Safe screening
an introduction and perspectives

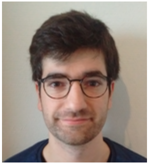
Clément Elvira

CentraleSupélec, SCEE group

9 décembre 2021

Joint work with...



Context and motivations

Inverse / Learning problems with linear models

Given

- Observation $\mathbf{y} \in \mathbb{R}^m$
- Linear model $\mathbf{M} : \mathbb{R}^n \longrightarrow \mathbb{R}^m$

Typical case: $m \ll n$

Goal: recover / find $\mathbf{x}_0$ such that $\mathbf{y} \simeq \mathbf{M}\mathbf{x}_0$

$$\underset{\mathbf{x}\in\mathbb{R}^n}{\arg\min} \; L(\mathbf{y}, \mathbf{M}\mathbf{x})$$

Inverse / Learning problems with linear models

Given

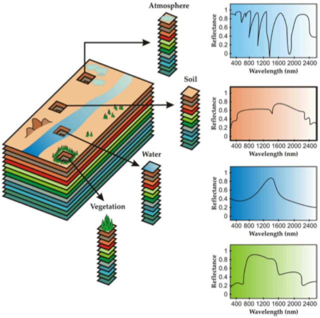
- Observation $\mathbf{y} \in \mathbb{R}^m$
- Linear model $\mathbf{M} \colon \mathbb{R}^n \longrightarrow \mathbb{R}^m$

Typical case: $m \ll n$

Goal: recover / find $\mathbf{x}_0$ such that $\mathbf{y} \simeq \mathbf{M}\mathbf{x}_0$

$$\operatorname*{arg\,min}_{\mathbf{x} \in \mathbb{R}^n} \; L(\mathbf{y}, \mathbf{M}\mathbf{x}) \overset{e.g.}{=} \frac{1}{2}\|\mathbf{y} - \mathbf{M}\mathbf{x}\|_2^2$$

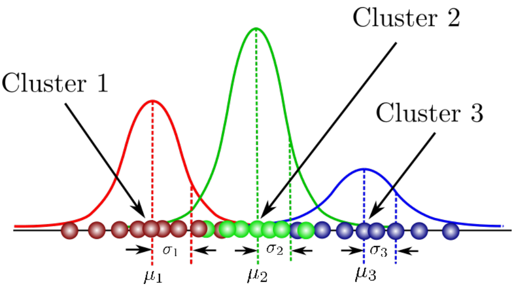
Example 1: Hyperspectral unmixing



Courtesy of J. Bioucas-Dias

- $\mathbf{y} \equiv$ one pixel
- $\mathbf{x} \equiv$ Proportions
- $\mathbf{M} \equiv$ Elementary spectrum

Example 2: Mixture model fitting



Courtesy of scikit-learn

- $\mathbf{y} \equiv$ Samples
- $\mathbf{x} \equiv$ Proportions
- $\mathbf{M} \equiv$ Gaussian curves with various parameterization

Inverse / Learning problems with linear models

Given

- Observation $\mathbf{y} \in \mathbb{R}^m$
- Linear model $\mathbf{M} : \mathbb{R}^n \longrightarrow \mathbb{R}^m$

Typical case: $m \ll n$

Goal: recover / find $\mathbf{x}_0$ such that $\mathbf{y} \simeq \mathbf{M}\mathbf{x}_0$

$$\operatorname*{arg\,min}_{\mathbf{x}\in\mathbb{R}^n} \; L(\mathbf{y}, \mathbf{M}\mathbf{x}) \overset{e.g.}{=} \frac{1}{2}\|\mathbf{y} - \mathbf{M}\mathbf{x}\|_2^2$$

Infinite number of solutions 😱

$\longrightarrow$ **ill posed problem**

Penalized problem

Penalized problem

$$\mathbf{x}^\star \in \arg\min_{\mathbf{x}\in\mathbb{R}^n} L(\mathbf{y}, \mathbf{M}\mathbf{x}) + R_{\mathrm{eg}}(\mathbf{x})$$

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

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

Penalized problem

Penalized problem

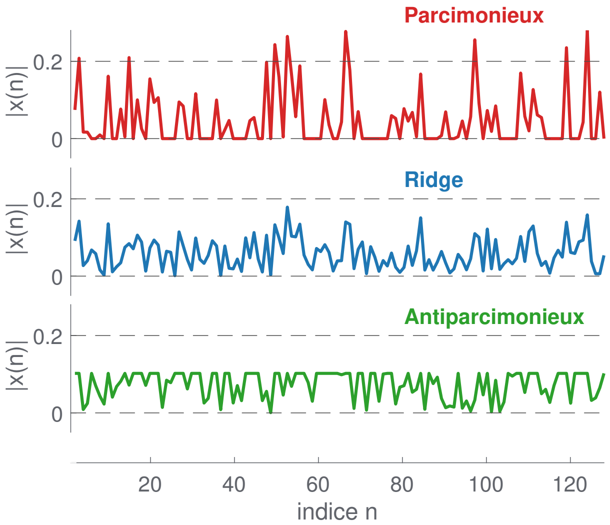
$$\mathbf{x}^\star \in \operatorname*{arg\,min}_{\mathbf{x}\in\mathbb{R}^n} L(\mathbf{y}, \mathbf{M}\mathbf{x}) + R_{\mathrm{eg}}(\mathbf{x})$$

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

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

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

Example of regularizers



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

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

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

Of particular interest: the Lasso problem 1/3

$$\mathbf{x}^\star \in \operatorname*{arg\,min}_{\mathbf{x}\in\mathbb{R}^n} \tfrac{1}{2}\|\mathbf{y} - \mathbf{M}\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_1 = P_\lambda(\mathbf{x})$$

known to promote "**sparse**" solutions

Of particular interest: the Lasso problem 1/3

$$\mathbf{x}^\star \in \operatorname*{arg\,min}_{\mathbf{x}\in\mathbb{R}^n} \ \tfrac{1}{2}\|\mathbf{y} - \mathbf{M}\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_1 = P_\lambda(\mathbf{x})$$

known to promote "**sparse**" solutions

Sparsity & the Lasso *(fuzzy version)* *Foucart and Rauhut, 2013*

$\mathbf{x}^\star$ has (at most) m nonzero entries

$\longrightarrow$ slightly more subtle if multiple minimizers

$\longrightarrow$ Solutions are sparse (recall $m \ll n$)

Of particular interest: the Lasso problem 2/3

Typical solver *(e.g. Ista Beck and Teboulle, 2009)*

Starting from $\mathbf{x}^{(0)}$, **repeat**

1. Evaluate residual error $\mathbf{r}^{(t)} = \mathbf{y} - \mathbf{M}\mathbf{x}^{(t)}$
2. Evaluate "residual correlations" $\mathbf{M}^\top \mathbf{r}^{(t)}$
3. Update $\mathbf{x}^{(t)}$ according to some rule

Of particular interest: the Lasso problem 2/3

Typical solver *(e.g. Ista Beck and Teboulle, 2009)*

Starting from $\mathbf{x}^{(0)}$, **repeat**	Complexity
1. Evaluate residual error $\mathbf{r}^{(t)} = \mathbf{y} - \mathbf{M}\mathbf{x}^{(t)}$	$\mathcal{O}(mn)$
2. Evaluate "residual correlations" $\mathbf{M}^\top\mathbf{r}^{(t)}$	$\mathcal{O}(mn)$
3. Update $\mathbf{x}^{(t)}$ according to some rule	TBD

Findings:

- Scale (at least) linearly with n
- Many calculations are useless 😱
  especially since $\mathbf{x}^\star$ is m-sparse

Of particular interest: the Lasso problem 2/3

Typical solver *(e.g. Ista Beck and Teboulle, 2009)*

Starting from $\mathbf{x}^{(0)}$, **repeat**	Complexity
1. Evaluate residual error $\mathbf{r}^{(t)} = \mathbf{y} - \mathbf{M}\mathbf{x}^{(t)}$	$\mathcal{O}(mn)$
2. Evaluate "residual correlations" $\mathbf{M}^{\top}\mathbf{r}^{(t)}$	$\mathcal{O}(mn)$
3. Update $\mathbf{x}^{(t)}$ according to some rule	TBD

Findings:

- Scale (at least) linearly with n
- Many calculations are useless 😱
  especially since $\mathbf{x}^\star$ is m-sparse
- Shall we remove "useless" columns of $\mathbf{M}$? 🤔

Column elimination as dimensionality reduction

🤔 Removing "useless" columns of $\mathbf{M}$: no impact on the solution ? 🤔

Fact

If $\mathbf{x}^\star(\ell) = 0$ for all $\ell \in \mathcal{S}$

Solving the Lasso

$$\iff$$

$$\begin{cases} \mathbf{s}^\star \in \underset{\mathbf{s}\in\mathbb{R}^{n-\mathrm{card}(\mathcal{S})}}{\arg\min} \ \frac{1}{2}\|\mathbf{y} - \mathbf{M}_{\backslash\mathcal{S}}\mathbf{s}\|_2^2 + \lambda\|\mathbf{s}\|_1 \\ \mathbf{x}^\star_{\backslash\mathcal{S}} = \mathbf{s}^\star \\ \mathbf{x}^\star_{\mathcal{S}} = 0 \end{cases}$$

Column elimination as dimensionality reduction

🤔 Removing "useless" columns of $\mathbf{M}$: no impact on the solution ? 🤔

Fact

If $\mathbf{x}^\star(\ell) = 0$ for all $\ell \in \mathcal{S}$

Solving the Lasso

$$\Longleftrightarrow$$

$$\begin{cases} \mathbf{s}^\star \in \underset{\mathbf{s}\in\mathbb{R}^{n-\mathrm{card}(\mathcal{S})}}{\arg\min} \ \frac{1}{2}\|\mathbf{y} - \mathbf{M}_{\setminus\mathcal{S}}\mathbf{s}\|_2^2 + \lambda\|\mathbf{s}\|_1 \\ \mathbf{x}^\star_{\setminus\mathcal{S}} = \mathbf{s}^\star \\ \mathbf{x}^\star_{\mathcal{S}} = 0 \end{cases}$$

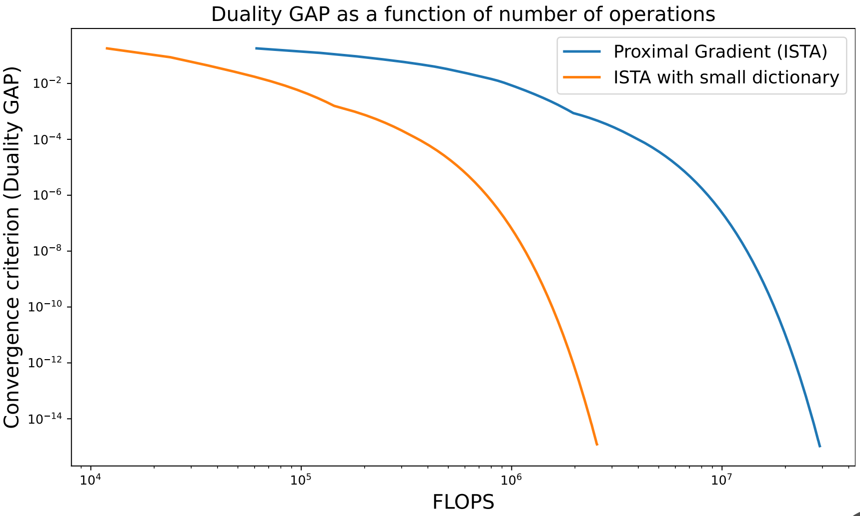
🎉 Equivalent **lower-dimensional** Lasso problem 🎉

Of particular interest: the Lasso problem 3/3

$(m, n) = (100, 150)$, "Gaussian dictionary", 20 repetitions

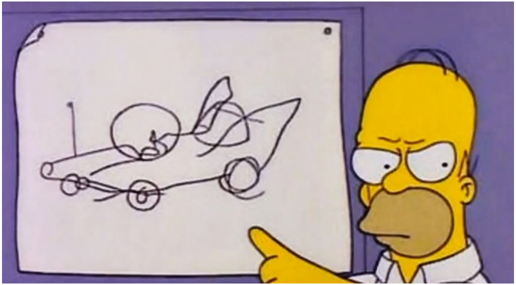
Of particular interest: the Lasso problem 3/3

$(m, n) = (100, 150)$, “Gaussian dictionary”, 20 repetitions



Towards safe screening

- Design of **screening tests** to detect **zero entries** in $\mathbf{x}^\star$
  Introduced by El ghaoui et al in 2013
- The tests must be "**safe**"
  In contract with strong screening
- The test must be computationally **cheap**
  Recall that we want to accelerate a solver



Safe screening for Lasso 101

The crux of screening: the residual error

- Say one has already found $\mathbf{x}^\star$
- Denote $\mathbf{r}^\star = \mathbf{y} - \mathbf{M}\mathbf{x}^\star$ the residual error
- Let $\mathbf{m}_\ell$ be an **unused** atom *(i.e., $\mathbf{x}^\star(\ell) = 0$)*

The crux of screening: the residual error

- Say one has already found $\mathbf{x}^\star$
- Denote $\mathbf{r}^\star = \mathbf{y} - \mathbf{M}\mathbf{x}^\star$ the residual error
- Let $\mathbf{m}_\ell$ be an **unused** atom *(i.e., $\mathbf{x}^\star(\ell) = 0$)*

$$P_\lambda(\mathbf{x}^\star + x_\ell \mathbf{m}_\ell) \;=\; \tfrac{1}{2}\|\mathbf{r}^\star - x_\ell \mathbf{m}_\ell\|_2^2 + \lambda\|\mathbf{x}^\star\|_1 + \lambda|x_\ell|$$

The crux of screening: the residual error

- Say one has already found $\mathbf{x}^\star$
- Denote $\mathbf{r}^\star = \mathbf{y} - \mathbf{M}\mathbf{x}^\star$ the residual error
- Let $\mathbf{m}_\ell$ be an **unused** atom *(i.e., $\mathbf{x}^\star(\ell) = 0$)*

$$
\begin{aligned}
P_\lambda(\mathbf{x}^\star + x_\ell \mathbf{m}_\ell) &= \tfrac{1}{2}\|\mathbf{r}^\star - x_\ell \mathbf{m}_\ell\|_2^2 + \lambda\|\mathbf{x}^\star\|_1 + \lambda|x_\ell| \\
&= P_\lambda(\mathbf{x}^\star) + \tfrac{1}{2}x_\ell^2 - \langle \mathbf{r}^\star, \mathbf{m}_\ell\rangle x_\ell + \lambda|x_\ell|
\end{aligned}
$$

The crux of screening: the residual error

- Say one has already found $\mathbf{x}^\star$
- Denote $\mathbf{r}^\star = \mathbf{y} - \mathbf{M}\mathbf{x}^\star$ the residual error
- Let $\mathbf{m}_\ell$ be an **unused** atom *(i.e., $\mathbf{x}^\star(\ell) = 0$)*

$$\begin{aligned} P_\lambda(\mathbf{x}^\star + x_\ell \mathbf{m}_\ell) &= \tfrac{1}{2}\|\mathbf{r}^\star - x_\ell \mathbf{m}_\ell\|_2^2 + \lambda\|\mathbf{x}^\star\|_1 + \lambda|x_\ell| \\ &= P_\lambda(\mathbf{x}^\star) + \tfrac{1}{2}x_\ell^2 - \langle \mathbf{r}^\star, \mathbf{m}_\ell\rangle x_\ell + \lambda|x_\ell| \end{aligned}$$

Tempting conclusion

$$|\langle \mathbf{r}^\star, \mathbf{m}_\ell\rangle| < \lambda \quad \Longrightarrow \quad \mathbf{x}^\star(\ell) = 0?$$

Outline

1. Screening rule even when $\mathbf{x}^\star$ has not been identified?

2. Mathematically grounded framework?

Suitable framework: Fenchel-Rockafellar duality – 1/2

Primal optimization problem *(the Lasso)*

$$\text{Find } \mathbf{x}^\star \in \operatorname*{arg\,min}_{\mathbf{x}\in\mathbb{R}^n} \; P_\lambda(\mathbf{x}) = \tfrac{1}{2}\|\mathbf{y} - \mathbf{M}\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_1$$

Suitable framework: Fenchel-Rockafellar duality – 1/2

Primal optimization problem *(the Lasso)*

$$\text{Find } \mathbf{x}^\star \in \underset{\mathbf{x}\in\mathbb{R}^n}{\arg\min}\ P_\lambda(\mathbf{x}) = \frac{1}{2}\|\mathbf{y} - \mathbf{M}\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_1$$

Dual optimization problem

$$\text{Find } \mathbf{r}^\star = \underset{\mathbf{r}\in\mathbb{R}^m}{\arg\max}\ D(\mathbf{r}) = \frac{1}{2}\|\mathbf{y}\|_2^2 - \frac{1}{2}\|\mathbf{y} - \mathbf{r}\|_2^2$$

Such that

$$|\langle \mathbf{r}, \mathbf{m}_j\rangle| \leq \lambda \qquad \forall j = 1 \dots, n$$

Denote $\mathcal{R}$ the constrained set

Strong duality: $\texttt{GAP}(\mathbf{x}^\star, \mathbf{r}^\star) = P_\lambda(\mathbf{x}^\star) - D(\mathbf{r}^\star) = 0$

Suitable framework: Fenchel-Rockafellar duality – 1/2

Primal optimization problem *(the Lasso)*

$$\text{Find } \mathbf{x}^\star \in \operatorname*{arg\,min}_{\mathbf{x}\in\mathbb{R}^n} \; P_\lambda(\mathbf{x}) = \tfrac{1}{2}\|\mathbf{y} - \mathbf{M}\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_1$$

Dual optimization problem

$$\text{Find } \mathbf{r}^\star = \operatorname*{arg\,min}_{\mathbf{r}\in\mathbb{R}^m} \; \|\mathbf{y} - \mathbf{r}\|_2^2$$

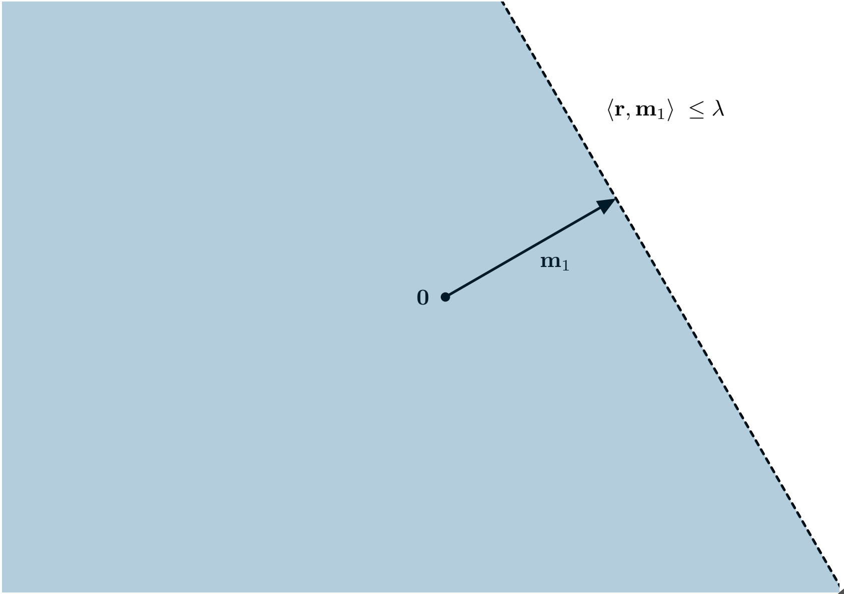
Such that

$$|\langle \mathbf{r}, \mathbf{m}_j \rangle| \leq \lambda \qquad \forall j = 1\ldots, n$$

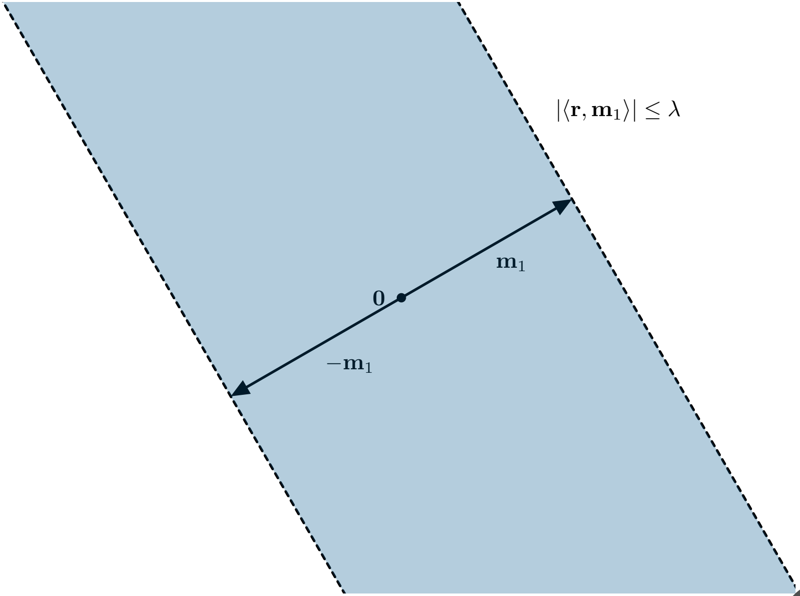
Denote $\mathcal{R}$ the constrained set

Strong duality: $\mathtt{GAP}(\mathbf{x}^\star, \mathbf{r}^\star) = P_\lambda(\mathbf{x}^\star) - D(\mathbf{r}^\star) = 0$

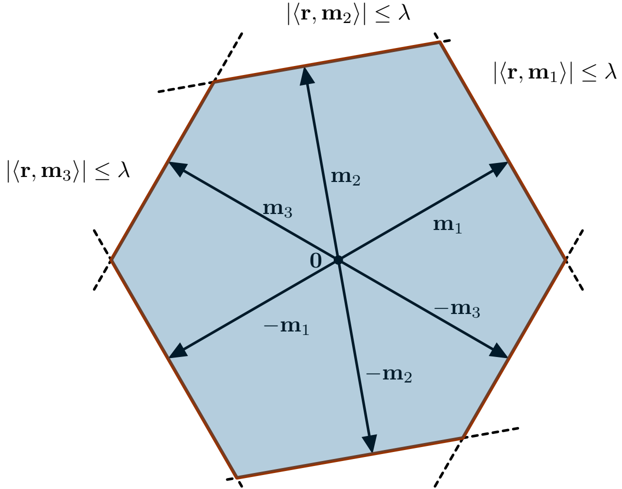
Geometric interpretation



Geometric interpretation

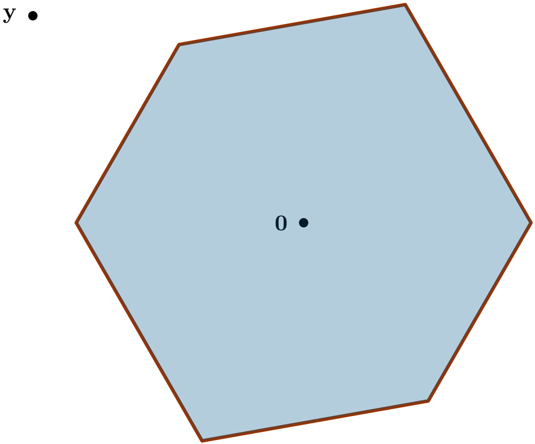


Geometric interpretation



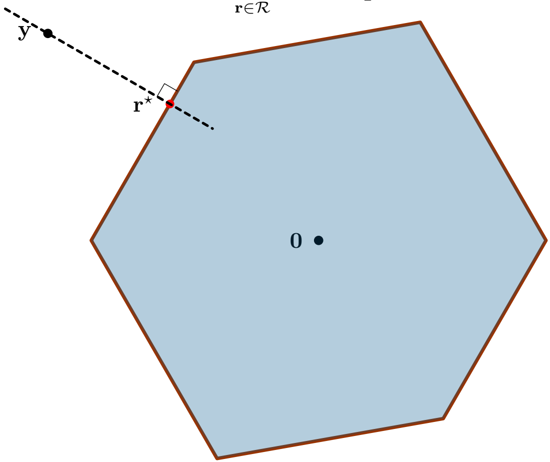
Geometric interpretation

$$\mathbf{r}^\star = \arg\min_{\mathbf{r}\in\mathcal{R}} \|\mathbf{y} - \mathbf{r}\|_2^2$$



Geometric interpretation

$$\mathbf{r}^\star = \arg\min_{\mathbf{r}\in\mathcal{R}} \|\mathbf{y}-\mathbf{r}\|_2^2$$



Suitable framework: Fenchel-Rockafellar duality – 2/2

Primal dual link

if $(\mathbf{x}^\star, \mathbf{r}^\star)$ is a couple of primal / dual solutions then

$$\mathbf{r}^\star = \mathbf{y} - \mathbf{M}\mathbf{x}^\star$$

Suitable framework: Fenchel-Rockafellar duality – 2/2

Primal dual link

if $(\mathbf{x}^\star, \mathbf{r}^\star)$ is a couple of primal / dual solutions then

$$\mathbf{r}^\star = \mathbf{y} - \mathbf{M}\mathbf{x}^\star$$

Optimality condition *(Fermat's rule)*

$(\mathbf{x}^\star, \mathbf{r}^\star)$ is a couple of primal / dual solutions **if and only if**

$$\nabla P_\lambda(\mathbf{x}^\star) = \mathbf{0}_n$$

Suitable framework: Fenchel-Rockafellar duality – 2/2

Primal dual link

if $(\mathbf{x}^\star, \mathbf{r}^\star)$ is a couple of primal / dual solutions then

$$\mathbf{r}^\star = \mathbf{y} - \mathbf{M}\mathbf{x}^\star$$

Optimality condition *(Fermat's rule)*

$(\mathbf{x}^\star, \mathbf{r}^\star)$ is a couple of primal / dual solutions **if and only if**

$$\cancel{\nabla P_\lambda(\mathbf{x}^\star) = \mathbf{0}_n} \qquad \text{(not differentiable)}$$

Suitable framework: Fenchel-Rockafellar duality – 2/2

Primal dual link

if $(\mathbf{x}^\star, \mathbf{r}^\star)$ is a couple of primal / dual solutions then

$$\mathbf{r}^\star = \mathbf{y} - \mathbf{M}\mathbf{x}^\star$$

Optimality condition *(Fermat's rule)*

$(\mathbf{x}^\star, \mathbf{r}^\star)$ is a couple of primal / dual solutions **if and only if**

$$\mathbf{0}_n \in \partial P_\lambda(\mathbf{x}^\star)$$

Suitable framework: Fenchel-Rockafellar duality – 2/2

Primal dual link

if $(\mathbf{x}^\star, \mathbf{r}^\star)$ is a couple of primal / dual solutions then

$$\mathbf{r}^\star = \mathbf{y} - \mathbf{M}\mathbf{x}^\star$$

Optimality condition *(Fermat's rule)*

$(\mathbf{x}^\star, \mathbf{r}^\star)$ is a couple of primal / dual solutions **if and only if**

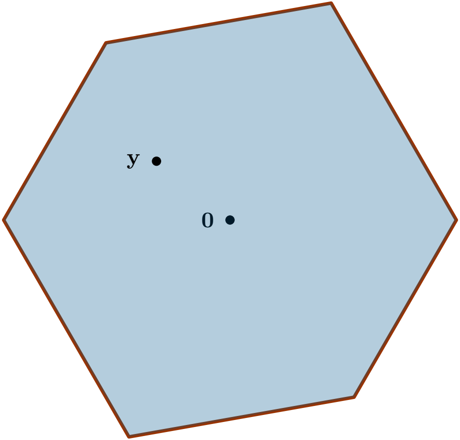
$$\langle \mathbf{r}^\star, \mathbf{m}_j \rangle = \mathrm{sign}\left(\mathbf{x}^\star(j)\right)\lambda$$

$\mathrm{sign}(0) \in [-1, 1]$

A first simple & safe screening rule

- Consider the case where $\mathbf{y} \in \mathcal{R}$

That is $|\langle \mathbf{y}, \mathbf{m}_j \rangle| \leq \lambda \quad (\forall j = 1, \ldots, n)$



A first simple & safe screening rule

- Consider the case where $\mathbf{y} \in \mathcal{R}$

That is $|\langle \mathbf{y}, \mathbf{m}_j \rangle| \leq \lambda \quad (\forall j = 1, \dots, n)$

- Hence $(\mathbf{0}_n, \mathbf{y})$ satisfies the Fermat's rule (CNS)

Rappel

$$\langle \mathbf{r}^\star, \mathbf{m}_j \rangle = \operatorname{sign}\left(\mathbf{x}^\star(j)\right)\lambda$$

$\operatorname{sign}(0) \in [-1, 1]$

A first simple & safe screening rule

- Consider the case where $\mathbf{y} \in \mathcal{R}$

  That is $|\langle \mathbf{y}, \mathbf{m}_j \rangle| \leq \lambda \quad (\forall j = 1, \ldots, n)$

- Hence $(\mathbf{0}_n, \mathbf{y})$ satisfies the Fermat's rule (CNS)

- **Conclusion:** $\mathbf{0}_n$ is a minimizer! 🎉

A first simple & safe screening rule

- Consider the case where $\mathbf{y} \in \mathcal{R}$

That is $|\langle \mathbf{y}, \mathbf{m}_j \rangle| \leq \lambda \quad (\forall j = 1, \dots, n)$

- Hence $(\mathbf{0}_n, \mathbf{y})$ satisfies the Fermat's rule (CNS)

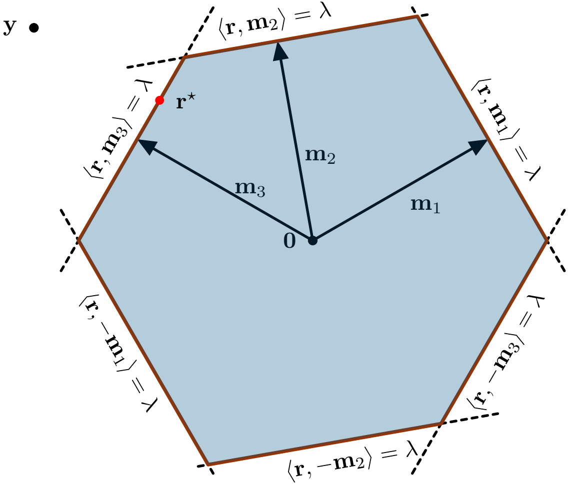
- **Conclusion:** $\mathbf{0}_n$ is a minimizer! 🎉

"All-zero" safe screening rule

$$\|\mathbf{M}^{\mathsf{T}}\mathbf{y}\|_\infty \leq \lambda \quad \Longrightarrow \quad \mathbf{x}^\star(\ell) = 0 \quad \forall \ell = 1, \dots, n$$

🤔 Practical interest?

Geometric interpretation vs Fermat's rule



Rule (simplified)

$$|\langle \mathbf{r}^\star, \mathbf{m}_j \rangle| \leq \lambda$$

Or

$$|\langle \mathbf{r}^\star, \mathbf{m}_j \rangle| = \lambda$$

Safe screening rule *(the real one)*

- **Finding:** $\mathbf{x}^\star(\ell) \neq 0 \Longrightarrow |\langle \mathbf{r}^\star, \mathbf{m}_\ell \rangle| = \lambda$

Safe screening rule *El Ghaoui et al. (2012)*

$$|\langle \mathbf{r}^\star, \mathbf{m}_\ell \rangle| < \lambda \quad \Longrightarrow \quad \mathbf{x}^\star(\ell) = 0$$

Safe screening rule *(the real one)*

- **Finding:** $\mathbf{x}^\star(\ell) \neq 0 \Longrightarrow |\langle \mathbf{r}^\star, \mathbf{m}_\ell \rangle| = \lambda$

Safe screening rule *El Ghaoui et al. (2012)*

$$|\langle \mathbf{r}^\star, \mathbf{m}_\ell \rangle| < \lambda \quad \Longrightarrow \quad \mathbf{x}^\star(\ell) = 0$$

- Nothing else than a **contraposition**
- Not a heuristic! *(hence the S-word)*
- Independent from the minimizer $\mathbf{x}^\star$
- Computationally simple

  only involves one inner product 😘

From safe region to safe sphere

😱 **Bad news:** Finding $\mathbf{r}^\star$ is as difficult as finding $\mathbf{x}^\star$ 😱

From safe region to safe sphere

😱 **Bad news:** Finding $\mathbf{r}^\star$ is as difficult as finding $\mathbf{x}^\star$ 😱

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

From safe region to safe sphere

Bad news: Finding $\mathbf{r}^\star$ is as difficult as finding $\mathbf{x}^\star$

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

Notion of "safe region" *El Ghaoui et al. (2012)*

A set $\mathcal{R}_s \subset \mathbb{R}^m$ is a Safe region iff $\mathbf{r}^\star \in \mathcal{R}_s$

From safe region to safe sphere

😱 **Bad news:** Finding $\mathbf{r}^\star$ is as difficult as finding $\mathbf{x}^\star$ 😱

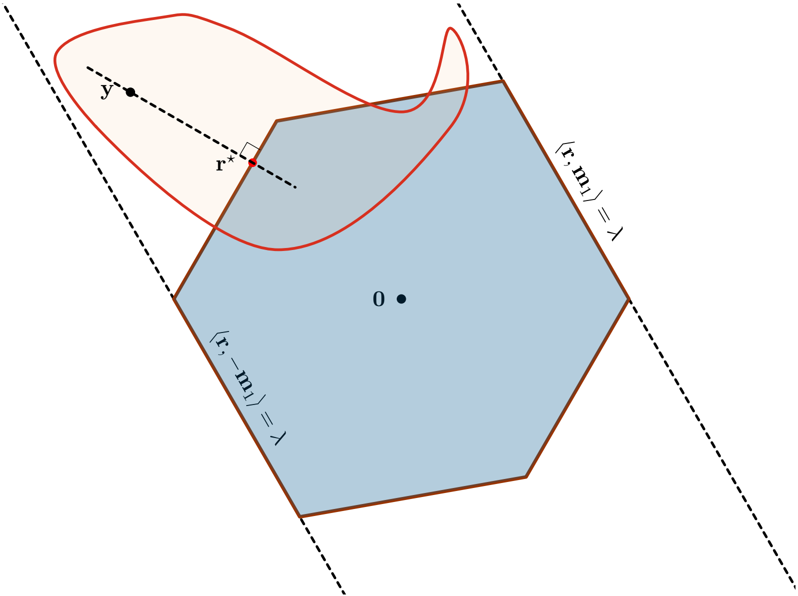
Idea: perform the test without computing $\mathbf{r}^\star$

Notion of "safe region" *El Ghaoui et al. (2012)*

A set $\mathcal{R}_s \subset \mathbb{R}^m$ is a <u>Safe region</u> iff $\mathbf{r}^\star \in \mathcal{R}_s$

$$\max_{\mathbf{r}\in\mathcal{R}_s} |\langle \mathbf{r}, \mathbf{m}_\ell \rangle| < \lambda \quad \Longrightarrow \quad |\langle \mathbf{r}^\star, \mathbf{m}_\ell \rangle| < \lambda$$

$$\Longrightarrow \quad \mathbf{x}^\star(\ell) = 0$$



$\mathbf{y}$
$\mathbf{r}^{\star}$
$\mathbf{0}$
$\langle \mathbf{r}, \mathbf{m}_1 \rangle = \lambda$
$\langle \mathbf{r}, -\mathbf{m}_1 \rangle = \lambda$

From safe region to safe sphere

Bad news: Finding $\mathbf{r}^\star$ is as difficult as finding $\mathbf{x}^\star$

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

Notion of "safe region" *El Ghaoui et al. (2012)*

A set $\mathcal{R}_s \subset \mathbb{R}^m$ is a <u>Safe region</u> iff $\mathbf{r}^\star \in \mathcal{R}_s$

$$\max_{\mathbf{r}\in\mathcal{R}_s} |\langle \mathbf{r}, \mathbf{m}_\ell \rangle| < \lambda \quad \Longrightarrow \quad |\langle \mathbf{r}^\star, \mathbf{m}_\ell \rangle| < \lambda$$

$$\Longrightarrow \quad \mathbf{x}^\star(\ell) = 0$$

<u>**Today's example: safe sphere**</u>

$$\mathcal{S} = \mathcal{B}(\mathbf{c}, r) \qquad \text{and} \qquad \max_{\mathbf{r}\in\mathcal{B}(\mathbf{c},r)} |\langle \mathbf{r}, \mathbf{m}_\ell \rangle| = |\langle \mathbf{c}, \mathbf{m}_\ell \rangle| + r\|\mathbf{m}_\ell\|_2$$

Closed-form expression!

Safe sphere design

Goal

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

Safe sphere design

Dual problem

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

$\longrightarrow$ **Projection** onto the convex set $\mathcal{R}$!

Safe sphere design

Dual problem

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

$\longrightarrow$ **Projection** onto the convex set $\mathcal{R}$!

If one **knows** some $\mathbf{r}_0 \in \mathcal{R}$, then by **definition**

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

$\Longrightarrow$ $\mathbf{r}^\star$ belongs to a **Sphere**!

Safe sphere design

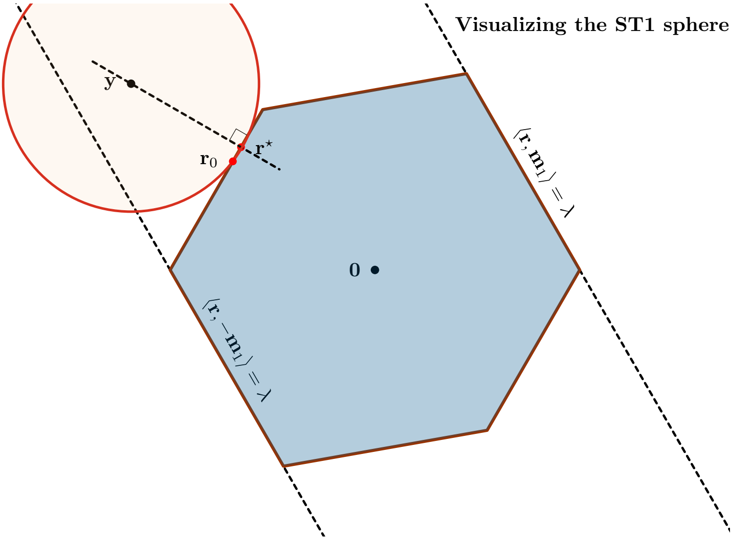
Goal

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

ST 1: Choose $\mathbf{r}_0 \in \mathcal{R}$ (*e.g.* $\mathbf{0}_m$)

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

typical use: done once for all before runtime



Visualizing the ST1 sphere
$\mathbf{y}$
$\mathbf{r}^\star$
$\mathbf{r}_0$
$\mathbf{0}$
$\langle \mathbf{r}, \mathbf{m}_1 \rangle = \lambda$
$\langle \mathbf{r}, -\mathbf{m}_1 \rangle = \lambda$

Safe sphere design

Goal

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

ST 1: Choose $\mathbf{r}_0 \in \mathcal{R}$ (*e.g.* $\mathbf{0}_m$)

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

typical use: done once for all before runtime

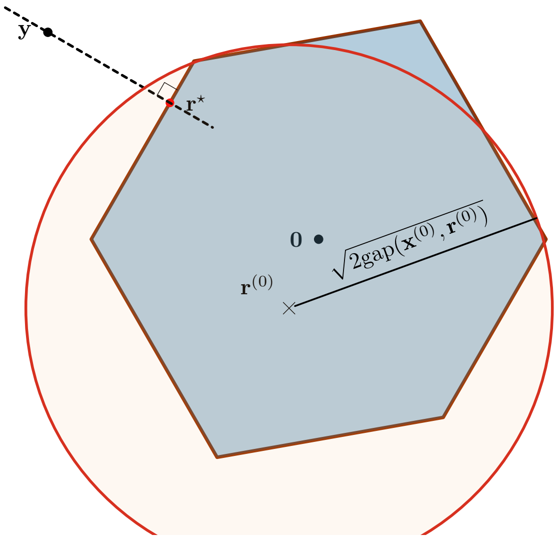
GAP sphere: Choose $\mathbf{x}_0 \in \mathbb{R}^n$ *[Fercoq et al, 2015]*

$$\begin{aligned} \mathbf{c} &= \Phi_{\mathcal{R}}(\mathbf{y} - \mathbf{M}\mathbf{x}_0) \\ r &= \sqrt{2\mathtt{GAP}(\mathbf{x}_0, \mathbf{c})} \end{aligned}$$

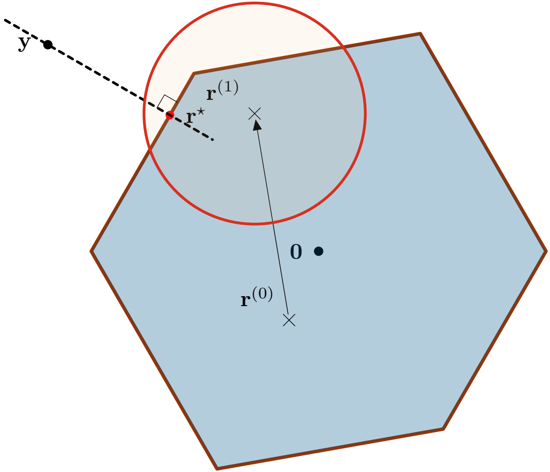
typical use:

- Dynamically: $\mathbf{x}_0 = \mathbf{x}^{(t)}$
- radius tends to 0

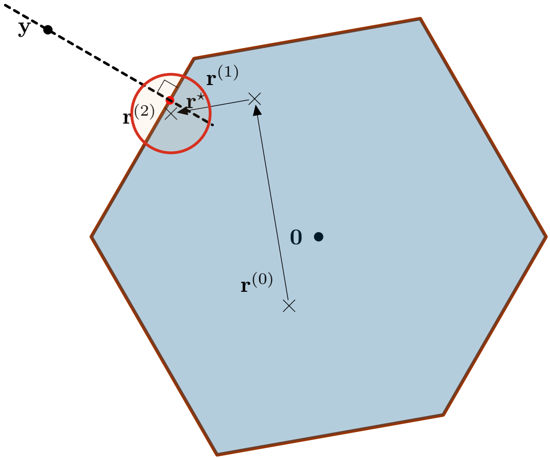
Visualizing the GAP sphere



Visualizing the GAP sphere



Visualizing the GAP sphere

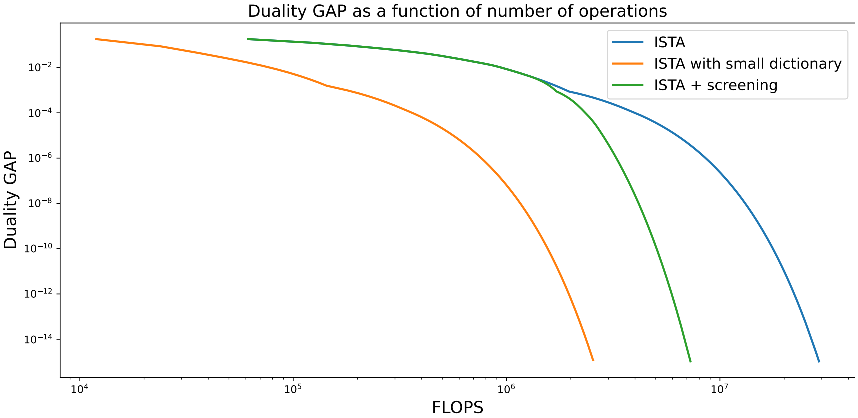


Numerical illustration

$(m, n) = (100, 150)$, "Gaussian dictionary", 20 repetitions

Numerical illustration

$(m, n) = (100, 150)$, "Gaussian dictionary", 20 repetitions



Selected lines of research

- Other geometries for **safe regions** *(e.g. dome region)*
  trade-off performance / complexity *Zhen et al, 2017*

Selected lines of research

- Other geometries for **safe regions** *(e.g. dome region)*
  trade-off performance / complexity *Zhen et al, 2017*

- **Beyond** the Lasso: *Ndiaye, 2018, Fraga-Dantas et al., 2021*
  sparse promoting convex optimization

$$\arg\min_{\mathbf{x}\in\mathbb{R}^n} \; f(\mathbf{Mx}) + \Omega(\mathbf{x})$$

Selected lines of research

- Other geometries for **safe regions** *(e.g. dome region)*
  trade-off performance / complexity *Zhen et al, 2017*

- **Beyond** the Lasso: *Ndiaye, 2018, Fraga-Dantas et al., 2021*
  sparse promoting convex optimization

$$\underset{\mathbf{x}\in\mathbb{R}^n}{\arg\min}\ f(\mathbf{Mx}) + \Omega(\mathbf{x})$$

- Accelerating “Greedy algorithm” *(e.g. conditional gradient method)*
  Reduce searching set *Sun and bach, 2020*

Selected lines of research

- Other geometries for **safe regions** *(e.g. dome region)*
  trade-off performance / complexity *Zhen et al, 2017*

- **Beyond** the Lasso: *Ndiaye, 2018, Fraga-Dantas et al., 2021*
  sparse promoting convex optimization

$$\operatorname*{arg\,min}_{\mathbf{x}\in\mathbb{R}^n} \; f(\mathbf{Mx}) + \Omega(\mathbf{x})$$

- Accelerating "Greedy algorithm" *(e.g. conditional gradient method)*
  Reduce searching set *Sun and bach, 2020*

- Machine learning: screening **data point** *Zhang et al., 2017*
  e.g., in SVM, not all points are relevant for evaluating the separating hyperplane

Beyond safe screening

Unconventional 1: Nonnegative least squares

Relevant in many signal processing applications

$$\mathbf{x}^\star \in \operatorname*{arg\,min}_{\mathbf{x}\in\mathbb{R}^n} \ \frac{1}{2}\|\mathbf{y} - \mathbf{M}\mathbf{x}\|_2^2 \quad \text{s.t.} \quad \mathbf{x} \in \mathbb{R}_+^n$$

- Known to promote **sparsity** (*Night sky theorem* *Byrne (2009)*)
- **Explicit** regularization *(not a norm)*

Unconventional 1: Nonnegative least squares

Relevant in many signal processing applications

$$\mathbf{x}^\star \in \operatorname*{arg\,min}_{\mathbf{x}\in\mathbb{R}^n} \ \frac{1}{2}\|\mathbf{y} - \mathbf{M}\mathbf{x}\|_2^2 \quad \text{s.t.} \quad \mathbf{x} \in \mathbb{R}_+^n$$

- Known to promote **sparsity** (*Night sky theorem* *Byrne (2009)*)
- **Explicit** regularization *(not a norm)*
- **Main difficulty & contribution:** design of feasible dual points

The mapping $\Phi\colon \mathbb{R}^m \longrightarrow \mathcal{R}$ in the previous slide

`In preparation`

Unconventional 2: Slope

Recent surge of interest for the **SLOPE**[1] problem

$$\mathbf{x}^\star \in \operatorname*{arg\,min}_{\mathbf{x}\in\mathbb{R}^n} \ \tfrac{1}{2}\|\mathbf{y}-\mathbf{M}\mathbf{x}\|_2^2 + \lambda \sum_{j=1}^{n} \gamma_j |\mathbf{x}|_{[j]}$$

Bogdan et al. (2015), Figueiredo and Nowak (2016), Su and Candès (2019),...

- $\gamma_1 \geq \cdots \geq \gamma_n$
- $|\mathbf{x}|_{[j]}$: j **largest** entry of $\mathbf{x}$ *(in absolute value)*
- Includes the Lasso as a special case
- Desirable **statistical** properties, may promote **sparsity** / **clustering** of the cœfficients *(see references above)*

[1] Also known as OSCAR / OWL regression

Unconventional 2: Slope (still)

- Dœs not fit existing frameworks due to the **sorting** operation
- **Sorting** makes screening even more desirable $\mathcal{O}(n \log n)$

Unconventional 2: Slope (still)

- Dœs not fit existing frameworks due to the **sorting** operation
- **Sorting** makes screening even more desirable $\mathcal{O}(n \log n)$

Theorem *[Elvira et Herzet, 2021]*

$\forall q \in \{1, \dots, n\}$ Then

$$\forall q \in \{1, \dots, n\} : \left|\mathbf{m}_\ell^T \mathbf{c}\right| + \sum_{k=p_q}^{q-1} \left|\mathbf{M}_{\backslash \ell}^T \mathbf{c}\right|_{[k]} < B_{q,p_q} \implies \mathbf{x}^\star(\ell) = 0$$

- Requires verifying n inequalities for a **single** entry

Unconventional 2: Slope (still)

- Dœs not fit existing frameworks due to the **sorting** operation
- **Sorting** makes screening even more desirable $\mathcal{O}(n \log n)$

Theorem *[Elvira et Herzet, 2021]*

$\forall q \in \{1, \dots, n\}$ let $p_q \in \{1, \dots, q\}$. Then

$$\forall q \in \{1, \dots, n\} : \ \left|\mathbf{m}_\ell^{\mathrm{T}} \mathbf{c}\right| + \sum_{k=p_q}^{q-1} \left|\mathbf{M}_{\backslash \ell}^{\mathrm{T}} \mathbf{c}\right|_{[k]} < B_{q,p_q} \implies \mathbf{x}^\star(\ell) = 0$$

- Requires verifying n inequalities for a **single** entry
- Defines a family of $n!$ different tests 🎉 😱

Unconventional 2: Slope (still)

- Dœs not fit existing frameworks due to the **sorting** operation
- **Sorting** makes screening even more desirable $\mathcal{O}(n \log n)$

Theorem *[Elvira et Herzet, 2021]*

$\forall q \in \{1, \dots, n\}$ let $p_q \in \{1, \dots, q\}$. Then

$$\forall q \in \{1, \dots, n\} : \ \left|\mathbf{m}_\ell^{\mathrm{T}}\mathbf{c}\right| + \sum_{k=p_q}^{q-1} \left|\mathbf{M}_{\backslash \ell}^{\mathrm{T}}\mathbf{c}\right|_{[k]} < B_{q,p_q} \implies \mathbf{x}^\star(\ell) = 0$$

- Requires verifying n inequalities for a **single** entry
- Defines a family of $n!$ different tests 🎉 😱
- **Contribution:** one method in $\mathcal{O}(n \log(n))$ to perform **them all**

`https://arxiv.org/abs/2110.11784`

Unconventional 3: ℓ_0-problem 😲

Relevant in some applications

$$\mathbf{x}^\star \in \arg\min_{\mathbf{x}\in\mathbb{R}^n} \ \tfrac{1}{2}\|\mathbf{y} - \mathbf{M}\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_0$$

- Known to promote sparsity
- **Not** convex ⚠️

Unconventional 3: ℓ_0-problem 😲

Relevant in some applications

$$\mathbf{x}^\star \in \arg\min_{\mathbf{x}\in\mathbb{R}^n} \tfrac{1}{2}\|\mathbf{y} - \mathbf{M}\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_0$$

- Known to promote sparsity
- **Not** convex ⚠️
- **Contribution:** New framework for safe screening

$\longrightarrow$ Focus on a specific solver: Branch and bounds

$\longrightarrow$ Safe screening for B&B: take decision on sub-nodes at **no computational cost**

$\longrightarrow$ Outperforms state-of-the-art commercial solvers

Curious? see preprint and code at
`https://arxiv.org/abs/2110.07308`

Screening beyond sparsity 1: Safe relaxing

- Using the knowledge of the position of 0 👍
- What about the knowledge of **nonzero** entries 🤔

Screening beyond sparsity 1: Safe relaxing

- Using the knowledge of the position of 0 👍
- What about the knowledge of **nonzero** entries 🤔

Say that one can **ensure** that $\mathbf{x}^\star(\ell) > 0$. Then

Solving the Lasso

$$\Longleftrightarrow$$

$$\arg\min_{\mathbf{x}\in\mathbb{R}^n} \frac{1}{2}\left\|\mathbf{y} - \mathbf{M}_{\backslash\ell}\mathbf{x}_{\backslash\ell} - \mathbf{m}_\ell x_\ell\right\|_2^2 + \lambda\|\mathbf{x}_{\backslash\ell}\|_1 + \lambda\mathbf{x}_\ell$$

Screening beyond sparsity 1: Safe relaxing

- Using the knowledge of the position of 0 👍
- What about the knowledge of **nonzero** entries 🤔

Say that one can **ensure** that $\mathbf{x}^\star(\ell) > 0$. Then

Solving the Lasso

$$\Longleftrightarrow$$

$$\operatorname*{arg\,min}_{\mathbf{x}_{\backslash\ell}\in\mathbb{R}^{n-1}} \operatorname*{arg\,min}_{\mathbf{x}_\ell\in\mathbb{R}} \frac{1}{2}\left\|\mathbf{y} - \mathbf{M}_{\backslash\ell}\mathbf{x}_{\backslash\ell} - \mathbf{m}_\ell x_\ell\right\|_2^2 + \lambda\|\mathbf{x}_{\backslash\ell}\|_1 + \lambda\mathbf{x}_\ell$$

Screening beyond sparsity 1: Safe relaxing

- Using the knowledge of the position of 0 👍
- What about the knowledge of **nonzero** entries 🤔

Say that one can **ensure** that $\mathbf{x}^\star(\ell) > 0$. Then

$$\text{Solving the Lasso}$$

$$\Longleftrightarrow$$

$$\begin{cases} \mathbf{x}^\star_{\setminus\ell} \in \arg\min_{\mathbf{x}_{\setminus\ell}\in\mathbb{R}^{n-1}} \frac{1}{2}\left\|\widetilde{\mathbf{y}} - \widetilde{\mathbf{M}}\mathbf{x}_{\setminus\ell}\right\|_2^2 + \widetilde{\lambda}\|\mathbf{x}_{\setminus\ell}\|_1 \\ x^\star_\ell = \mathbf{m}_\ell^\mathsf{T}(\mathbf{y} - \mathbf{M}_{\setminus\ell}\mathbf{x}^\star_{\setminus\ell}) - \lambda \end{cases}$$

Screening beyond sparsity 1: Safe relaxing

- Using the knowledge of the position of 0 👍
- What about the knowledge of **nonzero** entries 🤔

Say that one can **ensure** that $\mathbf{x}^\star(\ell) > 0$. Then

Solving the Lasso

$$\Longleftrightarrow$$

$$\begin{cases} \mathbf{x}^\star_{\backslash\ell} \in \arg\min_{\mathbf{x}_{\backslash\ell} \in \mathbb{R}^{n-1}} \frac{1}{2} \left\| \widetilde{\mathbf{y}} - \widetilde{\mathbf{M}} \mathbf{x}_{\backslash\ell} \right\|_2^2 + \widetilde{\lambda} \|\mathbf{x}_{\backslash\ell}\|_1 \\ x^\star_\ell = \mathbf{m}_\ell^T (\mathbf{y} - \mathbf{M}_{\backslash\ell} \mathbf{x}^\star_{\backslash\ell}) - \lambda \end{cases}$$

🎉 We end up with a **lower dimensional** Lasso problem 🎉

Tests, details and derivation available in
`https://arxiv.org/abs/2110.07281`

Screening beyond sparsity 2: Safe squeezing

So called "**anti-sparse**" problem

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

- Known to promote solutions with **saturated entries** (*i.e.*, equal to $\pm\alpha$)
- Under mild assumptions, **most** entries are **saturated**

Elvira and Herzet (2020)

Screening beyond sparsity 2: Safe squeezing

So called "**anti-sparse**" problem

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

- Known to promote solutions with **saturated entries** (*i.e.*, equal to $\pm\alpha$)
- Under mild assumptions, **most** entries are **saturated**

  Elvira and Herzet (2020)
- **Contribution:** A **test** to detect saturated entries

  Same nature as screening test
- **Main difficulty:** Resulting lower dimensional optimization problem is of different nature

Interested? paper and code available at
`https://doi.org/10.1109/tsp.2020.2995192`

Conclusion

Foundation: safe screening for “standard” sparse problem

- **Rationale:** It leads to an equivalent low dimensional problem
- **Ideal test:** leverage convex optimization to detect zero entries
- **Impact:** $\searrow$ computational complexity $\nearrow$ convergence properties
- **In practice:** plug and play tests with **low computational cost**

Conclusion

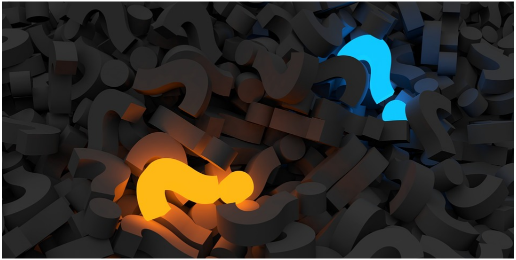
Foundation: safe screening for “standard” sparse problem

- **Rationale:** It leads to an equivalent low dimensional problem
- **Ideal test:** leverage convex optimization to detect zero entries
- **Impact:** $\searrow$ computational complexity $\nearrow$ convergence properties
- **In practice:** plug and play tests with **low computational cost**

Opening Pandora’s box

- Extension to **unconventional** problems
  $\longrightarrow$ new families of **convex regularizers**
  $\longrightarrow$ **non-convex** problems
- **Not only** zero entries can be detected!

Merci de votre attention!



stay tuned!

`https://c-elvira.github.io`