

# Un algorithme branch-and-bound pour résoudre exactement des problèmes d'optimisation parcimonieuse structurée

Gwenaël Samain<sup>1,2</sup>, Sébastien Bourguignon<sup>1</sup>, Jordan Ninin<sup>2</sup>

<sup>1</sup> École Centrale de Nantes, LS2N, équipe SIMS, F-44000 Nantes, France  
 {gwenael.samain,sebastien.bourguignon}@ls2n.fr

<sup>2</sup> ENSTA Bretagne, Lab-STICC, équipe MATRIX, F-29000 Brest, France  
 jordan.ninin@ensta-bretagne.fr

## 1 Introduction

Les problèmes d'ajustement de modèles de faible cardinalité (dits parcimonieux) ont trouvé de nombreuses applications, notamment en statistique [2], en optimisation de portefeuille [3] ou encore en traitement du signal [2]. Nous nous intéressons en particulier au problème de parcimonie structurée (voir [1] pour des applications pratiques) :

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

où  $\|u\|_0 := \text{Card}\{i | u_i \neq 0\}$  et  $P \in \{0;1\}^{R \times Q}$  est la matrice sélectionnant des groupes de variables sur lesquels porte la mesure de cardinalité. Ainsi  $P = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}$  dénote un problème à 2 groupes.  $\|P|x|\|_0 = 1$  signifie alors que soit  $|x_1| + |x_2| = 0$ , soit  $|x_1| + |x_3| = 0$ .

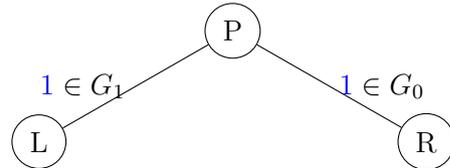
À notre connaissance, il existe des approches par relaxation convexe pour résoudre approximativement ce problème [1], mais aucune méthode de résolution exacte n'a encore été proposée. Nous proposons un algorithme de branch-and-bound dans ce but, qui généralise les travaux développés dans [2] dans le cas de modèles parcimonieux *simples* ( $P = I_Q$ ).

## 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 information se retrouve encodée dans les groupes *actifs* et *inactifs* : en connaissant le support des groupes de la solution on en déduit le support des variables  $x_i$ . Cette recherche peut se reformuler en introduisant trois ensembles d'indices définis comme suit :

$$P = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \quad \begin{array}{l} \bar{G} = \{1, 2\}, G_1 = \emptyset, G_0 = \emptyset \\ \bar{S} = \{1, 2, 3\}, S_1 = \emptyset, S_0 = \emptyset \end{array}$$



$$\begin{array}{ll} \bar{G} = \{2\}, G_1 = \{1\}, G_0 = \emptyset & \bar{G} = \{2\}, G_1 = \emptyset, G_0 = \{1\} \\ \bar{S} = \{1, 3\}, S_1 = \{2\}, S_0 = \emptyset & \bar{S} = \{3\}, S_1 = \emptyset, S_0 = \{1, 2\} \end{array}$$

FIG. 1 – Exemple de division d'un noeud

$G_0$  contient l'ensemble des indices des groupes forcés à être hors du support de la solution,  $G_1$  les indices des groupes forcés à être dans le support de la solution,  $\bar{G}$  les indices des groupes libres. Diviser l'espace de recherche à un nœud donné revient à prendre un élément de  $\bar{G}$  et à le mettre soit dans  $G_0$ , soit dans  $G_1$ .

À ces supports de groupe, on ajoute des supports de variables ( $S_0, S_1, \bar{S}$ ), tels que toutes les variables appartenant à un groupe de  $G_0$  sont dans  $S_0$ , toutes les variables qui n'appartiennent qu'à des groupes de  $G_1$  sont dans  $S_1$ , les autres sont dans  $\bar{S}$  (voir la Figure 1 pour un exemple). Ainsi, à un nœud donné, le sous-problème s'écrit sous la forme (2) :

$$\min_{x \in \mathbb{R}^Q} \frac{1}{2} \|y - A_{S_1} x_{S_1} - A_{\bar{S}} x_{\bar{S}}\|_2^2 + \mu \text{Card}(G_1) + \mu \|P_{\bar{G}} |x|\|_0 \quad \text{s.t.} \quad x_{S_0} = 0, \|x\|_\infty \leq M. \quad (2)$$

## 2.2 Les bornes inférieures et la stratégie de branchement

Sachant la contrainte  $\|x\|_\infty \leq M$ , on a  $\mu \|P_{\bar{G}} |x|\|_0 = \mu \sum_{i \in \bar{G}} \mathbf{1}_{P_i |x| \neq 0} \geq \mu \sum_{i \in \bar{G}} \frac{P_i |x|}{\|P_i\|_0 M}$ , avec  $P_i$  la  $i$ -ème ligne de  $P$ . On peut donc formuler le problème de borne inférieure comme la relaxation (3).

$$\min_x \frac{1}{2} \|y - A_{S_1} x_{S_1} - A_{\bar{S}} x_{\bar{S}}\|_2^2 + \mu |G_1| + \frac{\mu}{M} \sum_{i \in \bar{G}} \sum_{j \in \bar{S}} \frac{P_{ij} |x_j|}{(\|P_i\|_0)} \quad \text{s.t.} \quad \|x\|_\infty \leq M \quad (3)$$

En posant  $\alpha_j = \sum_{i \in \bar{G}} \frac{P_{ij}}{\|P_i\|_0}$  et  $\forall j \in \bar{S}, z_j = \alpha_j x_j$ , il vient  $\sum_{i \in \bar{G}} \sum_{j \in \bar{S}} \frac{P_{ij} |x_j|}{\|P_i\|_0} = \sum_{j \in \bar{S}} |x_j| \alpha_j = \sum_{j \in \bar{S}} |z_j| = \|z_{\bar{S}}\|_1$ . qu'on utilise pour écrire le problème (3) en  $z$  comme dans (4) :

$$\min_{z \in \mathbb{R}^Q} \frac{1}{2} \|y - \mathcal{A}z\|_2^2 + \frac{\mu}{M} \|z\|_1 \quad \text{s.t.} \quad z_{S_0} = 0, \forall j \in \bar{S}, z_j \in [-\alpha_j M, \alpha_j M], \|z_{S_1 \cup S_0}\|_\infty \leq M \quad (4)$$

où  $\forall j \in \bar{S}, \mathcal{A}_j = A_j / \alpha_j; \forall j \in S_0 \cup S_1, \mathcal{A}_j = A_j$  et  $z_j = x_j$ .

Le changement de variable de  $x$  à  $z$  permet l'utilisation de la méthode homotopique qui est très efficace quand le degré de parcimonie recherché est faible [2].

Pour savoir sur quel groupe brancher quand on divise un nœud, on étend la stratégie du maximum des amplitudes utilisées dans [2] au cas structuré en prenant le terme de régularisation de la borne inférieure :  $i = \arg \max_{i \in \bar{G}} \sum_{j \in \bar{S}} \frac{P_{ij} |x_j|}{\|P_i\|_0}$ .

## 3 Évaluation expérimentale

Dans cette étude, nous évaluons l'impact de l'utilisation d'une méthode exacte  $\ell_0$  structurée par rapport à une méthode exacte  $\ell_0$  non structurée. Cette évaluation porte sur la qualité des solutions obtenues ainsi que sur le temps de calcul. Pour cela, nous constituons des jeux de données synthétiques représentatifs de problèmes de sélection de variables en traitement du signal. 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. Structured sparsity through convex optimization. *Statistical Science*, 2012.
- [2] R. BEN MHENNI. *Méthodes de programmation en nombres mixtes pour l'optimisation parcimonieuse en traitement du signal*. Thèse de doctorat, École Centrale de Nantes, mai 2020.
- [3] 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.