Some computational aspects of off-the-grid inverse problems

Clément ELVIRA

CentraleSupélec Rennes – IETR Lab

AI WILD West workshop
January 29, 2026

Context

Spike deconvolution, machine learning, sketching
as sparse inverse problems over measures

Running example: spike deconvolution in 1D

- Signal model

$$\boldsymbol{x} = \sum_{\ell=1}^{3} x_\ell \delta_{\theta_\ell^\star} \quad \text{with} \quad \theta_\ell^\star \in [\theta_{\min}, \theta_{\max}]$$

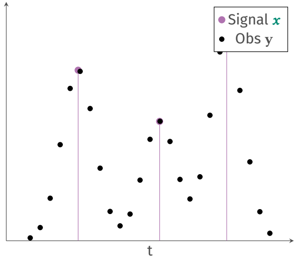
- Observation model

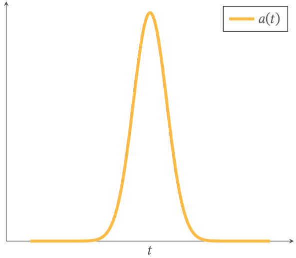
$$\begin{aligned} y(t) &= (a * \boldsymbol{x})(t) + \text{ noise } (t) \\ &= \sum_{\ell=1}^{3} x_\ell a(t - \theta_\ell^\star) + \text{ noise } (t) \end{aligned}$$

eventually sampled at $t_1, \dots, t_m$

- "Blurring", "point spread", "atom" function

$$a : \mathbb{R} \to \mathbb{R}$$





Sparse inverse problems over spaces of measures[1]

$$\text{Find } \hat{\boldsymbol{x}} \in \underset{\boldsymbol{x} \in \mathscr{R}(\Theta)}{\arg\min} \quad f(\mathscr{M}\boldsymbol{x}) + \lambda \|\boldsymbol{x}\|_{\mathrm{TV}} \qquad \text{with } \lambda > 0 \qquad (\mathrm{P}_\lambda)$$

Ingredient 1: Signal

- Signal set $\mathscr{R}(\Theta)$ *– Set of Radon measures*
- Parameter set $\Theta \subset \mathbb{R}^d$

Ingredient 2: Sensing

- "Atom" function $\boldsymbol{a} : \Theta \to \mathbb{R}^m$ *– assumed vanishing*
- Sensing operator $\mathscr{M} : \mathscr{R}(\Theta) \to \mathbb{R}^m$ *– linear, continuous, characterized by $\mathscr{M}\delta_\theta = \boldsymbol{a}(\theta)$*

Ingredient 3: Measure of success + regularization (signal model)

- "Smooth" loss function f *– latter required to be differentiable with Lipschitz gradient*
- "Sparsity" *– trough the TV-norm, see next slide*

---

[1] [Bredies & Pikkarainen, 2012], [Castro & Gamboa 2012]

TV-norm enforces sparsity

Representation theorem [Boyer & *al*, 2019]

If (P_λ) has a solution, then at least one is of the form

$$\hat{\boldsymbol{x}} = \sum_{\ell=1}^{\hat{k}} \hat{x}_\ell \delta_{\hat{\theta}_\ell} \qquad \text{with } \hat{k} \leq m$$

- Further condition exists to ensure uniqueness
- TV-norm of a discrete measure $\boldsymbol{x} = \sum_{\ell=1}^{k} x_\ell \delta_{\theta_\ell}$:

$$\|\boldsymbol{x}\|_{\mathrm{TV}} = \sum_{\ell=1}^{k} |x_\ell| = \| \begin{pmatrix} x_1 \\ \vdots \\ x_k \end{pmatrix} \|_1$$

Spike deconvolution trough sparse decomposition

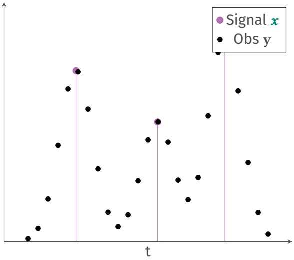
- Grid of admissible locations
  $\Theta_{\texttt{grid}} = \{\theta_1, \dots, \theta_n\}$

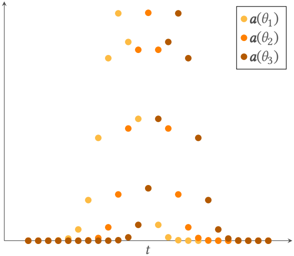
- Sampled"Blurring" function

$$\boldsymbol{a}(\theta) = \begin{pmatrix} a(t_1 - \theta) \\ \vdots \\ a(t_m - \theta) \end{pmatrix} \in \mathbb{R}^m$$

- Reconstruction model

$$\underbrace{\begin{pmatrix} y(t_1) \\ \vdots \\ y(t_m) \end{pmatrix}}_{\overset{\text{def.}}{=} \mathbf{y}} \simeq \underbrace{\begin{pmatrix} \vdots & & \vdots \\ \boldsymbol{a}(\theta_1) & \dots & \boldsymbol{a}(\theta_n) \\ \vdots & & \vdots \end{pmatrix}}_{\overset{\text{def.}}{=} \mathbf{A}} \overset{\text{def.}}{=} \mathbf{x}$$





The Lasso is an instance of (P_λ)

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

Rewriting of an **instance** of (P_λ) where

- $\Theta = \Theta_{\mathtt{grid}}$
- $\mathscr{R}(\Theta)$ set if discrete measures located in $\Theta_{\mathtt{grid}}$
- $\boldsymbol{a}(\theta_i)$ refers to the ith column of $\mathbf{A}$
- $f(\cdot)=\frac{1}{2}\|\mathbf{y}-\cdot\|_2^2$
- Solution at most m-sparse[2]

---

[2] Foucart, & Rauhut (2013). *A Mathematical Introduction to Compressive Sensing*

Off-the-grid Lasso (Blasso)

Problems with grid discretization:

- Ground truth parameters $\theta_\ell^\star$ may not lie on the grid
- Theoretical guarantees may not hold if the grid is too thin
- Computational burden / memory footprint

Solution: instance of (P_λ) over with

- $\Theta = [\theta_{\min}, \theta_{\max}]$
- $\mathscr{R}(\Theta)$ the set of Radon measures over Θ

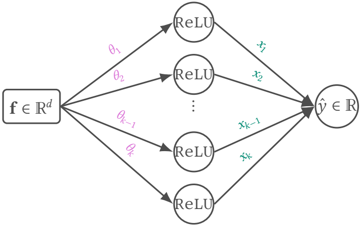
BLasso: $$\underset{\boldsymbol{x} \in \mathscr{R}(\Theta)}{\arg\min} \; \frac{1}{2}\|\mathbf{y} - \mathscr{M}\boldsymbol{x}\|_2^2 + \lambda\|\boldsymbol{x}\|_{\mathrm{TV}}$$

Applications: [Denoyelle & al, 2019], [Di Carlo & al, 2020], ...

Training over-parametrized neural networks

Over-parameterized **shallow neural network** with one hidden layer:

$$\begin{aligned} \mathrm{NN}: \ \mathbb{R}^d &\to \mathbb{R} \\ \mathbf{f} &\mapsto \textstyle\sum_{\ell=1}^{k} x_\ell \mathrm{ReLU}(\langle \theta_\ell \mid \mathbf{f} \rangle) \end{aligned}$$



Training problem: Ridge-regularized ERM rule

$$\underset{\substack{x_1,\dots,x_k \mathbb{R} \\ \theta_1,\dots,\theta_k \in \mathbb{R}^d}}{\arg\min} \ \frac{1}{m} \sum_{i=1}^{m} \left(y_i - \mathrm{NN}(\mathbf{f}_i) \right)^2 + \lambda \sum_{\ell=1}^{k} \left(\|\theta_\ell\|_2^2 + x_\ell^2 \right)$$

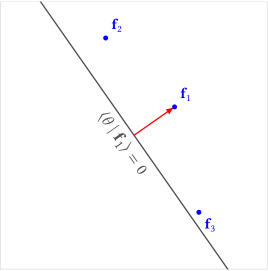
$\{(\mathbf{f}_i, y_i)\}_{i=1}^m$ *training set*

Blasso for NN

Equivalent[3] to an instance of (P_λ) with

- Θ = unit ball of $\mathbb{R}^d$ (hidden weights)
- Atom function

$$a(\theta) = \begin{pmatrix} \mathrm{ReLU}(\langle \theta \mid \mathbf{f}_1 \rangle) \\ \vdots \\ \mathrm{ReLU}(\langle \theta \mid \mathbf{f}_m \rangle) \end{pmatrix} \in \mathbb{R}^m$$

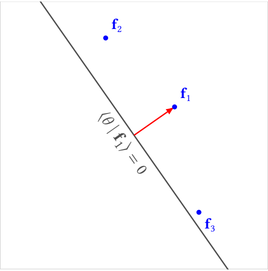


Blasso for NN

Equivalent[3] to an instance of (P_λ) with

- Θ = unit ball of $\mathbb{R}^d$ (hidden weights)
- Atom function

$$a(\theta) = \begin{pmatrix} \mathrm{ReLU}(\langle \theta \mid \mathbf{f}_1 \rangle) \\ \vdots \\ \mathrm{ReLU}(\langle \theta \mid \mathbf{f}_m \rangle) \end{pmatrix} \in \mathbb{R}^m$$

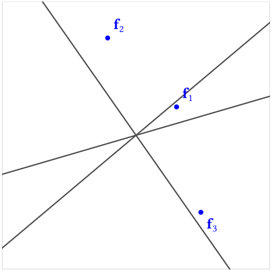


Blasso for NN

Equivalent[3] to an instance of (P_λ) with

- Θ = unit ball of $\mathbb{R}^d$ (hidden weights)
- Atom function

$$a(\theta) = \begin{pmatrix} \mathrm{ReLU}(\langle \theta \mid \mathbf{f}_1 \rangle) \\ \vdots \\ \mathrm{ReLU}(\langle \theta \mid \mathbf{f}_m \rangle) \end{pmatrix} \in \mathbb{R}^m$$



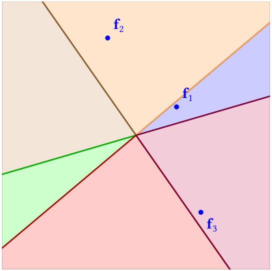
Blasso for NN

Equivalent[3] to an instance of (P_λ) with

- Θ = unit ball of $\mathbb{R}^d$ (hidden weights)
- Atom function

$$a(\theta) = \begin{pmatrix} \mathrm{ReLU}(\langle \theta \mid \mathbf{f}_1 \rangle) \\ \vdots \\ \mathrm{ReLU}(\langle \theta \mid \mathbf{f}_m \rangle) \end{pmatrix} \in \mathbb{R}^m$$

Linear on each cone



---

[3] provided that k is larger than the number hyperplane arrangements induced by the training set, see [Pilanci & and Ergen 2020]

Problem

Complexity issues associated to the resolution of (P_λ)

Main ingredients of typical solvers

Most solvers leverage a (fine) **grid** or **covering** of Θ and follows the following three steps pattern:

1. Maximize some “correlation” function

$$\begin{aligned} \texttt{Corr}(\cdot, \texttt{residual}) : \ & \Theta \to \mathbb{R}_+ \\ & \theta \mapsto |\langle \boldsymbol{a}(\theta) \mid \texttt{residual} \rangle| \end{aligned}$$

Specifically: $\texttt{residual} = \nabla f(\mathscr{M}\boldsymbol{x}^{(t)})$ *with* $\boldsymbol{x}^{(t)}$ *the measure at iteration* t

2. Solve a finite-dimensional counterpart / approximation of (P_λ) related to the grid / covering

3. Update or refine the grid / covering

Computational bottleneck

Resulting grid: ε^{-d} vertices

$\varepsilon \equiv$ grid step-size

1. Evaluating/storing atoms
2. Conditioning of finite-dimensional versions of (P$_\lambda$)
3. solver-specific issues

Question: can we lower this computational *burden* / *memory footprint*?

Solution

- Given $\mathscr{T} \subset \Theta$, if all solutions of ($\mathrm{P}_\lambda$) have no mass in $\mathscr{T}$, then one can “switch” to the resolution of

$$\operatorname*{arg\,min}_{\boldsymbol{x}\in\mathscr{R}(\Theta\setminus\mathscr{T})} \quad f(\mathscr{M}\boldsymbol{x}) + \lambda\|\boldsymbol{x}\|_{\mathrm{TV}} \quad \text{s.t.}$$

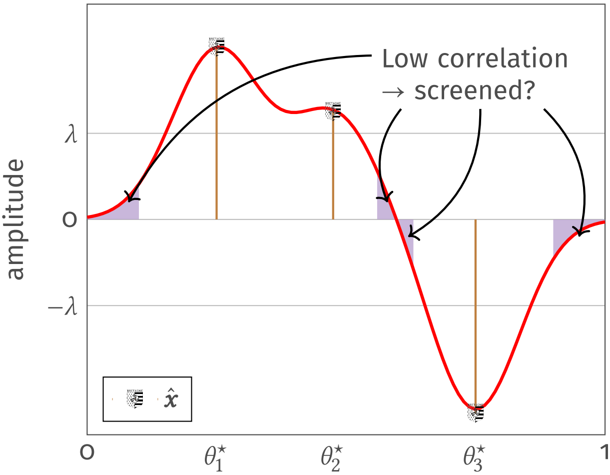
- A test to ***safely*** prune/screen any $\mathscr{T} \subseteq \Theta$

$$\varphi(\mathscr{T}) < \lambda \quad \Longrightarrow \quad \forall \text{ minimizers } \hat{\boldsymbol{x}} \,:\, \mathscr{T} \cap \operatorname{support}(\hat{\boldsymbol{x}}) = \emptyset$$

Infinite-dimensional counterpart of safe screening introduced by El Ghaoui and coauthors in 2010 for feature elimination

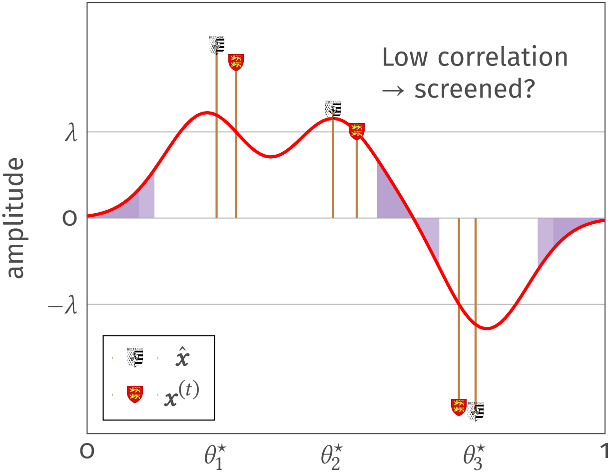
Intuition and rationale of the approach

“Correlation” function $\theta \mapsto \langle \boldsymbol{a}(\theta) \mid \mathbf{y} \rangle$



Intuition and rationale of the approach

“Correlation” function $\theta \mapsto \langle \boldsymbol{a}(\theta) \mid \mathbf{y} - \mathscr{M}\boldsymbol{x}^{(t)} \rangle$



Question: How small should be "$|\langle \boldsymbol{a}(\theta) \mid \mathbf{y} - \mathscr{M}\boldsymbol{x}^{(t)} \rangle|$" to *safely* eliminate θ?

To simplify the exposition, I will

- focus on the Least Squares loss

  results will remain valid under conditions that will be made explicit

- Assume that (P_λ) admits a unique solution $\hat{\boldsymbol{x}}$

  the general case requires replacing "$\hat{\boldsymbol{x}}$" by "$\forall$ minimizers $\hat{\boldsymbol{x}}$"

The Fermat's rule for (P_λ)[4]

The solution $\hat{\boldsymbol{x}}$ of (P_λ) verifies

$$\forall \theta \;:\; |\langle \boldsymbol{a}(\theta) \mid \mathbf{y} - \mathscr{M}\hat{\boldsymbol{x}} \rangle| \in \begin{cases} \{\lambda\} & \text{if } \theta \in \mathrm{support}(\hat{\boldsymbol{x}}) \\ [0, \lambda] & \text{otherwise} \end{cases}$$

---

[4] necessary and sufficient condition for optimality generalizing the so-called "$\nabla f(x) = 0$"

The Fermat's rule for (P_λ)[4]

The solution $\hat{\boldsymbol{x}}$ of (P_λ) verifies

$$\forall \theta \,:\, |\langle \boldsymbol{a}(\theta) \mid \mathbf{y} - \mathscr{M}\hat{\boldsymbol{x}} \rangle| \in \begin{cases} \{\lambda\} & \text{if } \theta \in \mathrm{support}(\hat{\boldsymbol{x}}) \\ [0, \lambda] & \text{otherwise} \end{cases}$$

Safe screening rule

$$|\langle \boldsymbol{a}(\theta) \mid \mathbf{y} - \mathscr{M}\hat{\boldsymbol{x}} \rangle| < \lambda \quad \Longrightarrow \quad \theta \notin \mathrm{support}(\hat{\boldsymbol{x}})$$

Independent from the primal solution

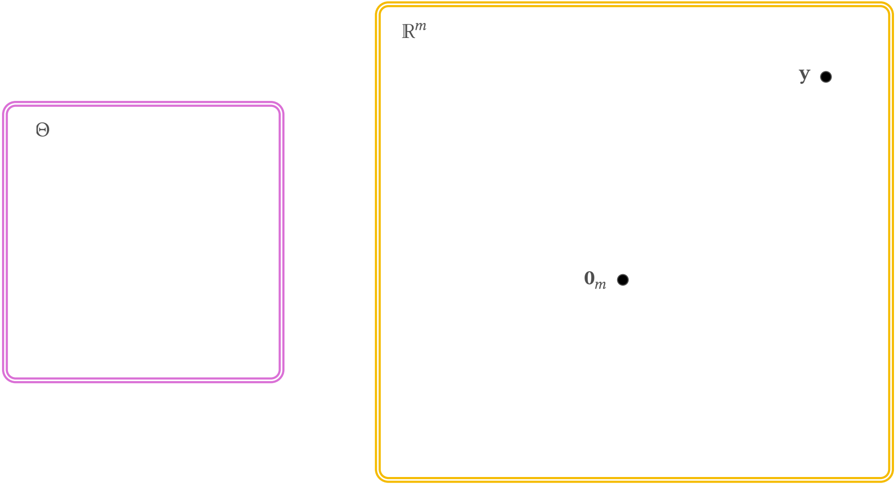
Safe

Replace "$\mathbf{y} - \mathscr{M}\hat{\boldsymbol{x}}$" by "$-\nabla f(\mathscr{M}\hat{\boldsymbol{x}})$" in the general case

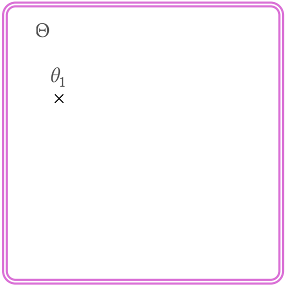
---

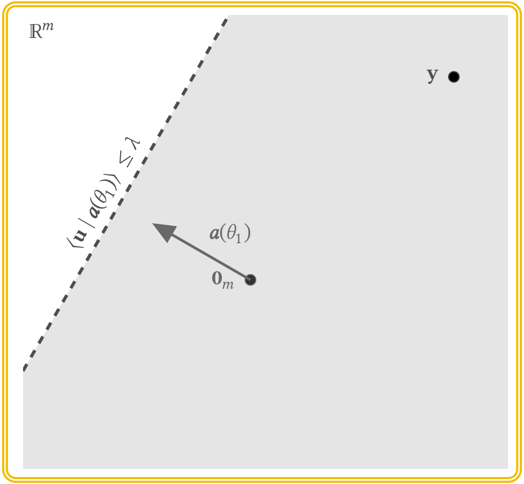
[4] necessary and sufficient condition for optimality generalizing the so-called "$\nabla f(x) = 0$"

A geometric view of safe screening

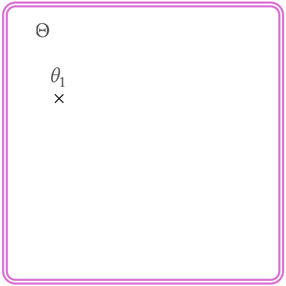


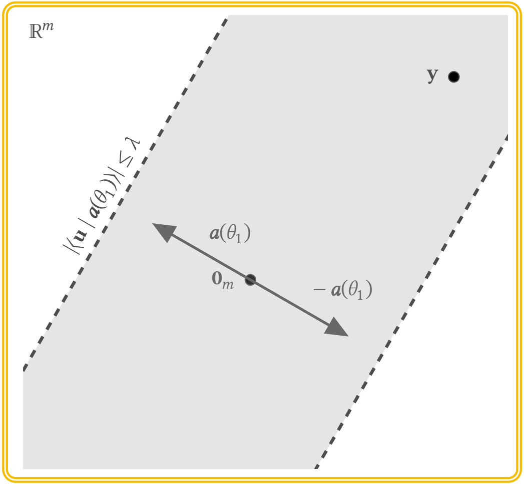
A geometric view of safe screening



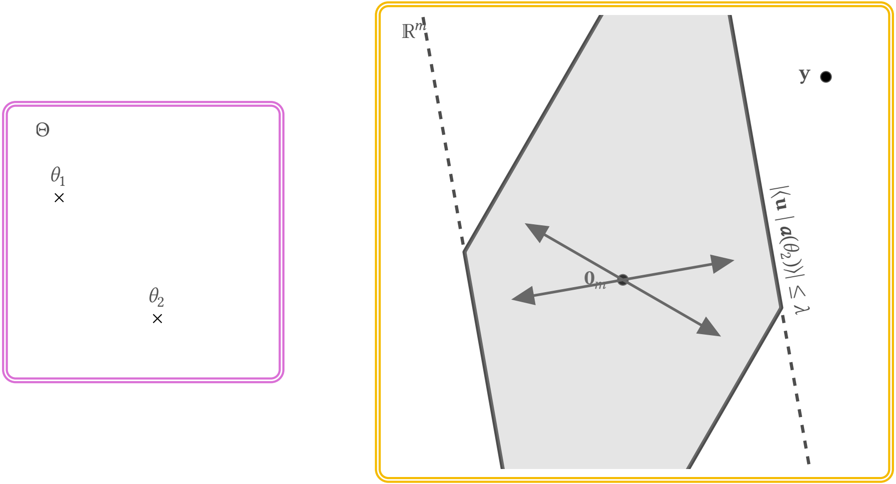


A geometric view of safe screening

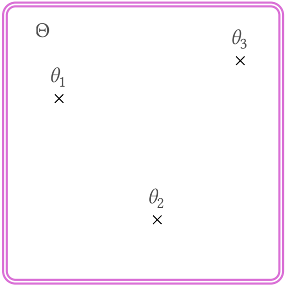


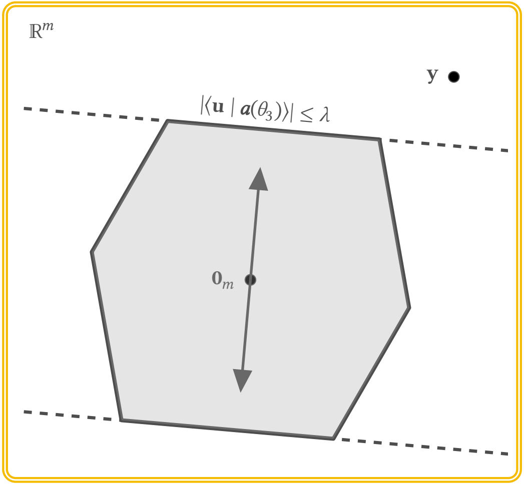


A geometric view of safe screening

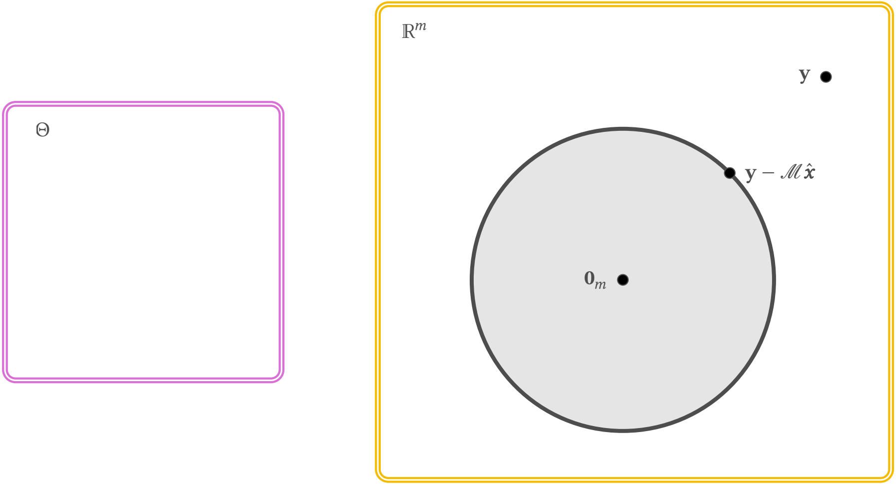


A geometric view of safe screening

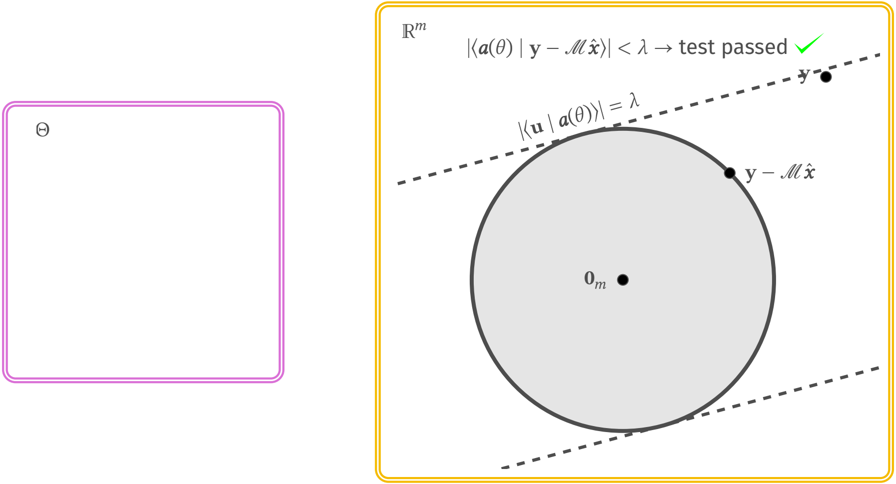




A geometric view of safe screening



A geometric view of safe screening



Two practical limitations

Safe screening rule *(reminder)*

$$|\langle \boldsymbol{a}(\theta) \mid \mathbf{y} - \mathscr{M}\hat{\boldsymbol{x}} \rangle| < \lambda \quad \Longrightarrow \quad \theta \notin \mathrm{support}(\hat{\boldsymbol{x}})$$

1. Require the knowledge of $\mathbf{y} - \mathscr{M}\hat{\boldsymbol{x}}$

   Evaluating $\mathbf{y} - \mathscr{M}\hat{\boldsymbol{x}}$ *amounts to solving* (P_λ)

2. Results exploitation ?

Addressing limitation 1 with safe regions

Residual error is the dual solution

Fact: $\mathbf{y} - \mathscr{M}\hat{\boldsymbol{x}}$ is *the unique* solution to the **dual problem**

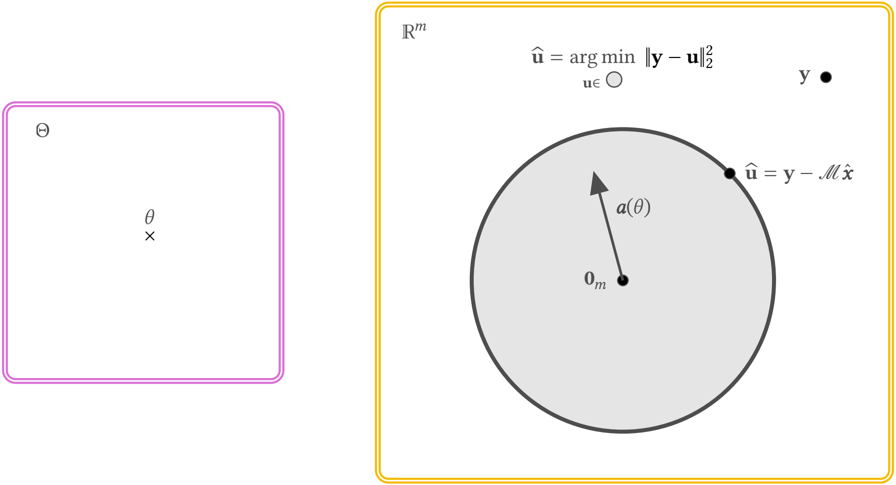
$$\widehat{\mathbf{u}} \overset{\text{def.}}{=} \underset{\mathbf{u}\in\mathbb{R}^m}{\arg\max}\ \frac{1}{2}\|\mathbf{y}\|_2^2 - \|\mathbf{y} - \mathbf{u}\|_2^2 \quad \text{s.t.} \quad |\langle \boldsymbol{a}(\theta) \mid \mathbf{u}\rangle| \leq \lambda\ \forall\theta \in \Theta \ \ (\mathsf{D}_\lambda)$$

- Equivalent to solve

$$\underset{\mathbf{u}\in\mathbb{R}^m}{\arg\min}\ \|\mathbf{y} - \mathbf{u}\|_2^2 \quad \text{s.t.} \quad |\langle \boldsymbol{a}(\theta) \mid \mathbf{u}\rangle| \leq \lambda\ \forall\theta \in \Theta$$

$\implies$ orthogonal projection

- $\widehat{\mathbf{u}} = -\nabla f(\mathscr{M}\hat{\boldsymbol{x}})$ in the general case


Θ
θ
$\mathbb{R}^m$
$\widehat{\mathbf{u}} = \arg\min_{\mathbf{u}\in ○} \|\mathbf{y}-\mathbf{u}\|_2^2$
$\mathbf{y}$
$\widehat{\mathbf{u}} = \mathbf{y} - \mathscr{M}\hat{\boldsymbol{x}}$
$\boldsymbol{a}(\theta)$
$\mathbf{0}_m$

Addressing limitation 1 with safe regions

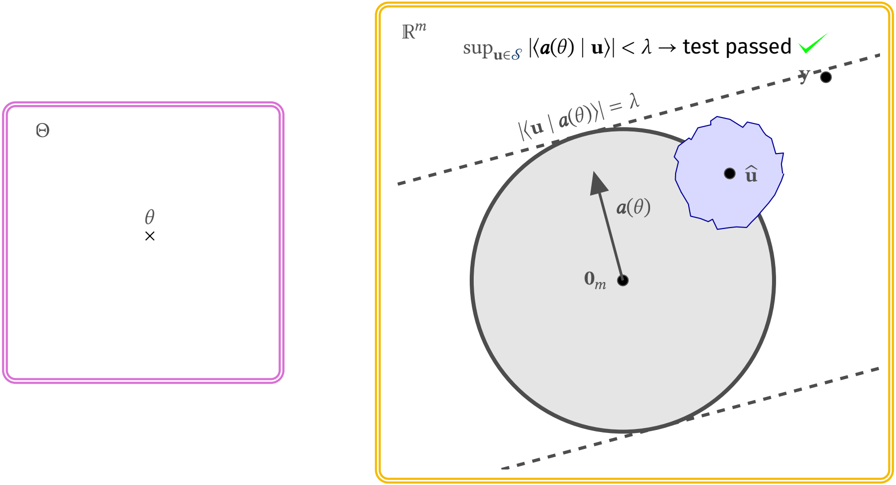
Definition: $\mathcal{S} \subseteq \mathbb{R}^m$ is said to be a **safe region**[5] iff $\widehat{\mathbf{u}} \in \mathcal{S}$

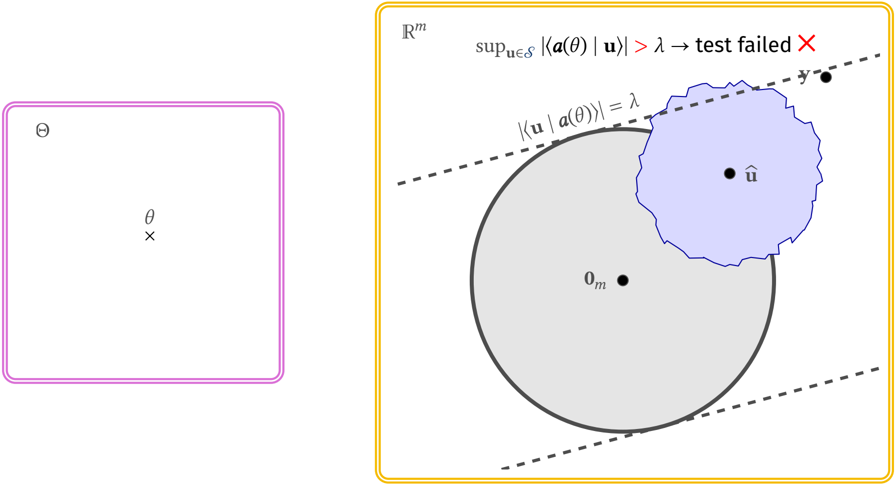
Relaxed test with safe region

$$|\langle \boldsymbol{a}(\theta) \mid \widehat{\mathbf{u}} \rangle| \leq \sup_{\mathbf{u} \in \mathcal{S}} |\langle \boldsymbol{a}(\theta) \mid \mathbf{u} \rangle| < \lambda \quad \Longrightarrow \quad \theta \notin \operatorname{support}(\hat{\boldsymbol{x}})$$

---

[5] Introduced by El Ghaoui *et al.* (2012) in the context of LASSO


Θ
θ
$\mathbb{R}^m$
$\sup_{\mathbf{u}\in\mathcal{S}} |\langle \boldsymbol{a}(\theta) \mid \mathbf{u}\rangle| < \lambda \rightarrow$ test passed ✓
$\mathbf{y}$
$|\langle \mathbf{u} \mid \boldsymbol{a}(\theta)\rangle| = \lambda$
$\widehat{\mathbf{u}}$
$\boldsymbol{a}(\theta)$
$\mathbf{0}_m$


Θ
θ
$\mathbb{R}^m$
$\sup_{\mathbf{u}\in\mathcal{S}} |\langle \boldsymbol{a}(\theta) \mid \mathbf{u}\rangle| > \lambda \to$ test failed ✗
$|\langle \mathbf{u} \mid \boldsymbol{a}(\theta)\rangle| = \lambda$
$\mathbf{y}$
$\widehat{\mathbf{u}}$
$\mathbf{0}_m$

Two imperatives on the choice of a safe region

1. **Effectiveness:** $\mathcal{S}$ should be as small as possible (inclusion sense)
   if $\widehat{\mathbf{u}} \in \mathcal{S}_{\text{small}} \subseteq \mathcal{S}_{\text{big}}$ then

$$|\langle \boldsymbol{a}(\theta) \mid \widehat{\mathbf{u}} \rangle| \leq \sup_{\mathbf{u} \in \mathcal{S}_{\text{small}}} |\langle \boldsymbol{a}(\theta) \mid \mathbf{u} \rangle| \leq \sup_{\mathbf{u} \in \mathcal{S}_{\text{big}}} |\langle \boldsymbol{a}(\theta) \mid \mathbf{u} \rangle|$$

2. **Efficiency:** The evaluation of

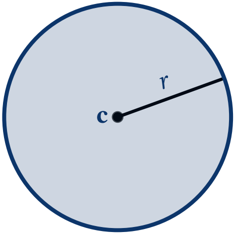
$$\sup_{\mathbf{u} \in \mathcal{S}} |\langle \boldsymbol{a}(\theta) \mid \mathbf{u} \rangle|$$

   should be done at *low* computational cost

Complexity[6]

Advocated geometry: ball

$$\mathcal{S} = \{\mathbf{u} \in \mathbb{R}^m : \ \|\mathbf{u} - \mathbf{c}\|_2 \leq r\}$$



In that case

$$\sup_{\mathbf{u} \in \mathcal{S}} |\langle \boldsymbol{a}(\theta) \mid \mathbf{u} \rangle| = |\langle \boldsymbol{a}(\theta) \mid \mathbf{c} \rangle| + r\|\boldsymbol{a}(\theta)\|_2$$

- ✓ Closed-form expression of the supremum
- ✓ Requires the evaluation of a single inner-product

---

[6] Other geometries considered in the literature: ellipsoids, (truncated) domes... all leading to different complexity/effectiveness trade-offs

The RYU Ball[7]

The RYU safe ball [Le *et al*, 2025]

Let $(\boldsymbol{x}, \mathbf{u})$ be a **feasible** primal-dual pair

Then, one can choose

$$\begin{aligned} \mathbf{c}_{\mathrm{RYU}} &= \frac{\mathbf{y} - \mathscr{M}\boldsymbol{x} \; + \mathbf{u}}{2} \\ r_{\mathrm{RYU}} &= \sqrt{\mathrm{GAP}(\boldsymbol{x}, \mathbf{u}) \; - \frac{1}{4}\| \; \mathbf{y} - \mathscr{M}\boldsymbol{x} \; - \mathbf{u}\|_2^2} \end{aligned}$$

In particular

$$\begin{aligned} \lim_{(\boldsymbol{x},\mathbf{u})\to(\hat{\boldsymbol{x}},\widehat{\mathbf{u}})} \; \mathbf{c}_{\mathrm{RYU}} &= \widehat{\mathbf{u}} \\ \lim_{(\boldsymbol{x},\mathbf{u})\to(\hat{\boldsymbol{x}},\widehat{\mathbf{u}})} \; r_{\mathrm{RYU}} &= 0 \end{aligned}$$

The RYU Ball[7]

The RYU safe ball [Le *et al*, 2025]

Let $(\boldsymbol{x}, \mathbf{u})$ be a **feasible** primal-dual pair

Then, one can choose

$$\mathbf{c}_{\mathrm{RYU}} = \frac{\overbrace{\mathbf{y} - \mathscr{M}\boldsymbol{x}}^{-\nabla f(\mathscr{M}\boldsymbol{x})} + \mathbf{u}}{2}$$

$$r_{\mathrm{RYU}} = \sqrt{\overset{\times L_f^{-1}}{\mathrm{GAP}(\boldsymbol{x}, \mathbf{u})} - \frac{1}{4}\|\overbrace{\mathbf{y} - \mathscr{M}\boldsymbol{x}}^{-\nabla f(\mathscr{M}\boldsymbol{x})} - \mathbf{u}\|_2^2}$$

In particular

$$\lim_{(\boldsymbol{x},\mathbf{u})\to(\hat{\boldsymbol{x}},\widehat{\mathbf{u}})} \mathbf{c}_{\mathrm{RYU}} = \widehat{\mathbf{u}}$$

$$\lim_{(\boldsymbol{x},\mathbf{u})\to(\hat{\boldsymbol{x}},\widehat{\mathbf{u}})} r_{\mathrm{RYU}} = 0$$

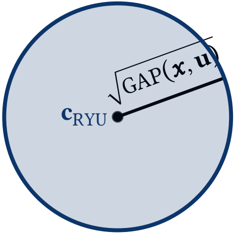
---

[7] Slightly modified result holds for any gradient Lipschitz loss f

To summarize: safe sphere screening test

Advocated geometry: ball

- Closed-form expression of the supremum
- Computationally cheap
- Collapses to $\{\widehat{\mathbf{u}}\}$ as $(\boldsymbol{x}, \mathbf{u}) \to (\hat{\boldsymbol{x}}, \widehat{\mathbf{u}})$



(Simplified) safe sphere screening rule

$$|\langle \boldsymbol{a}(\theta) \mid \mathbf{c}_{\mathrm{RYU}} \rangle| + \sqrt{\mathrm{GAP}(\mathbf{x}, \mathbf{u})} \|\boldsymbol{a}(\theta)\|_2 < \lambda \quad \Longrightarrow \quad \theta \notin \mathrm{support}(\hat{\boldsymbol{x}})$$

Addressing limitation 2 with joint screening

Joint screening tests

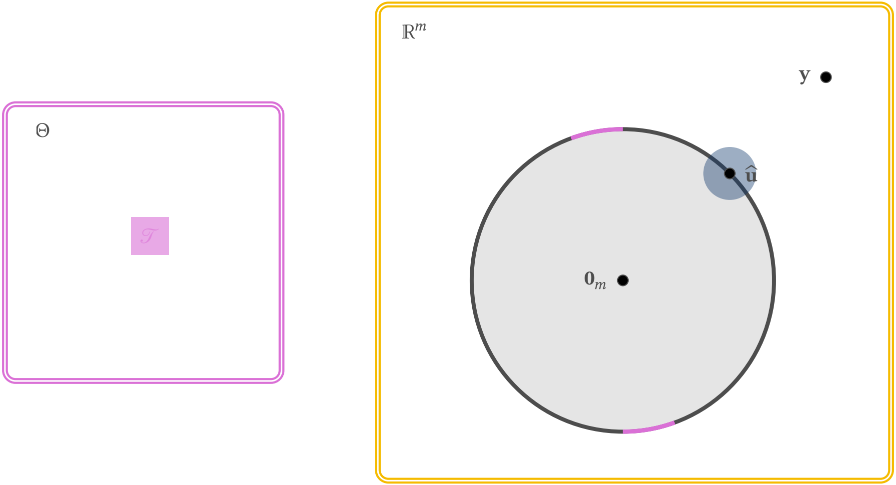
Rationale: tests are performed **jointly** for a **group of parameters**

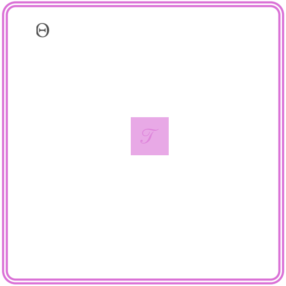
$$\mathscr{T} \subseteq \Theta$$

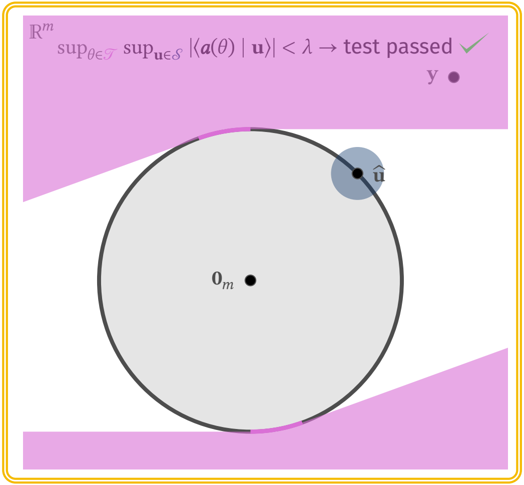
e.g.: an interval

Joint safe screening rule

$$\sup_{\theta \in \mathscr{T}} |\langle \boldsymbol{a}(\theta) \mid \mathbf{c} \rangle| + r\|\boldsymbol{a}(\theta)\|_2 < \lambda \quad \Longrightarrow \quad \mathscr{T} \cap \operatorname{support}(\hat{\boldsymbol{x}}) = \emptyset$$


Θ
$\mathscr{T}$
$\mathbb{R}^m$
$\mathbf{y}$
$\hat{\mathbf{u}}$
$\mathbf{0}_m$


Θ
$\mathscr{T}$


$\mathbb{R}^m$
$\sup_{\theta \in \mathscr{T}} \sup_{\mathbf{u} \in \mathcal{S}} |\langle \boldsymbol{a}(\theta) \mid \mathbf{u} \rangle| < \lambda \rightarrow$ test passed ✓
$\mathbf{y}$
$\hat{\mathbf{u}}$
$\mathbf{0}_m$

Challenge

The quantity

$$\sup_{\theta \in \mathscr{T}} \quad |\langle \boldsymbol{a}(\theta) \mid \mathbf{c} \rangle| \; + r \; \|\boldsymbol{a}(\theta)\|_2$$

cannot be evaluated in closed form in the general case
(because of the potential nonlinearity and non-convexity of $\boldsymbol{a}$)

Atom region

Proposed approach: Find $\mathscr{A} \subset \mathbb{R}^m$ such that

$$\{\boldsymbol{a}(\theta) :\ \theta \in \mathscr{T}\} \subseteq \mathscr{A}$$

Relaxed joint sphere screening rule

$$\sup_{\mathbf{a} \in \mathscr{A}} |\langle \mathbf{a} \mid \mathbf{c} \rangle| + r\|\mathbf{a}\|_2 < \lambda \quad \Longrightarrow \quad \mathscr{T} \cap \ \mathrm{support}(\hat{\boldsymbol{x}}) = \emptyset$$

admits a closed form solution if $\mathscr{A}$ has a “nice” geometry

Designing atom region should leverage analytical properties of the atom function[8]

Approach: exploit "regularity" of $\boldsymbol{a}$ — pick $\theta_0 \in \mathscr{T}$

$$\forall \theta \in \mathscr{T} \quad \|\boldsymbol{a}(\theta) - \boldsymbol{a}(\theta_0)\|_2 \leq L(\mathscr{T})\, \|\theta - \theta_0\|_2 \leq L(\mathscr{T})\, \mathrm{diam}(\mathscr{T})$$

"Ball" atom region (simplified)

$$\mathscr{A} \overset{\text{def.}}{=} \mathscr{B}(\boldsymbol{a}(\theta_0), L(\mathscr{T})\, \mathrm{diam}(\mathscr{T}))$$

---

[7] Other geometries includes "cylinders" or "ellipsoids", "union of balls", "cones", "domes", all exploiting specific one or several properties of $\boldsymbol{a}$

Implementable test with ball atom region and safe ball

Given $\mathscr{T} \subseteq \Theta$ and $\theta_0 \in \mathscr{T}$

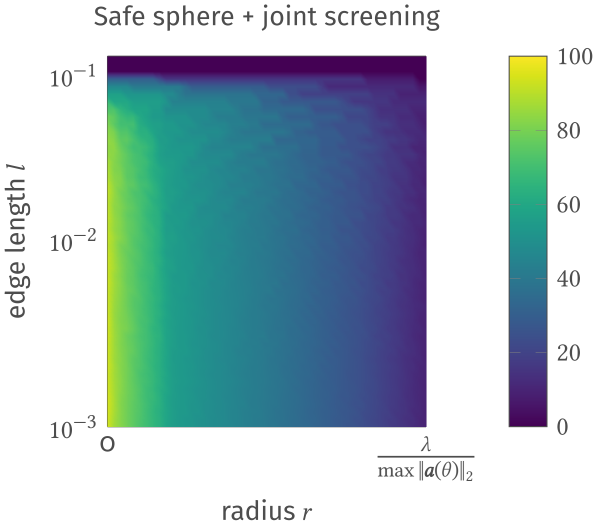
$$\underbrace{|\langle \boldsymbol{a}(\theta_0) \mid \mathbf{c}\rangle|}_{\xrightarrow{r \to 0} |\langle \boldsymbol{a}(\theta_0) \mid \widehat{\mathbf{u}}\rangle|} + \underbrace{L\|\mathbf{c}\|_2 \mathrm{diam}(\mathscr{T})}_{\xrightarrow{\mathscr{T} \to \{\theta_0\}} 0} + \underbrace{r(\|\mathbf{c}\|_2 + L(\mathscr{T})\mathrm{diam}(\mathscr{T}))}_{\xrightarrow{r \to 0} 0} < \lambda \quad \Longrightarrow \quad \mathscr{T} \cap \mathrm{support}(\hat{\boldsymbol{x}}) =$$

Numerical assessment of the method

- Sparse spike deconvolution
- "Gaussian" atom function with known variance
- BLASSO problem ($f = \frac{1}{2}\|\mathbf{y} - \cdot\|_2^2$)
- $\lambda = .5\lambda_{\max}$ ("intermediate" difficulty)

- Safe region: $\mathscr{S} = \mathrm{Ball}(\widehat{\mathbf{u}}, r)$
- Parameter region $\mathscr{T} = \{\theta_0\} + \left[-\frac{l}{2}, \frac{l}{2}\right]$ with l the (edge) length

Effectiveness VS Radius of atom region / safe ball



One step back

- ✓ **Foundation** of safe screening:

$$\theta \in \mathrm{support}(\hat{\boldsymbol{x}}) \quad \Longrightarrow \quad \theta \text{ is a maximizer of } \theta \mapsto |\langle \boldsymbol{a}(\theta) \mid \widehat{\mathbf{u}} \rangle|$$

- ✓ **Implementation:**

$$|\langle \boldsymbol{a}(\theta) \mid \widehat{\mathbf{u}} \rangle| < \lambda \quad \Longrightarrow \quad \theta \text{ is } \textbf{not} \text{ a maximizer of } \theta \mapsto |\langle \boldsymbol{a}(\theta) \mid \widehat{\mathbf{u}} \rangle|$$

can be interpreted as a “zeroth-order” optimality condition[9]

---

[9] zero because it does not involves gradient r subdifferential

Towards a first order safe screening rule

If $\boldsymbol{a}$ is assumed differentiable, we have

$$\forall \theta \in \mathrm{Int}(\Theta) \qquad \theta \in \mathrm{support}(\hat{\boldsymbol{x}}) \quad \Longrightarrow \quad \nabla_\theta \langle \boldsymbol{a}(\theta) \mid \widehat{\mathbf{u}} \rangle = \mathbf{0}_d$$

and thus

$$\nabla_\theta \langle \boldsymbol{a}(\theta) \mid \widehat{\mathbf{u}} \rangle \neq \mathbf{0}_d \quad \Longrightarrow \quad \theta \notin \mathrm{support}(\hat{\boldsymbol{x}})$$

reminiscent to a "local" technique used in [Flinth, 2023]

Towards a first order safe screening rule

If $\boldsymbol{a}$ is assumed differentiable, we have

$$\forall \theta \in \mathrm{Int}(\Theta) \qquad \theta \in \mathrm{support}(\hat{\boldsymbol{x}}) \quad \Longrightarrow \quad \nabla_\theta \langle \boldsymbol{a}(\theta) \mid \widehat{\mathbf{u}} \rangle = \mathbf{0}_d$$

and thus

$$\nabla_\theta \langle \boldsymbol{a}(\theta) \mid \widehat{\mathbf{u}} \rangle \neq \mathbf{0}_d \quad \Longrightarrow \quad \theta \notin \mathrm{support}(\hat{\boldsymbol{x}})$$

reminiscent to a "local" technique used in [Flinth, 2023]

Relaxed first-order safe screening rule

If $\mathscr{T} \subset \mathrm{Int}(\Theta)$ then

$$\inf_{\theta \in \mathscr{T}} \inf_{\mathbf{u} \in \mathcal{S}} \|\nabla_\theta \langle \boldsymbol{a}(\theta) \mid \mathbf{u} \rangle\| > 0 \quad \Longrightarrow \quad \mathscr{T} \cap \mathrm{support}(\hat{\boldsymbol{x}}) = \emptyset$$

Numerical assessment

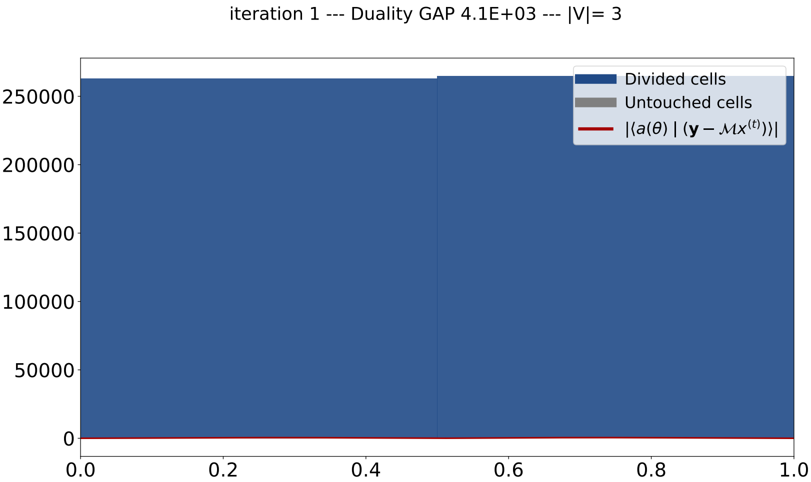
- **Setup:** running example
- **Solver:** adaptive Refinement Algorithm [Flinth, 2023]

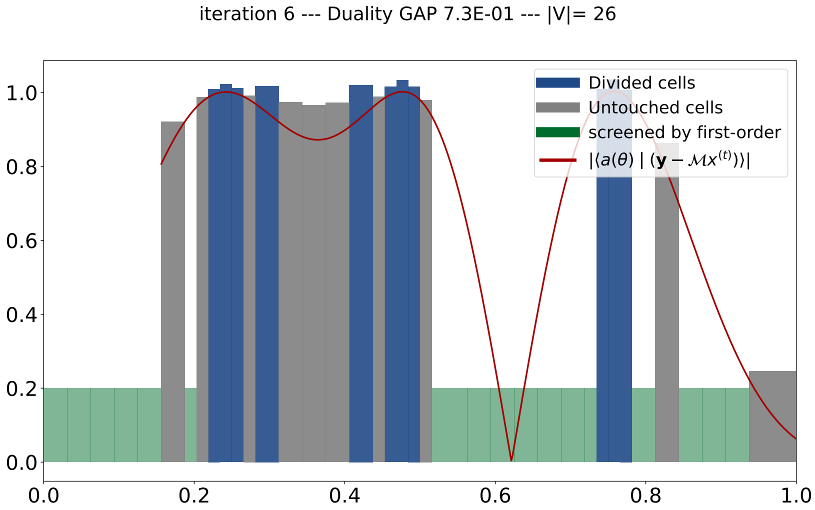
  iterative refinement of a grid
- **Safe region:** RYU ball with parameter $(\boldsymbol{x}^{(t)}, \kappa^{(t)}(\mathbf{y} - \mathscr{M}\boldsymbol{x}^{(t)}))$

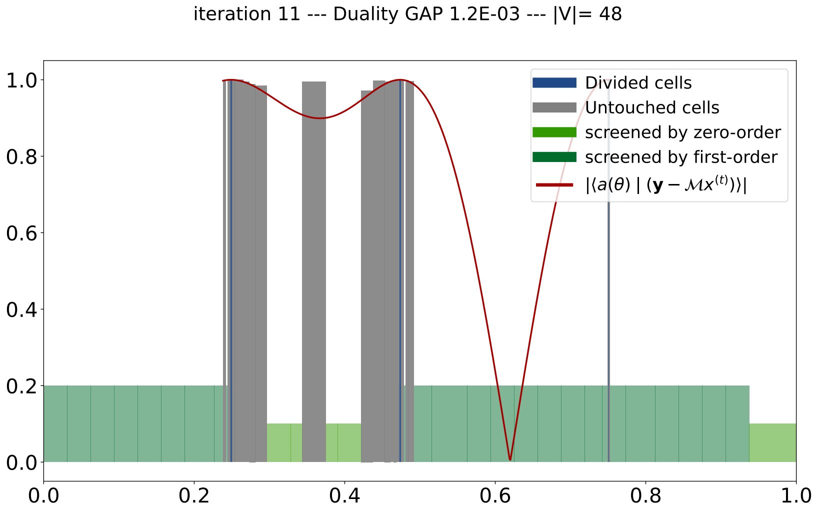
  t *is the iteration number*

  $\kappa^{(t)}$ *is a multiplicative factor to ensure dual feasibility*
- **Parameter region:** same as solver

  of the form $\theta_0 + [-l/2, l/2]$, l *is the edge length*


iteration 1 --- Duality GAP 4.1E+03 --- |V|= 3
Divided cells
Untouched cells
$|\langle a(\theta) \mid (\mathbf{y} - \mathcal{M}x^{(t)})\rangle|$
250000
200000
150000
100000
50000
0
0.0
0.2
0.4
0.6
0.8
1.0


iteration 6 --- Duality GAP 7.3E-01 --- |V|= 26


iteration 11 --- Duality GAP 1.2E-03 --- |V|= 48

Advertisement

arXiv Preprint + Python / C++ Toolbox

available within weeks

Stay tuned!
`https://c-elvira.github.io/`

Concluding words

Includes work with

Any question?

Appendices

Remark: tests inclusion

- If $\theta \in \mathrm{Int}(\Theta)$, we have

$$\|\nabla_\theta \langle \boldsymbol{a}(\theta) \mid \widehat{\mathbf{u}} \rangle\| > 0 \quad \Longrightarrow \quad |\langle \boldsymbol{a}(\theta) \mid \widehat{\mathbf{u}} \rangle| < \lambda$$

  i.e.,

  θ passes the “First-order” test $\implies$ θ passes the “zero-order” test

- Does not hold anymore when comparing relaxed tests!