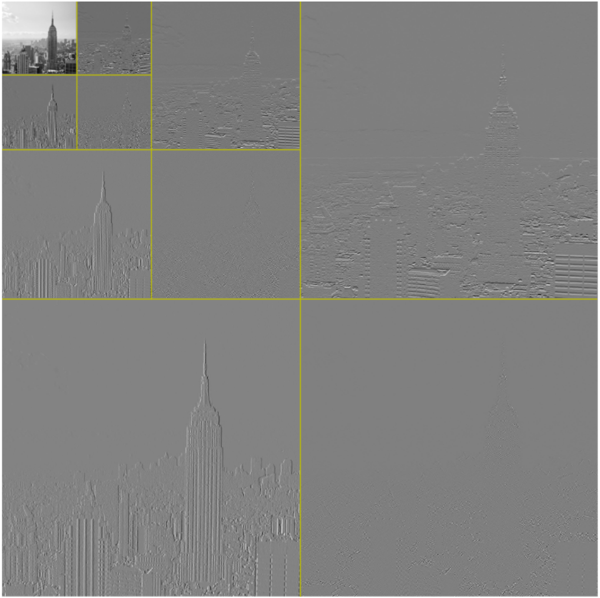
Problem: Sampling is time consuming (annoying, kids move . . .)

Images are compressible (sparse in wavelet basis)

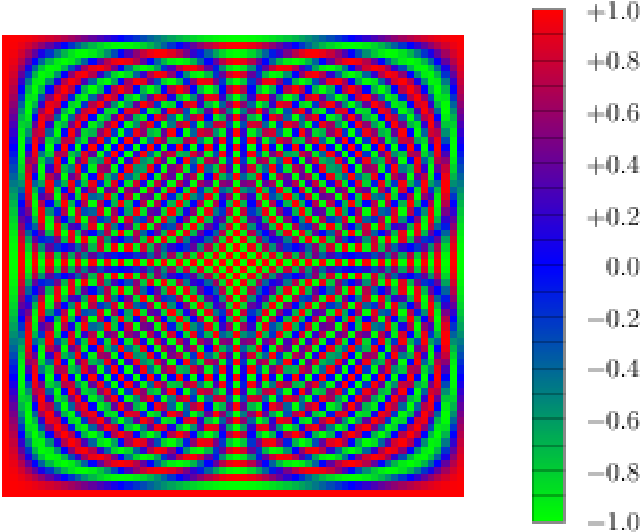
Can we recover compressible signals from less data?

2D wavelet transform

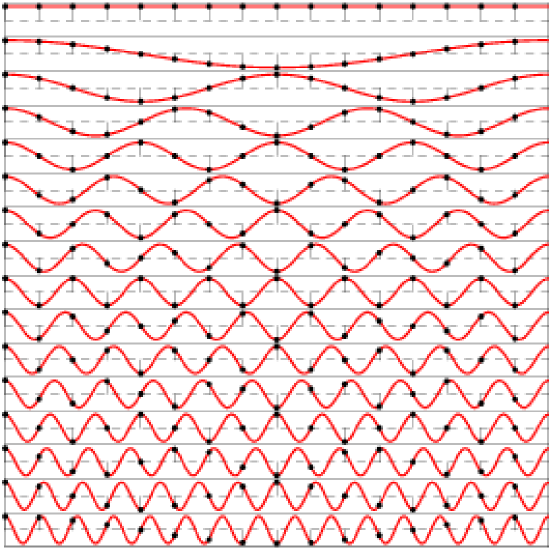
2D wavelet transform



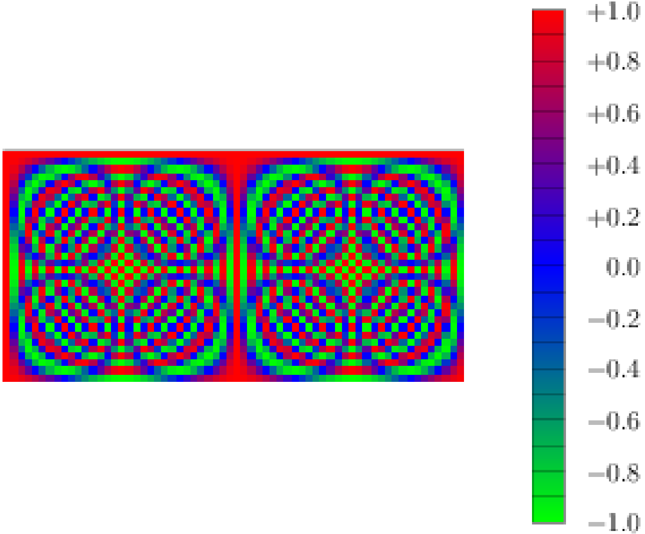
Full DFT matrix



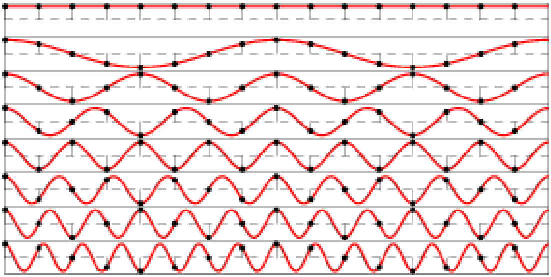
Full DFT matrix



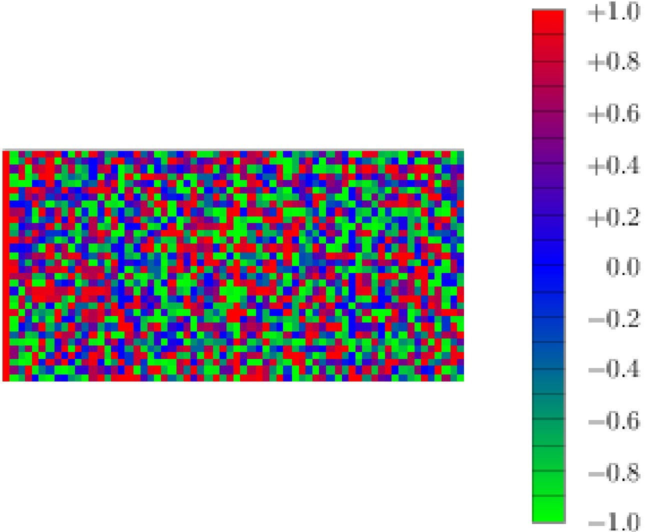
Regular subsampling



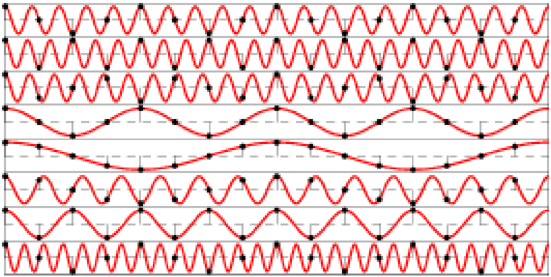
Regular subsampling



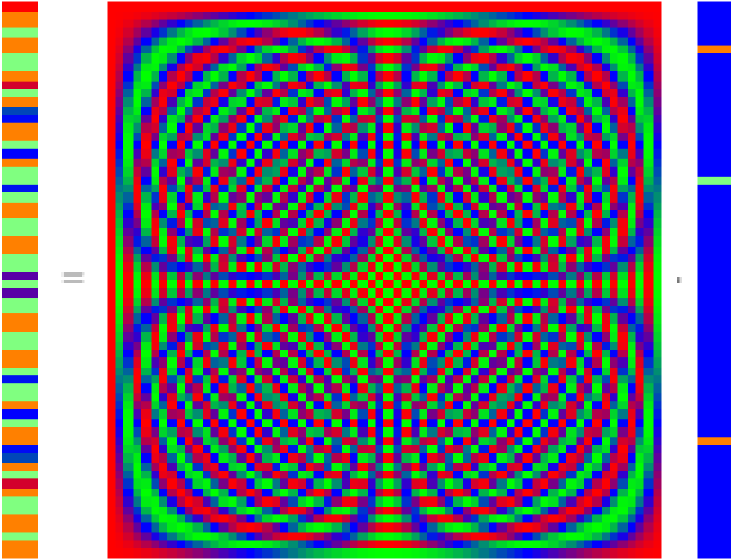
Random subsampling



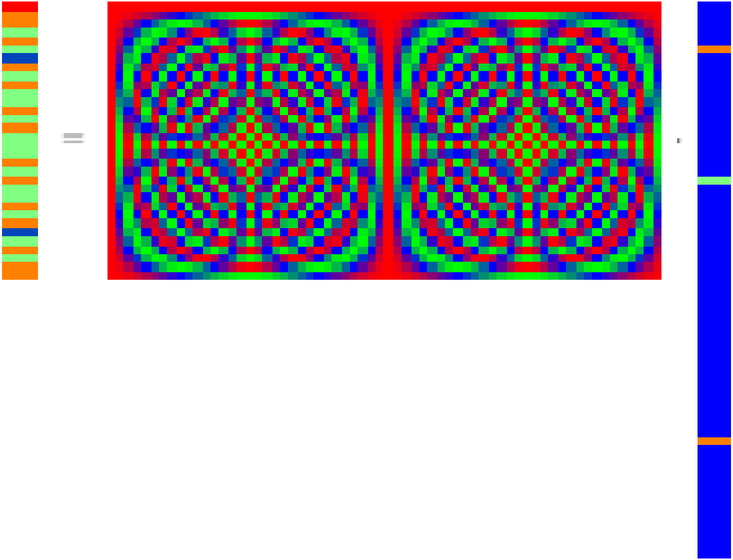
Random subsampling



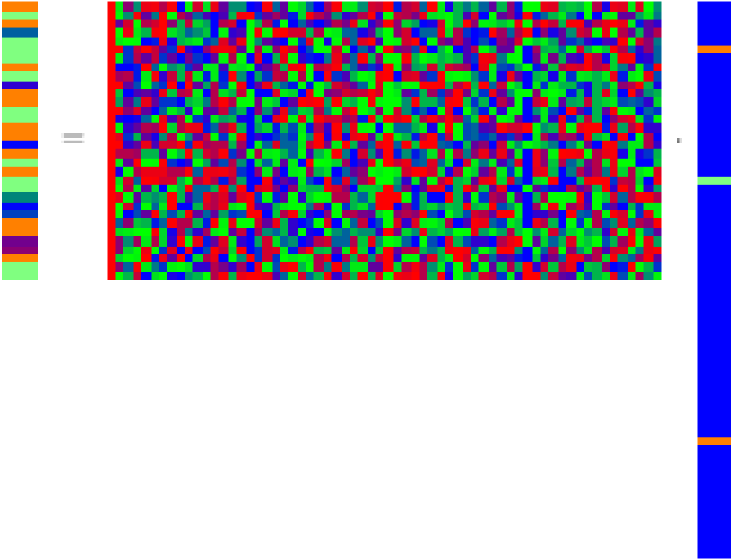
Toy example



Regular subsampling



Random subsampling



Linear inverse problems

Linear inverse problem

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

Linear measurements, $A \in \mathbb{R}^{m \times n}$

$$\vec{y}[i] = \langle A_{i:}, \vec{x} \rangle, \qquad 1 \leq i \leq n,$$

Aim: Recover data signal $\vec{x} \in \mathbb{R}^m$ from data $\vec{y} \in \mathbb{R}^n$

We need $n \geq m$, otherwise the problem is underdetermined

If $n < m$ there are infinite solutions $\vec{x} + \vec{w}$ where $\vec{w} \in \operatorname{null}(A)$

Sparse recovery

Aim: Recover sparse $\vec{x}$ from linear measurements

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

When is the problem well posed?

There shouldn't be two sparse vectors $\vec{x}_1$ and $\vec{x}_2$ such that $A\vec{x}_1 = A\vec{x}_2$

Spark

The spark of a matrix is the smallest subset of columns that is linearly dependent

Spark

The spark of a matrix is the smallest subset of columns that is linearly dependent

Let $\vec{y} := A\vec{x}^*$, where $A \in \mathbb{R}^{m\times n}$, $\vec{y} \in \mathbb{R}^n$ and $\vec{x}^* \in \mathbb{R}^m$ is a sparse vector with s nonzero entries

The vector $\vec{x}^*$ is the only vector with sparsity s consistent with the data, i.e. it is the solution of

$$\min_{\vec{x}} ||\vec{x}||_0 \qquad \text{subject to} \quad A\vec{x} = \vec{y}$$

for any choice of $\vec{x}^*$ if and only if

$$\text{spark}(A) > 2s$$

Proof

Equivalent statements

- For any $\vec{x}^*$, $\vec{x}^*$ is the only vector with sparsity s consistent with the data

Proof

Equivalent statements

- For any $\vec{x}^*$, $\vec{x}^*$ is the only vector with sparsity s consistent with the data

- For any pair of s-sparse vectors $\vec{x}_1$ and $\vec{x}_2$

$$A\left(\vec{x}_1 - \vec{x}_2\right) \neq \vec{0}$$

Proof

Equivalent statements

- For any $\vec{x}^*$, $\vec{x}^*$ is the only vector with sparsity s consistent with the data

- For any pair of s-sparse vectors $\vec{x}_1$ and $\vec{x}_2$

$$A(\vec{x}_1 - \vec{x}_2) \neq \vec{0}$$

- For any pair of subsets of s indices T_1 and T_2

$$A_{T_1 \cup T_2} \vec{\alpha} \neq \vec{0} \qquad \text{for any } \vec{\alpha} \in \mathbb{R}^{|T_1 \cup T_2|}$$

Proof

Equivalent statements

- For any $\vec{x}^*$, $\vec{x}^*$ is the only vector with sparsity s consistent with the data

- For any pair of s-sparse vectors $\vec{x}_1$ and $\vec{x}_2$

$$A\left(\vec{x}_1 - \vec{x}_2\right) \neq \vec{0}$$

- For any pair of subsets of s indices T_1 and T_2

$$A_{T_1 \cup T_2} \vec{\alpha} \neq \vec{0} \qquad \text{for any } \vec{\alpha} \in \mathbb{R}^{|T_1 \cup T_2|}$$

- All submatrices with at most $2s$ columns have no nonzero vectors in their null space

Proof

Equivalent statements

- For any $\vec{x}^*$, $\vec{x}^*$ is the only vector with sparsity s consistent with the data

- For any pair of s-sparse vectors $\vec{x}_1$ and $\vec{x}_2$

$$A(\vec{x}_1 - \vec{x}_2) \neq \vec{0}$$

- For any pair of subsets of s indices T_1 and T_2

$$A_{T_1 \cup T_2}\vec{\alpha} \neq \vec{0} \qquad \text{for any } \vec{\alpha} \in \mathbb{R}^{|T_1 \cup T_2|}$$

- All submatrices with at most $2s$ columns have no nonzero vectors in their null space

- All submatrices with at most $2s$ columns are full rank

Restricted-isometry property

Robust version of spark

If two s-sparse vectors $\vec{x}_1$, $\vec{x}_2$ are far, then $A\vec{x}_1$, $A\vec{x}_2$ should be far

The linear operator should preserve distances (be an isometry) when restricted to act upon sparse vectors

Restricted-isometry property

A satisfies the restricted isometry property (RIP) with constant κ_s if

$$(1 - \kappa_s) \left|\left|\vec{x}\right|\right|_2 \leq \left|\left|A\vec{x}\right|\right|_2 \leq (1 + \kappa_s) \left|\left|\vec{x}\right|\right|_2$$

for any s-sparse vector $\vec{x}$

If A satisfies the RIP for a sparsity level $2s$ then for any s-sparse $\vec{x}_1$, $\vec{x}_2$

$$\left|\left|\vec{y}_2 - \vec{y}_1\right|\right|_2$$

Restricted-isometry property

A satisfies the restricted isometry property (RIP) with constant κ_s if

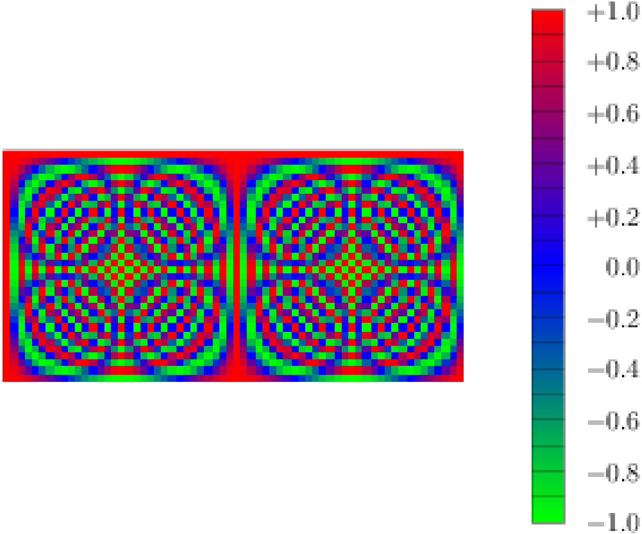
$$(1 - \kappa_s) \, ||\vec{x}||_2 \leq ||A\vec{x}||_2 \leq (1 + \kappa_s) \, ||\vec{x}||_2$$

for any s-sparse vector $\vec{x}$

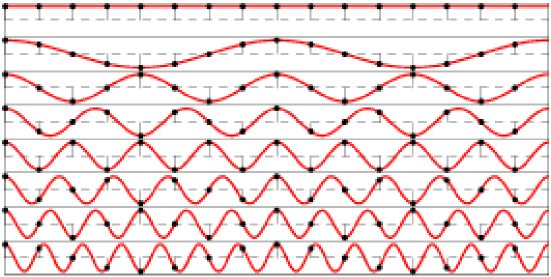
If A satisfies the RIP for a sparsity level $2s$ then for any s-sparse $\vec{x}_1$, $\vec{x}_2$

$$\begin{aligned} ||\vec{y}_2 - \vec{y}_1||_2 &= A \left(\vec{x}_1 - \vec{x}_2 \right) \\ &\geq (1 - \kappa_{2s}) \, ||\vec{x}_2 - \vec{x}_1||_2 \end{aligned}$$

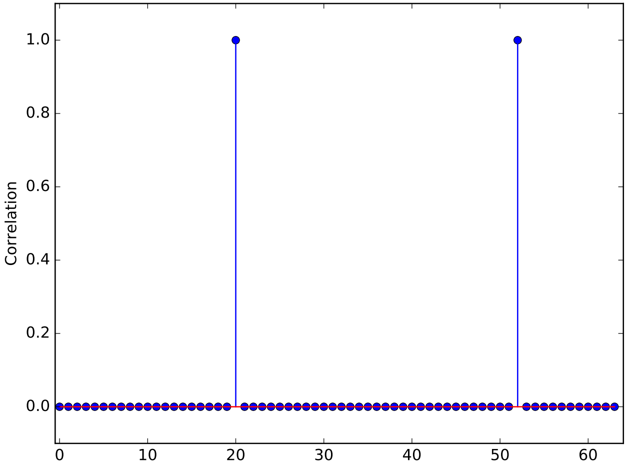
Regular subsampling



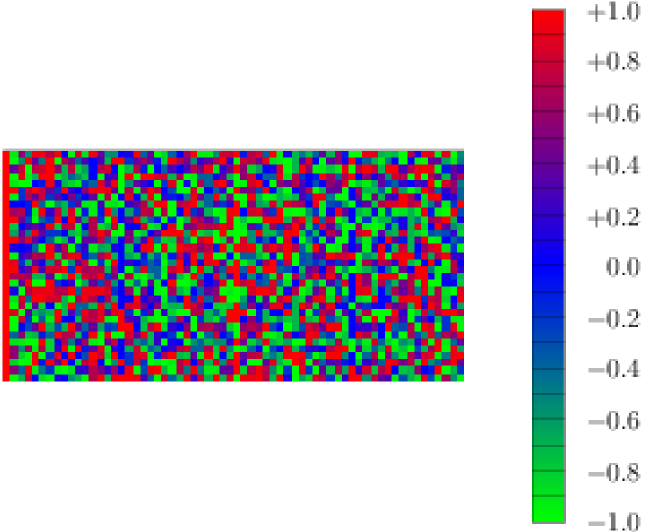
Regular subsampling



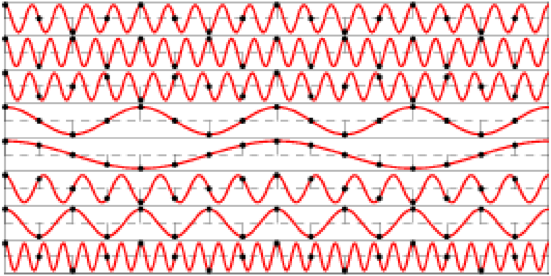
Correlation with column 20



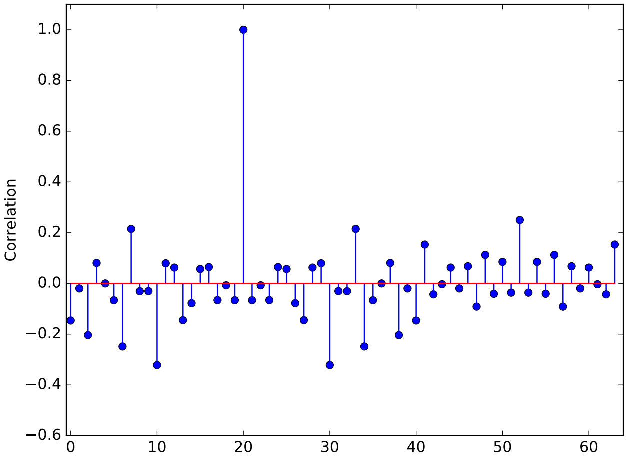
Random subsampling



Random subsampling



Correlation with column 20



Restricted-isometry property

Deterministic matrices tend to not satisfy the RIP

It is NP-hard to check if spark or RIP hold

Random matrices satisfy RIP with high probability

We prove it for Gaussian iid matrices, ideas in proof for random Fourier matrices are similar

Restricted-isometry property for Gaussian matrices

Let $\mathbf{A} \in \mathbb{R}^{m \times n}$ be a random matrix with iid standard Gaussian entries

$\frac{1}{\sqrt{m}}\mathbf{A}$ satisfies the RIP for a constant κ_s with probability $1 - \frac{C_2}{n}$ as long as

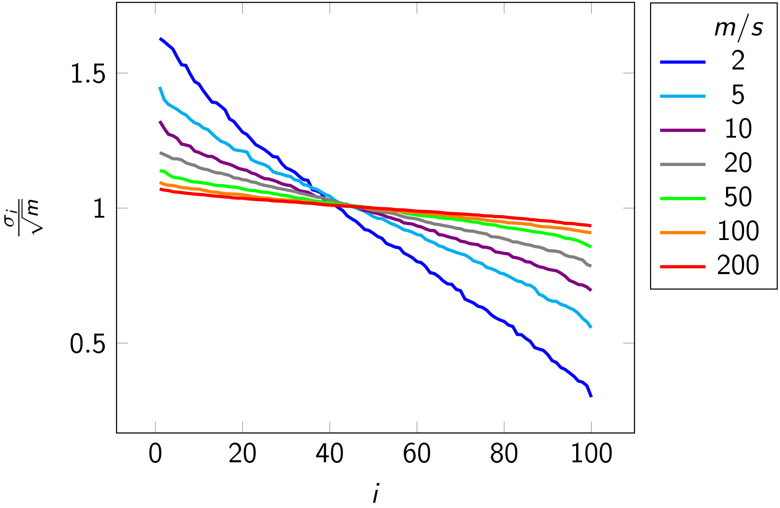
$$m \geq \frac{C_1 s}{\kappa_s^2} \log\left(\frac{n}{s}\right)$$

for two fixed constants $C_1, C_2 > 0$

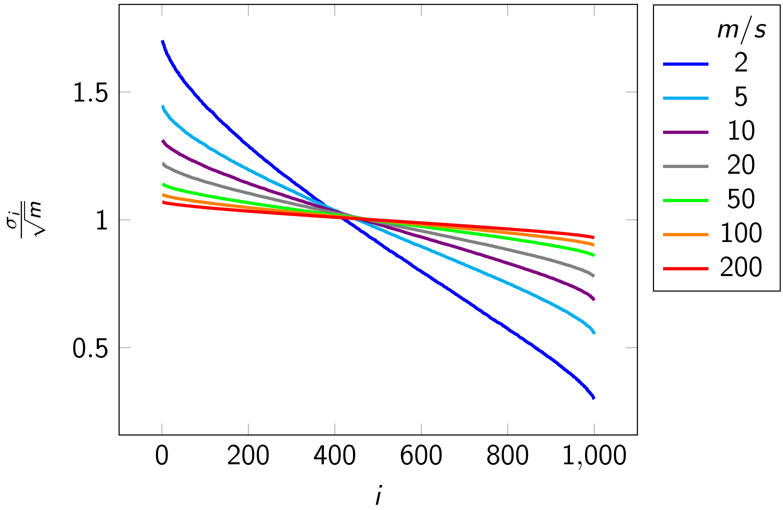
Proof

For a fixed support T of size s bounds follow from bounds on singular values of Gaussian matrices

Singular values of $m \times s$ matrix, $s = 100$



Singular values of $m \times s$ matrix, $s = 1000$



Proof

For a fixed submatrix the singular values are bounded by

$$\sqrt{m}\,(1-\kappa_s) \leq \boldsymbol{\sigma_s} \leq \boldsymbol{\sigma_1} \leq \sqrt{m}\,(1+\kappa_s)$$

with probability at least

$$1 - 2\left(\frac{12}{\kappa_s}\right)^s \exp\left(-\frac{m\kappa_s^2}{32}\right)$$

For any vector $\vec{x}$ with support T

$$\sqrt{1-\kappa_s}\,||\vec{x}||_2 \leq \frac{1}{\sqrt{m}}\,||\mathbf{A}\vec{x}||_2 \leq \sqrt{1+\kappa_s}\,||\vec{x}||_2$$

Union bound

For any events $S_1, S_2, \ldots, S_n$ in a probability space

$$\mathrm{P}\left(\cup_i S_i\right) \leq \sum_{i=1}^{n} \mathrm{P}\left(S_i\right).$$

Proof

Number of different supports of size s

Proof

Number of different supports of size s

$$\binom{n}{s} \leq \left(\frac{en}{s}\right)^s$$

Proof

Number of different supports of size s

$$\binom{n}{s} \leq \left(\frac{en}{s}\right)^s$$

By the union bound

$$\sqrt{1-\kappa_s}\,||\vec{x}||_2 \leq \frac{1}{\sqrt{m}}\,||\mathbf{A}\vec{x}||_2 \leq \sqrt{1+\kappa_s}\,||\vec{x}||_2$$

holds for any s-sparse vector $\vec{x}$ with probability at least

$$\begin{aligned}
&1-2\left(\frac{en}{s}\right)^s\left(\frac{12}{\kappa_s}\right)^s \exp\left(-\frac{m\kappa_s^2}{32}\right)\\
&= 1-\exp\left(\log 2 + s + s\log\left(\frac{n}{s}\right) + s\log\left(\frac{12}{\kappa_s}\right) - \frac{m\kappa_s^2}{2}\right)\\
&\leq 1-\frac{C_2}{n} \qquad \text{as long as} \quad m \geq \frac{C_1 s}{\kappa_s^2}\log\left(\frac{n}{s}\right)
\end{aligned}$$

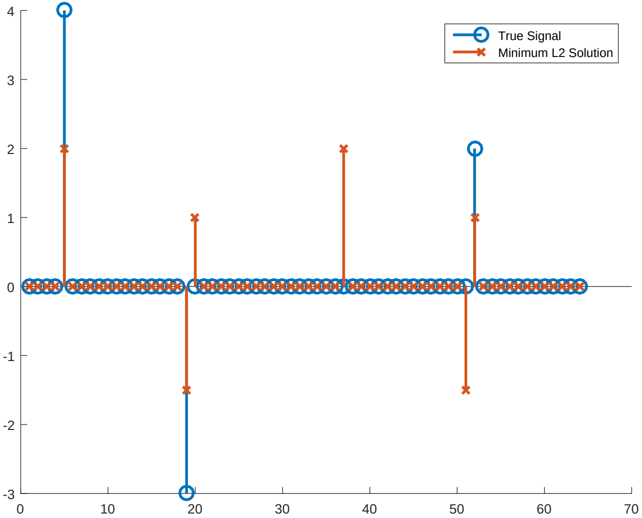
Sparse recovery via ℓ_1-norm minimization

ℓ_0-"norm" minimization is intractable

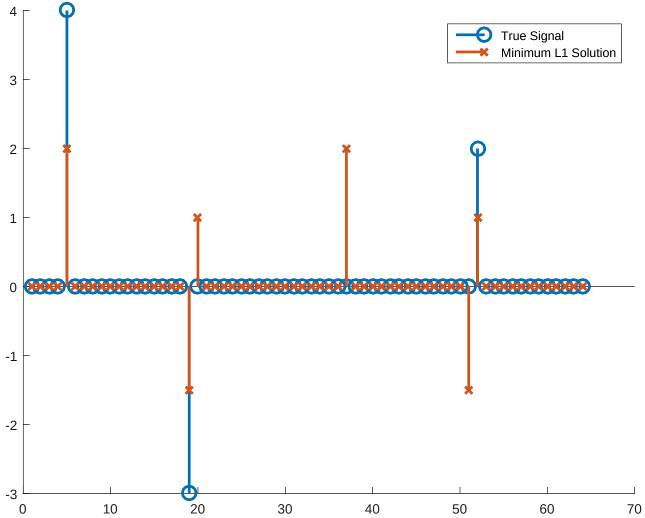
(As usual) we can minimize ℓ_1 norm instead, estimate $\vec{x}_{\ell_1}$ is the solution to

$$\min_{\vec{x}} ||\vec{x}||_1 \qquad \text{subject to} \quad A\vec{x} = \vec{y}$$

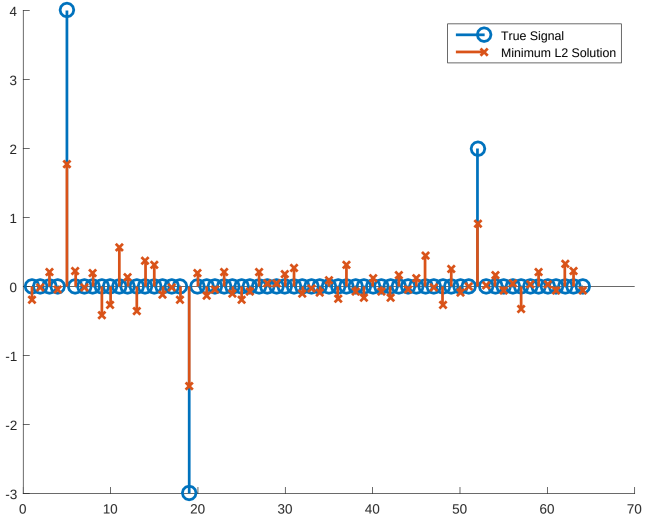
Minimum ℓ_2-norm solution (regular subsampling)



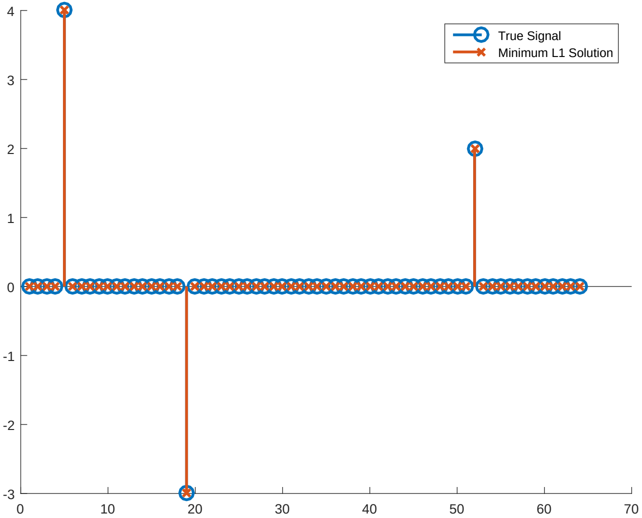
Minimum ℓ_1-norm solution (regular subsampling)



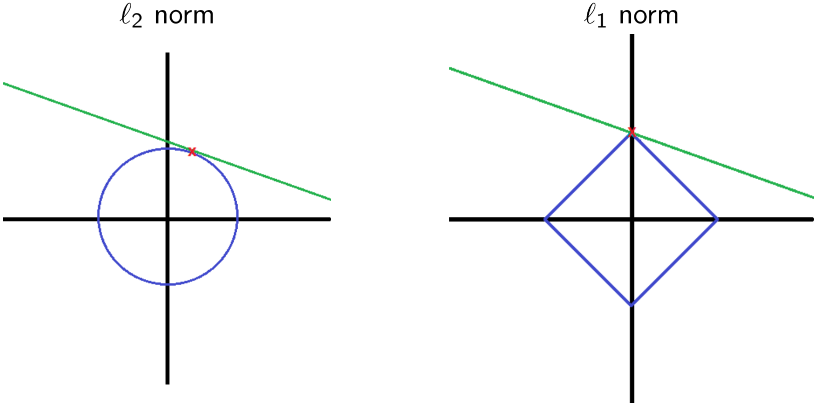
Minimum ℓ_2-norm solution (random subsampling)



Minimum ℓ_1-norm solution (random subsampling)



Geometric intuition



Sparse recovery via ℓ_1-norm minimization

If the signal is sparse in a transform domain then

$$\min_{\vec{x}} ||\vec{c}||_1 \qquad \text{subject to} \qquad AW\vec{c} = \vec{y}$$

If we want to recover the original $\vec{c}^*$ then AW should satisfy the RIP

Sparse recovery via ℓ_1-norm minimization

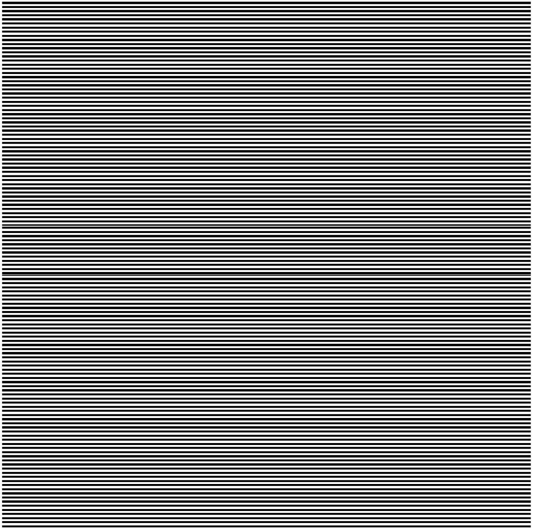
If the signal is sparse in a transform domain then

$$\min_{\vec{x}} ||\vec{c}||_1 \qquad \text{subject to} \qquad AW\vec{c} = \vec{y}$$

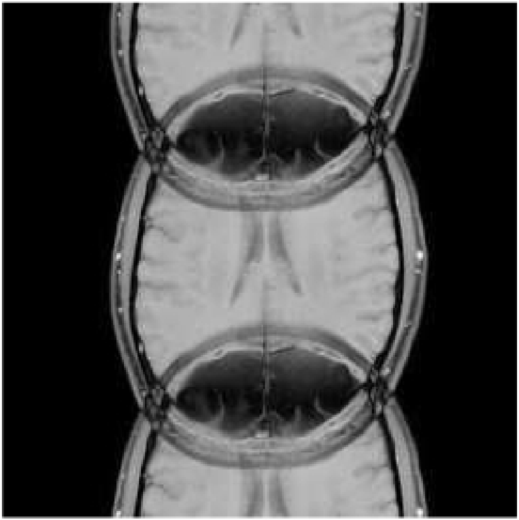
If we want to recover the original $\vec{c}^*$ then AW should satisfy the RIP

However, we might be fine with any $\vec{c}'$ such that $A\vec{c}' = \vec{x}^*$

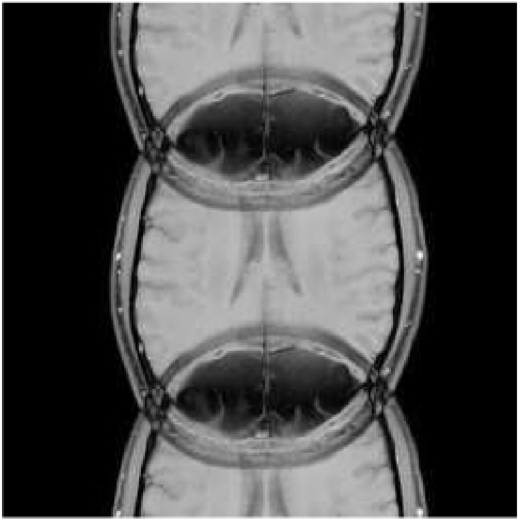
Regular subsampling



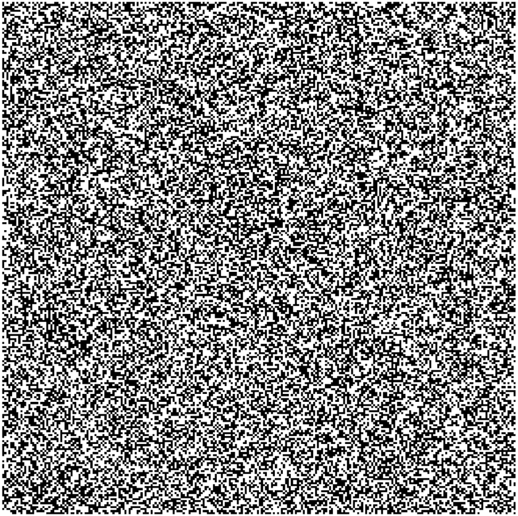
Minimum ℓ_2-norm solution (regular subsampling)



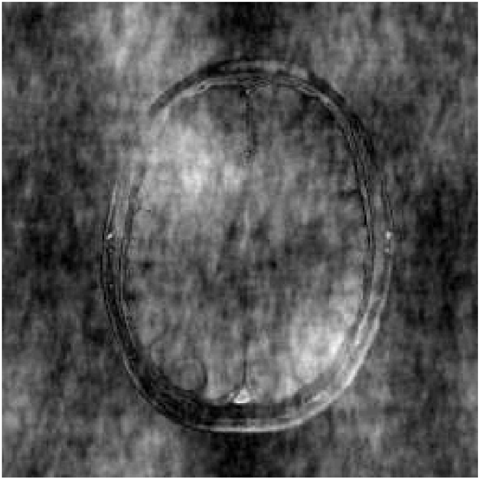
Minimum ℓ_1-norm solution (regular subsampling)



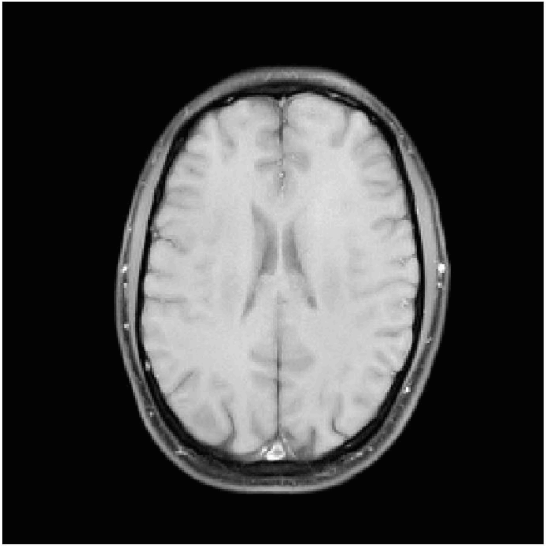
Random subsampling



Minimum ℓ_2-norm solution (random subsampling)



Minimum ℓ_1-norm solution (random subsampling)



Compressed sensing

Convex constrained problems

Analyzing optimization-based methods

Convex sets

A convex set $\mathcal{S}$ is any set such that for any $\vec{x}, \vec{y} \in \mathcal{S}$ and $\theta \in (0,1)$

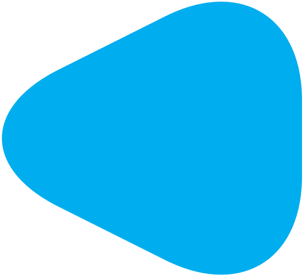
$$\theta \vec{x} + (1-\theta)\, \vec{y} \in \mathcal{S}$$

The intersection of convex sets is convex

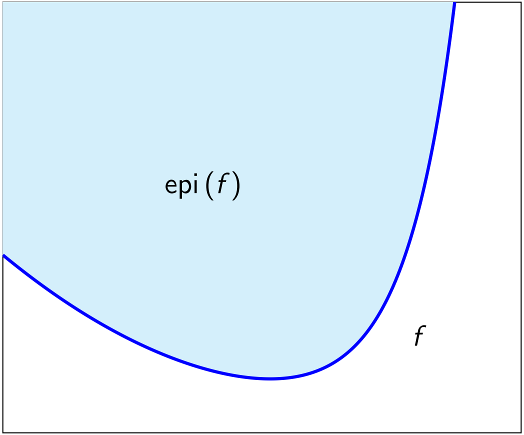
Convex vs nonconvex

Nonconvex

Convex



Epigraph



A function is convex if and only if its epigraph is convex

Projection onto convex set

The projection of any vector $\vec{x}$ onto a non-empty closed convex set $\mathcal{S}$

$$\mathcal{P}_{\mathcal{S}}\left(\vec{x}\right) := \arg\min_{\vec{y} \in \mathcal{S}} ||\vec{x} - \vec{y}||_2$$

exists and is unique

Proof

Assume there are two distinct projections $\vec{y}_1 \neq \vec{y}_2$

Consider

$$\vec{y}\,' := \frac{\vec{y}_1 + \vec{y}_2}{2}$$

$\vec{y}\,'$ belongs to $\mathcal{S}$ (why?)

Proof

$$\left\langle \vec{x} - \vec{y}\,', \vec{y}_1 - \vec{y}\,' \right\rangle = \left\langle \vec{x} - \frac{\vec{y}_1 + \vec{y}_2}{2}, \vec{y}_1 - \frac{\vec{y}_1 + \vec{y}_2}{2} \right\rangle$$
$$= \left\langle \frac{\vec{x} - \vec{y}_1}{2} + \frac{\vec{x} - \vec{y}_2}{2}, \frac{\vec{x} - \vec{y}_1}{2} - \frac{\vec{x} - \vec{y}_2}{2} \right\rangle$$

Proof

$$\begin{aligned}
\left\langle \vec{x} - \vec{y}\,', \vec{y}_1 - \vec{y}\,' \right\rangle &= \left\langle \vec{x} - \frac{\vec{y}_1 + \vec{y}_2}{2}, \vec{y}_1 - \frac{\vec{y}_1 + \vec{y}_2}{2} \right\rangle \\
&= \left\langle \frac{\vec{x} - \vec{y}_1}{2} + \frac{\vec{x} - \vec{y}_2}{2}, \frac{\vec{x} - \vec{y}_1}{2} - \frac{\vec{x} - \vec{y}_2}{2} \right\rangle \\
&= \frac{1}{4}\left(||\vec{x} - \vec{y}_1||^2 + ||\vec{x} - \vec{y}_2||^2 \right) \\
&= 0
\end{aligned}$$

Proof

$$\left\langle \vec{x} - \vec{y}\,', \vec{y}_1 - \vec{y}\,' \right\rangle = \left\langle \vec{x} - \frac{\vec{y}_1 + \vec{y}_2}{2}, \vec{y}_1 - \frac{\vec{y}_1 + \vec{y}_2}{2} \right\rangle$$
$$= \left\langle \frac{\vec{x} - \vec{y}_1}{2} + \frac{\vec{x} - \vec{y}_2}{2}, \frac{\vec{x} - \vec{y}_1}{2} - \frac{\vec{x} - \vec{y}_2}{2} \right\rangle$$
$$= \frac{1}{4}\left(||\vec{x} - \vec{y}_1||^2 + ||\vec{x} - \vec{y}_2||^2 \right)$$
$$= 0$$

By Pythagoras' theorem

$$||\vec{x} - \vec{y}_1||_2^2$$

Proof

$$\begin{aligned}\left\langle \vec{x} - \vec{y}\,', \vec{y}_1 - \vec{y}\,' \right\rangle &= \left\langle \vec{x} - \frac{\vec{y}_1 + \vec{y}_2}{2}, \vec{y}_1 - \frac{\vec{y}_1 + \vec{y}_2}{2} \right\rangle \\ &= \left\langle \frac{\vec{x} - \vec{y}_1}{2} + \frac{\vec{x} - \vec{y}_2}{2}, \frac{\vec{x} - \vec{y}_1}{2} - \frac{\vec{x} - \vec{y}_2}{2} \right\rangle \\ &= \frac{1}{4}\left(||\vec{x} - \vec{y}_1||^2 + ||\vec{x} - \vec{y}_2||^2 \right) \\ &= 0\end{aligned}$$

By Pythagoras' theorem

$$||\vec{x} - \vec{y}_1||_2^2 = \left|\left|\vec{x} - \vec{y}\,'\right|\right|_2^2 + \left|\left|\vec{y}_1 - \vec{y}\,'\right|\right|_2^2$$

Proof

$$\begin{aligned}\left\langle \vec{x} - \vec{y}\,', \vec{y}_1 - \vec{y}\,' \right\rangle &= \left\langle \vec{x} - \frac{\vec{y}_1 + \vec{y}_2}{2}, \vec{y}_1 - \frac{\vec{y}_1 + \vec{y}_2}{2} \right\rangle \\ &= \left\langle \frac{\vec{x} - \vec{y}_1}{2} + \frac{\vec{x} - \vec{y}_2}{2}, \frac{\vec{x} - \vec{y}_1}{2} - \frac{\vec{x} - \vec{y}_2}{2} \right\rangle \\ &= \frac{1}{4}\left(||\vec{x} - \vec{y}_1||^2 + ||\vec{x} - \vec{y}_2||^2 \right) \\ &= 0\end{aligned}$$

By Pythagoras' theorem

$$\begin{aligned}||\vec{x} - \vec{y}_1||_2^2 &= \left|\left|\vec{x} - \vec{y}\,'\right|\right|_2^2 + \left|\left|\vec{y}_1 - \vec{y}\,'\right|\right|_2^2 \\ &= \left|\left|\vec{x} - \vec{y}\,'\right|\right|_2^2 + \left|\left|\frac{\vec{y}_1 - \vec{y}_2}{2}\right|\right|_2^2\end{aligned}$$

Proof

$$\left\langle \vec{x} - \vec{y}\,', \vec{y}_1 - \vec{y}\,' \right\rangle = \left\langle \vec{x} - \frac{\vec{y}_1 + \vec{y}_2}{2}, \vec{y}_1 - \frac{\vec{y}_1 + \vec{y}_2}{2} \right\rangle$$

$$= \left\langle \frac{\vec{x} - \vec{y}_1}{2} + \frac{\vec{x} - \vec{y}_2}{2}, \frac{\vec{x} - \vec{y}_1}{2} - \frac{\vec{x} - \vec{y}_2}{2} \right\rangle$$

$$= \frac{1}{4}\left(||\vec{x} - \vec{y}_1||^2 + ||\vec{x} - \vec{y}_2||^2 \right)$$

$$= 0$$

By Pythagoras' theorem

$$||\vec{x} - \vec{y}_1||_2^2 = \left|\left|\vec{x} - \vec{y}\,'\right|\right|_2^2 + \left|\left|\vec{y}_1 - \vec{y}\,'\right|\right|_2^2$$

$$= \left|\left|\vec{x} - \vec{y}\,'\right|\right|_2^2 + \left|\left|\frac{\vec{y}_1 - \vec{y}_2}{2}\right|\right|_2^2$$

$$> \left|\left|\vec{x} - \vec{y}\,'\right|\right|_2^2$$

Convex combination

Given n vectors $\vec{x}_1, \vec{x}_2, \dots, \vec{x}_n \in \mathbb{R}^n$,

$$\vec{x} := \sum_{i=1}^{n} \theta_i \vec{x}_i$$

is a convex combination of $\vec{x}_1, \vec{x}_2, \dots, \vec{x}_n$ if

$$\theta_i \geq 0, \quad 1 \leq i \leq n$$

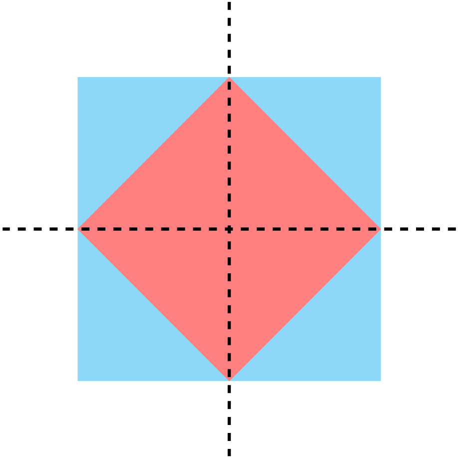
$$\sum_{i=1}^{n} \theta_i = 1$$

Convex hull

The convex hull of $\mathcal{S}$ is the set of convex combinations of points in $\mathcal{S}$

The ℓ_1-norm ball is the convex hull of the intersection between the ℓ_0 “norm" ball and the ℓ_∞-norm ball

ℓ_1-norm ball



$\mathcal{B}_{\ell_1} \subseteq \mathcal{C}\left(\mathcal{B}_{\ell_0} \cap \mathcal{B}_{\ell_\infty}\right)$

Let $\vec{x} \in \mathcal{B}_{\ell_1}$

Set $\theta_i := |\vec{x}[i]|$, $\theta_0 = 1 - \sum_{i=1}^{n} \theta_i$

$\sum_{i=0}^{n} \theta_i = 1$ by construction, $\theta_i \geq 0$ and

$$\begin{aligned}\theta_0 &= 1 - \sum_{i=1}^{n+1} \theta_i \\ &= 1 - ||\vec{x}||_1 \\ &\geq 0 \qquad \text{because } \vec{x} \in \mathcal{B}_{\ell_1}\end{aligned}$$

$\mathcal{B}_{\ell_1} \subseteq \mathcal{C}\left(\mathcal{B}_{\ell_0} \cap \mathcal{B}_{\ell_\infty}\right)$

Let $\vec{x} \in \mathcal{B}_{\ell_1}$

Set $\theta_i := |\vec{x}[i]|$, $\theta_0 = 1 - \sum_{i=1}^{n} \theta_i$

$\sum_{i=0}^{n} \theta_i = 1$ by construction, $\theta_i \geq 0$ and

$$\begin{aligned} \theta_0 &= 1 - \sum_{i=1}^{n+1} \theta_i \\ &= 1 - ||\vec{x}||_1 \\ &\geq 0 \qquad \text{because } \vec{x} \in \mathcal{B}_{\ell_1} \end{aligned}$$

$\vec{x} \in \mathcal{B}_{\ell_0} \cap \mathcal{B}_{\ell_\infty}$ because

$$\vec{x} = \sum_{i=1}^{n} \theta_i \operatorname{sign}\left(\vec{x}[i]\right) \vec{e}_i + \theta_0 \vec{0}$$

$\mathcal{C}\left(\mathcal{B}_{\ell_0} \cap \mathcal{B}_{\ell_\infty}\right) \subseteq \mathcal{B}_{\ell_1}$

Let $\vec{x} \in \mathcal{C}\left(\mathcal{B}_{\ell_0} \cap \mathcal{B}_{\ell_\infty}\right)$, then

$$\vec{x} = \sum_{i=1}^{m} \theta_i \vec{y}_i$$

$\mathcal{C}\left(\mathcal{B}_{\ell_0} \cap \mathcal{B}_{\ell_\infty}\right) \subseteq \mathcal{B}_{\ell_1}$

Let $\vec{x} \in \mathcal{C}\left(\mathcal{B}_{\ell_0} \cap \mathcal{B}_{\ell_\infty}\right)$, then

$$\vec{x} = \sum_{i=1}^{m} \theta_i \vec{y}_i$$

$$||\vec{x}||_1$$

$\mathcal{C}\left(\mathcal{B}_{\ell_0} \cap \mathcal{B}_{\ell_\infty}\right) \subseteq \mathcal{B}_{\ell_1}$

Let $\vec{x} \in \mathcal{C}\left(\mathcal{B}_{\ell_0} \cap \mathcal{B}_{\ell_\infty}\right)$, then

$$\vec{x} = \sum_{i=1}^{m} \theta_i \vec{y}_i$$

$$||\vec{x}||_1 \leq \sum_{i=1}^{m} \theta_i \, ||\vec{y}_i||_1 \quad \text{by the Triangle inequality}$$

$\mathcal{C}\left(\mathcal{B}_{\ell_0} \cap \mathcal{B}_{\ell_\infty}\right) \subseteq \mathcal{B}_{\ell_1}$

Let $\vec{x} \in \mathcal{C}\left(\mathcal{B}_{\ell_0} \cap \mathcal{B}_{\ell_\infty}\right)$, then

$$\vec{x} = \sum_{i=1}^{m} \theta_i \vec{y}_i$$

$$\begin{aligned} \left|\left|\vec{x}\right|\right|_1 &\leq \sum_{i=1}^{m} \theta_i \left|\left|\vec{y}_i\right|\right|_1 \quad \text{by the Triangle inequality} \\ &\leq \sum_{i=1}^{m} \theta_i \left|\left|\vec{y}_i\right|\right|_\infty \quad \vec{y}_i \text{ only has one nonzero entry} \end{aligned}$$

$\mathcal{C}\left(\mathcal{B}_{\ell_0} \cap \mathcal{B}_{\ell_\infty}\right) \subseteq \mathcal{B}_{\ell_1}$

Let $\vec{x} \in \mathcal{C}\left(\mathcal{B}_{\ell_0} \cap \mathcal{B}_{\ell_\infty}\right)$, then

$$\vec{x} = \sum_{i=1}^{m} \theta_i \vec{y}_i$$

$$\begin{aligned}
\left|\left|\vec{x}\right|\right|_1 &\leq \sum_{i=1}^{m} \theta_i \left|\left|\vec{y}_i\right|\right|_1 \quad \text{by the Triangle inequality} \\
&\leq \sum_{i=1}^{m} \theta_i \left|\left|\vec{y}_i\right|\right|_\infty \quad \vec{y}_i \text{ only has one nonzero entry} \\
&\leq \sum_{i=1}^{m} \theta_i
\end{aligned}$$

$\mathcal{C}\left(\mathcal{B}_{\ell_0} \cap \mathcal{B}_{\ell_\infty}\right) \subseteq \mathcal{B}_{\ell_1}$

Let $\vec{x} \in \mathcal{C}\left(\mathcal{B}_{\ell_0} \cap \mathcal{B}_{\ell_\infty}\right)$, then

$$\vec{x} = \sum_{i=1}^{m} \theta_i \vec{y}_i$$

$$\begin{aligned}
||\vec{x}||_1 &\leq \sum_{i=1}^{m} \theta_i \, ||\vec{y}_i||_1 && \text{by the Triangle inequality} \\
&\leq \sum_{i=1}^{m} \theta_i \, ||\vec{y}_i||_\infty && \vec{y}_i \text{ only has one nonzero entry} \\
&\leq \sum_{i=1}^{m} \theta_i \\
&\leq 1
\end{aligned}$$

Convex optimization problem

$$f_0, f_1, \ldots, f_m, h_1, \ldots, h_p : \mathbb{R}^n \to \mathbb{R}$$

$$\begin{aligned} \text{minimize} \quad & f_0(\vec{x}) \\ \text{subject to} \quad & f_i(\vec{x}) \leq 0, \quad 1 \leq i \leq m, \\ & h_i(\vec{x}) = 0, \quad 1 \leq i \leq p, \end{aligned}$$

Definitions

- A feasible vector is a vector that satisfies all the constraints
- A solution is any vector $\vec{x}^*$ such that for all feasible vectors $\vec{x}$

$$f_0(\vec{x}) \geq f_0(\vec{x}^*)$$

- If a solution exists $f(\vec{x}^*)$ is the optimal value or optimum of the problem

Convex optimization problem

The optimization problem is convex if

- f_0 is convex
- $f_1, \ldots, f_m$ are convex
- $h_1, \ldots, h_p$ are affine, i.e. $h_i(\vec{x}) = \vec{a}_i^T \vec{x} + b_i$ for some $\vec{a}_i \in \mathbb{R}^n$ and $b_i \in \mathbb{R}$

Linear program

$$
\begin{aligned}
\text{minimize} \quad & \vec{a}^T \vec{x} \\
\text{subject to} \quad & \vec{c_i}^T \vec{x} \leq d_i, \quad 1 \leq i \leq m \\
& A\vec{x} = \vec{b}
\end{aligned}
$$

ℓ_1-norm minimization as an LP

The optimization problem

$$\begin{aligned} \text{minimize} \quad & ||\vec{x}||_1 \\ \text{subject to} \quad & A\vec{x} = \vec{b} \end{aligned}$$

can be recast as the LP

$$\begin{aligned} \text{minimize} \quad & \sum_{i=1}^{m} \vec{t}[i] \\ \text{subject to} \quad & \vec{t}[i] \geq \vec{e_i}^{\,T} \vec{x} \\ & \vec{t}[i] \geq -\vec{e_i}^{\,T} \vec{x} \\ & A\vec{x} = \vec{b} \end{aligned}$$

Proof

Solution to ℓ_1-norm min. problem: $\vec{x}^{\ell_1}$

Solution to linear program: $\left(\vec{x}^{\text{lp}}, \vec{t}^{\text{lp}}\right)$

Set $\vec{t}^{\ell_1}[i] := \left|\vec{x}^{\ell_1}[i]\right|$

$\left(\vec{x}^{\ell_1}, \vec{t}^{\ell_1}\right)$ is feasible for linear program

$$\left|\left|\vec{x}^{\ell_1}\right|\right|_1 = \sum_{i=1}^{m} \vec{t}^{\ell_1}[i]$$

Proof

Solution to ℓ_1-norm min. problem: $\vec{x}^{\,\ell_1}$

Solution to linear program: $\left(\vec{x}^{\,\text{lp}}, \vec{t}^{\,\text{lp}}\right)$

Set $\vec{t}^{\,\ell_1}[i] := \left|\vec{x}^{\,\ell_1}[i]\right|$

$\left(\vec{x}^{\,\ell_1}, \vec{t}^{\,\ell_1}\right)$ is feasible for linear program

$$\begin{aligned} \left|\left|\vec{x}^{\,\ell_1}\right|\right|_1 &= \sum_{i=1}^{m} \vec{t}^{\,\ell_1}[i] \\ &\geq \sum_{i=1}^{m} \vec{t}^{\,\text{lp}}[i] \quad \text{by optimality of } \vec{t}^{\,\text{lp}} \end{aligned}$$

Proof

Solution to ℓ_1-norm min. problem: $\vec{x}^{\,\ell_1}$

Solution to linear program: $\left(\vec{x}^{\,\text{lp}}, \vec{t}^{\,\text{lp}}\right)$

Set $\vec{t}^{\,\ell_1}[i] := \left|\vec{x}^{\,\ell_1}[i]\right|$

$\left(\vec{x}^{\,\ell_1}, \vec{t}^{\,\ell_1}\right)$ is feasible for linear program

$$\begin{aligned}
\left|\left|\vec{x}^{\,\ell_1}\right|\right|_1 &= \sum_{i=1}^{m} \vec{t}^{\,\ell_1}[i] \\
&\geq \sum_{i=1}^{m} \vec{t}^{\,\text{lp}}[i] \quad \text{by optimality of } \vec{t}^{\,\text{lp}} \\
&\geq \left|\left|\vec{x}^{\,\text{lp}}\right|\right|_1
\end{aligned}$$

Proof

Solution to ℓ_1-norm min. problem: $\vec{x}^{\ell_1}$

Solution to linear program: $\left(\vec{x}^{\text{lp}}, \vec{t}^{\text{lp}}\right)$

Set $\vec{t}^{\ell_1}[i] := \left|\vec{x}^{\ell_1}[i]\right|$

$\left(\vec{x}^{\ell_1}, \vec{t}^{\ell_1}\right)$ is feasible for linear program

$$\begin{aligned}
\left|\left|\vec{x}^{\ell_1}\right|\right|_1 &= \sum_{i=1}^{m} \vec{t}^{\ell_1}[i] \\
&\geq \sum_{i=1}^{m} \vec{t}^{\text{lp}}[i] \quad \text{by optimality of } \vec{t}^{\text{lp}} \\
&\geq \left|\left|\vec{x}^{\text{lp}}\right|\right|_1
\end{aligned}$$

$\vec{x}^{\text{lp}}$ is a solution to the ℓ_1-norm min. problem

Proof

Set $\vec{t}^{\ell_1}[i] := \left|\vec{x}^{\ell_1}[i]\right|$

$$\sum_{i=1}^{m} t_i^{\ell_1} = \left|\left|\vec{x}^{\ell_1}\right|\right|_1$$

Proof

Set $\vec{t}^{\,\ell_1}[i] := \left|\vec{x}^{\,\ell_1}[i]\right|$

$$\begin{aligned}\sum_{i=1}^{m} t_i^{\ell_1} &= \left|\left|\vec{x}^{\,\ell_1}\right|\right|_1 \\ &\leq \left|\left|\vec{x}^{\,\text{lp}}\right|\right|_1 \quad \text{by optimality of } \vec{x}^{\,\ell_1}\end{aligned}$$

Proof

Set $\vec{t}^{\,\ell_1}[i] := \left|\vec{x}^{\,\ell_1}[i]\right|$

$$\begin{aligned} \sum_{i=1}^{m} t_i^{\ell_1} &= \left|\left|\vec{x}^{\,\ell_1}\right|\right|_1 \\ &\leq \left|\left|\vec{x}^{\,\text{lp}}\right|\right|_1 \quad \text{by optimality of } \vec{x}^{\,\ell_1} \\ &\leq \sum_{i=1}^{m} \vec{t}^{\,\text{lp}}[i] \end{aligned}$$

Proof

Set $\vec{t}^{\ell_1}[i] := \left|\vec{x}^{\ell_1}[i]\right|$

$$\begin{aligned} \sum_{i=1}^{m} t_i^{\ell_1} &= \left|\left|\vec{x}^{\ell_1}\right|\right|_1 \\ &\leq \left|\left|\vec{x}^{\text{lp}}\right|\right|_1 \quad \text{by optimality of } \vec{x}^{\ell_1} \\ &\leq \sum_{i=1}^{m} \vec{t}^{\text{lp}}[i] \end{aligned}$$

$\left(\vec{x}^{\ell_1}, \vec{t}^{\ell_1}\right)$ is a solution to the linear problem

Quadratic program

For a positive semidefinite matrix $Q \in \mathbb{R}^{n \times n}$

$$\begin{aligned} \text{minimize} \quad & \vec{x}^T Q \vec{x} + \vec{a}^T \vec{x} \\ \text{subject to} \quad & \vec{c_i}^T \vec{x} \leq d_i, \quad 1 \leq i \leq m, \\ & A\vec{x} = \vec{b} \end{aligned}$$

ℓ_1-norm regularized least squares as a QP

The optimization problem

$$\text{minimize} \quad ||A\vec{x} - y||_2^2 + \vec{\alpha}\,||\vec{x}||_1$$

can be recast as the QP

$$\begin{aligned} \text{minimize} \quad & \vec{x}^T A^T A\vec{x} - 2\vec{y}^T\vec{x} + \vec{\alpha}\sum_{i=1}^{n} \vec{t}[i] \\ \text{subject to} \quad & \vec{t}[i] \geq \vec{e_i}^{\,T}\vec{x} \\ & \vec{t}[i] \geq -\vec{e_i}^{\,T}\vec{x} \end{aligned}$$

Lagrangian

The Lagrangian of a canonical optimization problem is

$$L(\vec{x}, \vec{\alpha}, \vec{\nu}) := f_0(\vec{x}) + \sum_{i=1}^{m} \vec{\alpha}[i]\, f_i(\vec{x}) + \sum_{j=1}^{p} \vec{\nu}[j]\, h_j(\vec{x}),$$

$\vec{\alpha} \in \mathbb{R}^m, \vec{\nu} \in \mathbb{R}^p$ are called Lagrange multipliers or dual variables

If $\vec{x}$ is feasible and $\vec{\alpha}[i] \geq 0$ for $1 \leq i \leq m$

$$L(\vec{x}, \vec{\alpha}, \vec{\nu}) \leq f_0(\vec{x})$$

Lagrange dual function

The Lagrange dual function of the problem is

$$l(\vec{\alpha}, \vec{\nu}) := \inf_{\vec{x} \in \mathbb{R}^n} f_0(\vec{x}) + \sum_{i=1}^{m} \vec{\alpha}[i] f_i(\vec{x}) + \sum_{j=1}^{p} \vec{\nu}[j] h_j(\vec{x})$$

Let p^* be an optimum of the optimization problem

$$l(\vec{\alpha}, \vec{\nu}) \leq p^*$$

as long as $\vec{\alpha}[i] \geq 0$ for $1 \leq i \leq n$

Dual problem

The dual problem of the (primal) optimization problem is

$$\begin{aligned} \text{maximize} \quad & l\left(\vec{\alpha}, \vec{\nu}\right) \\ \text{subject to} \quad & \vec{\alpha}[i] \geq 0, \quad 1 \leq i \leq m. \end{aligned}$$

The dual problem is always convex, even if the primal isn't!

Maximum/supremum of convex functions

Pointwise maximum of m convex functions $f_1, \ldots, f_m$

$$f_{\max}(x) := \max_{1 \le i \le m} f_i(x)$$

is convex

Pointwise supremum of a family of convex functions indexed by a set $\mathcal{I}$

$$f_{\sup}(x) := \sup_{i \in \mathcal{I}} f_i(x)$$

is convex

Proof

For any $0 \leq \theta \leq 1$ and any $\vec{x}, \vec{y} \in \mathbb{R}$,

$$f_{\text{sup}}\left(\theta\vec{x} + (1-\theta)\,\vec{y}\right) = \sup_{i \in \mathcal{I}} f_i\left(\theta\vec{x} + (1-\theta)\,\vec{y}\right)$$

Proof

For any $0 \leq \theta \leq 1$ and any $\vec{x}, \vec{y} \in \mathbb{R}$,

$$\begin{aligned} f_{\sup}\left(\theta\vec{x} + (1-\theta)\,\vec{y}\right) &= \sup_{i\in\mathcal{I}} f_i\left(\theta\vec{x} + (1-\theta)\,\vec{y}\right) \\ &\leq \sup_{i\in\mathcal{I}} \theta f_i\left(\vec{x}\right) + (1-\theta)\, f_i\left(\vec{y}\right) \quad \text{by convexity of the } f_i \end{aligned}$$

Proof

For any $0 \leq \theta \leq 1$ and any $\vec{x}, \vec{y} \in \mathbb{R}$,

$$\begin{aligned} f_{\text{sup}}\left(\theta\vec{x} + (1-\theta)\,\vec{y}\right) &= \sup_{i \in \mathcal{I}} f_i\left(\theta\vec{x} + (1-\theta)\,\vec{y}\right) \\ &\leq \sup_{i \in \mathcal{I}} \theta f_i\left(\vec{x}\right) + (1-\theta)\, f_i\left(\vec{y}\right) \quad \text{by convexity of the } f_i \\ &\leq \theta \sup_{i \in \mathcal{I}} f_i\left(\vec{x}\right) + (1-\theta) \sup_{j \in \mathcal{I}} f_j\left(\vec{y}\right) \end{aligned}$$

Proof

For any $0 \leq \theta \leq 1$ and any $\vec{x}, \vec{y} \in \mathbb{R}$,

$$\begin{aligned}
f_{\text{sup}}\left(\theta\vec{x} + (1-\theta)\,\vec{y}\right) &= \sup_{i \in \mathcal{I}} f_i\left(\theta\vec{x} + (1-\theta)\,\vec{y}\right) \\
&\leq \sup_{i \in \mathcal{I}} \theta f_i\left(\vec{x}\right) + (1-\theta)\, f_i\left(\vec{y}\right) \quad \text{by convexity of the } f_i \\
&\leq \theta \sup_{i \in \mathcal{I}} f_i\left(\vec{x}\right) + (1-\theta) \sup_{j \in \mathcal{I}} f_j\left(\vec{y}\right) \\
&= \theta f_{\text{sup}}\left(\vec{x}\right) + (1-\theta)\, f_{\text{sup}}\left(\vec{y}\right)
\end{aligned}$$

Weak duality

If p^* is a primal optimum and d^* a dual optimum

$$d^* \leq p^*$$

Strong duality

For convex problems

$$d^* = p^*$$

under very weak conditions

LPs: The primal optimum is finite

General convex programs (Slater's condition):

There exists a point that is strictly feasible

$$f_i(\vec{x}) < 0 \quad 1 \leq i \leq m$$

ℓ_1-norm minimization

The dual problem of

$$\min_{\vec{x}} ||\vec{x}||_1 \qquad \text{subject to} \quad A\vec{x} = \vec{y}$$

is

$$\max_{\vec{\nu}} \vec{y}^T \vec{\nu} \qquad \text{subject to} \quad \left|\left|A^T \vec{\nu}\right|\right|_\infty \leq 1$$

Proof

Lagrangian $L(\vec{x}, \vec{\nu}) = ||\vec{x}||_1 + \vec{\nu}^T (\vec{y} - A\vec{x})$

Lagrange dual function

$$l(\vec{\alpha}, \vec{\nu}) := \inf_{\vec{x} \in \mathbb{R}^n} ||\vec{x}||_1 - (A^T \vec{\nu})^T \vec{x} + \vec{\nu}^T \vec{y}$$

Proof

Lagrangian $L(\vec{x}, \vec{\nu}) = ||\vec{x}||_1 + \vec{\nu}^T (\vec{y} - A\vec{x})$

Lagrange dual function

$$l(\vec{\alpha}, \vec{\nu}) := \inf_{\vec{x} \in \mathbb{R}^n} ||\vec{x}||_1 - (A^T \vec{\nu})^T \vec{x} + \vec{\nu}^T \vec{y}$$

If $A^T \vec{\nu}[i] > 1$?

Proof

Lagrangian $L(\vec{x}, \vec{\nu}) = ||\vec{x}||_1 + \vec{\nu}^T(\vec{y} - A\vec{x})$

Lagrange dual function

$$l(\vec{\alpha}, \vec{\nu}) := \inf_{\vec{x} \in \mathbb{R}^n} ||\vec{x}||_1 - (A^T\vec{\nu})^T\vec{x} + \vec{\nu}^T\vec{y}$$

If $A^T\vec{\nu}[i] > 1$? We can set $\vec{x}[i] \to \infty$ and $l(\vec{\alpha}, \vec{\nu}) \to -\infty$

Proof

Lagrangian $L(\vec{x}, \vec{\nu}) = ||\vec{x}||_1 + \vec{\nu}^T (\vec{y} - A\vec{x})$

Lagrange dual function

$$l(\vec{\alpha}, \vec{\nu}) := \inf_{\vec{x} \in \mathbb{R}^n} ||\vec{x}||_1 - (A^T \vec{\nu})^T \vec{x} + \vec{\nu}^T \vec{y}$$

If $A^T \vec{\nu}[i] > 1$? We can set $\vec{x}[i] \to \infty$ and $l(\vec{\alpha}, \vec{\nu}) \to -\infty$

If $\left|\left|A^T \vec{\nu}\right|\right|_\infty \leq 1$?

$$(A^T \vec{\nu})^T \vec{x}$$

Proof

Lagrangian $L(\vec{x}, \vec{\nu}) = ||\vec{x}||_1 + \vec{\nu}^T (\vec{y} - A\vec{x})$

Lagrange dual function

$$l(\vec{\alpha}, \vec{\nu}) := \inf_{\vec{x} \in \mathbb{R}^n} ||\vec{x}||_1 - (A^T \vec{\nu})^T \vec{x} + \vec{\nu}^T \vec{y}$$

If $A^T \vec{\nu}[i] > 1$? We can set $\vec{x}[i] \to \infty$ and $l(\vec{\alpha}, \vec{\nu}) \to -\infty$

If $\left|\left|A^T \vec{\nu}\right|\right|_\infty \leq 1$?

$$(A^T \vec{\nu})^T \vec{x} \leq ||\vec{x}||_1 \left|\left|A^T \vec{\nu}\right|\right|_\infty \leq ||\vec{x}||_1$$

Proof

Lagrangian $L(\vec{x}, \vec{\nu}) = ||\vec{x}||_1 + \vec{\nu}^T (\vec{y} - A\vec{x})$

Lagrange dual function

$$l(\vec{\alpha}, \vec{\nu}) := \inf_{\vec{x} \in \mathbb{R}^n} ||\vec{x}||_1 - (A^T \vec{\nu})^T \vec{x} + \vec{\nu}^T \vec{y}$$

If $A^T \vec{\nu}[i] > 1$? We can set $\vec{x}[i] \to \infty$ and $l(\vec{\alpha}, \vec{\nu}) \to -\infty$

If $\left|\left|A^T \vec{\nu}\right|\right|_\infty \leq 1$?

$$(A^T \vec{\nu})^T \vec{x} \leq ||\vec{x}||_1 \left|\left|A^T \vec{\nu}\right|\right|_\infty \leq ||\vec{x}||_1$$

so $l(\vec{\alpha}, \vec{\nu}) = \vec{\nu}^T \vec{y}$

Strong duality

The solution $\vec{\nu}^*$ to

$$\max_{\vec{\nu}} \vec{y}^T \vec{\nu} \qquad \text{subject to} \quad \left|\left|A^T \vec{\nu}\right|\right|_\infty \leq 1$$

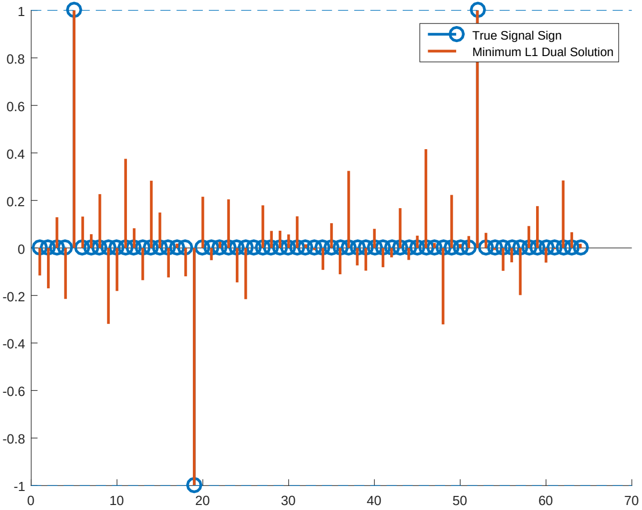
satisfies

$$(A^T \vec{\nu}^*)[i] = \text{sign}(\vec{x}^*[i]) \quad \text{for all } \vec{x}^*[i] \neq 0$$

for all solutions $\vec{x}^*$ to the primal problem

$$\min_{\vec{x}} ||\vec{x}||_1 \qquad \text{subject to} \quad A\vec{x} = \vec{y}$$

Dual solution



Proof

By strong duality

$$
\begin{aligned}
||\vec{x}^*||_1 &= \vec{y}^T \vec{\nu}^* \\
&= (A\vec{x}^*)^T \vec{\nu}^* \\
&= (\vec{x}^*)^T (A^T \vec{\nu}^*) \\
&= \sum_{i=1}^{m} (A^T \vec{\nu}^*)[i] \vec{x}^*[i]
\end{aligned}
$$

By Hölder's inequality

$$
||\vec{x}^*||_1 \geq \sum_{i=1}^{m} (A^T \vec{\nu}^*)[i] \vec{x}^*[i]
$$

with equality if and only if

$$
(A^T \vec{\nu}^*)[i] = \operatorname{sign}(\vec{x}^*[i]) \quad \text{for all } \vec{x}^*[i] \neq 0
$$

Another algorithm for sparse recovery

Aim: Find nonzero locations of a sparse vector $\vec{x}$ from $\vec{y} = A\vec{x}$

Insight: We have access to inner products of $\vec{x}$ and $A^T\vec{w}$ for any $\vec{w}$

$$\vec{y}^T\vec{w}$$

Another algorithm for sparse recovery

Aim: Find nonzero locations of a sparse vector $\vec{x}$ from $\vec{y} = A\vec{x}$

Insight: We have access to inner products of $\vec{x}$ and $A^T\vec{w}$ for any $\vec{w}$

$$\vec{y}^T\vec{w} = (A\vec{x})^T\vec{w}$$

Another algorithm for sparse recovery

Aim: Find nonzero locations of a sparse vector $\vec{x}$ from $\vec{y} = A\vec{x}$

Insight: We have access to inner products of $\vec{x}$ and $A^T\vec{w}$ for any $\vec{w}$

$$\begin{aligned} \vec{y}^T\vec{w} &= (A\vec{x})^T\vec{w} \\ &= \vec{x}^T(A^T\vec{w}) \end{aligned}$$

Another algorithm for sparse recovery

Aim: Find nonzero locations of a sparse vector $\vec{x}$ from $\vec{y} = A\vec{x}$

Insight: We have access to inner products of $\vec{x}$ and $A^T\vec{w}$ for any $\vec{w}$

$$\begin{aligned} \vec{y}^T\vec{w} &= (A\vec{x})^T\vec{w} \\ &= \vec{x}^T(A^T\vec{w}) \end{aligned}$$

Idea: Maximize $A^T\vec{w}$, bounding magnitude of entries by 1

Another algorithm for sparse recovery

Aim: Find nonzero locations of a sparse vector $\vec{x}$ from $\vec{y} = A\vec{x}$

Insight: We have access to inner products of $\vec{x}$ and $A^T\vec{w}$ for any $\vec{w}$

$$\begin{aligned}\vec{y}^T\vec{w} &= (A\vec{x})^T\vec{w} \\ &= \vec{x}^T(A^T\vec{w})\end{aligned}$$

Idea: Maximize $A^T\vec{w}$, bounding magnitude of entries by 1

Entries where $\vec{x}$ is nonzero should saturate to 1 or -1

Compressed sensing

Convex constrained problems

Analyzing optimization-based methods

Analyzing optimization-based methods

Best case scenario: Primal solution has closed form

Otherwise: Use dual solution to characterize primal solution

Minimum ℓ_2-norm solution

Let $A \in \mathbb{R}^{m \times n}$ be a full rank matrix such that $m < n$

For any $\vec{y} \in \mathbb{R}^n$ the solution to the optimization problem

$$\arg\min_{\vec{x}} ||\vec{x}||_2 \qquad \text{subject to} \qquad A\vec{x} = \vec{y}.$$

is

$$\begin{aligned} \vec{x}^* &:= VS^{-1}U^T\vec{y} \\ &= A^T\left(A^TA\right)^{-1}\vec{y} \end{aligned}$$

where $A = USV^T$ is the SVD of A

Proof

$$\vec{x} = \mathcal{P}_{\text{row}(A)}\,\vec{x} + \mathcal{P}_{\text{row}(A)^{\perp}}\,\vec{x}$$

Since A is full rank V, $\mathcal{P}_{\text{row}(A)}\,\vec{x} = V\vec{c}$ for some vector $\vec{c} \in \mathbb{R}^n$

$$A\vec{x} = A\mathcal{P}_{\text{row}(A)}\,\vec{x}$$

Proof

$$\vec{x} = \mathcal{P}_{\text{row}(A)}\,\vec{x} + \mathcal{P}_{\text{row}(A)^\perp}\,\vec{x}$$

Since A is full rank V, $\mathcal{P}_{\text{row}(A)}\,\vec{x} = V\vec{c}$ for some vector $\vec{c} \in \mathbb{R}^n$

$$\begin{aligned} A\vec{x} &= A\mathcal{P}_{\text{row}(A)}\,\vec{x} \\ &= USV^T V\vec{c} \end{aligned}$$

Proof

$$\vec{x} = \mathcal{P}_{\text{row}(A)}\,\vec{x} + \mathcal{P}_{\text{row}(A)^{\perp}}\,\vec{x}$$

Since A is full rank V, $\mathcal{P}_{\text{row}(A)}\,\vec{x} = V\vec{c}$ for some vector $\vec{c} \in \mathbb{R}^n$

$$\begin{aligned} A\vec{x} &= A\mathcal{P}_{\text{row}(A)}\,\vec{x} \\ &= USV^T V\vec{c} \\ &= US\vec{c} \end{aligned}$$

Proof

$$\vec{x} = \mathcal{P}_{\text{row}(A)}\,\vec{x} + \mathcal{P}_{\text{row}(A)^\perp}\,\vec{x}$$

Since A is full rank V, $\mathcal{P}_{\text{row}(A)}\,\vec{x} = V\vec{c}$ for some vector $\vec{c} \in \mathbb{R}^n$

$$\begin{aligned} A\vec{x} &= A\mathcal{P}_{\text{row}(A)}\,\vec{x} \\ &= USV^T V\vec{c} \\ &= US\vec{c} \end{aligned}$$

$A\vec{x} = \vec{y}$ is equivalent to $US\vec{c} = \vec{y}$ and $\vec{c} = S^{-1}U^T\vec{y}$

Proof

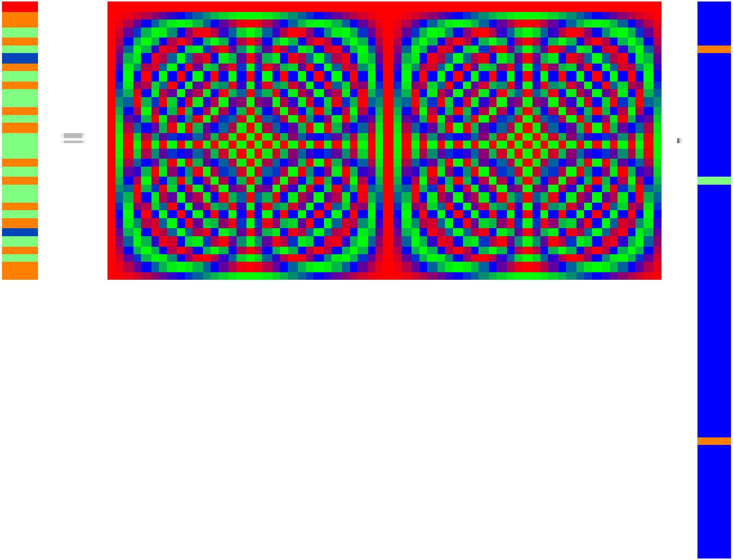
For all feasible vectors $\vec{x}$

$$\mathcal{P}_{\text{row}(A)}\,\vec{x} = VS^{-1}U^{T}\vec{y}$$

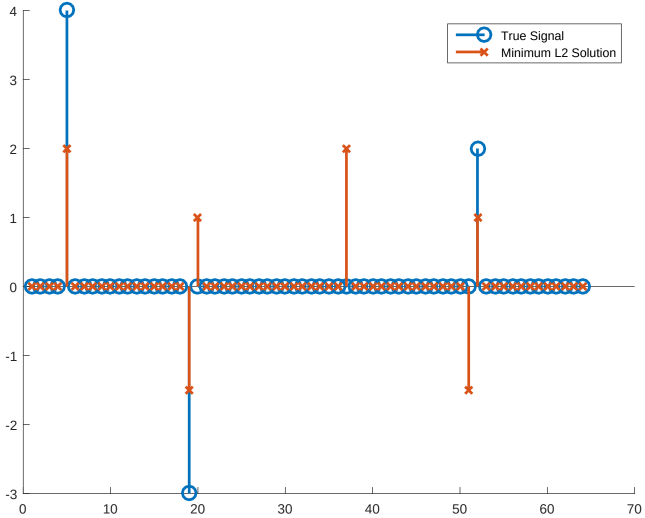
By Pythagoras' theorem, minimizing $||\vec{x}||_2$ is equivalent to minimizing

$$||\vec{x}||_2^2 = \left|\left|\mathcal{P}_{\text{row}(A)}\,\vec{x}\right|\right|_2^2 + \left|\left|\mathcal{P}_{\text{row}(A)^{\perp}}\,\vec{x}\right|\right|_2^2$$

Regular subsampling



Minimum ℓ_2-norm solution (regular subsampling)



Regular subsampling

$$A := \frac{1}{\sqrt{2}} \begin{bmatrix} F_{m/2} & F_{m/2} \end{bmatrix}$$

$$F_{m/2}^{*} F_{m/2} = I$$

$$F_{m/2} F_{m/2}^{*} = I$$

$$\vec{x} := \begin{bmatrix} \vec{x}_{\text{up}} \\ \vec{x}_{\text{down}} \end{bmatrix}$$

Regular subsampling

$$\vec{x}_{\ell_2} = \arg\min_{A\vec{x}=\vec{y}} ||\vec{x}||_2$$

Regular subsampling

$$
\begin{aligned}
\vec{x}_{\ell_2} &= \arg\min_{A\vec{x}=\vec{y}} ||\vec{x}||_2 \\
&= A^T \left(A^T A\right)^{-1} \vec{y}
\end{aligned}
$$

Regular subsampling

$$\vec{x}_{\ell_2} = \arg \min_{A\vec{x}=\vec{y}} ||\vec{x}||_2$$

$$= A^T \left(A^T A\right)^{-1} \vec{y}$$

$$= \frac{1}{\sqrt{2}} \begin{bmatrix} F^*_{m/2} \\ F^*_{m/2} \end{bmatrix} \left(\frac{1}{\sqrt{2}} \begin{bmatrix} F_{m/2} & F_{m/2} \end{bmatrix} \frac{1}{\sqrt{2}} \begin{bmatrix} F^*_{m/2} \\ F^*_{m/2} \end{bmatrix} \right)^{-1} \frac{1}{\sqrt{2}} \begin{bmatrix} F_{m/2} & F_{m/2} \end{bmatrix} \begin{bmatrix} \vec{x}_{\text{up}} \\ \vec{x}_{\text{down}} \end{bmatrix}$$

Regular subsampling

$$
\begin{aligned}
\vec{x}_{\ell_2} &= \arg \min_{A\vec{x}=\vec{y}} ||\vec{x}||_2 \\
&= A^T \left(A^T A\right)^{-1} \vec{y} \\
&= \frac{1}{\sqrt{2}} \begin{bmatrix} F_{m/2}^* \\ F_{m/2}^* \end{bmatrix} \left(\frac{1}{\sqrt{2}} \begin{bmatrix} F_{m/2} & F_{m/2} \end{bmatrix} \frac{1}{\sqrt{2}} \begin{bmatrix} F_{m/2}^* \\ F_{m/2}^* \end{bmatrix} \right)^{-1} \frac{1}{\sqrt{2}} \begin{bmatrix} F_{m/2} & F_{m/2} \end{bmatrix} \begin{bmatrix} \vec{x}_{\text{up}} \\ \vec{x}_{\text{down}} \end{bmatrix} \\
&= \frac{1}{2} \begin{bmatrix} F_{m/2}^* \\ F_{m/2}^* \end{bmatrix} \left(\frac{1}{2} \left[F_{m/2} F_{m/2}^* + F_{m/2} F_{m/2}^* \right] \right)^{-1} \left(F_{m/2} \vec{x}_{\text{up}} + F_{m/2} \vec{x}_{\text{down}} \right)
\end{aligned}
$$

Regular subsampling

$$
\begin{aligned}
\vec{x}_{\ell_2} &= \arg \min_{A\vec{x}=\vec{y}} ||\vec{x}||_2 \\
&= A^T \left(A^T A\right)^{-1} \vec{y} \\
&= \frac{1}{\sqrt{2}} \begin{bmatrix} F^*_{m/2} \\ F^*_{m/2} \end{bmatrix} \left(\frac{1}{\sqrt{2}} \begin{bmatrix} F_{m/2} & F_{m/2} \end{bmatrix} \frac{1}{\sqrt{2}} \begin{bmatrix} F^*_{m/2} \\ F^*_{m/2} \end{bmatrix} \right)^{-1} \frac{1}{\sqrt{2}} \begin{bmatrix} F_{m/2} & F_{m/2} \end{bmatrix} \begin{bmatrix} \vec{x}_{\text{up}} \\ \vec{x}_{\text{down}} \end{bmatrix} \\
&= \frac{1}{2} \begin{bmatrix} F^*_{m/2} \\ F^*_{m/2} \end{bmatrix} \left(\frac{1}{2} \left[F_{m/2} F^*_{m/2} + F_{m/2} F^*_{m/2} \right] \right)^{-1} \left(F_{m/2} \vec{x}_{\text{up}} + F_{m/2} \vec{x}_{\text{down}} \right) \\
&= \frac{1}{2} \begin{bmatrix} F^*_{m/2} \\ F^*_{m/2} \end{bmatrix} I^{-1} \left(F_{m/2} \vec{x}_{\text{up}} + F_{m/2} \vec{x}_{\text{down}} \right)
\end{aligned}
$$

Regular subsampling

$$
\begin{aligned}
\vec{x}_{\ell_2} &= \arg\min_{A\vec{x}=\vec{y}} ||\vec{x}||_2 \\
&= A^T \left(A^T A\right)^{-1} \vec{y} \\
&= \frac{1}{\sqrt{2}} \begin{bmatrix} F_{m/2}^* \\ F_{m/2}^* \end{bmatrix} \left(\frac{1}{\sqrt{2}} \begin{bmatrix} F_{m/2} & F_{m/2} \end{bmatrix} \frac{1}{\sqrt{2}} \begin{bmatrix} F_{m/2}^* \\ F_{m/2}^* \end{bmatrix} \right)^{-1} \frac{1}{\sqrt{2}} \begin{bmatrix} F_{m/2} & F_{m/2} \end{bmatrix} \begin{bmatrix} \vec{x}_{\text{up}} \\ \vec{x}_{\text{down}} \end{bmatrix} \\
&= \frac{1}{2} \begin{bmatrix} F_{m/2}^* \\ F_{m/2}^* \end{bmatrix} \left(\frac{1}{2} \left[F_{m/2} F_{m/2}^* + F_{m/2} F_{m/2}^* \right] \right)^{-1} \left(F_{m/2} \vec{x}_{\text{up}} + F_{m/2} \vec{x}_{\text{down}} \right) \\
&= \frac{1}{2} \begin{bmatrix} F_{m/2}^* \\ F_{m/2}^* \end{bmatrix} I^{-1} \left(F_{m/2} \vec{x}_{\text{up}} + F_{m/2} \vec{x}_{\text{down}} \right) \\
&= \frac{1}{2} \begin{bmatrix} F_{m/2}^* \left(F_{m/2} \vec{x}_{\text{up}} + F_{m/2} \vec{x}_{\text{down}} \right) \\ F_{m/2}^* \left(F_{m/2} \vec{x}_{\text{up}} + F_{m/2} \vec{x}_{\text{down}} \right) \end{bmatrix}
\end{aligned}
$$

Regular subsampling

$$
\begin{aligned}
\vec{x}_{\ell_2} &= \arg\min_{A\vec{x}=\vec{y}} ||\vec{x}||_2 \\
&= A^T \left(A^T A\right)^{-1} \vec{y} \\
&= \frac{1}{\sqrt{2}} \begin{bmatrix} F^*_{m/2} \\ F^*_{m/2} \end{bmatrix} \left(\frac{1}{\sqrt{2}} \begin{bmatrix} F_{m/2} & F_{m/2} \end{bmatrix} \frac{1}{\sqrt{2}} \begin{bmatrix} F^*_{m/2} \\ F^*_{m/2} \end{bmatrix} \right)^{-1} \frac{1}{\sqrt{2}} \begin{bmatrix} F_{m/2} & F_{m/2} \end{bmatrix} \begin{bmatrix} \vec{x}_{\text{up}} \\ \vec{x}_{\text{down}} \end{bmatrix} \\
&= \frac{1}{2} \begin{bmatrix} F^*_{m/2} \\ F^*_{m/2} \end{bmatrix} \left(\frac{1}{2} \left[F_{m/2} F^*_{m/2} + F_{m/2} F^*_{m/2} \right] \right)^{-1} \left(F_{m/2} \vec{x}_{\text{up}} + F_{m/2} \vec{x}_{\text{down}} \right) \\
&= \frac{1}{2} \begin{bmatrix} F^*_{m/2} \\ F^*_{m/2} \end{bmatrix} I^{-1} \left(F_{m/2} \vec{x}_{\text{up}} + F_{m/2} \vec{x}_{\text{down}} \right) \\
&= \frac{1}{2} \begin{bmatrix} F^*_{m/2} \left(F_{m/2} \vec{x}_{\text{up}} + F_{m/2} \vec{x}_{\text{down}} \right) \\ F^*_{m/2} \left(F_{m/2} \vec{x}_{\text{up}} + F_{m/2} \vec{x}_{\text{down}} \right) \end{bmatrix} \\
&= \frac{1}{2} \begin{bmatrix} \vec{x}_{\text{up}} + \vec{x}_{\text{down}} \\ \vec{x}_{\text{up}} + \vec{x}_{\text{down}} \end{bmatrix}
\end{aligned}
$$

Minimum ℓ_1-norm solution

Problem: $\arg\min_{A\vec{x}=\vec{y}} ||\vec{x}||_1$ doesn't have a closed form

Instead we can use a dual variable to certify optimality

Dual solution

The solution $\vec{\nu}^*$ to

$$\max_{\vec{\nu}} \vec{y}^T \vec{\nu} \qquad \text{subject to} \quad \left|\left|A^T \vec{\nu}\right|\right|_\infty \leq 1$$

satisfies

$$(A^T \vec{\nu}^*)[i] = \text{sign}(\vec{x}^*[i]) \quad \text{for all } \vec{x}^*[i] \neq 0$$

where $\vec{x}^*[i]$ is a solution to the primal problem

$$\min_{\vec{x}} ||\vec{x}||_1 \qquad \text{subject to} \quad A\vec{x} = \vec{y}$$

Dual certificate

If there exists a vector $\vec{\nu} \in \mathbb{R}^n$ such that

$$\begin{aligned} (A^T\vec{\nu})[i] &= \operatorname{sign}(\vec{x}^*[i]) \qquad \text{if } \vec{x}^*[i] \neq 0 \\ \left|(A^T\vec{\nu})[i]\right| &< 1 \qquad \text{if } \vec{x}^*[i] = 0 \end{aligned}$$

then $\vec{x}^*$ is the unique solution to the primal problem

$$\min_{\vec{x}} ||\vec{x}||_1 \qquad \text{subject to} \quad A\vec{x} = \vec{y}$$

as long as the submatrix A_T is full rank

Proof 1

$\vec{\nu}$ is feasible for the dual problem, so for any primal feasible $\vec{x}$

$$||\vec{x}||_1 \geq \vec{y}^T \vec{\nu}$$

Proof 1

$\vec{\nu}$ is feasible for the dual problem, so for any primal feasible $\vec{x}$

$$||\vec{x}||_1 \geq \vec{y}^T \vec{\nu} = (A\vec{x}^*)^T \vec{\nu}$$

Proof 1

$\vec{\nu}$ is feasible for the dual problem, so for any primal feasible $\vec{x}$

$$\begin{aligned} ||\vec{x}||_1 \geq \vec{y}^T\vec{\nu} &= (A\vec{x}^*)^T\vec{\nu} \\ &= (\vec{x}^*)^T(A^T\vec{\nu}) \end{aligned}$$

Proof 1

$\vec{\nu}$ is feasible for the dual problem, so for any primal feasible $\vec{x}$

$$\begin{aligned}||\vec{x}||_1 \geq \vec{y}^T \vec{\nu} &= (A\vec{x}^*)^T \vec{\nu} \\ &= (\vec{x}^*)^T (A^T \vec{\nu}) \\ &= \sum_{i \in T} \vec{x}^*[i] \operatorname{sign}(\vec{x}^*[i])\end{aligned}$$

Proof 1

$\vec{\nu}$ is feasible for the dual problem, so for any primal feasible $\vec{x}$

$$\begin{aligned} ||\vec{x}||_1 \geq \vec{y}^T \vec{\nu} &= (A\vec{x}^*)^T \vec{\nu} \\ &= (\vec{x}^*)^T (A^T \vec{\nu}) \\ &= \sum_{i \in T} \vec{x}^*[i] \, \text{sign}(\vec{x}^*[i]) \\ &= ||\vec{x}^*||_1 \end{aligned}$$

$\vec{x}^*$ must be a solution

Proof 2

$A^T \vec{\nu}$ is a subgradient of the ℓ_1 norm at $\vec{x}^*$

For any other feasible vector $\vec{x}$

$$||\vec{x}||_1 \geq ||\vec{x}^*||_1 + \left(A^T \vec{\nu}\right)^T (\vec{x} - \vec{x}^*)$$

Proof 2

$A^T \vec{\nu}$ is a subgradient of the ℓ_1 norm at $\vec{x}^*$

For any other feasible vector $\vec{x}$

$$\begin{aligned} ||\vec{x}||_1 &\geq ||\vec{x}^*||_1 + \left(A^T \vec{\nu}\right)^T (\vec{x} - \vec{x}^*) \\ &= ||\vec{x}^*||_1 + \vec{\nu}^T (A\vec{x} - A\vec{x}^*) \end{aligned}$$

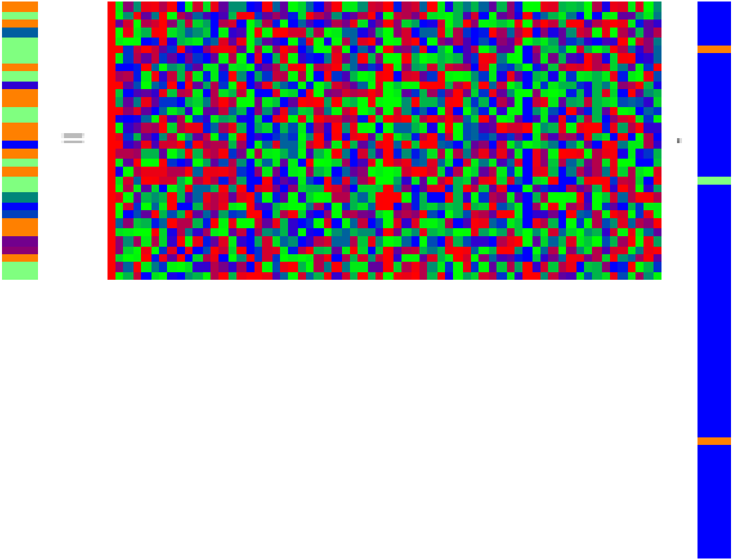
Proof 2

$A^T \vec{\nu}$ is a subgradient of the ℓ_1 norm at $\vec{x}^*$

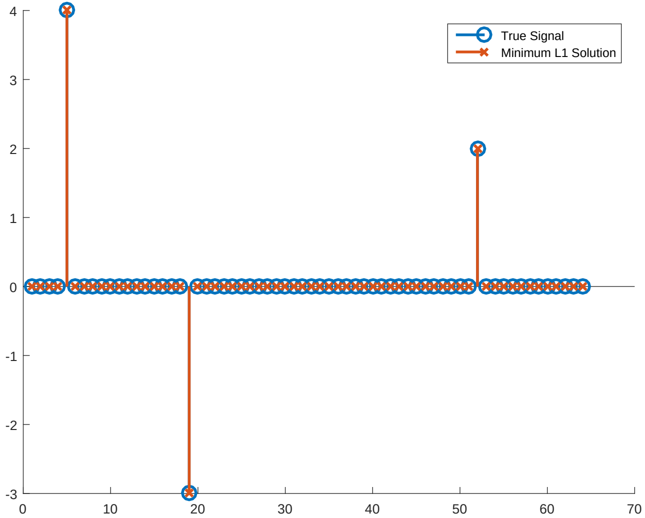
For any other feasible vector $\vec{x}$

$$
\begin{aligned}
||\vec{x}||_1 &\geq ||\vec{x}^*||_1 + \left(A^T \vec{\nu}\right)^T (\vec{x} - \vec{x}^*) \\
&= ||\vec{x}^*||_1 + \vec{\nu}^T (A\vec{x} - A\vec{x}^*) \\
&= ||\vec{x}^*||_1
\end{aligned}
$$

Random subsampling



Minimum ℓ_1-norm solution (random subsampling)



Exact sparse recovery via ℓ_1-norm minimization

Assumption: There exists a signal $\vec{x}^* \in \mathbb{R}^m$ with s nonzeros such that

$$\mathbf{A}\vec{x}^* = \vec{y}$$

for a random $\mathbf{A} \in \mathbb{R}^{m \times n}$ (random Fourier, Gaussian iid, ...)

Exact recovery: If the number of measurements satisfies

$$m \geq C' s \log n$$

the solution of the problem

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

is the original signal with probability at least $1 - \frac{1}{n}$

Proof

Show that dual certificate always exists

We need

$$\mathbf{A}_T^T \vec{\nu} = \text{sign}(\vec{x}_T^*) \qquad \textcolor{red}{s \text{ constraints}}$$

$$\left|\left|\mathbf{A}_{T^c}^T \vec{\nu}\right|\right|_\infty < 1$$

Idea: Impose $\mathbf{A}_T \vec{\nu} = \text{sign}(\vec{x}^*)$ and minimize $\left|\left|\mathbf{A}_{T^c}^T \vec{\nu}\right|\right|_\infty$

Problem: No closed-form solution

How about minimizing ℓ_2 norm?

Proof of exact recovery

Prove that dual certificate exists for any s-sparse $\vec{x}^*$

Dual certificate candidate: Solution of

$$\begin{aligned} \text{minimize} \quad & ||\vec{v}||_2 \\ \text{subject to} \quad & \mathbf{A}_T^T \vec{v} = \text{sign}\left(\vec{x}_T^*\right) \end{aligned}$$

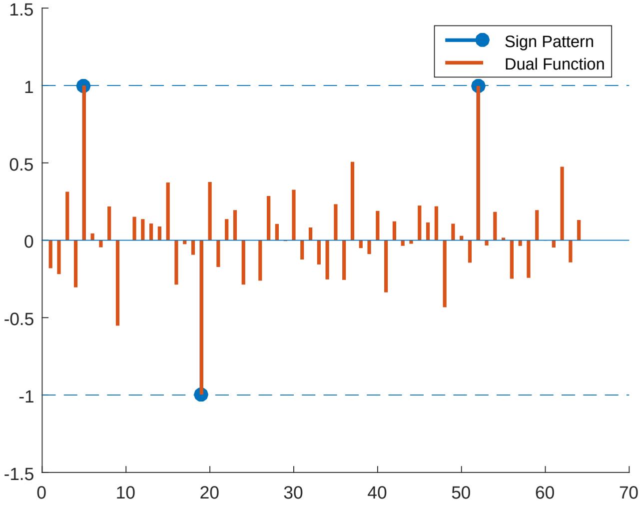
Closed-form solution $\vec{\boldsymbol{\nu}}_{\ell_2} := \mathbf{A}_T \left(\mathbf{A}_T^T \mathbf{A}_T\right)^{-1} \text{sign}\left(\vec{x}_T^*\right)$

$\mathbf{A}_T^T \mathbf{A}_T$ is invertible with high probability

We need to prove that $\mathbf{A}^T \vec{\boldsymbol{\nu}}_{\ell_2}$ satisfies

$$\left|\left|(\mathbf{A}^T \vec{\nu}_{\ell_2})_{T^c}\right|\right|_\infty < 1$$

Dual certificate



Proof of exact recovery

To control $(\mathbf{A}^T \vec{\nu}_{\ell_2})_{T^c}$, we need to bound

$$\mathbf{A}_i^T \left(\mathbf{A}_T^T \mathbf{A}_T \right)^{-1} \operatorname{sign}\left(\vec{x}_T^* \right)$$

for $i \in T^c$

Let $\vec{\mathbf{w}} := \left(\mathbf{A}_T^T \mathbf{A}_T \right)^{-1} \operatorname{sign}\left(\vec{x}_T^* \right)$

$|\mathbf{A}_i^T \vec{\mathbf{w}}|$ can be bounded using independence

Result then follows from union bound