Safe *squeezing* for antisparse coding

Clément Elvira — joint work with Cédric Herzet

CentraleSupélec, L2S, Inverse Problem Group (GPI)
4 Octobre 2019

Antisparse coding

Inverse / Learning problems

Given

- Acquisition matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ $\quad m < n$
- Observation $\mathbf{y} \simeq \mathbf{A}\mathbf{x}_0 \in \mathbb{R}^m$

Goal: recover $\mathbf{x}_0$ from

$$\arg\min_{\mathbf{x} \in \mathbb{R}^n} \frac{1}{2} \|\boldsymbol{y} - \mathbf{A}\mathbf{x}\|_2^2$$

Infinite number of solutions ☹

⟶ **ill posed problem**

Penalized problem

Penalized problem

$$\mathbf{x}^\star \in \arg\min_{\mathbf{x}\in\mathbb{R}^n} \frac{1}{2}\|\boldsymbol{y} - \mathbf{A}\mathbf{x}\|_2^2 + R_{\text{eg}}(\mathbf{x})$$

The choice of R_{eg} **should**

- reduce the number of solutions
- favor solutions with **desirable** properties
- allow for **fast** algorithms

Penalized problem

Penalized problem

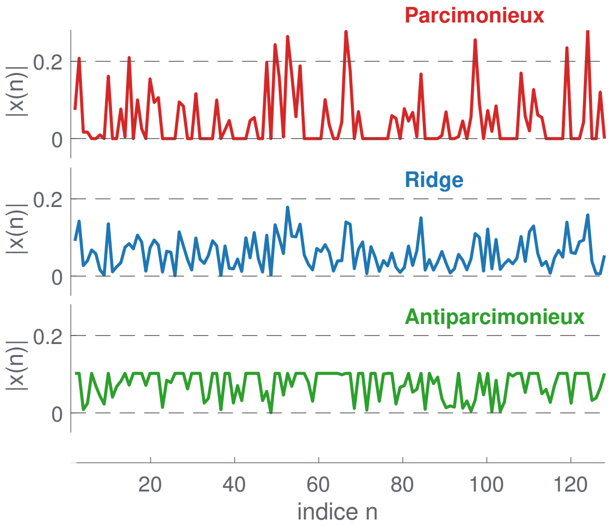
$$\mathbf{x}^\star \in \arg\min_{\mathbf{x}\in\mathbb{R}^n} \frac{1}{2}\|\boldsymbol{y} - \mathbf{A}\mathbf{x}\|_2^2 + R_{\text{eg}}(\mathbf{x})$$

The choice of R_{eg} **should**

- reduce the number of solutions
- favor solutions with **desirable** properties
- allow for **fast** algorithms

Popular choice of $R_{\text{eg}} \longrightarrow$ **convex** function

From sparse coding to antisparse coding



$R_{eg}(\mathbf{x}) = \lambda\|\mathbf{x}\|_1$
$\Rightarrow$ sparsity

$R_{eg}(\mathbf{x}) = \lambda\|\mathbf{x}\|_2^2$
$\Rightarrow$ energy

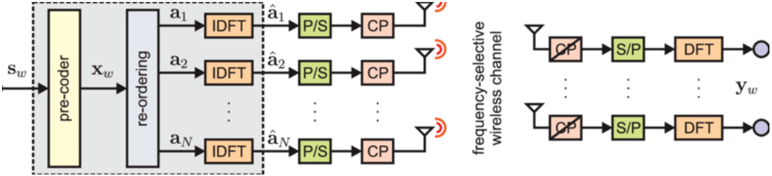
$R_{eg}(\mathbf{x}) = \lambda\|\mathbf{x}\|_\infty$
$\Rightarrow$ amplitude

Application 1 / 3

- Peak to Average Power Ratio (PAPR) reduction

Studer & Larsson (2013)

$$\forall \mathbf{x} \in \mathbb{R}^n, \quad \mathrm{PAPR}(\mathbf{x}) = \frac{n\|\mathbf{x}\|_\infty^2}{\|\mathbf{x}\|_2^2}$$



Courtesy of Studer and Larsson

$$\mathbf{s}_w = \mathbf{H}_w \mathbf{x}_w \qquad \mathbf{y}_w = \mathbf{H}_w \mathbf{x}_w + \mathbf{n}_w$$

Application 2 / 3

- **Robotic:** Uniform power allocation *Cadzow (1971)*

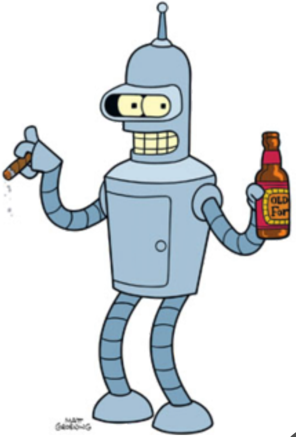
- Cinematic redundant system
- Uniform spread of electric power

$$\mathbf{y} = \mathbf{A} \begin{pmatrix} x_0^{(1)} \\ \vdots \\ x_p^{(1)} \\ \vdots \\ x_0^{(t)} \\ \vdots \\ x_p^{(t)} \end{pmatrix}$$

- **Robotic:** Uniform power allocation *Cadzow (1971)*

- Cinematic redundant system
- Uniform spread of electric power

$$\mathbf{y} = \mathbf{A} \begin{pmatrix} x_0^{(1)} \\ \vdots \\ x_p^{(1)} \\ \vdots \\ x_0^{(t)} \\ \vdots \\ x_p^{(t)} \end{pmatrix}$$

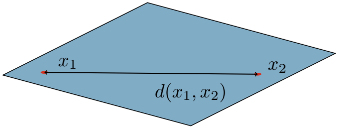


Application 3 / 3

ML: Approximate Nearest Neighbor search

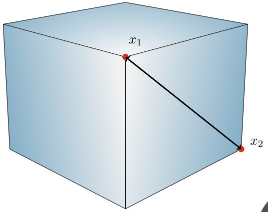
Jegou, Furon and Fuchs (2012)

Idea: Learn a **higher dimensional** representation



- $x(i) = \pm\alpha$
  $\Longrightarrow$ binarization + **privacy**

- binary distance = **XOR**
  $\Longrightarrow$ **faster**

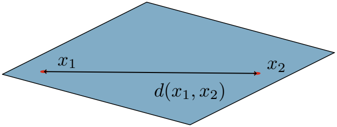


Application 3 / 3

ML: Approximate Nearest Neighbor search

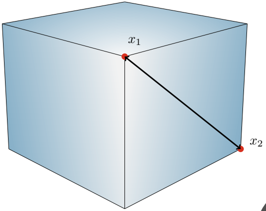
Jegou, Furon and Fuchs (2012)

Idea: Learn a **higher dimensional** representation



- $x(i) = \pm\alpha$
  $\Longrightarrow$ binarization + **privacy**

- binary distance = **XOR**
  $\Longrightarrow$ **faster**



Safe squeezing

for

antisparse coding

Computing antisparse representation

Optimization problem $$\mathbf{x}^\star \in \operatorname*{arg\,min}_{\mathbf{x}\in\mathbb{R}^n} \frac{1}{2}\|\boldsymbol{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_\infty$$

$\longrightarrow$ Convex, coercive

Optimization methods

Computing antisparse representation

Optimization problem $$\mathbf{x}^\star \in \underset{\mathbf{x}\in\mathbb{R}^n}{\arg\min}\ \frac{1}{2}\|\boldsymbol{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_\infty$$

$\longrightarrow$ Convex, coercive

Optimization methods

- **Heuristic to match the optimality conditions** *[Fuchs, 2011]*
  - **add / remove** entries from the **set** of **saturated** entries

Computing antisparse representation

Optimization problem $\mathbf{x}^\star \in \arg\min_{\mathbf{x}\in\mathbb{R}^n} \frac{1}{2}\|\boldsymbol{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_\infty$

$\longrightarrow$ Convex, coercive

Optimization methods

- **Heuristic to match the optimality conditions** *[Fuchs, 2011]*
  - **add / remove** entries from the **set** of **saturated** entries
- **Proximal Gradient: FITRA** *[Studer & Larsson, 2013]*
  - $\mathbf{x}^{(t+1)} = \mathrm{prox}_{\lambda\|\cdot\|_\infty}(\mathbf{x}^{(t)} - \alpha\nabla f(\mathbf{x}^{(t)}))$

Computing antisparse representation

Optimization problem $\mathbf{x}^\star \in \arg\min_{\mathbf{x}\in\mathbb{R}^n} \frac{1}{2}\|\boldsymbol{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_\infty$

$\longrightarrow$ Convex, coercive

Optimization methods

- **Heuristic to match the optimality conditions** *[Fuchs, 2011]*
  - **add / remove** entries from the **set** of **saturated** entries
- **Proximal Gradient: FITRA** *[Studer & Larsson, 2013]*
  - $\mathbf{x}^{(t+1)} = \mathrm{prox}_{\lambda\|\cdot\|_\infty}(\mathbf{x}^{(t)} - \alpha\nabla f(\mathbf{x}^{(t)}))$
- **Bayesian framework** *[Elvira et al., 2017]*
  - **Democratic prior** $\mathrm{p}(\mathbf{x}) \propto \exp\left(-\lambda\|\mathbf{x}\|_\infty\right)$
  - Gibbs sampler / Proximal MCMC

Connections with inverse problems involving sparsity

Lasso Find $\mathbf{x}^\star \in \arg\min_{\mathbf{x}\in\mathbb{R}^n} \frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_1$

$\longrightarrow$ Promotes **sparsity**

Unused feature

Safe **screening**

[El ghaoui et al., 2010] [Fercoq et al., 2015] [Fraga-Dantas et al., 2018] [Dorfler et al., 2019]

Connections with inverse problems involving sparsity

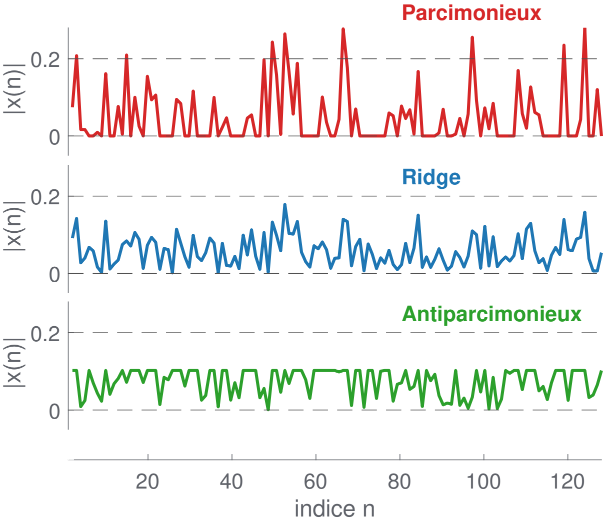
Lasso Find $\mathbf{x}^\star \in \underset{\mathbf{x}\in\mathbb{R}^n}{\arg\min} \; \frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_1$

$\longrightarrow$ Promotes **sparsity**

Unused feature $\longleftrightarrow$ **Saturation**

Safe **screening** $\longleftrightarrow$ Safe **squeezing**

From sparse coding to antisparse coding



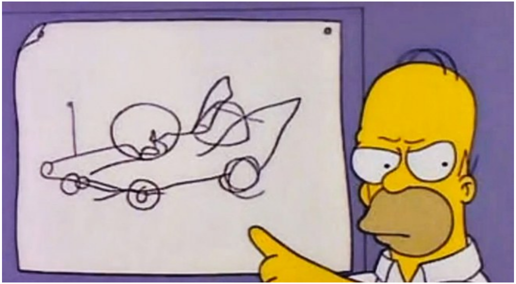
$R_{eg}(\mathbf{x}) = \lambda\|\mathbf{x}\|_1$
$\Rightarrow$ sparsity

$R_{eg}(\mathbf{x}) = \lambda\|\mathbf{x}\|_2^2$
$\Rightarrow$ energy

$R_{eg}(\mathbf{x}) = \lambda\|\mathbf{x}\|_\infty$
$\Rightarrow$ amplitude

Take home message

- It is possible to **dynamically** detect **saturated** entries
- It leads to consider an equivalent **lower dimensional** problem
- It provides **faster** algorithm at (almost) **no** additional **cost**
- It is experimentally validated



Notions of saturation

Recall $$\mathbf{x}^\star \in \arg\min_{\mathbf{x}\in\mathbb{R}^n} \frac{1}{2}\|\boldsymbol{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_\infty$$

Definition: **saturated** entry

entry i is saturated iff $\quad \mathbf{x}^\star(i) = \pm\|\mathbf{x}^\star\|_\infty$

Notions of saturation

Recall $$\mathbf{x}^\star \in \arg\min_{\mathbf{x}\in\mathbb{R}^n} \frac{1}{2}\|\boldsymbol{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_\infty$$

Definition: **saturated** entry

entry i is saturated iff $\mathbf{x}^\star(i) = \pm\|\mathbf{x}^\star\|_\infty$

Proposition *[Fuchs,2011]*

Generically, $\mathbf{x}^\star$ has at most $m-1$ **non saturated** entries

Similar results for "antisparse" Basis pursuit

$m-1$ can be **small** compared to n

$(n \gg m)$

Towards a lower dimensional problem

Let $\mathcal{I}_+^\star \triangleq \{i \mid \mathbf{x}^\star(i) = +\|\mathbf{x}^\star\|_\infty\}$ and $\mathcal{I}_-^\star \triangleq \{i \mid \mathbf{x}^\star(i) = -\|\mathbf{x}^\star\|_\infty\}$

Sets of saturated entries

Towards a lower dimensional problem

Let $\mathcal{I}_+^\star \triangleq \{i \mid \mathbf{x}^\star(i) = +\|\mathbf{x}^\star\|_\infty\}$ and $\mathcal{I}_-^\star \triangleq \{i \mid \mathbf{x}^\star(i) = -\|\mathbf{x}^\star\|_\infty\}$

Sets of saturated entries

If we **know** $\mathcal{I}_+ \subset \mathcal{I}_+^\star$ and $\mathcal{I}_- \subset \mathcal{I}_-^\star$

Define

- $\mathbf{B} = \mathbf{A}_{\mathcal{I}^c}$
- $\mathbf{s} = \sum_{\ell \in \mathcal{I}_+} \mathbf{a}_i - \sum_{\ell \in \mathcal{I}_-} \mathbf{a}_i$

Towards a lower dimensional problem

Let $\mathcal{I}_+^\star \triangleq \{i \mid \mathbf{x}^\star(i) = +\|\mathbf{x}^\star\|_\infty\}$ and $\mathcal{I}_-^\star \triangleq \{i \mid \mathbf{x}^\star(i) = -\|\mathbf{x}^\star\|_\infty\}$

Sets of saturated entries

If we **know** $\mathcal{I}_+ \subset \mathcal{I}_+^\star$ and $\mathcal{I}_- \subset \mathcal{I}_-^\star$

Define

- $\mathbf{B} = \mathbf{A}_{\mathcal{I}^c}$
- $\mathbf{s} = \sum_{\ell\in\mathcal{I}_+} \mathbf{a}_i - \sum_{\ell\in\mathcal{I}_-} \mathbf{a}_i$

Equivalent lower dimensional problem [To appear]

$$(\mathbf{q}^\star, w^\star) \in \underset{\mathbf{q}, w \in \mathbb{R}^{\mathrm{card}(\mathcal{I}^c)} \times \mathbb{R}}{\arg\min} \frac{1}{2}\|\boldsymbol{y} - \mathbf{B}\mathbf{q} - w\mathbf{s}\|_2^2 + \lambda w \quad \text{s.t.} \quad \|\mathbf{q}\|_\infty \leq w$$

Towards a lower dimensional problem

Let $\mathcal{I}_+^\star \triangleq \{i \mid \mathbf{x}^\star(i) = +\|\mathbf{x}^\star\|_\infty\}$ and $\mathcal{I}_-^\star \triangleq \{i \mid \mathbf{x}^\star(i) = -\|\mathbf{x}^\star\|_\infty\}$

Sets of saturated entries

If we **know** $\mathcal{I}_+ \subset \mathcal{I}_+^\star$ and $\mathcal{I}_- \subset \mathcal{I}_-^\star$

Define

- $\mathbf{B} = \mathbf{A}_{\mathcal{I}^c}$
- $\mathbf{s} = \sum_{\ell\in\mathcal{I}_+} \mathbf{a}_i - \sum_{\ell\in\mathcal{I}_-} \mathbf{a}_i$

Equivalent lower dimensional problem **[To appear]**

$$(\mathbf{q}^\star, w^\star) \in \underset{\mathbf{q}, w\in\mathbb{R}^{\mathrm{card}(\mathcal{I}^c)}\times\mathbb{R}}{\arg\min} \ \frac{1}{2}\|\boldsymbol{y} - \mathbf{B}\mathbf{q} - w\mathbf{s}\|_2^2 + \lambda w \quad \text{s.t.} \quad \|\mathbf{q}\|_\infty \leq w$$

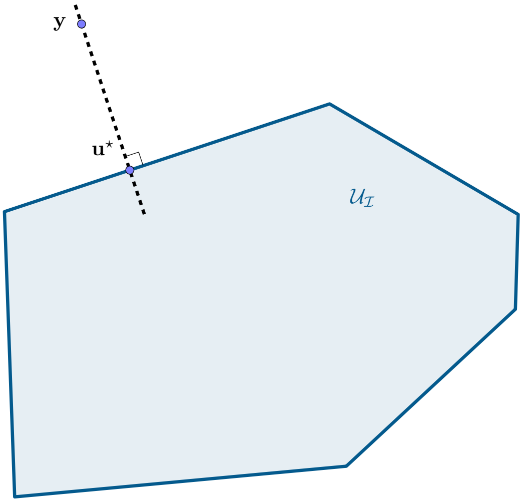
$\longrightarrow$ **Can we (dynamically) detect saturated entries?** $\longleftarrow$

Detecting saturated entries

Theorem *[to appear]*

Given a known **polytope** $\mathcal{U}_{\mathcal{I}}$

$$\text{Let} \qquad \mathbf{u}^\star = \operatorname*{arg\,min}_{\mathbf{u}\in\mathcal{U}_{\mathcal{I}}} \tfrac{1}{2}\|\boldsymbol{y} - \mathbf{u}\|_2^2$$


y
u⋆
$\mathcal{U}_{\mathcal{I}}$

Detecting saturated entries

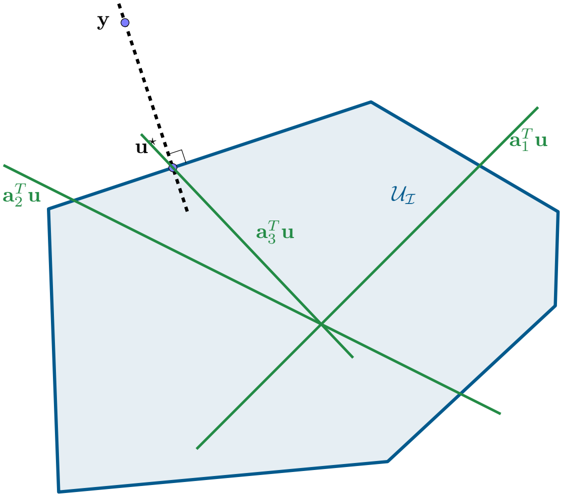
Theorem *[to appear]*

Given a known **polytope** $\mathcal{U}_{\mathcal{I}}$

$$\text{Let} \qquad \mathbf{u}^\star = \operatorname*{arg\,min}_{\mathbf{u}\in\mathcal{U}_{\mathcal{I}}} \tfrac{1}{2}\|\boldsymbol{y} - \mathbf{u}\|_2^2$$

$$\text{Then} \qquad \left|\mathbf{a}_i^\mathsf{T}\mathbf{u}^\star\right| > 0 \implies \mathbf{x}^\star(i) \text{ is saturated}$$

+ **sign** given by $\mathrm{sign}(\mathbf{a}_i^\mathsf{T}\mathbf{u}^\star)$


$\mathbf{y}$
$\mathbf{u}^\star$
$\mathbf{a}_1^T\mathbf{u}$
$\mathbf{a}_2^T\mathbf{u}$
$\mathbf{a}_3^T\mathbf{u}$
$\mathcal{U}_\mathcal{I}$

Detecting saturated entries

Theorem *[to appear]*

Given a known **polytope** $\mathcal{U}_{\mathcal{I}}$

$$\text{Let} \qquad \mathbf{u}^\star = \operatorname*{arg\,min}_{\mathbf{u}\in\mathcal{U}_{\mathcal{I}}} \tfrac{1}{2}\|\boldsymbol{y} - \mathbf{u}\|_2^2$$

$$\text{Then} \qquad \left|\mathbf{a}_i^{\mathsf{T}}\mathbf{u}^\star\right| > 0 \implies \mathbf{x}^\star(i) \ \text{ is saturated}$$

+ **sign** given by $\operatorname{sign}(\mathbf{a}_i^{\mathsf{T}}\mathbf{u}^\star)$

- **Not** a heuristic
- Computationally simple

From safe region to safe sphere

Finding $\mathbf{u}^\star$ is (almost) as difficult as finding $\mathbf{x}^\star$ ☹

From safe region to safe sphere

Finding $\mathbf{u}^\star$ is (almost) as difficult as finding $\mathbf{x}^\star$ ☹

Idea: perform the test without computing $\mathbf{u}^\star$

From safe region to safe sphere

Finding $\mathbf{u}^\star$ is (almost) as difficult as finding $\mathbf{x}^\star$ ☹

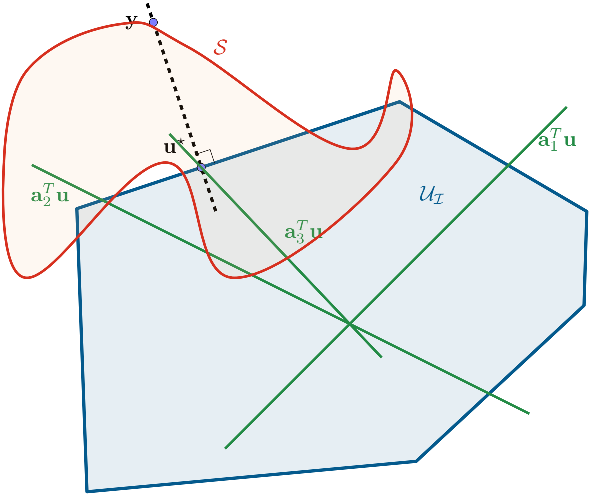
Idea: perform the test without computing $\mathbf{u}^\star$

$\longrightarrow$ Resort to a safe region

A subset $\mathcal{S}$ is called *Safe region* iff $\mathbf{u}^\star \in \mathcal{S}$ *[El Ghaoui et al., 2010]*

$$\min_{\mathbf{u}\in\mathcal{S}} \mathbf{a}_i^{\mathsf{T}}\mathbf{u} > 0 \quad \Longrightarrow \quad \mathbf{a}_i^{\mathsf{T}}\mathbf{u}^\star > 0$$

$$\Longrightarrow \quad \mathbf{x}^\star(i) \text{ is saturated}$$



$\mathbf{y}$
$\mathcal{S}$
$\mathbf{u}^\star$
$\mathbf{a}_1^T\mathbf{u}$
$\mathbf{a}_2^T\mathbf{u}$
$\mathbf{a}_3^T\mathbf{u}$
$\mathcal{U}_\mathcal{I}$

From safe region to safe sphere

Finding $\mathbf{u}^\star$ is (almost) as difficult as finding $\mathbf{x}^\star$ ☹

Idea: perform the test without computing $\mathbf{u}^\star$

$\longrightarrow$ Resort to a safe region

A subset $\mathcal{S}$ is called *Safe region* iff $\mathbf{u}^\star \in \mathcal{S}$ *[El Ghaoui et al., 2010]*

$$\min_{\mathbf{u}\in\mathcal{S}} \mathbf{a}_i{}^\mathsf{T}\mathbf{u} > 0 \quad \Longrightarrow \quad \mathbf{a}_i{}^\mathsf{T}\mathbf{u}^\star > 0$$

$$\Longrightarrow \quad \mathbf{x}^\star(i) \text{ is saturated}$$

Safe sphere:

$\mathcal{S} = \mathcal{B}(\mathbf{c}, r)$ and $\min_{\mathbf{u}\in\mathcal{B}(\mathbf{c},r)} \mathbf{a}_i{}^\mathsf{T}\mathbf{u} = \mathbf{a}_i{}^\mathsf{T}\mathbf{c} - r\|\mathbf{a}_i\|_2$

Close form solution!

Safe sphere design

Goal

Find $\mathbf{c}$ and r such that $\mathbf{u}^\star \in \mathcal{B}(\mathbf{c}, r)$

Safe sphere design

Dual problem

$$\text{Find} \quad \mathbf{u}^\star = \operatorname*{arg\,min}_{\mathbf{u}\in\mathcal{U}_{\mathcal{I}}} \|\mathbf{y} - \mathbf{u}\|_2^2$$

$\longrightarrow$ **Projection** onto the convex set $\mathcal{U}_{\mathcal{I}}$!

Safe sphere design

Dual problem

$$\text{Find} \quad \mathbf{u}^\star = \operatorname*{arg\,min}_{\mathbf{u}\in\mathcal{U}_\mathcal{I}} \|\mathbf{y} - \mathbf{u}\|_2^2$$

$\longrightarrow$ **Projection** onto the convex set $\mathcal{U}_\mathcal{I}$!

If one **knows** some $\mathbf{u}_0 \in \mathcal{U}_\mathcal{I}$, then by **definition**

$$\|\mathbf{y} - \mathbf{u}^\star\|_2^2 \leq \|\mathbf{y} - \mathbf{u}_0\|_2^2$$

$\longrightarrow$ $\mathbf{u}^\star$ belongs to a **Sphere**!

Safe sphere design

Goal

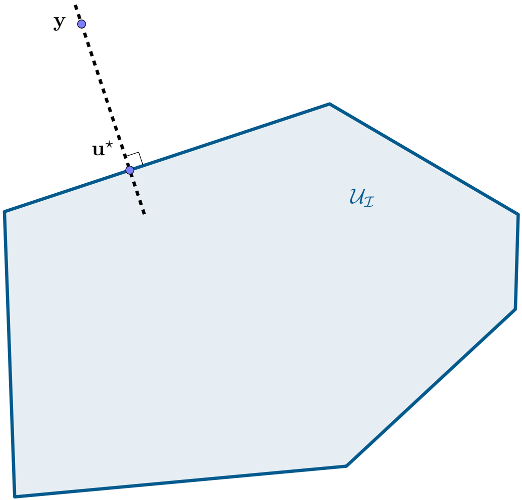
Find $\mathbf{c}$ and r such that $\mathbf{u}^\star \in \mathcal{B}(\mathbf{c}, r)$

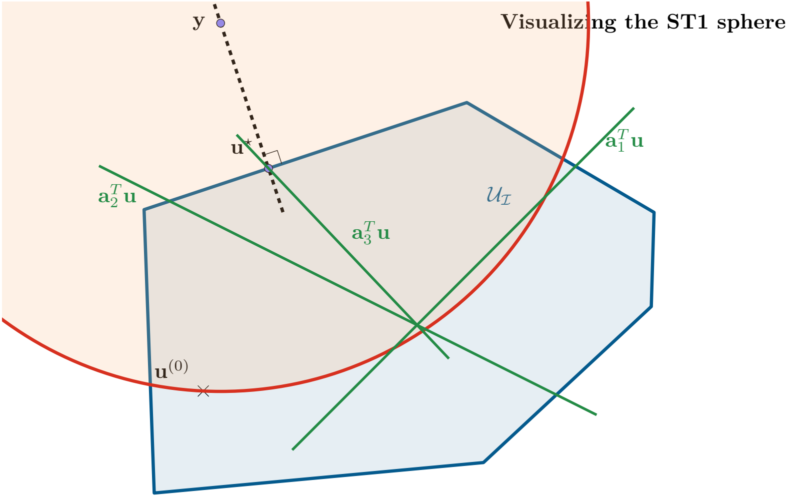
Choose $\mathbf{u}_0 \in \mathcal{U}_{\mathcal{I}}$

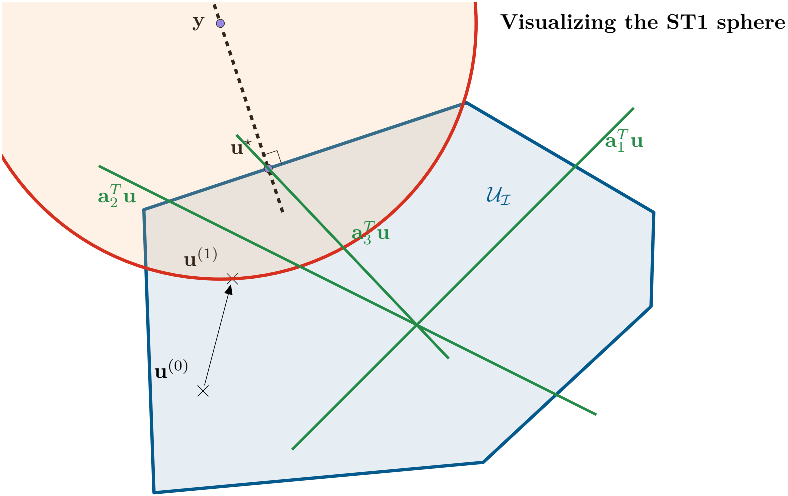
ST 1:

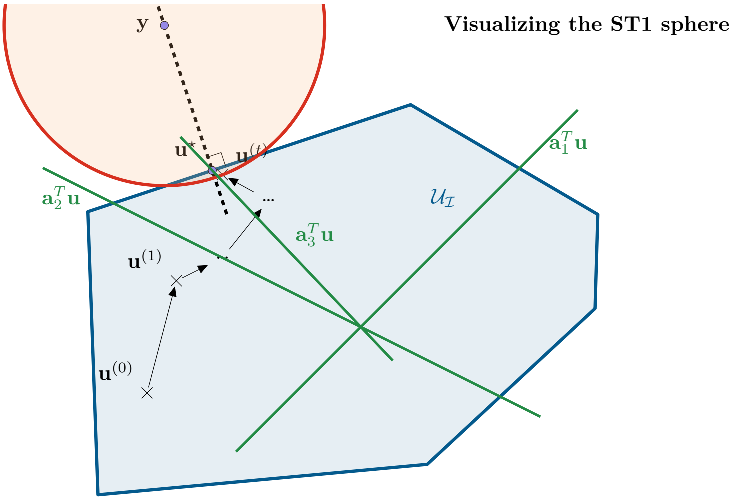
$$\mathbf{c} = \boldsymbol{y}$$
$$r = \|\boldsymbol{y} - \mathbf{u}_0\|_2$$

typical use: done once for all before runtime


y
u⋆
$\mathcal{U}_{\mathcal{I}}$


Visualizing the ST1 sphere
y
u⋆
$\mathbf{a}_1^T\mathbf{u}$
$\mathbf{a}_2^T\mathbf{u}$
$\mathbf{a}_3^T\mathbf{u}$
$\mathcal{U}_{\mathcal{I}}$
$\mathbf{u}^{(0)}$


Visualizing the ST1 sphere
$\mathbf{y}$
$\mathbf{u}^\star$
$\mathbf{a}_1^T\mathbf{u}$
$\mathbf{a}_2^T\mathbf{u}$
$\mathbf{a}_3^T\mathbf{u}$
$\mathcal{U}_\mathcal{I}$
$\mathbf{u}^{(1)}$
$\mathbf{u}^{(0)}$


Visualizing the ST1 sphere
y
u⋆
u(t)
a_1^T u
a_2^T u
a_3^T u
U_I
u(1)
u(0)
...

Safe sphere design

Goal

Find $\mathbf{c}$ and r such that $\mathbf{u}^\star \in \mathcal{B}(\mathbf{c}, r)$

Choose $\mathbf{u}_0 \in \mathcal{U}_\mathcal{I}$

ST 1:

$$\begin{aligned} \mathbf{c} &= \boldsymbol{y} \\ r &= \|\boldsymbol{y} - \mathbf{u}_0\|_2 \end{aligned}$$

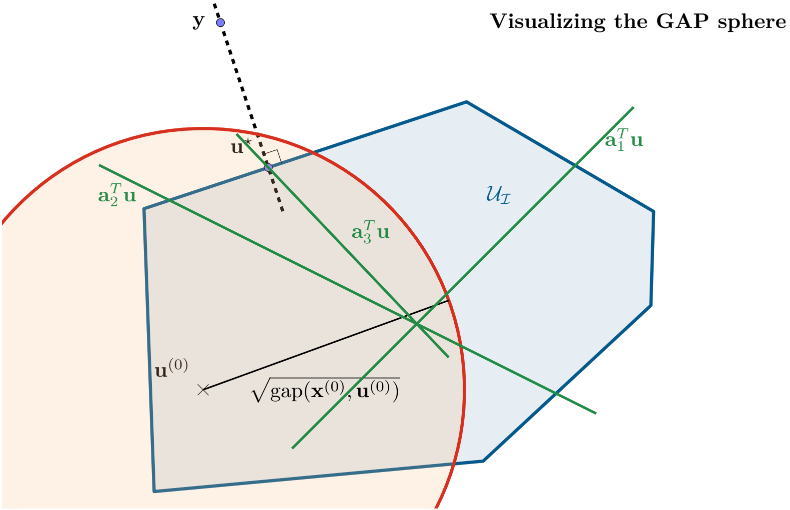
typical use: done once for all before runtime

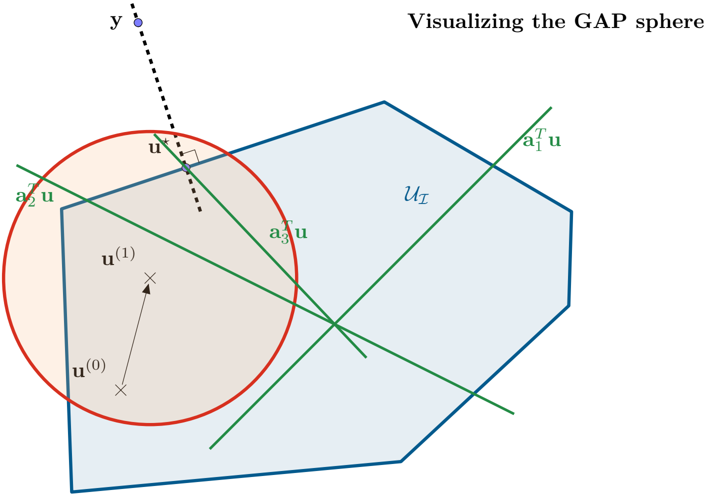
GAP sphere: *[Fercoq et al, 2015]*

$$\begin{aligned} \mathbf{c} &= \mathbf{u}_0 \\ r &= \sqrt{2\mathrm{gap}(\mathbf{x}_0, \mathbf{u}_0)} \end{aligned}$$

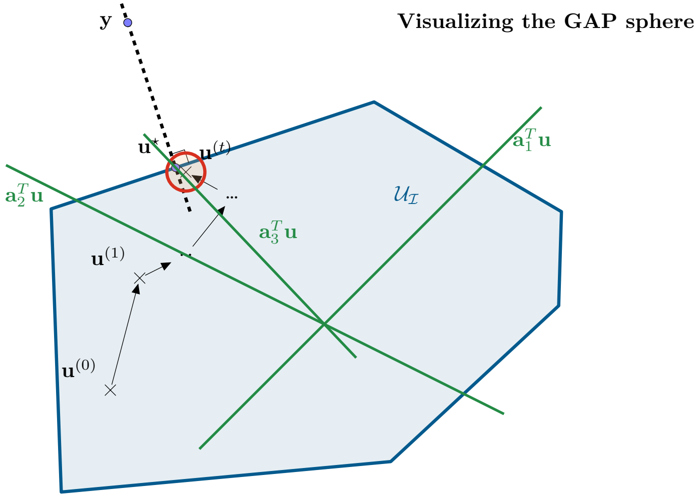
typical use:

- Dynamically
- $\mathbf{u}^{(t)} = \mathrm{proj}_{\mathcal{U}_\mathcal{I}}(\boldsymbol{y} - \mathbf{A}\mathbf{x}^{(t)})$
- radius tends to 0


Visualizing the GAP sphere
$\mathbf{y}$
$\mathbf{u}^\star$
$\mathbf{a}_1^T\mathbf{u}$
$\mathbf{a}_2^T\mathbf{u}$
$\mathbf{a}_3^T\mathbf{u}$
$\mathcal{U}_\mathcal{I}$
$\mathbf{u}^{(0)}$
$\sqrt{\mathrm{gap}(\mathbf{x}^{(0)}, \mathbf{u}^{(0)})}$



Visualizing the GAP sphere
y
u*
$a_1^T u$
$a_2^T u$
$a_3^T u$
$\mathcal{U}_{\mathcal{I}}$
$u^{(1)}$
$u^{(0)}$



Visualizing the GAP sphere
y
u⋆
u(t)
a1T u
a2T u
a3T u
U_I
u(1)
u(0)
...

Algorithms

Principle of Dynamic squeezing

$\mathbf{x}^{(0)} = \mathbf{0}_n, \mathcal{I}^{(0)} = \emptyset$;

$\mathbf{u}^{(0)} = \text{dualscal}(\boldsymbol{y})$;

$t = 1$ // iteration index

repeat

// Iterations of the optimization procedure

$(\mathbf{x}^{(t)}, \mathbf{u}^{(t)}) = \text{optim\_update}(\mathbf{x}^{(t-1)}, \mathbf{u}^{(t-1)}, \mathcal{I}^{(t)})$

// Update iteration index

$t = t + 1$

until *convergence criterion is met*;

Output: $\mathbf{x}^{(t)}, \mathcal{I}^{(t)}$

Principle of Dynamic squeezing

$\mathbf{x}^{(0)} = \mathbf{0}_n, \mathcal{I}^{(0)} = \emptyset$;

$\mathbf{u}^{(0)} = \mathrm{dualscal}(\boldsymbol{y})$;

$t = 1$ // iteration index

repeat

// Squeezing test

$(\mathbf{c}^{(t)}, r^{(t)}) = \mathrm{sphere\_param}(\mathbf{x}^{(t-1)}, \mathbf{u}^{(t-1)}, \mathcal{I}^{(t-1)})$;

$\mathcal{I}^{(t-½)} = \mathrm{squeezing\_test}(\mathbf{c}^{(t)}, r^{(t)})$;

$\mathcal{I}^{(t)} = \mathcal{I}^{(t-½)} \cup \mathcal{I}^{(t-1)}$;

// Iterations of the optimization procedure

$(\mathbf{x}^{(t)}, \mathbf{u}^{(t)}) = \mathrm{optim\_update}(\mathbf{x}^{(t-1)}, \mathbf{u}^{(t-1)}, \mathcal{I}^{(t)})$

// Update iteration index

$t = t + 1$

until *convergence criterion is met*;

Output: $\mathbf{x}^{(t)}, \mathcal{I}^{(t)}$

A word about optimization procedure

Optimization problem

$$(\mathbf{q}^\star, w^\star) \in \operatorname*{arg\,min}_{\mathbf{q}, w \in \mathbb{R}^{\mathrm{card}(\mathcal{I}^c)} \times \mathbb{R}} \frac{1}{2}\|\boldsymbol{y} - \mathbf{B}\mathbf{q} - w\mathbf{s}\|_2^2 + \lambda w \quad \text{s.t.} \begin{cases} +\mathbf{q} \leq w \\ -\mathbf{q} \leq w \end{cases}$$

Fitra is **not fitted** ☹

A word about optimization procedure

Optimization problem

$$(\mathbf{q}^\star, w^\star) \in \underset{\mathbf{q}, w \in \mathbb{R}^{\mathrm{card}(\mathcal{I}^c)} \times \mathbb{R}}{\arg\min} \ \frac{1}{2}\|\boldsymbol{y} - \mathbf{B}\mathbf{q} - w\mathbf{s}\|_2^2 + \lambda w \quad \text{s.t.} \begin{cases} +\mathbf{q} \leq w \\ -\mathbf{q} \leq w \end{cases}$$

Fitra is **not fitted** ☹

- Squeezed Fitra
  - **Projected gradient** algorithm
  - ⚠ Require the projection onto a cone
  - ⚠ Conditioning $\longrightarrow$ **scaled algorithm**

A word about optimization procedure

Optimization problem

$$(\mathbf{q}^\star, w^\star) \in \underset{\mathbf{q}, w \in \mathbb{R}^{\mathrm{card}(\mathcal{I}^c)} \times \mathbb{R}}{\arg\min} \frac{1}{2}\|\boldsymbol{y} - \mathbf{B}\mathbf{q} - w\mathbf{s}\|_2^2 + \lambda w \quad \text{s.t.} \begin{cases} +\mathbf{q} \leq w \\ -\mathbf{q} \leq w \end{cases}$$

Fitra is **not fitted** ☹

- Squeezed Fitra
  - **Projected gradient** algorithm
  - ⚠ Require the projection onto a cone
  - ⚠ Conditioning $\longrightarrow$ **scaled algorithm**

$$w\mathbf{s} \longleftarrow \widetilde{w}\frac{\mathbf{s}}{\|\mathbf{s}\|_2}$$

A word about optimization procedure

Optimization problem

$$(\mathbf{q}^\star, w^\star) \in \operatorname*{arg\,min}_{\mathbf{q}, w \in \mathbb{R}^{\mathrm{card}(\mathcal{I}^c)} \times \mathbb{R}} \frac{1}{2}\|\boldsymbol{y} - \mathbf{B}\mathbf{q} - w\mathbf{s}\|_2^2 + \lambda w \quad \text{s.t.} \begin{cases} +\mathbf{q} \leq w \\ -\mathbf{q} \leq w \end{cases}$$

Fitra is **not fitted** ☹

- Squeezed Fitra
  - **Projected gradient** algorithm
  - ⚠ Require the projection onto a cone
  - ⚠ Conditioning $\longrightarrow$ **scaled algorithm**

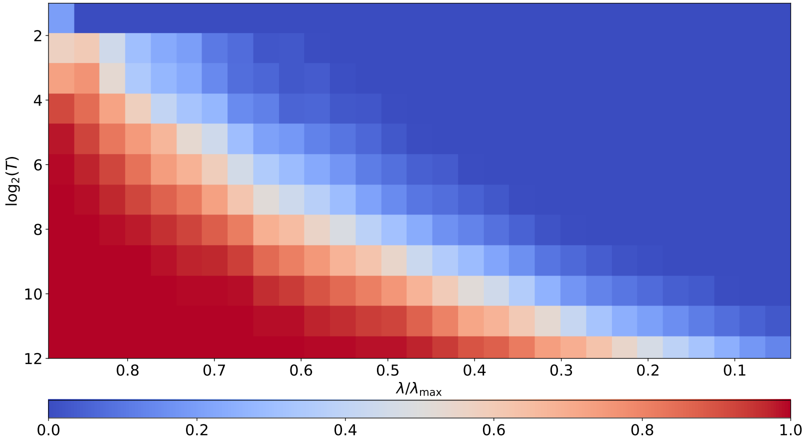
$$w\mathbf{s} \longleftarrow \widetilde{w}\frac{\mathbf{s}}{\|\mathbf{s}\|_2}$$

- Squeezed **Frank-Wolfe**

Numerical experiments

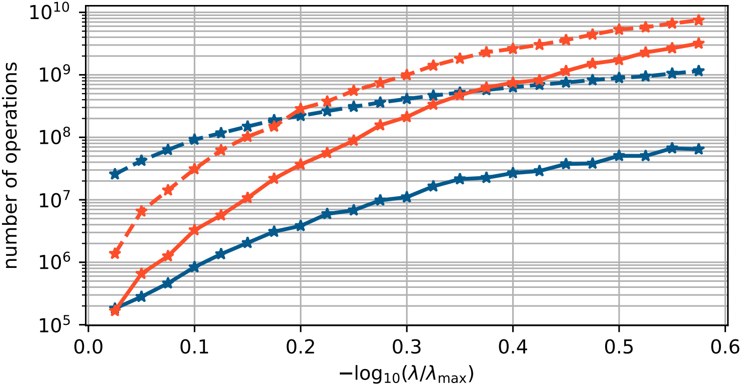
Percentage of screened variables of iteration

$\mathbf{A} \in \mathbb{R}^{100\times 150}$ $\mathbf{A}[i,j] \in [0,1]$ GAP sphere



Complexity savings

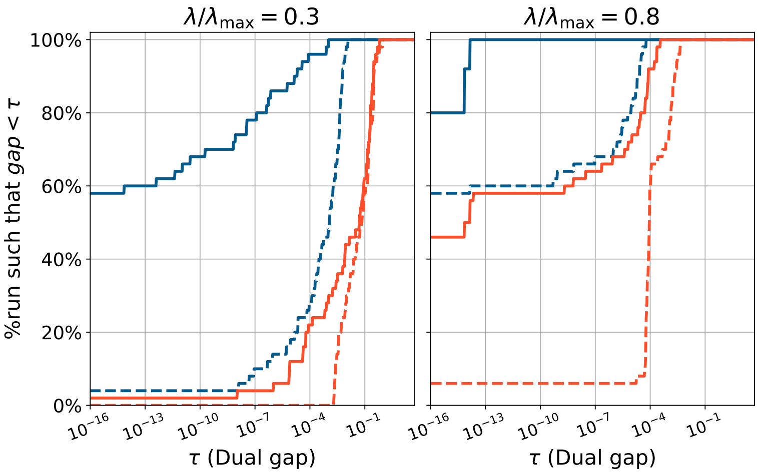
$\mathbf{A} \in \mathbb{R}^{100\times 150}$ $\mathbf{A}[i,j] \in [0,1]$ GAP sphere



- Fitra
- Squeezed Fitra
- Frank-Wolfe
- Squeezed Frank-wolfe

Benchmark

$\mathbf{A} \in \mathbb{R}^{100\times 150}$ $\mathbf{A}[i,j] \in [0,1]$ Budget: 10^8 operations



- Fitra
- Squeezed Fitra
- Frank-Wolfe
- Squeezed Frank-wolfe

Squeezing test - at no cost?

- **Computing u:** Dual scaling of residual vector

  $\mathcal{O}(1)$ ✓

- **Squeezing test:** inner product $\mathbf{a}_i^T\mathbf{u}$

  $\equiv$ gradient descent step $\rightarrow$ **already done** ✓

- **Squeezing test:** radius r

  $\equiv$ dual gap $\rightarrow$ **already computed to monitor convergence** ✓

- **Proximity operator:** sorting $\mathcal{O}(\tilde{n}\log(\tilde{n}))$ $\tilde{n}$ is decreasing here

  can be **faster** than computing the prox of the ℓ_∞-norm $\mathcal{O}(n)$

Conclusion - prospects

- It is possible to **dynamically** detect **saturated** entries
- It leads to an equivalent **low dimensional** problem
- We obtain **faster** algorithms at (almost) **no** additional **cost**

Prospects

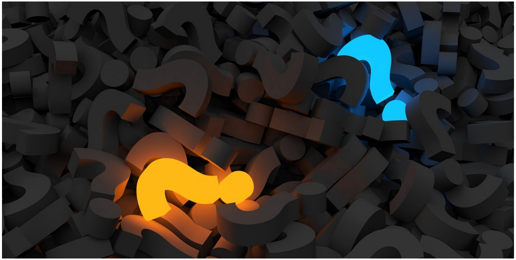
- Other safe regions (dome, truncated dome...)
- **Nesterov** acceleration?
- Extension to more **BLasso**? continuous dictionaries?

stay tuned!

`https://arxiv.org/abs/1911.07508`

Toolbox: `https://gitlab.inria.fr/celvira/safe-squeezing`

Merci de votre attention!



stay tuned!

https://arxiv.org/abs/1911.07508

Toolbox: https://gitlab.inria.fr/celvira/safe-squeezing

Ideal test to detect saturated entries

Theorem *[to appear]*

Let $$\mathbf{u}^\star = \underset{\mathbf{u}\in\mathcal{U}_\mathcal{I}}{\arg\max}\ \frac{1}{2}\|\boldsymbol{y}\|_2^2 - \frac{1}{2}\|\boldsymbol{y} - \mathbf{u}\|_2^2$$

with $\mathcal{U}_\mathcal{I}$ a known polytope

Then $$\left|\mathbf{a}_i{}^\top \mathbf{u}^\star\right| > 0 \implies \mathbf{x}^\star(i) \text{ is saturated}$$

+ sign given by $\mathrm{sign}(\mathbf{a}_i{}^\top\mathbf{u}^\star)$

Ideal test to detect saturated entries

Theorem *[to appear]*

Let $$\mathbf{u}^\star = \arg\max_{\mathbf{u}\in\mathcal{U}_{\mathcal{I}}} \tfrac{1}{2}\|\boldsymbol{y}\|_2^2 - \tfrac{1}{2}\|\boldsymbol{y} - \mathbf{u}\|_2^2$$

with $\mathcal{U}_{\mathcal{I}}$ a known polytope

Then $$\left|\mathbf{a}_i{}^{\mathsf{T}}\mathbf{u}^\star\right| > 0 \implies \mathbf{x}^\star(i) \text{ is saturated}$$

+ sign given by $\mathrm{sign}(\mathbf{a}_i{}^{\mathsf{T}}\mathbf{u}^\star)$

1. **Slater** conditions involves
   $$\exists \mathbf{v}_+^\star \text{ s.t. } \mathbf{v}_+^\star(i)(\mathbf{x}^\star(i) - w^\star) = 0$$

Ideal test to detect saturated entries

Theorem *[to appear]*

Let $$\mathbf{u}^\star = \operatorname*{arg\,max}_{\mathbf{u}\in\mathcal{U}_\mathcal{I}} \tfrac{1}{2}\|\boldsymbol{y}\|_2^2 - \tfrac{1}{2}\|\boldsymbol{y}-\mathbf{u}\|_2^2$$

with $\mathcal{U}_\mathcal{I}$ a known polytope

Then $$\left|\mathbf{a}_i{}^\mathsf{T}\mathbf{u}^\star\right| > 0 \implies \mathbf{x}^\star(i) \text{ is saturated}$$

+ sign given by $\mathrm{sign}(\mathbf{a}_i{}^\mathsf{T}\mathbf{u}^\star)$

1. **Slater** conditions involves
   $$\exists \mathbf{v}_+^\star \text{ s.t. } \mathbf{v}_+^\star(i)(\mathbf{x}^\star(i) - w^\star) = 0$$
2. 1st order **optimality conditions** involve
   $$\mathbf{v}_+^\star(i) = \mathbf{a}_i{}^\mathsf{T}\mathbf{u}^\star$$

Ideal test to detect saturated entries

Theorem *[to appear]*

Let $$\mathbf{u}^\star = \underset{\mathbf{u}\in\mathcal{U}_\mathcal{I}}{\arg\max}\ \tfrac{1}{2}\|\boldsymbol{y}\|_2^2 - \tfrac{1}{2}\|\boldsymbol{y} - \mathbf{u}\|_2^2$$

with $\mathcal{U}_\mathcal{I}$ a known polytope

Then $$\left|\mathbf{a}_i{}^\mathsf{T}\mathbf{u}^\star\right| > 0 \implies \mathbf{x}^\star(i) \text{ is saturated}$$

+ sign given by $\mathrm{sign}(\mathbf{a}_i{}^\mathsf{T}\mathbf{u}^\star)$

1. **Slater** conditions involves
   $$\exists \mathbf{v}_+^\star \text{ s.t. } \mathbf{v}_+^\star(i)(\mathbf{x}^\star(i) - w^\star) = 0$$
2. 1st order **optimality conditions** involve
   $$\mathbf{v}_+^\star(i) = \mathbf{a}_i{}^\mathsf{T}\mathbf{u}^\star$$
3. If $\mathbf{v}_+^\star(i) \neq 0$ then
   $$\mathbf{x}^\star(i) = +w^\star \text{ necessarily}$$

Ideal test to detect saturated entries

Theorem *[to appear]*

Let $$\mathbf{u}^\star = \arg\max_{\mathbf{u}\in\mathcal{U}_{\mathcal{I}}} \tfrac{1}{2}\|\boldsymbol{y}\|_2^2 - \tfrac{1}{2}\|\boldsymbol{y} - \mathbf{u}\|_2^2$$

with $\mathcal{U}_{\mathcal{I}}$ a known polytope

Then $$\left|\mathbf{a}_i{}^{\mathsf{T}}\mathbf{u}^\star\right| > 0 \implies \mathbf{x}^\star(i) \text{ is saturated}$$

+ sign given by $\mathrm{sign}(\mathbf{a}_i{}^{\mathsf{T}}\mathbf{u}^\star)$

1. **Slater** conditions involves
   $$\exists \mathbf{v}_+^\star \text{ s.t. } \mathbf{v}_+^\star(i)(\mathbf{x}^\star(i) - w^\star) = 0$$
2. 1st order **optimality conditions** involve
   $$\mathbf{v}_+^\star(i) = \mathbf{a}_i{}^{\mathsf{T}}\mathbf{u}^\star$$
3. If $\mathbf{v}_+^\star(i) \neq 0$ then
   $$\mathbf{x}^\star(i) = +w^\star \text{ necessarily}$$

+ $\mathbf{u}^\star$ cannot be orthogonal to all columns of $\mathbf{A}$.