La dualité convexe comme accélération d'un algorithme de *Branch-and-Bound* dédié à l'optimisation parcimonieuse

Gwenaël Samain[1,2], Sébastien Bourguignon[1], Jordan Ninin[2]

[1] École Centrale de Nantes, LS2N, SIMS, F-44000 Nantes, France
{gwenael.samain,sebastien.bourguignon}@ls2n.fr
[2] ENSTA Bretagne, Lab-STICC, MATRIX, F-29000 Brest, France
jordan.ninin@ensta-bretagne.fr

Mots-clés : *optimisation, branch-and-bound, dualité, relaxation convexe, parcimonie.*

1 Introduction

Les problèmes parcimonieux – c'est-à-dire les problèmes d'ajustement de modèle de faible cardinalité – ont trouvé de nombreuses applications, notamment en statistique [4], en optimisation de portefeuille [6] ou encore en traitement du signal [5]. Parmi ces problèmes parcimonieux, nous nous intéressons au problème (1).

$$\min_{x\in\mathbb{R}^Q} \quad \frac{1}{2}\|y - Ax\|_2^2 + \lambda\|x\|_0 \quad \text{s.t.} \quad \|x\|_\infty \leq M \tag{1}$$

avec $\|x\|_0 := \text{Card}(\{i|x_i \neq 0\}) = |\{i|x_i \neq 0\}|$.

Nous nous intéressons ici à la résolution exacte du problème (1) via un algorithme de branch-and-bound, à l'instar de [3]. En particulier, nous cherchons à exploiter la dualité convexe pour accélérer le temps de calcul. Dans la section 2 nous présentons les spécificités de l'algorithme de branch-and-bound utilisé, ainsi que l'utilisation de la dualité convexe comme technique d'accélération. Dans la section 3 nous présentons la méthode de test de cette technique.

2 Une méthode de branch-and-bound dédiée

2.1 Structuration de l'espace de recherche

La principale utilité de l'algorithme de branch-and-bound consiste à trouver le bon *support* de la solution, c'est-à-dire les indices des variables x_i non nulles à l'optimum de (1). Cette recherche peut se reformuler en introduisant trois ensembles d'indices définis comme suit :

- S_0 : l'ensemble des indices des variables forcées à être hors du support de la solution,
- S_1 : l'ensemble des indices des variables forcées à être dans le support de la solution,
- $\bar{S}$: l'ensemble des indices des variables libres / indéterminées.

Diviser l'espace de recherche à un nœud donné revient à prendre un élément de $\bar{S}$ et à le mettre soit dans S_0, soit dans S_1. Ainsi, à un nœud donné, le sous-problème s'écrit sous la forme (2).

$$\min_{x\in\mathbb{R}^Q} \quad \frac{1}{2}\|y - A_{S_1}x_{S_1} - A_{\bar{S}}x_{\bar{S}}\|_2^2 + \lambda\|x_{\bar{S}}\|_0 + \lambda|S_1| \quad \text{s.t.} \quad \|x\|_\infty \leq M \tag{2}$$

2.2 Les bornes inférieures

Les bornes inférieures au sein de l'algorithme de branch-and-bound s'obtiennent par relaxation continue convexe du problème (2) et se formulent comme l'optimisation du problème (3).

$$\min_{x\in\mathbb{R}^Q} P(x) := \quad \frac{1}{2}\|y - A_{S_1}x_{S_1} - A_{\bar{S}}x_{\bar{S}}\|_2^2 + \frac{\lambda}{M}\|x_{\bar{S}}\|_1 + \lambda|S_1| \quad \text{s.t.} \quad \|x\|_\infty \leq M \tag{3}$$

En effet, sous la contrainte de borne $\|x\|_\infty \leq M$, on a $\|x_{\bar{S}}\|_1 := \sum_{i\in\bar{S}} |x_i| \leq M\|x_{\bar{S}}\|_0$.

Depuis une vingtaine d'années, l'optimisation de critères comportant une norme ℓ_1 bénéficie de nombreux algorithmes dédiés efficaces [1]. La majorité du temps nécessaire à la résolution de (1) est imputable au temps de calcul des bornes inférieures. L'objectif est de réduire ce temps passé à les calculer. Nous proposons d'utiliser les propriétés de la dualité convexe, et notamment le théorème de Fenchel-Rockafellar [2]. Ce théorème nous permet d'exprimer le problème dual de (3) comme le problème (4).

$$\min_{w\in\mathbb{R}^N} D(w) := \tfrac{1}{2}(\|w+y\|_2^2 - \|y\|_2^2) + M\left(\|-A_{S_1}^T w\|_1 + \sum_{i\in\bar{S}} \max(0, |-A_i^T w| - \tfrac{\lambda}{M})\right) + \lambda|S_1| \quad (4)$$

Plus précisément, nous proposons d'utiliser la dualité faible : $\forall(x,w), P(x) + D(w) \geq 0$, pour réduire le temps de calcul des bornes inférieures par deux moyens complémentaires :

— l'utilisation d'une méthode de *screening* utilisant $D(w)$ [7], qui permet de réduire la dimension du problème (3) au cours de sa résolution en fixant des variables à 0 ou $\pm M$;

— la comparaison de $D(w)$ avec la meilleure solution connue z_U. En effet, si l'on a un w tel que $D(w) \leq -z_U$, alors $\min_x P(x) \geq z_U$, et on peut élaguer précocement le nœud.

3 Évaluation expérimentale

Dans cette étude, nous testons plusieurs algorithmes de l'état de l'art [1] dédiés à la résolution du problème (3) : une méthode homotopique, une méthode par ensemble actif, une méthode de descente par coordonnées ainsi que quelques algorithmes proximaux. L'objectif est de déterminer la méthode diminuant le plus rapidement $D(w)$.

Avec cette méthode, nous évaluons l'impact de l'utilisation de la dualité faible au sein d'un algorithme de branch-and-bound. Pour cela, nous constituons des jeux de données synthétiques représentatifs de problèmes de sélection de variables en statistique [4]. En faisant varier le niveau de parcimonie K, le nombre d'inconnues Q, le niveau de bruit dans les données et la structure plus ou moins complexe de la matrice A, nous évaluons la robustesse de la méthode proposée et son applicabilité à des données expérimentales.

Références

[1] F. Bach, R. Jenatton, J. Mairal et G. Obozinski. *Optimization with Sparsity-Inducing Penalties.* now publishers Inc, 2011.

[2] H. Bauschke et P. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Space.* Springer, 2011.

[3] R. Ben Mhenni, S. Bourguignon et J. Ninin. Global optimization for sparse solution of least squares problems. *Optimization Methods and Software*, 2021.

[4] D. Bertsimas, A. King et R. Mazumder. Best subset selection via a modern optimization lens. *The Annals of Statistics*, 2016.

[5] S. Bourguignon, J. Ninin, H. Carfantan et M. Mongeau. Exact sparse approximation problems via mixed-integer programming : formulations and computational performance. *IEEE Transactions on Signal Processing*, 2016.

[6] X. Cui, X. Zheng, S. S. Zhu et X. Sun. Convex relaxations and miqcqp reformulations for a class of cardinality-constrained portfolio selection problems. *Journal of Global Optimization*, 2013.

[7] A. Raj, J. Olbrich, B. Gärtner, B. Schölkopf et M. Jaggi. Screening rules for convex problems. In *9th NIPS Workshop on Optimization for Machine Learning*, 2016.