Méthodes de Newton pour l'optimisation composite non-lisse

- **Gilles Bareilles** (Univ. Grenoble Alpes, France)
- Franck Iutzeler (Univ. Grenoble Alpes, France)
- Jérôme Malick (CNRS & LJK, Grenoble, France)

Mots-clé : optimisation non-lisse, identification, opérateur proximal, méthode de Newton.

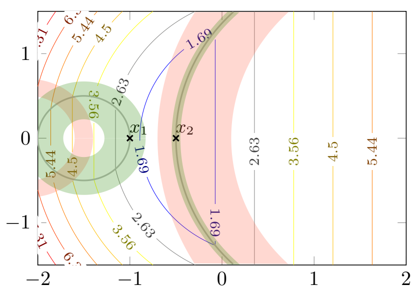
Résumé : Beaucoup de problèmes d'optimisation de machine learning ou de traitement du signal peuvent s'écrire comme la minimisation d'une composition de fonctions, non différentiable [1]. Cette non-différentiabilité est souvent structurée en une collection de variétés lisses où la fonction est lisse. Nous proposons une méthode permettant d'identifier le sous-espace d'un minimiseur et, suivant la philosophie SQP, en déduisons des algorithmes rapides.

Non différentiabilité et structure Nous considérons $\min_x F(x)$, où $F = g \circ c$, c est une application lisse et g une fonction non-différentiable non-convexe admettant un opérateur proximal explicite. Ces problèmes comprennent le maximum de fonctions lisses, la valeur propre maximale d'une matrice symétrique réelle, ou sa i-ème valeur propre (voir [2], [3] pour le cas valeur propre maximale).

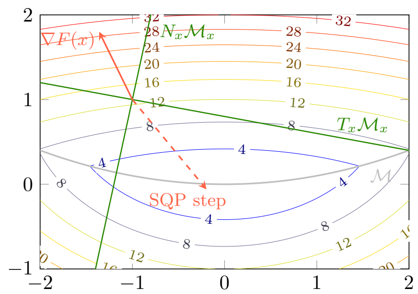
Problème équivalent La non différentiabilité est fréquemment structurée en une collection de variétés lisses où la fonction est lisse. Avec la connaissance du sous-espace final, le problème non différentiable se transforme en problème contraint différentiable, qui peut être résolu efficacement par une méthode SQP.

Cependant, la performance de la méthode SQP dépend fondamentalement de l'identification du bon sous-espace, qui n'est *jamais* connu avec certitude, et difficile à détecter de manière robuste. Les travaux existants supposent ce sous-espace connu a priori, ou estimé correctement par des heuristiques ([3, 4]).

Méthode proposée Nous proposons une approche originale basée sur l'analyse non-lisse pour déterminer la structure intéressante localement (fig. a), impliquant notamment l'opérateur proximal. Nous utilisons ensuite les méthodes classiques de SQP pour faire des pas efficaces (fig. b). Nous en déduisons un algorithme convergeant localement quadratiquement pour ces problèmes non-lisses.



(a) Identification de variété



(b) Iteration de Newton

Références :

[1] F. Iutzeler and J. Malick. Nonsmoothness in Machine Learning: specific structure, proximal identification, and applications. *Set-Valued and Variational Analysis*, Springer, 2020, 28 (4).

[2] A. Shapiro. On a Class of Nonsmooth Composite Functions. *Mathematics of Operations Research*, 28 (2003).

[3] D. Noll and P. Apkarian. Spectral bundle methods for non-convex maximum eigenvalue functions: second-order methods. *Mathematical Programming*, 104 (2005).

[4] X. Y. Han and A. S. Lewis. Survey Descent: A Multipoint Generalization of Gradient Descent for Nonsmooth Optimization. *preprint*, Nov. 2021, arxiv.