Quelles relaxations continues pour le critère ℓ_2-ℓ_0 ?

Emmanuel SOUBIES[1,2], Laure BLANC-FÉRAUD[2], Gilles AUBERT[3]

[1]Biomedical Imaging Group, EPFL, Suisse

[2]Université Côte d'Azur, CNRS, INRIA, Laboratoire I3S, 06903 Sophia Antipolis, France

[3]Université Côte d'Azur, CNRS, Laboratoire J.A. Dieudonné UMR 7351, 06100 Nice, France

emmanuel.soubies@epfl.ch,{blancf,gaubert}@unice.fr

Résumé – Depuis plus de deux décennies, de nombreuses pénalités continues (et généralement séparables) ont été proposées afin d'approcher (relaxer) la pseudo-norme ℓ_0. Bien que certaines «bonnes» propriétés pour ces pénalités aient été mises en évidence, le choix d'une relaxation par rapport à une autre reste peu clair. Une approche pour les comparer est de s'intéresser à la fidélité de la relaxation au problème initial. En d'autres termes, est-ce que cette dernière préserve les minimiseurs globaux du critère initial sans en ajouter de locaux ? Dans le contexte des moindres carrés pénalisés en norme-ℓ_0, nous avons récemment étudié cette question et défini une classe de pénalités qualifiés d'exactes. Dans cette communication, nous présentons ces résultats et les complétons par une étude des minimiseurs locaux qui sont éliminés par de telles relaxations. En particulier, nous montrons que la pénalité CEL0 est celle éliminant le plus de minimiseurs locaux.

Abstract – For more than two decades, several continuous (and generally separable) penalties approximating (relaxing) the ℓ_0-pseudo norm have been proposed. Although some "good" properties for such penalties have been highlighted, the choice of one relaxation rather than another one remains unclear. One approach to compare them is to investigate their fidelity to the initial problem. In other words, do they preserve global minimizers of the initial criteria without adding new local ones? Within the context of the ℓ_0 penalized least squares, we have recently studied this question resulting in a class of penalties said exact. In this communication, we present these results and complete them with a study concerning the local minimizers eliminated by such relaxations. In particular, we show that the CEL0 penalty is the one removing the largest number of local minimizers.

1 Introduction

Dans cet article, nous nous intéressons aux relaxations *continues* du critère moindres carrés pénalisé en norme-ℓ_0 :

$$\mathrm{G}_{\ell_0}(\mathrm{x}) := \frac{1}{2}\|A\mathrm{x} - d\|_2^2 + \lambda\|\mathrm{x}\|_0, \tag{1}$$

où $A \in \mathbb{R}^{M\times N}$, $M \ll N$, $d \in \mathbb{R}^M$, $\lambda > 0$ et $\|\cdot\|_0$ est la pseudo-norme ℓ_0 dénombrant le nombre de composantes non-nulles du signal. Il n'est maintenant plus à rappeler l'importance de ce critère pour le traitement du signal et des images, ne serait-ce qu'à la vue de la vaste littérature sur le sujet. En particulier, de nombreux travaux se sont intéressés à des relaxations non-convexes *continues* de G_{ℓ_0} de la forme :

$$\tilde{\mathrm{G}}_\lambda(\mathrm{x}) := \frac{1}{2}\|A\mathrm{x} - d\|_2^2 + \Phi(\mathrm{x}), \tag{2}$$

où $\Phi = \sum_i \phi_i$ est une pénalité continue et séparable approchant $\lambda\|\cdot\|_0$. On peut citer, par exemple, les pénalités *log-sum* [1], ℓ_p*-norm* ($p \in]0,1[$) [4], *SCAD* [3] ou encore *MCP* [8]. Il existe cependant peu de travaux s'intéressant aux relations entre les minimiseurs des fonctionnelles G_{ℓ_0} et $\tilde{\mathrm{G}}_\lambda$ et notamment répondant aux deux questions suivantes :

1. $\tilde{\mathrm{G}}_\lambda$ préserve-t-elle les minimiseurs globaux de G_{ℓ_0} ?
2. $\tilde{\mathrm{G}}_\lambda$ ajoute-t-elle de nouveaux minimiseurs locaux ?

Afin d'y apporter une réponse, nous avons déterminé dans [7] des conditions *nécessaires et suffisantes* sur les ϕ_i définissant Φ pour que, quelles que soient les données $d \in \mathbb{R}^M$, les deux propriétés suivantes soient vérifiées :

$$\arg\min_{\mathrm{x}\in\mathbb{R}^N} \tilde{\mathrm{G}}_\lambda(\mathrm{x}) = \arg\min_{\mathrm{x}\in\mathbb{R}^N} \mathrm{G}_{\ell_0}(\mathrm{x}) \tag{P1}$$

$$\hat{\mathrm{x}} \text{ minimiseur local de } \tilde{\mathrm{G}}_\lambda \Rightarrow \hat{\mathrm{x}} \text{ minimiseur local de } \mathrm{G}_{\ell_0} \tag{P2}$$

Une classe de relaxations continues, dites *exactes*, est ainsi définie. Par ailleurs, notons que d'après (P2) $\tilde{\mathrm{G}}_\lambda$ peut potentiellement éliminer des minimiseurs locaux (non-globaux) de G_{ℓ_0}, ce qui est une propriété intéressante dans ce contexte d'optimisation non-convexe.

Dans ce travail, après avoir rappelé le résultat principal de [7], nous nous intéressons à l'étude des minimiseurs de G_{ℓ_0} éliminés par $\tilde{\mathrm{G}}_\lambda$. En particulier, des illustrations numériques ainsi qu'une justification théorique des comportements observés sont présentées.

Notations : $\|a_i\|_2 \neq 0$ correspond à la norme de la i-ème colonne de $A \in \mathbb{R}^{N\times M}$; $\mathbb{I}_N = \{1,\dots,N\}$; $\sigma(\mathrm{x}) \subseteq \mathbb{I}_N$ définit le support de $\mathrm{x} \in \mathbb{R}^N$; A_ω dénote la restriction de $A \in \mathbb{R}^{N\times M}$ aux colonnes indexées par les éléments de $\omega \subseteq \mathbb{I}_N$ et la notation A_ω^T est utilisée pour $(A_\omega)^T$ (et non $(A^T)_\omega$) ; $\sharp\mathcal{E}$ définit la

cardinalité de l'ensemble $\mathcal{E}$; $\mathbb{1}_{\{X\}}$ représente la fonction indicatrice sur X ; e_i dénote le i-ème vecteur de la base canonique de l'ensemble considéré.

2 Une classe de pénalités exactes

En considérant des pénalités continues $\Phi = \sum_i \phi_i$ où :
— $\forall i \in \mathbb{I}_N$, ϕ_i est localement Lipschitz et C^2 sur $\mathbb{R}$ sauf en un nombre fini de points (on notera cet ensemble de points $B^i \subset \mathbb{R}$),
— $\forall u \in B^i$, les limites $\lim_{v\to u^-} \phi_i(v)$ et $\lim_{v\to u^+} \phi_i(v)$ existent, sont finies, et sont différentes,
nous avons le résultat suivant.

Théorème 1 ([7]). *Soient $A \in \mathbb{R}^{M\times N}$ ($\|a_i\|_2 \neq 0\ \forall i \in \mathbb{I}_N$) et $\lambda > 0$. Les propriétés* (P1) *et* (P2) *sont vérifiées $\forall d \in \mathbb{R}^M$ si et seulement si Φ vérifie les conditions suivantes : $\forall i \in \mathbb{I}_N$,*

$$\phi_i(0) = 0, \tag{3a}$$
$$\forall u \in \mathbb{R}\backslash]\beta^{i-}, \beta^{i+}[,\ \phi_i(u) = \lambda, \tag{3b}$$
$$\forall u \in]\beta^{i-}, \beta^{i+}[\backslash\{0\},\ \phi_i(u) > \phi_{\text{CEL0}}(\|a_i\|_2, \lambda; u), \tag{3c}$$
$$\forall u \in B^i \setminus \{0\},\ \lim_{\substack{v\to u\\ v<u}} \phi_i'(v) > \lim_{\substack{v\to u\\ v>u}} \phi_i'(v), \tag{3d}$$
$$\forall u \in]\beta^{i-}, \beta^{i+}[\backslash B^i, \begin{cases} \phi_i''(u) \leq -\|a_i\|_2^2 \text{ et} \\ \forall \varepsilon > 0,\ \exists v_\varepsilon \in (u-\varepsilon, u+\varepsilon) \\ \quad \text{t.q. } \phi_i''(v_\varepsilon) < -\|a_i\|_2^2 \end{cases} \tag{3e}$$

où $\beta^{i-} \in [-\sqrt{2\lambda}/\|a_i\|_2, 0[$, $\beta^{i+} \in]0, \sqrt{2\lambda}/\|a_i\|_2]$, B^i est un sous-ensemble de $[\beta^{i-}, \beta^{i+}]$ contenant un nombre fini de points pour lesquels ϕ_i n'est pas différentiable ($0 \in B^i$) et ϕ_{CEL0} dénote la pénalité CEL0 [1] *[6].*

Le théorème 1 établit des conditions *nécessaires et suffisantes* sur Φ pour vérifier les deux propriétés (P1) et (P2). Plus précisément, seules les trois premières conditions sont nécessaires et suffisantes pour avoir (P1) alors que les cinq sont nécessaires et suffisantes pour avoir à la fois (P1) et (P2) [7]. Ainsi, on peut voir que les caractéristiques requises pour les ϕ_i sont d'être nulles et non-différentiables à l'origine, constantes égales à λ pour $|u|$ grand et strictement plus grandes que la pénalité CEL0 [6] ailleurs. D'autre part, les deux dernières conditions imposent aux ϕ_i d'être suffisamment concave là où elles ne sont pas constantes. La figure 1 illustre ces conditions.

Enfin, notons que CEL0 n'est autre que la limite inférieure de la classe de pénalités exactes résultant du théorème 1.

3 Étude des minimiseurs stricts de $\tilde{\mathrm{G}}_\lambda$

Tout au long de cette section, nous considérons $A \in \mathbb{R}^{M\times N}$ ($\|a_i\|_2 \neq 0\ \forall i \in \mathbb{I}_N$), $\lambda > 0$ et $\Phi = \sum_i \phi_i$ vérifiant (3). Sous ces conditions, nous nous intéressons aux minimiseurs locaux *stricts* de $\tilde{\mathrm{G}}_\lambda$ dont nous rappelons la définition ci-après.

1. voir table 1 pour une définition.

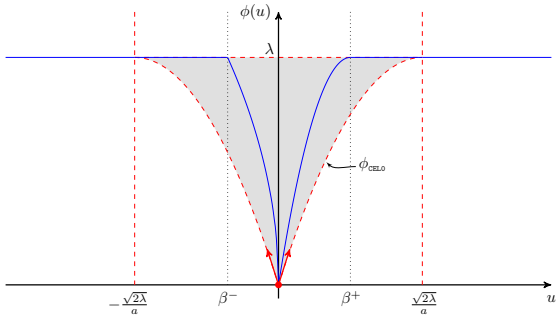

FIGURE 1 – Illustrations (1D) des conditions du théorème 1. Les courbes rouges ainsi que le point en 0 représentent les conditions (3). Un exemple de pénalité vérifiant ces conditions est tracé en bleu.

Définition 1. *$\hat{\mathrm{x}}$ est un minimiseur local strict de f s'il existe un voisinage $\mathcal{V}$ de $\hat{\mathrm{x}}$ tel que $\forall \mathrm{y} \in \mathcal{V} \setminus \{\hat{\mathrm{x}}\},\ f(\hat{\mathrm{x}}) < f(\mathrm{y})$.*

Nous introduisons ensuite $\mathcal{S}_{\tilde{\mathrm{G}}_\lambda}$ (resp. $\mathcal{S}_{\mathrm{G}_{\ell_0}}$), l'ensemble des minimiseurs locaux *stricts* de $\tilde{\mathrm{G}}_\lambda$ (resp. G_{ℓ_0}). Rappelons que $\mathcal{S}_{\mathrm{G}_{\ell_0}}$ a été étudié dans [5] et qu'il a aussi été montré dans ce même article que les minimiseurs globaux de G_{ℓ_0} étaient stricts. Il est donc naturel de s'intéresser aux propriétés de $\mathcal{S}_{\tilde{\mathrm{G}}_\lambda}$. La proposition suivante établit un lien entre les ensembles $\mathcal{S}_{\tilde{\mathrm{G}}_\lambda}$ et $\mathcal{S}_{\mathrm{G}_{\ell_0}}$.

Proposition 1. $\mathcal{S}_{\tilde{\mathrm{G}}_\lambda} \subseteq \mathcal{S}_{\mathrm{G}_{\ell_0}}$

Démonstration. Étant donné que sous (3), (P2) est vérifiée, la preuve est une simple adaptation de [6, Corollaire 4.9]. □

Corollaire 1. *Les minimiseurs globaux de $\tilde{\mathrm{G}}_\lambda$ sont stricts.*

Démonstration. Conséquence du fait que les minimiseurs globaux de G_{ℓ_0} sont stricts [5] et que (P1) est vérifiée. □

Le corollaire précédent motive l'étude de $\mathcal{S}_{\tilde{\mathrm{G}}_\lambda}$ et en particulier l'étude de l'inclusion réciproque de la proposition 1. Nous allons voir qu'elle n'est généralement pas vérifiée et que certains minimiseurs locaux stricts de G_{ℓ_0} sont éliminés par $\tilde{\mathrm{G}}_\lambda$.

3.1 Illustrations Numériques

Dans cette section, nous quantifions numériquement les minimiseurs locaux stricts de $\tilde{\mathrm{G}}_\lambda$ pour les pénalités de la table 1. Toutes vérifient les conditions (3) dès lors que les paramètres les définissants sont choisis comme indiqué dans cette même table [7]. Notons que les pénalités n'étant pas constantes à l'infini (*e.g.* log-sum, ℓ_p-norm) ne peuvent pas vérifier (3) et que capped-ℓ_1 ou SCAD peuvent uniquement vérifier (3a)–(3c) et donc seulement (P1) [7]. Elles ne sont donc pas considérées dans cette étude. Dans la suite, les paramètres des pénalités sont choisis de sorte à obtenir celles étant les plus proches de la limite inférieure CEL0 tout en vérifiant (3). De fait, les paramètres considérés *dépendent de* λ.

D'après la proposition 1, quantifier $\sharp\mathcal{S}_{\tilde{\mathrm{G}}_\lambda}$ peut être réalisé en comptant les points de $\mathcal{S}_{\mathrm{G}_{\ell_0}}$ qui sont des points critiques [2] de

2. Être un point critique de $\tilde{\mathrm{G}}_\lambda$ est équivalent à $0 \in \partial\tilde{\mathrm{G}}_\lambda(\hat{\mathrm{x}})$, où $\partial\tilde{\mathrm{G}}_\lambda(\hat{\mathrm{x}})$ définit le gradient généralisé au sens de Clarke [2].

Pénalité	Paramètres	Définition ϕ_i	Conditions pour (3)
CEL0	aucun	$\lambda - \frac{\|a_i\|_2^2}{2}\left(\lvert \mathrm{x}_i\rvert - \frac{\sqrt{2\lambda}}{\|a_i\|_2}\right)^2 \mathbb{1}_{\{\lvert \mathrm{x}_i\rvert \geq \frac{\sqrt{2\lambda}}{\|a_i\|_2}\}}$	limite inférieure de la classe définie par (3)
ℓ_p-Trq	$\theta_i,\ \varepsilon_i > 0$ $p_i \in]0,1[$	$\lambda \min\left\{(\theta_i\lvert \mathrm{x}_i\rvert + \varepsilon_i)^{p_i}, 1 + \varepsilon_i^{p_i}\right\} - \lambda\varepsilon_i^{p_i}$	$\theta_i \geq \|a_i\|_2 \max\left\{\frac{(1+\varepsilon_i^{p_i})^{\frac{2-p_i}{2p_i}}}{\sqrt{(p_i-p_i^2)\lambda}}, \frac{(1+\varepsilon_i^{p_i})^{\frac{1}{p_i}} - \varepsilon_i}{\sqrt{2\lambda}}\right\}$
Log-Trq	$\theta_i,\ \varepsilon_i > 0$	$\lambda \min\left\{\log(\theta_i\lvert \mathrm{x}_i\rvert + \varepsilon_i), 1 + \log(\varepsilon_i)\right\} - \lambda\log(\varepsilon_i)$	$\theta_i \geq \|a_i\|_2 \max\left\{\frac{\varepsilon_i e^1}{\sqrt{\lambda}}, \frac{\varepsilon_i(e^1-1)}{\sqrt{2\lambda}}\right\}$

TABLE 1 – Pénalités CEL0 [6], ℓ_p-tronquée [7] et Log-Tronquée utilisées pour les illustrations numériques. La dernière colonne donne des conditions permettant de vérifier (3). La pénalité MCP [8] est ici représentée par le choix de paramètres définissant CEL0.

$\tilde{\mathrm{G}}_\lambda$. Cependant, notons que si $\hat{\mathrm{x}} \in \mathcal{S}_{\mathrm{G}_{\ell_0}}$ est un point critique de $\tilde{\mathrm{G}}_\lambda$, cela n'implique pas pour autant qu'il appartienne à $\mathcal{S}_{\tilde{\mathrm{G}}_\lambda}$. Ainsi nous aurons une borne supérieure de $\sharp\mathcal{S}_{\tilde{\mathrm{G}}_\lambda}$ qui est généralement atteinte. Déterminer tous les éléments de $\mathcal{S}_{\mathrm{G}_{\ell_0}}$ peut être aisément réalisé en résolvant [5]

$$A_\omega^T A_\omega \mathrm{x}_\omega = A_\omega^T d \quad \text{et} \quad \mathrm{x}_{\mathbb{I}_N \setminus \omega} = 0, \tag{4}$$

pour tous les supports ω appartenant à

$$\Omega_{max} = \bigcup_{r=0}^{M} \Omega_r \ \text{ où } \ \Omega_r = \{\omega : \sharp\omega = r = \mathrm{rang}(A_\omega)\} \tag{5}$$

qui est un ensemble fini. Cependant, la cardinalité de Ω_{max} explose rapidement avec M et N et nous nous restreindrons donc à des problèmes de très petite taille ($M = 5$ et $N = 10$).

Remarque 1. *Notons que l'on a $\sharp\mathcal{S}_{\mathrm{G}_{\ell_0}} \leq \sharp\Omega_{max}$ — i.e. deux supports $(\omega, \omega') \in \Omega_{max}^2$ peuvent conduirent au même minimiseur (local) $\hat{\mathrm{x}}$ — et que l'égalité ou la stricte inégalité dépend généralement de d. Cependant, lorsque $\mathrm{rang}(A) = M$ l'égalité est vérifiée pour tout $d \in \mathbb{R}^M \setminus \mathcal{D}$ avec*

$$\mathcal{D} := \bigcup_{\omega \in \Omega_{max}} \bigcup_{i=1}^{\sharp\omega} \{d \in \mathbb{R}^M : \left\langle e_i, (A_\omega^T A_\omega)^{-1} A_\omega^T d \right\rangle = 0\}. \tag{6}$$

D'autre part $\mathbb{R}^M \setminus \mathcal{D}$ contient un sous-ensemble ouvert dense dans $\mathbb{R}^M$ d'après [5, Lemme 3.7]. Ainsi, nous nous plaçons dans cette configuration pour la suite de ce travail afin d'assurer que tous les minimiseurs (locaux) obtenus par (4)-(5) *sont distincts.*

Remarque 2. *G_{ℓ_0} admet le même nombre de minimiseurs stricts quelque soit λ [5, Remarque 5].*

La figure 2 présente des diagrammes dénombrant le nombre de points de $\mathcal{S}_{\mathrm{G}_{\ell_0}}$ qui sont des points critiques de $\tilde{\mathrm{G}}_\lambda$ en fonction de λ. La couleur correspond à la proportion, parmi 10 000 générations de $A \in \mathbb{R}^{5\times 10}$ et $d \in \mathbb{R}^5$, où le nombre de points critiques indiqué en ordonnées pour le λ donné en abscisses a été obtenu. Les courbes en traits plein et tirets correspondent respectivement à la valeur moyenne et l'écart-type en fonction de λ. On observe clairement qu'un certain nombre des 638 minimiseurs stricts de G_{ℓ_0} sont éliminés par les relaxations continues exactes considérées. Aussi, CEL0 semble être celle qui en élimine le plus. Enfin, quelle que soit la pénalité considérée, on observe le même comportement pour des λ suffisamment grands ou petits. En particulier, pour λ grand, on observe que $\sharp\mathcal{S}_{\tilde{\mathrm{G}}_\lambda} = 1$ et $\sharp\mathcal{S}_{\tilde{\mathrm{G}}_\lambda} \leq 252 = \binom{10}{5}$ pour λ petit.

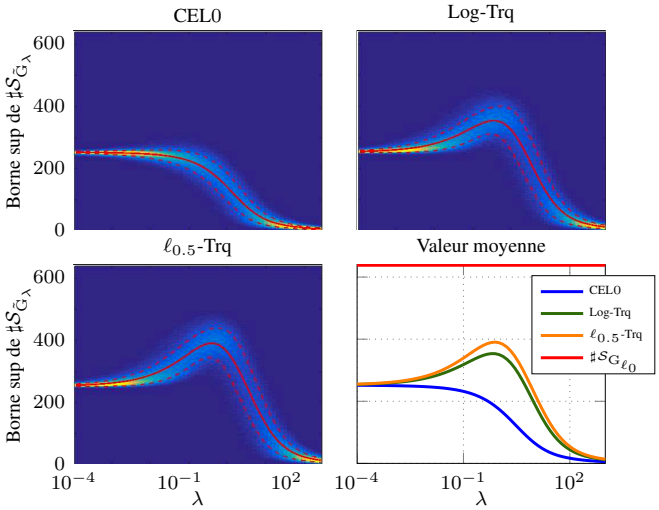

FIGURE 2 – Nombre d'éléments de $\mathcal{S}_{\mathrm{G}_{\ell_0}}$ qui sont des points critiques de $\tilde{\mathrm{G}}_\lambda$ en fonction de λ pour 10 000 générations de $A \in \mathbb{R}^{5\times 10}$ et $d \in \mathbb{R}^5$. Les diagrammes représentent la proportion d'occurrence (bleu sombre$=0$ et jaune≥ 500) parmi les 10 000 générations et les courbes la valeur moyenne (traits plein) et l'écart type (tirets).

3.2 Justification théorique

Nous nous intéressons maintenant à la justification théorique d'une partie des comportements observés sur les expériences numériques menées dans la section précédente. La proposition suivante donne une première justification de l'élimination de minimiseurs de G_{ℓ_0} par les relaxations exactes $\tilde{\mathrm{G}}_\lambda$.

Proposition 2 ([7]). *Soit $\hat{\mathrm{x}} \in \mathcal{S}_{\mathrm{G}_{\ell_0}}$. Alors,*

$$\exists i \in \sigma(\hat{\mathrm{x}}) \ \ t.q. \ \ \hat{\mathrm{x}}_i \in]\beta^{i-}, \beta^{i+}[\implies \hat{\mathrm{x}} \notin \mathcal{S}_{\tilde{\mathrm{G}}_\lambda}. \tag{7}$$

Afin de construire $\tilde{\mathrm{G}}_\lambda$ de manière à éliminer le plus grand nombre de minimiseurs locaux stricts de G_{ℓ_0}, il convient donc de choisir β^{i-} (resp. β^{i+}) le plus petit (resp. grand) possible tout en respectant les conditions (3). En observant (3), on comprend alors pourquoi CEL0, qui admet le plus grand intervalle $]\beta^{i-}, \beta^{i+}[\ \forall i$ est celle qui élimine le plus de minimiseurs locaux.

Afin d'étudier les comportements pour λ grand et proche de 0, nous considérons maintenant [3] β^{i-} et β^{i+} dépendant de λ.

3. Comme cela était le cas dans les expériences numériques précédentes.

On définit alors $\beta_\lambda^{i-} \in [-\sqrt{2\lambda}/\|a_i\|_2, 0[$ et $\beta_\lambda^{i+} \in]0, \sqrt{2\lambda}/\|a_i\|_2]$, respectivement décroissants et croissants avec λ et vérifiant

$$\lim_{\lambda\to+\infty} \beta_\lambda^{i-} = -\infty, \quad \lim_{\lambda\to+\infty} \beta_\lambda^{i+} = +\infty. \tag{8}$$

D'autre part, les demi-tangentes des ϕ_i en 0, que nous noterons $\delta^{i-} = \lim_{u\to 0^-} \phi_i'(u)$ et $\delta^{i+} = \lim_{u\to 0^+} \phi_i'(u)$, interviennent aussi dans « l'élimination » de minimiseurs locaux stricts par $\tilde{\mathrm{G}}_\lambda$. On peut alors voir que la condition (3c) impose d'avoir [7]

$$\delta^{i-} \leq -\sqrt{2\lambda}\|a_i\|_2 \quad \text{et} \quad \delta^{i+} \geq \sqrt{2\lambda}\|a_i\|_2. \tag{9}$$

Tout comme pour les β^{i-} et β^{i+}, nous verrons dans la suite que plus les δ^{i-} et δ^{i+} sont proches des bornes données ci-dessus, plus le nombre de minimiseurs locaux stricts de $\tilde{\mathrm{G}}_\lambda$ (*i.e.* $\sharp\mathcal{S}_{\tilde{\mathrm{G}}_\lambda}$) est faible. Nous définissons donc $\delta_\lambda^{i-} \in]-\infty, -\sqrt{2\lambda}\|a_i\|_2]$ et $\delta_\lambda^{i+} \in [\sqrt{2\lambda}\|a_i\|_2, +\infty[$, respectivement décroissants et croissants avec λ et vérifiant

$$\lim_{\lambda\to 0} \delta_\lambda^{i-} = 0, \quad \lim_{\lambda\to 0} \delta_\lambda^{i+} = 0. \tag{10}$$

Avec de telles définitions, pour $A \in \mathbb{R}^{M\times N}$ et $d \in \mathbb{R}^M$ fixés, le nombre de minimiseurs locaux stricts de $\tilde{\mathrm{G}}_\lambda$ (*i.e.* $\sharp\mathcal{S}_{\tilde{\mathrm{G}}_\lambda}$) est dépendant de λ, contrairement à G_{ℓ_0} (cf. remarque 2) ce qui va dans le sens des expériences de la section 3.1. Nous nous intéressons maintenant au comportement de $\sharp\mathcal{S}_{\tilde{\mathrm{G}}_\lambda}$ lorsque λ tend respectivement vers 0 et $+\infty$.

Proposition 3. *Il existe $\lambda_0 > 0$ et $\lambda_\infty > 0$ tels que :*

1. $\forall\lambda \in]\lambda_\infty, +\infty[$, $\mathcal{S}_{\tilde{\mathrm{G}}_\lambda} = \{0_{\mathbb{R}^N}\}$,
2. $\forall\lambda \in]0, \lambda_0[$, $\mathcal{S}_{\tilde{\mathrm{G}}_\lambda} \subseteq \mathcal{S}_{\mathrm{G}_{\ell_0}} \cap \mathcal{X}_{\mathrm{LS}}$,

où

$$\mathcal{X}_{\mathrm{LS}} = \arg\min_{\mathrm{x}\in\mathbb{R}^N} \frac{1}{2}\|A\mathrm{x} - d\|_2^2. \tag{11}$$

Démonstration.

(1) Soit

$$\beta = \max_{\hat{\mathrm{x}}\in\mathcal{S}_{\mathrm{G}_{\ell_0}}} \left\{ \min_{i\in\sigma(\hat{\mathrm{x}})} |\hat{\mathrm{x}}_i| \right\}. \tag{12}$$

D'après (8), il existe $\lambda_\infty > 0$ tel que $\forall\lambda > \lambda_\infty$ et $\forall i \in \mathbb{I}_N$, $\beta_\lambda^{i-} \in]-\infty, -\beta[$ et $\beta_\lambda^{i+} \in]\beta, +\infty[$.

Ainsi, pour tout $\lambda > \lambda_\infty$ et tout $\hat{\mathrm{x}} \in \mathcal{S}_{\mathrm{G}_{\ell_0}} \setminus \{0_{\mathbb{R}^N}\}$, il existe $i \in \sigma(\hat{\mathrm{x}})$ tel que $\hat{\mathrm{x}}_i \in]\beta_\lambda^{i-}, \beta_\lambda^{i+}[$ montrant que $\hat{\mathrm{x}} \notin \mathcal{S}_{\tilde{\mathrm{G}}_\lambda}$ d'après la proposition 2. On en déduit que $\mathcal{S}_{\tilde{\mathrm{G}}_\lambda} = \{0_{\mathbb{R}^N}\}$.

(2) Tout d'abord rappelons que $\forall\hat{\mathrm{x}} \in \mathcal{S}_{\mathrm{G}_{\ell_0}}$ on a (4) avec $\omega = \sigma(\hat{\mathrm{x}})$ [5]. Par ailleurs, de tels points appartiennent également à $\mathcal{X}_{LS}$ *si et seulement si* :

$$A^TA\hat{\mathrm{x}} = A^Td \underset{\text{sous (4)}}{\iff} \forall i \notin \sigma(\hat{\mathrm{x}}),\ \langle a_i, d - A\hat{\mathrm{x}}\rangle = 0. \tag{13}$$

Posons maintenant

$$\delta = \min_{\hat{\mathrm{x}}\in\mathcal{S}_{\mathrm{G}_{\ell_0}}\setminus\mathcal{X}_{\mathrm{LS}}} \left\{ \max_{i\in\mathbb{I}_N\setminus\sigma(\hat{\mathrm{x}})} |\langle a_i, d - A\hat{\mathrm{x}}\rangle| \right\} \underset{(13)}{\neq} 0. \tag{14}$$

Alors, d'après (10), il existe $\lambda_0 > 0$ tel que $\forall\lambda < \lambda_0$ et $\forall i \in \mathbb{I}_N$, $\delta_\lambda^{i-} \in]-\delta, 0[$ et $\delta_\lambda^{i+} \in]0, \delta[$.

Ensuite, $\forall\hat{\mathrm{x}} \in \mathcal{S}_{\tilde{\mathrm{G}}_\lambda}$, on a d'après la proposition 1 que $\hat{\mathrm{x}} \in \mathcal{S}_{\mathrm{G}_{\ell_0}}$ et la contraposée de (7) nous assure que $\forall i \in \sigma(\hat{\mathrm{x}})$, $\hat{\mathrm{x}}_i \notin]\beta_\lambda^{i-}, \beta_\lambda^{i+}[$. D'autre part, comme $\hat{\mathrm{x}} \in \mathcal{S}_{\tilde{\mathrm{G}}_\lambda}$, alors il est nécessairement point critique de $\tilde{\mathrm{G}}_\lambda$. D'après (3), cela est équivalent à :

$$\forall i \in \mathbb{I}_N,\ \begin{cases} \langle a_i, d - A\hat{\mathrm{x}}\rangle \in \left[\delta_\lambda^{i-}, \delta_\lambda^{i+}\right] & \text{si } i \notin \sigma(\hat{\mathrm{x}}), \\ \langle a_i, d - A\hat{\mathrm{x}}\rangle = 0 & \text{si } i \in \sigma(\hat{\mathrm{x}}). \end{cases} \tag{15}$$

Nous avons utilisé ici le fait que $\partial\tilde{\mathrm{G}}_\lambda(\mathrm{x}) = A^T(A\mathrm{x} - d) + \prod_i \partial\phi_i(\mathrm{x}_i)$, que $\partial\phi_i(0) = [\delta_\lambda^{i-}, \delta_\lambda^{i+}]$, que $\partial\phi_i(u) = \{0\}\ \forall u \notin]\beta_\lambda^{i-}, \beta_\lambda^{i+}[$ et que $\forall i \in \sigma(\hat{\mathrm{x}}), \hat{\mathrm{x}}_i \notin]\beta_\lambda^{i-}, \beta_\lambda^{i+}[$.

D'après (13) et (14), il vient alors que pour tout $\lambda < \lambda_0$ et tout $\hat{\mathrm{x}} \in \mathcal{S}_{\tilde{\mathrm{G}}_\lambda} \subseteq \mathcal{S}_{\mathrm{G}_{\ell_0}}$, nécessairement $\hat{\mathrm{x}} \in \mathcal{X}_{LS}$ afin de respecter la première ligne de (15) (la deuxième étant vérifiée pour tout $\lambda > 0$ dès lors que $\hat{\mathrm{x}} \in \mathcal{S}_{\mathrm{G}_{\ell_0}}$ d'après (4)). □

Ce résultat vient expliquer les comportements observés sur la figure 2. En effet, sur cette figure, il est clair que pour λ suffisamment grand, $\mathcal{S}_{\tilde{\mathrm{G}}_\lambda}$ est réduit à un unique point qui est donc d'après la proposition précédente $\hat{\mathrm{x}} = 0_{\mathbb{R}^N}$. D'autre part, pour λ suffisamment petit, nous avions observé que les courbes convergeaient vers une certaine valeur. Sous les conditions de la remarque 1, nous avons

$$\hat{\mathrm{x}} \in (\mathcal{S}_{\mathrm{G}_{\ell_0}} \cap \mathcal{X}_{LS}) \implies \|\hat{\mathrm{x}}\|_0 = M, \tag{16}$$

et ainsi, d'après le deuxième point de la proposition 3, $\sharp\mathcal{S}_{\tilde{\mathrm{G}}_\lambda} \leq \sharp(\mathcal{S}_{\mathrm{G}_{\ell_0}} \cap \mathcal{X}_{LS}) = \sharp\Omega_M = \binom{N}{M}$. Cela correspond à ce que l'on peut observer sur la figure 2.

4 Conclusion

Dans ce travail, nous nous somme intéressés aux minimiseurs locaux stricts pour la classe de relaxations exactes de G_{ℓ_0} dérivée dans [7]. Nous avons fourni des justifications, numériques et théoriques, concernant l'élimination de minimiseurs locaux stricts de G_{ℓ_0} par de telles relaxations venant ainsi compléter les résultats de [7]. Par ailleurs, il ressort de cette étude que la pénalité CEL0 est celle permettant d'éliminer le plus de minimiseurs de G_{ℓ_0}.

Références

[1] E.J. Candes, M.B. Wakin, and S.P. Boyd. Enhancing sparsity by reweighted ℓ_1 minimization. *Journal of Fourier analysis and applications*, 2008.

[2] F. H. Clarke. *Optimization and nonsmooth analysis*. SIAM, 1990.

[3] J. Fan and R. Li. Variable selection via nonconcave penalized likelihood and its oracle properties. *J. of the American Statistical Association*, 2001.

[4] S. Foucart and M.J. Lai. Sparsest solutions of underdetermined linear systems via ℓ_q-minimization for $0 < q \leq 1$. *Applied and Computational Harmonic Analysis*, 2009.

[5] M. Nikolova. Description of the minimizers of least squares regularized with ℓ_0-norm. Uniqueness of the global minimizer. *SIAM Journal on Imaging Sciences*, 2013.

[6] E. Soubies, L. Blanc-Féraud, and G. Aubert. A Continuous Exact l0 Penalty (CEL0) for Least Squares Regularized Problem. *SIAM Journal on Imaging Sciences*, 2015.

[7] E. Soubies, L. Blanc-Féraud, and G. Aubert. A unified view of exact continuous penalties for l2-l0 minimization. *(submitted)*, 2016.

[8] C.H. Zhang. Nearly unbiased variable selection under minimax concave penalty. *The Annals of Statistics*, 2010.