Duality

DS-GA 1013 / MATH-GA 2824 Mathematical Tools for Data Science

https://cims.nyu.edu/~cfgranda/pages/MTDS_spring19/index.html

Carlos Fernandez-Granda

Motivating applications

Convex sets

Lagrangian duality

Proof of strong duality

Compressed sensing

Matrix completion

Compressed sensing

Goal: Recovering signals from small number of data

Arbitrary vector of dimension d cannot be recovered from $m < d$ linear measurements

However, signals of interest are highly structured

For example, images are sparse in wavelet basis

If signal is parametrized by $s < m$ parameters, recovery may be possible

We focus on simplified problem: recovering sparse vectors

Restricted-isometry property

Different sparse vectors should never produce similar data

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

This is true for random operators with high probability

Simplified problem

Recover sparse signal $\vec{x}_{\text{true}} \in \mathbb{R}^m$ from measurements

$$A\vec{x}_{\text{true}} = \vec{y}$$

where $\vec{y} \in \mathbb{R}^n$, $A \in \mathbb{R}^{m \times n}$, and $m < n$

Minimize nonzero entries subject to equality constraints?

Promoting sparsity

Toy problem: Find t such that

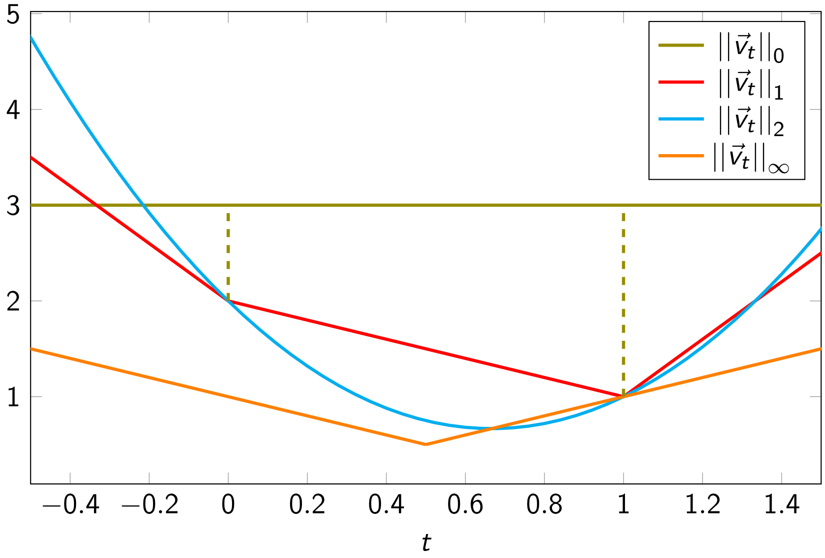
$$\vec{v}_t := \begin{bmatrix} t \\ t-1 \\ t-1 \end{bmatrix}$$

is sparse

Strategy: Minimize

$$f(t) := ||\vec{v}_t||$$

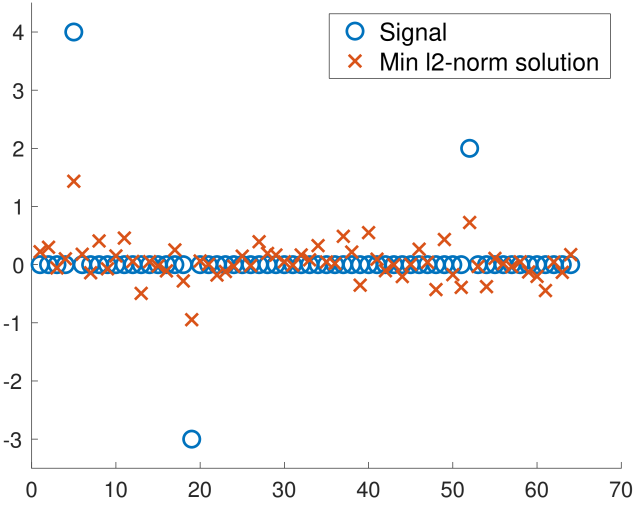
Promoting sparsity



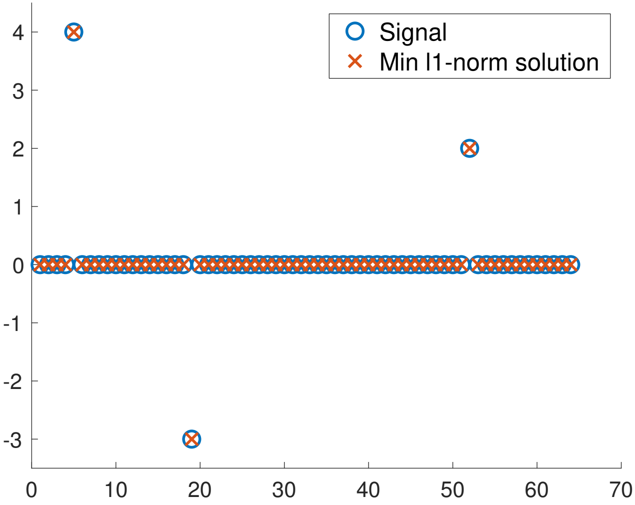
ℓ_1-norm minimization with equality constraints

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

ℓ_2-norm minimization with equality constraints



ℓ_1-norm minimization with equality constraints



Another algorithm for sparse recovery

Imagine we have access to inner products with sparse vector

$$\langle \vec{u}, \vec{x}_{\text{true}} \rangle$$

Strategy: Solve

$$\max_{\vec{u} \in \mathbb{R}^n} \langle \vec{u}, \vec{x}_{\text{true}} \rangle$$

Another algorithm for sparse recovery

Imagine we have access to inner products with sparse vector

$$\langle \vec{u}, \vec{x}_{\text{true}} \rangle$$

Strategy: Solve

$$\max_{\vec{u} \in \mathbb{R}^n} \langle \vec{u}, \vec{x}_{\text{true}} \rangle \quad \text{subject to} \;\; ||\vec{u}||_\infty \leq 1$$

Another algorithm for sparse recovery

If we have $\vec{y} = A\vec{x}_{\text{true}}$

$$\langle \vec{v}, \vec{y} \rangle$$

Another algorithm for sparse recovery

If we have $\vec{y} = A\vec{x}_{\text{true}}$

$$\langle \vec{v}, \vec{y} \rangle = \langle \vec{v}, A\vec{x}_{\text{true}} \rangle$$

Another algorithm for sparse recovery

If we have $\vec{y} = A\vec{x}_{\text{true}}$

$$\begin{aligned}\langle \vec{v}, \vec{y} \rangle &= \langle \vec{v}, A\vec{x}_{\text{true}} \rangle \\ &= \langle A^T \vec{v}, \vec{x}_{\text{true}} \rangle\end{aligned}$$

Another algorithm for sparse recovery

If we have $\vec{y} = A\vec{x}_{\text{true}}$

$$\begin{aligned} \langle \vec{v}, \vec{y} \rangle &= \langle \vec{v}, A\vec{x}_{\text{true}} \rangle \\ &= \langle A^T \vec{v}, \vec{x}_{\text{true}} \rangle \end{aligned}$$

We can solve

$$\max_{\vec{v} \in \mathbb{R}^m} \langle \vec{v}, \vec{y} \rangle \quad \text{subject to} \quad \left|\left| A^T \vec{v} \right|\right|_\infty \leq 1$$

Another algorithm for sparse recovery

If we have $\vec{y} = A\vec{x}_{\text{true}}$

$$\begin{aligned} \langle \vec{v}, \vec{y} \rangle &= \langle \vec{v}, A\vec{x}_{\text{true}} \rangle \\ &= \langle A^T \vec{v}, \vec{x}_{\text{true}} \rangle \end{aligned}$$

We can solve

$$\max_{\vec{v} \in \mathbb{R}^m} \langle \vec{v}, \vec{y} \rangle \quad \text{subject to} \quad \left|\left| A^T \vec{v} \right|\right|_\infty \leq 1$$

Equivalent to ℓ_1-norm minimization!

Matrix completion

Promoting low-rank structure

Finding low-rank matrices consistent with data is often very useful

Toy problem: Find t such that

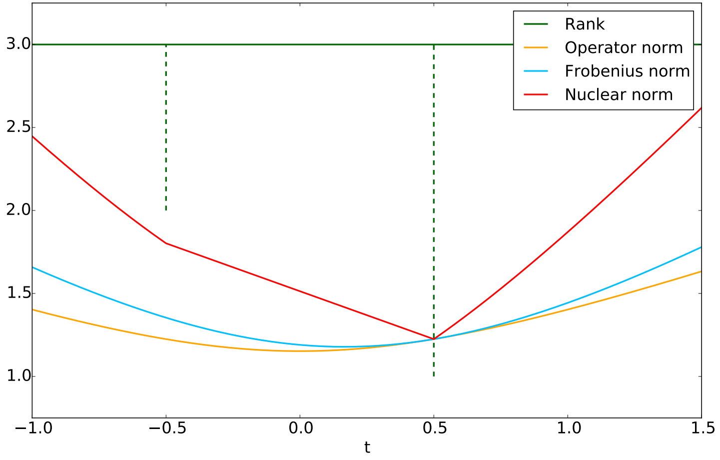
$$M(t) := \begin{bmatrix} 0.5 + t & 1 & 1 \\ 0.5 & 0.5 & t \\ 0.5 & 1 - t & 0.5 \end{bmatrix},$$

is low rank

Strategy: Minimize

$$f(t) := ||M(t)||$$

Promoting low-rank structure

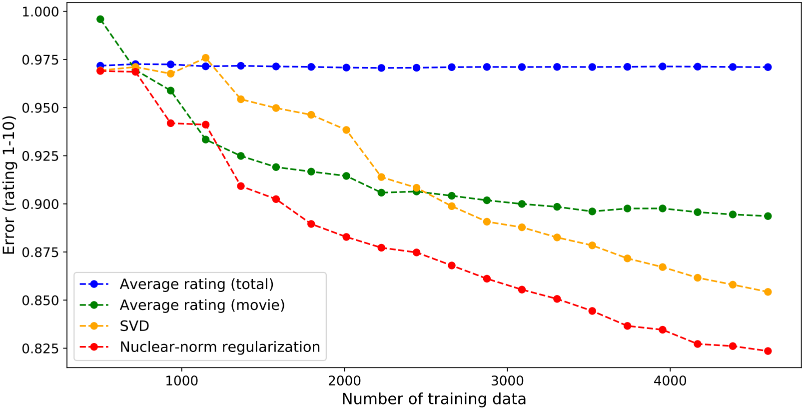


Nuclear-norm minimization for matrix completion

$\vec{y}$ contains the observed entries indexed by set Ω

$$\min_{X \in \mathbb{R}^{n_1 \times n_2}} ||X||_* \quad \text{such that } X_\Omega = \vec{y}$$

Results for movie dataset



Motivating applications

Convex sets

Lagrangian duality

Proof of strong duality

Compressed sensing

Matrix completion

Constrained optimization problem

$$\min_{\vec{x} \in \mathbb{R}^n} f(\vec{x}) \qquad \text{subject to } \vec{x} \in \mathcal{S} \subset \mathbb{R}^n$$

Any $\vec{x} \in \mathcal{S}$ is a feasible point

Convex functions

A function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if for any $\vec{x}, \vec{y} \in \mathbb{R}^n$ and any $\theta \in (0, 1)$

$$\theta f(\vec{x}) + (1 - \theta) f(\vec{y}) \geq f(\theta \vec{x} + (1 - \theta) \vec{y})$$

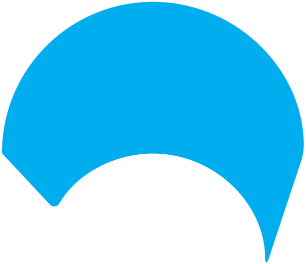
Convex sets

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

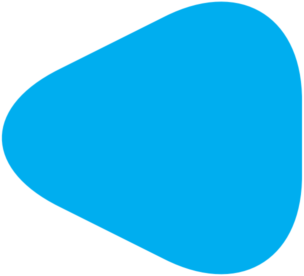
$$\theta \vec{x} + (1 - \theta)\, \vec{y} \in \mathcal{S}$$

Convex vs nonconvex

Nonconvex



Convex



Separating hyperplane

There exists a hyperplane separating any nonempty disjoint convex sets $\mathcal{S}_1$, $\mathcal{S}_2 \subset \mathbb{R}^n$

There exists $\vec{a} \in \mathbb{R}^n$, $b \in \mathbb{R}$ such that

$$\text{for all } \vec{x}_1 \in \mathcal{S}_1 \qquad \langle \vec{a}, \vec{x}_1 \rangle \leq b$$

$$\text{for all } \vec{x}_2 \in \mathcal{S}_2 \qquad \langle \vec{a}, \vec{x}_2 \rangle \leq b$$

Proof

Simplifying assumption:

$$||\vec{y}_2 - \vec{y}_1||_2 = \min_{\vec{x}_1 \in \mathcal{S}_1, \vec{x}_2 \in \mathcal{S}_2} ||\vec{x}_2 - \vec{x}_1||_2$$

Hyperplane orthogonal to $\vec{y}_2 - \vec{y}_1$ between $\vec{y}_1$ and $\vec{y}_2$:

$$h(\vec{x}) := \left\langle \vec{y}_2 - \vec{y}_1, \vec{x} - \frac{\vec{y}_1 + \vec{y}_2}{2} \right\rangle = 0$$

Goal: Show $h(\vec{x}) > 0$ for all $\mathcal{S}_2$

Proof

Assume that $h(\vec{u}) < 0$ for $\vec{u} \in \mathcal{S}_2$

$$h(\vec{u}) = \left\langle \vec{y}_2 - \vec{y}_1, \vec{u} - \frac{\vec{y}_1 + \vec{y}_2}{2} \right\rangle$$

Proof

Assume that $h(\vec{u}) < 0$ for $\vec{u} \in \mathcal{S}_2$

$$
\begin{aligned}
h(\vec{u}) &= \left\langle \vec{y}_2 - \vec{y}_1, \vec{u} - \frac{\vec{y}_1 + \vec{y}_2}{2} \right\rangle \\
&= \langle \vec{y}_2 - \vec{y}_1, \vec{u} - \vec{y}_2 \rangle + \left\langle \vec{y}_2 - \vec{y}_1, \frac{\vec{y}_2 - \vec{y}_1}{2} \right\rangle
\end{aligned}
$$

Proof

Assume that $h(\vec{u}) < 0$ for $\vec{u} \in \mathcal{S}_2$

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

Proof

Assume that $h(\vec{u}) < 0$ for $\vec{u} \in \mathcal{S}_2$

$$h(\vec{u}) = \left\langle \vec{y}_2 - \vec{y}_1, \vec{u} - \frac{\vec{y}_1 + \vec{y}_2}{2} \right\rangle$$
$$= \langle \vec{y}_2 - \vec{y}_1, \vec{u} - \vec{y}_2 \rangle + \left\langle \vec{y}_2 - \vec{y}_1, \frac{\vec{y}_2 - \vec{y}_1}{2} \right\rangle$$
$$= \langle \vec{y}_2 - \vec{y}_1, \vec{u} - \vec{y}_2 \rangle + \frac{1}{2} \left|\left|\vec{y}_2 - \vec{y}_1\right|\right|_2^2$$

so

$$\langle \vec{y}_2 - \vec{y}_1, \vec{u} - \vec{y}_2 \rangle < 0$$

Proof

$$\vec{y}_\theta := \theta \vec{u} + (1-\theta)\vec{y}_2 \in \mathcal{S}_2$$

$$||\vec{y}_\theta - \vec{y}_1||_2^2 = ||\theta(\vec{u} - \vec{y}_2) + \vec{y}_2 - \vec{y}_1||_2^2$$

Proof

$$\vec{y}_\theta := \theta \vec{u} + (1-\theta)\vec{y}_2 \in \mathcal{S}_2$$

$$\begin{aligned} \|\vec{y}_\theta - \vec{y}_1\|_2^2 &= \|\theta(\vec{u} - \vec{y}_2) + \vec{y}_2 - \vec{y}_1\|_2^2 \\ &= \|\vec{y}_2 - \vec{y}_1\|_2^2 + \theta^2 \|\vec{u} - \vec{y}_2\|_2^2 + 2\theta \langle \vec{y}_2 - \vec{y}_1, \vec{u} - \vec{y}_2 \rangle \end{aligned}$$

Proof

$$\vec{y}_\theta := \theta \vec{u} + (1-\theta)\vec{y}_2 \in \mathcal{S}_2$$

$$\begin{aligned} ||\vec{y}_\theta - \vec{y}_1||_2^2 &= ||\theta(\vec{u} - \vec{y}_2) + \vec{y}_2 - \vec{y}_1||_2^2 \\ &= ||\vec{y}_2 - \vec{y}_1||_2^2 + \theta^2 \, ||\vec{u} - \vec{y}_2||_2^2 + 2\theta \, \langle \vec{y}_2 - \vec{y}_1, \vec{u} - \vec{y}_2 \rangle \\ &= ||\vec{y}_2 - \vec{y}_1||_2^2 + g(\theta) \end{aligned}$$

Proof

$\vec{y}_\theta := \theta \vec{u} + (1-\theta)\vec{y}_2 \in \mathcal{S}_2$

$$\begin{aligned} ||\vec{y}_\theta - \vec{y}_1||_2^2 &= ||\theta(\vec{u} - \vec{y}_2) + \vec{y}_2 - \vec{y}_1||_2^2 \\ &= ||\vec{y}_2 - \vec{y}_1||_2^2 + \theta^2 \, ||\vec{u} - \vec{y}_2||_2^2 + 2\theta \, \langle \vec{y}_2 - \vec{y}_1, \vec{u} - \vec{y}_2 \rangle \\ &= ||\vec{y}_2 - \vec{y}_1||_2^2 + g(\theta) \end{aligned}$$

$g(0) = 0$ and $g'(0) = \langle \vec{y}_2 - \vec{y}_1, \vec{u} - \vec{y}_2 \rangle < 0$

Proof

$\vec{y}_\theta := \theta \vec{u} + (1 - \theta)\vec{y}_2 \in \mathcal{S}_2$

$$\begin{aligned} ||\vec{y}_\theta - \vec{y}_1||_2^2 &= ||\theta(\vec{u} - \vec{y}_2) + \vec{y}_2 - \vec{y}_1||_2^2 \\ &= ||\vec{y}_2 - \vec{y}_1||_2^2 + \theta^2 \, ||\vec{u} - \vec{y}_2||_2^2 + 2\theta \, \langle \vec{y}_2 - \vec{y}_1, \vec{u} - \vec{y}_2 \rangle \\ &= ||\vec{y}_2 - \vec{y}_1||_2^2 + g(\theta) \end{aligned}$$

$g(0) = 0$ and $g'(0) = \langle \vec{y}_2 - \vec{y}_1, \vec{u} - \vec{y}_2 \rangle < 0$

For small enough θ $\vec{y}_\theta$ is closer to $\vec{y}_1$ than $\vec{y}_2$

Hyperplanes are convex

Let $\mathcal{H} := \left\{ \vec{x} \mid A\vec{x} = \vec{b} \right\}$

For any $\vec{x}, \vec{y} \in \mathcal{H}$ and any $\theta \in (0, 1)$

$$A\left(\theta\vec{x} + (1 - \theta)\,\vec{y}\right) =$$

Hyperplanes are convex

Let $\mathcal{H} := \left\{\vec{x} \mid A\vec{x} = \vec{b}\right\}$

For any $\vec{x}, \vec{y} \in \mathcal{H}$ and any $\theta \in (0, 1)$

$$A\left(\theta\vec{x} + (1-\theta)\,\vec{y}\right) = \theta A\vec{x} + (1-\theta)\,A\vec{y}$$

Hyperplanes are convex

Let $\mathcal{H} := \left\{ \vec{x} \mid A\vec{x} = \vec{b} \right\}$

For any $\vec{x}, \vec{y} \in \mathcal{H}$ and any $\theta \in (0, 1)$

$$\begin{aligned} A\left(\theta\vec{x} + (1-\theta)\,\vec{y}\right) &= \theta A\vec{x} + (1-\theta)\,A\vec{y} \\ &= \vec{b} \end{aligned}$$

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

Sublevel sets

$$\mathcal{S}_\gamma := \{\vec{x} \mid f(\vec{x}) \leq \gamma\}$$

Sublevel sets of convex functions are convex

Let $\vec{x}$, $\vec{y} \in \mathcal{S}_\gamma$

$$f\left(\theta\vec{x} + (1-\theta)\,\vec{y}\right)$$

Sublevel sets of convex functions are convex

Let $\vec{x}$, $\vec{y} \in \mathcal{S}_\gamma$

$$f\left(\theta \vec{x} + (1-\theta)\,\vec{y}\right) \leq \theta f\left(\vec{x}\right) + (1-\theta)\, f\left(\vec{y}\right)$$

Sublevel sets of convex functions are convex

Let $\vec{x}$, $\vec{y} \in \mathcal{S}_\gamma$

$$\begin{aligned} f\left(\theta\vec{x} + (1-\theta)\,\vec{y}\right) &\leq \theta f\left(\vec{x}\right) + (1-\theta)\, f\left(\vec{y}\right) \\ &\leq \gamma \end{aligned}$$

Intersection of convex sets

If $\mathcal{S}_1, \ldots, \mathcal{S}_m$ are convex, $\cap_{i=1}^m \mathcal{S}_i$ is convex

Constrained optimization

Any optimization problem of the form,

$$\min_{\vec{x} \in \mathbb{R}^n} f_0(\vec{x}) \qquad \text{subject to } f_i(\vec{x}) \leq 0, \quad 1 \leq i \leq k,$$
$$A\vec{x} = \vec{b},$$

where $A \in \mathbb{R}^{m \times n}$, $\vec{b} \in \mathbb{R}^m$, has a convex feasibility set

Optimization problem with equality constraints

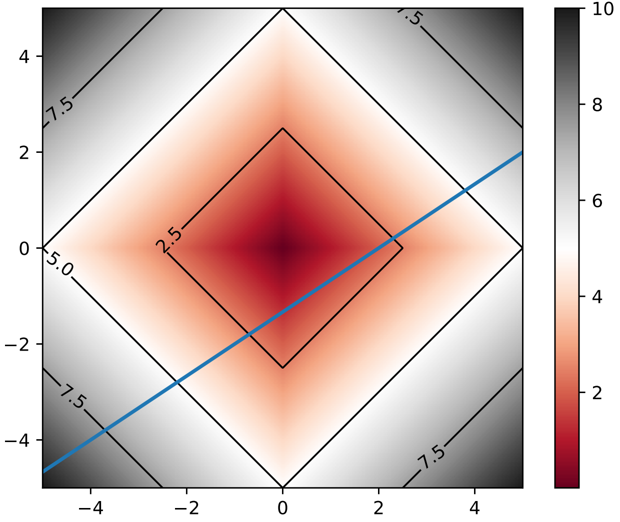
$A \in \mathbb{R}^{m \times n}$, $\vec{b} \in \mathbb{R}^m$

$$\min_{\vec{x} \in \mathbb{R}^n} f(\vec{x}) \qquad \text{subject to } A\vec{x} = \vec{b}$$

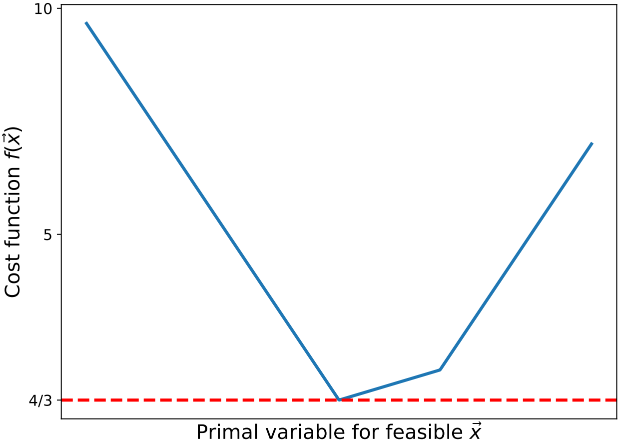
Example

$$\min_{\vec{x} \in \mathbb{R}^2} ||\vec{x}||_1 \quad \text{subject to } 2\vec{x}[1] - 3\vec{x}[2] = 4$$

Example



Example


10
5
4/3
Cost function $f(\vec{x})$
Primal variable for feasible $\vec{x}$

Lagrangian

$$L(\vec{x}, \vec{\alpha}) := f(\vec{x}) + \vec{\alpha}^T \left(\vec{b} - A\vec{x}\right)$$

$\vec{\alpha} \in \mathbb{R}^m$ is called a Lagrange multiplier

Lagrangian

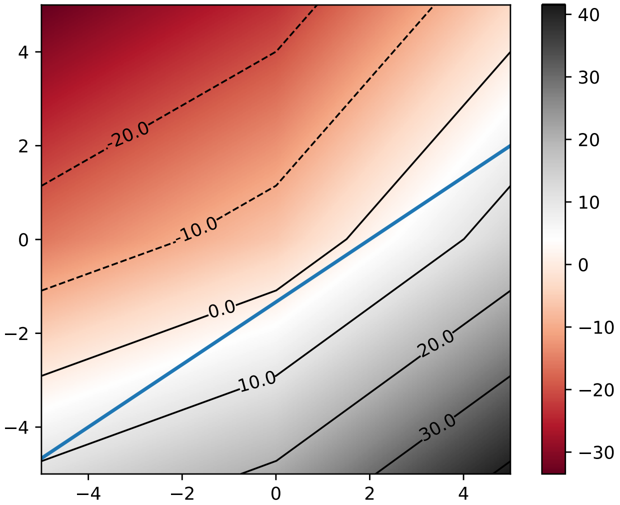
At any feasible point the Lagrangian is equal to the cost function

$$L(\vec{x}, \vec{\alpha}) = f(\vec{x})$$

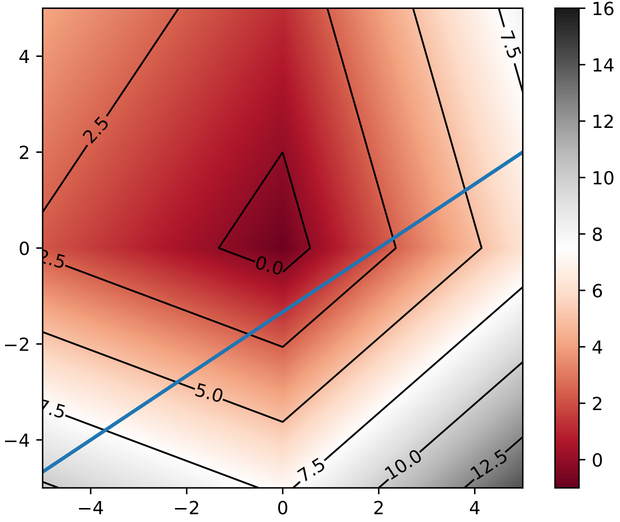
Example

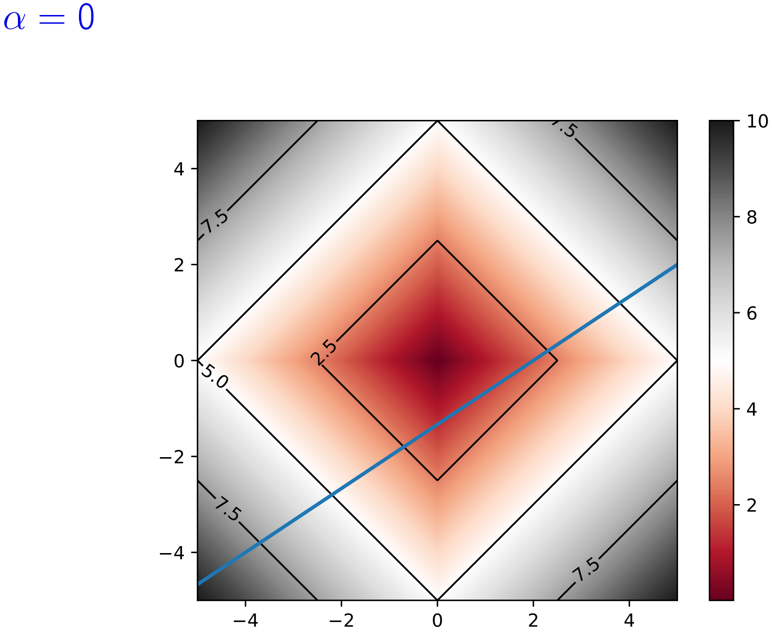
$$\mathcal{L}\left(\vec{x}, \alpha\right) = ||\vec{x}||_1 + \alpha(4 - 2\vec{x}[1] + 3\vec{x}[2])$$

$\alpha = -1.5$



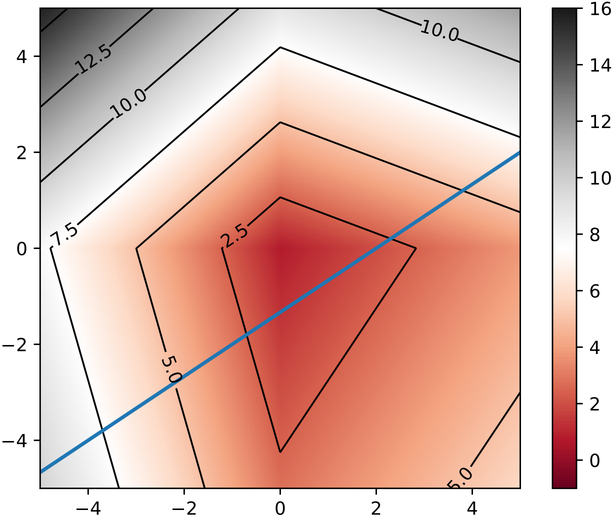
$\alpha = -0.2$



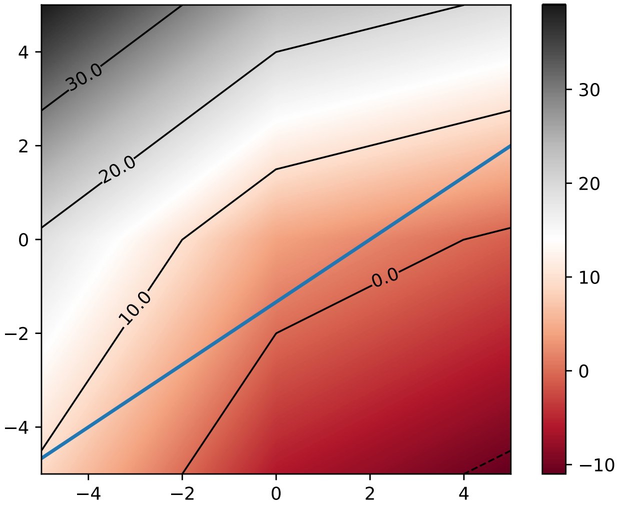


$\alpha = 0$
7.5
7.5
2.5
5.0
7.5
7.5
4
2
0
−2
−4
−4
−2
0
2
4
10
8
6
4
2

$\alpha = 0.2$



$\alpha = 1.0$



Lagrange dual function

$$g(\vec{\alpha}) := \inf_{\vec{x} \in \mathbb{R}^n} L(\vec{x}, \vec{\alpha})$$

Lagrange dual function as a lower bound

Let p^* be a minimum of the primal, for any $\vec{\alpha}$

$$p^*$$

Lagrange dual function as a lower bound

Let p^* be a minimum of the primal, for any $\vec{\alpha}$

$$p^* = f(\vec{x}^*)$$

Lagrange dual function as a lower bound

Let p^* be a minimum of the primal, for any $\vec{\alpha}$

$$\begin{aligned} p^* &= f(\vec{x}^*) \\ &= L(\vec{x}^*, \vec{\alpha}) \end{aligned}$$

Lagrange dual function as a lower bound

Let p^* be a minimum of the primal, for any $\vec{\alpha}$

$$\begin{aligned} p^* &= f(\vec{x}^*) \\ &= L(\vec{x}^*, \vec{\alpha}) \\ &\geq g(\vec{\alpha}) \end{aligned}$$

Dual problem

$$\max_{\alpha \in \mathbb{R}^m} g(\alpha)$$

$-g(\alpha) := \sup_{\vec{x} \in \mathbb{R}^n} L(\vec{x}, \vec{\alpha})$ is a pointwise supremum of linear functions

Maximum/supremum of convex functions

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

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

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 $\vec{x}, \vec{y} \in \mathbb{R}$,

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

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 $\vec{x}, \vec{y} \in \mathbb{R}$,

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

Weak duality

Let d^* be a maximum of the dual problem

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

Strong duality

Let d^* be a maximum of the dual problem

$$d^* = p^*$$

Not so obvious...

Norm minimization

The Lagrange dual function of

$$\min_{\vec{x} \in \mathbb{R}^n} ||\vec{x}|| \qquad \text{subject to } A\vec{x} = \vec{b}$$

where $A \in \mathbb{R}^{m \times n}$, $\vec{b} \in \mathbb{R}^m$, equals

$$\max_{\vec{\alpha} \in \mathbb{R}^m} \langle \vec{\alpha}, \vec{b} \rangle \qquad \text{subject to } \left|\left| A^T \vec{\alpha} \right|\right|_d \leq 1$$

$$||\vec{y}||_d := \max_{||\vec{x}|| \leq 1} \langle \vec{y}, \vec{x} \rangle$$

Proof

$$L(\vec{x}, \vec{\alpha}) := ||\vec{x}|| + \vec{\alpha}^T \left(\vec{b} - A\vec{x}\right)$$

Proof

$$
\begin{aligned}
L(\vec{x},\vec{\alpha}) &:= ||\vec{x}|| + \vec{\alpha}^T\left(\vec{b} - A\vec{x}\right) \\
&= ||\vec{x}|| - \langle A^T\vec{\alpha}, \vec{x}\rangle + \vec{\alpha}^T\vec{b}
\end{aligned}
$$

Proof

$$
\begin{aligned}
L(\vec{x}, \vec{\alpha}) &:= ||\vec{x}|| + \vec{\alpha}^T \left(\vec{b} - A\vec{x} \right) \\
&= ||\vec{x}|| - \langle A^T \vec{\alpha}, \vec{x} \rangle + \vec{\alpha}^T \vec{b} \\
&= \left(1 - \left\langle A^T \vec{\alpha}, \frac{\vec{x}}{||\vec{x}||} \right\rangle \right) ||\vec{x}|| + \vec{\alpha}^T \vec{b}
\end{aligned}
$$

Proof

$$
\begin{aligned}
L(\vec{x},\vec{\alpha}) &:= ||\vec{x}|| + \vec{\alpha}^T\left(\vec{b} - A\vec{x}\right)\\
&= ||\vec{x}|| - \langle A^T\vec{\alpha}, \vec{x}\rangle + \vec{\alpha}^T\vec{b}\\
&= \left(1 - \left\langle A^T\vec{\alpha}, \frac{\vec{x}}{||\vec{x}||}\right\rangle\right)||\vec{x}|| + \vec{\alpha}^T\vec{b}\\
&\geq a\left(1 - \left|\left|A^T\vec{\alpha}\right|\right|_d\right) + \vec{\alpha}^T\vec{b}
\end{aligned}
$$

$$
\begin{aligned}
&a := ||\vec{x}||\\
&\vec{u} := \arg\max_{||\vec{x}||\leq 1}\left\langle A^T\vec{\alpha}, \vec{x}\right\rangle \text{ so that } \left\langle A^T\vec{\alpha}, \vec{u}\right\rangle = \left|\left|A^T\vec{\alpha}\right|\right|_d
\end{aligned}
$$

Proof

$$
\begin{aligned}
L(\vec{x}, \vec{\alpha}) &:= ||\vec{x}|| + \vec{\alpha}^T \left(\vec{b} - A\vec{x}\right) \\
&= ||\vec{x}|| - \langle A^T \vec{\alpha}, \vec{x} \rangle + \vec{\alpha}^T \vec{b} \\
&= \left(1 - \left\langle A^T \vec{\alpha}, \frac{\vec{x}}{||\vec{x}||} \right\rangle\right) ||\vec{x}|| + \vec{\alpha}^T \vec{b} \\
&\geq a \left(1 - \left|\left|A^T \vec{\alpha}\right|\right|_d\right) + \vec{\alpha}^T \vec{b} \\
&= a \left(1 - \langle A^T \vec{\alpha}, \vec{u} \rangle\right) + \vec{\alpha}^T \vec{b}
\end{aligned}
$$

$$
\begin{aligned}
&a := ||\vec{x}|| \\
&\vec{u} := \arg \max_{||\vec{x}|| \leq 1} \left\langle A^T \vec{\alpha}, \vec{x} \right\rangle \text{ so that } \left\langle A^T \vec{\alpha}, \vec{u} \right\rangle = \left|\left|A^T \vec{\alpha}\right|\right|_d
\end{aligned}
$$

Proof

$$
\begin{aligned}
L(\vec{x}, \vec{\alpha}) &:= ||\vec{x}|| + \vec{\alpha}^T \left(\vec{b} - A\vec{x}\right) \\
&= ||\vec{x}|| - \langle A^T \vec{\alpha}, \vec{x} \rangle + \vec{\alpha}^T \vec{b} \\
&= \left(1 - \left\langle A^T \vec{\alpha}, \frac{\vec{x}}{||\vec{x}||} \right\rangle\right) ||\vec{x}|| + \vec{\alpha}^T \vec{b} \\
&\geq a \left(1 - \left|\left|A^T \vec{\alpha}\right|\right|_d\right) + \vec{\alpha}^T \vec{b} \\
&= a \left(1 - \langle A^T \vec{\alpha}, \vec{u} \rangle\right) + \vec{\alpha}^T \vec{b} \\
&= L(a\vec{u}, \vec{\alpha})
\end{aligned}
$$

$$
\begin{aligned}
a &:= ||\vec{x}|| \\
\vec{u} &:= \arg \max_{||\vec{x}|| \leq 1} \left\langle A^T \vec{\alpha}, \vec{x} \right\rangle \text{ so that } \left\langle A^T \vec{\alpha}, \vec{u} \right\rangle = \left|\left|A^T \vec{\alpha}\right|\right|_d
\end{aligned}
$$

Proof

$$L(a\vec{u}, \vec{\alpha}) = a\left(1 - \left|\left|A^T\vec{\alpha}\right|\right|_d\right)$$

Proof

$$L(a\vec{u}, \vec{\alpha}) = a\left(1 - \left|\left|A^T\vec{\alpha}\right|\right|_d\right)$$

- If $\left|\left|A^T\vec{\alpha}\right|\right|_d > 1$ when $a \to -\infty$ $L(a\vec{u}, \vec{\alpha}) \to \infty$

Proof

$$L(a\vec{u}, \vec{\alpha}) = a\left(1 - \left|\left|A^T\vec{\alpha}\right|\right|_d\right)$$

- If $\left|\left|A^T\vec{\alpha}\right|\right|_d > 1$ when $a \to -\infty$ $L(a\vec{u}, \vec{\alpha}) \to \infty$
- If $\left|\left|A^T\vec{\alpha}\right|\right|_d \leq 1$ then minimum achieved by $a := 0$

Proof

$$L(a\vec{u}, \vec{\alpha}) = a\left(1 - \left|\left|A^T\vec{\alpha}\right|\right|_d\right)$$

- If $\left|\left|A^T\vec{\alpha}\right|\right|_d > 1$ when $a \to -\infty$ $L(a\vec{u}, \vec{\alpha}) \to \infty$
- If $\left|\left|A^T\vec{\alpha}\right|\right|_d \leq 1$ then minimum achieved by $a := 0$

$$g(\vec{\alpha}) = \begin{cases} \vec{\alpha}^T\vec{b} & \text{if } \left|\left|A^T\vec{\alpha}\right|\right|_d \leq 1, \\ -\infty & \text{otherwise.} \end{cases}$$

ℓ_1-norm minimization

Let $A \in \mathbb{R}^{m \times n}$, $\vec{b} \in \mathbb{R}^m$. The dual of

$$\min_{\vec{x} \in \mathbb{R}^n} ||\vec{x}||_1 \qquad \text{subject to} \quad A\vec{x} = \vec{b}$$

is

$$\max_{\vec{\alpha} \in \mathbb{R}^m} \langle \vec{\alpha}, \vec{y} \rangle \qquad \text{subject to} \ \left|\left| A^T \vec{\alpha} \right|\right|_\infty \leq 1$$

Example

The dual of

$$\min_{\vec{x} \in \mathbb{R}^2} ||\vec{x}||_1 \quad \text{subject to } 2\vec{x}[1] - 3\vec{x}[2] = 4$$

is

Example

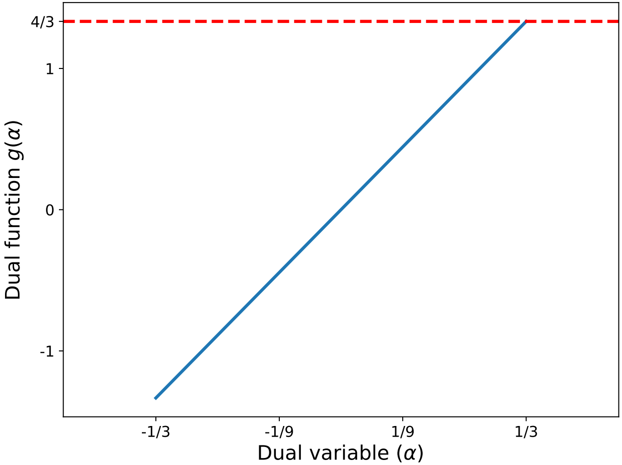
The dual of

$$\min_{\vec{x} \in \mathbb{R}^2} ||\vec{x}||_1 \quad \text{subject to } 2\vec{x}[1] - 3\vec{x}[2] = 4$$

is

$$\max_{\vec{\alpha} \in \mathbb{R}^m} \; 4\alpha \qquad \text{subject to } \; |\alpha| \leq \frac{1}{3}$$

Example



ℓ_1-norm minimization

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

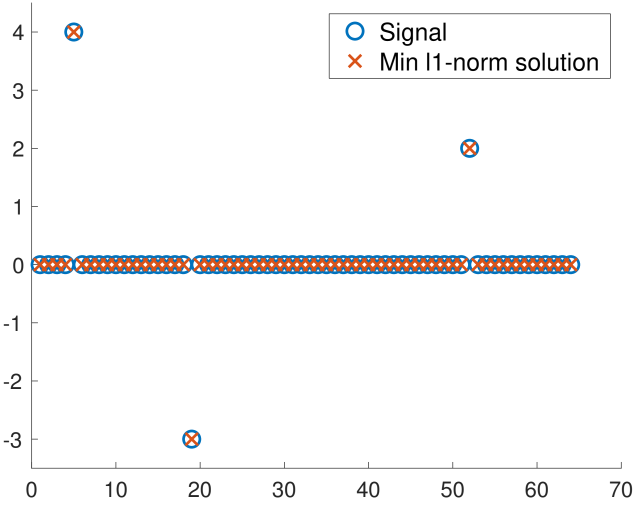
$$\max_{\vec{\alpha} \in \mathbb{R}^m} \langle \vec{\alpha}, \vec{y} \rangle \qquad \text{subject to } \left|\left| A^T \vec{\alpha} \right|\right|_\infty \leq 1$$

satisfies

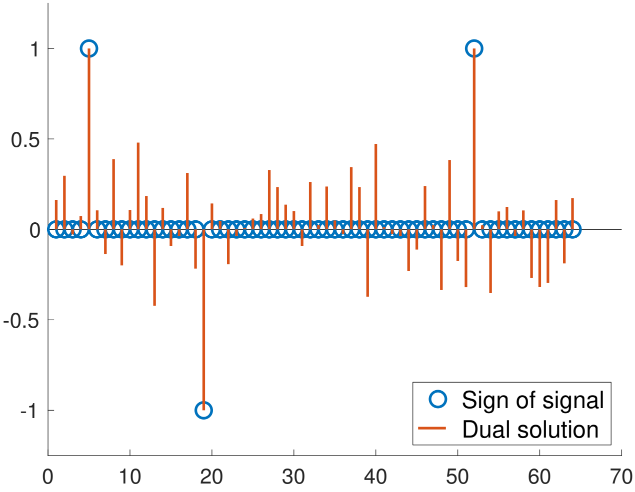
$$(A^T \vec{\alpha}^*)[i] = \text{sign}(\vec{x}^*[i]) \quad \text{for all } \vec{x}^*[i] \neq 0$$

for any solution $\vec{x}^*$ to the primal problem

ℓ_1-norm minimization with equality constraints



ℓ_1-norm minimization with equality constraints



Proof

By strong duality

$$||\vec{x}^*||_1 = \vec{y}^T \vec{\alpha}^*$$

Proof

By strong duality

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

Proof

By strong duality

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

Proof

By strong duality

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

Proof

By strong duality

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

By Hölder's inequality

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

Proof

By strong duality

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

By Hölder's inequality

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

Equality if and only if

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

Motivating applications

Convex sets

Lagrangian duality

Proof of strong duality

Compressed sensing

Matrix completion

Proof of strong duality

$$\mathcal{A} := \left\{ (\vec{v}, t) \mid \vec{b} - A\vec{x} = \vec{v} \quad \text{and} \quad f(\vec{x}) \leq t \quad \text{for some } \vec{x} \in \mathbb{R}^n \right\}$$

Proof of strong duality

$$\mathcal{A} := \left\{ (\vec{v}, t) \mid \vec{b} - A\vec{x} = \vec{v} \quad \text{and} \quad f(\vec{x}) \leq t \quad \text{for some } \vec{x} \in \mathbb{R}^n \right\}$$

$$p^* = \inf \left\{ t \mid (\vec{0}, t) \in \mathcal{A} \right\}$$

Proof of strong duality

$$\mathcal{A} := \left\{ (\vec{v}, t) \mid \vec{b} - A\vec{x} = \vec{v} \quad \text{and} \quad f(\vec{x}) \leq t \quad \text{for some } \vec{x} \in \mathbb{R}^n \right\}$$

$$p^* = \inf \left\{ t \mid (\vec{0}, t) \in \mathcal{A} \right\}$$

$$g(\vec{\alpha}) = \inf \left\{ \langle \vec{\alpha}, \vec{v} \rangle + t \mid (\vec{v}, t) \in \mathcal{A} \right\}$$

Geometrically

The hyperplane

$$\langle \vec{\alpha}, \vec{v} \rangle + t = g(\vec{\alpha})$$

is a supporting hyperplane to $\mathcal{A}$

Geometrically

The hyperplane

$$\langle \vec{\alpha}, \vec{v} \rangle + t = g(\vec{\alpha})$$

is a supporting hyperplane to $\mathcal{A}$

Implies weak duality

$$p^* = \langle \vec{\alpha}, \vec{0} \rangle + p^* \geq g(\vec{\alpha})$$

Example

$$\min_{\vec{x} \in \mathbb{R}^2} ||\vec{x}||_1 \quad \text{subject to } 2\vec{x}[1] - 3\vec{x}[2] = 4$$

Example

$$\min_{\vec{x} \in \mathbb{R}^2} ||\vec{x}||_1 \quad \text{subject to } 2\vec{x}[1] - 3\vec{x}[2] = 4$$

Fix $v := 4 - 2\vec{x}[1] + 3\vec{x}[2]$, then

$$||\vec{x}||_1 = |\vec{x}[1]| + \left| \frac{v - 4 + 2\vec{x}[1]}{3} \right|$$

Example

$$\min_{\vec{x}\in\mathbb{R}^2} ||\vec{x}||_1 \quad \text{subject to } 2\vec{x}[1] - 3\vec{x}[2] = 4$$

Fix $v := 4 - 2\vec{x}[1] + 3\vec{x}[2]$, then

$$||\vec{x}||_1 = |\vec{x}[1]| + \left|\frac{v - 4 + 2\vec{x}[1]}{3}\right|$$

Piecewise linear function with two kinks at $\vec{x}[1] = 0$ and $\vec{x}[1] = (4 - v)/2$

Example

$$\min_{\vec{x} \in \mathbb{R}^2} ||\vec{x}||_1 \quad \text{subject to } 2\vec{x}[1] - 3\vec{x}[2] = 4$$

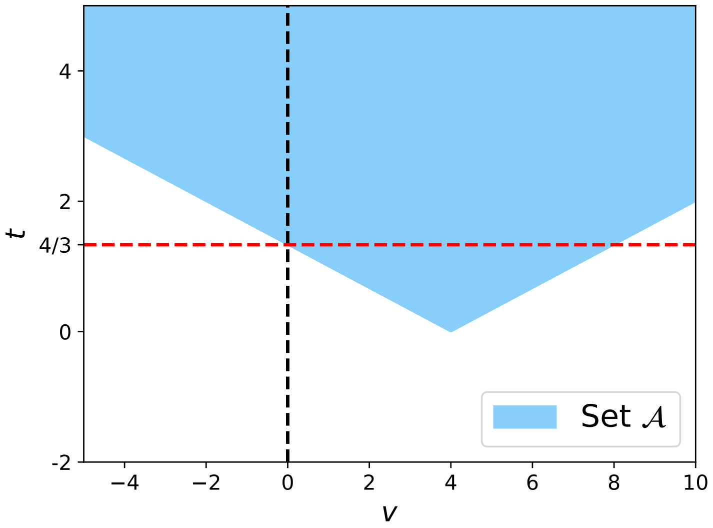
Fix $v := 4 - 2\vec{x}[1] + 3\vec{x}[2]$, then

$$||\vec{x}||_1 = |\vec{x}[1]| + \left| \frac{v - 4 + 2\vec{x}[1]}{3} \right|$$

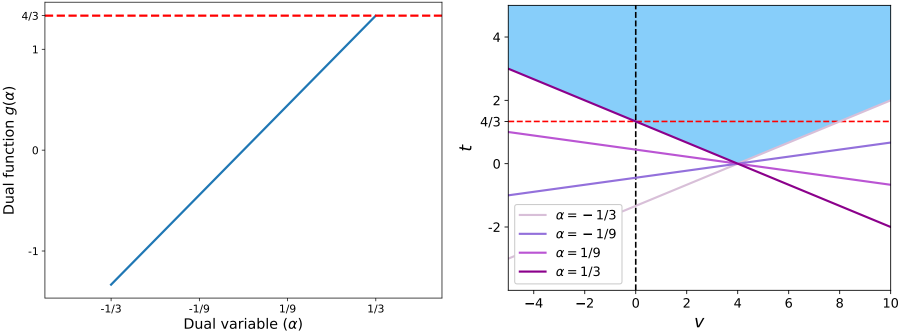
Piecewise linear function with two kinks at $\vec{x}[1] = 0$ and $\vec{x}[1] = (4 - v)/2$

$$\min_{v = 4 - 2\vec{x}[1] + 3\vec{x}[2]} ||\vec{x}||_1 = \min \left\{ \left| \frac{v - 4}{3} \right|, \left| \frac{v - 4}{2} \right| \right\}$$

Example



Example



$$\langle \vec{\alpha}, \vec{v} \rangle + t = g(\vec{\alpha})$$

$\mathcal{A}$ is convex

Let $(\vec{v}_1, t_1)$, $(\vec{v}_2, t_2) \in \mathcal{A}$

There exists $\vec{x}_1$ and $\vec{x}_2$ such that

$$\vec{v}_1 = \vec{b} - A\vec{x}_1, \qquad f(\vec{x}_1) \le t_1$$
$$\vec{v}_2 = \vec{b} - A\vec{x}_2, \qquad f(\vec{x}_2) \le t_2$$

This implies

$$\theta\vec{v}_1 + (1-\theta)\vec{v}_2 = \vec{b} - A(\vec{v}_1\vec{x}_1 + (1-\theta)\vec{x}_2)$$

and by convexity of f

$$\begin{aligned} f(\theta\vec{x}_1 + (1-\theta)\vec{x}_2) &\le \theta f(\vec{x}_1) + (1-\theta)f(\vec{x}_2) \\ &\le \theta t_1 + (1-\theta)t_2, \end{aligned}$$

so $\theta(\vec{v}_1, t_1) + (1-\theta)(\vec{v}_2, t_2) \in \mathcal{A}$

Another convex set

$$\mathcal{B} := \left\{ (\vec{0}, t) \mid t < p^* \right\}$$

Another convex set

$$\mathcal{B} := \left\{ (\vec{0}, t) \mid t < p^* \right\}$$

$\mathcal{A}$ and $\mathcal{B}$ are disjoint

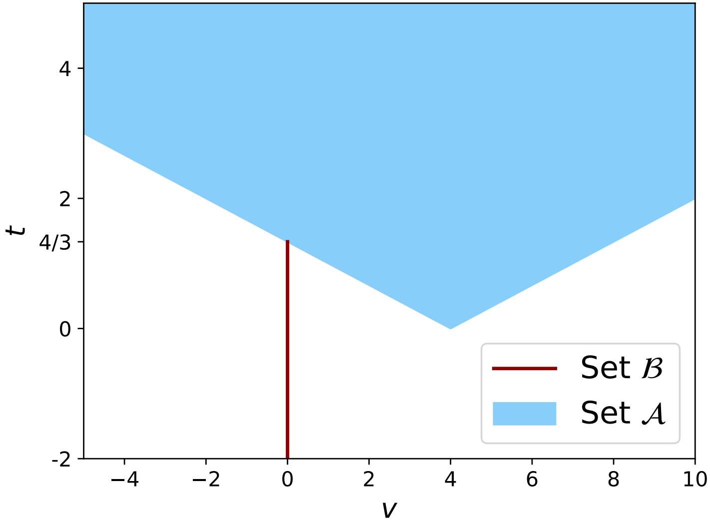
Another convex set

$$\mathcal{B} := \left\{ (\vec{0}, t) \mid t < p^* \right\}$$

$\mathcal{A}$ and $\mathcal{B}$ are disjoint

If $t \in \mathcal{A} \cap \mathcal{B}$ there exists $\vec{x}$ such that $f(\vec{x}) \leq t < p^*$

Example



Separating hyperplane

There exists a hyperplane separating $\mathcal{A}$ and $\mathcal{B}$

There exists $\vec{w} \in \mathbb{R}^m$ and $z \in \mathbb{R}$ such that

$$\vec{w}^T \vec{v} + zt \geq q \quad \text{for all } (\vec{v}, t) \in \mathcal{A}$$

$$\vec{w}^T \vec{v} + zt \leq q \quad \text{for all } (\vec{v}, t) \in \mathcal{B}$$

Assume $z > 0$ ($z < 0$ is impossible, argument for $z = 0$ is similar)

Separating hyperplane

$$z^{-1}\vec{w}^T\vec{v} + t \geq z^{-1}q \quad \text{for all } (\vec{v}, t) \in \mathcal{A}$$

$$z^{-1}\vec{w}^T\vec{v} + t \leq z^{-1}q \quad \text{for all } (\vec{v}, t) \in \mathcal{B}$$

Separating hyperplane

$$z^{-1}\vec{w}^T\vec{v} + t \le z^{-1}q \quad \text{for all } (\vec{v}, t) \in \mathcal{B}$$

implies

$$p^* \le z^{-1}q$$

Separating hyperplane

$$z^{-1}\vec{w}^T\vec{v} + t \geq z^{-1}q \quad \text{for all } (\vec{v}, t) \in \mathcal{A}$$

implies

$$\mathcal{L}(z^{-1}\vec{w}, \vec{x}) = f(\vec{x}) + z^{-1}\vec{w}^T(\vec{b} - A\vec{x})$$

Separating hyperplane

$$z^{-1}\vec{w}^T\vec{v} + t \geq z^{-1}q \quad \text{for all } (\vec{v}, t) \in \mathcal{A}$$

implies

$$\begin{aligned} \mathcal{L}(z^{-1}\vec{w}, \vec{x}) &= f(\vec{x}) + z^{-1}\vec{w}^T(\vec{b} - A\vec{x}) \\ &= z^{-1}\vec{w}^T\vec{v} + t \end{aligned}$$

Separating hyperplane

$$z^{-1}\vec{w}^T\vec{v} + t \geq z^{-1}q \quad \text{for all } (\vec{v}, t) \in \mathcal{A}$$

implies

$$\begin{aligned} \mathcal{L}(z^{-1}\vec{w}, \vec{x}) &= f(\vec{x}) + z^{-1}\vec{w}^T(\vec{b} - A\vec{x}) \\ &= z^{-1}\vec{w}^T\vec{v} + t \\ &\geq z^{-1}q \end{aligned}$$

Separating hyperplane

$$z^{-1}\vec{w}^T\vec{v} + t \geq z^{-1}q \quad \text{for all } (\vec{v}, t) \in \mathcal{A}$$

implies

$$\begin{aligned}
\mathcal{L}(z^{-1}\vec{w}, \vec{x}) &= f(\vec{x}) + z^{-1}\vec{w}^T(\vec{b} - A\vec{x}) \\
&= z^{-1}\vec{w}^T\vec{v} + t \\
&\geq z^{-1}q \\
&\geq p^* \quad \text{for all } \vec{x}!
\end{aligned}$$

Separating hyperplane

$$z^{-1}\vec{w}^T\vec{v} + t \geq z^{-1}q \quad \text{for all } (\vec{v}, t) \in \mathcal{A}$$

implies

$$\begin{aligned}
\mathcal{L}(z^{-1}\vec{w}, \vec{x}) &= f(\vec{x}) + z^{-1}\vec{w}^T(\vec{b} - A\vec{x}) \\
&= z^{-1}\vec{w}^T\vec{v} + t \\
&\geq z^{-1}q \\
&\geq p^* \quad \text{for all } \vec{x}!
\end{aligned}$$

$$p^* \geq d^*$$

Separating hyperplane

$$z^{-1}\vec{w}^T\vec{v} + t \geq z^{-1}q \quad \text{for all } (\vec{v}, t) \in \mathcal{A}$$

implies

$$\begin{aligned}\mathcal{L}(z^{-1}\vec{w}, \vec{x}) &= f(\vec{x}) + z^{-1}\vec{w}^T(\vec{b} - A\vec{x}) \\ &= z^{-1}\vec{w}^T\vec{v} + t \\ &\geq z^{-1}q \\ &\geq p^* \quad \text{for all } \vec{x}!\end{aligned}$$

$$p^* \geq d^* \geq g(z^{-1}\vec{w})$$

Separating hyperplane

$$z^{-1}\vec{w}^T\vec{v} + t \geq z^{-1}q \quad \text{for all } (\vec{v}, t) \in \mathcal{A}$$

implies

$$\begin{aligned}\mathcal{L}(z^{-1}\vec{w}, \vec{x}) &= f(\vec{x}) + z^{-1}\vec{w}^T(\vec{b} - A\vec{x}) \\ &= z^{-1}\vec{w}^T\vec{v} + t \\ &\geq z^{-1}q \\ &\geq p^* \quad \text{for all } \vec{x}!\end{aligned}$$

$$p^* \geq d^* \geq g(z^{-1}\vec{w}) := \inf_{\vec{x}} \mathcal{L}(z^{-1}\vec{w}, \vec{x})$$

Separating hyperplane

$$z^{-1}\vec{w}^T\vec{v} + t \geq z^{-1}q \quad \text{for all } (\vec{v}, t) \in \mathcal{A}$$

implies

$$\begin{aligned}
\mathcal{L}(z^{-1}\vec{w}, \vec{x}) &= f(\vec{x}) + z^{-1}\vec{w}^T(\vec{b} - A\vec{x}) \\
&= z^{-1}\vec{w}^T\vec{v} + t \\
&\geq z^{-1}q \\
&\geq p^* \quad \text{for all } \vec{x}!
\end{aligned}$$

$$p^* \geq d^* \geq g(z^{-1}\vec{w}) := \inf_{\vec{x}} \mathcal{L}(z^{-1}\vec{w}, \vec{x}) \geq p^*$$

Motivating applications

Convex sets

Lagrangian duality

Proof of strong duality

Compressed sensing

Matrix completion

Compressed sensing

Goal: Recovering signals from small number of data

Arbitrary vector of dimension d cannot be recovered from $m < d$ linear measurements

However, signals of interest are highly structured

For example, images are sparse in wavelet basis

If signal is parametrized by $s < m$ parameters, recovery may be possible

We focus on simplified problem: recovering sparse vectors

Exact recovery

Let $\mathbf{A} \in \mathbb{R}^{m \times d}$ have iid standard Gaussian entries

Let $\vec{x}_{\text{true}} \in \mathbb{R}^d$ have s nonzero entries

If $\mathbf{A}\vec{x}_{\text{true}} = \vec{y}$, then $\vec{x}_{\text{true}}$ is the unique solution of the problem

$$\min_{\vec{x} \in \mathbb{R}^d} ||\vec{x}||_1 \qquad \text{subject to} \quad A\vec{x} = \vec{y}$$

with probability at least $1 - \frac{1}{d}$ as long as

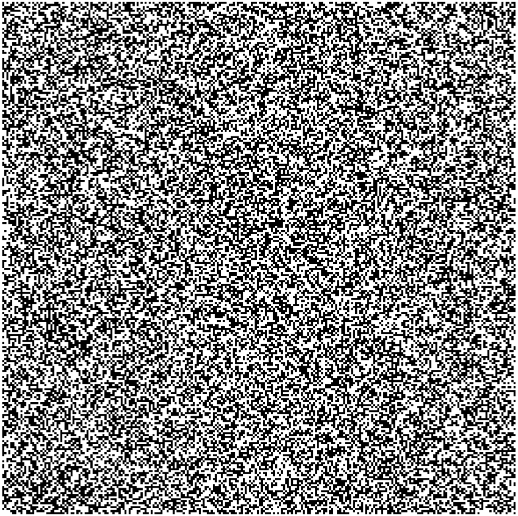
$$m \geq C s \log d$$

Sparsity in a transform domain

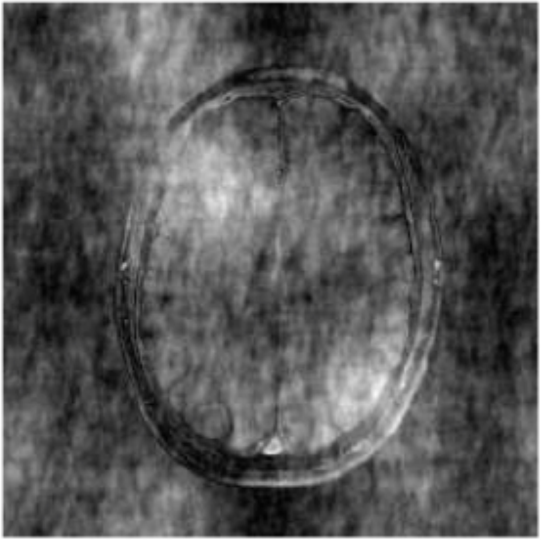
If $\vec{x}$ is sparse in the wavelet domain, $\vec{x}_{\text{true}} = W\vec{c}_{\text{true}}$

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

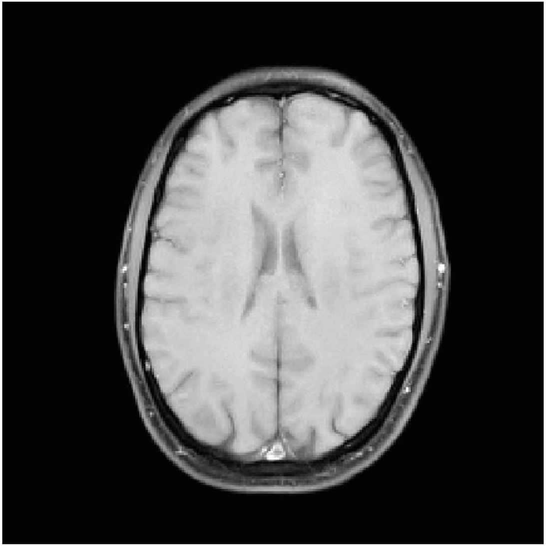
Undersampling pattern



Direct reconstruction



Min. ℓ_1-norm estimate (wavelet coefficients)



How to prove exact recovery

Need to prove that no other $\vec{x}$ such that $A\vec{x} = \vec{y}$ has smaller ℓ_1 norm than $\vec{x}_{\text{true}}$

Idea: Use duality

How to prove exact recovery

Assume there exists a feasible vector $\vec{\alpha}'$ for the dual

$$\max_{\vec{\alpha} \in \mathbb{R}^m} \langle \vec{\alpha}, \vec{y} \rangle \qquad \text{subject to } \left|\left| A^T \vec{\alpha} \right|\right|_\infty \leq 1$$

such that

$$\left|\left| \vec{x}_{\text{true}} \right|\right|_1 = \langle \vec{\alpha}', \vec{y} \rangle$$

How to prove exact recovery

Assume there exists a feasible vector $\vec{\alpha}'$ for the dual

$$\max_{\vec{\alpha} \in \mathbb{R}^m} \langle \vec{\alpha}, \vec{y} \rangle \qquad \text{subject to } \left|\left| A^T \vec{\alpha} \right|\right|_\infty \leq 1$$

such that

$$||\vec{x}_{\text{true}}||_1 = \langle \vec{\alpha}', \vec{y} \rangle$$

then by weak duality, for any feasible $\vec{x}$

$$\begin{aligned} ||\vec{x}||_1 &\geq \langle \vec{\alpha}', \vec{y} \rangle \\ &= ||\vec{x}_{\text{true}}||_1 \end{aligned}$$

Proof strategy

Show that for any sparse vector $\vec{x}_{\text{true}}$, there exists $\vec{\alpha}\,'$ such that $||\vec{x}_{\text{true}}||_1 = \langle \vec{\alpha}\,', \vec{y}\,\rangle$

ℓ_1-norm minimization

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

$$\max_{\vec{\alpha} \in \mathbb{R}^m} \langle \vec{\alpha}, \vec{b} \rangle \qquad \text{subject to} \; \left|\left| A^T \vec{\alpha} \right|\right|_\infty \leq 1$$

satisfies

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

for any solution $\vec{x}^*$ to the primal problem

Proof strategy

Show that for any sparse vector $\vec{x}_{\text{true}}$, there exists $\vec{\alpha}'$ such that

$$\left|\left|A^T \vec{\alpha}'\right|\right|_\infty \leq 1$$

$$(A^T \vec{\alpha}')[i] = \text{sign}(\vec{x}_{\text{true}}[i]) \quad \text{for all } \vec{x}_{\text{true}}[i] \neq 0$$

Subdifferential of ℓ_1 norm

$\vec{g}$ is a subgradient of the ℓ_1 norm at $\vec{x} \in \mathbb{R}^n$ if and only if

$$\vec{g}[i] = \operatorname{sign}(x[i]) \qquad \text{if } x[i] \neq 0$$

$$|\vec{g}[i]| \leq 1 \qquad \text{if } \vec{x}[i] = 0$$

Proof strategy

Show that for any sparse vector $\vec{x}_{\text{true}}$, there exists $\vec{\alpha}\,'$ such that

$$\left|\left|A^T\vec{\alpha}\,'\right|\right|_\infty \leq 1$$

$$(A^T\vec{\alpha}\,')[i] = \text{sign}(\vec{x}_{\text{true}}[i]) \quad \text{for all } \vec{x}_{\text{true}}[i] \neq 0$$

Proof strategy

Show that for any sparse vector $\vec{x}_{\text{true}}$, there exists $\vec{\alpha}\,'$ such that

$$\left|\left|A^T \vec{\alpha}\,'\right|\right|_\infty \leq 1$$

$$(A^T \vec{\alpha}\,')[i] = \text{sign}(\vec{x}_{\text{true}}[i]) \quad \text{for all } \vec{x}_{\text{true}}[i] \neq 0$$

Alternative justification: $\vec{g} := A^T \vec{\alpha}\,'$ is a subgradient of the ℓ_1 norm at $\vec{x}_{\text{true}}$ so for any $\vec{x}$ such that $A\vec{x} = \vec{y}$

$$||\vec{x}||_1 \geq ||\vec{x}_{\text{true}}||_1 + \langle \vec{g}, \vec{x} - \vec{x}_{\text{true}} \rangle$$

Proof strategy

Show that for any sparse vector $\vec{x}_{\text{true}}$, there exists $\vec{\alpha}\,'$ such that

$$\left|\left|A^T \vec{\alpha}\,'\right|\right|_\infty \leq 1$$

$$(A^T \vec{\alpha}\,')[i] = \text{sign}(\vec{x}_{\text{true}}[i]) \quad \text{for all } \vec{x}_{\text{true}}[i] \neq 0$$

Alternative justification: $\vec{g} := A^T \vec{\alpha}\,'$ is a subgradient of the ℓ_1 norm at $\vec{x}_{\text{true}}$ so for any $\vec{x}$ such that $A\vec{x} = \vec{y}$

$$\begin{aligned} ||\vec{x}||_1 &\geq ||\vec{x}_{\text{true}}||_1 + \langle \vec{g}, \vec{x} - \vec{x}_{\text{true}} \rangle \\ &= ||\vec{x}_{\text{true}}||_1 + \langle A^T \vec{\alpha}\,', \vec{x} - \vec{x}_{\text{true}} \rangle \end{aligned}$$

Proof strategy

Show that for any sparse vector $\vec{x}_{\text{true}}$, there exists $\vec{\alpha}\,'$ such that

$$\left|\left|A^T\vec{\alpha}\,'\right|\right|_\infty \leq 1$$

$$(A^T\vec{\alpha}\,')[i] = \text{sign}(\vec{x}_{\text{true}}[i]) \quad \text{for all } \vec{x}_{\text{true}}[i] \neq 0$$

Alternative justification: $\vec{g} := A^T\vec{\alpha}\,'$ is a subgradient of the ℓ_1 norm at $\vec{x}_{\text{true}}$ so for any $\vec{x}$ such that $A\vec{x} = \vec{y}$

$$\begin{aligned} ||\vec{x}||_1 &\geq ||\vec{x}_{\text{true}}||_1 + \langle \vec{g}, \vec{x} - \vec{x}_{\text{true}} \rangle \\ &= ||\vec{x}_{\text{true}}||_1 + \langle A^T\vec{\alpha}\,', \vec{x} - \vec{x}_{\text{true}} \rangle \\ &= ||\vec{x}_{\text{true}}||_1 + \langle \vec{\alpha}\,', A(\vec{x} - \vec{x}_{\text{true}}) \rangle \end{aligned}$$

Proof strategy

Show that for any sparse vector $\vec{x}_{\text{true}}$, there exists $\vec{\alpha}\,'$ such that

$$\left|\left|A^T\vec{\alpha}\,'\right|\right|_\infty \leq 1$$

$$(A^T\vec{\alpha}\,')[i] = \text{sign}(\vec{x}_{\text{true}}[i]) \quad \text{for all } \vec{x}_{\text{true}}[i] \neq 0$$

Alternative justification: $\vec{g} := A^T\vec{\alpha}\,'$ is a subgradient of the ℓ_1 norm at $\vec{x}_{\text{true}}$ so for any $\vec{x}$ such that $A\vec{x} = \vec{y}$

$$\begin{aligned}
||\vec{x}||_1 &\geq ||\vec{x}_{\text{true}}||_1 + \langle \vec{g}, \vec{x} - \vec{x}_{\text{true}} \rangle \\
&= ||\vec{x}_{\text{true}}||_1 + \langle A^T\vec{\alpha}\,', \vec{x} - \vec{x}_{\text{true}} \rangle \\
&= ||\vec{x}_{\text{true}}||_1 + \langle \vec{\alpha}\,', A(\vec{x} - \vec{x}_{\text{true}}) \rangle \\
&= ||\vec{x}_{\text{true}}||_1 + \langle \vec{\alpha}\,', \vec{y} - \vec{y} \rangle
\end{aligned}$$

Proof strategy

Show that for any sparse vector $\vec{x}_{\text{true}}$, there exists $\vec{\alpha}\,'$ such that

$$\left|\left|A^T\vec{\alpha}\,'\right|\right|_{\infty} \leq 1$$

$$(A^T\vec{\alpha}\,')[i] = \text{sign}(\vec{x}_{\text{true}}[i]) \quad \text{for all } \vec{x}_{\text{true}}[i] \neq 0$$

Alternative justification: $\vec{g} := A^T\vec{\alpha}\,'$ is a subgradient of the ℓ_1 norm at $\vec{x}_{\text{true}}$ so for any $\vec{x}$ such that $A\vec{x} = \vec{y}$

$$\begin{aligned}
||\vec{x}||_1 &\geq ||\vec{x}_{\text{true}}||_1 + \langle \vec{g}, \vec{x} - \vec{x}_{\text{true}} \rangle \\
&= ||\vec{x}_{\text{true}}||_1 + \langle A^T\vec{\alpha}\,', \vec{x} - \vec{x}_{\text{true}} \rangle \\
&= ||\vec{x}_{\text{true}}||_1 + \langle \vec{\alpha}\,', A(\vec{x} - \vec{x}_{\text{true}}) \rangle \\
&= ||\vec{x}_{\text{true}}||_1 + \langle \vec{\alpha}\,', \vec{y} - \vec{y} \rangle \\
&= ||\vec{x}_{\text{true}}||_1
\end{aligned}$$

Dual certificate for ℓ_1-norm minimization

For $\vec{x}_{\text{true}} \in \mathbb{R}^d$ with support $\mathcal{S}$ such that $A\vec{x}_{\text{true}} = \vec{y}$

Assume the submatrix $A_{\mathcal{S}}$ is full rank

If there exists $\vec{\alpha}_{\text{cert}} \in \mathbb{R}^m$ such that $\vec{g}_{\text{cert}} := A^T \vec{\alpha}_{\text{cert}}$ satisfies

$$\vec{g}_{\text{cert}}[i] = \text{sign}(\vec{x}_{\text{true}}[i]) \qquad \text{if } \vec{x}_{\text{true}}[i] \neq 0 \tag{1}$$

$$|\vec{g}_{\text{cert}}[i]| < 1 \qquad \text{if } \vec{x}_{\text{true}}[i] = 0 \tag{2}$$

then $\vec{x}_{\text{true}}$ is the unique solution to the ℓ_1-norm minimization problem

Proof

For any feasible $\vec{x} \in \mathbb{R}^d$, let $\vec{h} := \vec{x} - \vec{x}_{\text{true}}$ (so $A\vec{h} = \vec{0}$)

Proof

For any feasible $\vec{x} \in \mathbb{R}^d$, let $\vec{h} := \vec{x} - \vec{x}_{\text{true}}$ (so $A\vec{h} = \vec{0}$)

If $A_{\mathcal{S}}$ is full rank $\vec{h}_{\mathcal{S}^c} \neq 0$ unless $\vec{h} = \vec{0}$

Proof

For any feasible $\vec{x} \in \mathbb{R}^d$, let $\vec{h} := \vec{x} - \vec{x}_{\text{true}}$ (so $A\vec{h} = \vec{0}$)

If $A_{\mathcal{S}}$ is full rank $\vec{h}_{\mathcal{S}^c} \neq 0$ unless $\vec{h} = \vec{0}$

The strict inequality implies

$$\left|\left|\vec{h}_{\mathcal{S}^c}\right|\right|_1 > \vec{g}_{\text{cert}}^{\,T} \vec{h}_{\mathcal{S}^c}$$

Proof

For any feasible $\vec{x} \in \mathbb{R}^d$, let $\vec{h} := \vec{x} - \vec{x}_{\text{true}}$ (so $A\vec{h} = \vec{0}$)

If $A_{\mathcal{S}}$ is full rank $\vec{h}_{\mathcal{S}^c} \neq 0$ unless $\vec{h} = \vec{0}$

The strict inequality implies

$$\left|\left|\vec{h}_{\mathcal{S}^c}\right|\right|_1 > \vec{g}_{\text{cert}}^{\,T}\vec{h}_{\mathcal{S}^c}$$

Then

$$||\vec{x}||_1 = \left|\left|\vec{x}_{\text{true}} + \mathcal{P}_{\mathcal{S}}\left(\vec{h}\right)\right|\right|_1 + \left|\left|\vec{h}_{\mathcal{S}^c}\right|\right|_1 \quad \text{because } \vec{x}_{\text{true}} \text{ is supported on } \mathcal{S}$$

Proof

For any feasible $\vec{x} \in \mathbb{R}^d$, let $\vec{h} := \vec{x} - \vec{x}_{\text{true}}$ (so $A\vec{h} = \vec{0}$)

If $A_{\mathcal{S}}$ is full rank $\vec{h}_{\mathcal{S}^c} \neq 0$ unless $\vec{h} = \vec{0}$

The strict inequality implies

$$\left|\left|\vec{h}_{\mathcal{S}^c}\right|\right|_1 > \vec{g}_{\text{cert}}^{\,T} \vec{h}_{\mathcal{S}^c}$$

Then

$$\begin{aligned} ||\vec{x}||_1 &= \left|\left|\vec{x}_{\text{true}} + \mathcal{P}_{\mathcal{S}}\left(\vec{h}\right)\right|\right|_1 + \left|\left|\vec{h}_{\mathcal{S}^c}\right|\right|_1 \quad \text{because } \vec{x}_{\text{true}} \text{ is supported on } \mathcal{S} \\ &> ||\vec{x}_{\text{true}}||_1 + \vec{g}_{\text{cert}}^{\,T} \mathcal{P}_{\mathcal{S}}\left(\vec{h}\right) + \vec{g}_{\text{cert}}^{\,T} \mathcal{P}_{\mathcal{S}^c}\left(\vec{h}\right) \\ &> ||\vec{x}_{\text{true}}||_1 + \vec{g}_{\text{cert}}^{\,T} \vec{h} \end{aligned}$$

Proof

For any feasible $\vec{x} \in \mathbb{R}^d$, let $\vec{h} := \vec{x} - \vec{x}_{\text{true}}$ (so $A\vec{h} = \vec{0}$)

If $A_{\mathcal{S}}$ is full rank $\vec{h}_{\mathcal{S}^c} \neq 0$ unless $\vec{h} = \vec{0}$

The strict inequality implies

$$\left|\left|\vec{h}_{\mathcal{S}^c}\right|\right|_1 > \vec{g}_{\text{cert}}^{\,T}\vec{h}_{\mathcal{S}^c}$$

Then

$$\begin{aligned}
||\vec{x}||_1 &= \left|\left|\vec{x}_{\text{true}} + \mathcal{P}_{\mathcal{S}}\left(\vec{h}\right)\right|\right|_1 + \left|\left|\vec{h}_{\mathcal{S}^c}\right|\right|_1 \quad \text{because } \vec{x}_{\text{true}} \text{ is supported on } \mathcal{S} \\
&> ||\vec{x}_{\text{true}}||_1 + \vec{g}_{\text{cert}}^{\,T}\mathcal{P}_{\mathcal{S}}\left(\vec{h}\right) + \vec{g}_{\text{cert}}^{\,T}\mathcal{P}_{\mathcal{S}^c}\left(\vec{h}\right) \\
&> ||\vec{x}_{\text{true}}||_1 + \vec{g}_{\text{cert}}^{\,T}\vec{h} \\
&= ||\vec{x}_{\text{true}}||_1 + \left(A^T\vec{\alpha}_{\text{cert}}\right)^T\vec{h}
\end{aligned}$$

Proof

For any feasible $\vec{x} \in \mathbb{R}^d$, let $\vec{h} := \vec{x} - \vec{x}_{\text{true}}$ (so $A\vec{h} = \vec{0}$)

If $A_{\mathcal{S}}$ is full rank $\vec{h}_{\mathcal{S}^c} \neq 0$ unless $\vec{h} = \vec{0}$

The strict inequality implies

$$\left|\left|\vec{h}_{\mathcal{S}^c}\right|\right|_1 > \vec{g}_{\text{cert}}^{\,T} \vec{h}_{\mathcal{S}^c}$$

Then

$$\begin{aligned}
||\vec{x}||_1 &= \left|\left|\vec{x}_{\text{true}} + \mathcal{P}_{\mathcal{S}}\left(\vec{h}\right)\right|\right|_1 + \left|\left|\vec{h}_{\mathcal{S}^c}\right|\right|_1 \quad \text{because } \vec{x}_{\text{true}} \text{ is supported on } \mathcal{S} \\
&> ||\vec{x}_{\text{true}}||_1 + \vec{g}_{\text{cert}}^{\,T} \mathcal{P}_{\mathcal{S}}\left(\vec{h}\right) + \vec{g}_{\text{cert}}^{\,T} \mathcal{P}_{\mathcal{S}^c}\left(\vec{h}\right) \\
&> ||\vec{x}_{\text{true}}||_1 + \vec{g}_{\text{cert}}^{\,T} \vec{h} \\
&= ||\vec{x}_{\text{true}}||_1 + \left(A^T \vec{\alpha}_{\text{cert}}\right)^T \vec{h} \\
&= ||\vec{x}_{\text{true}}||_1 + \vec{\alpha}_{\text{cert}}^{\,T} A \vec{h}
\end{aligned}$$

Proof

For any feasible $\vec{x} \in \mathbb{R}^d$, let $\vec{h} := \vec{x} - \vec{x}_{\text{true}}$ (so $A\vec{h} = \vec{0}$)

If $A_{\mathcal{S}}$ is full rank $\vec{h}_{\mathcal{S}^c} \neq 0$ unless $\vec{h} = \vec{0}$

The strict inequality implies

$$\left|\left|\vec{h}_{\mathcal{S}^c}\right|\right|_1 > \vec{g}_{\text{cert}}^{\,T}\vec{h}_{\mathcal{S}^c}$$

Then

$$\begin{aligned}
||\vec{x}||_1 &= \left|\left|\vec{x}_{\text{true}} + \mathcal{P}_{\mathcal{S}}\left(\vec{h}\right)\right|\right|_1 + \left|\left|\vec{h}_{\mathcal{S}^c}\right|\right|_1 \quad \text{because } \vec{x}_{\text{true}} \text{ is supported on } \mathcal{S} \\
&> ||\vec{x}_{\text{true}}||_1 + \vec{g}_{\text{cert}}^{\,T}\mathcal{P}_{\mathcal{S}}\left(\vec{h}\right) + \vec{g}_{\text{cert}}^{\,T}\mathcal{P}_{\mathcal{S}^c}\left(\vec{h}\right) \\
&> ||\vec{x}_{\text{true}}||_1 + \vec{g}_{\text{cert}}^{\,T}\vec{h} \\
&= ||\vec{x}_{\text{true}}||_1 + \left(A^T\vec{\alpha}_{\text{cert}}\right)^T\vec{h} \\
&= ||\vec{x}_{\text{true}}||_1 + \vec{\alpha}_{\text{cert}}^{\,T}A\vec{h} \\
&= ||\vec{x}_{\text{true}}||_1
\end{aligned}$$

Proof of exact recovery

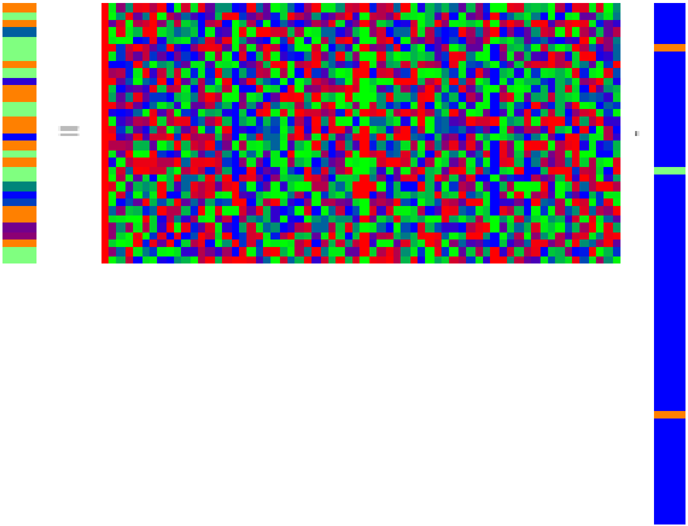
Goal: Build dual certificate

Interpolate sign pattern using vector in row space of $\mathbf{A}$

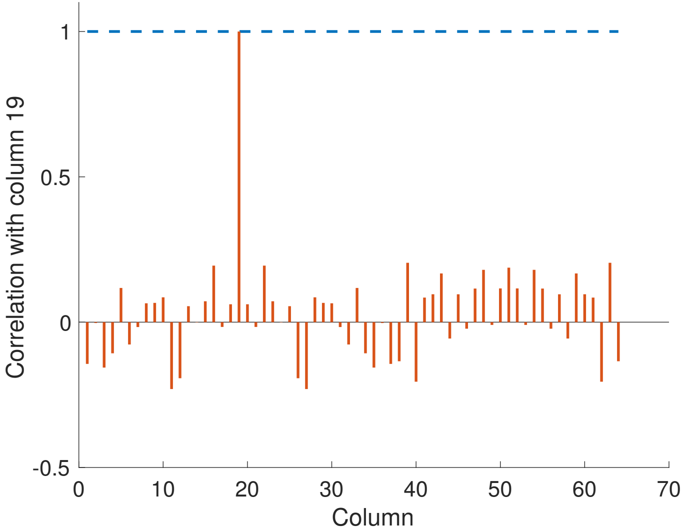
Consider correlation vector

$$\vec{\mathbf{c}}_i := \mathbf{A}^T \mathbf{A}_i, \quad 1 \leq i \leq m$$

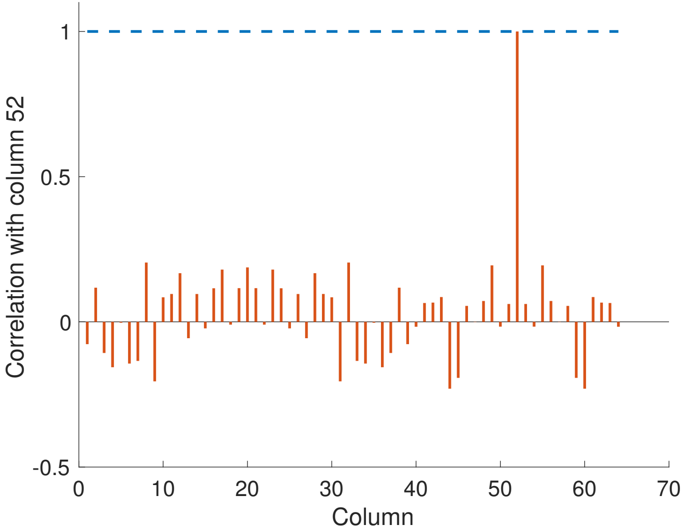
x2 random undersampling (Fourier)



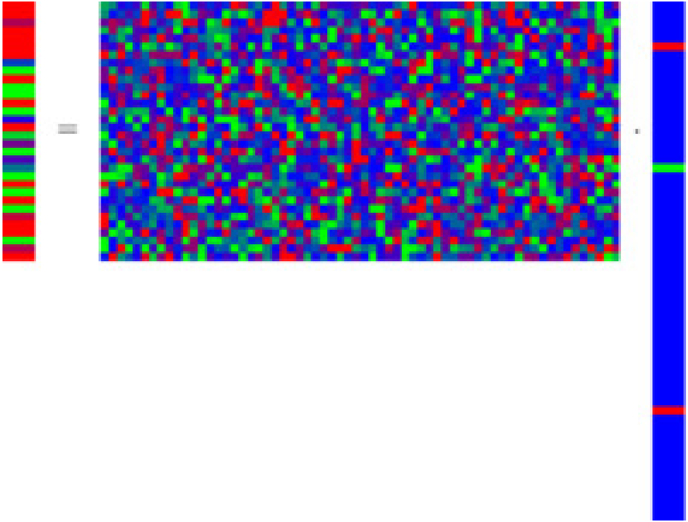
x2 random undersampling (Fourier)



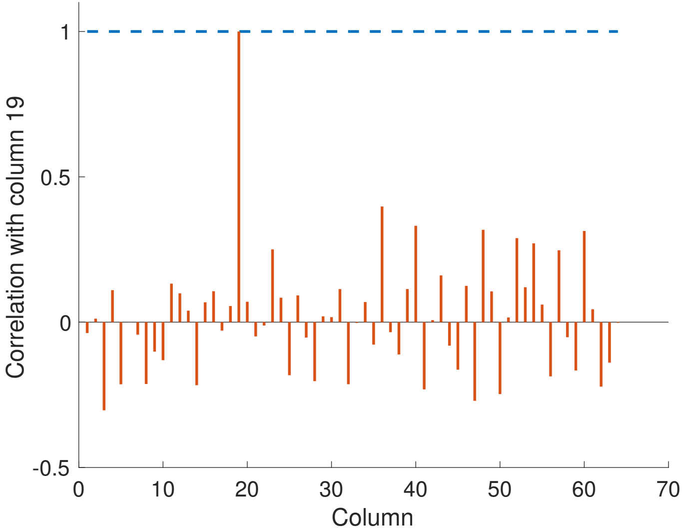
x2 random undersampling (Fourier)



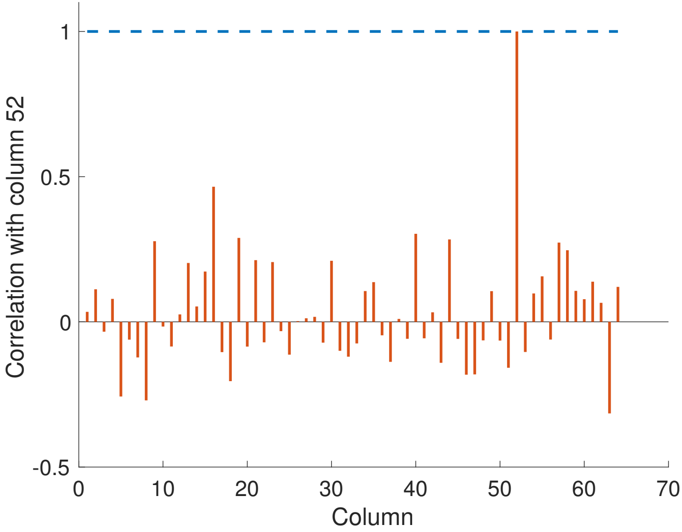
x2 random Gaussian measurements



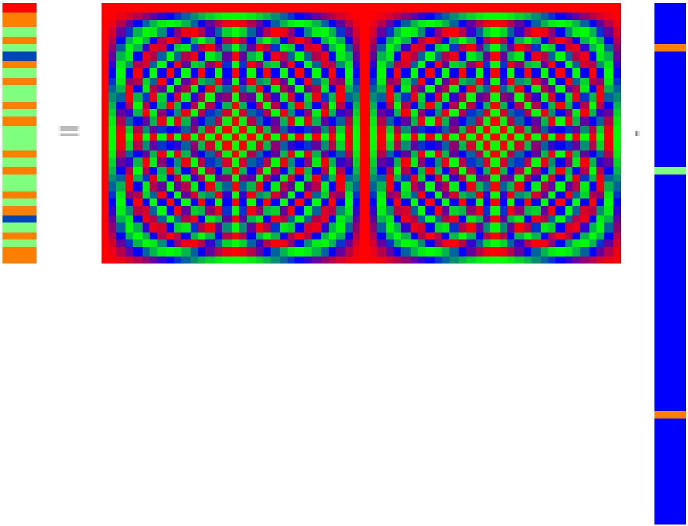
x2 random Gaussian measurements



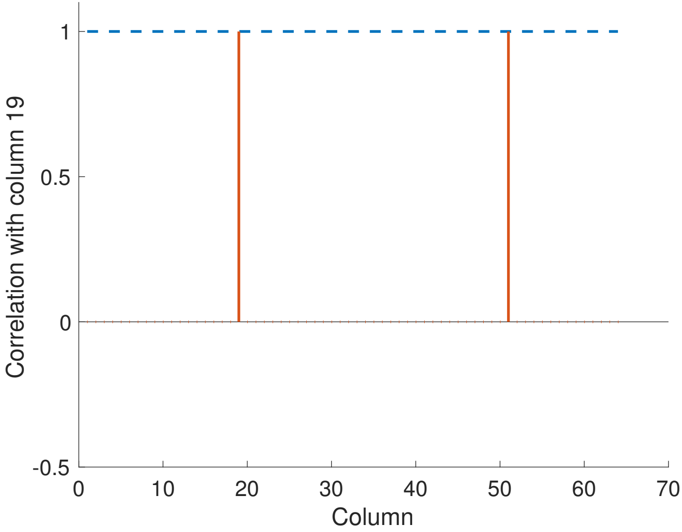
x2 random Gaussian measurements



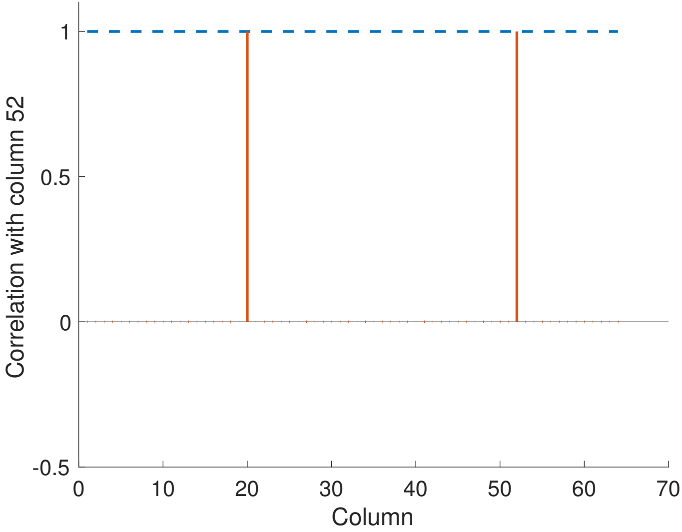
x2 regular undersampling (Fourier)



x2 regular undersampling (Fourier)



x2 regular undersampling (Fourier)



Proof of exact recovery

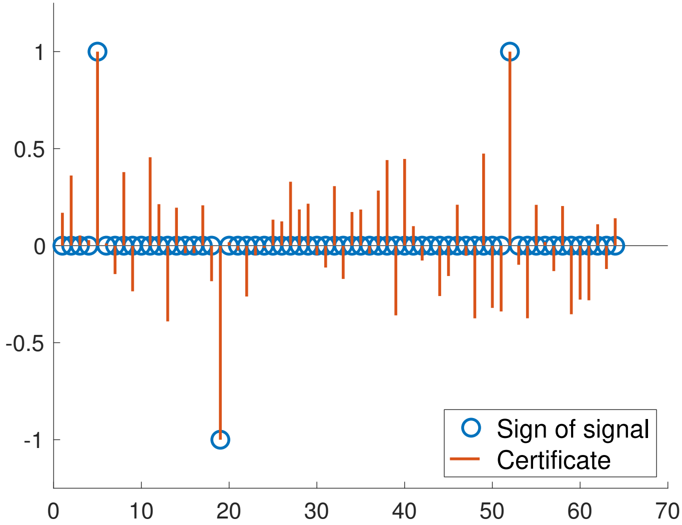
Idea: Use correlation vectors to interpolate

$$\vec{\mathbf{g}}_{\text{cert}} := \sum_{i \in \mathcal{S}} \mathbf{w}_i \vec{\mathbf{c}}_i$$

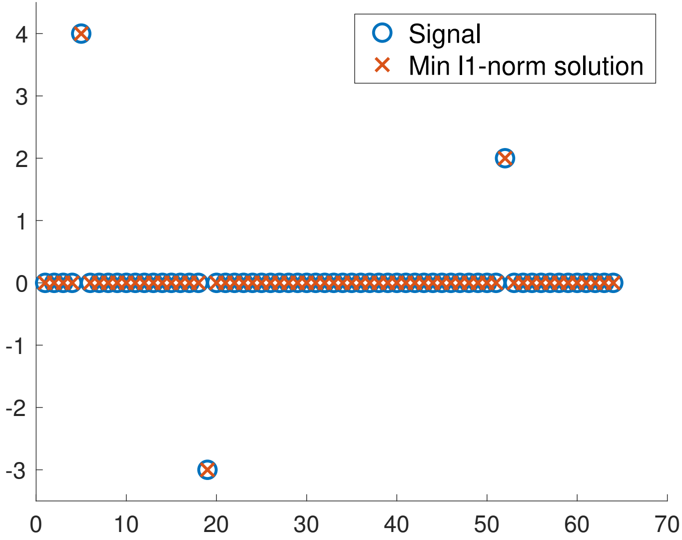
where weights $\mathbf{w}_i$, $i \in \mathcal{S}$ are set so that for all $j \in \mathcal{S}$

$$\operatorname{sign}\left(\vec{x}_{\text{true}}\right)[j] = \vec{\mathbf{g}}_{\text{cert}}[j]$$

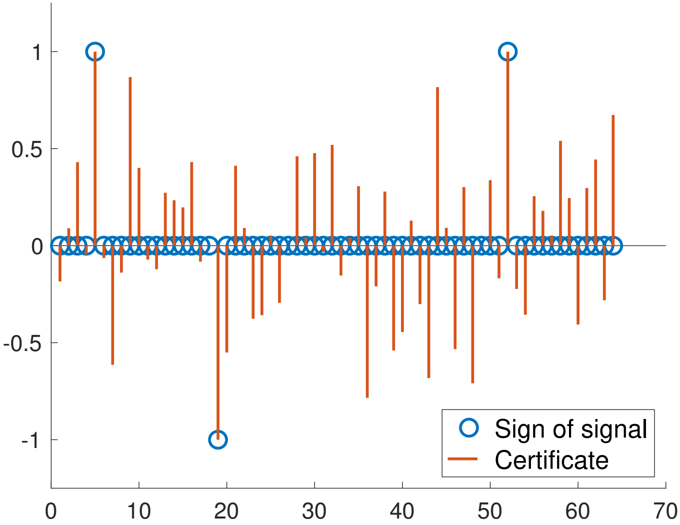
x2 random undersampling (Fourier)



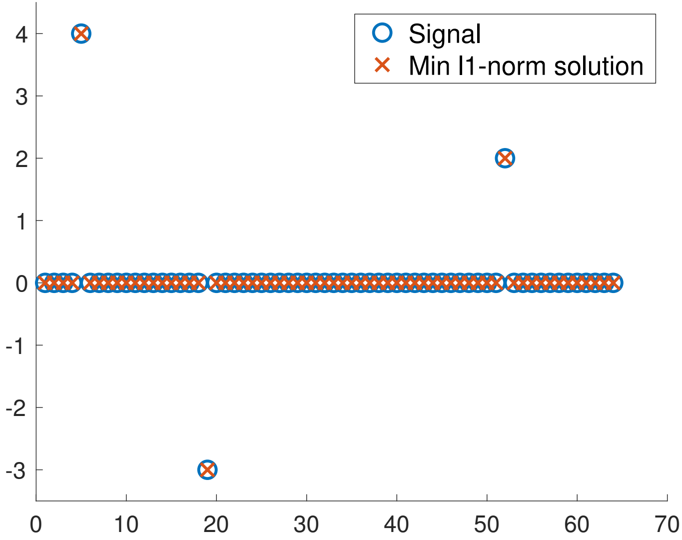
x2 random undersampling (Fourier)



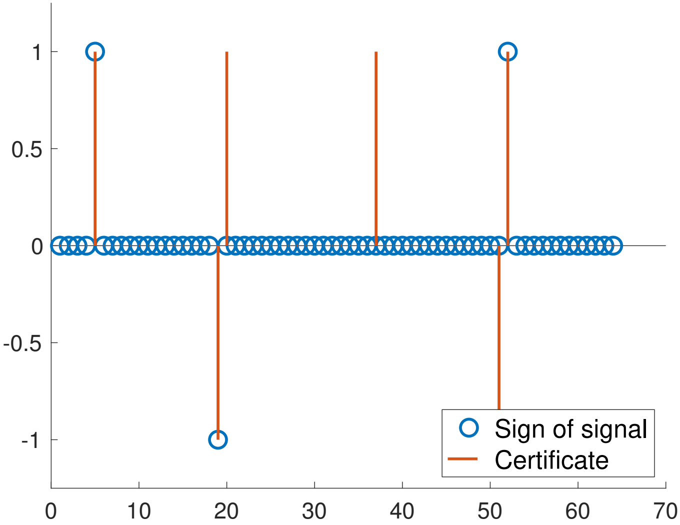
x2 random Gaussian measurements



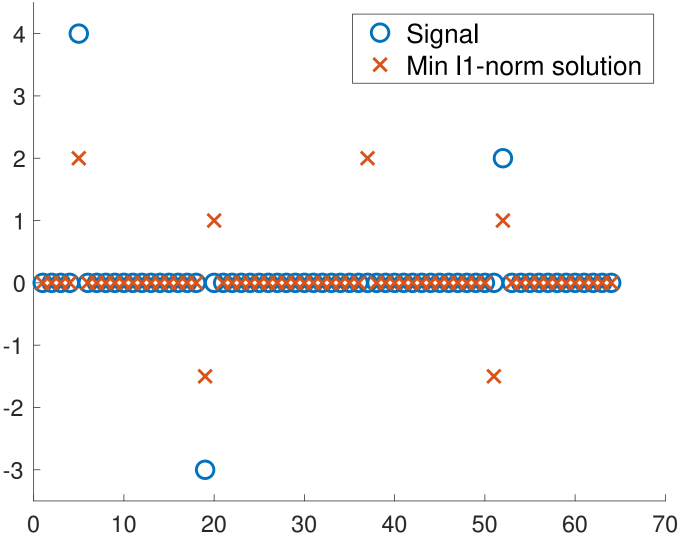
x2 random Gaussian measurements



x2 regular undersampling (Fourier)



x2 regular undersampling (Fourier)



Proof of exact recovery

Challenge: Analyzing certificate for all sign patterns

$$\text{sign}\,(\vec{x}_{\text{true}})_{\mathcal{S}} = (\sum_{i \in \mathcal{S}} \mathbf{w}_i \vec{\mathbf{c}}_i)_{\mathcal{S}}$$

Proof of exact recovery

Challenge: Analyzing certificate for all sign patterns

$$\text{sign}\,(\vec{x}_{\text{true}})_{\mathcal{S}} = (\sum_{i \in \mathcal{S}} \mathbf{w}_i \vec{\mathbf{c}}_i)_{\mathcal{S}}$$
$$= \sum_{i \in \mathcal{S}} \mathbf{w}_i \mathbf{A}_{\mathcal{S}}^T \mathbf{A}_i$$

Proof of exact recovery

Challenge: Analyzing certificate for all sign patterns

$$\begin{aligned}\operatorname{sign}(\vec{x}_{\text{true}})_{\mathcal{S}} &= \Big(\sum_{i\in\mathcal{S}} \mathbf{w}_i \vec{\mathbf{c}}_i\Big)_{\mathcal{S}} \\ &= \sum_{i\in\mathcal{S}} \mathbf{w}_i \mathbf{A}_{\mathcal{S}}^T \mathbf{A}_i \\ &= \mathbf{A}_{\mathcal{S}}^T \mathbf{A}_{\mathcal{S}} \vec{\mathbf{w}}\end{aligned}$$

Proof of exact recovery

Challenge: Analyzing certificate for all sign patterns

$$\begin{aligned}\operatorname{sign}(\vec{x}_{\text{true}})_{\mathcal{S}} &= \Big(\sum_{i\in\mathcal{S}} \mathbf{w}_i \vec{\mathbf{c}}_i\Big)_{\mathcal{S}} \\ &= \sum_{i\in\mathcal{S}} \mathbf{w}_i \mathbf{A}_{\mathcal{S}}^T \mathbf{A}_i \\ &= \mathbf{A}_{\mathcal{S}}^T \mathbf{A}_{\mathcal{S}} \vec{\mathbf{w}}\end{aligned}$$

Solving for $\vec{\mathbf{w}}$ yields

$$\vec{\mathbf{w}} := \left(\mathbf{A}_{\mathcal{S}}^T \mathbf{A}_{\mathcal{S}}\right)^{-1} \operatorname{sign}(\vec{x}_{\text{true}})_{\mathcal{S}}$$

Certificate candidate

$$\vec{\mathbf{g}}_{\text{cert}} = \sum_{i \in \mathcal{S}} \mathbf{w}_i \vec{\mathbf{c}}_i$$

Certificate candidate

$$\vec{\mathbf{g}}_{\text{cert}} = \sum_{i \in \mathcal{S}} \mathbf{w}_i \vec{\mathbf{c}}_i$$
$$= \mathbf{A}^T \mathbf{A}_{\mathcal{S}} \vec{\mathbf{w}}_{\text{cert}}$$

Certificate candidate

$$
\begin{aligned}
\vec{\mathbf{g}}_{\text{cert}} &= \sum_{i \in \mathcal{S}} \mathbf{w}_i \vec{\mathbf{c}}_i \\
&= \mathbf{A}^T \mathbf{A}_{\mathcal{S}} \vec{\mathbf{w}}_{\text{cert}} \\
&= \mathbf{A}^T \mathbf{A}_{\mathcal{S}} \left(\mathbf{A}_{\mathcal{S}}^T \mathbf{A}_{\mathcal{S}} \right)^{-1} \operatorname{sign} \left(\vec{x}_{\text{true}} \right)_{\mathcal{S}}
\end{aligned}
$$

Is $\mathbf{A}_{\mathcal{S}}^T \mathbf{A}_{\mathcal{S}}$ invertible?

Certificate candidate

$$\begin{aligned}\vec{\mathbf{g}}_{\text{cert}} &= \sum_{i \in \mathcal{S}} \mathbf{w}_i \vec{\mathbf{c}}_i \\ &= \mathbf{A}^T \mathbf{A}_{\mathcal{S}} \vec{\mathbf{w}}_{\text{cert}} \\ &= \mathbf{A}^T \mathbf{A}_{\mathcal{S}} \left(\mathbf{A}_{\mathcal{S}}^T \mathbf{A}_{\mathcal{S}} \right)^{-1} \operatorname{sign}\left(\vec{x}_{\text{true}}\right)_{\mathcal{S}}\end{aligned}$$

Is $\mathbf{A}_{\mathcal{S}}^T \mathbf{A}_{\mathcal{S}}$ invertible?

What about

$$|\vec{g}_{\text{cert}}[i]| < 1 \qquad \text{if } \vec{x}_{\text{true}}[i] = 0 \quad ?$$

Singular values of a Gaussian matrix

Let $\mathbf{M}$ be a $m \times s$ matrix with iid standard Gaussian entries such that $m > s$

For any fixed $\epsilon > 0$, the singular values of $\mathbf{M}$ satisfy

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

with probability at least $1 - 2\left(\frac{12}{\epsilon}\right)^s \exp\left(-\frac{m\epsilon^2}{32}\right)$

Singular values of $A_{\mathcal{S}}$

$\boldsymbol{\sigma}_s$ is the smallest singular value of $A_{\mathcal{S}}$

Setting $\epsilon := 0.5$, let $\mathcal{E}$ denote the event that

$$0.5\sqrt{m} \leq \boldsymbol{\sigma_s} \leq \boldsymbol{\sigma_1} \leq 1.5\sqrt{m}.$$

Singular values of $A_{\mathcal{S}}$

$\boldsymbol{\sigma}_s$ is the smallest singular value of $A_{\mathcal{S}}$

Setting $\epsilon := 0.5$, let $\mathcal{E}$ denote the event that

$$0.5\sqrt{m} \leq \boldsymbol{\sigma}_{\mathbf{s}} \leq \boldsymbol{\sigma}_{\mathbf{1}} \leq 1.5\sqrt{m}.$$

For a constant C'

$$\mathrm{P}\left(\mathcal{E}\right) \geq 1 - \exp\left(-C'\frac{m}{s}\right)$$

Singular values of $A_{\mathcal{S}}$

$\boldsymbol{\sigma}_s$ is the smallest singular value of $A_{\mathcal{S}}$

Setting $\epsilon := 0.5$, let $\mathcal{E}$ denote the event that

$$0.5\sqrt{m} \leq \boldsymbol{\sigma}_{\mathbf{s}} \leq \boldsymbol{\sigma}_{\mathbf{1}} \leq 1.5\sqrt{m}.$$

For a constant C'

$$\mathrm{P}\left(\mathcal{E}\right) \geq 1 - \exp\left(-C'\frac{m}{s}\right)$$

Conditioned on $\mathcal{E}$ $\mathbf{A}_{\mathcal{S}}$ is full rank, so $\mathbf{A}_{\mathcal{S}}^T\mathbf{A}_{\mathcal{S}}$ is invertible

Singular values of $A_{\mathcal{S}}$

$\boldsymbol{\sigma}_s$ is the smallest singular value of $A_{\mathcal{S}}$

Setting $\epsilon := 0.5$, let $\mathcal{E}$ denote the event that

$$0.5\sqrt{m} \leq \boldsymbol{\sigma}_{\mathbf{s}} \leq \boldsymbol{\sigma}_{\mathbf{1}} \leq 1.5\sqrt{m}.$$

For a constant C'

$$\mathrm{P}\left(\mathcal{E}\right) \geq 1 - \exp\left(-C'\frac{m}{s}\right)$$

Conditioned on $\mathcal{E}$ $\mathbf{A}_{\mathcal{S}}$ is full rank, so $\mathbf{A}_{\mathcal{S}}^T \mathbf{A}_{\mathcal{S}}$ is invertible

$\vec{g}_{\text{cert}}$ interpolates the sign

What about $|\vec{g}_{\text{cert}}[i]| < 1$ on $\mathcal{S}^c$?

$$\begin{aligned}\vec{\boldsymbol{\alpha}}_{\text{cert}} &:= \mathbf{A}_{\mathcal{S}} \vec{\mathbf{w}}_{\text{cert}} \\ &= \mathbf{A}_{\mathcal{S}} \left(\mathbf{A}_{\mathcal{S}}^T \mathbf{A}_{\mathcal{S}} \right)^{-1} \text{sign}\left(\vec{x}_{\text{true}} \right)_{\mathcal{S}}\end{aligned}$$

$$\vec{\mathbf{g}}_{\text{cert}} = \mathbf{A}^T \vec{\boldsymbol{\alpha}}_{\text{cert}}$$

What about $|\vec{g}_{\text{cert}}[i]| < 1$ on $\mathcal{S}^c$?

$$\begin{aligned}\vec{\boldsymbol{\alpha}}_{\text{cert}} &:= \mathbf{A}_{\mathcal{S}} \vec{\mathbf{w}}_{\text{cert}} \\ &= \mathbf{A}_{\mathcal{S}} \left(\mathbf{A}_{\mathcal{S}}^T \mathbf{A}_{\mathcal{S}} \right)^{-1} \text{sign}\left(\vec{x}_{\text{true}}\right)_{\mathcal{S}}\end{aligned}$$

$$\vec{\mathbf{g}}_{\text{cert}} = \mathbf{A}^T \vec{\boldsymbol{\alpha}}_{\text{cert}}$$

Let $\mathbf{USV}^T$ be the SVD of $\mathbf{A}_{\mathcal{S}}$, conditioned on $\mathcal{E}$

$$||\vec{\boldsymbol{\alpha}}_{\text{cert}}||_2 = \left|\left|\mathbf{US}^{-1}\mathbf{V}^T \text{sign}\left(\vec{x}_{\text{true}}\right)_{\mathcal{S}}\right|\right|_2$$

What about $|\vec{g}_{\text{cert}}[i]| < 1$ on $\mathcal{S}^c$?

$$\begin{aligned}\vec{\boldsymbol{\alpha}}_{\text{cert}} &:= \mathbf{A}_{\mathcal{S}} \vec{\mathbf{w}}_{\text{cert}} \\ &= \mathbf{A}_{\mathcal{S}} \left(\mathbf{A}_{\mathcal{S}}^T \mathbf{A}_{\mathcal{S}} \right)^{-1} \operatorname{sign}\left(\vec{x}_{\text{true}}\right)_{\mathcal{S}}\end{aligned}$$

$$\vec{\mathbf{g}}_{\text{cert}} = \mathbf{A}^T \vec{\boldsymbol{\alpha}}_{\text{cert}}$$

Let $\mathbf{USV}^T$ be the SVD of $\mathbf{A}_{\mathcal{S}}$, conditioned on $\mathcal{E}$

$$\begin{aligned}\left|\left|\vec{\boldsymbol{\alpha}}_{\text{cert}}\right|\right|_2 &= \left|\left|\mathbf{US}^{-1}\mathbf{V}^T \operatorname{sign}\left(\vec{x}_{\text{true}}\right)_{\mathcal{S}}\right|\right|_2 \\ &\leq \frac{\left|\left|\operatorname{sign}\left(\vec{x}_{\text{true}}\right)_{\mathcal{S}}\right|\right|_2}{\boldsymbol{\sigma}_{\mathbf{s}}}\end{aligned}$$

What about $|\vec{g}_{\text{cert}}[i]| < 1$ on $\mathcal{S}^c$?

$$
\begin{aligned}
\vec{\boldsymbol{\alpha}}_{\text{cert}} &:= \mathbf{A}_{\mathcal{S}} \vec{\mathbf{w}}_{\text{cert}} \\
&= \mathbf{A}_{\mathcal{S}} \left(\mathbf{A}_{\mathcal{S}}^T \mathbf{A}_{\mathcal{S}} \right)^{-1} \operatorname{sign}\left(\vec{x}_{\text{true}}\right)_{\mathcal{S}}
\end{aligned}
$$

$$
\vec{\mathbf{g}}_{\text{cert}} = \mathbf{A}^T \vec{\boldsymbol{\alpha}}_{\text{cert}}
$$

Let $\mathbf{USV}^T$ be the SVD of $\mathbf{A}_{\mathcal{S}}$, conditioned on $\mathcal{E}$

$$
\begin{aligned}
\left|\left|\vec{\boldsymbol{\alpha}}_{\text{cert}}\right|\right|_2 &= \left|\left| \mathbf{US}^{-1}\mathbf{V}^T \operatorname{sign}\left(\vec{x}_{\text{true}}\right)_{\mathcal{S}} \right|\right|_2 \\
&\leq \frac{\left|\left|\operatorname{sign}\left(\vec{x}_{\text{true}}\right)_{\mathcal{S}}\right|\right|_2}{\boldsymbol{\sigma}_{\mathbf{s}}} \\
&\leq 2\sqrt{\frac{s}{m}}
\end{aligned}
$$

What about $|\vec{g}_{\text{cert}}[i]| < 1$ on $\mathcal{S}^c$?

$\mathbf{A}_i^T \vec{v} / \left|\left|\vec{v}\right|\right|_2$ is a standard Gaussian

What about $|\vec{g}_{\text{cert}}[i]| < 1$ on $\mathcal{S}^c$?

$\mathbf{A}_i^T \vec{v} / ||\vec{v}||_2$ is a standard Gaussian

For a standard Gaussian $\mathbf{u}$ and any $t > 0$

$$\mathrm{P}\left(|\mathbf{u}| \geq t\right) \leq 2 \exp\left(-\frac{t^2}{2}\right)$$

What about $|\vec{g}_{\text{cert}}[i]| < 1$ on $\mathcal{S}^c$?

$\mathbf{A}_i^T \vec{v} / \left|\left|\vec{v}\right|\right|_2$ is a standard Gaussian

For a standard Gaussian $\mathbf{u}$ and any $t > 0$

$$\mathrm{P}\left(|\mathbf{u}| \geq t\right) \leq 2 \exp\left(-\frac{t^2}{2}\right)$$

$$\begin{aligned} \mathrm{P}\left(\left|\mathbf{A}_i^T \vec{v}\right| \geq 1\right) &= \mathrm{P}\left(\frac{\left|\mathbf{A}_i^T \vec{v}\right|}{\left|\left|\vec{v}\right|\right|_2} \geq \frac{1}{\left|\left|\vec{v}\right|\right|_2}\right) \\ &\leq 2 \exp\left(-\frac{1}{2\left|\left|\vec{v}\right|\right|_2^2}\right) \end{aligned}$$

What about $|\vec{g}_{\text{cert}}[i]| < 1$ on $\mathcal{S}^c$?

$$\mathrm{P}\left(\left|\mathbf{A}_i^T \vec{\boldsymbol{\alpha}}_{\text{cert}}\right| \geq 1 \,|\, \mathcal{E}\right) = \mathrm{P}\left(\left|\mathbf{A}_i^T \vec{v}\right| \geq 1 \quad \text{for} \qquad ||\vec{v}||_2 \leq 2\sqrt{\frac{s}{m}}\right)$$

$$\leq 2 \exp\left(-\frac{m}{8s}\right)$$

What about $|\vec{g}_{\text{cert}}[i]| < 1$ on $\mathcal{S}^c$?

$$\mathrm{P}\left(\left|\mathbf{A}_i^T \vec{\boldsymbol{\alpha}}_{\text{cert}}\right| \geq 1 \,|\, \mathcal{E}\right) = \mathrm{P}\left(\left|\mathbf{A}_i^T \vec{v}\right| \geq 1 \quad \text{for} \qquad \|\vec{v}\|_2 \leq 2\sqrt{\frac{s}{m}}\right)$$
$$\leq 2\exp\left(-\frac{m}{8s}\right)$$

$$\mathrm{P}\left(\left|\mathbf{A}_i^T \vec{\boldsymbol{\alpha}}_{\text{cert}}\right| \geq 1\right) \leq \mathrm{P}\left(\left|\mathbf{A}_i^T \vec{\boldsymbol{\alpha}}_{\text{cert}}\right| \geq 1 \,|\, \mathcal{E}\right) + \mathrm{P}\left(\mathcal{E}^c\right)$$
$$\leq 2\exp\left(-\frac{m}{8s}\right) + \exp\left(-C'\frac{m}{s}\right)$$

What about $|\vec{g}_{\text{cert}}[i]| < 1$ on $\mathcal{S}^c$?

$$\mathrm{P}\left(\left|\mathbf{A}_i^T \vec{\alpha}_{\text{cert}}\right| \geq 1 \,|\, \mathcal{E}\right) = \mathrm{P}\left(\left|\mathbf{A}_i^T \vec{v}\right| \geq 1 \quad \text{for} \qquad ||\vec{v}||_2 \leq 2\sqrt{\frac{s}{m}}\right)$$

$$\leq 2 \exp\left(-\frac{m}{8s}\right)$$

$$\mathrm{P}\left(\left|\mathbf{A}_i^T \vec{\alpha}_{\text{cert}}\right| \geq 1\right) \leq \mathrm{P}\left(\left|\mathbf{A}_i^T \vec{\alpha}_{\text{cert}}\right| \geq 1 \,|\, \mathcal{E}\right) + \mathrm{P}\left(\mathcal{E}^c\right)$$

$$\leq 2 \exp\left(-\frac{m}{8s}\right) + \exp\left(-C' \frac{m}{s}\right)$$

By the union bound

$$\mathrm{P}\left(\bigcup_{i \in \mathcal{S}^c} \left\{\left|\mathbf{A}_i^T \vec{\alpha}_{\text{cert}}\right| \geq 1\right\}\right) \leq n\left(2 \exp\left(-\frac{m}{8s}\right) + \exp\left(-C' \frac{m}{s}\right)\right)$$

What about $|\vec{g}_{\text{cert}}[i]| < 1$ on $\mathcal{S}^c$?

$$\mathrm{P}\left(\left|\mathbf{A}_i^T \vec{\alpha}_{\text{cert}}\right| \geq 1 \mid \mathcal{E}\right) = \mathrm{P}\left(\left|\mathbf{A}_i^T \vec{v}\right| \geq 1 \quad \text{for} \qquad \|\vec{v}\|_2 \leq 2\sqrt{\frac{s}{m}}\right)$$
$$\leq 2\exp\left(-\frac{m}{8s}\right)$$

$$\mathrm{P}\left(\left|\mathbf{A}_i^T \vec{\alpha}_{\text{cert}}\right| \geq 1\right) \leq \mathrm{P}\left(\left|\mathbf{A}_i^T \vec{\alpha}_{\text{cert}}\right| \geq 1 \mid \mathcal{E}\right) + \mathrm{P}\left(\mathcal{E}^c\right)$$
$$\leq 2\exp\left(-\frac{m}{8s}\right) + \exp\left(-C'\frac{m}{s}\right)$$

By the union bound

$$\mathrm{P}\left(\bigcup_{i \in \mathcal{S}^c} \left\{\left|\mathbf{A}_i^T \vec{\alpha}_{\text{cert}}\right| \geq 1\right\}\right) \leq n\left(2\exp\left(-\frac{m}{8s}\right) + \exp\left(-C'\frac{m}{s}\right)\right)$$
$$\leq 1 - \frac{1}{n} \qquad \text{if } m \geq Cs\log n$$

Matrix completion as an inverse problem

$$\begin{bmatrix} 1 & ? & 5 \\ ? & 3 & 2 \end{bmatrix}$$

For a fixed sampling pattern, underdetermined system of equations

Matrix completion as an inverse problem

$$\begin{bmatrix} 1 & ? & 5 \\ ? & 3 & 2 \end{bmatrix}$$

For a fixed sampling pattern, underdetermined system of equations

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} Y_{11} \\ Y_{21} \\ Y_{12} \\ Y_{22} \\ Y_{13} \\ Y_{23} \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 5 \\ 2 \end{bmatrix}$$

Isn't this completely ill posed?

Assumption: Matrix is low rank, depends on $\approx r\,(m+n)$ parameters

As long as data $>$ parameters recovery is possible (in principle)

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

Can we complete this matrix by minimizing rank?

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

Measurements

We must see an entry in each row/column at least

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ ? & ? & ? & ? \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 \\ ? \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix}$$

Assumption: Random sampling (usually doesn't hold in practice!)

Can we complete this matrix from random samples?

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 23 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Can we complete this matrix from random samples?

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 2 & 3 & 4 & 5 \end{bmatrix}$$

Can we complete this matrix from random samples?

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 2 & 3 & 4 & 5 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 & 4 \end{bmatrix}$$

Incoherence

A matrix is incoherent if its singular vectors must be spread out

For $1/\sqrt{n} \leq \mu \leq 1$

$$\max_{1 \leq i \leq r, 1 \leq j \leq m} |U_{ij}| \leq \mu$$

$$\max_{1 \leq i \leq r, 1 \leq j \leq n} |V_{ij}| \leq \mu$$

for the left $U_1, \ldots, U_r$ and right $V_1, \ldots, V_r$ singular vectors

Common assumption in theoretical analysis

Nuclear-norm minimization for matrix completion

$\vec{y}$ contains the observed entries indexed by set Ω

$$\min_{X \in \mathbb{R}^{n_1 \times n_2}} ||X||_* \quad \text{such that } X_\Omega = \vec{y}$$

Challenge: Prove that this works

Nuclear-norm minimization for matrix completion

$\vec{y}$ contains the observed entries indexed by set Ω

$$\min_{X \in \mathbb{R}^{n_1 \times n_2}} ||X||_* \quad \text{such that } X_\Omega = \vec{y}$$

Challenge: Prove that this works

Use duality!

Norm minimization

The Lagrange dual function of

$$\min_{\vec{x} \in \mathbb{R}^n} ||\vec{x}|| \qquad \text{subject to } A\vec{x} = \vec{b}$$

where $A \in \mathbb{R}^{m \times n}$, $\vec{b} \in \mathbb{R}^m$, equals

$$\max_{\vec{\alpha} \in \mathbb{R}^m} \langle \vec{\alpha}, \vec{b} \rangle \qquad \text{subject to } \left|\left| A^T \vec{\alpha} \right|\right|_d \leq 1$$

$$||\vec{y}||_d := \max_{||\vec{x}|| \leq 1} \langle \vec{y}, \vec{x} \rangle$$

Dual of nuclear-norm minimization

Adjoint of operator $X \to X_\Omega$ is M_Ω

$M_\Omega(\vec{b})$ contains $\vec{b}$ in entries indexed by Ω and zeros elsewhere

Dual of nuclear-norm minimization

Adjoint of operator $X \to X_\Omega$ is M_Ω

$M_\Omega(\vec{b})$ contains $\vec{b}$ in entries indexed by Ω and zeros elsewhere

Proof: For any A and $\vec{b}$

$$\langle A_\Omega, \vec{b} \rangle = \langle A, M_\Omega(\vec{b}) \rangle$$

Dual norm of nuclear norm

For any matrix $A \in \mathbb{R}^{m \times n}$,

$$||A||_* = \max_{\{||B|| \leq 1 \mid B \in \mathbb{R}^{m \times n}\}} \langle A, B \rangle$$

$$||A||_d := \max_{||B||_* \leq 1} \langle A, B \rangle$$

Dual norm of nuclear norm

For any matrix $A \in \mathbb{R}^{m \times n}$,

$$||A||_* = \max_{\{||B|| \leq 1 \mid B \in \mathbb{R}^{m \times n}\}} \langle A, B \rangle$$

$$\begin{aligned} ||A||_d &:= \max_{||B||_* \leq 1} \langle A, B \rangle \\ &= ||A|| \max_{||B||_* \leq 1} \left\langle \frac{A}{||A||}, B \right\rangle \end{aligned}$$

Dual norm of nuclear norm

For any matrix $A \in \mathbb{R}^{m \times n}$,

$$||A||_* = \max_{\{||B|| \leq 1 \mid B \in \mathbb{R}^{m \times n}\}} \langle A, B \rangle$$

$$\begin{aligned} ||A||_d &:= \max_{||B||_* \leq 1} \langle A, B \rangle \\ &= ||A|| \max_{||B||_* \leq 1} \left\langle \frac{A}{||A||}, B \right\rangle \\ &= ||A|| \end{aligned}$$

Dual of nuclear-norm minimization

Let Ω be a subset of m entries, and $\vec{y} \in \mathbb{R}^m$

The dual of

$$\min_{X \in \mathbb{R}^{n_1 \times n_2}} ||X||_* \quad \text{such that } X_\Omega = \vec{y}$$

is

$$\max_{\vec{\alpha} \in \mathbb{R}^m} \langle \vec{\alpha}, \vec{y} \rangle \qquad \text{subject to } ||M_\Omega(\vec{\alpha})|| \leq 1$$

How to prove exact recovery

Assume there exists a feasible vector $\vec{\alpha}'$ for the dual

$$\max_{\vec{\alpha} \in \mathbb{R}^m} \langle \vec{\alpha}, \vec{y} \rangle \qquad \text{subject to } \ ||M_\Omega(\vec{\alpha})|| \leq 1$$

such that

$$||X_{\text{true}}||_* = \langle \vec{\alpha}', \vec{y} \rangle$$

How to prove exact recovery

Assume there exists a feasible vector $\vec{\alpha}\,'$ for the dual

$$\max_{\vec{\alpha} \in \mathbb{R}^m} \langle \vec{\alpha}, \vec{y} \rangle \qquad \text{subject to } \ ||M_\Omega(\vec{\alpha})|| \leq 1$$

such that

$$||X_{\text{true}}||_* = \langle \vec{\alpha}\,', \vec{y} \rangle$$

then by weak duality, for any feasible X

$$\begin{aligned} ||X||_* &\geq \langle \vec{\alpha}\,', \vec{y} \rangle \\ &= ||X_{\text{true}}||_* \end{aligned}$$

Analogy with ℓ_1 norm

- ℓ_1 norm $\rightarrow$ nuclear norm
- ℓ_∞ norm $\rightarrow$ operator norm
- UV^T $\rightarrow$ sign pattern of true sparse signal

How to prove exact recovery

Since

$$\begin{aligned} \langle \vec{\alpha}, \vec{y} \rangle &= \langle M_\Omega(\vec{\alpha}), M_\Omega(\vec{y}) \rangle \\ &= \langle M_\Omega(\vec{\alpha}), X_{\text{true}} \rangle \end{aligned}$$

for

$$||X_{\text{true}}||_* = \langle \vec{\alpha}, \vec{y}\, \rangle = \langle M_\Omega(\vec{\alpha}), X_{\text{true}} \rangle$$

$G := M_\Omega(\vec{\alpha})$ must be of the form

$$G := UV^T + W$$

where

$$||W|| \leq 1 \qquad U^T W = 0 \qquad W\, V = 0$$

How to prove exact recovery

$$\langle G, X_{\text{true}} \rangle = \langle UV^T + W, X_{\text{true}} \rangle$$

How to prove exact recovery

$$\begin{aligned}\langle G, X_{\text{true}} \rangle &= \langle UV^T + W, X_{\text{true}} \rangle \\ &= \text{tr}\left(X_{\text{true}}^T W + X_{\text{true}}^T UV^T \right)\end{aligned}$$

How to prove exact recovery

$$
\begin{aligned}
\langle G, X_{\text{true}} \rangle &= \langle UV^T + W, X_{\text{true}} \rangle \\
&= \text{tr}\left(X_{\text{true}}^T W + X_{\text{true}}^T UV^T \right) \\
&= \text{tr}\left(VSU^T W + VU^T USV^T \right)
\end{aligned}
$$

How to prove exact recovery

$$\begin{aligned}
\langle G, X_{\text{true}} \rangle &= \langle UV^T + W, X_{\text{true}} \rangle \\
&= \operatorname{tr}\left(X_{\text{true}}^T W + X_{\text{true}}^T UV^T \right) \\
&= \operatorname{tr}\left(VSU^T W + VU^T USV^T \right) \\
&= \operatorname{tr}(S)
\end{aligned}$$

How to prove exact recovery

$$
\begin{aligned}
\langle G, X_{\text{true}} \rangle &= \langle UV^T + W, X_{\text{true}} \rangle \\
&= \operatorname{tr}\left(X_{\text{true}}^T W + X_{\text{true}}^T UV^T \right) \\
&= \operatorname{tr}\left(VSU^T W + VU^T USV^T \right) \\
&= \operatorname{tr}\left(S \right) \\
&= ||X_{\text{true}}||_*
\end{aligned}
$$

Subdifferential of the nuclear norm

Let $X \in \mathbb{R}^{m \times n}$ be a rank-r matrix with SVD USV^T, where $U \in \mathbb{R}^{m \times r}$, $V \in \mathbb{R}^{n \times r}$ and $S \in \mathbb{R}^{r \times r}$

A matrix G is a subgradient of the nuclear norm at X if and only if

$$G := UV^T + W$$

where W satisfies

$$\begin{aligned} \|W\| &\leq 1 \\ U^T W &= 0 \\ W\,V &= 0 \end{aligned}$$

Alternative justification

$G := M_\Omega(\vec{\alpha})$ is a subgradient of the nuclear norm at X_{true}

For any X such that $X_\Omega = (X_{\text{true}})_\Omega$

$$||X||_* \geq ||X_{\text{true}}||_* + \langle X - X_{\text{true}}, G \rangle$$

Alternative justification

$G := M_\Omega(\vec{\alpha})$ is a subgradient of the nuclear norm at X_{true}

For any X such that $X_\Omega = (X_{\text{true}})_\Omega$

$$\begin{aligned} ||X||_* &\geq ||X_{\text{true}}||_* + \langle X - X_{\text{true}}, G \rangle \\ &= ||X_{\text{true}}||_* + \langle (X - X_{\text{true}})_\Omega, G_\Omega \rangle \end{aligned}$$

Alternative justification

$G := M_\Omega(\vec{\alpha})$ is a subgradient of the nuclear norm at X_{true}

For any X such that $X_\Omega = (X_{\text{true}})_\Omega$

$$\begin{aligned}||X||_* &\geq ||X_{\text{true}}||_* + \langle X - X_{\text{true}}, G \rangle \\ &= ||X_{\text{true}}||_* + \langle (X - X_{\text{true}})_\Omega, G_\Omega \rangle \\ &= ||X_{\text{true}}||_*\end{aligned}$$

Alternative justification

$G := M_\Omega(\vec{\alpha})$ is a subgradient of the nuclear norm at X_{true}

For any X such that $X_\Omega = (X_{\text{true}})_\Omega$

$$\begin{aligned}||X||_* &\geq ||X_{\text{true}}||_* + \langle X - X_{\text{true}}, G\rangle \\ &= ||X_{\text{true}}||_* + \langle (X - X_{\text{true}})_\Omega, G_\Omega\rangle \\ &= ||X_{\text{true}}||_*\end{aligned}$$

If $||W|| < 1$, under a certain constraint on sampling pattern, existence of G implies that X_{true} is the unique solution

Example

$$X_{\text{true}} := \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} a & b & b \end{bmatrix}$$

$$= \begin{bmatrix} a & b & b \\ a & b & b \\ a & b & b \end{bmatrix}, \quad a \in (0,1),\ b := \sqrt{\frac{1-a^2}{2}}.$$

Example

$$X_{\text{true}} := \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} a & b & b \end{bmatrix}$$
$$= \begin{bmatrix} a & b & b \\ a & b & b \\ a & b & b \end{bmatrix}, \quad a \in (0,1),\ b := \sqrt{\frac{1-a^2}{2}}.$$

USV^T of X_{true} is given by

$$U = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \qquad S = 1, \qquad V = \begin{bmatrix} a \\ b \\ b \end{bmatrix}.$$

Example

$$\begin{bmatrix} ? & b & b \\ a & ? & b \\ a & b & ? \end{bmatrix}$$

Question: For what values of a does nuclear-norm minimization work?

Strategy

Build

$$G := UV^T + W$$

supported on Ω such that

$$||W|| \leq 1 \qquad U^T W = 0 \qquad W\,V = 0$$

Dual certificate

G is supported on Ω so $G_{\Omega^c} = \vec{0}$ and

$$W_{\Omega^c} = -(UV^T)_{\Omega^c}$$

Dual certificate

G is supported on Ω so $G_{\Omega^c} = \vec{0}$ and

$$W_{\Omega^c} = -(UV^T)_{\Omega^c}$$

since

$$UV^T = \frac{1}{\sqrt{3}} \begin{bmatrix} a & b & b \\ a & b & b \\ a & b & b \end{bmatrix}$$

Dual certificate

G is supported on Ω so $G_{\Omega^c} = \vec{0}$ and

$$W_{\Omega^c} = -(UV^T)_{\Omega^c}$$

since

$$UV^T = \frac{1}{\sqrt{3}} \begin{bmatrix} a & b & b \\ a & b & b \\ a & b & b \end{bmatrix}$$

this implies

$$W = \frac{1}{\sqrt{3}} \begin{bmatrix} -a & w_3 & w_5 \\ w_1 & -b & w_6 \\ w_2 & w_4 & -b \end{bmatrix}$$

Dual certificate

$$W = \frac{1}{\sqrt{3}} \begin{bmatrix} -a & w_3 & w_5 \\ w_1 & -b & w_6 \\ w_2 & w_4 & -b \end{bmatrix} \qquad U = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \qquad S = 1, \qquad V = \begin{bmatrix} a \\ b \\ b \end{bmatrix}$$

$U^T W = 0$ and $W\, V = 0$ implies

Dual certificate

$$W = \frac{1}{\sqrt{3}} \begin{bmatrix} -a & w_3 & w_5 \\ w_1 & -b & w_6 \\ w_2 & w_4 & -b \end{bmatrix} \qquad U = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \qquad S = 1, \qquad V = \begin{bmatrix} a \\ b \\ b \end{bmatrix}$$

$U^T W = 0$ and $W\, V = 0$ implies

$$w_1 + w_2 = a$$

Dual certificate

$$W = \frac{1}{\sqrt{3}} \begin{bmatrix} -a & w_3 & w_5 \\ w_1 & -b & w_6 \\ w_2 & w_4 & -b \end{bmatrix} \qquad U = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \qquad S = 1, \qquad V = \begin{bmatrix} a \\ b \\ b \end{bmatrix}$$

$U^T W = 0$ and $W\ V = 0$ implies

$$w_1 + w_2 = a$$
$$w_3 + w_4 = b$$

Dual certificate

$$W = \frac{1}{\sqrt{3}} \begin{bmatrix} -a & w_3 & w_5 \\ w_1 & -b & w_6 \\ w_2 & w_4 & -b \end{bmatrix} \qquad U = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \qquad S = 1, \qquad V = \begin{bmatrix} a \\ b \\ b \end{bmatrix}$$

$U^T W = 0$ and $W\ V = 0$ implies

$$\begin{aligned} w_1 + w_2 &= a \\ w_3 + w_4 &= b \\ w_5 + w_6 &= b \end{aligned}$$

Dual certificate

$$W = \frac{1}{\sqrt{3}} \begin{bmatrix} -a & w_3 & w_5 \\ w_1 & -b & w_6 \\ w_2 & w_4 & -b \end{bmatrix} \qquad U = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \qquad S = 1, \qquad V = \begin{bmatrix} a \\ b \\ b \end{bmatrix}$$

$U^T W = 0$ and $W\ V = 0$ implies

$$\begin{aligned} w_1 + w_2 &= a \\ w_3 + w_4 &= b \\ w_5 + w_6 &= b \\ w_3 + w_5 &= \frac{a^2}{b} \end{aligned}$$

Dual certificate

$$W = \frac{1}{\sqrt{3}} \begin{bmatrix} -a & w_3 & w_5 \\ w_1 & -b & w_6 \\ w_2 & w_4 & -b \end{bmatrix} \qquad U = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \qquad S = 1, \qquad V = \begin{bmatrix} a \\ b \\ b \end{bmatrix}$$

$U^T W = 0$ and $W\,V = 0$ implies

$$\begin{aligned} w_1 + w_2 &= a \\ w_3 + w_4 &= b \\ w_5 + w_6 &= b \\ w_3 + w_5 &= \frac{a^2}{b} \\ aw_1 + bw_6 &= b^2 \end{aligned}$$

Dual certificate

$$W = \frac{1}{\sqrt{3}} \begin{bmatrix} -a & w_3 & w_5 \\ w_1 & -b & w_6 \\ w_2 & w_4 & -b \end{bmatrix} \qquad U = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \qquad S = 1, \qquad V = \begin{bmatrix} a \\ b \\ b \end{bmatrix}$$

$U^T W = 0$ and $W\ V = 0$ implies

$$\begin{aligned} w_1 + w_2 &= a \\ w_3 + w_4 &= b \\ w_5 + w_6 &= b \\ w_3 + w_5 &= \frac{a^2}{b} \\ aw_1 + bw_6 &= b^2 \\ aw_2 + bw_4 &= b^2 \end{aligned}$$

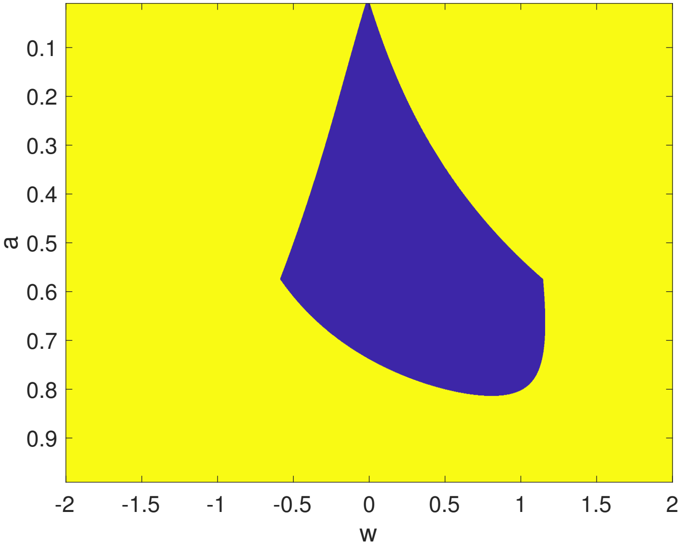
Dual certificate

Equations are dependent, fixing $w_1 := w$

$$W = \frac{1}{\sqrt{3}} \begin{bmatrix} -a & a - \frac{wb}{a} & \frac{wb}{a} \\ w & -b & b - w \\ \frac{a^2}{b} - w & b - \frac{a^2}{b} + w & -b \end{bmatrix},$$

Nuclear-norm minimization works if $||W|| < 1$ *for any* w

In the blue region $||W|| < 1$



Results

Nuclear-norm minimization fails if singular vector is too spiky

For example, if $a = 0.82$ ($b = 0.4047$) the solution is

$$X^* := \begin{bmatrix} 0.8095 & 0.82 & 0.82 \\ 0.4047 & 0.4047 & 0.4047 \\ 0.4047 & 0.4047 & 0.4047 \end{bmatrix},$$

where $||X^*||_* = 1.7320 < 1.7321 = ||X_{\text{true}}||_*$