Vous avez dit parcimonie ?

Jean-Baptiste Hiriart-Urruty [1]

Résumé. La "parcimonie" (*sparsity* en anglais) est une notion qui intervient dans les Sciences des données ; c'est une propriété requise ou recherchée dans des problèmes de modélisation. Une fonction importante de parcimonie est celle qui dénombre le nombre de composantes non nulles dans un vecteur ; elle est analysée en détail ici. Trois contextes sont abordés successivement : celui de $\mathbb{R}^n$ (le plus fouillé), celui matriciel de $\mathcal{M}_{m,n}(\mathbb{R})$ (abordé), et celui fonctionnel des espaces de Lebesgue L^p (à peine évoqué).

Un exemple de motivation : comment Netflix recommande les films.

Typiquement on a une matrice $Y = [Y_{ij}]$ de taille $M \times N$ contenant les préférences des utilisateurs : les lignes représentent les films, les colonnes les utilisateurs, et Y_{ij} une note donnée par l'utilisateur j au film i. Comme tous les utilisateurs n'ont pas vu - et donc n'ont pas noté - tous les films, la matrice Y est très "creuse". Le but est d'inférer les notes manquantes dans Y à partir de celles connues, et ensuite de s'en servir pour recommander des films aux utilisateurs. Un modèle répandu pour réaliser cet objectif est de chercher une matrice X de rang faible qui minimise, par exemple, $\|Y - SX\|^2$, où S est un masque (*i.e.*, une matrice carrée d'ordre M) qui met à 0 les coefficients inconnus dans Y. La notion de parcimonie réside dans cette contrainte de "rang le plus petit possible".

1ère partie : Contexte de $\mathbb{R}^n$.

On comprend, même si on n'est pas versé dans le sujet, que pour un vecteur $x = (x_1, x_2, ..., x_n)$ de $\mathbb{R}^n$, plus il y a de composantes x_i nulles, plus il est considéré comme creux ou parcimonieux. Il est donc assez naturel d'étudier la fonction

$$x = (x_1, x_2, ..., x_n) \in \mathbb{R}^n \mapsto \|x\|_0 = \mathrm{card}\,\{i : x_i \neq 0\}\,. \qquad (1)$$

Cette fonction numérique (à valeurs entières $0, 1, ..., n$) de plusieurs variables mesure, d'une certaine façon, la parcimonie d'un vecteur x (ou "creusitude" comme pourrait dire un homme ou femme politique). Notons tout de

1. Université Paul Sabatier de Toulouse
118 Route de Narbonne
31062 Toulouse Cedex 09
Mél. jbhu@math.univ-toulouse.fr
https ://www.math.univ-toulouse.fr/~jbhu/

suite une faiblesse : une composante $x_i = 10^6$ ou $x_i = 10^{-2}$ contribue de la même façon (c'est-à-dire en ajoutant 1) à la valeur de $\|x\|_0$; il faudra donc travailler de préférence sur des ensembles bornés de vecteurs x. Observons aussi l'homogénéité (d'ordre 0) de la fonction $\|.\|_0$, c'est-à-dire : $\|rx\|_0 = \|x\|_0$ pour tout réel $r \neq 0$; cette propriété peut être perçue comme un défaut ou, au contraire, une opportunité dans l'utilisation de $\|.\|_0$ en Analyse ou Optimisation. C'est en tout cas une fonction bien intéressante que nous avons là. Nous allons en apprendre plus sur elle, sous forme de questions-réponses.

Q1. *Quelles propriétés, utiles pour l'analyse et l'optimisation de cette fonction* $\|.\|_0$ *pouvons-nous proposer ?*

En voici quelques-unes.

- Contrairement à ce que la notation pourrait laisser penser, la fonction $\|.\|_0$ *n'est pas une norme*, ni même une quasi-norme ou une semi-norme (qui correspondent à des définitions mathématiques précises) ; certes, elle est sous-additive :

$$\|x+y\|_0 \leqslant \|x\|_0 + \|y\|_0 \text{ pour tout } x, y\,;$$

elle vérifie $\|x\|_0 = 0$ si et seulement si $x = 0$; mais il lui manque la propriété d'homogénéité d'ordre 1 (par exemple, $\|2x\|_0 = \|x\|_0 \neq 2\,\|x\|_0$ si $x \neq 0$). L'appellation "pseudo-norme ℓ_0" pour $\|.\|_0$ est acceptée (et très utilisée), mais elle ne correspond à aucune définition mathématique (universelle).

Malgré tout, si on définit

$$d_0 : \mathbb{R}^n \times \mathbb{R}^n \mapsto d_0(x,y) = \|x-y\|_0\,, \tag{2}$$

comme on le ferait pour construire une distance à partir d'une norme, on obtient une distance : d_0 est la *distance de* HAMMING, très utilisée dans l'étude des codes (Informatique, Théorie du signal).

- Bien sûr, la fonction $\|.\|_0$ n'est pas continue, elle est même bien chahutée. Pourtant elle jouit d'une propriété de "continuité à moitié", fort utile dès qu'il s'agit de la minimiser sur un ensemble : La fonction $\|.\|_0$ est *semicontinue inférieurement* (en abrégé *s.c.i.*) sur $\mathbb{R}^n$. Qu'est-ce que cela veut dire ? Nous allons l'exprimer de trois façons, une analytique et deux géométriques.

Analytiquement, la semicontinuité inférieure de $\|.\|_0$ (en $x \in \mathbb{R}^n$) s'exprime par la relation :

$$\liminf_{\nu \to +\infty} \|x_\nu\|_0 \geqslant \|x\|_0 \text{ dès lors que } \lim_{\nu \to +\infty} x_\nu = x. \tag{3}$$

En clair, la valeur de l'entier $\|x_\nu\|_0$ ne peut que chuter à la limite.

D'une manière géométrique équivalente, cela s'exprime aussi par le fait que les ensembles de sous-niveau de $\|.\|_0$,

$$\{x \in \mathbb{R}^n : \|x\|_0 \leqslant r\}\,,\ r \geqslant 0, \tag{4}$$

sont toujours fermés (dans $\mathbb{R}^n$), ou bien que l'épigraphe de $\|.\|_0$ (littéralement, "ce qui est au-dessus du graphe de $\|.\|_0$"),

$$\{(x,r) \in \mathbb{R}^n \times \mathbb{R} : \|x\|_0 \leqslant r\} \tag{5}$$

est un fermé (de $\mathbb{R}^n \times \mathbb{R}$).

L'approche par ensembles de sous-niveau est sans doute la plus rapide puisque ces ensembles sont des réunions finies de sous-espaces vectoriels.

- La fonction $\|.\|_0$ est *localement constante sur un ouvert dense* Ω de $\mathbb{R}^n$. Il existe en effet un ouvert dense Ω de $\mathbb{R}^n$, plus précisément $\Omega = \{x : \|x\|_0 = n\}$, avec la propriété suivante : pour tout $x_0 \in \Omega$, on peut trouver un voisinage ouvert V de x_0 tel que $\|.\|_0$ soit constante sur V. Malheureusement - ou heureusement - les choses intéressantes se passent sur l'ensemble fermé complémentaire Ω^c, c'est-à-dire "à la croisée des valeurs différentes de $\|x\|_0$".

- Conséquence des propriétés énoncées au-dessus pour l'optimisation.

Si on a à minimiser $x \mapsto \|x\|_0$ sur un ensemble-contrainte S de $\mathbb{R}^n$, alors *tout point de S est un minimiseur local.* Surprenant, non ? Bien sûr, cela ne serait pas possible avec une fonction à minimiser qui serait continue : Si tout point de $\mathbb{R}^n$ est un minimiseur local de la fonction continue $f : \mathbb{R}^n \to \mathbb{R}$, alors f est constante.

- La fonction $\|.\|_0$ est "séparable" en les variables x_i, au sens suivant :

$$\|(x_1, x_2, ..., x_n)\|_0 = \sum_{i=1}^{n} \gamma(x_i),$$

où γ est la fonction basique définie par : $\gamma(u) = 1$ si $u \neq 0$, $\gamma(u) = 0$ si $u = 0$. Propriété partagée avec la fonction-norme $\|(x_1, x_2, ..., x_n)\|_1 = \sum_{i=1}^{n} |x_i|$. Cette séparabilité en les variables aide dans les calculs d'optimisation relatifs à ces fonctions, en voici un exemple. Dans des procédures algorithmiques intervenant dans les problèmes où la parcimonie est un objectif ou une contrainte, on est conduit à "régulariser" la fonction $\|.\|_0$ de la manière suivante : pour un paramètre $\varepsilon > 0$ donné, on construit

$$\begin{aligned} x &= (x_1, x_2, ..., x_n) \in \mathbb{R}^n \mapsto c_\varepsilon(x) = \inf_{u \in \mathbb{R}^n} \left\{ \|u\|_0 + \frac{1}{2\varepsilon} \|x - u\|^2 \right\} \\ &= \inf_{u \in \mathbb{R}^n} \left\{ \sum_{i=1}^{n} \left[\gamma(u_i) + \frac{1}{2\varepsilon}(x_i - u_i)^2 \right] \right\} \\ &= \sum_{i=1}^{n} \inf_{v \in \mathbb{R}} \left[\gamma(v) + \frac{1}{2\varepsilon}(x_i - v)^2 \right]. \end{aligned} \tag{6}$$

Ainsi, il n'y a plus qu'un calcul sur les fonctions d'une seule variable à faire. De fait,

$$\gamma_\varepsilon(x_i) = \inf_{v\in\mathbb{R}} \left[\gamma(v) + \frac{1}{2\varepsilon}(x_i - v)^2\right] = \begin{cases} \frac{1}{2\varepsilon}x_i^2 \text{ si } |x_i| \leqslant \sqrt{2\varepsilon}, \\ 1 \text{ si } |x_i| \geqslant \sqrt{2\varepsilon}. \end{cases}$$

Le graphe de γ_ε est une parabole écrêtée, dont la courbure en 0, $\gamma_\varepsilon''(0) = 1/\varepsilon$, croît vers l'infini à mesure que ε diminue vers zéro.

En conséquence, $x \mapsto c_\varepsilon(x) = \sum_{i=1}^n \gamma_\varepsilon(x_i)$ est une fonction *continue* jouant le rôle d'approximation de la fonction $x \mapsto \|x\|_0$.

Q2. *Quels sont les liens de* $\|.\|_0$ *avec les normes usuelles* ℓ^p, $p > 0$ *?*

On rappelle que $\|x\|_p = \left(\sum_{i=1}^n |x_i|^p\right)^{1/p}$ est une norme seulement pour $p \geqslant 1$. La notation adoptée $\|x\|_\infty = \max_{i=1,\ldots,n} |x_i|$ est cohérente, même si ∞ n'est pas une valeur, puisque $\|x\|_\infty = \lim_{p\to\infty} \|x\|_p$ (exercice classique). Il est donc naturel de se poser la question : Quid du comportement de $\|x\|_p$ quand $p \to 0$?

- Pour $x \in \mathbb{R}^n$ fixé, $\|x\|_p$ a bien une limite quand $p \to 0$ ... mais ce n'est pas $\|x\|_0$. Pour cela, on commence par voir ce qui se passe dans $\mathbb{R}^2$: pour $a > 0$ et $b > 0$, $\|(a, b)\|_p \to \sqrt{ab}$ quand $p \to 0$. Puis, d'une manière générale, pour $x = (x_1, x_2, ..., x_n)$ dans $\mathbb{R}^n$ avec $x_i > 0$ pour tout i, on montre que

$$\begin{aligned} \lim_{p\to 0} \|x\|_p &= (x_1 x_2 ... x_n)^{1/n} \qquad (7) \\ &= \exp\left(\frac{1}{n}\sum_{i=1}^n \ln x_i\right) \end{aligned}$$

(ladite moyenne géométrique des x_i).

Pour $\left(\|x\|_p\right)^p = \sum_{i=1}^n |x_i|^p$, les choses sont bien plus faciles. On démontre aisément que

$$\lim_{p\to 0} \left(\|x\|_p\right)^p = \|x\|_0 . \qquad (8)$$

De ce fait, la notation $\|.\|_0$ n'est pas bonne ($\|.\|_0^0$ serait plus logique)... mais comme elle est très répandue, on l'utilise quand même. Autres notations utilisées dans la littérature : $c(x)$ (fonction de comptage ou de cardinalité de x), $nnz(x)$ (nombre de non-zéros dans x).

Q3. Un optimiseur pense utiliser $x \mapsto \|x\|_p$ ou $\left(\|x\|_p\right)^p$, avec p petit, comme substitut de la fonction (difficile à appréhender directement) $x \mapsto \|x\|_0$. *Quels seraient les avantages et les inconvénients d'une telle approche ?*

- Utiliser $\left(\|x\|_p\right)^p$ ou $\|x\|_p$, avec p petit (disons $0,5 \geqslant p \to 0$) comme approximation de $\|x\|_0$ peut être une idée. On gagne la continuité de la fonction en jeu, mais même avec elle on a perdu la convexité (que l'on avait pour $p \geqslant 1$).

Q4. Pour un entier k compris entre 1 et n, on définit $x = (x_1, x_2, ..., x_n) \mapsto \|x\|_{(k)}$ = somme des k plus grandes valeurs parmi les $|x_i|$. *Comment utiliser $\|.\|_{(k)}$ comme aide dans les problèmes où $\|.\|_0$ sert de mesure de parcimonie ?*

Il se trouve que $\|.\|_{(k)}$ est une norme sur $\mathbb{R}^n$ (facile à démontrer, seule l'inégalité triangulaire mérite une petite démonstration) ; sa boule-unité $B_{(k)}$ est un polyèdre convexe compact de $\mathbb{R}^n$. Ainsi, on a une suite de normes polyédrales $\|.\|_{(k)}$ qui va interpolant de $\|.\|_{(1)}$ $(= \|x\|_\infty)$ à $\|.\|_{(n)}$ $(= \|x\|_1)$; elles ont des qualités et une utilité différentes de celles des normes usuelles $\|.\|_p$.

Exemple avec $n = 3$ et $k = 2$. La partie de $B_{(2)}$ correspondant aux composantes positives est un polytope de $\mathbb{R}^3$ avec 4 sommets (de $B_{(2)}$) : $(1, 0, 0), (0, 1, 0), (0, 0, 1)$ et $(1/2, 1/2, 1/2)$. Curieux que dessiner ce morceau de polyèdre pose des problèmes aux étudiants... Le nombre total de sommets de $B_{(2)}$ dans $\mathbb{R}^3$ est de 14. Une étude plus détaillée des normes $\|.\|_{(k)}$ se trouve dans [2].

- Prenons un calculateur qui n'a accès qu'à $x \mapsto \|x\|_{(k)}$ et aux normes usuelles $x \mapsto \|x\|_p$. Comment alors remplacer la contrainte d'inégalité (difficile) $\|x\|_0 \leqslant k$ par une (ou des) contrainte(s) plus facilement exploitable(s) où n'interviendrait que ce à quoi a accès le calculateur ?

On a l'équivalence suivante, facile à démontrer :

$$(\|x\|_0 \leqslant k) \Leftrightarrow \Big(\|x\|_1 - \|x\|_{(k)} = 0 \text{ (ou bien } \leqslant 0)\Big). \tag{9}$$

Ainsi, la contrainte de parcimonie $\|x\|_0 \leqslant k$ a été remplacée par une équation (ou inéquation) exprimée à l'aide de deux normes polyédrales.

Q5. Les sommets de la boule-unité pour la norme ℓ^1 sont très "parcimonieuses", puisque ce sont les points dont toutes les composantes sont nulles, sauf une qui vaut 1. Dans les applications, on utilise $\|.\|_1$ comme substitut à $\|x\|_0$, car "elle favorise la parcimonie" des solutions, propriété recherchée avec l'intervention de $\|x\|_0$. *Quelle propriété mathématique reliant les fonctions $\|.\|_0$ et $\|.\|_1$ pourrait expliquer ces résultats satisfaisants ?*

En remplaçant $x \in \mathbb{R}^n \mapsto \|x\|_0$ par $x \in \mathbb{R}^n \mapsto \|x\|_1$ dans son critère d'optimisation, les variables x restant confinées (elles aussi !) dans le polyèdre convexe compact $K_r = \{x \in \mathbb{R}^n : \|x\|_\infty \leqslant r\}$, les utilisateurs dans des sciences diverses (dont la Géophysique) avaient constaté depuis longtemps que les solutions x obtenues étaient parcimonieuses, c'est-à-dire avec beaucoup de composantes nulles. Une propriété mathématique expliquant cela est celle de "relaxation" suivante.

Sur $K_r = \{x \in \mathbb{R}^n : \|x\|_\infty \leqslant r\}$, *l'enveloppe convexe de la fonction* $\|.\|_0$ (c'est-à-dire la plus grande fonction convexe minorant $\|.\|_0$ sur cet ensemble) *est la fonction* $\frac{1}{r}\|.\|_1$ (c'est un résultat mis en évidence par M. Fazel en 2002). Visuellement, à deux dimensions ($n = 2$), c'est très facile à voir. On pourrait résumer cela en disant que ce sont "les bienfaits de la relaxation"...

2ème partie : Contexte de l'espace des matrices $\mathcal{M}_{m,n}(\mathbb{R})$.

On pourrait penser que, pour une matrice, être "creuse" ou "parcimonieuse" signifie comporter beaucoup de termes nuls. C'est effectivement le cas en Analyse numérique. Mais ce n'est pas le cas dans des problèmes d'optimisation intervenant dans les Sciences des données. Dans ces problèmes, où les variables sont des matrices, la fonction-objectif ou bien les fonctions définissant les contraintes font apparaître une veille connaissance : le *rang* d'une matrice, soit la fonction $A \in \mathcal{M}_{m,n}(\mathbb{R}) \mapsto$ rang *de* A. Ainsi, dans ce type de problèmes, l'objectif d'avoir "une matrice A parcimonieuse" s'exprime comme celui d'avoir A de "rang le plus petit possible".

Q6. *Quel changement de variables permet de relier le rang à la fonction de parcimonie* $\|.\|_0$ *(traitée en 1ère partie) ?*

On désigne par p le plus petit des entiers m et n. La fonction $\sigma \in \mathbb{R}^p \mapsto \|\sigma\|_0$ est celle définie et étudiée dans la première partie.

Ce qui permet de passer du contexte de $\mathbb{R}^n$ à celui de $\mathcal{M}_{m,n}(\mathbb{R})$ est la *décomposition en valeurs singulières (DVS) d'une matrice* (notion importante !). Si $A \in \mathcal{M}_{m,n}(\mathbb{R})$, il existe des matrices orthogonales U, V (de tailles *ad hoc*), une matrice pseudo-diagonale D de même format que A, telles que

$$A = UDV.$$

D pseudo-diagonale signifie que les termes d_{ij} sont nuls si $i \neq j$. Les termes pseudo-diagonaux d_{ii} de D sont les valeurs singulières σ_i de A, c'est-à-dire les racines carrées des valeurs propres de AA^T (ou A^TA). On les range habituellement en ordre décroissant : $\sigma_1 \geqslant \sigma_1 \geqslant ... \geqslant \sigma_p$. La relation fondamentale suivante fait le lien entre la première et la deuxième partie de cette étude :

$$\text{rang} A = \|(\sigma_1, \sigma_2, ..., \sigma_p)\|_0 . \tag{10}$$

Bref, le rang de A est le nombre de valeurs singulières de A non nulles [2]. D'ailleurs la plateforme de logiciels Matlab calcule le rang d'une matrice de cette manière-là. Pour en savoir davantage sur les valeurs singulières, nous invitons le lecteur à consulter la référence [1].

---

2. Ou encore le nombre de valeurs propres strictement positives de la matrice positive AA^T (ou A^TA). D'ailleurs, rang A = rang AA^T = rang A^TA.

Q7. *Quelles propriétés, utiles pour l'analyse et l'optimisation de cette fonction* rang(.) *peut-on en déduire ?*

Les propriétés de $A \mapsto \text{rang}A$ utiles pour son analyse et son optimisation sont parallèles à celles listées dans la première partie. En voici quelques-unes :

- Le rang est une fonction semicontinue inférieurement : Si $A_\nu \to A$ quand $\nu \to +\infty$, alors

$$\lim_{\nu \to +\infty} \inf (\text{rang}A_\nu) \geqslant \text{rang}A. \tag{11}$$

(En bref, "Le rang ne peut que chuter à la limite" [3]).

- Si on a à minimiser $A \mapsto \text{rang}A$ sur un ensemble-contrainte $\mathcal{S}$ de $\mathcal{M}_{m,n}(\mathbb{R})$, *tout point de $\mathcal{S}$ est un minimiseur local.*

- Si $\text{rang}A = k < p$, alors dans tout voisinage de A il y a une matrice de rang $k+1, k+1, ..., p$.

- Exemple de propriété de $S_k = \{A \in \mathcal{M}_n(\mathbb{R}) : \text{rang}A \leqslant k\}$, pour $m = n$: S_k est une variété semi-algébrique de dimension $(2n-k)k$. De fait, S_k est un "infâme chewing-gum" difficile à traiter directement.

- Les normes matricielles usuelles s'expriment facilement à l'aide des valeurs singulières : si $\sigma_1 \geqslant \sigma_1 \geqslant ... \geqslant \sigma_p \geqslant 0$ sont les valeurs singulières de A,

$$\begin{aligned}
\|A\|_* &= \sum_{i=1}^{p} \sigma_i \text{ est la norme (dite) "nucléaire" de } A; \\
\|A\|_{sp} &= \max_{i=1,...,p} \sigma_i \text{ est la norme spectrale de } A; \\
\|A\|_F &= \sqrt{\sum_{i=1}^{p} \sigma_i^2} \text{ est la norme de } \textsc{Frobenius} \text{ de } A.
\end{aligned}$$

Ce sont simplement les normes $\ell^1, \ell^\infty, \ell^2$ de $\sigma = (\sigma_1, \sigma_2, ..., \sigma_p) \in \mathbb{R}^p$.

- "Relaxation" de la fonction rang(.) : sur $\mathcal{K}_r = \left\{A \in \mathcal{M}_{m,n}(\mathbb{R}) : \|A\|_{sp} \leqslant r\right\}$, l'enveloppe convexe de la fonction rang(.) (c'est-à-dire la plus grande fonction convexe minorant rang(.) sur cet ensemble) est la fonction $\frac{1}{r}\|.\|_*$.

"Tout ce que vous voulez savoir sur le rang, du point de vue variationnel, et que vous n'osez demander... " se trouve dans le papier-revue [3].

3ème partie : Contexte des espaces de Lebesgue L^p.

Le Calcul variationnel ou la théorie du Contrôle font apparaître la nécessité d'avoir des solutions (des fonctions ici) qui soient parcimonieuses, c'est-à-dire

3. On mémorise facilement cette propriété en pensant à l'exemple suivant : une matrice A_θ avec deux vecteurs-colonnes faisant un léger angle θ, qui à la limite (quand $\theta \to 0$) devient une matrice avec deux vecteurs colinéaires (pour $\theta = 0$).

prenant la valeur 0 le plus souvent possible. La réflexion et la recherche à cet égard concernent ici les fonctions des espaces $L^p(\Omega, \mathcal{F}, \mu)$, où μ est une mesure de probabilité. Pour $p > 0$, on rappelle que $\|f\|_p = \left(\int_\Omega |f(x)|^p \, d\mu \right)^{1/p}$.

Q8. *Quel jeu d'hypothèses sur f fixé permettrait qu'on ait une limite de $\|f\|_p$ quand $p \to 0$? une limite de $\left(\|f\|_p\right)^p$ quand $p \to 0$?*

Soit $f \in L^1(\Omega)$. Alors on a :

$$\lim_{p \to 0} \|f\|_p = \lim_{p \to 0} \left(\int_\Omega |f(x)|^p \, d\mu \right)^{1/p} = \exp\left[\int_\Omega \ln|f| \;\, d\mu \right]. \qquad (12)$$

Coucou revoilà la moyenne géométrique de f... Plusieurs jeux d'hypothèses sont possibles pour cela :

- aucune supplémentaire, si on pose $\exp(-\infty) = 0$;
- en supposant que $|f|$ et $\ln|f| \in L^1(\Omega)$;
- en supposant que $\|f\|_r < +\infty$ pour un certain $r > 0$ (et en posant $\exp(-\infty) = 0$).

De manière similaire, plusieurs jeux d'hypothèses sur f assurent que

$$\lim_{p \to 0} \left(\|f\|_p\right)^p = \int_\Omega 1_{\{x \,:\, f(x) \neq 0\}} d\mu = \mu\{x : f(x) \neq 0\}. \qquad (13)$$

Q9. Au vu des résultats qui précèdent, *quelle "mesure de parcimonie" pourrait-on proposer pour une fonction $f \in L^p(\Omega)$?*

Au vu de la relation (13) et des résultats des deux premières parties, une façon de mesurer la parcimonie de f est de considérer $\mu\{x : f(x) \neq 0\}$. Ainsi, le modèle général, dans un contexte de dimension infinie, prenant en compte la parcimonie du résultat, est

$$\left\{ \begin{array}{c} \text{Minimiser } \Phi(f) + K \;\mu\{x : f(x) \neq 0\} \quad (K > 0) \\ f \in \mathcal{U} \end{array} \right.$$

parallèle à celui en dimension finie

$$\left\{ \begin{array}{c} \text{Minimiser } \varphi(x) + K \;\|x\|_0 \\ x \in \mathcal{X}. \end{array} \right. .$$

Quant à savoir si $\|f\|_0 = \mu\{x : f(x) \neq 0\}$ peut être "convexifiée" ou "relaxée" en $\frac{1}{r}\|f\|_1$ sur la boule $\{f : \|f\|_\infty \leqslant r\}$, cela fait partie des développements sur cette "mesure de parcimonie fonctionnelle" $f \mapsto \|f\|_0$ qui nous éloignent de notre propos...

Références

1. O. Bordellès, *Les valeurs singulières.* Quadrature n° 80, 26 − 32 (2011).

2. M. Gaudioso et J.-B. Hiriart-Urruty, *Deforming* $\|.\|_1$ *into* $\|.\|_\infty$ *via polyhedral norms : a pedestrian approach.* Prépublication (Août 2020).

3. J.-B. Hiriart-Urruty et Hai Yen Le, *A variational approach of the rank function.* TOP (Journal of the Spanish Society of Statistics and Operations Research) Vol. 21, n° 2, 207 − 240 (2013).