

Une nouvelle méthode d'accélération pour LASSO par élimination sûre de variables

Thu-Le Tran, Clément Elvira, Hong-Phuong Dang, Cédric Herzet

▶ **To cite this version:**

Thu-Le Tran, Clément Elvira, Hong-Phuong Dang, Cédric Herzet. Une nouvelle méthode d'accélération pour LASSO par élimination sûre de variables. CAP 2022 - Conférence sur l'Apprentissage automatique, Jul 2022, Vannes, France. pp.1-6. hal-03806044

HAL Id: hal-03806044

https://centralesupelec.hal.science/hal-03806044

Submitted on 7 Oct 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Une nouvelle méthode d'accélération pour LASSO par élimination sûre de variables

Thu-Le Tran[1], Clément Elvira[2], Hong-Phuong Dang[3], et Cédric Herzet[4]

[1]Univ Rennes, IRMAR - UMR 6625, F-35000 Rennes, France
[2]IETR UMR CNRS 6164, CentraleSupelec Rennes Campus, 35576 Cesson Sévigné, France
[3]LaTIM, INSERM-UMR1101, Univ. de Bretagne Occidentale, Brest, France
[4]Inria Rennes - Bretagne Atlantique, Rennes, France

Résumé

Nous présentons une nouvelle région de sûreté (*safe region*) pour la mise en oeuvre de techniques d'"élimination sûre de variables" (*safe screening*) pour le problème LASSO. La région proposée a une géométrie de dôme et exploite une représentation canonique de l'ensemble des demi-espaces contenant l'ensemble réalisable du problème dual. Nous montrons que la région proposée est toujours incluse dans la "sphère GAP" et le "dôme GAP" introduits par Fercoq *et al.* (et strictement sous certaines conditions non restrictives), tout en induisant la même complexité calculatoire. Les expériences numériques démontrent que notre nouvelle région de sûreté permet de concevoir des règles d'élimination de variables plus efficaces que celles basées sur les régions GAP et conduit à une accélération significative de la résolution de LASSO.

Mots-clef : LASSO, réduction de dimensionalité, élimination de variables, safe screening.

1 Introduction

Le terme "représentations parcimonieuses" désigne une méthodologie populaire en apprentissage automatique et en traitement du signal consistant à décomposer un vecteur $\mathbf{y} \in \mathbb{R}^m$ en une combinaison linéaire de quelques colonnes (ou *atomes*) d'une matrice (ou *dictionnaire*) $\mathbf{A} = [\mathbf{a}_1, \dots, \mathbf{a}_n] \in \mathbb{R}^{m \times n}$. Une approche classique pour calculer une représentation parcimonieuse d'un vecteur $\mathbf{y} \in \mathbb{R}^m$ consiste à résoudre le problème LASSO :

$$\mathbf{x}^\star \in \underset{\mathbf{x}\in\mathbb{R}^n}{\arg\min}\ P(\mathbf{x}) \triangleq \tfrac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_1 \qquad (1)$$

où $\lambda > 0$ [1]. Il existe de nombreux algorithmes pour approcher numériquement ce problème, voir par exemple [2–4]. Néanmoins, lorsque les dimensions du problème augmentent, la résolution de (1) peut induire des complexités calculatoires importantes et mettre à l'épreuve l'efficacité des techniques d'optimisation classiques. Ce constat a motivé le développement de nouvelles méthodes d'accélération pour LASSO.

Une approche de référence dans ce domaine est la méthode proposée par El Ghaoui *et al.* dans [5] et connue sous le nom d'"élimination sûre de variables" (*safe screening* en anglais). Cette méthode consiste à concevoir des tests simples permettant d'identifier les entrées nulles (c'est-à-dire, les variables inactives) des minimiseurs de (1). Cette connaissance peut ensuite être exploitée afin de réduire (de manière potentiellement significative) la dimension du problème d'optimisation en éliminant les colonnes du dictionnaire pondérées par zéro. Ce type de méthodes a suscité un vif intérêt de la communauté scientifique ces dernières années, voir par exemple [6–17].

Les méthodologies standard d'élimination sûre de variables exploitent le concept de "région de sûreté" (*safe region* en anglais), à savoir un sous-ensemble de l'espace dual contenant la solution du problème dual de LASSO [18, Section 4]. Le choix d'une région de sûreté pertinente est crucial pour l'obtention de bonnes performances d'élimination de variables. En effet, il existe d'une part un lien entre le volume de la région de sûreté et l'efficacité du test résultant [18, Lemma 1]. D'autre part, le coût calculatoire du test d'élimination est généralement lié à la géométrie de la région de sûreté. Par conséquent, les régions de géométrie "simple", telles que les sphères [5–9] ou les dômes [12, 13], sont généralement privilégiées dans la littérature.

Une méthodologie de référence pour trouver un bon

compromis entre ces deux exigences a été proposée dans [8] : les auteurs ont introduit deux nouvelles régions de sûreté (appelées "sphère GAP" et "dôme GAP") dont les rayons sont proportionnels au saut de dualité atteint par le couple primal-dual utilisé pour construire la région. Une propriété importante de ces régions est donc la suivante : leur rayon tend vers 0 lorsque le couple primal-dual considéré dans leur construction tend vers une solution primale-duale du problème LASSO.

Dans ce papier, nous présentons une nouvelle région de sûreté possédant également cette propriété. Nous démontrons que la région proposée est *toujours* incluse dans les deux régions GAP mentionnées ci-dessus tout en induisant la même charge calculatoire.

Le reste de l'article est organisé de la façon suivante. Les notations utilisées tout au long du papier sont définies dans la prochaine partie. La partie 3 introduit le problème LASSO et les méthodes de suppression sûre de variables. Dans la partie 4, nous détaillons la construction de notre nouvelle région de sûreté. Les simulations numériques décrites dans la partie 5 illustrent la pertinence de l'approche proposée et démontrent des gains significatifs par rapport aux tests basés sur les régions GAP.

2 Notations

Les lettres grasses minuscules (par exemple $\mathbf{x}$) et majuscules (par exemple $\mathbf{A}$) sont respectivement utilisées pour désigner les vecteurs et les matrices. Le symbole $\mathbf{0}_n$ fait référence au vecteur nul de $\mathbb{R}^n$. $\langle \cdot, \cdot \rangle$ désigne le produit scalaire usuel entre deux vecteurs. La $i^{\text{ème}}$ entrée du vecteur $\mathbf{x}$ est notée $\mathbf{x}(i)$. Les lettres majuscules rondes (par exemple $\mathcal{S}$) sont utilisées pour représenter des ensembles.

3 Autour du problème Lasso

Dans cette partie, nous présentons quelques éléments d'analyse convexe particularisés au problème (1) et rappelons les principaux ingrédients des méthodes d'élimination sûre de variables pour LASSO.

3.1 Dual et conditions d'optimalité

Notons d'abord que le problème (1) admet au moins un minimiseur puisque la fonction $P(\cdot)$ est continue, propre et coercive [19, Theorem 2.14]. Le problème dual de (1) s'écrit

$$\mathbf{u}^\star = \arg\max_{\mathbf{u} \in \mathcal{U}} \; D(\mathbf{u}) \triangleq \tfrac{1}{2}\|\mathbf{y}\|_2^2 - \tfrac{1}{2}\|\mathbf{y} - \mathbf{u}\|_2^2 \quad (2)$$

où $\mathcal{U} \triangleq \{\mathbf{u} \in \mathbb{R}^m : \|\mathbf{A}^{\mathrm{T}}\mathbf{u}\|_\infty \leq \lambda\}$ est l'ensemble réalisable du dual [20, Appendix A]. Puisque $\mathcal{U}$ est fermé et que $D(\cdot)$ est strictement concave, le problème (2) admet un unique maximiseur $\mathbf{u}^\star$.

Comme (1) est convexe et la condition de Slater est vérifiée, il y a dualité forte entre les problèmes (1) et (2), c'est-à-dire

$$\forall \mathbf{x} \in \mathbb{R}^n, \forall \mathbf{u} \in \mathcal{U} : \; \mathrm{gap}(\mathbf{x}, \mathbf{u}) \triangleq P(\mathbf{x}) - D(\mathbf{u}) \geq 0 \quad (3)$$

avec égalité si et seulement si $(\mathbf{x}, \mathbf{u})$ est une solution primale-duale, voir par exemple [19, Theorem A.2]. De plus, tout couple optimal $(\mathbf{x}^\star, \mathbf{u}^\star)$ doit vérifier les conditions suivantes [20, Section 2] :

$$\mathbf{u}^\star = \mathbf{y} - \mathbf{A}\mathbf{x}^\star \quad (4)$$

et

$$\langle \mathbf{a}_i, \mathbf{u}^\star \rangle = \begin{cases} \lambda \; \mathrm{sign}(\mathbf{x}^\star(i)) & \text{si } \mathbf{x}^\star(i) \neq 0, \\ s \in [-\lambda, \lambda] & \text{sinon} \end{cases} \quad (5)$$

pour tout $i = 1, \ldots, n$. On déduit facilement de ces conditions que $\mathbf{x}^\star = \mathbf{0}_n$ est l'unique solution de (1) si et seulement si

$$\lambda \geq \lambda_{\max} \triangleq \|\mathbf{A}^{\mathrm{T}}\mathbf{y}\|_\infty \quad (6)$$

où $\|\cdot\|_\infty$ désigne la norme ℓ_∞ d'un vecteur.

3.2 Élimination sûre de variables

Les tests d'élimination sûre de variables découlent directement de la condition d'optimalité donnée en (5) et s'écrivent :

$$|\langle \mathbf{a}_i, \mathbf{u}^\star \rangle| < \lambda \implies \mathbf{x}^\star(i) = 0. \quad (7)$$

Si l'inégalité du côté gauche de (7) est vérifiée pour un indice i, la colonne correspondante de $\mathbf{A}$ peut donc être éliminée sans impacter la valeur du minimum de (1).

Bien que le calcul de $\mathbf{u}^\star$ soit (généralement) aussi complexe que la résolution de (1), une version relâchée de (7) peut être obtenue lorsqu'une région $\mathcal{R} \subseteq \mathbb{R}^m$ contenant $\mathbf{u}^\star$ (appelée "région de sûreté" ou "*safe region*" en anglais) est connue. Plus précisément, si $\mathbf{u}^\star \in \mathcal{R}$, on déduit de (7) l'implication suivante :

$$\max_{\mathbf{u} \in \mathcal{R}} |\langle \mathbf{a}_i, \mathbf{u} \rangle| < \lambda \implies \mathbf{x}^\star(i) = 0. \quad (8)$$

Du point de vue de l'efficacité de détection des zéros de $\mathbf{x}^\star$, la région $\mathcal{R}$ se doit d'être choisie aussi petite que possible. En particulier, si $\mathcal{R} \subseteq \mathcal{R}'$ alors

$$\max_{\mathbf{u}\in\mathcal{R}'} |\langle \mathbf{a}_i, \mathbf{u}\rangle| < \lambda \implies \max_{\mathbf{u}\in\mathcal{R}} |\langle \mathbf{a}_i, \mathbf{u}\rangle| < \lambda. \quad (9)$$

En d'autres termes, le test d'élimination de variables se basant sur la région $\mathcal{R}$ détectera toujours au moins autant de zéros que celui exploitant la région $\mathcal{R}'$.

Du point de vue de la complexité, le coût du test (8) dépend principalement de l'évaluation de la quantité $\max_{\mathbf{u}\in\mathcal{R}} |\langle \mathbf{a}_i, \mathbf{u}\rangle|$. Pour réduire le coût de cette opération, une stratégie simple consiste à construire des régions de sûreté ayant des géométries "bien choisies". Les deux régions les plus plébiscitées dans la littérature sont les sphères et les dômes.

Les régions de type "sphère" sont définies par un centre $\mathbf{c} \in \mathbb{R}^m$ et un rayon $R \geq 0$:

$$\mathcal{R} = \mathcal{B}(\mathbf{c}, R) \triangleq \{\mathbf{u} \in \mathbb{R}^m : \|\mathbf{u} - \mathbf{c}\|_2 \leq R\}. \quad (10)$$

Ainsi, lorsque $\mathcal{R}$ est une sphère, le côté gauche de (8) devient :

$$\max_{\mathbf{u}\in\mathcal{B}(\mathbf{c},R)} |\langle \mathbf{a}_i, \mathbf{u}\rangle| = |\langle \mathbf{a}_i, \mathbf{c}\rangle| + R\|\mathbf{a}_i\|_2. \quad (11)$$

En d'autres mots, la solution du problème de maximisation peut être calculée en forme fermée. Son évaluation ne nécessite alors que le calcul d'un produit scalaire entre $\mathbf{a}_i$ et $\mathbf{c}$.

Un autre choix populaire de géométrie est la région en forme de "dôme". Un dôme est défini comme l'intersection d'une sphère et d'un demi-espace :

$$\mathcal{R} = \mathcal{D}(\mathbf{c}, R, \mathbf{g}, \delta) \triangleq \mathcal{B}(\mathbf{c}, R) \cap \mathcal{H}(\mathbf{g}, \delta) \quad (12)$$

où $\mathbf{g} \in \mathbb{R}^m$, $\delta \in \mathbb{R}$ et [1]

$$\mathcal{H}(\mathbf{g}, \delta) \triangleq \{\mathbf{u} \in \mathbb{R}^m \colon \langle \mathbf{g}, \mathbf{u}\rangle \leq \delta\}. \quad (13)$$

Dans ce cas, le côté gauche de (8) admet également une solution en forme fermée, voir [20, Lemma. 3]. Plus précisément, on a

$$\max_{\mathbf{u}\in\mathcal{D}} |\langle \mathbf{a}_i, \mathbf{u}\rangle| = \max\left(\max_{\mathbf{u}\in\mathcal{D}} \langle \mathbf{a}_i, \mathbf{u}\rangle, \max_{\mathbf{u}\in\mathcal{D}} \langle -\mathbf{a}_i, \mathbf{u}\rangle\right)$$

et pour tout vecteur $\mathbf{a} \in \mathbb{R}^m \setminus \{\mathbf{0}_m\}$ et dôme $\mathcal{D} \neq \emptyset$, on vérifie que

$$\max_{\mathbf{u}\in\mathcal{D}} \langle \mathbf{a}, \mathbf{u}\rangle = \langle \mathbf{a}, \mathbf{c}\rangle + R\|\mathbf{a}\|_2 f(\psi_1, \psi_2), \quad (14)$$

où

$$f(\psi_1, \psi_2) = \begin{cases} 1 & \text{si } \psi_1 \leq \psi_2, \\ \psi_1\psi_2 + \sqrt{1-\psi_1^2}\sqrt{1-\psi_2^2} & \text{sinon} \end{cases}$$

$$\psi_1 = \frac{\langle \mathbf{a}, \mathbf{g}\rangle}{\|\mathbf{a}\|_2\|\mathbf{g}\|_2}, \quad \psi_2 = \min\left(\frac{\delta - \langle \mathbf{g}, \mathbf{c}\rangle}{R\|\mathbf{g}\|_2}, 1\right).$$

Le coût associé à l'implémentation de (14) est principalement dominé par l'évaluation des quantités $\langle \mathbf{g}, \mathbf{c}\rangle$, $\langle \mathbf{a}, \mathbf{c}\rangle$ et $\langle \mathbf{a}, \mathbf{g}\rangle$. Les tests basés sur un dôme sont donc plus coûteux à mettre en œuvre que les tests basés sur des sphères. Toutefois, comme la géométrie des dômes permet (généralement) d'obtenir des régions de sûreté plus petites, les performances de détection sont aussi potentiellement bien meilleures.

3.3 Régions GAP

Pour finir la partie 3, nous présentons les expressions des deux régions de sûreté proposées dans [8] et aujourd'hui considérées comme "état de l'art". La première région prend la forme d'une sphère et est connue sous le nom de "sphère GAP" dans la littérature. Elle est définie par le choix suivant de paramètres :

$$\mathbf{c} = \mathbf{u} \quad (15)$$

$$R = \sqrt{2\text{gap}(\mathbf{x}, \mathbf{u})} \quad (16)$$

où $(\mathbf{x}, \mathbf{u})$ est un couple primal-dual réalisable et $\text{gap}(\mathbf{x}, \mathbf{u})$ désigne le saut de dualité rappelé à l'équation (3). La seconde région a la géométrie d'un dôme et est connue sous le nom de "dôme GAP". Elle est définie par l'ensemble de paramètres suivants :

$$\mathbf{c} = \tfrac{1}{2}(\mathbf{y} + \mathbf{u}) \quad (17)$$

$$R = \tfrac{1}{2}\|\mathbf{y} - \mathbf{u}\|_2 \quad (18)$$

$$\mathbf{g} = \mathbf{y} - \mathbf{c} \quad (19)$$

$$\delta = \langle \mathbf{g}, \mathbf{c}\rangle + \text{gap}(\mathbf{x}, \mathbf{u}) - R^2 \quad (20)$$

où $(\mathbf{x}, \mathbf{u})$ est un couple primal-dual réalisable. Dans la suite, afin de souligner explicitement leur dépendance vis-à-vis du choix du couple primal-dual réalisable choisi, nous dénoterons (avec un léger abus de notation) la sphère GAP et le dôme GAP respectivement comme suit : $\mathcal{B}_{\texttt{gap}}(\mathbf{x}, \mathbf{u})$ et $\mathcal{D}_{\texttt{gap}}(\mathbf{x}, \mathbf{u})$.

Bien que le résultat ne soit pas rendu explicite dans [8], on démontre facilement l'inclusion [21, Appendix B]

$$\mathcal{D}_{\texttt{gap}}(\mathbf{x}, \mathbf{u}) \subseteq \mathcal{B}_{\texttt{gap}}(\mathbf{x}, \mathbf{u}). \quad (21)$$

Les deux régions GAP voient également leur rayon [2] tendre vers 0 lors que le couple $(\mathbf{x}, \mathbf{u})$ tend vers une

1. Lorsque $\mathbf{g} = \mathbf{0}_m$, notre définition implique que $\mathcal{H}(\mathbf{g}, \delta)$ est soit égal à $\mathbb{R}^m$ si $\delta \geq 0$, soit vide si $\delta < 0$.

2. Voir (31) pour une définition du rayon dans le cas général.

solution primale-duale de LASSO. Notons que dans le cas particulier où $(\mathbf{x}, \mathbf{u}) = (\mathbf{x}^\star, \mathbf{u}^\star)$, les deux régions GAP sont réduites au singleton $\{\mathbf{u}^\star\}$ et les tests d'élimination à (7).

4 Le dôme de Hölder

Dans cette partie, nous présentons notre nouvelle région de sûreté baptisée "dôme de Hölder". La construction du dôme exploite une caractérisation de la forme canonique des éléments de l'ensemble suivant :

$$\mathcal{G} \triangleq \{(\mathbf{g}, \delta) \in \mathbb{R}^m \times \mathbb{R} \colon \langle \mathbf{g}, \mathbf{u} \rangle \leq \delta \ \forall \mathbf{u} \in \mathcal{U}\}, \qquad (22)$$

c'est-à-dire l'ensemble (des paramètres) des demi-espaces contenant l'ensemble réalisable $\mathcal{U}$. Dans la suite, nous appellerons "*demi-espace à coupe duale*" tout demi-espace généré par un élément de $\mathcal{G}$. Le lemme suivant fournit une représentation canonique des éléments de $\mathcal{G}$:

Lemme 1.

$$\mathcal{G} = \{(\mathbf{A}\mathbf{x}, \delta) \colon \mathbf{x} \in \mathbb{R}^n, \delta \geq \lambda \|\mathbf{x}\|_1\}. \qquad (23)$$

Une preuve de ce résultat est disponible dans [21, Appendix A-A].

Nous décrivons maintenant le premier résultat de cet article qui est une nouvelle région de sûreté de type dôme, nommée "dôme de Hölder". La définition de ce dôme est formulée dans le théorème suivant :

Théorème 1. *Soient* $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{u} \in \mathcal{U}$ *et*

$$\mathbf{c} = \tfrac{1}{2}(\mathbf{y} + \mathbf{u}) \qquad (24)$$

$$R = \tfrac{1}{2}\|\mathbf{y} - \mathbf{u}\|_2 \qquad (25)$$

$$\mathbf{g} = \mathbf{A}\mathbf{x} \qquad (26)$$

$$\delta = \lambda \|\mathbf{x}\|_1. \qquad (27)$$

Alors

$$\mathbf{u}^\star \in \mathcal{D}(\mathbf{c}, R, \mathbf{g}, \delta). \qquad (28)$$

Le lecteur trouvera la preuve de ce résultat dans [21, Appendix A-B]. Remarquons que la sûreté de la région $\mathcal{H}(\mathbf{A}\mathbf{x}, \lambda\|\mathbf{x}\|_1)$ peut être vue comme une simple conséquence de l'inégalité de Hölder :

$$\langle \mathbf{A}\mathbf{x}, \mathbf{u}^\star \rangle = \langle \mathbf{x}, \mathbf{A}^{\mathrm{T}}\mathbf{u}^\star \rangle \leq \|\mathbf{x}\|_1 \|\mathbf{A}^{\mathrm{T}}\mathbf{u}^\star\|_\infty \leq \lambda \|\mathbf{x}\|_1,$$

où la dernière inégalité provient du fait que $\mathbf{u}^\star$ est réalisable pour le problème (2). Cette observation sous-tend l'appellation "dôme de Hölder" pour la région de sûreté proposée. Le lecteur remarquera également que, à l'image des régions GAP, le dôme de Hölder est complètement spécifié par le choix d'un couple primal-dual réalisable $(\mathbf{x}, \mathbf{u})$. Dans la suite, nous utiliserons donc la notation $\mathcal{D}_{\texttt{Hölder}}(\mathbf{x}, \mathbf{u})$ pour le désigner.

Notons que la définition de $\mathcal{D}_{\texttt{Hölder}}(\mathbf{x}, \mathbf{u})$ ne diffère de celle de $\mathcal{D}_{\texttt{gap}}(\mathbf{x}, \mathbf{u})$ que par le choix du demi-espace $\mathcal{H}(\mathbf{g}, \delta)$, où la caractérisation canonique de $\mathcal{G}$ donnée dans le Lemme 1 est directement exploitée. Notre prochain résultat montre que ce choix est bénéfique au regard de la taille de la région de sûreté :

Théorème 2. *Pour tout* $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{u} \in \mathcal{U}$ *:*

$$\mathcal{D}_{\texttt{Hölder}}(\mathbf{x}, \mathbf{u}) \subseteq \mathcal{D}_{\texttt{gap}}(\mathbf{x}, \mathbf{u}). \qquad (29)$$

Par ailleurs, si $P(\mathbf{x}) < P(\mathbf{0}_n)$ *et* $(\mathbf{x}, \mathbf{u})$ *n'est pas une solution primale-duale, alors l'inclusion est stricte.*

La preuve de ce résultat est détaillée dans [21, Appendix A-C]. Le Théorème 2 montre que le dôme de Hölder est *nécessairement* contenu dans le dôme GAP. Par conséquent, (9) implique que les tests d'élimination sûre basés sur le dôme proposé seront toujours au moins aussi performants que ceux basés sur les régions GAP. Nous notons également que la condition "$P(\mathbf{x}) < P(\mathbf{0}_n)$", assurant que $\mathcal{D}_{\texttt{Hölder}}(\mathbf{x}, \mathbf{u}) \subset \mathcal{D}_{\texttt{gap}}(\mathbf{x}, \mathbf{u})$, n'est pas contraignante et est généralement vérifiée en pratique. C'est par exemple le cas lorsque les itérés $\{\mathbf{x}^{(t)}\}_{t=1}^{\infty}$ d'une procédure d'optimisation initialisée à $\mathbf{x}^{(0)} = \mathbf{0}_n$ et diminuant la fonction de coût de manière monotone sont utilisés pour définir le point primal $\mathbf{x}$ intervenant dans la construction du dôme.

5 Expériences numériques

Cette partie présente une étude empirique démontrant la pertinence du dôme de Hölder introduit dans la partie 4.[3] Pour les deux expériences, notre configuration de simulation est la suivante. Les dimensions du problème sont fixées à $(m, n) = (100, 500)$. Des nouvelles réalisations de $\mathbf{A}$ et $\mathbf{y}$ sont générées pour chaque répétition de l'expérience. L'observation $\mathbf{y}$ est simulée selon une loi uniforme sur la sphère unité de $\mathbb{R}^m$ et le dictionnaire $\mathbf{A}$ satisfait soit : *i)* les entrées sont des réalisations i.i.d. d'une loi Normale centrée réduite ; *ii)* $\mathbf{A}$ a une structure Tœplitz, c'est-à-dire les colonnes sont des versions translatées d'une courbe gaussienne. Dans les deux cas, les colonnes de $\mathbf{A}$ sont ensuite normalisées de sorte que $\|\mathbf{a}_i\|_2 = 1$ pour tout i. Nous appellerons respectivement ces deux configurations "Gaussien" et "Tœplitz" dans la suite.

3. Le code est disponible à l'adresse suivante : `https://gitlab.inria.fr/cherzet/holder-safe`.

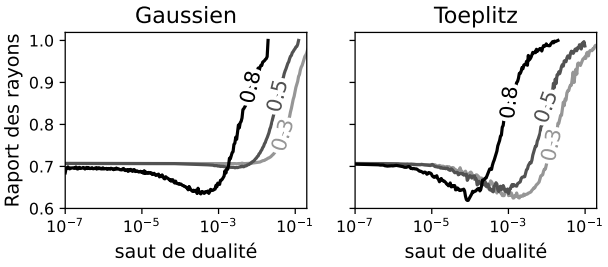


FIGURE 1 – Valeur du rapport (30) en fonction du saut de dualité obtenu par $(\mathbf{x}, \mathbf{u})$ pour les deux dictionnaires. Chaque courbe correspond à une valeur du ratio $\lambda/\lambda_{\max}$ (la valeur est indiquée le long de la courbe).

5.1 Comparaison des rayons

Dans une première expérience, nous étudions la différence de taille entre les dômes de Hölder et GAP. Plus précisément, nous évaluons le rapport

$$\frac{\texttt{Rad}(\mathcal{D}_{\texttt{Hölder}}(\mathbf{x}, \mathbf{u}))}{\texttt{Rad}(\mathcal{D}_{\texttt{gap}}(\mathbf{x}, \mathbf{u}))} \tag{30}$$

pour différents choix de $\mathbf{x} \in \mathbb{R}^n$ et $\mathbf{u} \in \mathcal{U}$, où $\texttt{Rad}(\mathcal{S})$ désigne le rayon de l'ensemble $\mathcal{S}$:

$$\texttt{Rad}(\mathcal{S}) \triangleq \sup_{\mathbf{u}, \mathbf{u}' \in \mathcal{S}} \tfrac{1}{2} \|\mathbf{u}' - \mathbf{u}\|_2. \tag{31}$$

La Figure 1 indique la valeur moyenne du rapport (30) en fonction du saut de dualité obtenu par $(\mathbf{x}, \mathbf{u})$ pour les deux dictionnaires considérés et trois valeurs du rapport $\lambda/\lambda_{\max}$. Les résultats ont été obtenus en faisant la moyenne sur 50 réalisations. Comme attendu (cf. Théorème 2), le rapport des rayons est toujours inférieur à 1. On observe également que le rayon du dôme de Hölder est jusqu'à 0.6 plus petit que celui du dôme GAP. Curieusement, toutes les courbes semblent converger pour nos scénarios de simulation vers une valeur du rapport des rayons proche de 0.7 lorsque le saut de dualité tend vers zéro. La confirmation théorique que ce comportement reste vrai dans des cas généraux est une question ouverte que nous traiterons dans de prochains travaux.

5.2 Profils de performance

Dans cette seconde expérience, nous comparons les gains de précision obtenus dans la résolution de LASSO par différentes méthodes d'élimination sûre pour un budget calculatoire donné. À cette fin, nous comparons trois variantes de FISTA –une méthode standard pour résoudre (1), voir [3]– où chaque itération de l'algorithme est précédée d'un test d'élimination sûre de

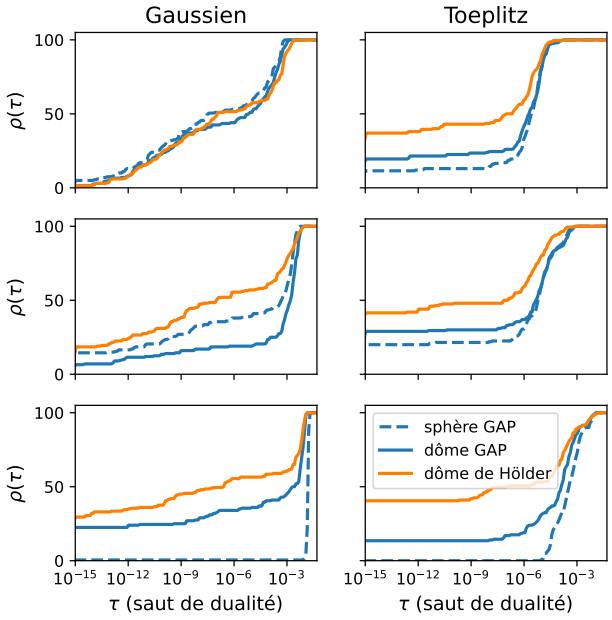


FIGURE 2 – Profils de performance des tests d'élimination sûre de variables avec différentes régions de sûreté. Chaque ligne correspond à une valeur différente du rapport $\lambda/\lambda_{\max}$. De haut en bas : $0.3, 0.5, 0.8$.

variables basé sur *i)* la sphère GAP, *ii)* le dôme GAP, *iii)* le dôme de Hölder. Plus précisément, à chaque itération t, un test est effectué avec une région de sûreté calculée à partir du couple $(\mathbf{x}^{(t)}, \mathbf{u}^{(t)})$ où $\mathbf{x}^{(t)}$ est l'itéré courant de FISTA et $\mathbf{u}^{(t)}$ est obtenu par une opération de "*dual scaling*" de $\mathbf{y} - \mathbf{A}\mathbf{x}^{(t)}$ (voir [5, Section 3.3]). Nous utilisons les profils de performance de Dolan-Moré [22] pour comparer les trois méthodes.

Nous exécutons chaque méthode avec un budget calculatoire prescrit (le nombre d'opérations en virgule flottante) sur 200 instances du problème (1). Nous évaluons ensuite la probabilité (empirique) $\rho(\tau)$ qu'un algorithme atteigne un saut de dualité inférieur à τ avec le budget disponible. Pour chaque configuration, le budget est ajusté de sorte que $\rho(10^{-7}) = 50\%$ pour la méthode utilisant le dôme de Hölder.

La Figure 2 montre les profils de performance pour les deux dictionnaires considérés et différentes valeurs du rapport $\lambda/\lambda_{\max}$. On peut constater que, pour ces scénarios de simulation, les tests basés sur le dôme de Hölder améliorent de manière significative la précision moyenne de résolution de LASSO dans toutes les configurations considérées sauf une. Nous soulignons que le cas où les trois régions de sûreté conduisent à des résultats comparables (dictionnaire Gaussien et $\lambda/\lambda_{\max} = 0.3$) doit être interprété de la façon sui-

vante : même si les dômes Hölder et GAP permettent de meilleures performances de détection que la sphère GAP, les profils résultent d'un compromis entre l'efficacité du test et sa complexité calculatoire. L'analyse de nos résultats de simulation révèle que pour cette configuration particulière, davantage d'itérations sont effectuées (en moyenne) avec la sphère GAP pour le même budget, malgré des performances de détection moindre. La précision de résolution de LASSO s'en trouve par conséquent potentiellement améliorée.

6 Conclusion

Dans cet article, nous avons introduit une nouvelle région de sûreté (*safe region* en anglais) de type dôme qui peut être utilisée pour construire des tests d'élimination sûre de variables et ainsi accélérer la résolution du problème LASSO. Nous avons montré que le dôme proposé est toujours inclus (potentiellement strictement) dans la sphère et le dôme GAP, deux régions de sûreté considérées comme état de l'art. Les simulations numériques réalisées ont démontré que la méthodologie proposée permet des gains de précision significatifs pour un budget calculatoire donné.

Références

[1] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*. Springer New York, 2013.

[2] M. A. T. Figueiredo, R. D. Nowak, and S. J. Wright, "Gradient projection for sparse reconstruction : Application to compressed sensing and other inverse problems," *IEEE J. Sel. Top. Signal Process.*, vol. 1, no. 4, pp. 586–597, 2007.

[3] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imag. Sc.*, vol. 2, no. 1, pp. 183–202, 2009.

[4] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2011.

[5] L. El Ghaoui, V. Viallon, and T. Rabbani, "Safe feature elimination for the Lasso and sparse supervised learning problems," *Pacific J. Optim.*, vol. 8, no. 4, pp. 667–698, 2012.

[6] Z. Xiang, H. Xu, and P. J. Ramadge, "Learning sparse representations of high dimensional data on large scale dictionaries," *Proc. of Adv. in Neural Inform. Process. Systems*, vol. 24, 2011.

[7] A. Bonnefoy, V. Emiya, L. Ralaivola, and R. Gribonval, "A dynamic screening principle for the Lasso," in *Proc. Europ. Signal Process. Conf. (EUSIPCO)*, 2014, pp. 6–10.

[8] O. Fercoq, A. Gramfort, and J. Salmon, "Mind the duality gap : safer rules for the Lasso," in *Proc. Int. Conf. Machine Learning (ICML)*, 2015, pp. 333–342.

[9] C. Herzet and A. Malti, "Safe screening tests for Lasso based on firmly non-expansiveness," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, 2016, pp. 4732–4736.

[10] C. Herzet, C. Dorffer, and A. Drémeau, "Gather and conquer : Region-based strategies to accelerate safe screening tests," *IEEE Trans. on Signal Process.*, vol. 67, no. 12, pp. 3300–3315, 2019.

[11] L. Dai and K. Pelckmans, "An ellipsoid based, two-stage screening test for BPDN," in *Proc. Europ. Signal Process. Conf. (EUSIPCO)*, 2012, pp. 654–658.

[12] Z. J. Xiang and P. J. Ramadge, "Fast Lasso screening tests based on correlations," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*. IEEE, 2012, pp. 2137–2140.

[13] J. Liu, Z. Zhao, J. Wang, and J. Ye, "Safe screening with variational inequalities and its application to Lasso," in *Proc. Int. Conf. Machine Learning (ICML)*, 2014, pp. 289–297.

[14] E. Ndiaye, O. Fercoq, A. Gramfort, and J. Salmon, "Gap safe screening rules for sparsity enforcing penalties," *J. Mach. Learn. Res.*, vol. 18, no. 1, pp. 4671–4703, 2017.

[15] C. Elvira and C. Herzet, "Safe squeezing for antisparse coding," *IEEE Trans. on Signal Process.*, vol. 68, pp. 3252–3265, 2020.

[16] ——, "Safe rules for the identification of zeros in the solutions of the SLOPE problem," arXiv :2110.11784, 2021.

[17] T. Guyard, C. Herzet, and C. Elvira, "Screen & relax : accelerating the resolution of Elastic-Net by safe identification of the solution support," arXiv :2110.07281, 2021.

[18] Z. J. Xiang, Y. Wang, and P. J. Ramadge, "Screening tests for Lasso problems," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 5, pp. 1008–1027, 2017.

[19] A. Beck, *First-Order Methods in Optimization*. Philadelphia, PA : Society for Industrial and Applied Mathematics, 2017.

[20] Z. J. Xiang, Y. Wang, and P. J. Ramadge, "Screening tests for Lasso problems," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 5, pp. 1008–1027, 2016.

[21] T.-L. Tran, C. Elvira, H.-P. Dang, and C. Herzet, "Beyond GAP screening for lasso by exploiting new dual cutting half-spaces with supplementary material," arXiv :2203.00987, 2022.

[22] E. D. Dolan and J. J. Moré, "Benchmarking optimization software with performance profiles," *Math. Program.*, vol. 91, no. 2, pp. 201–213, 2002.