



HAL
open science

Optimisation robuste des réseaux de télécommunications

Olivier Klopfenstein

► **To cite this version:**

Olivier Klopfenstein. Optimisation robuste des réseaux de télécommunications. Modélisation et simulation. Université de Technologie de Compiègne, 2008. Français. NNT: . tel-00321868

HAL Id: tel-00321868

<https://theses.hal.science/tel-00321868>

Submitted on 16 Sep 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE DE DOCTORAT EN INFORMATIQUE

OPTIMISATION ROBUSTE DE RÉSEAUX DE TÉLÉCOMMUNICATIONS

Olivier Klopfenstein

Université de Technologie de Compiègne - Laboratoire Heudiasyc UMR CNRS 6599
Spécialité Technologies de l'Information et des Systèmes

Thèse soutenue le 02 juillet 2008

Jury :

Walid BEN AMEUR	TELECOM SudParis	(examineur)
J. Frédéric BONNANS	INRIA - École Polytechnique	(rapporteur)
Jacques CARLIER	Université de Technologie de Compiègne	(examineur, président du jury)
Michel MINOUX	Université Paris 6	(rapporteur)
Dritan NACE	Université de Technologie de Compiègne	(directeur de thèse)
Adam OUOROU	France Télécom	(examineur)

Remerciements

Mes premières pensées vont à Bernard Liau, mon responsable d'équipe à France Télécom il y a quelques années, qui a validé, encouragé et rendu possible ce projet de thèse. Marie-Hélène Briens, qui lui a succédé, a continué de soutenir ce travail et d'en garantir l'aboutissement. Dritan Nace, de l'Université de Technologie de Compiègne, a accepté d'assurer la responsabilité scientifique de cette thèse. Son suivi, sérieux et empressé, parfois exigeant, m'a été une aide précieuse. La forme et le contenu du document final lui doivent beaucoup.

Les professeurs J. Frédéric Bonnans, de l'INRIA - École Polytechnique, et Michel Minoux, de l'Université Paris 6, m'ont fait l'honneur d'accepter d'être rapporteurs de ce travail. Walid Ben Ameer, de l'Institut National des Télécommunications, Jacques Carlier, de l'Université de Technologie de Compiègne, et Adam Ouorou, de France Télécom, ont bien voulu en être les examinateurs. Jacques Carlier a en particulier présidé le jury.

Enfin, Clotilde Durand-Gasselin, François Perreau et David Savourey m'ont spontanément proposé de relire le manuscrit, et j'ai eu la faiblesse d'abuser de leur inconscience. Nombre de scories ont ainsi été éliminées du document.

À chacun, j'adresse mes plus sincères remerciements.

(Paris, juillet 2008.)

Table des matières

Résumé	vii
Résumé étendu	viii
Extended abstract	x
Notations	xii
1 Préliminaires	1
1.1 Généralités sur l'optimisation en contexte incertain	1
1.1.1 Programmation stochastique	2
1.1.2 Optimisation robuste	3
1.2 Quelques modèles	3
1.2.1 Optimisation robuste à partir de scénarios	3
1.2.2 Un cadre plus général	5
1.2.3 Liens avec la programmation sous contraintes probabilistes	6
I Approches robustes pour les problèmes sous contraintes probabilistes	9
2 Sous-ensembles d'événements pour l'optimisation robuste	11
2.1 Sous-ensembles optimaux d'événements	11
2.2 Propriétés spécifiques dans un contexte convexe	14
2.2.1 Ensemble convexe d'événements	14
2.2.2 Un cas particulier	15
2.3 Ensembles polyédriques	17
2.4 Sous-ensembles d'incertitudes pour la programmation linéaire	23
2.5 Un exemple simple	24
3 Utilisation du modèle robuste de Bertsimas et Sim pour la PLNE	29
3.1 Modèle d'incertitude et sur-conservatisme	29
3.2 Complexité du problème robuste	31
3.3 Aspects polyédriques généraux	32
3.3.1 Polyèdre associé au relâché continu	33
3.3.2 Polyèdre associé au problème robuste en nombres entiers	35
3.3.3 Autres inégalités	40
3.4 Le problème de sac-à-dos robuste 0-1	41
3.4.1 Étude polyédrique	42
3.4.2 Séparation des inégalités de couvertures robustes	46
3.4.3 Incertitude sur la capacité du sac	47
3.4.4 Tests numériques	48
3.5 Conclusion	51

4	Une approche robuste pour le problème de sac-à-dos sous contrainte probabiliste	55
4.1	Travaux antérieurs sur la programmation sous contraintes probabilistes avec nombres entiers	55
4.2	Le problème de sac-à-dos sous contraintes probabilistes	57
4.3	Lien entre RKP et CCKP	58
4.4	Algorithme pour la résolution approchée de CCKP	62
4.5	Conclusion	66
5	Généralisation aux problèmes combinatoires sous contraintes probabilistes	67
5.1	Modèles mathématiques	67
5.2	Lien entre RILP et CCILP	69
5.3	Complexité théorique et pratique de RILP	71
5.3.1	Relations de complexité entre RILP et ILP	71
5.3.2	Une formulation linéaire en nombres entiers de RILP	72
5.3.3	Description des ensembles d’incertitudes	74
5.4	Algorithmes efficaces pour CCILP	75
5.4.1	Calcul de la probabilité de réalisabilité d’une solution	75
5.4.2	Difficultés et restrictions	78
5.4.3	Algorithmes	79
5.4.4	Bornes supérieures pour CCILP	82
5.4.5	Cas particulier du sac-à-dos	83
5.5	Tests numériques	85
5.5.1	Description des instances et des méthodes de résolution	86
5.5.2	Le problème de sac-à-dos	87
5.5.3	Problèmes de sac-à-dos multi-dimensionnels	90
5.5.4	Synthèse	94
5.6	Conclusion	94
II Résolution optimale de problèmes combinatoires sous contraintes probabilistes		99
6	Résolution optimale de problèmes combinatoires sous contraintes probabilistes	101
6.1	Notations et préliminaires	101
6.1.1	Cadre principal	101
6.1.2	Extensions de l’analyse	102
6.2	Description de X à l’aide d’inégalités linéaires	102
6.2.1	Modèles déterministes équivalents	102
6.2.2	Inégalités linéaires pour caractériser X	103
6.2.3	Idée de généralisation à un contexte non-linéaire	105
6.3	Étude des scénarios basiques	106
6.3.1	Structure de l’ensemble des scénarios basiques	106
6.3.2	Construction de “bons” scénarios basiques	107
6.3.3	Illustration sur un exemple simple	110
6.4	Algorithme de résolution	112
6.4.1	Séparation des inégalités (6.4)	112
6.4.2	Calcul de $D_j(I_1)$	114
6.4.3	Calcul des probabilités	116

6.4.4	Contraintes de probabilités jointes	117
6.5	Tests numériques	117
6.5.1	Problèmes de sac-à-dos multi-dimensionnels	118
6.5.2	Un problème de localisation avec contrainte de capacité	122
6.6	Conclusion	124
III Applications		127
7	Localisation de points de mesure dans un réseau	129
7.1	Introduction	129
7.2	Description du problème	130
7.2.1	Notations et modèles mathématiques	130
7.2.2	Approche pessimiste	131
7.2.3	Construction des intervalles de trafic	131
7.3	Un algorithme efficace	132
7.3.1	Forme polynomiale du problème robuste	132
7.3.2	Algorithme de résolution	134
7.3.3	Hypothèses probabilistes	135
7.4	Tests sur un réseau IP de France Télécom	135
7.4.1	Problème spécifique traité	135
7.4.2	Données SNMP	136
7.4.3	Résultats et analyses	137
7.5	Pistes pour une analyse empirique et la simulation de Monte-Carlo	139
7.6	Conclusion	141
8	Dimensionnement de réseaux d'accès	143
8.1	Introduction	143
8.2	Problème traité et modèles mathématiques	144
8.2.1	Modèle déterministe de base	144
8.2.2	Modélisation des incertitudes	146
8.3	Hypothèses probabilistes	147
8.3.1	Présentation d'un cadre gaussien	147
8.3.2	Modèle déterministe équivalent	149
8.3.3	Prise en compte de corrélations	150
8.3.4	Exemple d'un autre contexte probabiliste : l'inégalité de Hoeffding	150
8.4	Relaxation du problème probabiliste pour le cas gaussien	151
8.5	Tests numériques	153
8.5.1	Algorithme de résolution	153
8.5.2	Description des données	153
8.5.3	Résultats obtenus	154
8.6	Conclusion	156
9	Conclusions et perspectives	157
	Publications	159
Annexes		163
A	Amélioration des résultats de la partie 3.3.2	163

B	Notes sur le calcul de probabilités	165
B.1	Distribution de la somme de deux variables aléatoires	165
B.2	Différents théorèmes de la limite centrale	165
B.3	Validité de l'inégalité de Hoeffding pour des v.a. corrélées	166
C	Annexes du chapitre 5	169
C.1	Note sur un modèle robuste simpliste	169
C.2	Propriétés de la valeur de $RILP(\gamma)$ en fonction de γ	170
C.3	Bornes supérieures pour CCILP	171
D	Optimisation avec contraintes probabilistes et VaR/CVaR	173
D.1	Concepts de VaR et de CVaR	173
D.2	Lien avec les contraintes probabilistes	174
	Bibliographie	175

Résumé : Cette thèse est consacrée à la prise en compte de données incertaines dans les problèmes d'optimisation. On se concentre sur la programmation mathématique sous contraintes probabilistes, dont le but est de trouver la meilleure solution qui sera réalisable avec une probabilité minimale garantie. Par ailleurs, on s'intéresse à la prise en compte de variables de décisions entières, qui sont souvent requises en pratique.

Pour résoudre de tels problèmes combinatoires sous contraintes probabilistes, on s'appuie d'abord sur l'optimisation robuste. Les liens théoriques entre ces deux familles de méthodes sont mis en évidence. À partir de modèles robustes appropriés, des algorithmes de résolution heuristique sont définis. On s'intéresse ensuite à la résolution optimale de problèmes combinatoires sous contraintes probabilistes. Des tests numériques illustrent les méthodes présentées et montrent leur efficacité pratique. Enfin, deux applications au domaine des télécommunications sont développées. Elles concernent toutes deux la localisation de fonctions dans un réseau.

Abstract : This thesis deals with taking uncertain data into account in optimization problems. Our focus is on mathematical programs with chance constraints. In this approach, the goal is to find the best solution, given a unfeasibility probability tolerance. Furthermore, we concentrate on integer variables, since they are often required for practical applications.

To solve such chance-constrained combinatorial problems, we first rely on robust optimization. The theoretical links between these two families of approaches are studied. Based on appropriate robust models, solution algorithms to chance-constrained combinatorial problems are designed. Then, the optimal solution of such problems is also investigated. Specific theoretical results and algorithms are given to reach optimality. Finally, two specific telecommunication applications are developed. Both of them deal with location in a network.

Résumé étendu

Le plus souvent dans la pratique, les techniques de recherche opérationnelle utilisées supposent une parfaite connaissance de données d'entrée. Or il est clair que cela n'est pas le cas pour un grand nombre d'applications. Considérer des incertitudes dans des programmes mathématiques n'est pas un souci nouveau. Dès les années 50, des auteurs tels que G.B. Dantzig envisagent la prise en compte d'incertitudes sur les données d'un programme linéaire. Si le sujet n'est pas neuf, il est vaste, et des développements récents lui ont donné un nouveau souffle. C'est dans cet élan que s'inscrit cette thèse.

Divers modèles ont été proposés pour rendre compte de manière aussi satisfaisante que possible d'incertitudes sur les données. Le présent travail a principalement concerné la programmation sous contraintes probabilistes avec variables de décisions entières. Dans un programme sous contrainte probabiliste, le but est de trouver la meilleure solution qui satisfait les contraintes définies avec une probabilité minimum garantie (par exemple, la meilleure solution réalisable avec une probabilité 0,95). Faire de ce modèle le pivot de cette thèse relève d'une double motivation. La première vient de la nature des problèmes de décisions rencontrés en pratique. La plupart d'entre eux se modélisent de façon pertinente avec des contraintes probabilistes. Par ailleurs, nombre de problèmes nécessitent la prise en compte de variables de décision discrètes (programmation en nombres entiers), surtout dans un domaine tel que les télécommunications. La seconde motivation a été le faible nombre de contributions dédiées à la résolution de tels problèmes dans la littérature à ce jour.

Un des axes pour aborder la résolution de programmes en nombres entiers sous contraintes probabilistes a été l'optimisation robuste. Cette dernière consiste à trouver la meilleure solution réalisable pour un ensemble d'événements donné. Plusieurs articles assez récents ont en effet mis en évidence le lien existant entre ces deux approches de l'incertitude. Après les quelques préliminaires du chapitre 1, les chapitres 2 à 5 traitent directement de la manière d'utiliser des modèles d'optimisation robuste pour approcher des contraintes probabilistes. L'idéal serait de pouvoir construire un ensemble d'événements E "optimal", c'est-à-dire tel que la meilleure solution réalisable pour E soit solution du problème sous contraintes probabilistes associé. La motivation principale pour une telle méthodologie est la fréquente simplicité de résolution des problèmes robustes. Le chapitre 2 décrit quelques propriétés d'un tel ensemble E . Il illustre cependant la difficulté de construire E de manière optimale : ceci est en fait équivalent à résoudre le problème sous contraintes probabilistes. On devra donc vraisemblablement se contenter d'algorithmes approchés (heuristiques) avec de telles approches robustes. Ceci n'invalide toutefois pas la démarche, qui reste intéressante en pratique, vue la difficulté d'abord des problèmes combinatoires sous contraintes probabilistes.

Le chapitre 3 se concentre sur un modèle spécifique d'optimisation robuste très populaire, proposé par Bertsimas et Sim en 2003. Bien qu'étant directement utilisable pour des programmes comportant des variables entières, cet aspect a été relativement peu étudié jusque-là. Des résultats de complexité et des caractérisations polyédriques sont donnés. Le problème fondamental du sac-à-dos robuste fait l'objet de développements spécifiques. On montre comment des résultats classiques, connus pour le problème de sac-à-dos avec poids connus, s'étendent au contexte robuste, où les poids sont supposés incertains. En particulier, les inégalités classiques de couverture sont transposées pour le modèle robuste ; on montre qu'elles peuvent définir des facettes du sac-à-dos robuste dans de nombreux cas.

Le chapitre suivant revient à l'objet principal de la thèse : la programmation sous contraintes probabilistes avec variables entières. On montre comment le modèle robuste de Bertsimas et Sim peut être utilisé pour fournir de bonnes solutions au problème de sac-à-dos sous contrainte probabiliste. Des affinités théoriques entre le modèle robuste et le modèle sous contraintes

probabilistes sont mises en évidence. L'algorithme de résolution développé pour le problème sous contrainte probabiliste est de faible complexité, et résout le problème à l'optimum dans certains cas simples. Le chapitre 5 étend cette approche aux problèmes généraux avec nombres entiers. Un modèle robuste original est introduit, dont les affinités avec les contraintes probabilistes sont étudiées. Avec des idées semblables à celles du chapitre 4, un algorithme de résolution est défini, qui repose sur la résolution successive de plusieurs problèmes robustes. Des tests numériques nombreux montrent la pertinence de l'approche en pratique. Ce type d'algorithme itératif peut être appliqué sur des instances de grandes tailles (plusieurs centaines, voire plusieurs milliers, de variables entières). En effet, le modèle robuste proposé est de complexité équivalente au problème nominal sans incertitudes : il suffit donc en pratique d'être capable de résoudre plusieurs fois d'affilée un problème nominal.

Jusque-là, des approches heuristiques ont été proposées pour résoudre des problèmes combinatoires sous contraintes probabilistes à partir de modèles robustes. La motivation principale était d'ordre pratique : il est indispensable de pouvoir apporter des solutions à ces problèmes, qui peuvent concrètement être de très grande taille. L'objet du chapitre 6 est la résolution optimale de ces problèmes. Une formulation linéaire implicite est donnée. Elle comporte un nombre exponentiel d'inégalités, dont la séparation est étudiée. Un algorithme de *branch-and-bound* est défini et testé numériquement. Il permet de résoudre à l'optimum des problèmes comportant jusqu'à quelques centaines de variables 0-1. La méthode proposée est particulièrement adaptée au cas où les variables aléatoires du problème sont toutes identiquement distribuées, à une constante additive près. Elle est néanmoins utilisable dans des contextes très généraux.

Enfin, les chapitres 7 et 8 sont consacrés à deux applications dans le domaine des télécommunications. La première concerne le placement de points de mesures du trafic dans un réseau. Le problème est modélisé de manière originale et très pertinente à l'aide d'une contrainte probabiliste. Cette dernière permet d'assurer une bonne stabilité dans le temps à la décision prise : les points de mesure sélectionnés resteront valides malgré les inévitables variations futures du trafic. Des tests numériques sur un cœur de réseau IP international de France Télécom montrent l'intérêt, et même la nécessité d'une telle approche par rapport à d'autres méthodes plus naïves (reposant par exemple sur des valeurs moyennes de trafic). Enfin, la deuxième application traite d'un problème de planification d'investissements futurs. Il s'agit de décider des sites de raccordement pour un ensemble de sites clients, de manière à minimiser les coûts de raccordement et d'équipement des sites. Ces opérations de planification s'appuient sur des prévisions de trafic qui se révèlent souvent loin de la réalité observée *a posteriori*. La prise en compte d'incertitudes est donc cruciale. Des modèles sous contraintes probabilistes sont proposés, et des algorithmes de résolution heuristiques sont définis.

Extended abstract

The most widely used methods of operations research are based on data which are supposed to be perfectly known. However, most of the time, this latter assumption is not satisfied. The need for taking account of data uncertainty in mathematical programming is not a recent concern. Authors such as G.B. Dantzig have already dealt with such issues in the fifties. While being not new, this field of research is wide, and recent developments have brought a renewed interest for it.

Many different approaches have been proposed to deal with uncertainty. The present work has focused on chance-constrained integer programming. In a chance-constrained program, the goal is to find the best solution being feasible according to a target probability (for instance, the best solution feasible with probability 0.95). There are two main motivations for dealing with chance-constrained integer programs. First, many real-life problems can be modeled with chance constrained in a very suitable way. Secondly, they frequently involve integer decision variables, and despite the critical importance of this aspect in practice, very few earlier works deal with general combinatorial chance-constrained problems.

Robust optimization is one of the tractable ways of dealing with chance constraints, especially with integer variables. The aim of robust optimization is to find the best solution feasible for a pre-defined set of events. Several recent works have underlined the link existing between robust optimization and chance-constrained programming. After some preliminaries in Chapter 1, chapters 2 to 5 deal with using robust optimization models for obtaining good solution to chance-constrained ones. Ideally, we would like to build the “optimal” set E of events such that an optimal solution of the associated robust problem is optimal also for the chance-constrained problem. The motivation for such an approach is that, very often, robust optimization models are quite tractable (especially when compared to chance-constrained ones). Chapter 2 gives some properties of such a set E . The difficulties of building an optimal set of events are highlighted : it is in fact equivalent to finding the optimal solution of the chance-constrained problem. This shows that robust approaches to chance-constrained problems will probably remain approximate. However, robust approaches remain interesting in practice because of their usual tractability.

Chapter 3 focuses on a specific and quite popular robust model proposed by Bertsimas and Sim in 2003. While being usable directly for integer programming, this aspect has received very few attention by the past. Some complexity results and polyhedral characteristics are given. The robust knapsack problem has been specifically studied, because of its fundamental importance for integer programming. We show how classical results for the knapsack problem can be adapted to the robust context, when weights are supposed uncertain. In particular, the widely known cover inequalities are transposed for the robust model. They are shown to define facets of the robust polyhedron in many cases.

Chapter 4 comes back to the main topic of this thesis : chance constraints with integer variables. We show how the robust model of Bertsimas and Sim can be used to design tractable solution heuristics for the chance-constrained knapsack problem. Some theoretical affinity between the chance-constrained model and its robust version are underlined. The proposed algorithm has low complexity, and solves the chance-constrained problem to optimality in some simple cases. Chapter 5 extends this approach to general integer programming. A new robust model is introduced, and its relations with chance-constrained models are studied. With ideas very similar to those of Chapter 4, a solution heuristic is designed for chance-constrained integer programs. Numerous computational tests show the relevance of the approach in practice. It can be used for large combinatorial problems; indeed, the approach is usable as soon as several nominal problem instances (*i.e.* without uncertainty) can be solved successively.

Until now, only heuristic algorithms have been proposed to solve combinatorial chance-constrained problems, using robust optimization. The main motivation for this work was practical tractability. Chapter 6 is devoted to optimal resolution of these problems. An implicit linear formulation of chance-constrained 0-1 programs is given. It contains an exponential number of inequalities, whose separation is investigated. A branch-and-bound solution algorithm is then defined. Numerical tests show that medium-size problems can be solved to optimality with the proposed algorithm (up to a few hundreds of 0-1 variables). The method is well adapted when random variables are independent and identically distributed, up to an additive constant. However, it is usable in far more general contexts.

Finally, chapters 7 and 8 are devoted to telecommunication applications. Chapter 7 deals with selecting points of traffic measurement in a network. The aim is to ensure some robustness to traffic variability for the decision. Using a chance constraint, an original model is given. It ensures that the computed solution will remain valid in the future despite traffic variations. Numerical tests on a France Télécom IP backbone are performed : they illustrate the interest of our model, especially when compared with more naive approaches. Then, Chapter 8 deals with planning future investments in an access network. Core sites have to be chosen to provide access to clients, while minimizing network costs (equipments). Uncertainty is considered on traffic forecasts, which are often far from reality in many cases. Chance-constrained models are given, and efficient heuristic algorithms are designed.

Notations

On récapitule les notations mathématiques les plus couramment utilisées dans ce document :

$n \in \mathbb{N}^*$, $m \in \mathbb{N}^*$

$I = \{1, \dots, n\}$, $J = \{1, \dots, m\}$

$x \in \mathbb{R}^n$ est le vecteur des variables

$A \in \mathbb{R}^{m \times n}$ est la matrice des contraintes (m lignes, n colonnes)

$b \in \mathbb{R}^m$ est le vecteur (colonne) des seconds membres

$r^+ = \max\{r, 0\}$, pour tout réel $r \in \mathbb{R}$

\mathbb{P} est la mesure de probabilité

\mathbb{E} est l'espérance

$\text{var}(X)$ désigne la variance d'une variable aléatoire X

Φ désigne la fonction de répartition de la loi normale standard $\mathcal{N}(0, 1)$

$\varepsilon \in]0, 1[$ est une tolérance sur la non-réalisabilité d'une solution

$\text{conv}(E)$ est l'enveloppe convexe de l'ensemble E

$\|v\|_p$ désigne la norme L_p de $v \in \mathbb{R}^n$, avec $p \in \mathbb{N}^*$: $\|v\|_p = (\sum_{i=1}^n |v_i|^p)^{1/p}$.

$\text{cl}(E)$ est la fermeture de l'ensemble E

$\mathbb{1}_E$ est la fonction caractéristique de l'ensemble E : $\mathbb{1}_E(x) = \begin{cases} 1 & \text{si } x \in E, \\ 0 & \text{sinon} \end{cases}$

C_n^p désigne le coefficient binomial des entiers naturels n et $p \leq n$: $C_n^p = \frac{n!}{p!(n-p)!}$

Pour toute paire de vecteurs u et v de même taille n , on note :

$u \cdot v = \sum_{i \in I} u_i v_i$ (produit scalaire euclidien canonique, parfois noté plus simplement encore uv)

$u \otimes v$ le vecteur $(u_i v_i)_{i \in I}$

$u \oslash v$ le vecteur $(u_i / v_i)_{i \in I}$

Enfin, on donne la signification de quelques abréviations courantes :

“s.c.” signifie “sous les contraintes”

“t.q.” signifie “tel que”

Chapitre 1

Préliminaires

1.1 Généralités sur l'optimisation en contexte incertain

Le plus souvent, les modèles et algorithmes d'aide à la décision supposent implicitement que les données d'entrée sont connues de manière exacte. Pourtant, la plupart des problèmes rencontrés en pratique peuvent difficilement être traités dans ce cadre. Le désir de tenir compte d'incertitudes sur les données d'entrée d'un problème d'optimisation n'est pas nouveau, on renvoie par exemple au travail de Dantzig effectué sur ce sujet en 1955 [28]. Il n'en est pas moins vrai que ce thème de recherche reste des plus actifs aujourd'hui encore, et a connu tout récemment un fort regain d'intérêt. Les avancées récentes ont ouvert des perspectives nouvelles qui devraient être fructueuses tant d'un point de vue théorique que pratique.

Derrière le concept de “donnée incertaine” se cachent en fait des réalités très diverses, issues de la grande variété des applications envisagées. Il faut sans doute renoncer à une typologie complète des incertitudes existant et des modèles mathématiques permettant d'en rendre compte. De nombreux éléments de la littérature proposent différentes classifications (voir e.g. [48, 72, 77]). Deux éléments fondamentaux semblent devoir être distingués : l'ensemble des événements pouvant survenir, et la caractérisation probabiliste de ces événements. Ce deuxième élément est un raffinement du premier. Il est des cas où seul l'ensemble des événements est connu de manière satisfaisante. Plus généralement, une caractérisation probabiliste rigoureusement exacte est souvent impossible à garantir, même si des approximations empiriques peuvent en pratique s'avérer satisfaisantes.

Exemple 1 : Si on utilise un instrument de mesure dont l'incertitude est de 10%, on en déduit facilement l'ensemble des valeurs que peut *a priori* prendre une donnée obtenue par son biais. Par contre, on ne connaît pas la loi de probabilité associée à l'intervalle de valeurs. Supposons maintenant que l'on dispose d'un historique de mesures suffisamment important, on pourrait en déduire une loi de probabilité empirique qui décrirait de façon satisfaisante le comportement de notre instrument de mesure.

Exemple 2 : Dans le cas où l'on cherche à pallier les pannes simples de liens dans un réseau, on peut aisément énumérer l'ensemble des événements pouvant survenir. La caractérisation probabiliste de ces pannes peut pourtant s'avérer difficile.

Si l'on ne connaît pas de caractérisation stochastique de l'incertitude, on se place dans le cadre de l'*optimisation robuste*. Dans ce cas, on cherche la meilleure solution qui sera réalisable pour l'ensemble d'événements envisagé. On effectue alors plutôt une optimisation prenant en compte le pire cas pouvant survenir. À l'inverse, si l'on a accès à des lois de probabilités, on

se trouve dans le contexte de la *programmation stochastique*. Ce domaine de la programmation mathématique a donné naissance à plusieurs modèles. Le plus couramment rencontré dans la littérature est la programmation multi-étapes avec recours, dont l’objectif est d’optimiser le coût en moyenne sur un ensemble de périodes successives.

Une courte parenthèse s’impose ici à propos d’une autre approche très courante de l’incertitude : l’*analyse de sensibilité*. Alors que la programmation stochastique et l’optimisation robuste incluent les incertitudes dans le processus de décision, l’analyse de sensibilité cherchera plutôt à analyser les propriétés d’une solution donnée. En particulier, elle s’attache à déterminer le domaine de validité d’une solution proposée. Ainsi, l’incertitude n’est pas à proprement parler prise en compte dans les algorithmes d’optimisation. Elle intervient plutôt dans une étude complémentaire post-optimisation. Puisque notre intérêt se concentre principalement dans cette thèse sur les problèmes en nombres entiers, on renvoie à [81] pour une analyse de la complexité de l’analyse de sensibilité pour les programmes 0-1.

1.1.1 Programmation stochastique

Dans ce cadre, on associe des lois de probabilités aux données d’entrée du problème d’optimisation. On est ainsi conduit à une modélisation fine de l’incertitude. Ce domaine de la programmation mathématique a bénéficié de nombreux travaux depuis une cinquantaine d’années. Pour plus de détails sur la programmation stochastique, voir par exemple [46, 16].

Quelques critiques peuvent être formulées à l’égard de l’optimisation stochastique (voir [48]). D’abord, comme déjà souligné, il peut être difficile voire impossible en pratique de déterminer les lois de probabilité à utiliser. On notera néanmoins que cet écueil potentiel a été pris en compte dans un certain nombre de travaux antérieurs, où une incertitude est aussi considérée sur les lois de probabilité utilisées. Une autre faiblesse concerne l’optimisation en moyenne d’un objectif donné, souvent préconisée (cf. la programmation multi-étapes avec recours). Si une telle approche peut parfois se justifier, la solution qui en résulte peut se révéler très mauvaise pour beaucoup de scénarios possibles. Une manière de pallier ce comportement indésirable est de recourir à un autre modèle de la programmation stochastique : la programmation sous contraintes probabilistes (*chance-constrained programming*). Le but est de trouver la meilleure solution dont on pourra garantir qu’elle est réalisable avec une probabilité cible donnée. Le coût de la solution résultante, peut-être moins satisfaisant en moyenne, sera cependant garanti avec une probabilité maîtrisée. Une telle approche s’avère complémentaire d’une optimisation en moyenne. En effet, considérons par exemple une problématique de planification d’investissements sur plusieurs années. S’il est important de garantir assez finement le coût à court terme, le manque de visibilité sur le long terme rend une optimisation en moyenne pertinente au-delà.

La programmation sous contraintes probabilistes est très pertinente pour beaucoup d’applications. Elle généralise notamment les concepts de *Value at Risk* (VaR) et de *Conditional Value at Risk* (CVaR, aussi appelée *expected shortfall*) (voir par exemple [67, 45] et l’annexe D), très utilisés en gestion du risque. Pourtant, elle a été relativement peu utilisée par le passé, sans doute du fait des difficultés théoriques qui s’y attachent. Cette dernière remarque nous mène à notre dernière observation concernant la programmation stochastique. Souvent, les techniques de résolution se révèlent complexes et difficiles à mettre en œuvre de manière satisfaisante sur des problèmes de grandes tailles. En particulier, la prise en compte de variables entières dans des programmes stochastiques a été plutôt moins étudiée. Pourtant, de nombreux problèmes de décision sont par nature combinatoires.

1.1.2 Optimisation robuste

Quand aucune information probabiliste n'est disponible, l'incertitude est caractérisée par la seule connaissance d'un ensemble d'événements pouvant survenir. On distingue souvent (voir e.g. [48, 77]) :

- l'optimisation de la valeur dans le pire cas : la solution robuste calculée doit par exemple minimiser le coût maximal engendré par les scénarios envisagés. On parle alors généralement de problème de "coût minmax".
- la minimisation de l'écart maximal avec la solution optimale de chacun des scénarios envisagés. Ceci n'est *a priori* réalisable que si ces scénarios sont en nombre fini (et restreint). On parle alors de problèmes de "regret minmax".

Un des intérêts distinctifs de l'optimisation robuste est de pouvoir être utilisée, bien souvent, pour traiter des problèmes combinatoires. [59] propose d'ailleurs une bibliographie des usages de l'optimisation robuste pour de tels problèmes.

L'optimisation robuste souffre aussi de plusieurs faiblesses. D'abord, classiquement, elle ne permet pas la modélisation fine des incertitudes par des lois de probabilité. Mais les reproches les plus couramment faits à l'encontre de l'optimisation robuste concernent le "conservatisme" des solutions obtenues (*over-conservatism*). En effet, une optimisation au pire cas risque d'induire un coût élevé, alors que le pire cas en question n'a peut-être qu'une probabilité très faible de se produire. Cet équilibre entre la valeur de l'objectif et la réalisabilité de la solution est un des aspects les plus cruciaux de l'optimisation avec données incertaines. De nombreux travaux ont été motivés par cette question, et plusieurs réponses ont été apportées. Par exemple, [55] introduit deux types de robustesse : pour les auteurs, une solution est dite *solution-robuste* si elle garde une valeur d'objectif assez bonne pour tout événement. D'autre part, une solution est dite *modèle-robuste* si elle demeure réalisable dans presque tous les scénarios. Plus récemment, plusieurs publications ont proposé des modèles robustes relâchés, qui permettent un meilleur compromis entre la valeur d'une solution et sa réalisabilité, comme par exemple [33, 9, 13].

1.2 Quelques modèles

Afin de clarifier certaines idées précédemment évoquées, quelques modèles mathématiques très classiques sont ici décrits succinctement.

1.2.1 Optimisation robuste à partir de scénarios

Il est fréquent que les modèles d'optimisation robuste s'appuient sur un ensemble discret fini de scénarios possibles. Kouvelis et Yu [48] ont largement étudié cette famille de modèles ; on présente rapidement ci-dessous le cadre de leur étude. Le but est, à partir de cette approche simple, de mettre en évidence les questions majeures posées par l'optimisation en contexte incertain.

Soit S l'ensemble discret fini des scénarios. Pour tout scénario $s \in S$, on note f_s la fonction objectif à minimiser. Par ailleurs, X_s désigne l'ensemble des solutions admissibles pour le scénario s . Ainsi, $X(S) = \bigcap_{s \in S} X_s$ est l'ensemble des solutions réalisables pour tout scénario. Alors, le modèle de coût minmax associé s'écrit :

$$\min_{x \in X(S)} \max_{s \in S} f_s(x).$$

De plus, on note x_s^* la solution optimale pour le scénario s : $x_s^* \in \arg \min_{x \in X_s} f_s(x)$. Alors

le modèle de regret minmax est :

$$\min_{x \in X(S)} \max_{s \in S} \{f_s(x) - f_s(x_s^*)\}.$$

Ces modèles conduisent à des solutions très robustes, dans le sens où elles seront réalisables quel que soit le scénario se produisant (cf. $x \in X(S)$). On parle de *conservatisme* de la solution. Une telle propriété n'est pas toujours pertinente, dans la mesure où certains scénarios peuvent en pratique être bien moins probables que d'autres. Est-il alors utile de payer un éventuel sur-coût pour tenir compte d'un événement qui n'a que très peu de chances d'arriver ?

La modélisation proposée par Mulvey, Vanderbei et Zenios [55] repose elle aussi sur un ensemble fini de scénarios. Elle diffère cependant du travail de [48] par l'introduction de notions probabilistes simples. Les auteurs attribuent à chaque scénario s une probabilité p_s d'occurrence, et s'intéressent plus spécifiquement à l'optimisation de l'espérance du coût (*i.e.* $\min_x \sum_{s \in S} p_s f_s(x)$). La question du conservatisme de la solution est traitée, et les auteurs cherchent à relâcher cette propriété. Pour cela, ils ont recours à une fonction de pénalité sur la non-réalisabilité éventuelle d'une solution. Un des intérêts de [55] est de mettre très clairement en évidence les deux objectifs contradictoires intervenant souvent dans l'optimisation en contexte incertain. D'un côté, on cherche une solution "souvent" réalisable (robustesse ou conservatisme de la solution). Mais d'un autre côté, on veut aussi une solution de coût aussi faible que possible. Pourtant, la réponse apportée demeure partielle et assez largement insatisfaisante. La définition de la fonction de pénalité sur la non-réalisabilité pose question. De plus, les auteurs ont finalement recours à une combinaison linéaire des deux objectifs (coût et pénalité) : si cette modélisation a le mérite de la simplicité, la manière de choisir les coefficients de la combinaison linéaire reste problématique. Il serait peut-être plus satisfaisant d'avoir recours à des approches purement bi-objectifs conduisant à une description du front de Pareto associé.

Les approches de [48] et [55] sont bien adaptées quand les scénarios peuvent être facilement définis et énumérés. Pourtant, c'est loin d'être toujours le cas. Pour de nombreuses applications, il n'y a souvent pas de sens immédiat à vouloir définir un ensemble fini de scénarios possibles. Cela peut arriver notamment quand les grandeurs incertaines sont par nature continues (par exemple le trafic dans un réseau de télécommunications). Une idée qui peut sembler naturelle est d'avoir recours à des méthodes d'échantillonnage de type Monte-Carlo. Mais il faut s'interroger sur le sens théorique d'une telle démarche, qui demande à être clarifié. Par ailleurs, d'une manière générale, l'obtention de garanties théoriques à partir de méthodes d'échantillonnage conduit à la définition d'un très grand nombre de scénarios : ceci rend très difficile l'utilisation directe de modèles tels que ceux de [48]. Une discussion sur les inconvénients liés à certaines approches s'appuyant sur des méthodes d'échantillonnage peut être trouvée dans [58] (voir aussi [7] pour des observations similaires).

Finalement, malgré l'apparente simplicité des modèles classiques de [48], des difficultés théoriques et algorithmiques existent. En effet, beaucoup de problèmes robustes ainsi modélisés ont une complexité bien plus importante que leur correspondant classique (non-robuste). Parmi plusieurs exemples détaillés dans [48], les auteurs montrent en particulier que le problème de plus court chemin robuste, quand les poids sur les arcs sont incertains, est NP-difficile même si le nombre de scénarios est restreint à 2 ($|S| = 2$). Cette non-conservation de la complexité du problème initial, surprenante de prime abord, conduit à des difficultés importantes de mise en œuvre d'une telle méthodologie robuste en pratique.

Ainsi, cette brève réflexion sur les modèles étudiés dans [48] a mis en évidence les questions centrales suivantes :

1. comment gérer le compromis entre réalisabilité et coût d'une solution robuste ?
2. Comment définir l'ensemble des événements auxquels une solution devrait être robuste ?
3. Peut-on trouver une méthodologie d'optimisation en contexte incertain garantissant une complexité théorique et pratique modérée ?

Aucune approche ne semble pouvoir répondre de manière satisfaisante à l'ensemble de ces questions pour tous les domaines d'application. Il est néanmoins important de les garder à l'esprit pour définir des démarches pertinentes.

1.2.2 Un cadre plus général

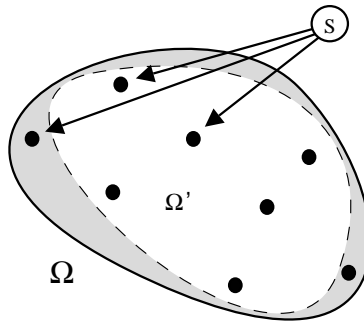
L'extension naturelle des approches précédentes consiste à travailler avec un ensemble général Ω d'événements (non-discret *a priori*). À chaque événement $\omega \in \Omega$, on associe l'ensemble X_ω des solutions réalisables. Si l'on note $X(\Omega) = \bigcap_{\omega \in \Omega} X_\omega$, le modèle robuste de coût minmax est :

$$\min_{x \in X(\Omega)} \max_{\omega \in \Omega} f(x, \omega)$$

où $f(x, \omega)$ est de coût de la solution x si l'événement ω survient. Il faut remarquer qu'une telle généralisation rend difficile une approche de type regret minmax. Comme précédemment, le problème du conservatisme de la solution ainsi calculée se pose. On souhaite toujours obtenir une solution qui (i) soit de coût "assez" faible, et (ii) soit réalisable pour "la plupart" des éventualités envisagées. Une manière de faire est de s'intéresser seulement à un sous-ensemble assez grand d'événements $\Omega' \subset \Omega$, et de résoudre le problème robuste associé :

$$\min_{x \in X(\Omega')} \max_{\omega \in \Omega'} f(x, \omega)$$

où $X(\Omega') = \bigcap_{\omega \in \Omega'} X_\omega$. La figure ci-dessous donne une représentation des ensembles d'incertitude Ω et Ω' , ainsi que de S tel qu'utilisé précédemment.



La difficulté fondamentale consiste à bien définir l'ensemble d'incertitude Ω' . Une démarche naturelle pour mesurer la réalisabilité d'une solution consiste à introduire une mesure \mathbb{P} de probabilité associée à Ω ($\mathbb{P}(\Omega) = 1$). On veut alors Ω' assez grand, par exemple : $\mathbb{P}(\Omega') \geq 1 - \varepsilon$, avec $\varepsilon \in]0, 1[$ donné, généralement petit. Ce sous-ensemble doit également être choisi de manière à entraîner un coût robuste associé aussi faible que possible. Le problème suivant est donc particulièrement intéressant :

$$\min_{\Omega' \subset \Omega : \mathbb{P}(\Omega') \geq 1 - \varepsilon} \left\{ \min_{x \in X(\Omega')} \max_{\omega \in \Omega'} f(x, \omega) \right\}. \quad (1.1)$$

Travailler avec un sous-ensemble d'événements $\Omega' \subset \Omega$ est une idée qui a été récemment beaucoup utilisée, voir par exemple [33, 9, 13]. Chaque contribution présente une manière spécifique de définir le sous-ensemble Ω' . Dans le cadre de la programmation linéaire robuste, [11] procure une vision unifiée de la plupart de ces travaux.

1.2.3 Liens avec la programmation sous contraintes probabilistes

Les approches robustes évoquées ci-dessus sont en lien direct avec la programmation sous contraintes probabilistes (*chance-constrained programming*, voir e.g. [66, 16, 46, 67, 45]). Avec nos notations, ce modèle peut s'écrire :

$$\min_{x \in X, z \in \mathbb{R}} \left\{ z \mid \mathbb{P}(z \geq f(x, \omega) \text{ et } x \in X_\omega) \geq 1 - \varepsilon \right\}. \quad (1.2)$$

Cette approche de l'incertitude a été introduite en 1959 par Charnes et Cooper [23]. En toute généralité, ce modèle donne lieu à des programmes mathématiques complexes. Une des sources majeures de difficulté est la non-convexité potentielle de l'ensemble des solutions réalisables. Beaucoup de contributions ont donné, jusque récemment, des conditions dans lesquelles cette convexité est assurée. Néanmoins, des exemples simples et frappants révèlent que l'ensemble des solutions peut être particulièrement difficile à manipuler.

Exemple : Considérons une variable aléatoire (a, b) telle que : $\mathbb{P}((a, b) = (1, 1)) = 1/2$ et $\mathbb{P}((a, b) = (-1, 0)) = 1/2$. Soit $X = \{x \in \mathbb{R} \mid \mathbb{P}(ax \geq b) \geq 1/2\}$. On a : $X =]-\infty, 0] \cup [1, +\infty[$. On voit que X n'est même pas connexe.

On renvoie à [45] pour un récapitulatif de cas particuliers de programmes sous contraintes probabilistes connus pour être convexes. On notera que, classiquement, les travaux sur la programmation sous contraintes probabilistes se sont essentiellement intéressés à des problèmes linéaires dont les coefficients sont des variables aléatoires normalement distribuées, et le plus souvent indépendantes [75, 74, 84]. En effet, on peut dans ce cas écrire un équivalent déterministe du problème.

Le modèle (1.1) apparaît en fait comme une ré-écriture équivalente du problème sous contrainte probabiliste (1.2) :

Lemme 1 *Il existe z tel que (x, z) est une solution optimale de (1.2) si, et seulement si, il existe $\Omega' \subseteq \Omega$ tel que (Ω', x) est une solution optimale de (1.1).*

Preuve : (1.1) peut s'écrire de manière équivalente comme :

$$\min_{\Omega' \subseteq \Omega : \mathbb{P}(\Omega') \geq 1 - \varepsilon, x \in X(\Omega'), z \in \mathbb{R}} \left\{ z \mid z \geq \max_{\omega \in \Omega'} f(x, \omega) \right\} \quad (1.3)$$

Montrons que (x, z) est solution réalisable de (1.2) si et seulement si (Ω', x, z) est solution réalisable de (1.3). Soit (x, z) solution réalisable de (1.2). Soit $\Omega' = \{\omega \in \Omega \mid z \geq f(x, \omega) \text{ et } x \in X_\omega\}$. Cette définition implique que $x \in X(\Omega')$. On sait également que $\mathbb{P}(\Omega') \geq 1 - \varepsilon$. On en déduit que (Ω', x, z) est une solution réalisable de (1.3).

Réciproquement, soit (Ω', x, z) une solution réalisable de (1.3). Alors :

$$\begin{aligned} \mathbb{P}(z \geq f(x, \omega) \text{ et } x \in X_\omega) &\geq \mathbb{P}(z \geq f(x, \omega) \text{ et } x \in X_\omega \text{ et } \omega \in \Omega') \\ &= \mathbb{P}(x \in X_\omega \text{ et } \omega \in \Omega') \\ &= \mathbb{P}(\omega \in \Omega'), \text{ puisque } x \in X(\Omega') \end{aligned}$$

On en déduit : $\mathbb{P}(z \geq f(x, \omega) \text{ et } x \in X_\omega) \geq 1 - \varepsilon$. Ceci montre que (x, z) est une solution réalisable de (1.2).

Comme chacun des deux problèmes (1.2) et (1.3) impliquent la même fonction objectif, on en déduit la correspondance des solutions optimales entre les deux problèmes. Le résultat avec

(1.1) s'ensuit immédiatement. \square

La programmation sous contrainte probabiliste est un modèle classique du domaine de la programmation stochastique. Les quelques observations ci-dessus montrent les liens qui relient ce modèle au domaine de l'optimisation robuste. Alors que la programmation sous contrainte probabiliste est connue pour être particulièrement difficile dans le cas général, on peut espérer avoir recours à l'optimisation robuste pour approcher ce problème de manière efficace.

Le lien entre modèles robustes et stochastiques a été souligné et utilisé de diverses manières dans plusieurs publications récentes [8, 9, 13, 24, 58].

Première partie

**Approches robustes pour les
problèmes sous contraintes
probabilistes**

Chapitre 2

Sous-ensembles d'événements pour l'optimisation robuste

Motivations : On a montré dans le chapitre précédent le lien qui existe entre contraintes probabilistes et optimisation robuste. La logique de cette dernière approche est de trouver la meilleure solution valide pour un ensemble d'événements donné. Idéalement, on souhaiterait définir cet ensemble d'événements de manière à résoudre, ou au moins approcher, un problème d'optimisation sous contraintes probabilistes. Cette réflexion se poursuit dans le présent chapitre. On se concentre sur la manière de définir les ensembles d'événements, et l'impact que cela a sur l'ensemble des solutions robustes associé.

Étant donné un ensemble Ω d'événements possibles, la question principale qui nous intéresse ici est la caractérisation de bons sous-ensembles $\Omega' \subset \Omega$ pour résoudre le problème robuste associé :

$$\min_{x \in X(\Omega')} \max_{\omega \in \Omega'} f(x, \omega).$$

On rappelle que $X(\Omega')$ est l'ensemble des solutions réalisables pour tout événement de Ω' . Ω' doit être assez grand pour assurer une bonne réalisabilité de la solution robuste correspondante. Mais cet ensemble doit aussi permettre un gain au niveau du pire coût de la solution : $\max_{\omega \in \Omega'} f(x, \omega)$. Ces deux objectifs sont bien sûr contradictoires, et nécessitent un compromis.

2.1 Sous-ensembles optimaux d'événements

En lien avec la programmation sous contraintes probabilistes, on rappelle que le meilleur sous-ensemble $\Omega' \subset \Omega$ est solution du problème d'optimisation suivant (voir § 1.2.2 et 1.2.3) :

$$\min_{\Omega' : \mathbb{P}(\Omega') \geq 1 - \varepsilon} \left\{ \min_{x \in X(\Omega')} \max_{\omega \in \Omega'} f(x, \omega) \right\}. \quad (2.1)$$

Sans perte de généralité, on suppose que les ensembles de solution $\{X_\omega\}_{\omega \in \Omega}$ s'écrivent à l'aide d'une fonction scalaire $g : \mathbb{R}^n \times \Omega \rightarrow \mathbb{R} : X_\omega = \{x \in \mathbb{R}^n | g(x, \omega) \geq 0\}$. On observera que le cadre considéré est extrêmement général. Il inclut en particulier les cas de programmation avec variables entières, puisqu'aucune hypothèse spécifique sur g n'est imposée. On pourrait écrire indifféremment : $X_\omega = \{x \in \mathcal{X} | g(x, \omega) \geq 0\}$, où $\mathcal{X} \subseteq \mathbb{Z}^r \times \mathbb{R}^{n-r}$ ($r \leq n$). D'autre part,

le plus souvent dans la pratique, les problèmes d'optimisation sont écrits à l'aide de plusieurs contraintes scalaires : $g_1(x, \omega) \geq 0, \dots, g_m(x, \omega) \geq 0$. On se ramène facilement à notre cadre en écrivant $g(x, \omega) = \min\{g_1(x, \omega), \dots, g_m(x, \omega)\}$. Observer que certaines propriétés des g_j peuvent rester valables pour g . Par exemple, si les $g_j(x, \cdot)$ sont concaves, alors $g(x, \cdot)$ est aussi concave (se souvenir que l'intersection d'ensembles convexes est convexe).

En conformité avec les notations utilisées jusqu'ici, pour tout sous-ensemble $F \subseteq \Omega$, $X(F) = \bigcap_{\omega \in F} X_\omega$ est l'ensemble des solutions réalisables pour tout événement de F . Notons également : $X = \bigcup_{\omega \in \Omega} X_\omega$. Quelques observations simples peuvent être faites pour se familiariser avec le rapport entre un ensemble d'événements et l'ensemble des solutions robustes correspondant :

Lemme 2 Soit $F_1 \subseteq \Omega$ et $F_2 \subseteq \Omega$:

- (i) $F_1 \subseteq F_2 \Rightarrow X(F_1) \supseteq X(F_2)$
- (ii) $X(F_1 \cup F_2) = X(F_1) \cap X(F_2)$
- (iii) $X(F_1 \cap F_2) \supseteq X(F_1) \cup X(F_2)$

Preuve : (i) est évident. Pour (ii), observer que (i) implique : $X(F_1 \cup F_2) \subseteq X(F_1)$ et $X(F_1 \cup F_2) \subseteq X(F_2)$, et donc : $X(F_1 \cup F_2) \subseteq X(F_1) \cap X(F_2)$. Alors, pour tout $x \in X(F_1) \cap X(F_2)$, on vérifie aisément que $x \in X(F_1 \cup F_2)$. Pour (iii), il suffit de voir qu'une solution réalisable pour tout F_1 (resp. F_2) est aussi réalisable pour $F_1 \cap F_2$. Donc $X(F_1) \subseteq X(F_1 \cap F_2)$ et $X(F_2) \subseteq X(F_1 \cap F_2)$. \square

Pour le point (iii), seule l'inclusion donnée est vraie en général. La plupart du temps : $X(F_1 \cap F_2) \neq X(F_1) \cup X(F_2)$. En effet, considérons par exemple que Ω est composé de trois scénarios : $\Omega = \{\omega_1, \omega_2, \omega_3\}$. Supposons que $F_1 = \{\omega_1, \omega_2\}$, et $F_2 = \{\omega_2, \omega_3\}$. On a : $F_1 \cap F_2 = \{\omega_2\}$. On voit que si une solution x est réalisable pour F_1 ou pour F_2 , alors elle l'est pour le scénario ω_2 , intersection de F_1 et F_2 . En revanche, une solution réalisable pour le seul scénario ω_2 n'est pas forcément réalisable aussi pour ω_1 , et donc pour F_1 .

Lemme 3 Supposons que le problème sous contrainte probabiliste (1.2) admet une solution optimale. Il existe $x^* \in X$ et $v \in \mathbb{R}$ tels que $\Omega' = \{\omega \in \Omega \mid f(x^*, \omega) \leq v \text{ et } g(x^*, \omega) \geq 0\}$ est une solution optimale de (2.1).

Preuve : D'après le lemme 1, si (1.2) admet une solution optimale, alors (2.1) également. Soit (Ω^*, x^*) une solution optimale de (2.1). On note v la valeur optimale associée. On vérifie aisément que $\Omega^* \subseteq \Omega'$. On en déduit que $\mathbb{P}(\Omega') \geq 1 - \varepsilon$. Comme de plus $x^* \in X(\Omega')$, (Ω', x^*) est une solution réalisable pour (2.1). D'autre part : $\max_{\omega \in \Omega'} f(x^*, \omega) \leq v = \min_{x \in X(\Omega^*)} \max_{\omega \in \Omega^*} f(x, \omega)$. Donc, puisque $x^* \in X(\Omega')$: $\min_{x \in X(\Omega')} \max_{\omega \in \Omega'} f(x, \omega) \leq v = \min_{x \in X(\Omega^*)} \max_{\omega \in \Omega^*} f(x, \omega)$. Ceci assure que (Ω', x^*) est une solution optimale de (2.1). \square

Ce résultat simple fournit des informations intéressantes quant à la forme d'un ensemble optimal d'événements. Si par exemple, pour tout $x \in X$, l'ensemble $\{\omega \in \Omega \mid g(x, \omega) \geq 0\}$ est un polyèdre, alors on sait qu'il existe une solution polyédrique de (2.1).

Pour résoudre le problème (2.1) (équivalent à un programme sous contrainte probabiliste), une idée naturelle consiste à faire "grandir" progressivement l'ensemble d'événements Ω' en espérant converger vers un ensemble optimal. Ce sera la stratégie appliquée dans les chapitres 4 et 5. Néanmoins, on montre ci-dessous qu'une telle approche peut malheureusement converger vers un ensemble non-optimal.

On introduit le concept de solution *maximale* de (2.1) :

Définition 1 Un sous-ensemble $F \subseteq \Omega$ est dit solution maximale de (2.1) si :

- (i) il existe une solution x^* au problème : $\min_{x \in X(F)} \max_{\omega \in F} f(x, \omega)$;
(ii) pour tout autre sous-ensemble $F' \supset F$:

$$\min_{x \in X(F')} \max_{\omega \in F'} f(x, \omega) > \min_{x \in X(F)} \max_{\omega \in F} f(x, \omega).$$

Ainsi, F est dit maximal s'il ne peut pas être agrandi sans augmenter le coût de la solution robuste associée. Cette caractéristique sera en particulier vérifiée par l'ensemble-limite de la stratégie de résolution évoquée plus haut. Deux observations permettent de clarifier les relations entre solution optimale et solution maximale :

- dans de nombreux cas, il existe une solution optimale de (2.1) qui est maximale. C'est clair par exemple lorsque Ω est fini. Si Ω n'est pas fini, mais s'il existe une solution x^* unique au problème sous contrainte probabiliste, l'affirmation reste valide. En effet, il suffit alors de considérer l'ensemble Ω' décrit dans le lemme 3. Si cet ensemble pouvait encore être étendu, il serait associé à une nouvelle solution optimale x' .
- En revanche, une solution maximale de probabilité $1 - \varepsilon$ n'est pas nécessairement optimale pour (2.1), même dans des cas très simples.

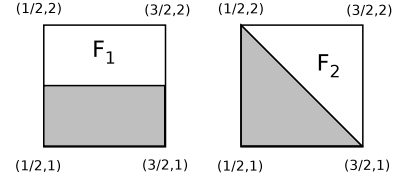
L'exemple suivant illustre cette dernière affirmation.

Exemple : Considérons le problème robuste suivant :

$$\begin{aligned} \min \quad & x_1 + 2x_2 \\ \text{s.c.} \quad & a_1x_1 + a_2x_2 \geq 1, \\ & x \geq 0. \end{aligned}$$

Les coefficients a_1 et a_2 sont incertains, et appartiennent respectivement aux intervalles $[1/2, 3/2]$ et $[1, 2]$. Ainsi, $\Omega = [1/2, 3/2] \times [1, 2]$. Supposons que chaque coefficient est une variable aléatoire uniformément distribué sur son intervalle de définition. Soit $\varepsilon = 1/2$. On introduit :

$$\begin{cases} F_1 = \{a \in \Omega \mid a_2 \geq 3/2\} \\ F_2 = \{a \in \Omega \mid a_1 + a_2 \geq 5/2\} \end{cases}$$



Observer que : $\mathbb{P}(F_1) = \mathbb{P}(F_2) = 1/2 = 1 - \varepsilon$. Soit $V(F_1)$ (resp. $V(F_2)$) la valeur de l'objectif pour la meilleure solution robuste à tous les événements de F_1 (resp. F_2).

- Pour tout $a \in F_1$, observer que : $ax \geq 1/2x_1 + 3/2x_2$, puisque $a_1 \geq 1/2$ et $a_2 \geq 3/2$. Ainsi, x est une solution robuste à tous les événements de F_1 si, et seulement si : $1/2x_1 + 3/2x_2 \geq 1$. En conséquence : $V(F_1) = \min\{x_1 + 2x_2 \mid 1/2x_1 + 3/2x_2 \geq 1, x \geq 0\} = 4/3$.

- Montrons que F_1 est maximal. Considérons un quelconque $\tilde{a} \in \Omega \setminus F_1$, i.e. : $\tilde{a} \in [1/2, 3/2] \times [1, 3/2[$. On note $\tilde{F}_1 = F_1 \cup \{\tilde{a}\}$, et on a : $V(\tilde{F}_1) = \min\{x_1 + 2x_2 \mid 1/2x_1 + 3/2x_2 \geq 1, \tilde{a}x \geq 1, x \geq 0\}$. Observer que : $\tilde{a}x \leq 3/2x_1 + \tilde{a}_2x_2$. On en déduit : $V(\tilde{F}_1) \geq \min\{x_1 + 2x_2 \mid 1/2x_1 + 3/2x_2 \geq 1, 3/2x_1 + \tilde{a}_2x_2 \geq 1, x \geq 0\} = V'$. Soit \tilde{x} le point d'intersection des droites $\{x \in \mathbb{R} \mid 1/2x_1 + 3/2x_2 = 1\}$ et $\{x \in \mathbb{R} \mid 3/2x_1 + \tilde{a}_2x_2 = 1\}$, on peut vérifier que : $V' = \tilde{x}_1 + 2\tilde{x}_2$.

Par ailleurs, des calculs faciles montrent que : $\tilde{x}_1 = (6 - 4\tilde{a}_2)/(9 - 2\tilde{a}_2)$ et $\tilde{x}_2 = 4/(9 - 2\tilde{a}_2)$. En conséquence : $V' = 2 - 4/(9 - 2\tilde{a}_2)$. Comme on a supposé $\tilde{a}_2 < 3/2$: $V' > 4/3 = V(F_1)$. Ceci assure que : $V(\tilde{F}_1) > V(F_1)$. Ainsi, on a montré que F_1 ne pouvait pas être étendu sans détériorer la valeur de la solution robuste associée : F_1 est maximal.

- Finalement, montrons que F_1 n'est pas optimal. Ceci se fait par comparaison avec F_2 . Pour calculer $V(F_2)$, on écrit : $V(F_2) = \min\{x_1 + 2x_2 \mid \forall a \in \text{conv}\{a^{(1)}, a^{(2)}, a^{(3)}\} : ax \geq 1, x \geq 0\}$, où $a^{(1)} = (1/2, 2)$, $a^{(2)} = (3/2, 1)$, $a^{(3)} = (3/2, 2)$. Il est clair que : $V(F_2) \geq \min\{x_1 + 2x_2 \mid a^{(1)}x \geq$

$1, a^{(2)}x \geq 1, a^{(3)}x \geq 1, x \geq 0\}$. Mais on montre aussi facilement que si x satisfait $a^{(1)}x \geq 1, a^{(2)}x \geq 1$ et $a^{(3)}x \geq 1$, alors x satisfait $ax \geq 1$ pour tout $a \in \text{conv}\{a^{(1)}, a^{(2)}, a^{(3)}\}$. Ceci assure : $V(F_2) \leq \min\{x_1 + 2x_2 \mid a^{(1)}x \geq 1, a^{(2)}x \geq 1, a^{(3)}x \geq 1, x \geq 0\}$.

En définitive : $V(F_2) = \min\{x_1 + 2x_2 \mid a^{(1)}x \geq 1, a^{(2)}x \geq 1, a^{(3)}x \geq 1, x \geq 0\} = \min\{x_1 + 2x_2 \mid 1/2x_1 + 2x_2 \geq 1, 3/2x_1 + x_2 \geq 1, 3/2x_1 + 2x_2 \geq 1, x \geq 0\}$. La résolution graphique de ce programme linéaire permet de vérifier que : $V(F_2) = 6/5 < V(F_1)$.

2.2 Propriétés spécifiques dans un contexte convexe

On suppose dans cette partie que Ω est un ensemble convexe.

2.2.1 Ensemble convexe d'événements

Lemme 4 *Supposons que pour tout $x \in X$, $g(x, \cdot)$ est concave sur Ω . Soit $F \subseteq \Omega$, on a : $X(F) = X(\text{conv}(F))$.*

Preuve : Comme $\text{conv}(F) \supset F$, on a : $X(\text{conv}(F)) \subseteq X(F)$ (cf. lemme 2). De plus, soit $x \in X(F)$. Pour tout $(\omega_1, \omega_2) \in F^2$, et pour tout $\lambda \in [0, 1]$, par concavité de $g(x, \cdot)$: $g(x, \lambda\omega_1 + (1 - \lambda)\omega_2) \geq \lambda g(x, \omega_1) + (1 - \lambda)g(x, \omega_2) \geq 0$. Ceci montre que $x \in X(\text{conv}(F))$. Alors : $X(F) \subseteq X(\text{conv}(F))$, et donc : $X(F) = X(\text{conv}(F))$. \square

Proposition 1 *Supposons que pour tout $x \in X$, $g(x, \cdot)$ est concave sur Ω , et $f(x, \cdot)$ est convexe sur Ω . Alors, il existe une solution optimale Ω' de (2.1) qui est convexe.*

Preuve : Ce résultat peut être vu comme une conséquence du lemme 3. On se propose cependant de montrer plus directement que toute solution peut être remplacée par son enveloppe convexe.

Supposons que Ω' est une solution optimale non-convexe de (2.1), on montre que $\text{conv}(\Omega')$ est aussi une solution optimale de (2.1). On a de manière évidente : $\mathbb{P}(\text{conv}(\Omega')) \geq \mathbb{P}(\Omega') \geq 1 - \varepsilon$. De plus, d'après le lemme 4 : $X(\Omega') = X(\text{conv}(\Omega'))$. Finalement, soit $x \in X(\Omega')$. Par convexité de $f(x, \cdot)$, on voit facilement que : $\max_{\text{conv}(\Omega')} f(x, \cdot) \leq \max_{\Omega'} f(x, \cdot)$. Par ailleurs, puisque $\Omega' \subset \text{conv}(\Omega')$: $\max_{\Omega'} f(x, \cdot) \leq \max_{\text{conv}(\Omega')} f(x, \cdot)$. Ainsi : $\max_{\Omega'} f(x, \cdot) = \max_{\text{conv}(\Omega')} f(x, \cdot)$. On a donc prouvé que : $\min_{x \in X(\text{conv}(\Omega'))} \max_{\omega \in \text{conv}(\Omega')} f(x, \omega) = \min_{x \in X(\Omega')} \max_{\omega \in \Omega'} f(x, \omega)$. \square

Cette observation montre qu'il est inutile de travailler avec des ensembles d'événements non-convexes lorsque les fonctions f et g ont les propriétés décrites. De plus, ce résultat explicite le fait que les problèmes sous contraintes probabilistes peuvent en théorie être remplacés de manière équivalente par un problème robuste convexe. Pourtant, il est bien connu que ces problèmes sous contraintes probabilistes peuvent être non-convexes (voir l'exemple du paragraphe 1.2.3). Par ailleurs, noter que dans la proposition précédente, les contraintes de convexité ne concernent que la forme des ensembles d'événements réalisables pour une solution donnée (les fonctions $g(x, \cdot)$ et $f(x, \cdot)$). Le résultat reste donc valide quelle que soit la nature de $g(\cdot, \omega)$ et $f(\cdot, \omega)$. Ceci concerne par exemple l'ensemble auquel appartient la variable x , qui peut être (partiellement) discret (cf. programmation avec nombres entiers).

On a donc montré que, dans bon nombre de cas pratiques, on pouvait se restreindre à travailler avec des sous-ensembles d'événements $\Omega' \subseteq \Omega$ convexes. Dans ce contexte, un certain nombre de propriétés peuvent permettre de simplifier grandement la résolution pratique du problème robuste associé à Ω' .

On suppose à partir de maintenant que Ω est inclus dans un espace vectoriel de dimension finie d .

Définition 2 Soit C un ensemble convexe, $c \in C$ est appelé point extrême de C si c ne peut pas être écrit comme la combinaison convexe de deux autres points de C . En d'autres termes, pour tout $(c_1, c_2) \in C^2$ et pour tout $\lambda \in]0, 1[$: $c = \lambda c_1 + (1 - \lambda)c_2 \Leftrightarrow c_1 = c_2 = c$.

On notera $\mathcal{G}(C)$ l'ensemble des points extrêmes de C . On rappelle ce résultat classique de la théorie des ensembles convexes (voir par exemple [69], corollaire 18.5.1) :

Théorème 1 Un ensemble C convexe, fermé et borné est l'enveloppe convexe de ses points extrêmes : $C = \text{conv}(\mathcal{G}(C))$.

Lemme 5 Supposons que pour tout $x \in X$, $g(x, \cdot)$ est concave sur Ω . Soit $\Omega' \subset \Omega$ fermé, convexe et borné : $X(\mathcal{G}(\Omega')) = X(\Omega')$.

Preuve : Comme $\mathcal{G}(\Omega') \subseteq \Omega'$, on a : $X(\mathcal{G}(\Omega')) \supseteq X(\Omega')$. Maintenant, considérons $x \in X(\mathcal{G}(\Omega'))$. Soit $\omega \in \Omega'$. D'après le théorème de Carathéodory, il existe $d+1$ points extrêmes $\{\omega_k\}_{k \in \{1, \dots, d+1\}}$ dans $\mathcal{G}(\Omega')$, et $\lambda \in \mathbb{R}_+^{d+1}$ tel que $\sum_{k=1}^{d+1} \lambda_k = 1$ et : $\omega = \sum_{k=1}^{d+1} \lambda_k \omega_k$. Alors, par concavité de $g(x, \cdot)$: $g(x, \omega) \geq \sum_{k=1}^{d+1} \lambda_k g(x, \omega_k) \geq 0$, puisque $x \in X(\mathcal{G}(\Omega'))$. Ceci prouve que $x \in X_\omega$. Comme ceci est vrai pour tout $\omega \in \Omega'$: $x \in X(\Omega')$. Ainsi, on a montré que $X(\mathcal{G}(\Omega')) \subseteq X(\Omega')$. \square

On peut alors montrer :

Proposition 2 Supposons que pour tout $x \in X$, $g(x, \cdot)$ est concave sur Ω , et $f(x, \cdot)$ est convexe sur Ω . Soit $\Omega' \subset \Omega$ fermé, convexe et borné, on a :

$$\min_{x \in X(\Omega')} \max_{\omega \in \Omega'} f(x, \omega) = \min_{x \in X(\mathcal{G}(\Omega'))} \max_{\omega \in \mathcal{G}(\Omega')} f(x, \omega).$$

Preuve : D'après le lemme 5, on a : $X(\mathcal{G}(\Omega')) = X(\Omega')$. Soit $x \in X(\Omega')$, on montre que $\max_{\Omega'} f(x, \cdot) = \max_{\mathcal{G}(\Omega')} f(x, \cdot)$. Puisque $\mathcal{G}(\Omega') \subset \Omega'$: $\max_{\Omega'} f(x, \cdot) \geq \max_{\mathcal{G}(\Omega')} f(x, \cdot)$. Soit $\omega \in \Omega'$, il existe $d+1$ points extrêmes $\{\omega_k\}_{k \in \{1, \dots, d+1\}}$ dans $\mathcal{G}(\Omega')$, et $\lambda \in \mathbb{R}_+^{d+1}$ tel que $\sum_{k=1}^{d+1} \lambda_k = 1$ et : $\omega = \sum_{k=1}^{d+1} \lambda_k \omega_k$ (cf. théorème de Carathéodory). Alors, par convexité de $f(x, \cdot)$: $f(x, \omega) \leq \sum_{k=1}^{d+1} \lambda_k f(x, \omega_k) \leq \max_{\mathcal{G}(\Omega')} f(x, \cdot)$. Comme ceci est vrai pour n'importe quel $\omega \in \Omega'$, on a : $\max_{\Omega'} f(x, \cdot) \leq \max_{\mathcal{G}(\Omega')} f(x, \cdot)$. Ceci achève la preuve. \square

Ainsi, on peut se restreindre aux seuls points extrêmes de Ω' (convexe) pour assurer une robustesse de la solution à Ω' dans son ensemble. Cette observation a fréquemment été exploitée dans la littérature (voir par exemple le travail de [5, 6]).

2.2.2 Un cas particulier

La proposition 2 a montré que, sous certaines hypothèses de convexité, on peut se restreindre aux points extrêmes de l'ensemble des scénarios envisagés. On montre ici que pour un certain contexte fréquent en pratique, le nombre de scénarios vraiment utiles peut théoriquement encore être réduit.

On reprend d'abord quelques éléments classiques de l'optimisation lagrangienne ; on renvoie par exemple à [53, 69], notamment pour les preuves. On note $J = \{1, \dots, m\}$ l'ensemble des contraintes. Considérons le problème :

$$\min \{ \ell(x) \mid \forall j \in J, h_j(x) \geq 0, \text{ et } x \in \mathbb{R}^n \}. \quad (2.2)$$

Le lagrangien associé est : $L(x, \lambda) = \ell(x) - \sum_{j \in J} \lambda_j h_j(x)$.

Définition 3 Soient $\bar{x} \in \mathbb{R}^n$ et $\bar{\lambda} \geq 0$, on dit que $(\bar{x}, \bar{\lambda})$ est un point-col de L si :

- (i) $\forall x \in \mathbb{R}^n, L(\bar{x}, \bar{\lambda}) \leq L(x, \bar{\lambda})$,
- (ii) $\forall \lambda \geq 0, L(\bar{x}, \bar{\lambda}) \geq L(\bar{x}, \lambda)$.

On a le résultat suivant, très général (voir [53], partie 2.2, théorème 3) :

Théorème 2 Si $(\bar{x}, \bar{\lambda})$ est un point-col de L , alors \bar{x} est un optimum global du problème (2.2).

Ce théorème admet une réciproque sous certaines hypothèses (voir [53], partie 2.5, théorème 5) :

Théorème 3 Supposons que :

- (i) les fonctions $\{h_j\}_{j \in J}$ sont concaves, et ℓ est convexe ;
- (ii) il existe $x \in \mathbb{R}^n$ tel que, pour tout $j \in J$, $h_j(x) > 0$.

Alors, si le problème (2.2) admet une solution optimale \bar{x} , il existe un multiplicateur de Lagrange $\bar{\lambda} \geq 0$ tel que $(\bar{x}, \bar{\lambda})$ est un point-col de L .

Dans notre contexte d'optimisation robuste, le résultat suivant montre qu'en théorie, lorsque g peut être écrit à l'aide de plusieurs contraintes g_j possédant certaines propriétés de linéarité, alors on peut encore diminuer le nombre de scénarios à considérer. Les hypothèses de ce résultat sont fréquemment vérifiées en pratique.

Une fonction $f(., .)$ de $\mathbb{R}^n \times \Omega$ dans \mathbb{R} est dite *convexe-affine* si (i) pour tout $x \in \mathbb{R}^n$, $f(x, .)$ est affine, et si (ii) pour tout $\omega \in \Omega$, $f(., \omega)$ est convexe. La notion de fonction concave-affine s'en déduit immédiatement.

Proposition 3 Soit $\{g_j(., .)\}_{j \in J}$ une famille de fonctions concaves-affines, de $\mathbb{R}^n \times \Omega$ dans \mathbb{R} , telle que $g(., .) = \min_{j \in J} g_j(., .)$. On suppose aussi que $f(., .)$ est convexe-affine, de $\mathbb{R}^n \times \Omega$ dans \mathbb{R} . Soit $\Omega' \subset \Omega$ un polytope, on suppose que $X(\Omega')$ est borné et contient un point \check{x} tel que : $\forall \omega \in \Omega', g(\check{x}, \omega) > 0$. Il existe $\{\bar{\omega}_j\}_{j \in J} \subset \Omega'$ tel que :

$$\min_{x \in X(\Omega')} \max_{\omega \in \Omega'} f(x, \omega) = \min \left\{ \max_{j \in J} f(x, \bar{\omega}_j) \mid \forall j \in J, g_j(x, \bar{\omega}_j) \geq 0 \right\}.$$

Preuve : On rappelle qu'une fonction convexe à valeurs réelles sur un ouvert est continue en tout point de cet ouvert (voir e.g. [34]). Comme Ω' est un polytope, $\mathcal{G}(\Omega')$ contient un nombre fini d'éléments : on note $\mathcal{G}(\Omega') = \{\omega_k\}_{k \in \{1, \dots, N\}}$. Pour simplifier la preuve, on suppose que f ne dépend pas de l'événement ω : $f(x, \omega) = f(x)$. Ceci se fait sans perte de généralité, puisque le problème : $\min_{x \in X(\Omega')} \max_{\omega \in \Omega'} f(x, \omega)$ s'écrit de manière équivalente : $\min_{z, x \in X(\Omega')} \{z \mid \forall \omega \in \Omega' : z - f(x, \omega) \geq 0\}$. Il suffit alors de considérer comme variable $x' = (x, z)$, et une fonction de contrainte supplémentaire $g_{m+1}(x', \omega) = z - f(x, \omega)$.

D'après la proposition 2 : $\min_{x \in X(\Omega')} f(x) = \min_{x \in X(\mathcal{G}(\Omega'))} f(x)$. Le lagrangien associé à ce problème d'optimisation est noté : $L(x, \lambda) = f(x) - \sum_{j \in J} \sum_{k \in \{1, \dots, N\}} \lambda_{j,k} g_j(x, \omega_k)$. Puisque les fonctions $g_j(., \omega_k)$ sont concaves sur \mathbb{R}^n , elles sont continues : l'ensemble $X(\mathcal{G}(\Omega'))$ est fermé. On a de plus supposé que cet ensemble était borné. On sait par ailleurs que f est convexe sur \mathbb{R}^n , donc continue : il existe une solution optimale $x^* \in X(\mathcal{G}(\Omega'))$ au problème $\min_{x \in X(\mathcal{G}(\Omega'))} f(x)$.

Observer que le point $\check{x} \in X(\mathcal{G}(\Omega'))$ satisfait : $g_j(\check{x}, \omega_k) > 0$ pour tous $j \in J$ et $k \in \{1, \dots, N\}$. Le théorème 3 s'applique alors : il existe un multiplicateur de Lagrange $\lambda^* \geq 0$ tel que (x^*, λ^*) est un point-col du lagrangien L . On suppose qu'il existe j tel que $\lambda_j^* > 0$ (si ce n'est pas le cas, $L(x, \lambda^*) = f(x)$: on peut optimiser sans contrainte). Comme $g_j(x^*, .)$ est affine,

on peut écrire : $g_j(x^*, \cdot) = \tilde{g}_j(x^*, \cdot) + \beta_j(x^*)$, où $\tilde{g}_j(x^*, \cdot)$ est linéaire et $\beta_j(x^*)$ est une constante. En posant $S_j^* = \sum_{k \in \{1, \dots, N\}} \lambda_{j,k}^*$ et $\tilde{\lambda}_{j,k} = \lambda_{j,k}^* / S_j^*$, on a :

$$\begin{aligned} L(x^*, \lambda^*) &= f(x^*) - \sum_{j \in J} S_j^* \cdot \tilde{g}_j(x^*, \sum_{k \in \{1, \dots, N\}} \tilde{\lambda}_{j,k} \omega_k) - \sum_{j \in J} S_j^* \cdot \beta_j(x^*) \\ &= f(x^*) - \sum_{j \in J} S_j^* \cdot g_j(x^*, \sum_{k \in \{1, \dots, N\}} \tilde{\lambda}_{j,k} \omega_k) \end{aligned}$$

Introduisons l'événement : $\bar{\omega}_j = \sum_{k \in \{1, \dots, N\}} \tilde{\lambda}_{j,k} \omega_k$. Comme $\sum_{k \in \{1, \dots, N\}} \tilde{\lambda}_{j,k} = 1$, d'après la convexité de $\Omega' : \bar{\omega}_j \in \Omega'$. On vérifie alors facilement que (x^*, S^*) est un point-col du lagrangien : $\tilde{L}(x, S) = f(x) - \sum_{j \in J} S_j \cdot g_j(x, \bar{\omega}_j)$. D'après le théorème 2, ceci signifie que x^* est une solution optimale de $\min\{f(x) \mid \forall j, g_j(x, \bar{\omega}_j) \geq 0\}$. \square

On garantit donc qu'en théorie, quelle que soit la forme du polytope Ω' (et en particulier son nombre de points extrêmes), le problème robuste peut s'écrire à l'aide de seulement m scénarios, dont chacun est utilisé sur une seule des m contraintes. Cependant, la détermination de ces scénarios peut ne pas être immédiate en pratique.

Remarque : On trouvera d'autres résultats de discrétisation similaires dans [17] (voir les propositions 5.100 et suivantes dans cet ouvrage). Ils sont établis sous des hypothèses plus faibles et dans un cadre plus général, où l'on ne suppose pas une décomposition de g sous la forme de m fonctions $\{g_j(\cdot, \cdot)\}_{j \in J}$. En contre-partie, les discrétisations obtenues concernent l'ensemble du problème : dans le contexte de la proposition 3, cela signifie autant de répliques des m contraintes que de scénarios de discrétisation. La proposition 3 garantit, sous des hypothèses certes plus restrictives, que le problème discrétisé peut s'écrire avec seulement m contraintes.

2.3 Ensembles polyédriques

On considère de nouveau dans cette partie que Ω est convexe et inclus dans un espace vectoriel de dimension finie d . Il est intéressant d'étudier les ensembles polyédriques, qui jouent un rôle important en programmation mathématique, en particulier pour la programmation linéaire. On déduit du lemme 5 ce corollaire :

Corollaire 1 *Supposons que :*

- pour tout $x \in X$, $g(x, \cdot)$ est concave sur Ω ,
- pour tout $\omega \in \Omega$, X_ω est un polyèdre.

Soit $F \subset \Omega$, si F est un polytope, alors $X(F)$ est un polyèdre.

Preuve : D'après le lemme 5, $X(F) = X(\mathcal{G}(F))$. Si F est un polytope, alors $\mathcal{G}(F)$ est de cardinalité N finie : $\mathcal{G}(F) = \{\omega_k\}_{k \in \{1, \dots, N\}}$. En conséquence : $X(F) = X(\mathcal{G}(F)) = \bigcap_{k \in \{1, \dots, N\}} X_{\omega_k}$. Donc $X(F)$ est un polyèdre, en tant qu'intersection d'un nombre fini de polyèdres. \square

On peut s'interroger sur la validité d'une sorte de réciproque de ce résultat : si l'on considère un ensemble polyédrique V de solutions, existe-t-il un polyèdre F d'événements tel que $V = X(F)$? En d'autres termes, on veut savoir si un ensemble polyédrique de solutions dans \mathbb{R}^n peut être décrit comme l'ensemble des solutions robustes à un polyèdre d'événements dans Ω (ou de manière équivalente dans notre contexte, comme l'ensemble robuste à un nombre fini de scénarios, cf. proposition 2).

En inversant les rôles de x et de ω dans le corollaire 1, on obtient :

Lemme 6 *Supposons que :*

- pour tout $\omega \in \Omega$, $g(\cdot, \omega)$ est concave sur \mathbb{R}^n ,
- pour tout $x \in X$, $\Omega_x = \{\omega \in \Omega | g(x, \omega) \geq 0\}$ est un polyèdre.

Soit $V \subset X$, si V est un polytope, alors l'ensemble maximal d'événements auquel V est robuste est un polyèdre. Autrement dit, il existe un polyèdre $F \subset \Omega$ tel que :

- (i) $V \subseteq X(F)$,
- (ii) pour tout $\omega \in \Omega \setminus F$, $V \not\subseteq X_\omega$.

Preuve : En suivant le même raisonnement que pour la preuve du corollaire 1, on montre que $\bigcap_{x \in V} \Omega_x = \bigcap_{x \in \mathcal{G}(V)} \Omega_x$ est un polyèdre. Considérons alors $F = \bigcap_{x \in V} \Omega_x$: il est clair que $V \subseteq X(F)$. En effet, soit $x \in V$: pour tout $\omega \in F$, par construction, $\omega \in \Omega_x$ et donc $g(x, \omega) \geq 0$.

Par ailleurs, considérons $\omega \in \Omega \setminus F$ quelconque. Si $V \subseteq X_\omega$, on a : $\forall x \in V, g(x, \omega) \geq 0$. Ceci signifie : $\omega \in \bigcap_{x \in V} \Omega_x = F$, ce qui est une contradiction. D'où le point (ii). \square

Corollaire 2 *Supposons que :*

- pour tout $\omega \in \Omega$, $g(\cdot, \omega)$ est concave sur \mathbb{R}^n ,
- pour tout $x \in X$, $\Omega_x = \{\omega \in \Omega | g(x, \omega) \geq 0\}$ est un polyèdre.

Soit $F_1 \subset \Omega$ tel que $X(F_1)$ est un polytope, alors il existe un polyèdre $F_2 \supseteq F_1$ tel que $X(F_2) = X(F_1)$.

Preuve : D'après le lemme 6, il existe un polyèdre $F_2 \subseteq \Omega$ tel que $X(F_1) \subseteq X(F_2)$, et F_2 est "maximal". Cette dernière propriété assure que $F_1 \subseteq F_2$, et donc d'après le lemme 2 : $X(F_1) \supseteq X(F_2)$. \square

Le résultat ci-dessus impose une forme polyédrique aux ensembles de robustesse associés à chaque solution (les ensembles $\{\Omega_x\}_{x \in X}$). Dans la suite de cette partie, on s'intéresse à des hypothèses de convexité différentes, permettant d'obtenir le même type de résultat. On rappelle quelques résultats classiques de l'analyse convexe (voir par exemple [69]).

Définition 4 *Soit $C \subset \mathbb{R}^n$ un ensemble convexe fermé, un demi-espace est dit porteur pour C s'il contient C et si sa frontière contient un point de C . Si un demi-espace porteur est unique au point $x \in C$, il est dit tangent à C .*

On sait que (voir [69], théorème 18.8) :

Théorème 4 *Un ensemble convexe fermé $C \subset \mathbb{R}^n$ de dimension n est l'intersection des demi-espaces fermés qui lui sont tangents.*

On introduit la distance classique de Hausdorff entre deux ensembles S_1 et S_2 :

$$d_H(S_1, S_2) = \max \left\{ \max_{x \in S_1} \min_{y \in S_2} \|x - y\|, \max_{y \in S_2} \min_{x \in S_1} \|x - y\| \right\}.$$

Soit $\mathcal{S}(\mathbb{R}^n)$ la collection de tous les ensembles fermés de \mathbb{R}^n , $d_H(\cdot, \cdot)$ est une distance sur $\mathcal{S}(\mathbb{R}^n)$ (observer que $d_H(S, cl(S)) = 0$ pour tout ensemble S) :

- pour $(S_1, S_2) \in \mathcal{S}(\mathbb{R}^n)^2$: $S_1 = S_2 \Leftrightarrow d_H(S_1, S_2) = 0$,
- pour $(S_1, S_2) \in \mathcal{S}(\mathbb{R}^n)^2$, $d_H(S_1, S_2) = d_H(S_2, S_1)$,
- pour $(S_1, S_2, S_3) \in \mathcal{S}(\mathbb{R}^n)^3$, $d_H(S_1, S_2) \leq d_H(S_1, S_3) + d_H(S_3, S_2)$. En effet :

$$\begin{aligned} & \forall (u, v) \in \mathcal{S}_3^2 : \|x - y\| \leq \|x - u\| + \|u - v\| + \|v - y\| \\ \Rightarrow & \forall (u, v) \in \mathcal{S}_3^2 : \min_{y \in S_2} \|x - y\| \leq \|x - u\| + \|u - v\| + \min_{y \in S_2} \|v - y\| \\ \Rightarrow & \forall v \in S_3 : \min_{y \in S_2} \|x - y\| \leq \min_{u \in S_3} (\|x - u\| + \|u - v\|) + \min_{y \in S_2} \|v - y\| \\ \Rightarrow & \forall v \in S_3 : \max_{x \in S_1} \min_{y \in S_2} \|x - y\| \leq 0 + \max_{x \in S_1} \min_{u \in S_3} \|x - u\| + \min_{y \in S_2} \|v - y\| \\ \Rightarrow & \max_{x \in S_1} \min_{y \in S_2} \|x - y\| \leq \max_{x \in S_1} \min_{u \in S_3} \|x - u\| + \max_{v \in S_3} \min_{y \in S_2} \|v - y\| \end{aligned}$$

et donc : $\max_{x \in S_1} \min_{y \in S_2} \|x - y\| \leq d_H(S_1, S_3) + d_H(S_3, S_2)$. En inversant les rôles de S_1 et S_2 , on a aussi : $\max_{x \in S_2} \min_{y \in S_1} \|x - y\| \leq d_H(S_2, S_3) + d_H(S_3, S_1)$, ce qui achève la démonstration.

Définition 5 Soit $(S_k)_{k \in \mathbb{N}}$ une suite d'ensembles. \bar{S} est appelé la limite de (S_k) si \bar{S} est fermé et : $d_H(S_k, \bar{S}) \rightarrow 0$.

On introduit la fonction :

$$G : \Omega \longrightarrow \mathcal{S}(\mathbb{R}^n) \\ \omega \longmapsto X_\omega = g(\cdot, \omega)^{-1}(\mathbb{R}_+) = \{x \in \mathbb{R}^n \mid g(x, \omega) \geq 0\} .$$

Pour tous $x \in \mathbb{R}^n$ et $r > 0$, $B(x, r)$ désigne la boule fermée de centre x et de rayon r .

Théorème 5 Supposons que :

- (i) $g(\cdot, \cdot)$ est continue sur $\mathbb{R}^n \times \Omega$,
- (ii) pour tout $\omega \in \Omega$, si x est un point intérieur de $G(\omega)$, alors $g(x, \omega) > 0$,
- (iii) pour tout $\omega \in \Omega$, $X_\omega = G(\omega)$ est convexe,
- (iv) il existe $\delta > 0$ tel que, pour tout $\omega \in \Omega$, on peut trouver $x \in G(\omega)$ tel que $B(x, \delta) \subset G(\omega)$,
- (v) il existe $\Delta > 0$ tel que, pour tout $\omega \in \Omega$, $G(\omega) \subset B(0, \Delta)$.

Alors G est continue sur Ω .

Preuve : Soit $\bar{\omega} \in \Omega$, montrons que G est continue en $\bar{\omega}$. Si ce n'est pas le cas, il existe $\varepsilon > 0$ tel que, pour tout $\eta > 0$, il existe $\omega \in \Omega$ tel que $\|\omega - \bar{\omega}\| < \eta$ et $d_H(G(\omega), G(\bar{\omega})) \geq \varepsilon$. Alors on peut construire une suite $(\omega_k)_{k \in \mathbb{N}}$ dans Ω telle que : $\omega_k \rightarrow \bar{\omega}$ et $d_H(G(\omega_k), G(\bar{\omega})) \geq \varepsilon$.

• Soit un quelconque $k \in \mathbb{N}$. Puisque tous les ensembles considérés sont fermés (cf. (i)), $d_H(G(\omega_k), G(\bar{\omega})) \geq \varepsilon$ implique l'existence de $x \in \mathbb{R}^n$ tel que, ou bien $x \in G(\omega_k)$ et $B(x, \varepsilon) \cap G(\bar{\omega}) = \emptyset$, ou bien $x \in G(\bar{\omega})$ et $B(x, \varepsilon) \cap G(\omega_k) = \emptyset$. Ainsi, en notant $r = \varepsilon\delta/\Delta$, on peut trouver x_k satisfaisant :

- ou bien $B(x_k, r) \subset G(\omega_k)$ et $B(x_k, r) \cap G(\bar{\omega}) = \emptyset$,
- ou bien $B(x_k, r) \subset G(\bar{\omega})$ et $B(x_k, r) \cap G(\omega_k) = \emptyset$.

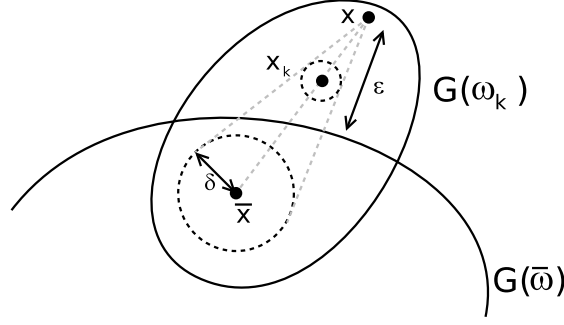
En effet, montrons cela dans le premier cas, quand $x \in G(\omega_k)$ et $B(x, \varepsilon) \cap G(\bar{\omega}) = \emptyset$. D'après l'hypothèse (v), il existe $\bar{x} \in G(\omega_k)$ tel que $B(\bar{x}, \delta) \subset G(\omega_k)$. D'après la convexité de $G(\omega_k)$, tout point $\lambda x + (1 - \lambda)\bar{x}$ est dans $G(\omega_k)$, pour $\lambda \in [0, 1]$. Notons $x_k = \lambda_k x + (1 - \lambda_k)\bar{x}$ pour $\lambda_k \in [0, 1]$. Considérons l'ensemble \mathcal{B} tel que $y \in \mathcal{B}$ si et seulement s'il existe $\bar{y} \in B(\bar{x}, \delta)$ vérifiant : $y = \lambda_k x + (1 - \lambda_k)\bar{y}$. Comme $x \in G(\omega_k)$ et $B(\bar{x}, \delta) \subset G(\omega_k)$, la convexité de $G(\omega_k)$ implique que $\mathcal{B} \subset G(\omega_k)$. De plus, pour tout $y \in \mathcal{B}$:

$$\begin{aligned} x_k - y &= (x_k - \bar{x}) + (\bar{x} - \bar{y}) + (\bar{y} - y) = \lambda_k(x - \bar{x}) + (\bar{x} - \bar{y}) + \lambda_k(\bar{y} - x) \\ &= (1 - \lambda_k)(\bar{x} - \bar{y}). \end{aligned}$$

Ceci montre que : $\mathcal{B} = B(x_k, (1 - \lambda_k)\delta)$. Donc, pour tout $y \in \mathcal{B}$:

$$\|x - y\| \leq \|x - x_k\| + \|x_k - y\| \leq (1 - \lambda_k)(\|x - \bar{x}\| + \delta) \leq (1 - \lambda_k)(\Delta - \delta + \delta) = (1 - \lambda_k)\Delta.$$

Ainsi, en posant $\lambda_k = 1 - \varepsilon/\Delta$, on garantit : $\mathcal{B} \subset B(x, \varepsilon)$, et donc : $\mathcal{B} \cap G(\bar{\omega}) = \emptyset$. En définitive, observer que $\mathcal{B} = B(x_k, r)$.



De la même manière, si $x \in G(\bar{\omega})$ et $B(x, \varepsilon) \cap G(\omega_k) = \emptyset$ (deuxième cas), on peut trouver x_k tel que $B(x_k, r) \subset G(\bar{\omega})$ et $B(x_k, r) \cap G(\omega_k) = \emptyset$.

• On construit ainsi une suite $(x_k)_{k \in \mathbb{N}}$. D'après l'hypothèse (iv), cette suite est bornée et on peut donc en extraire une sous-suite convergente $(x_{\varphi(k)})$, de limite \bar{x} (cf. théorème de Bolzano-Weierstrass). Comme on a supposé que $g(\cdot, \cdot)$ est continue : $g(x_{\varphi(k)}, \omega_{\varphi(k)}) \cdot g(x_{\varphi(k)}, \bar{\omega}) \rightarrow g(\bar{x}, \bar{\omega})^2 \geq 0$, et comme par construction $g(x_{\varphi(k)}, \omega_{\varphi(k)}) \cdot g(x_{\varphi(k)}, \bar{\omega}) < 0$, on a nécessairement : $g(\bar{x}, \bar{\omega}) = 0$. D'après l'hypothèse (ii), ceci signifie que \bar{x} est sur la frontière de $G(\bar{\omega})$. Mais comme la suite (x_k) a été construite de manière à être toujours à distance au moins $r > 0$ de cette frontière, ce n'est pas possible (contradiction). \square

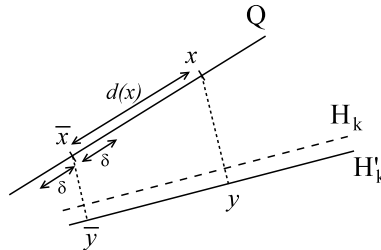
Soit une inégalité $ax \geq b$ valide pour un polyèdre P . L'ensemble $F = \{x \in P \mid ax = b\}$ est une *facette* si $\dim(F) = \dim(P) - 1$.

Lemme 7 Soit \mathcal{C} une collection d'ensembles convexes de \mathbb{R}^n , et supposons que $P = \bigcap_{S \in \mathcal{C}} S$ est un polyèdre de dimension pleine. Soit $ax \geq b$ une inégalité définissant une facette de P , on note $Q = \{x \in P \mid ax = b\}$. Il existe une suite $(S_k) \subset \mathcal{C}$ telle que :

$$\forall x \in Q, \min_{y \notin S_k} \|x - y\| \longrightarrow 0.$$

Preuve : Soit \bar{x} un point intérieur de Q : il existe $\delta > 0$ tel que \bar{x} est à une distance au moins δ de la frontière de Q . Pour tout $k \in \mathbb{N}^*$, il existe $S_k \in \mathcal{C}$ tel que $B(\bar{x}, 1/k) \not\subseteq S_k$. Il existe donc un demi-espace H_k tangent à S_k et vérifiant $B(\bar{x}, 1/k) \not\subseteq H_k$. Observer que $P \subseteq S_k \subseteq H_k$.

Soit maintenant $x \in Q$ quelconque, on note : $d(x) = \|x - \bar{x}\|$. Comme $B(\bar{x}, 1/k) \not\subseteq H_k$, il existe $\bar{y} \notin H_k$ tel que $\|\bar{y} - \bar{x}\| \leq 1/k$. On note H'_k l'hyper-plan parallèle à la frontière de H_k et passant par \bar{y} : $H'_k = \{x \in \mathbb{R}^n \mid a \cdot (x - \bar{y}) = 0\}$. Soit $y \in H'_k$ tel que $(y - x)$ et $(\bar{y} - \bar{x})$ sont co-linéaires : $y - x = \alpha \cdot (\bar{y} - \bar{x})$. Puisque $P \subseteq H_k$, on a forcément $\alpha \geq 0$.



On introduit, pour tout $\lambda > 0$, le point : $x(\lambda) = \bar{x} + \frac{\lambda}{d(x)} \cdot (\bar{x} - x)$. Ce point est donc à distance λ de \bar{x} dans la direction $(\bar{x} - x)$. Ceci implique en particulier que si $\lambda \leq \delta$, alors $x(\lambda) \in Q$. On

définit : $y(\lambda) = \bar{y} + \frac{\lambda}{d(x)} \cdot (\bar{y} - y)$. Par construction, $y(\lambda) \in H'_k$. Par ailleurs :

$$\begin{aligned} x(\lambda) - y(\lambda) &= (x(\lambda) - \bar{x}) + (\bar{x} - \bar{y}) + (\bar{y} - y(\lambda)) \\ &= \frac{\lambda}{d(x)} \cdot (\bar{x} - x) + (\bar{x} - \bar{y}) - \frac{\lambda}{d(x)} \cdot (\bar{y} - y) \\ &= \left(1 + \frac{\lambda}{d(x)}\right) \cdot (\bar{x} - \bar{y}) + \frac{\lambda}{d(x)} \cdot (y - x) \\ &= \left(1 + \frac{\lambda}{d(x)}[1 - \alpha]\right) \cdot (\bar{x} - \bar{y}). \end{aligned}$$

On se place dans le cas $\lambda = \delta$, ce qui assure $x(\lambda) \in P$: $x(\delta) - y(\delta) = \left(1 + \frac{\delta}{d(x)}[1 - \alpha]\right) \cdot (\bar{x} - \bar{y})$. Puisque $P \subseteq H_k$, les vecteurs $x(\delta) - y(\delta)$ et $\bar{x} - \bar{y}$ doivent être de même directions, donc : $\left(1 + \frac{\delta}{d(x)}[1 - \alpha]\right) \geq 0$. On en déduit : $\alpha \leq 1 + d(x)/\delta$. Alors :

$$\|y - x\| = \alpha \cdot \|\bar{y} - \bar{x}\| \leq \left(1 + \frac{d(x)}{\delta}\right) \cdot \frac{1}{k}.$$

Pour finir, remarquer que $y \notin H_k$. La suite $(S_k)_{k \in \mathbb{N}}$ est donc telle que :

$$\min_{y \notin S_k} \|x - y\| \leq \left(1 + \frac{d(x)}{\delta}\right) \cdot \frac{1}{k} \longrightarrow 0.$$

Ceci est vrai pour tout $x \in Q$. \square

Proposition 4 *Supposons que :*

- (i) $g(., .)$ est continue sur $\mathbb{R}^n \times \Omega$,
- (ii) pour tout $\omega \in \Omega$, si x est un point intérieur de X_ω , alors $g(x, \omega) > 0$,
- (iii) $g(., .)$ est concave-concave sur $\mathbb{R}^n \times \Omega$,
- (iv) il existe $\delta > 0$ tel que, pour tout $\omega \in \Omega$, on peut trouver $x \in X_\omega$ tel que $B(x, \delta) \subset X_\omega$,
- (v) il existe $\Delta > 0$ tel que, pour tout $\omega \in \Omega$, $X_\omega \subset B(0, \Delta)$.

Soit $F_1 \subset \Omega$ un ensemble fermé. Si $X(F_1)$ est un polyèdre, alors il existe un polytope $F_2 \subset \Omega$ tel que $X(F_2) = X(F_1)$. Si F_1 est convexe, alors $F_2 \subseteq F_1$.

Preuve : On suppose sans perte de généralité que $X(F_1)$ est de dimension pleine ; si ce n'est pas le cas dans \mathbb{R}^n , on se ramène à cette hypothèse en considérant un sous-espace vectoriel plus petit. Soit $\Xi = \{X_\omega\}_{\omega \in F_1}$. On note Q_p , pour $p \in \{1, \dots, N\}$, les facettes de $X(F_1)$; si $a_p x \geq b_p$ est l'inégalité caractérisant Q_p , $Q_p = \{x \in X(F_1) | a_p x = b_p\}$. Pour tout $p \in \{1, \dots, N\}$, d'après le lemme 7, il existe une suite $(X_{\omega_k^p})_{k \in \mathbb{N}} \subset \Xi$ telle que : $\forall x \in Q_p, \min_{y \notin X_{\omega_k^p}} \|x - y\| \longrightarrow 0$. Soit $(\omega_{\varphi(k)}^p)_{k \in \mathbb{N}}$ une sous-suite convergente de $(\omega_k^p)_{k \in \mathbb{N}}$, et notons $\bar{\omega}^p \in F_1$ sa limite. D'après le théorème 5, G est continue, et donc la suite $(X_{\omega_{\varphi(k)}^p})$ tend vers $X_{\bar{\omega}^p}$. Cet ensemble limite satisfait : $X(F_1) \subseteq X_{\bar{\omega}^p}$, puisque $\bar{\omega}^p \in F_1$. De plus, pour tout $x \in Q_p$: $\lim_{k \rightarrow \infty} \left\{ \min_{y \notin X_{\omega_k^p}} \|x - y\| \right\} = \min_{y \notin X_{\bar{\omega}^p}} \|x - y\|$, donc : $\min_{y \notin X_{\bar{\omega}^p}} \|x - y\| = 0$. Q_p est donc une facette de $X_{\bar{\omega}^p}$, puisque $Q_p \subseteq X(F_1) \subseteq X_{\bar{\omega}^p}$ et pour tout $x \in Q_p$, $\min_{y \notin X_{\bar{\omega}^p}} \|x - y\| = 0$ (cette dernière condition assure que Q_p est "au bord" de $X_{\bar{\omega}^p}$).

On a alors :

$$X(F_1) \subseteq \bigcap_{p \in \{1, \dots, N\}} X_{\bar{\omega}^p} \subseteq \bigcap_{p \in \{1, \dots, N\}} \{x \in \mathbb{R}^m | a_p x \geq b_p\} = X(F_1)$$

et donc : $X(F_1) = \bigcap_{p \in \{1, \dots, N\}} X_{\bar{\omega}^p}$. Finalement, on note $F_2 = \text{conv}\{\bar{\omega}^p\}_{p \in \{1, \dots, N\}}$. F_2 est un polytope, et puisque $\mathcal{G}(F_2) = \{\bar{\omega}^p\}_{p \in \{1, \dots, N\}}$, le lemme 5 implique : $X(F_2) = \bigcap_{p \in \{1, \dots, N\}} X_{\bar{\omega}^p} =$

$X(F_1)$. Si F_1 est convexe, $F_2 \subseteq F_1$ est évident par construction, ce qui conclut la preuve. \square

Ainsi, le corollaire 2 et la proposition 4 assurent, sous des hypothèses différentes, que si l'on décrit tous les ensembles polyédriques d'événements dans Ω , on est sûr aussi de décrire tous les polyèdres de solutions robustes dans X . Ces résultats ne peuvent s'appliquer directement à des cas où des variables de décision entières sont utilisées. Ils peuvent cependant être étendus à ce contexte en se ramenant à la relaxation continue correspondante.

On illustre ceci avec l'extension suivant de la proposition 4 :

Proposition 5 Soit \mathcal{X} un sous-ensemble quelconque de \mathbb{R}^n . Supposons que pour tout $\omega \in \Omega$: $X_\omega = \{x \in \mathcal{X} \mid g(x, \omega) \geq 0\}$. On note : $\tilde{X}_\omega = \{x \in \mathbb{R}^n \mid g(x, \omega) \geq 0\}$. Supposons par ailleurs que :

- (i) $g(\cdot, \cdot)$ est continue sur $\mathbb{R}^n \times \Omega$,
- (ii) pour tout $\omega \in \Omega$, si x est un point intérieur de \tilde{X}_ω , alors $g(x, \omega) > 0$,
- (iii) $g(\cdot, \cdot)$ est concave-concave sur $\mathbb{R}^n \times \Omega$,
- (iv) il existe $\delta > 0$ tel que, pour tout $\omega \in \Omega$, on peut trouver $x \in \tilde{X}_\omega$ tel que $B(x, \delta) \subset \tilde{X}_\omega$,
- (v) il existe $\Delta > 0$ tel que, pour tout $\omega \in \Omega$, $\tilde{X}_\omega \subset B(0, \Delta)$.

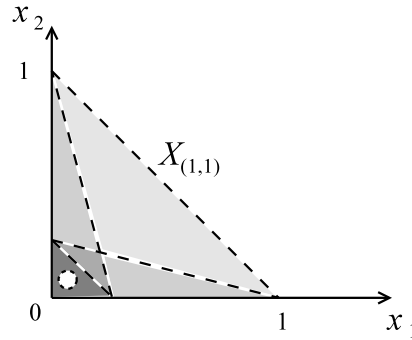
Soit $F_1 \subset \Omega$ un ensemble fermé. Si $\tilde{X}(F_1) = \{x \in \mathbb{R}^n \mid \forall \omega \in F_1, g(x, \omega) \geq 0\}$ est un polyèdre, alors il existe un polytope $F_2 \subset \Omega$ tel que $X(F_2) = X(F_1)$. Si F_1 est convexe, alors $F_2 \subseteq F_1$.

Preuve : D'après la proposition 4, il existe un polytope $F_2 \subseteq F_1$ tel que $\tilde{X}(F_2) = \tilde{X}(F_1)$. Ceci assure : $\tilde{X}(F_2) \cap \mathcal{X} = \tilde{X}(F_1) \cap \mathcal{X}$, et donc : $X(F_2) = X(F_1)$. \square

On termine cette partie par un exemple simple illustrant le sens de la proposition 4.

Exemple : On considère $\Omega = [1, 2] \times [1, 2]$, et $X_\omega = \{x \geq 0 \mid \omega_1^2 \cdot x_1 + \omega_2^2 \cdot x_2 \leq 1\}$. Il est clair que $X = \bigcup_{\omega \in \Omega} X_\omega = \{x \geq 0 \mid x_1 + x_2 \leq 1\}$. Par ailleurs, d'après le lemme 5, on a : $X(\Omega) = X_{(1,1)} \cap X_{(1,2)} \cap X_{(2,1)} \cap X_{(2,2)} = X_{(2,2)} = \{x \geq 0 \mid x_1 + x_2 \leq 1/4\}$.

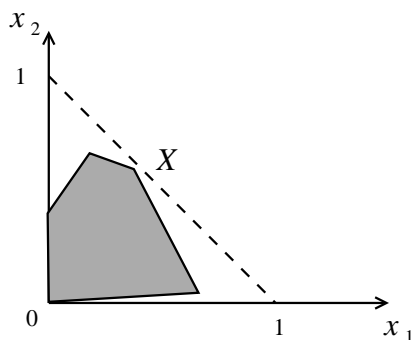
On représente les 4 ensembles extrémaux $X_{(1,1)}$, $X_{(1,2)}$, $X_{(2,1)}$ et $X_{(2,2)}$ sur la figure suivante :



On vérifie que les hypothèses de la proposition 5 sont satisfaites :

- (i) on remarque que $g(x, \omega) = \omega_1^2 \cdot x_1 + \omega_2^2 \cdot x_2$: $g(\cdot, \cdot)$ est bien continue sur $\mathbb{R}^2 \times \Omega$;
- (ii) pour tout $\omega \in \Omega$, un point x à l'intérieur de X_ω satisfait $g(x, \omega) > 0$;
- (iii) $g(\cdot, \cdot)$ est bien concave-concave également ;
- (iv) $X(\Omega)$ contient une boule de rayon non-nul (voir la boule blanche sur la figure), c'est donc vrai aussi pour tout X_ω , $\omega \in \Omega$;
- (v) l'ensemble $X = \bigcup_{\omega \in \Omega} X_\omega$ est borné.

Considérons alors un polyèdre \mathcal{P} quelconque de X , par exemple :



La proposition 4 nous garantit que :

- ou bien \mathcal{P} ne correspond à aucun $X(F)_{F \subseteq \Omega}$ (ce n'est pas un ensemble de solutions robustes),
- ou bien il existe un polyèdre $F \subseteq \Omega$ tel que $\mathcal{P} = X(F)$.

Il est clair ici que nous sommes dans le premier cas.

2.4 Sous-ensembles d'incertitudes pour la programmation linéaire

Le but de cette partie est de présenter rapidement quelques approches de la littérature pour définir des sous-ensembles de Ω dans le cadre de la programmation linéaire. On considère un système de contraintes $Ax \leq b$ où les coefficients de A sont incertains. On s'appuie sur l'analyse de [11], qui propose une vision unifiée de plusieurs approches robustes de la littérature telles que [9], [33] et [13].

Une adaptation du résultat principal de [11] est proposée ci-après ; on renvoie le lecteur à l'article initial pour un énoncé plus général. Ce résultat montre que considérer un sous-ensemble d'événements Ω' est en fait équivalent à introduire un terme de protection dans les contraintes linéaires du problème.

Soit $b \in \mathbb{R}^m$ le second membre. On considère qu'un événement est une matrice de contraintes $A \in \mathbb{R}^{m \times n}$. L'ensemble correspondant des solutions réalisables est : $X_A = \{x \in \mathcal{X} | Ax \leq b\}$, où \mathcal{X} est un sous-ensemble quelconque de \mathbb{R}^n . On introduit deux matrices de référence : la première, notée \underline{A} , correspond aux valeurs minimales que peuvent prendre les coefficients de A ; la deuxième, notée \overline{A} , correspond aux valeurs maximales de ces mêmes coefficients. On a bien sûr $\underline{A} \leq \overline{A}$. On note enfin : $\delta = \overline{A} - \underline{A} \geq 0$. Pour simplifier l'exposé, on suppose par la suite que $\delta > 0$.

Soit \otimes et \oslash les opérateurs définis sur $\mathbb{R}^n \times \mathbb{R}^n$, à valeurs dans \mathbb{R}^n , tels que : $w = u \otimes v \Leftrightarrow \forall i, w_i = u_i \cdot v_i$, et : $w = u \oslash v \Leftrightarrow \forall i, w_i = u_i / v_i$. Noter que $u \oslash v$ n'a de sens que si toutes les composantes de v sont non-nulles. On note :

$$\begin{cases} X^P = \{x \in \mathcal{X} | \forall j \in J, \underline{A}_j x + \|\delta_j \otimes x\| \leq b_j\} \\ X^R = \{x \in \mathcal{X} | \forall A \in \mathbb{R}^{m \times n} \text{ t.q. } \forall j \in J, \|(A_j - \underline{A}_j) \oslash \delta_j\|^* \leq 1 : Ax \leq b\} \end{cases}$$

où $\|\cdot\|$ est une norme quelconque, et $\|\cdot\|^*$ est la norme duale associée (on rappelle que : $\|u\|^* = \max_{\|x\| \leq 1} u \cdot x$). Ainsi, X^P est défini en partant du système de contraintes $\underline{A} \cdot x \leq b$, et en introduisant à chaque ligne j un terme de protection $\|\delta_j \otimes x\|$. D'autre part, X^R est l'ensemble des solutions robustes à tous les événements de $\Omega' = \{A \in \mathbb{R}^{m \times n} | \forall j \in J, \|(A_j - \underline{A}_j) \oslash \delta_j\|^* \leq 1\}$. X^P a l'avantage d'être écrit de manière beaucoup plus compacte. En revanche, la définition de X^R permet d'avoir une compréhension plus claire de l'ensemble des événements contre lesquels on se prémunit.

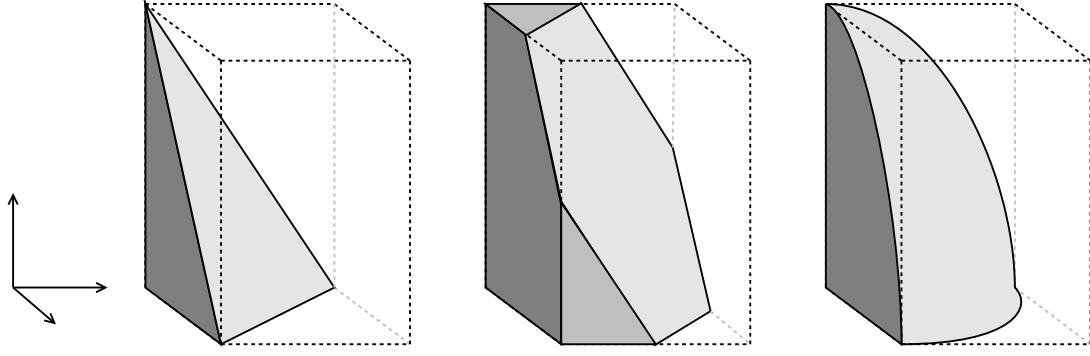


FIG. 2.1 – Ensembles d’incertitude pour trois approches robustes différentes.

Théorème 6 (Bertsimas, Pachamanova, Sim, 2004) $X^P = X^R$.

Preuve : Soit $j \in J$. Soit $\mathcal{S}_j = \{s \in \mathbb{R}^n \mid \|s \otimes \delta_j\|^* \leq 1\}$. Soit $x \in \mathcal{X}$ quelconque :

$$\left(A_j x \leq b_j, \forall A \text{ t.q. } (A - \underline{A})_j \in \mathcal{S}_j \right) \Leftrightarrow \max_{(A - \underline{A})_j \in \mathcal{S}_j} \{A_j x\} \leq b_j.$$

De plus, on a : $\max_{(A - \underline{A})_j \in \mathcal{S}_j} A_j x = \underline{A}_j x + \max_{s \in \mathcal{S}_j} s x$. On introduit : $e = s \otimes \delta_j$, et on a en conséquence : $s = e \otimes \delta_j$. Observer alors que :

$$\max_{s \in \mathcal{S}_j} s x = \max_{\|e\|^* \leq 1} (e \otimes \delta_j) \cdot x = \max_{\|e\|^* \leq 1} e \cdot (x \otimes \delta_j).$$

On en déduit finalement :

$$\max_{s \in \mathcal{S}_j} s x = (\|x \otimes \delta_j\|^*)^* = \|x \otimes \delta_j\|$$

puisque $(\|\cdot\|^*)^* = \|\cdot\|$. \square

Par exemple, considérons la norme L_1 , notée $\|\cdot\|_1$. Sa norme duale est : $\|\cdot\|_1^* = \|\cdot\|_\infty$. Dans ce cas, on obtient : $\Omega' = \Omega$. Dans le cas de la norme L_2 , on a : $\|\cdot\|_2^* = \|\cdot\|_2$. Dans ce cas : $\Omega' = \left\{ A \in \mathbb{R}^{m \times n} \mid \forall j \in J, \sum_{i \in I} (A_{ji} - \underline{A}_{ji})^2 / \delta_{ji}^2 \leq 1 \right\}$ (ellipsoïde).

Supposons que l’on a seulement 3 données scalaires incertaines. La figure 2.1 illustre les ensembles d’incertitude pris en compte pour 3 modèles robustes différents. Le premier, à gauche, correspond à un terme de protection écrit avec la norme L_∞ . Le deuxième correspond au modèle de [12, 13] (avec un paramètre $\Gamma = 1, 5$). Finalement, le troisième ensemble (à droite) est obtenu en travaillant avec la norme L_2 ; c’est sur une norme semblable que le travail de [8, 9] s’appuie.

Ces ensembles peuvent bien sûr être dilatés ou rétractés en considérant un multiple de la norme considérée. Par exemple, intéressons-nous à la norme $\lambda \|\cdot\|_\infty$, avec $\lambda > 0$. Quand λ augmente, la taille de la pyramide (cf. figure 2.1 à gauche) diminue. On représente sur la figure 2.2 les ensembles obtenus avec $\lambda = 2/3$ et $\lambda = 2$. Pour $\lambda = 2/3$, observer que l’ensemble de robustesse dépasse le cadre du pavé Ω . Ceci n’est pas très pertinent en pratique, puisqu’on a supposé que les données appartiennent forcément à Ω . On remarquera, par contraste, que les ensembles d’incertitudes de [12, 13] respectent mieux cette hypothèse (cf. figure 2.1, au milieu).

2.5 Un exemple simple

On rappelle que notre but est de trouver $\Omega' \subseteq \Omega$ tel que $\mathbb{P}(\Omega') \geq 1 - \varepsilon$ et assurant un coût robuste associé aussi faible que possible. Idéalement, on aimerait donc résoudre le problème

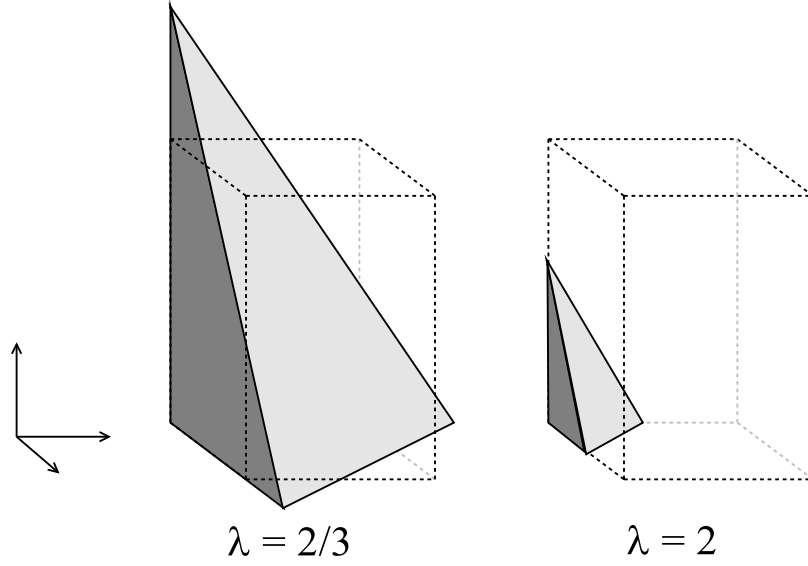


FIG. 2.2 – Dilatations de l'ensemble d'incertitude engendré par l'usage de la norme L_∞ .

(2.1). Au travers de l'exemple développé ci-après, on montre que la définition d'un tel sous-ensemble Ω' d'événements dépend fortement du problème considéré.

Considérons le programme linéaire suivant :

$$\begin{aligned} \min \quad & x_1 + 2x_2 \\ \text{s.c.} \quad & a_1x_1 + a_2x_2 \geq 1, \\ & x \geq 0. \end{aligned} \tag{2.3}$$

Chaque coefficient scalaire $a_i \geq 0$ est supposé incertain. Ainsi, l'ensemble Ω des événements est l'ensemble des valeurs possibles du vecteur a . Pour $a \in \Omega$, on note $V(a)$ la valeur optimale de (2.3).

Supposons que $a_1 \neq 0$ et $a_2 \neq 0$. On introduit $\rho_1 = 1/a_1$ et $\rho_2 = 2/a_2$, qui sont les ratios des coefficients de l'objectif et de la contrainte pour chaque variable $x_{1,2}$. Soit $k = \arg \min\{\rho_i\}_{i \in \{1,2\}}$, une solution optimale de (2.3) est :

$$\begin{cases} x_k^* = 1/a_k, \\ x_i^* = 0, \quad \text{pour } i \neq k. \end{cases}$$

Alors, la valeur optimale est $V(a) = \rho_k$. Pour tout $\alpha \geq 0$, soit S_α l'ensemble des vecteurs défini par : $S_\alpha = \{a \geq 0 \mid V(a) = \alpha\}$ (courbe de niveau α de la valeur du problème). De l'observation précédente, on déduit que $V(a) = \alpha$ si et seulement si $\min \rho_i = \alpha$. Donc :

$$S_\alpha = \{a \geq 0 \mid \begin{cases} a_1 = 1/\alpha \\ a_2 \leq 2/\alpha \end{cases} \text{ ou } \begin{cases} a_1 \leq 1/\alpha \\ a_2 = 2/\alpha \end{cases} \}.$$

Certaines de ces courbes sont dessinées sur la figure 2.3.

Avant d'entamer l'étude de notre exemple, il convient de remarquer que pour tout $F \subseteq \Omega$:

$$\min_{x \geq 0} \{x_1 + 2x_2 \mid \forall a \in F, a \cdot x \geq 1\} \geq \max_{a \in F} \min_{x \geq 0} \{x_1 + 2x_2 \mid a \cdot x \geq 1\} = \max_{a \in F} V(a).$$

Pour un ensemble d'incertitude F donné, le terme de droite, facile à obtenir à partir de la figure 2.3, donne donc une borne sur la valeur robuste. La démarche pour déterminer Ω' sera alors toujours la même :

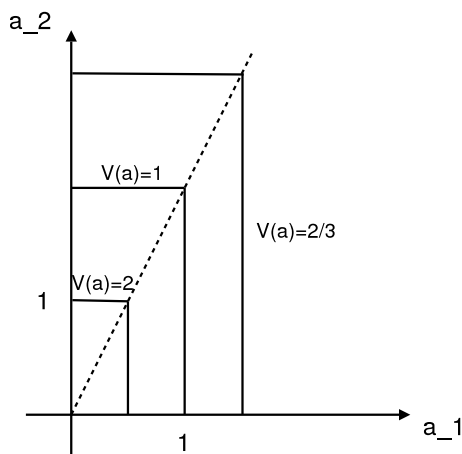


FIG. 2.3 – Courbes de niveaux de la valeur-objectif du problème (2.3).

- (i) construire d’abord Ω' , de mesure $\mathbb{P}(\Omega') = 1 - \varepsilon$, qui minimise la borne $\max_{a \in \Omega'} V(a)$;
- (ii) puis observer que l’inégalité ci-dessus est en fait vérifiée à l’égalité.

Un ensemble vérifiant (i) et (ii) est nécessairement optimal.

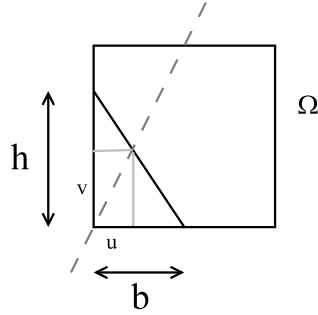
Supposons maintenant que chaque coefficient a_i appartient à un intervalle $[\underline{a}_i, \underline{a}_i + 1]$. Ainsi, Ω est un carré de côté de longueur 1 : $\Omega = [\underline{a}_1, \underline{a}_1 + 1] \times [\underline{a}_2, \underline{a}_2 + 1]$. Chaque coefficient est supposé être une variable aléatoire uniformément distribuée sur son intervalle de définition. Considérons $\varepsilon = 1/4$. Trois cas différents sont illustrés sur la figure 2.4 : chacun est caractérisé par des valeurs $(\underline{a}_1, \underline{a}_2)$ différentes, donc un carré Ω différent. Dans chaque cas, notre but va maintenant être de trouver $\Omega' \subseteq \Omega$ solution optimale de (2.1). Les constructions proposées sont très géométriques, et s’appuient en bonne partie sur une intuition visuelle. Elles ne prétendent pas à une parfaite rigueur mathématique.

Dans le cas de gauche, Ω est complètement à droite de la ligne en pointillés. En conséquence, toutes les lignes de niveau $S(\alpha)$ qui intersectent Ω sont verticales. On en déduit que Ω' est le rectangle obtenu en enlevant de Ω les événements “sur la gauche”. L’ensemble ainsi construit satisfait clairement (i) et (ii).

Dans le cas de droite, Ω est au contraire complètement à gauche de la ligne en pointillés. On voit alors que toutes les courbes de niveau de $S(\alpha)$ qui intersectent Ω sont horizontales. On en déduit que Ω' est le rectangle obtenu en enlevant les éléments “en bas” de Ω .

Finalement, le cas restant est un cas intermédiaire pour lequel Ω' est obtenu en retirant un triangle à Ω . Sa construction est plus complexe que dans les deux cas précédents. Nous la détaillons maintenant. On s’intéresse, de manière arbitraire, à un ensemble Ω' délimité dans Ω par une simple droite. On suppose que $\Omega \setminus \Omega'$ est un triangle rectangle de base b et de hauteur h . On veut avoir : $bh/2 = \varepsilon = 1/4$. Par ailleurs, l’hypoténuse du triangle rectangle intersecte la ligne en pointillés en un point (u, v) dont les coordonnées sont, relativement à Ω :

$$\begin{cases} v = 2u \\ v = -h/b \cdot v + h \end{cases}$$



Après simplifications, on obtient : $b^2v - b/2 + u/2 = 0$. Ceci n'est possible que si le discriminant associé à ce polynôme du second degré est positif : $\Delta = 1/4 - 2uv \geq 0$. Ceci implique : $u \leq 1/4$. Par ailleurs, pour assurer le plus bas coût possible sur l'ensemble des événements de Ω' , u doit être maximisé (cf. condition (i) ; observer en effet que le coût "diminue le long de la droite en pointillés"). On suppose donc maintenant que $u = 1/4$; ceci implique que $b = 1/2$ et $h = 1$.

Pour le moment, la seule garantie de l'ensemble Ω' construit est que son aire (ou sa probabilité) est bien égale à $1 - \varepsilon$. Montrons maintenant que c'est en fait une solution optimale de notre problème. On peut montrer que le maximum des coûts de chaque événement de Ω' est égal au coût de la solution robuste à tous les événements de Ω' (condition (ii)). D'après la forme des courbes de niveaux (voir figure 2.3), le coût maximal $\max_{a \in \Omega'} V(a)$ associé aux événements de Ω' est obtenu simplement à l'intersection de la ligne en pointillés et de Ω' . En s'appuyant sur le lemme 5, le coût de la solution robuste aux événements de Ω' s'obtient en résolvant le problème : $\min \{x_1 + 2x_2 \mid (\underline{a}_1 + b)x_1 + \underline{a}_2x_2 \geq 1, \underline{a}_1x_1 + (\underline{a}_2 + h)x_2 \geq 1, x \geq 0\}$. On ne détaille pas ici les calculs.

Par ailleurs, d'après la proposition 1, le sous-ensemble optimal Ω' peut être cherché convexe. Or il semble clair que tout ensemble convexe $F \subseteq \Omega$ différent de Ω' , et de mesure au moins $1 - \varepsilon = 3/4$, est tel que :

$$\max_{a \in F} V(a) > \max_{a \in \Omega'} V(a).$$

Ainsi la condition (i) est vérifiée.

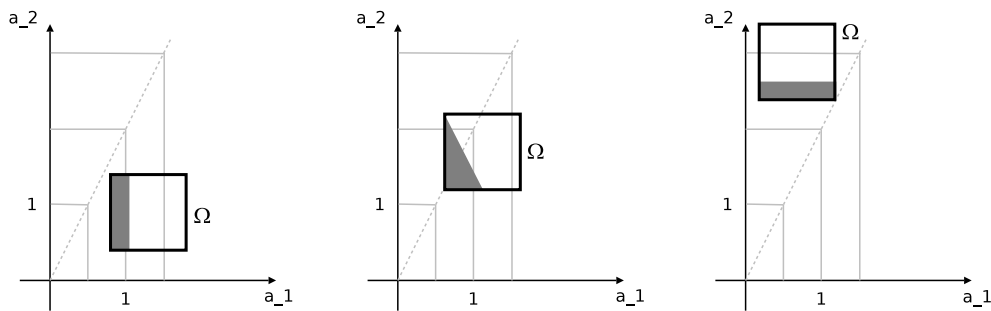


FIG. 2.4 – Trois sous-ensembles optimaux de Ω pour le problème (2.3) et $\varepsilon = 1/4$.

Chapitre 3

Utilisation du modèle robuste de Bertsimas et Sim pour la PLNE

Motivations : Bertsimas et Sim ont proposé un modèle robuste linéaire qui a acquis une légitime popularité, due d'abord à sa simplicité de mise en œuvre. Ce modèle est directement utilisable pour les problèmes linéaires en nombres entiers. Cet aspect spécifique, relativement peu exploité dans la littérature à ce jour, est le sujet de ce chapitre. On se propose en particulier de mettre en évidence quelques propriétés polyédriques de ce modèle robuste. La structure du sac-à-dos robuste, fondamentale, est plus particulièrement étudiée.

On rappelle que $I = \{1, \dots, n\}$ et $J = \{1, \dots, m\}$. Dans ce chapitre, on s'intéresse au programme linéaire en nombres entiers suivant :

$$\begin{aligned} \max \quad & cx \\ \text{s.c.} \quad & Ax \leq b, \\ & x \in \mathbb{N}^n. \end{aligned} \tag{3.1}$$

$c \in \mathbb{R}^n$ est le vecteur des coefficients de l'objectif, A est une matrice à m lignes et n colonnes, et $b \in \mathbb{R}^m$ est le vecteur des seconds membres. On suppose que seuls les coefficients de la matrice A sont incertains. Ce cadre n'est en rien restrictif. En effet, si l'on suppose que les seconds membres $\{b_j\}_{j \in J}$ sont aussi sujets à incertitudes, il suffit de rajouter une variable artificielle $x_{n+1} = 1$ pour se ramener à notre problème : pour tout $j \in J$, $A_j x \leq b_j$ devient $A_j x - b_j x_{n+1} \leq 0$. Le cas où l'objectif doit aussi être considéré incertain est traité aisément en introduisant une variable z et en écrivant : $\max\{z \mid z - cx \leq 0, Ax \leq b, x \in \mathbb{N}^n\}$.

Par ailleurs, beaucoup de résultats s'étendent directement au cas de la programmation linéaire mixte en nombres entiers, avec $x \in \mathbb{N}^r \times \mathbb{R}^{n-r}$ ($r \leq n$). Dans un souci de simplicité de l'exposé, on ne considérera ici que des variables entières.

3.1 Modèle d'incertitude et sur-conservatisme

Chaque coefficient A_{ji} est supposé pouvoir prendre ses valeurs dans un intervalle borné $[\underline{A}_{ji}, \overline{A}_{ji}] \subset \mathbb{R}$. Dans un souci de simplification, on suppose que tous les coefficients sont incertains, c'est-à-dire : $\underline{A}_{ji} < \overline{A}_{ji}$ pour tout $(i, j) \in I \times J$. On notera : $\delta_{ji} = \overline{A}_{ji} - \underline{A}_{ji} > 0$. Tous les

résultats peuvent être facilement écrits dans le cas où seulement une partie des coefficients sont incertains.

Une manière évidente de trouver une solution robuste à l'ensemble des événements pouvant survenir est de résoudre :

$$\begin{aligned} \max \quad & cx \\ \text{s.c.} \quad & \overline{A}x \leq b, \\ & x \in \mathbb{N}^n. \end{aligned}$$

On est sûr que la solution obtenue sera réalisable quoi qu'il arrive ([79] s'appuie sur cette vision de la robustesse, avec des hypothèses plus générales que celles considérées ici). Pourtant, l'objectif associé à cette solution a de fortes chances d'être relativement faible. Or la qualité médiocre de la valeur de la solution est due à la prise en compte du pire événement \overline{A} , alors que celui-ci a peut-être de très faibles chances de se produire. On parle de "sur-conservatisme" pour désigner ces situations où la solution calculée est bien robuste (conservatisme), mais en fait même trop robuste puisqu'elle impose une perte au niveau de l'objectif qui ne se justifie pas forcément.

Différents modèles ont été proposés dans la littérature pour faire face à ce problème de sur-conservatisme pour la programmation linéaire robuste [8, 13]. Bertsimas et Sim [13] ont défini une approche particulièrement intéressante pour notre contexte. Nous en détaillons ci-après les principales caractéristiques.

À chaque contrainte $j \in J$, on associe un paramètre $\Gamma_j \in \{0, \dots, n\}$ (supposé entier dans un premier temps). On cherche ensuite une solution qui reste réalisable si $n - \Gamma_j$ coefficients de la contrainte j prennent leur valeur la plus favorable (valeur minimale), tandis que les Γ_j coefficients restant prennent leur pire valeur (valeur maximale). Mathématiquement, la contrainte robuste j s'écrit donc :

$$\sum_{i \notin S} \underline{A}_{ji} x_i + \sum_{i \in S} \overline{A}_{ji} x_i \leq b_j, \quad \forall S \subseteq I \text{ t.q. } |S| = \Gamma_j. \quad (3.2)$$

On remarque par exemple que si $\Gamma_j = 0$, alors on a simplement : $\underline{A}_j x \leq b_j$. Ceci correspond au scénario le plus optimiste pour la contrainte j . À l'autre extrémité, si $\Gamma_j = n$, on obtient : $\overline{A}_j x \leq b_j$. Ceci correspond au scénario le plus pessimiste pour cette contrainte. On peut ainsi, au travers de Γ_j , paramétrer le conservatisme de la solution obtenue.

Le modèle robuste complet s'écrit donc :

$$\begin{aligned} \max \quad & cx \\ \text{s.c.} \quad & \sum_{i \notin S} \underline{A}_{ji} x_i + \sum_{i \in S} \overline{A}_{ji} x_i \leq b_j, \quad \forall S \subseteq I \text{ t.q. } |S| = \Gamma_j, \quad \forall j \in J, \\ & x \in \mathbb{N}^n. \end{aligned} \quad (3.3)$$

On peut montrer facilement que ce programme de taille exponentielle s'écrit en fait de manière équivalente (voir [13]) :

$$\begin{aligned} \max \quad & cx \\ \text{s.c.} \quad & \sum_{i \in I} \underline{A}_{ji} x_i + z_j \Gamma_j + \sum_{i \in I} p_{ji} \leq b_j, \quad \forall j \in J, \\ & z_j + p_{ji} \geq \delta_{ji} x_i, \quad \forall (i, j) \in I \times J, \\ & x_i \in \mathbb{N}, p_{ji} \geq 0, z_j \geq 0, \quad \forall (i, j) \in I \times J. \end{aligned} \quad (3.4)$$

Théorème 7 (Bertsimas et Sim, 2004) *x est une solution réalisable de (3.3) si, et seulement si, il existe z et p tels que (x, z, p) est une solution réalisable de (3.4).*

(3.4) est donc une écriture de taille polynomiale du problème robuste paramétré par Γ . On remarque à ce stade qu'il n'est pas nécessaire que les paramètres $\{\Gamma_j\}_{j \in J}$ soient entiers pour

que (3.4) ait un sens. À partir de maintenant, on supposera simplement $\Gamma_j \in [0, n]$ pour tout $j \in J$.

On a déjà mis en évidence deux raisons majeures pour s'intéresser à ce modèle robuste. La première concerne la modélisation : le paramètre Γ donne un contrôle sur la robustesse de la solution. Ceci est essentiel dans le cadre d'une réflexion sur le sur-conservatisme. La deuxième raison est plus algorithmique : le programme linéaire en nombres entiers (3.1), nominal, est transformé en un autre programme linéaire en nombres entiers (3.4), robuste, de taille raisonnable. Cette caractéristique a des conséquences pratiques importantes pour la résolution du problème robuste. On peut en effet espérer s'appuyer sur les logiciels efficaces qui existent de nos jours pour la PLNE (programmation linéaire en nombres entiers).

D'autres spécificités de cette approche robuste méritent d'être soulignées. On peut mentionner le fait que le problème robuste obtenu conserve souvent une complexité théorique comparable à celle du problème nominal correspondant. On reviendra plus loin sur cet aspect, déjà abordé dans [12] pour le cas où seuls les coefficients $\{c_i\}_{i \in I}$ de l'objectif sont incertains. Par ailleurs, Bertsimas et Sim proposent une analyse probabiliste de leur modèle robuste [13]. Ceci est particulièrement intéressant : on peut notamment faire le lien avec la programmation sous contraintes probabilistes. Cet aspect sera également abordé ultérieurement (voir chapitre 4).

Ce modèle robuste a initialement été proposé pour la programmation linéaire pure (*i.e.* avec x continu). Les auteurs ont cependant mis en évidence que leur approche s'appliquait très bien au cas de la programmation linéaire mixte en nombres entiers. Des tests numériques sur quelques problèmes combinatoires sont présentés dans [13]. À notre connaissance, les autres contributions sur l'application du modèle robuste à la PLNE se sont concentrées sur le cas où seuls les coefficients $\{c_i\}_{i \in I}$ de l'objectif sont incertains. [12] fournit des résultats de complexité. [64] s'est focalisé sur les aspects probabilistes, en fournissant des bornes plus faciles à calculer que celles de [13]. Toujours pour ce cas spécifique des coefficients $\{c_i\}_{i \in I}$, [2] étudie des reformulations plus efficaces.

De manière surprenante, à notre connaissance, aucun autre travail n'a été mené pour le cas où les coefficients de la matrice de contraintes A sont incertains. C'est ce contexte très général qui nous intéresse ici.

3.2 Complexité du problème robuste

On se restreint dans cette partie au problème robuste suivant, avec variables 0-1 :

$$\begin{aligned} \max \quad & cx \\ \text{s.c.} \quad & \sum_{i \in I} A_{ji} x_i + z_j \Gamma_j + \sum_{i \in I} p_{ji} \leq b_j, \quad \forall j \in J, \\ & z_j + p_{ji} \geq \delta_{ji} x_i, \quad \forall (i, j) \in I \times J, \\ & x_i \in \{0, 1\}, p_{ji} \geq 0, z_j \geq 0, \quad \forall (i, j) \in I \times J. \end{aligned} \tag{3.5}$$

L'analyse de complexité proposée ne semble pas pouvoir s'étendre facilement au cas des nombres entiers généraux (3.4). La seule chose sûre est que la complexité de ce dernier problème est au moins celle de (3.5). Le résultat qui suit est une généralisation de celui donné dans [12]. La preuve s'appuie sur des idées similaires à celles utilisées dans cet article.

Théorème 8 *Le problème robuste (3.5) peut être résolu via $(n+1)^m$ problèmes nominaux (3.1).*

Preuve : Il existe une solution optimale (x^*, z^*, p^*) du problème robuste (3.5) qui vérifie : $\forall (i, j) \in I \times J, p_{ji}^* = (\delta_{ji} x_i^* - z_j^*)^+$ (on rappelle que $(\cdot)^+ = \max\{\cdot, 0\}$). On en déduit que le

programme suivant fournit une solution optimale à (3.5) :

$$\begin{aligned} & \max && cx \\ \text{s.c.} & && \sum_{i \in I} \underline{A}_{ji} x_i + \Gamma_j z_j + \sum_{i \in I} (\delta_{ji} x_i - z_j)^+ \leq b_j, \quad \forall j \in J, \\ & && x \in \{0, 1\}^n, z \geq 0. \end{aligned}$$

Soit $j \in J$. On note : $f_j(z_j, x) = \Gamma_j z_j + \sum_{i \in I} (\delta_{ji} x_i - z_j)^+$ (terme de protection de la contrainte j). Comme $z_j \geq 0$ et $x_i \in \{0, 1\}$, on a : $(\delta_{ji} x_i - z_j)^+ = (\delta_{ji} - z_j)^+ x_i$. Afin d'éviter des jeux d'écriture complexes, on suppose sans perte de généralité que les valeurs $\{\delta_{ji}\}_{i \in I}$ sont triées en ordre croissant, *i.e.* : $\delta_{j1} \leq \delta_{j2} \leq \dots \leq \delta_{jn}$. Alors, si $z_j \in [\delta_{jk}, \delta_{j,k+1}[$: $\sum_{i \in I} (\delta_{ji} - z_j)^+ x_i = \sum_{i \geq k+1} (\delta_{ji} - z_j) x_i$. Ceci signifie :

$$f_j(z_j, x) = \begin{cases} \Gamma_j z_j + \sum_{i \in I} (\delta_{ji} - z_j) x_i, & \text{si } z_j < \delta_{j1}, \\ \Gamma_j z_j + \sum_{i \geq k+1} (\delta_{ji} - z_j) x_i, & \text{s'il existe } k \in I \text{ t.q. } z_j \in [\delta_{jk}, \delta_{j,k+1}[, \\ \Gamma_j z_j, & \text{si } z_j \geq \delta_{jn}. \end{cases}$$

La contrainte j s'écrit donc : $\underline{A}_j x + f_j(z_j, x) \leq b_j$. Soit $x \in \{0, 1\}^n$ quelconque. $f_j(\cdot, x)$ est continue sur \mathbb{R}^+ et linéaire sur $[0, \delta_{j1}]$, sur chaque intervalle $[\delta_{jk}, \delta_{j,k+1}]$, pour $k \in \{1, \dots, n-1\}$, et sur $[\delta_{jn}, +\infty[$. De plus, $\lim_{+\infty} f_j(\cdot, x) = +\infty$. En conséquence, $f_j(\cdot, x)$ atteint son minimum sur \mathbb{R}^+ en l'un des points de $Z_j = \{0\} \cup \{\delta_{ji}\}_{i \in I}$. (Cette conclusion est indépendante de l'ordre supposé sur les $\{\delta_{ji}\}_{i \in I}$.)

Revenons au problème global (3.5). On note V sa valeur optimale, et on introduit la valeur :

$$\begin{aligned} V(z) = & \min && cx \\ & \text{s.c.} && \underline{A}_j x + f_j(z_j, x) \leq b_j, \quad \forall j \in J, \\ & && x \in \{0, 1\}^n. \end{aligned}$$

De manière immédiate : $V = \max_{z \geq 0} V(z) \geq \max_{z \in Z_1 \times \dots \times Z_m} V(z)$. D'autre part, soit $z \geq 0$ tel que $z_j \notin Z_j$ pour un $j \in J$. Soit \bar{x} la solution associée à $V(z)$. D'après les réflexions ci-dessus, il existe $\bar{z}_j \in Z_j$ tel que $f_j(\bar{z}_j, \bar{x}) \leq f_j(z_j, \bar{x}_j)$. On en déduit qu'il existe $\bar{z} \in Z_1 \times \dots \times Z_m$ tel que $V(\bar{z}) \geq V(z)$. Ceci montre que : $\max_{z \in Z_1 \times \dots \times Z_m} V(z) \geq \max_{z \geq 0} V(z)$. On en conclut : $V = \max_{z \in Z_1 \times \dots \times Z_m} V(z)$.

Pour finir, observer que chaque problème $V(z)$ est une instance du problème nominal (3.1). Ceci achève la preuve. \square

Cette preuve montre en fait que, dans un contexte plus général, si l'on note n_j le nombre de coefficients incertains de la contrainte j ($n_j \leq n$), alors le problème robuste (3.5) peut être résolu *via* $\prod_{j \in J} (n_j + 1)$ problèmes nominaux (3.1). Dans [12], les auteurs ont supposé que seuls les coefficients de l'objectif étaient incertains. Ceci est équivalent à considérer qu'une seule contrainte du problème nominal est sujette à incertitudes. On retrouve alors leur résultat.

Le théorème 8 n'assure pas que la complexité du problème nominal est conservée dans le problème robuste. En particulier, un problème nominal polynomial peut *a priori* devenir NP-difficile dans sa version robuste. Cependant, dans certains cas où le nombre de contraintes affectées par l'incertitude est fixe, alors cette conservation de la polynomialité a lieu (avec un facteur $(n+1)^{\bar{m}}$, où \bar{m} est fixe).

3.3 Aspects polyédriques généraux

Soit une inégalité $\alpha x \leq \beta$ valide pour un polyèdre \mathcal{P} . L'ensemble $F = \{x \in \mathcal{P} \mid \alpha x = \beta\}$ est une *face* de \mathcal{P} . Cette face est propre si $F \neq \phi$ et $F \neq \mathcal{P}$. Elle est appelée *facette* si $\dim(F) =$

$\dim(\mathcal{P}) - 1$. Pour plus de détails sur la théorie des polyèdres, on renvoie le lecteur aux ouvrages classiques sur le sujet (par exemple [56]).

Dans toute cette partie, on se concentrera sur un indice $j \in J$ fixé : on étudie ainsi la transformation robuste d'une seule contrainte nominale j . Par ailleurs, on supposera que Γ_j est entier. De plus, on élimine les cas triviaux où $\Gamma_j = 0$ ou $\Gamma_j = n$. Ainsi, on supposera : $\Gamma_j \in \{1, \dots, n - 1\}$.

3.3.1 Polyèdre associé au relâché continu

On s'intéresse d'abord au polyèdre :

$$\tilde{\mathcal{P}}_j = \left\{ x \in \mathbb{R}_+^n \mid \sum_{i \notin S} \underline{A}_{ji} x_i + \sum_{i \in S} \bar{A}_{ji} x_i \leq b_j, \forall S \subseteq I \text{ t.q. } |S| = \Gamma_j \right\}.$$

Lemme 8 *Supposons que $b_j > 0$. Alors $\dim(\tilde{\mathcal{P}}_j) = n$ et pour tout $i \in I$, l'inégalité $x_i \geq 0$ définit une facette de $\tilde{\mathcal{P}}_j$.*

Preuve : Pour tout $k \in I$, on note e^k le vecteur tel que $e_i^k = 0$ si $i \neq k$, et $e_k^k = 1$. Soit $i \in I$. F_i désigne la face engendrée par l'inégalité $x_i \geq 0$: $F_i = \{x \in \tilde{\mathcal{P}}_j \mid x_i = 0\}$. Soit $k \neq i$. Si $\bar{A}_{jk} \leq 0$, on pose $x^k = e^k$. Si $\bar{A}_{jk} > 0$, on définit $x^k = b_j / \bar{A}_{jk} \cdot e^k$. Dans les deux cas, x^k appartient à F_i . Alors $\{0\} \cup \{x^k\}_{k \neq i}$ est un ensemble de n points affinement indépendants appartenant à F_i : $\dim(F_i) \geq n - 1$. Comme $\dim(F_i) < \dim(\tilde{\mathcal{P}}_j) \leq n$, on a : $\dim(\tilde{\mathcal{P}}_j) = n$ et $\dim(F_i) = n - 1$. \square

Lemme 9 *Supposons que soit $\underline{A}_j \geq 0$ et $b_j > 0$. Pour tout $S \subseteq I$ tel que $|S| = \Gamma_j$, l'inégalité triviale :*

$$\sum_{i \notin S} \underline{A}_{ji} x_i + \sum_{i \in S} \bar{A}_{ji} x_i \leq b_j \quad (3.6)$$

définit une facette de $\tilde{\mathcal{P}}_j$.

Preuve : On rappelle que d'après nos hypothèses, $\underline{A}_j \geq 0$ implique que $\bar{A}_j > 0$. On va d'abord montrer que la face engendrée par (3.6) est de dimension $n - 1$. Pour tout $k \in I$, on note e^k le vecteur tel que $e_i^k = 0$ si $i \neq k$, et $e_k^k = 1$. On obtient Γ_j points linéairement indépendants en considérant $x^k = (b_j / \bar{A}_{jk}) \cdot e^k$ pour tout $k \in S$. Ces points appartiennent à $\tilde{\mathcal{P}}_j$ et satisfont : $\sum_{i \notin S} \underline{A}_{ji} x_i^k + \sum_{i \in S} \bar{A}_{ji} x_i^k = b_j$.

Considérons le point : $\bar{x} = 1/\Gamma_j \cdot \sum_{k \in S} x^k$. Soit $k \notin S$, on construit : $x^k = \alpha_1 \bar{x} + \alpha_2 e^k$. Montrons qu'il existe $\alpha_1 > 0$ et $\alpha_2 > 0$ tels que : $x^k \in \tilde{\mathcal{P}}_j$ et x^k satisfait (3.6) à l'égalité. Cette dernière condition impose : $\alpha_1 b_j + \alpha_2 \underline{A}_{jk} = b_j$. Si $\underline{A}_{jk} = 0$, alors $\alpha_1 = 1$ et α_2 peut être choisi quelconque strictement positif. Si $\underline{A}_{jk} \neq 0$, on a : $\alpha_2 = b_j(1 - \alpha_1) / \underline{A}_{jk}$. Vérifions dans ce cas qu'il existe $\alpha_1 \in]0, 1[$ tel que $x^k \in \tilde{\mathcal{P}}_j$.

Soit $S' \subseteq I$ tel que $|S'| = \Gamma_j$, et montrons qu'il existe $\alpha_1 \in]0, 1[$ tel que : $\sum_{i \notin S'} \underline{A}_{ji} x_i^k + \sum_{i \in S'} \bar{A}_{ji} x_i^k \leq b_j$. Supposons dans un premier temps que $k \notin S'$. Alors on a : $\sum_{i \notin S'} \underline{A}_{ji} x_i^k + \sum_{i \in S'} \bar{A}_{ji} x_i^k \leq \alpha_1 b_j + \alpha_2 \cdot (\sum_{i \notin S'} \underline{A}_{ji} e_i^k + \sum_{i \in S'} \bar{A}_{ji} e_i^k) = \alpha_1 b_j + \alpha_2 \underline{A}_{jk} = b_j$, quel que soit $\alpha_1 \in]0, 1[$. Supposons maintenant que $k \in S'$. Ceci implique que $S' \neq S$; comme ces deux ensembles ont le même cardinal, il existe $l \in S \setminus S'$. Alors :

$$\begin{aligned} \sum_{i \notin S'} \underline{A}_{ji} x_i^k + \sum_{i \in S'} \bar{A}_{ji} x_i^k &\leq \alpha_1 \cdot [(\Gamma_j - 1)b_j / \Gamma_j + b_j \underline{A}_{jl} / (\bar{A}_{jl} \Gamma_j)] \\ &\quad + \alpha_2 \cdot [\sum_{i \notin S'} \underline{A}_{ji} e_i^k + \sum_{i \in S'} \bar{A}_{ji} e_i^k] \\ &= \alpha_1 \cdot [(\Gamma_j - 1)b_j / \Gamma_j + b_j \underline{A}_{jl} / (\bar{A}_{jl} \Gamma_j)] + \alpha_2 \cdot \bar{A}_{jk} \\ &= \alpha_1 \cdot [(\Gamma_j - 1)b_j / \Gamma_j + b_j \underline{A}_{jl} / (\bar{A}_{jl} \Gamma_j)] + (1 - \alpha_1) b_j \cdot \bar{A}_{jk} / \underline{A}_{jk} \end{aligned}$$

On note $Q_1 = (\Gamma_j - 1)b_j/\Gamma_j + b_j \underline{A}_{jl}/(\overline{A}_{jl}\Gamma_j)$, $Q_2 = b_j \cdot \overline{A}_{jk}/\underline{A}_{jk}$, et on souhaite assurer que : $\alpha_1 Q_1 + (1 - \alpha_1)Q_2 \leq b_j$, avec $\alpha_1 \in]0, 1[$. Comme $Q_2 > b_j$ et $Q_1 < b_j$, un tel α_1 existe.

On a travaillé avec S' fixé, et donc α_1 est une quantité positive qui dépend de S' ; on la note $\alpha_1(S')$. Le point x^k est alors défini comme : $x^k = \alpha_1^* \bar{x} + \alpha_2 e^k$, où $\alpha_1^* = \min_{S'} \alpha_1(S') > 0$.

En définitive, on a construit n points $\{x^k\}_{k \in I}$ linéairement indépendants appartenant à $\tilde{\mathcal{P}}_j$. Ceci montre que la face engendrée par (3.6) est de dimension $n - 1$, ce qui en fait une facette de $\tilde{\mathcal{P}}_j$. \square

Par des procédés similaires, on pourrait montrer :

Lemme 10 *Supposons que $\overline{A}_j \leq 0$ et $b_j < 0$. Alors $\dim(\tilde{\mathcal{P}}_j) = n$, et les inégalités (3.6) définissent des facettes de $\tilde{\mathcal{P}}_j$.*

On a supposé dans les résultats précédents que les coefficients d'une même contrainte avaient toujours tous le même signe. Si ce n'est pas le cas, l'analyse devient plus compliquée, et les inégalités (3.6) ne définissent pas toutes des facettes. C'est ce qu'illustre l'exemple suivant.

Exemple : Considérons l'inégalité : $a_1 x_1 + a_2 x_2 \leq 2$. Les coefficients a_1 et a_2 sont incertains, avec pour intervalles associés : $a_1 \in [1, 2]$, $a_2 \in [-2, -1]$. Considérons $\Gamma = 1$, le système définissant le polyèdre robuste est :

$$\begin{cases} 2x_1 - 2x_2 \leq 2 \\ x_1 - x_2 \leq 2 \end{cases} .$$

On voit que la deuxième inégalité est dominée par la première, et ne définit donc pas une facette du polyèdre robuste.

Néanmoins, la plupart des problèmes rencontrés en pratique conduiront à considérer des coefficients de même signe pour une contrainte donnée.

Lemme 11 *Supposons que soit $b_j > 0$ et $\underline{A}_j \geq 0$, soit $b_j < 0$ et $\overline{A}_j \leq 0$. Soit $\hat{x} \in \mathbb{R}_+^n$ tel que, pour tout $i \in I$:*

$$\hat{x}_i = \frac{1}{\delta_{ji}} \cdot \frac{b_j}{\sum_{k \in I} \underline{A}_{jk}/\delta_{jk} + \Gamma_j} .$$

\hat{x} est le seul point extrême de $\tilde{\mathcal{P}}_j$ tel que $\hat{x} > 0$.

Preuve : Les hypothèses considérées nous placent dans le cadre des précédents lemmes; en particulier, $\tilde{\mathcal{P}}_j$ est de dimension pleine, donc est non-vide. L'intersection des facettes définies par les inégalités (3.6) est caractérisée par :

$$\forall S \subseteq I \text{ t.q. } |S| = \Gamma_j : \sum_{i \notin S} \underline{A}_{ji} x_i + \sum_{i \in S} \overline{A}_{ji} x_i = b_j \Leftrightarrow \sum_{i \in I} \underline{A}_{ji} x_i + \sum_{i \in S} \delta_{ji} x_i = b_j .$$

Puisque $\Gamma_j \in \{1, \dots, n - 1\}$, ce système comporte au moins n équations linéairement indépendantes. Ainsi, il y a au plus une solution au système, qui est en fait donnée par les coordonnées \hat{x}_i définies dans le lemme. Si $b_j > 0$ et $\underline{A}_j \geq 0$, alors il est clair que $\hat{x} > 0$. Si $b_j < 0$ et $\overline{A}_j \leq 0$, on a pour tout $k \in I$: $\underline{A}_{jk} + \delta_{jk} \leq 0$, ce qui signifie : $\underline{A}_{jk}/\delta_{jk} \leq -1$. Alors : $\sum_{k \in I} \underline{A}_{jk}/\delta_{jk} \leq n$. Comme on a supposé $\Gamma_j \leq n - 1$, on obtient : $\sum_{k \in I} \underline{A}_{jk}/\delta_{jk} + \Gamma_j < 0$. Ceci assure que $\hat{x} > 0$.

Finalement, remarquer que les autres points extrêmes impliquent nécessairement les autres facettes $\{x \in \tilde{\mathcal{P}}_j | x_k = 0\}$, pour $k \in J$. \square

Ainsi, $\tilde{\mathcal{P}}_j$ peut être vu comme l'intersection d'une pyramide de sommet \hat{x} avec \mathbb{R}_+^n , comme illustré sur la figure 3.1.

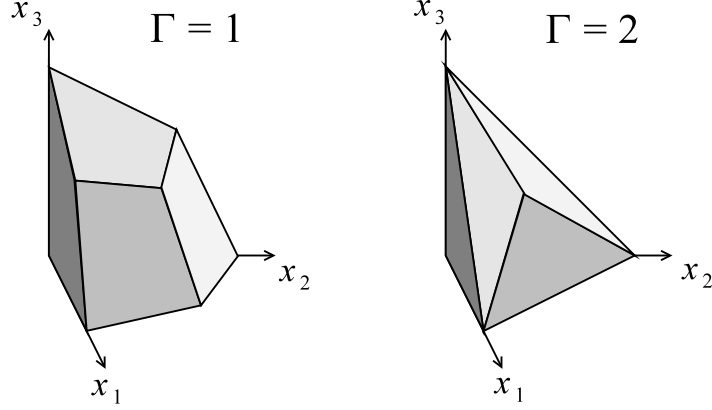


FIG. 3.1 – Polyèdres robustes des solutions réalisables ($n = 3$).

3.3.2 Polyèdre associé au problème robuste en nombres entiers

Dans cette partie, on s'intéresse au polyèdre :

$$\mathcal{P}_j = \text{conv} \left\{ x \in \mathbb{N}^n \mid \sum_{i \notin S} \underline{A}_{ji} x_i + \sum_{i \in S} \bar{A}_{ji} x_i \leq b_j, \forall S \subseteq I \text{ t.q. } |S| = \Gamma_j \right\}.$$

La description de ce polyèdre est importante en vue de la résolution de problèmes robustes en nombres entiers, notamment à l'aide d'algorithmes de plans coupants ou de *branch-and-cut*. En pratique, on a souvent des informations disponibles sur le polyèdre associé au problème nominal (3.1). Il pourrait être intéressant de ré-utiliser cette connaissance pour aider à la caractérisation de \mathcal{P}_j .

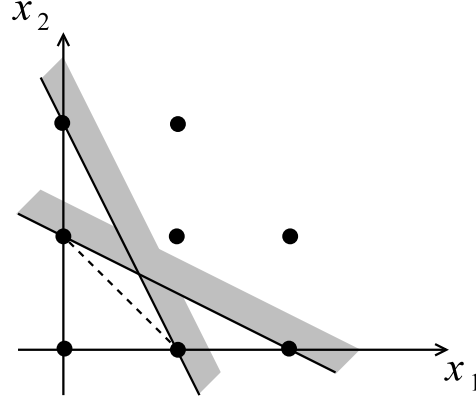
Soit $S \subseteq I$, on note $A_j(S)$ le vecteur dont les coefficients sont : $A_{ji}(S) = \bar{A}_{ji}$ si $i \in S$, $A_{ji}(S) = \underline{A}_{ji}$ si $i \notin S$. On peut déjà faire la simple observation suivante :

Lemme 12 *Soit $S \subseteq I$ tel que $|S| = \Gamma_j$. Toute inégalité valide pour $\{x \in \mathbb{N}^n \mid A_j(S).x \leq b_j\}$ est également valide pour \mathcal{P}_j .*

Ainsi, la caractérisation du polyèdre nominal associé au scénario $A_j(S)$ peut être ré-utilisée pour le problème robuste. Évidemment, les inégalités ainsi obtenues ne suffisent pas à décrire \mathcal{P}_j . On cherche dans ce qui suit à mettre en évidence des inégalités valides pour \mathcal{P}_j , mais non-valides pour les polyèdres associés aux $\{A_j(S)\}_{S \subseteq I, |S| = \Gamma_j}$.

Exemple : Considérons le polyèdre : $\{x \in \mathbb{N}^2 \mid a_1 x_1 + a_2 x_2 \leq 2\}$. Les coefficients a_1 et a_2 sont supposés prendre leurs valeurs dans l'intervalle $[1, 2]$. Prenons $\Gamma = 1$, le polyèdre robuste associé est :

$$\left\{ x \in \mathbb{N}^2 \mid 2x_1 + x_2 \leq 2 \text{ et } x_1 + 2x_2 \leq 2 \right\}.$$



La ligne en pointillés correspond à une facette du polyèdre robuste. Elle n'est évidemment valide pour aucun des ensembles $\{x \in \mathbb{N}^2 \mid 2x_1 + x_2 \leq 2\}$ et $\{x \in \mathbb{N}^2 \mid x_1 + 2x_2 \leq 2\}$.

Proposition 6 Soit $I' \subseteq I$ tel que $|I'| \leq \Gamma_j$. L'inégalité suivante est valide pour le polyèdre $\tilde{\mathcal{P}}_j$ du relâché continu :

$$\sum_{i \in I'} \bar{A}_{ji} x_i + \sum_{i \notin I'} \left(\underline{A}_{ji} + \frac{\Gamma_j - |I'|}{n - |I'|} \delta_{ji} \right) x_i \leq b_j. \quad (3.7)$$

Preuve : Soit $I' \subset I$ tel que $|I'| \leq \Gamma_j$. Supposons d'abord que $|I'| \leq \Gamma_j - 1$. Sommons toutes les contraintes robustes (3.2) telles que pour tout $i \in I'$, le coefficient de la contrainte soit \bar{A}_{ji} ; on obtient :

$$\sum_{i \in I'} C_{n-|I'|}^{\Gamma_j - |I'|} \cdot \bar{A}_{ji} x_i + \sum_{i \notin I'} \left[C_{n-|I'|}^{\Gamma_j - |I'|} \cdot \underline{A}_{ji} + C_{n-|I'| - 1}^{\Gamma_j - |I'| - 1} \cdot \delta_{ji} \right] x_i \leq C_{n-|I'|}^{\Gamma_j - |I'|} \cdot b_j.$$

En effet, il y a $C_{n-|I'|}^{\Gamma_j - |I'|}$ contraintes différentes telles que tous les coefficients $i \in I'$ sont égaux à leur valeur maximale \bar{A}_{ji} . Par ailleurs, soit $i \notin I'$: parmi les contraintes sélectionnées, il y en a $C_{n-|I'| - 1}^{\Gamma_j - |I'| - 1}$ dont le coefficient i est $\bar{A}_{ji} = \underline{A}_{ji} + \delta_{ji}$; les autres sont telles que ce coefficient est \underline{A}_{ji} . Le résultat est obtenu en divisant l'inégalité ci-dessus par $C_{n-|I'|}^{\Gamma_j - |I'|}$.

Pour finir, dans le cas où $|I'| = \Gamma_j$, l'inégalité (3.7) n'est rien d'autre qu'une contrainte robuste (3.2). \square

Les inégalités (3.7) peuvent être utilisées comme base pour construire des inégalités plus fortes, valides pour le polyèdre entier \mathcal{P}_j :

Proposition 7 Notons $\alpha = \min_{i \in I, \underline{A}_{ji} \neq 0} \{\delta_{ji} / |\underline{A}_{ji}|\}$, $\alpha > 0$:

(i) si $\underline{A}_j \in \mathbb{N}^n$, et si $\alpha \leq 1$, alors l'inégalité suivante est valide pour \mathcal{P}_j :

$$\sum_{i \in I} \underline{A}_{ji} x_i \leq \left\lfloor \frac{nb_j}{n + \alpha \Gamma_j} \right\rfloor; \quad (3.8)$$

(ii) si $(-\underline{A}_j) \in \mathbb{N}^n$, et si $\alpha \leq 1$, alors l'inégalité suivante est valide pour \mathcal{P}_j :

$$\sum_{i \in I} \underline{A}_{ji} x_i \leq \left\lfloor \frac{nb_j}{n - \alpha \Gamma_j} \right\rfloor. \quad (3.9)$$

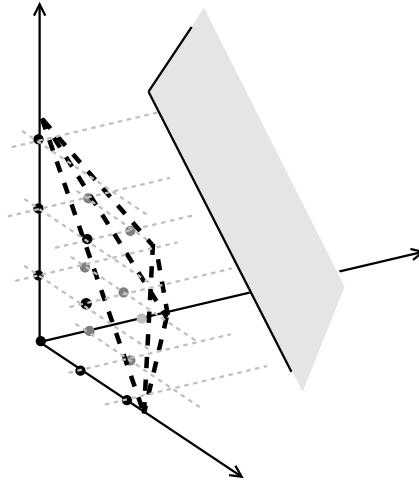
Preuve : Considérons l'inégalité (3.7) obtenue pour $I' = \phi$, on a : $\sum_{i \in I} (\underline{A}_{ji} + \delta_{ji} \cdot \Gamma_j / n) x_i \leq b_j$. Si $\underline{A}_j \in \mathbb{N}^n$, pour tout $i \in I : \delta_{ji} \geq \alpha \underline{A}_{ji}$. Alors l'inégalité suivante est aussi valide pour $\tilde{\mathcal{P}}_j : \sum_{i \in I} \underline{A}_{ji} (1 + \alpha \cdot \Gamma_j / n) x_i \leq b_j$. En divisant par cette quantité l'inégalité, et en arrondissant le membre de droite, on obtient (i).

Dans le cas où \underline{A}_j est un vecteur d'entiers négatifs, on a pour tout $i \in I : \delta_{ji} \geq -\alpha \underline{A}_{ji}$. On a alors l'inégalité suivante, valide pour $\tilde{\mathcal{P}}_j : \sum_{i \in I} \underline{A}_{ji} (1 - \Gamma_j / n \cdot \alpha) x_i \leq b_j$. Comme on suppose $\alpha \leq 1$, on a : $1 - \Gamma_j / n \cdot \alpha > 0$. On obtient alors (ii) comme précédemment. \square

Ces inégalités sont très générales. En contrepartie, elles sont le plus souvent assez faibles.

Exemple (suite) : Considérons l'exemple précédent, on a : $\alpha = 1$. D'après la proposition 7, l'inégalité suivante est valide : $2x_1 + 2x_2 \leq \lceil 4 / (2 + \alpha) \rceil = 2$. On obtient ainsi l'inégalité en pointillés sur la figure précédente.

Il semble difficile de donner des caractérisations polyédriques fortes dans un cas complètement général. Dans ce qui suit, on se concentre sur le cas particulier où toutes les données incertaines $\{A_{ji}\}_{i \in I}$ varient dans le même intervalle (mêmes valeurs minimale et maximale). On s'attachera à donner la valeur maximale que la somme $\sum_{i \in I} x_i$ peut prendre dans ce cas. Pour arriver au résultat, la démarche sera d'examiner les plans $\{x \in \mathbb{R}^n \mid \sum_{i \in I} x_i = \gamma\}$, et de voir pour quelles valeurs de γ ce plan intersecte le polyèdre \mathcal{P}_j .



Le lemme suivant est un résultat intermédiaire utile pour la preuve du théorème 9 :

Lemme 13 Soient $\bar{x} \in \mathbb{N}^*$, $c \in \{0, \dots, n\}$ et $\alpha > 0$. On note : $\mathcal{S} = \{S \subseteq I : |S| = \Gamma_j\}$, et $H = \{x \in \mathbb{N}^n \mid \sum_{i \in I} x_i = n\bar{x} - c\}$. Alors :

$$\min_{x \in H} \max_{S \in \mathcal{S}} \left\{ \sum_{i \notin S} x_i + (1 + \alpha) \sum_{i \in S} x_i \right\} \geq (n + \alpha \Gamma_j)(\bar{x} - 1) + (1 + \alpha) \min\{\Gamma_j, n - c\} + (n - c - \Gamma_j)^+$$

Preuve : Soit $x \in H$. Pour tout $S \in \mathcal{S}$, on note $g(x, S) = \sum_{i \notin S} x_i + (1 + \alpha) \sum_{i \in S} x_i$. Le maximum de $g(x, \cdot)$ est obtenu en : $S^*(x) \in \arg \max_{S \in \mathcal{S}} \sum_{i \in S} x_i$. $S^*(x)$ est simplement l'ensemble des indices correspondant aux Γ_j plus grandes valeurs de $\{x_i\}_{i \in I}$.

Il s'agit maintenant de trouver $x \in H$ tel que la somme des Γ_j plus grandes composantes de x est la plus faible possible. Ceci implique que les valeurs $\{x_i\}_{i \in I}$ sont aussi "équilibrées"

que possibles. On montre que $x \in \arg \min_H g(\cdot, S^*(\cdot))$ si les composantes de x sont à valeurs dans $\{\bar{x} - 1, \bar{x}\}$. En effet, considérons $x' \in H$ ne satisfaisant pas cette propriété. Supposons qu'il existe $k \in I$ tel que $x'_k \geq \bar{x} + 1$. Comme $x' \in H$, on a : $\sum_{i \neq k} x'_i = n\bar{x} - c - x'_k \leq (n-1)\bar{x} - c - 1 < (n-1)(\bar{x} - 1)$. Donc il existe $l \neq k$ tel que $x'_l < \bar{x} - 1$, soit : $x'_l \leq \bar{x} - 2$. On construit alors $x'' \in H$ égal à x' excepté pour les composantes k et l : $x''_k = x'_k - 1$ et $x''_l = x'_l + 1$. On voit que : $g(x'', S^*(x'')) < g(x', S^*(x'))$, donc $x' \notin \arg \min_H g(\cdot, S^*(\cdot))$. La démarche est la même dans le cas où il existe $k \in I$ tel que $x'_k \leq \bar{x} - 2$.

On a donc montré qu'un élément $x \in \arg \min_H g(\cdot, S^*(\cdot))$ a ses composantes dans $\{\bar{x} - 1, \bar{x}\}$. Alors, nécessairement, x a c composantes de valeur $\bar{x} - 1$, et $n - c$ de valeur \bar{x} . La valeur correspondante de $g(x, S^*(x))$ fournit le résultat, puisque :

$$g(x, S^*(x)) = (1 + \alpha) \left[\Gamma_j(\bar{x} - 1) + \min\{\Gamma_j, n - c\} \right] + \left[(n - \Gamma_j)(\bar{x} - 1) + (n - c - \Gamma_j)^+ \right]$$

La borne est obtenue après simplification de l'expression. \square

Théorème 9 *Supposons que $b_j > 0$, et que pour tout $i \in I$: $\underline{A}_{ji} = 1$ et $\delta_{ji} = \alpha > 0$. On note : $\bar{x} = \lceil b_j / (n + \alpha\Gamma_j) \rceil$, et : $\Delta = (n + \alpha\Gamma_j)\bar{x} - b_j$. Soit :*

$$Q = \begin{cases} \lceil \Delta \rceil, & \text{si } \Delta \leq n - \Gamma_j, \\ \left\lceil \frac{\Delta + \alpha(n - \Gamma_j)}{1 + \alpha} \right\rceil, & \text{sinon.} \end{cases}$$

Q est un entier compris entre 0 et n , et l'inégalité :

$$\sum_{i \in I} x_i \leq n\bar{x} - Q \tag{3.10}$$

définit une face propre de \mathcal{P}_j . De plus, si $Q \notin \{0, n\}$, c'est une facette de \mathcal{P}_j .

Preuve : On a montré dans la proposition 6 que : $\sum_{i \in I} x_i \leq nb_j / (n + \alpha\Gamma_j)$ est valide pour $\tilde{\mathcal{P}}_j$ (considérer $I' = \emptyset$), donc pour \mathcal{P}_j . Ceci assure que $\sum_{i \in I} x_i \leq n\bar{x}$ est bien valide pour \mathcal{P}_j . Notre but est de trouver un entier $c \in \mathbb{N}$ aussi grand que possible de manière à ce que $\sum_{i \in I} x_i \leq n\bar{x} - c$ reste valide pour \mathcal{P}_j . Observer que, si l'on augmente progressivement c à partir de $c = 0$, l'inégalité est valide tant que l'hyperplan $\{x \in \mathbb{R}^n \mid \sum_{i \in I} x_i = n\bar{x} - c\}$ ne "rencontre" pas \mathcal{P}_j , i.e. ne contient pas de point de ce polyèdre. Dit autrement, on cherche c aussi petit que possible de manière à ce que cet hyperplan contienne au moins un point de \mathcal{P}_j .

Soit $x \in \mathbb{N}^n$ tel que : $\sum_{i \in I} x_i = n\bar{x} - c$. D'après le lemme 13, il existe $S \subset I$ tel que $|S| = \Gamma_j$ et :

$$\sum_{i \notin S} x_i + (1 + \alpha) \sum_{i \in S} x_i \geq (n + \alpha\Gamma_j)(\bar{x} - 1) + (1 + \alpha) \min\{\Gamma_j, n - c\} + (n - c - \Gamma_j)^+$$

On s'intéresse à la différence entre cette borne et la quantité b_j , noté $f(c)$:

$$f(c) = (n + \alpha\Gamma_j)(\bar{x} - 1) + (1 + \alpha) \min\{\Gamma_j, n - c\} + (n - c - \Gamma_j)^+ - b_j$$

Si $c \leq n - \Gamma_j$, on a : $f(c) = (n + \alpha\Gamma_j)\bar{x} - b_j - c = \Delta - c$. Si $c \geq n - \Gamma_j$, on a : $f(c) = (n + \alpha\Gamma_j)\bar{x} - b_j + \alpha(n - \Gamma_j) - (1 + \alpha)c = \Delta + \alpha(n - \Gamma_j) - (1 + \alpha)c$. La fonction f est donc décroissante et s'annule pour une valeur, notée c^* , qui vaut Δ si $\Delta \leq n - \Gamma_j$, et $(\Delta + \alpha(n - \Gamma_j)) / (1 + \alpha)$ sinon. Noter que $c^* \geq 0$. De plus, en remarquant que $\Delta \leq n + \alpha\Gamma_j$, on vérifie que $c^* \leq n$.

Quand $c < c^*$, alors : $\sum_{i \notin S} x_i + (1 + \alpha) \sum_{i \in S} x_i > b_j$. Ceci prouve qu'il n'y a aucun point entier de l'hyperplan $\{x \in \mathbb{R}^n \mid \sum_{i \in I} x_i = n\bar{x} - c\}$ appartenant à \mathcal{P}_j . Puisqu'on cherche c entier, la condition $c < c^*$ s'écrit de manière équivalente : $c < \lceil c^* \rceil$.

Considérons maintenant le cas $c = \lceil c^* \rceil = Q$, et construisons un point entier $x \in \mathbb{N}^n$ tel que $\sum_{i \in I} x_i = n\bar{x} - Q$ et $x \in \mathcal{P}_j$. Le point x est défini par :

$$\begin{cases} x_i = \bar{x} - 1, & \text{si } i \leq Q, \\ x_i = \bar{x}, & \text{sinon.} \end{cases}$$

On vérifie bien que $\sum_{i \in I} x_i = n\bar{x} - Q$. Soit $S \subset I$ tel que $|S| = \Gamma_j$. La quantité $\sum_{i \notin S} x_i + (1 + \alpha) \sum_{i \in S} x_i$ est maximale lorsque S contient un maximum d'éléments i de valeur \bar{x} . On montre ainsi que : $\sum_{i \notin S} x_i + (1 + \alpha) \sum_{i \in S} x_i \leq f(Q) + b_j$ (en effet, en reprenant les notations de la preuve du lemme 13, ceci est équivalent à : $g(x, S) \leq g(x, S^*(x))$). De plus, comme $Q \geq c^*$ et f est décroissante, $f(Q) \leq 0$. On a donc : $\sum_{i \notin S} x_i + (1 + \alpha) \sum_{i \in S} x_i \leq b_j$. Ainsi : $x \in \mathcal{P}_j$. Ceci montre que l'inégalité (3.10) définit une face propre de \mathcal{P}_j .

Enfin, si $Q \notin \{0, n\}$, on peut définir n points linéairement indépendants sur cette face : il suffit de faire varier l'ensemble, de taille Q , des éléments i tels que $x_i = \bar{x} - 1$. On montre donc que la face propre est une facette de \mathcal{P}_j . \square

Le théorème 9 s'étend immédiatement aux cas où les valeurs nominales sont toutes égales et strictement positives : $\forall i \in I, \underline{A}_{ji} = a > 0$, et si les variations satisfont la même propriété : $\forall i \in I, \delta_{ji} = d$. L'inégalité (3.10) peut aussi fournir des inégalités valides si ces propriétés ne sont pas respectées. Par exemple, si les coefficients de \underline{A}_j sont strictement positifs, considérer le problème relâché avec, pour tout i , $\underline{A}'_{ji} = \min_{k \in I} \underline{A}_{jk}$, et $\alpha = \min_{i \in I} \delta_{ji} / \underline{A}'_{ji}$. Mais on n'a plus de garantie sur la qualité des inégalités obtenues.

Le même type de résultat s'obtient dans le cas où on travaille avec des coefficients négatifs :

Théorème 10 *Supposons que $b_j < 0$, et que pour tout $i \in I$: $\underline{A}_{ji} = -1$ et $\delta_{ji} = \alpha \in]0, 1]$. On note : $\underline{x} = \lfloor -b_j / (n - \alpha\Gamma_j) \rfloor$, et : $\Delta = -(n - \alpha\Gamma_j)\underline{x} - b_j$. Soit :*

$$Q = \begin{cases} \left\lceil \frac{\Delta}{1 - \alpha} \right\rceil, & \text{si } \alpha \neq 1 \text{ et } \Delta \leq (1 - \alpha)\Gamma_j, \\ \lceil \Delta + \alpha\Gamma_j \rceil, & \text{si } \Delta > (1 - \alpha)\Gamma_j, \\ 0, & \text{si } \alpha = 1 \text{ et } \Delta = 0. \end{cases}$$

Q est un entier entre 0 et n , et l'inégalité :

$$\sum_{i \in I} x_i \geq n\underline{x} + Q \tag{3.11}$$

définit une face propre de \mathcal{P}_j . De plus, si $Q \notin \{0, n\}$, c'est une facette de \mathcal{P}_j .

Idée de preuve : La preuve est complètement similaire à celle du théorème 9. En notant $H = \{x \in \mathbb{N}^n \mid \sum_{i \in I} x_i = n\bar{x} + c\}$, il faut d'abord observer que :

$$\max_{x \in H} \min_{S \in \mathcal{S}} \left\{ \sum_{i \notin S} x_i + (1 - \alpha) \sum_{i \in S} x_i \right\} \leq f(c) - b_j$$

avec : $f(c) = (1 - \alpha) \left[\Gamma_j \underline{x} + \min\{\Gamma_j, c\} \right] + (n - \Gamma_j)\underline{x} + (c - \Gamma_j)^+ + b_j$. La fonction f est croissante ; il faut en trouver le zéro c^* sur \mathbb{R}_+ . Ce zéro est tel que $\lceil c^* \rceil = Q$. On montre enfin que $\{x \in \mathbb{N}^n \mid \sum_{i \in I} x_i = n\bar{x} + Q\}$ contient des points de \mathcal{P}_j . \square

On donne pour finir le même type de résultat pour le cas où la valeur minimale de chaque coefficient incertain est nulle.

Théorème 11 *Supposons que $b_j > 0$, et que pour tout $i \in I : \underline{A}_{ji} = 0$ et $\delta_{ji} = \alpha > 0$. On note : $\bar{x} = \lceil b_j / (\Gamma_j \alpha) \rceil$, et :*

$$Q = \lceil \Gamma_j (\bar{x} - 1) + n - b_j / \alpha \rceil$$

Q est un entier entre $n - \Gamma_j$ et n , et l'inégalité :

$$\sum_{i \in I} x_i \leq n \bar{x} - Q \quad (3.12)$$

définit une face propre de \mathcal{P}_j . De plus, si $Q \neq n$, c'est une facette de \mathcal{P}_j .

Idée de preuve : Comme précédemment, on note $H = \{x \in \mathbb{N}^n \mid \sum_{i \in I} x_i = n \bar{x} - c\}$ et on montre que :

$$\min_{x \in H} \max_{S \subseteq S} \left\{ \alpha \sum_{i \in S} x_i \right\} \geq \alpha \Gamma_j (\bar{x} - 1) + \alpha \min\{\Gamma_j, n - c\}.$$

On cherche alors le zéro de la fonction décroissante : $f(c) = \alpha \Gamma_j (\bar{x} - 1) + \alpha \min\{\Gamma_j, n - c\} - b_j$. Cette fonction s'annule en $c^* = \Gamma_j (\bar{x} - 1) + n - b_j / \alpha \in [n - \Gamma_j, n]$. On a : $Q = \lceil c^* \rceil$. On montre enfin que $\{x \in \mathbb{N}^n \mid \sum_{i \in I} x_i = n \bar{x} - Q\}$ contient des points de \mathcal{P}_j . \square

Dans les trois théorèmes qui précèdent, on suppose toujours que les valeurs minimales et maximales ont toutes le même signe (soit toutes positives, soit toutes négatives). En effet, comme déjà remarqué dans la partie 3.3.1, la structure du polyèdre peut être très différente dans le cas où ce signe peut changer (voir l'exemple suivant le lemme 10).

Exemple : Considérons deux sous-réseaux liés entre eux par $n = 5$ liens. Chaque lien peut être équipé d'un ou plusieurs modules de capacité $c = 25$. On veut assurer une connectivité avec une capacité au moins $K = 100$ entre les deux sous-réseaux. Cette propriété doit en particulier être garantie pour toute panne simultanée de $\Gamma = 2$ liens. L'objectif est d'acheter le nombre minimum de modules pour équiper l'ensemble des liens.

On veut donc résoudre :

$$\begin{aligned} \min \quad & \sum_{i=1}^5 x_i \\ \text{s.c.} \quad & \min_{S:|S|=3} \sum_{i \in S} c x_i \geq K, \quad \Leftrightarrow \quad \min \quad \sum_{i=1}^5 x_i \\ & x \in \mathbb{N}^5 \quad \quad \quad \text{s.c.} \quad \max_{S:|S|=3} \sum_{i \in S} -x_i \leq -K/c, \\ & \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad x \in \mathbb{N}^5. \end{aligned}$$

On voit qu'on rentre dans le cadre du théorème 10, avec $\alpha = 1$ et $\Gamma = 2$. On calcule facilement : $\underline{x} = 1$, $\Delta = 1$ et $Q = 3$. On en déduit la valeur optimale du problème : $\sum_{i=1}^5 x_i = 8$. Ceci signifie qu'il faudra investir 8 modules sur l'ensemble des 5 liens.

3.3.3 Autres inégalités

Dans ce qui suit, on note : $\alpha = \min_{i \in I} \delta_{ji} / \underline{A}_{ji}$.

Proposition 8 *Soit $q = \lfloor n / \Gamma \rfloor$. Quel que soit $p \in \{1, \dots, q\}$, et pour tout $R \subseteq I$ tel que $|R| = n - p \Gamma$:*

- l'inégalité suivante est valide pour $\tilde{\mathcal{P}}_j$:

$$p \sum_{i \in R} \underline{A}_{ji} x_i + \sum_{i \notin R} [p \underline{A}_{ji} + \delta_{ji}] x_i \leq p b_j; \quad (3.13)$$

- si $p + \alpha > 0$ et $\underline{A}_j \in \mathbb{Z}^n$, l'inégalité suivante est valide pour \mathcal{P}_j :

$$\sum_{i \in R} \lfloor p \underline{A}_{ji} / (p + \alpha) \rfloor x_i + \sum_{i \notin R} \underline{A}_{ji} x_i \leq \lfloor p b_j / (p + \alpha) \rfloor. \quad (3.14)$$

Preuve : On peut trouver $\{S_1, \dots, S_p\} \subseteq I$ tel que $|S_k| = \Gamma$ pour tout $k \in \{1, \dots, p\}$, et : $k_1 \neq k_2 \Rightarrow S_{k_1} \cap S_{k_2} = \emptyset$. Alors, en sommant les p inégalités (3.2) correspondant aux ensembles S_k , on obtient la première inégalité. La deuxième est obtenue simplement par une procédure d'arrondi de Chvátal-Gomory. \square

Si $p = q = 1$, c'est-à-dire si $\Gamma > n/2$, les inégalités (3.13) sont simplement les inégalités de base (3.2). Si Γ divise n et $p = q$, les inégalités (3.14) sont équivalentes à celles des propositions 6 et 7 pour $I' = \emptyset$. Observer enfin que ces inégalités (3.14) sont séparables en temps $O(n \log(n))$. En effet, étant donnée une solution fractionnaire \tilde{x} , l'inégalité la plus violée est obtenue pour un ensemble R de cardinal $r = n - p\Gamma$ et de poids minimal, en considérant les poids $\{\lfloor q \underline{A}_{ji} / (q + \alpha) \rfloor \tilde{x}_i\}_{i \in I}$. Il suffit donc de trier ces poids en ordre croissant, et de sélectionner les r plus petits.

Avec des idées similaires, on obtient aussi les inégalités suivantes :

Proposition 9 Soit $q = \lfloor n / (n - \Gamma) \rfloor$. Quel que soit $p \in \{1, \dots, q\}$, et pour tout $R \subseteq I$ tel que $|R| = n - p(n - \Gamma)$:

- l'inégalité suivante est valide pour $\tilde{\mathcal{P}}_j$:

$$p \sum_{i \in R} (\underline{A}_{ji} + \delta_{ji}) x_i + \sum_{i \notin R} (p \underline{A}_{ji} + (p - 1) \delta_{ji}) x_i \leq p b_j; \quad (3.15)$$

- si $p + \alpha(p - 1) > 0$ et $\underline{A}_j \in \mathbb{Z}^n$, l'inégalité suivante est valide pour \mathcal{P}_j :

$$\sum_{i \in R} \left\lfloor \underline{A}_{ji} \cdot \frac{1 + \alpha}{1 + \alpha(1 - 1/p)} \right\rfloor x_i + \sum_{i \notin R} \underline{A}_{ji} x_i \leq \left\lfloor \frac{b_j}{1 + \alpha(1 - 1/p)} \right\rfloor. \quad (3.16)$$

Preuve : On peut trouver des sous-ensembles $\{\bar{S}_1, \dots, \bar{S}_p\} \subseteq I$ tels que $|\bar{S}_k| = n - \Gamma$ pour tout k , et : $k_1 \neq k_2 \Rightarrow \bar{S}_{k_1} \cap \bar{S}_{k_2} = \emptyset$. Alors, en sommant les p inégalités de base (3.2) correspondant aux ensembles $S_k = I \setminus \bar{S}_k$, on obtient l'inégalité (3.15). (3.16) en découle en divisant par $p + \alpha(p - 1)$ et en arrondissant les coefficients. \square

Observer que dans la proposition ci-dessus, si $p = q = 1$ (i.e. si $\Gamma < n/2$), alors $|R| = \Gamma$ et (3.15) est équivalent aux inégalités (3.2) de la formulation initiale. Dans ce cas, l'arrondi effectué pour (3.16) ne change pas l'inégalité. Ces inégalités n'ont donc d'intérêt que si $\Gamma \geq n/2$. Enfin, on peut vérifier que si $n - \Gamma$ est un diviseur de n , alors (3.16) revient aux inégalités de la proposition 7 pour $I' = \emptyset$.

Remarque : Les inégalités données ci-dessus pour \mathcal{P}_j sont obtenues en arrondissant les coefficients fractionnaires susceptibles d'apparaître. Cette procédure peut classiquement être améliorée par l'utilisation d'une fonction super-additive appropriée (cf. e.g. [56] ; voir annexe A).

3.4 Le problème de sac-à-dos robuste 0-1

La programmation linéaire en nombres 0-1 est un cas particulier fréquent en pratique de la programmation linéaire en nombres entiers. Il est donc important de donner des résultats

spécifiques à ce contexte. On se concentre sur le problème de sac-à-dos, dont le polyèdre s'écrit classiquement :

$$\mathcal{K} = \text{conv} \left\{ x \in \{0, 1\}^n \mid \sum_{i \in I} w_i x_i \leq c \right\} \quad (3.17)$$

où $w_i > 0$ est le poids de l'élément $i \in I$, et $c > 0$ est la capacité du sac. Si l'on considère un vecteur $p \in \mathbb{R}_+^n$ de profits, le problème de sac-à-dos est : $\max\{px \mid x \in \mathcal{K}\}$.

Il faut remarquer que les contraintes de signe sur les poids et sur la capacité du sac ne sont pas vraiment restrictives. En effet, si $w_i < 0$, alors on peut écrire : $w_i x_i = (-w_i)(1 - x_i) + w_i$. La variable $x'_i = 1 - x_i$ est encore 0-1, et elle est associée à un coefficient positif $(-w_i)$. Ainsi, l'étude d'une contrainte de sac-à-dos permet en fait une caractérisation de tout type d'inégalité dans un problème 0-1.

Pour une étude en profondeur du problème de sac-à-dos et de ses variantes, on renvoie par exemple à [52, 47]. Le problème est connu pour être NP-difficile au sens faible : il existe un algorithme de résolution par programmation dynamique, de complexité pseudo-polynomiale. Les approches fondées sur une analyse polyédrique de \mathcal{K} ne sont pas, en général, les plus efficaces en pratiques pour ce problème. Pourtant, puisque les contraintes de sac-à-dos apparaissent dans de nombreux autres problèmes plus complexes (par exemple avec des contraintes de capacité dans un réseau), il est important de pouvoir caractériser \mathcal{K} pour mettre en œuvre des algorithmes de *branch-and-cut* efficaces.

Comme précédemment, nous considérerons que les coefficients $\{w_i\}_{i \in I}$ sont tous incertains : pour tout $i \in I$, $w_i \in [\underline{w}_i, \overline{w}_i]$. Les données \underline{w} et \overline{w} sont supposées entières : $\underline{w} \in \mathbb{N}^n$ et $\overline{w} \in \mathbb{N}^n$. Pour $\Gamma \in \{0, \dots, n\}$, le polyèdre robuste est :

$$\mathcal{K}(\Gamma) = \text{conv} \left\{ x \in \{0, 1\}^n \mid \forall S \subseteq I \text{ t.q. } |S| = \Gamma : \sum_{i \notin S} \underline{w}_i x_i + \sum_{i \in S} \overline{w}_i x_i \leq c \right\}. \quad (3.18)$$

Observer que $\mathcal{K}(\Gamma)$ est un polyèdre de sac-à-dos multi-dimensionnel, de structure très particulière. Pour tout $S \subseteq I$, on note :

$$\mathcal{K}_S = \left\{ x \in \{0, 1\}^n \mid \sum_{i \in S} \overline{w}_i x_i + \sum_{i \notin S} \underline{w}_i x_i \leq c \right\}. \quad (3.19)$$

On a bien sûr, comme déjà dit auparavant :

$$\mathcal{K}(\Gamma) = \text{conv} \left(\bigcap_{S \subseteq I, |S| = \Gamma} \mathcal{K}_S \right). \quad (3.20)$$

Ainsi, toute inégalité valide pour les polyèdres de sac-à-dos classiques \mathcal{K}_S seront valides aussi pour $\mathcal{K}(\Gamma)$. Notre objectif est de fournir des caractérisations plus spécifiques de $\mathcal{K}(\Gamma)$. En particulier, on montre que dans le cadre du modèle robuste étudié, plusieurs résultats classiques pour le sac-à-dos s'étendent assez directement.

3.4.1 Étude polyédrique

À partir de la caractérisation (3.18), on déduit facilement :

Lemme 14 *Soit $\Gamma \geq 1$, $\mathcal{K}(\Gamma)$ est de dimension pleine si et seulement si, pour tout $i \in I$, $\overline{w}_i \leq c$.*

À partir de maintenant, on suppose qu'aucun des poids $\{\bar{w}_i\}_{i \in I}$ n'excède la capacité c .

Quand on s'intéresse au polyèdre du sac-à-dos classique, les inégalités valides les plus fréquemment utilisées sont les inégalités de couvertures. Un ensemble $\mathcal{C} \subseteq I$ est appelé *couverture* si : $\sum_{i \in \mathcal{C}} w_i > c$. Dans ce cas, l'inégalité : $\sum_{i \in \mathcal{C}} x_i \leq |\mathcal{C}| - 1$ est valide pour \mathcal{K} . L'idée très simple est que les éléments d'une couverture ne peuvent pas être présents tous simultanément dans le sac. Une couverture \mathcal{C} est dite *minimale* si pour tout $i \in \mathcal{C}$, $\mathcal{C} \setminus \{i\}$ n'est pas une couverture. Clairement, les inégalités obtenues à partir des couvertures minimales dominent les autres.

Une inégalité de couverture peut être "liftée" séquentiellement, mais ce processus requiert la résolution de nouveaux problèmes de sac-à-dos [37, 38, 56]. Il existe pourtant une manière très simple de renforcer les inégalités de couverture. Étant donnée une couverture \mathcal{C} , on appelle couverture *étendue* l'ensemble : $E(\mathcal{C}) = \mathcal{C} \cup \{k \in I \mid w_k \geq \max_{i \in \mathcal{C}} w_i\}$. L'inégalité suivante est alors valide pour \mathcal{K} : $\sum_{i \in E(\mathcal{C})} x_i \leq |\mathcal{C}| - 1$. Ces inégalités de couvertures étendues sont assez fortes, et définissent souvent des facettes de \mathcal{K} (voir e.g. [56]). Les tests numériques de [32] montrent de plus que ces inégalités sont très utiles en pratique pour la résolution de problèmes de sac-à-dos multi-dimensionnels.

Ces concepts classiques peuvent être adaptés à notre contexte robuste :

Définition 6 *Un ensemble $\mathcal{C} \subseteq I$ est appelé couverture robuste s'il existe $S \subseteq \mathcal{C}$ tel que $|S| = \min\{\Gamma, |\mathcal{C}|\}$ et :*

$$\sum_{i \in \mathcal{C} \setminus S} \underline{w}_i + \sum_{i \in S} \bar{w}_i > c.$$

Définition 7 *Une couverture robuste $\mathcal{C} \subseteq I$ est dite minimale si pour tout $i \in \mathcal{C}$, $\mathcal{C} \setminus \{i\}$ n'est pas une couverture robuste.*

Pour toute couverture robuste \mathcal{C} , l'inégalité suivante est valide pour $\mathcal{K}(\Gamma)$:

$$\sum_{i \in \mathcal{C}} x_i \leq |\mathcal{C}| - 1.$$

Ceci est clair, puisque cette inégalité est valide pour au moins un des ensembles \mathcal{K}_S (cf. (3.20)). Une couverture robuste \mathcal{C} peut être étendue en $E(\mathcal{C})$ défini par :

$$E(\mathcal{C}) = \begin{cases} \mathcal{C} \cup \{i \in I \mid \bar{w}_i \geq \max_{k \in \mathcal{C}} \bar{w}_k\}, & \text{si } |\mathcal{C}| \leq \Gamma \\ \mathcal{C} \cup \{i \in I \mid \underline{w}_i \geq \max_{k \in \mathcal{C}} \underline{w}_k, \text{ et } : \bar{w}_i \geq \max_{k \in \mathcal{C}} \bar{w}_k\}, & \text{si } |\mathcal{C}| \geq \Gamma + 1 \end{cases} \quad (3.21)$$

Proposition 10 *Soient \mathcal{C} une couverture robuste et $E(\mathcal{C})$ son extension, l'inégalité suivante est valide pour $\mathcal{K}(\Gamma)$:*

$$\sum_{i \in E(\mathcal{C})} x_i \leq |\mathcal{C}| - 1. \quad (3.22)$$

Preuve : Montrons que tout $\mathcal{C}' \subseteq E(\mathcal{C})$ tel que $|\mathcal{C}'| = |\mathcal{C}|$ est une couverture robuste. Soit $k \in E(\mathcal{C}) \setminus \mathcal{C}$. Observer que pour tout $i \in \mathcal{C}$, l'ensemble $(\mathcal{C} \setminus \{i\}) \cup \{k\}$ est une couverture robuste : cela suit directement de (3.21) et de la définition 6. Ainsi, \mathcal{C}' peut être obtenu à partir de \mathcal{C} en effectuant une succession de remplacements d'éléments de \mathcal{C} par des éléments de $E(\mathcal{C}) \setminus \mathcal{C}$. Ceci montre que \mathcal{C}' est une couverture robuste.

Supposons maintenant qu'il existe un point $x \in \mathcal{K}(\Gamma) \cap \{0, 1\}^n$ tel que $\sum_{i \in E(\mathcal{C})} x_i \geq |\mathcal{C}|$. Considérons un vecteur $\bar{x} \in \{0, 1\}^n$ tel que $\bar{x} \leq x$ et $\sum_{i \in E(\mathcal{C})} \bar{x}_i = |\mathcal{C}|$: $\bar{x} \in \mathcal{K}(\Gamma)$, puisque

$x \in \mathcal{K}(\Gamma)$. Soit enfin \mathcal{C}' l'ensemble dont le vecteur caractéristique est \bar{x} , *i.e.* : $i \in \mathcal{C}' \Leftrightarrow \bar{x}_i = 1$. Comme $|\mathcal{C}'| = |\mathcal{C}|$, \mathcal{C}' est une couverture robuste, et en conséquence : $\sum_{i \in \mathcal{C}'} x_i \leq |\mathcal{C}| - 1$ (contradiction). \square

Alors que les inégalités de couvertures robustes sont valides pour certains ensembles \mathcal{K}_S , ce n'est plus le cas avec des couvertures étendues. Les exemples suivants illustrent ce fait, en utilisant les deux types d'extension définis par (3.21).

Exemple 1 : Soit un problème de sac-à-dos robuste à $n = 3$ éléments de poids respectifs $w_1 \in [10, 12]$, $w_2 \in [11, 13]$ et $w_3 \in [12, 14]$. Soit $c = 24$ la capacité du sac. Avec $\Gamma = 2$, le polyèdre robuste est l'enveloppe convexe de l'ensemble suivant :

$$\begin{cases} 10x_1 + 13x_2 + 14x_3 \leq 24, & (i) \\ 12x_1 + 11x_2 + 14x_3 \leq 24, & (ii) \\ 12x_1 + 13x_2 + 12x_3 \leq 24, & (iii) \\ x_1, x_2, x_3 \in \{0, 1\}. \end{cases}$$

$\mathcal{C} = \{1, 2\}$ est une couverture robuste. En observant que $|\mathcal{C}| = \Gamma$, cette couverture peut être étendue en $E(\mathcal{C}) = \{1, 2, 3\}$. Ceci implique que l'inégalité $x_1 + x_2 + x_3 \leq 1$ est valide pour le problème robuste. Noter cependant qu'elle n'est satisfaite pour aucun des sac-à-dos classiques associés respectivement aux inégalités (i), (ii) et (iii).

Exemple 2 : Soit un autre problème de sac-à-dos robuste à 3 éléments, avec pour poids respectifs $w_1 \in [10, 13]$, $w_2 \in [11, 14]$ et $w_3 \in [12, 15]$. Soit $c = 23$ la capacité du sac. Avec $\Gamma = 1$, le polyèdre robuste est l'enveloppe convexe de l'ensemble :

$$\begin{cases} 13x_1 + 11x_2 + 12x_3 \leq 23, & (i) \\ 10x_1 + 14x_2 + 12x_3 \leq 23, & (ii) \\ 10x_1 + 11x_2 + 15x_3 \leq 23, & (iii) \\ x_1, x_2, x_3 \in \{0, 1\}. \end{cases}$$

$\mathcal{C} = \{1, 2\}$ est une couverture robuste. Puisque $|\mathcal{C}| = \Gamma + 1$, son extension est : $E(\mathcal{C}) = \{1, 2, 3\}$. Ceci conduit à l'inégalité : $x_1 + x_2 + x_3 \leq 1$. Bien que valide pour le problème robuste global, cette inégalité n'est valide pour aucun des sac-à-dos classiques associés respectivement à (i), (ii) et (iii).

On a ainsi montré que la procédure d'extension décrite conduira vraisemblablement à des inégalités qui ne seront valides pour aucun des ensembles \mathcal{K}_S , $|S| = \Gamma$. C'est une indication de leur intérêt pratique éventuel. Pourtant, cette observation n'est plus valable pour les extensions $E(\mathcal{C})$ de cardinal inférieur à Γ :

Lemme 15 *Soit \mathcal{C} une couverture robuste. Si $|E(\mathcal{C})| \leq \Gamma$, alors il existe $S \subseteq I$, avec $|S| = \Gamma$, tel que l'inégalité (3.22) est valide pour \mathcal{K}_S .*

Preuve : Soit $S \supseteq E(\mathcal{C})$ de cardinal Γ . \mathcal{C} est une couverture classique pour l'ensemble \mathcal{K}_S , et de plus $|\mathcal{C}| \leq \Gamma$. Donc, d'après (3.21), et puisque $E(\mathcal{C}) \subseteq S$, $E(\mathcal{C})$ est une couverture étendue classique pour le polyèdre $\text{conv}(\mathcal{K}_S)$. \square

Cette remarque apporte un bémol quant à la force supposée des couvertures robustes étendues. Cependant, la plupart des résultats polyédriques classiques disponibles pour les inégalités de couverture étendues (voir e.g. [56]) peuvent être transposés au contexte robuste qui est le nôtre :

Proposition 11 *Supposons que l'ensemble I est ordonné de telle sorte que $i < j \Rightarrow \begin{cases} \underline{w}_i \geq \underline{w}_j \\ \overline{w}_i \geq \overline{w}_j \end{cases}$. Soit $\mathcal{C} = \{i_1, \dots, i_r\}$ une couverture robuste minimale, avec $i_1 < i_2 < \dots < i_r$. Si l'une des conditions suivantes est réalisée, alors (3.22) définit une facette de $\mathcal{K}(\Gamma)$:*

- (i) $\mathcal{C} = I$;
- (ii) $E(\mathcal{C}) = I$, et $(\mathcal{C} \setminus \{i_1, i_2\}) \cup \{1\}$ n'est pas une couverture robuste ;
- (iii) $E(\mathcal{C}) = \mathcal{C}$, et $(\mathcal{C} \setminus \{i_1\}) \cup \{p\}$ n'est pas une couverture robuste, où $p = \min\{i \in I \setminus \mathcal{C}\}$;
- (iv) $\mathcal{C} \subset E(\mathcal{C}) \subset I$, et ni $(\mathcal{C} \setminus \{i_1, i_2\}) \cup \{1\}$, ni $(\mathcal{C} \setminus \{i_1\}) \cup \{p\}$, avec $p = \min\{i \in I \setminus \mathcal{C}\}$, ne sont des couvertures robustes.

Preuve : Soit $U \subseteq I$, on note $x^U \in \{0, 1\}^n$ le vecteur caractéristique de U : $x_i^U = 1 \Leftrightarrow i \in U$. Pour toute couverture robuste $\mathcal{C} = \{i_1, \dots, i_r\}$, considérons les ensembles $\{U_j\}_{j \in I}$ suivants :

- a. pour tout $j \in \mathcal{C}$, $U_j = \mathcal{C} \setminus \{j\}$,
- b. pour tout $j \in E(\mathcal{C}) \setminus \mathcal{C}$, $U_j = (\mathcal{C} \setminus \{i_1, i_2\}) \cup \{j\}$,
- c. et pour tout $j \in I \setminus E(\mathcal{C})$, $U_j = (\mathcal{C} \setminus \{i_1\}) \cup \{j\}$.

Quel que soit $j \in I$, observer que : $\sum_{i \in E(\mathcal{C})} x_i^{U_j} = |E(\mathcal{C}) \cap U_j| = |\mathcal{C}| - 1$. De plus, les points $\{x^{U_j}\}_{j \in I}$ sont linéairement indépendants. Vérifions que ces points appartiennent à $\mathcal{K}(\Gamma)$.

Dans le cas (i), les ensembles de a. fournissent $|\mathcal{C}| = n$ points de $\mathcal{K}(\Gamma)$, puisque \mathcal{C} est une couverture robuste minimale. Sous la condition (ii), puisque $(\mathcal{C} \setminus \{i_1, i_2\}) \cup \{1\}$ n'est pas une couverture robuste, pour tout $j \notin \mathcal{C}$, U_j n'est pas une couverture robuste. En effet, $\underline{w}_j \leq \underline{w}_1$ et $\overline{w}_j \leq \overline{w}_1$. Ainsi, les ensembles a. et b. fournissent $|E(\mathcal{C})| = n$ points de $\mathcal{K}(\Gamma)$.

Pour le cas (iii), pour tout $j \notin E(\mathcal{C})$, les ensembles U_j ne sont pas des couvertures robustes. En effet, comme précédemment, on a : $\underline{w}_j \leq \underline{w}_p$ et $\overline{w}_j \leq \overline{w}_p$. Comme $E(\mathcal{C}) = \mathcal{C}$, les ensembles définis par a. et c. correspondent bien à n points de $\mathcal{K}(\Gamma)$. Enfin, dans le cas (iv), les points correspondant aux ensembles a., b. et c. appartiennent tous à $\mathcal{K}(\Gamma)$. \square

La condition sur l'ordre que doivent respecter les poids de I est restrictive. On donne ci-après une caractérisation polyédrique valide dans le cas général.

Proposition 12 *Supposons (sans perte de généralité) que l'ensemble I est ordonné de telle sorte que $i < j \Rightarrow \overline{w}_i \geq \overline{w}_j$. Soit $\mathcal{C} = \{i_1, \dots, i_r\}$ une couverture robuste minimale, avec $i_1 < i_2 < \dots < i_r$. Si l'une des conditions suivantes est réalisée, alors (3.22) définit une facette de $\mathcal{K}(\Gamma)$:*

- (i) $\mathcal{C} = I$;
- (ii) $E(\mathcal{C}) = I$, $|\mathcal{C}| \leq \Gamma + 1$, et $(\mathcal{C} \setminus \{i_1, i_2\}) \cup \{1\}$ n'est pas une couverture robuste ;
- (iii) $E(\mathcal{C}) = \mathcal{C}$, $|\mathcal{C}| \leq \Gamma$, et $(\mathcal{C} \setminus \{i_1\}) \cup \{p\}$ n'est pas une couverture robuste, où $p = \min\{i \in I \setminus \mathcal{C}\}$;
- (iv) $\mathcal{C} \subset E(\mathcal{C}) \subset I$, $|\mathcal{C}| \leq \Gamma$, et ni $(\mathcal{C} \setminus \{i_1, i_2\}) \cup \{1\}$, ni $(\mathcal{C} \setminus \{i_1\}) \cup \{p\}$, avec $p = \min\{i \in I \setminus \mathcal{C}\}$, ne sont des couvertures robustes.

Preuve : On considère à nouveau les ensembles a., b. et c. introduits dans la preuve précédente. Comme précédemment, les points $\{x^{U_j}\}_{j \in I}$ sont linéairement indépendants et vérifient : $\sum_{i \in E(\mathcal{C})} x_i^{U_j} = |\mathcal{C}| - 1$.

Dans le cas (i), les points associés aux ensembles a. suffisent. Dans le cas (ii), $(\mathcal{C} \setminus \{i_1, i_2\}) \cup \{1\}$ n'est pas une couverture robuste. Comme cet ensemble est de cardinal inférieur à Γ , cela signifie : $\sum_{i \in (\mathcal{C} \setminus \{i_1, i_2\}) \cup \{1\}} \overline{w}_i \leq c$. Ceci implique que pour tout $j \notin \mathcal{C}$: $\sum_{i \in U_j} \overline{w}_i \leq c$. Ceci montre que U_j n'est pas une couverture robuste. Ainsi, les $n - |\mathcal{C}|$ points définis à l'aide des ensembles b. sont dans $\mathcal{K}(\Gamma)$. En ajoutant les points issus des ensembles a., on obtient n points linéairement indépendants de $\mathcal{K}(\Gamma)$.

Considérons le cas (iii). Soit $j \notin E(\mathcal{C})$, l'ensemble correspondant de c . satisfait : $|U_j| \leq \Gamma$. De plus, selon l'ordre supposé sur les poids maximaux : $\sum_{i \in U_j} \bar{w}_i \leq \sum_{i \in \mathcal{C} \setminus \{i_1\}} \bar{w}_i + \bar{w}_p$. Puisque $(\mathcal{C} \setminus \{i_1\}) \cup \{p\}$ n'est pas une couverture robuste, on en déduit : $\sum_{i \in U_j} \bar{w}_i \leq c$. U_j n'est donc pas une couverture robuste. Ainsi, les ensembles c . fournissent $n - |\mathcal{C}|$ points de $\mathcal{K}(\Gamma)$, auxquels on ajoute les points a . pour obtenir n points linéairement indépendants de $\mathcal{K}(\Gamma)$. Finalement, le cas (iv) peut être traité de manière complètement similaire : les ensembles associés à a ., b . et c . sont tous dans $\mathcal{K}(\Gamma)$. \square

Observons qu'en plus de fournir des inégalités valides fortes pour le problème robuste, le cadre théorique développé permet de les caractériser d'une manière globale. Ainsi, il n'est pas nécessaire d'énumérer les ensembles \mathcal{K}_S à la recherche d'inégalités valides. Ceci est essentiel en pratique.

3.4.2 Séparation des inégalités de couvertures robustes

Dans le cas du polyèdre de sac-à-dos classique \mathcal{K} , la séparation des inégalités de couverture est également un problème de sac-à-dos. Notons $\tilde{\mathcal{K}} = \{x \in [0, 1]^n \mid \sum_{i \in I} w_i x_i \leq c\}$ le polyèdre du relâché continu. Pour tout point fractionnaire $\tilde{x} \in \tilde{\mathcal{K}}$, une inégalité de couverture parmi les plus violées est obtenue en résolvant le problème suivant :

$$\begin{aligned} \min \quad & \sum_{i \in I} (1 - \tilde{x}_i) r_i \\ \text{s.c.} \quad & \sum_{i \in I} w_i r_i \geq c + 1, \\ & r \in \{0, 1\}^n. \end{aligned}$$

Si la valeur de ce problème est strictement inférieure à 1, alors une inégalité violée est trouvée. Sinon, il n'y a pas d'inégalité de couverture violée. Par ailleurs, si pour tout $i \in I$, $\tilde{x}_i < 1$, la couverture obtenue est minimale.

Ce problème de séparation peut être adapté au contexte robuste pour le polyèdre $\mathcal{K}(\Gamma)$. Soit $\tilde{\mathcal{K}}(\Gamma)$ le polyèdre du relâché linéaire de $\mathcal{K}(\Gamma)$. Étant donné un point fractionnaire $\tilde{x} \in \tilde{\mathcal{K}}(\Gamma)$, une inégalité de couverture robuste parmi les plus violées peut être générée en résolvant :

$$\begin{aligned} \min \quad & \sum_{i \in I} (1 - \tilde{x}_i) r_i \\ \text{s.c.} \quad & \sum_{i \in I} [\underline{w}_i + (\bar{w}_i - \underline{w}_i) s_i] r_i \geq c + 1, \\ & \sum_{i \in I} s_i \leq \Gamma, \\ & r \in \{0, 1\}^n, s \in \{0, 1\}^n. \end{aligned} \tag{3.23}$$

Ce programme mathématique peut être linéarisé :

$$\begin{aligned} \min \quad & \sum_{i \in I} (1 - \tilde{x}_i) r_i \\ \text{s.c.} \quad & \sum_{i \in I} [\underline{w}_i r_i + (\bar{w}_i - \underline{w}_i) t_i] \geq c + 1, \\ & \sum_{i \in I} s_i \leq \Gamma, \\ & t_i \leq r_i, \quad \forall i \in I, \\ & t_i \leq s_i, \quad \forall i \in I, \\ & r \in \{0, 1\}^n, s \in [0, 1]^n, t \geq 0. \end{aligned} \tag{3.24}$$

Lemme 16 *Il existe s tel que (r, s) est une solution optimale de (3.23) si, et seulement si, il existe (\tilde{s}, \tilde{t}) tel que $(r, \tilde{s}, \tilde{t})$ est une solution optimale de (3.24).*

Preuve : On montre dans un premier temps que pour toute solution réalisable $(r, \tilde{s}, \tilde{t})$ de (3.24), il existe (s', t') tel que (r, s', t') est réalisable également pour (3.24), avec $s' \in \{0, 1\}^n$. Soit $(r, \tilde{s}, \tilde{t})$ réalisable pour (3.24). On note $T \subseteq I$ l'ensemble des indices correspondant aux

Γ plus grands éléments de $\{(\bar{w}_i - \underline{w}_i)r_i\}_{i \in I}$. On définit $s' \in \{0, 1\}^n$ par : $s'_i = 1$ si $i \in T$, $s'_i = 0$ sinon. On pose également, pour tout $i \in I$: $t'_i = r_i s'_i = \min\{r_i, s'_i\}$. On a : $\sum_{i \in I} (\bar{w}_i - \underline{w}_i)t'_i = \sum_{i \in I} (\bar{w}_i - \underline{w}_i)r_i s'_i \geq \sum_{i \in I} (\bar{w}_i - \underline{w}_i)r_i \tilde{t}_i$. En effet, s' caractérise les Γ plus grands éléments de $\{(\bar{w}_i - \underline{w}_i)r_i\}_{i \in I}$, et $\sum_{i \in I} \tilde{t}_i \leq \Gamma$. Alors, en observant que $r_i \tilde{t}_i = \tilde{t}_i$, on obtient : $\sum_{i \in I} (\bar{w}_i - \underline{w}_i)t'_i \geq \sum_{i \in I} (\bar{w}_i - \underline{w}_i)\tilde{t}_i$. Ceci implique que (r, s', t') est réalisable pour (3.24).

Montrons maintenant qu'il existe s tel que (r, s) est une solution réalisable de (3.23) si, et seulement si, il existe (\tilde{s}, \tilde{t}) tel que $(r, \tilde{s}, \tilde{t})$ est une solution réalisable de (3.24). Soit (r, s) une solution réalisable de (3.23), alors \tilde{t} défini par $\tilde{t}_i = r_i s_i$ est tel que (r, s, \tilde{t}) est une solution réalisable de (3.24) ($\tilde{s} = s$). Réciproquement, soit $(r, \tilde{s}, \tilde{t})$ une solution réalisable de (3.24). D'après ce qui a été montré plus haut, il existe (s', t') tel que $s' \in \{0, 1\}^n$ et (r, s', t') est réalisable pour (3.24). On peut en outre supposer sans perte de généralité que pour tout $i \in I$: $t'_i = r_i s'_i = \min\{r_i, s'_i\}$. On en déduit que (r, s') est réalisable pour (3.23).

En définitive, comme les deux problèmes (3.23) et (3.24) ont la même fonction objectif, on a le résultat. \square

La valeur optimale de (3.24) est strictement inférieure à 1 si et seulement si une inégalité de couverture robuste violée existe. Dans ce cas, le vecteur de solution r caractérise la couverture correspondante. Ce problème de séparation est NP-difficile, puisque $\Gamma = 0$ nous ramène au sac-à-dos classique. Comme dans le cas des inégalités de couvertures classiques, on a :

Lemme 17 *Si pour tout $i \in I$, $\tilde{x}_i < 1$, alors la couverture robuste produite par le programme (3.24) est minimale.*

En pratique, le problème (3.24) reste encore difficile à résoudre de manière exacte. On propose l'heuristique suivante, qui est une adaptation naturelle de l'algorithme glouton utilisé pour le sac-à-dos classique (cf. [52]).

Heuristique pour la séparation des inégalités de couvertures robustes

Étape 0 : Soient $V = 0$, $\mathcal{C} = \phi$.

Pour tout $i \in I$, soient $\alpha_i = (1 - \tilde{x}_i)/\bar{w}_i$ et $\beta_i = (1 - \tilde{x}_i)/\underline{w}_i$.

Étape 1 : Pour tout $k \in \{1, \dots, \Gamma\}$, soit $L(k)$ l'indice du k^{eme} plus petit coefficient de $\{\alpha_i\}_{i \in I}$.

Pour $k = 1$ à $k = \Gamma$:

$\mathcal{C} \leftarrow \mathcal{C} \cup \{L(k)\}$

$V \leftarrow V + \bar{w}_{L(k)}$

si $V > c$, STOP : \mathcal{C} est une couverture robuste

$k \leftarrow k + 1$

Étape 2 : Pour tout $k \in \{1, \dots, n - \Gamma\}$, soit $L'(k)$ l'indice du k^{eme} plus petit coefficient de $\{\beta_i\}_{i \notin \mathcal{C}}$.

Pour $k = 1$ à $k = n - \Gamma$:

$\mathcal{C} \leftarrow \mathcal{C} \cup \{L'(k)\}$

$V \leftarrow V + \underline{w}_{L'(k)}$

si $V > c$, STOP : \mathcal{C} est une couverture robuste

$k \leftarrow k + 1$

3.4.3 Incertitude sur la capacité du sac

Jusqu'ici, on a supposé que la capacité c du sac était connue avec précision. Dans ce paragraphe, on suppose que ce n'est plus le cas : $c \in [\underline{c}, \bar{c}]$, avec $\bar{c} > \underline{c}$. On se ramène au cas étudié

jusqu'ici en supposant que la capacité du sac est fixée à \bar{c} , et en ajoutant un nouvel élément (artificiel) de poids w_{n+1} compris entre $\underline{w}_{n+1} = 0$ et $\bar{w}_{n+1} = \bar{c} - \underline{c}$. Ce nouvel élément a pour seul fin de modéliser l'incertitude sur la capacité du sac.

Plus précisément, soit $I' = I \cup \{n+1\}$. Pour un paramètre de robustesse $\Gamma \in \{0, \dots, n+1\}$, un remplissage $x \in \{0, 1\}^n$ sera dit réalisable s'il existe $x_{n+1} \in \{0, 1\}$ tel que :

$$\forall S \subseteq I' \text{ t.q. } |S| = \Gamma : \sum_{i \in I' \setminus S} \underline{w}_i x_i + \sum_{i \in S} \bar{w}_i x_i \leq \bar{c}.$$

L'analyse polyédrique produite plus haut s'applique directement ici.

3.4.4 Tests numériques

Les inégalités de couvertures robustes présentées ont été testées numériquement sur de nombreuses instances de sac-à-dos robuste. Ces instances sont générées aléatoirement selon les règles suivantes :

- pour tout $i \in I$, le poids minimal \underline{w}_i et le profit p_i associé sont choisis entiers dans l'intervalle $[100, 1000]$;
- chaque incertitude sur les poids est d'environ 10% la valeur minimale : $\bar{w}_i = \underline{w}_i + \lceil \underline{w}_i/10 \rceil$;
- la capacité c est choisie entière entre les valeurs $1/3 \cdot \sum_{i \in I} \underline{w}_i$ et $2/3 \cdot \sum_{i \in I} \underline{w}_i$.

Aucune incertitude n'est considérée sur c . L'implémentation s'appuie sur la ré-écriture suivante du problème robuste (cf. [13] ; voir aussi la partie 3.1 et le modèle (3.4)) :

$$\mathcal{K}(\Gamma) = \text{conv} \left\{ x \in \{0, 1\}^n \mid \exists u \geq 0, v \geq 0 \text{ s.t. } \begin{array}{l} \sum_{i \in I} \bar{w}_i x_i + \sum_{i \in I} u_i + \Gamma v \leq c \\ \forall i \in I, u_i + v \geq \hat{w}_i x_i \end{array} \right\}.$$

Tous les programmes mathématiques sont résolus à l'aide de Cplex 10.0, sur un ordinateur équipé d'un processeur Intel(R) Xeon(TM) à 2,8 GHz et de 2 Go de RAM. Toutes les procédures de génération automatique d'inégalités valides de Cplex sont désactivées.

Sac-à-dos classique : Avant de décrire les tests sur le modèle robuste, il convient de regarder l'impact des inégalités de couvertures classiques, avec ou sans extensions, sur le problème de sac-à-dos. On rappelle que les inégalités de couvertures ont été très largement utilisées en pratique pour améliorer la résolution de nombreux problèmes appliqués (tels que multiflots monoroutés, localisations avec capacités, etc.). Dans ce paragraphe, on considère temporairement le problème de sac-à-dos classique :

$$\max_{x \in \{0, 1\}} \left\{ \sum_{i \in I} p_i x_i \mid \sum_{i \in I} w_i x_i \leq c \right\}.$$

Des instances sont générées aléatoirement selon les règles suivantes, qui sont cohérentes avec celles déjà énoncées pour le problème robuste :

- pour tout $i \in I$, le poids w_i et le profit p_i sont des entiers aléatoirement choisis dans $[100, 1000]$;
- la capacité c est un entier choisi entre $1/3 \cdot \sum_{i \in I} w_i$ et $2/3 \cdot \sum_{i \in I} w_i$.

Dix instances de 100 éléments sont considérées. Les inégalités de couverture sont générées de manière systématique à tous les nœuds de calcul, en utilisant pour leur séparation l'algorithme glouton classique. Les résultats moyens observés sont donnés dans le tableau 3.1. Les entêtes de colonnes ont la signification suivante : NN désigne le nombre de nœuds de l'arbre de branchement, T est le temps total en secondes ; NC est le nombre total d'inégalités ajoutées au modèle,

aucune inégalité		couvertures			couvertures étendues			
NN	T	NC	NN	T	NC	NE	NN	T
50,3	0,01	26,2	38,6	0,05	18,4	17,7	27,3	0,03

TAB. 3.1 – Tests des inégalités de couvertures sur des instances de sac-à-dos classique ($n = 100$).

tandis que NE précise le nombre d’inégalités qui ont fait l’objet d’une procédure d’extension. L’usage de ces inégalités réduit le nombre de nœuds de calcul, mais augmente le temps total de résolution.

Sac-à-dos robuste : Les problèmes de sac-à-dos robuste sont résolus en générant les inégalités de manière “agressive” : dès qu’une inégalité de couverture robuste violée est trouvée, on l’ajoute au problème. La séparation est effectuée de manière approchée à l’aide de l’heuristique gloutonne présentée dans la partie 3.4.2. On se concentre sur des instances à $n = 100$ éléments. Plusieurs valeurs du paramètre Γ sont considérées ; pour chaque valeur de Γ , dix instances sont résolues, et les résultats moyens observés sont reportés dans le tableau 3.2. Comme précédemment, NN désigne le nombre de nœuds de l’arbre de branchement, T est le temps total en secondes ; NC est le nombre total d’inégalités ajoutées au modèle, tandis que NE précise le nombre d’inégalités qui ont fait l’objet d’une procédure d’extension ; finalement, ST est le temps consacré à la séparation des inégalités de couvertures robustes (en secondes). Comme on s’y attendait, le plus souvent, la taille de l’arbre de branchement est diminuée par l’ajout d’inégalités, et l’utilisation d’extensions des couvertures robustes tend à améliorer encore cette tendance. Noter que les extensions concernent une proportion importante des inégalités générées. L’impact des coupes sur le nombre de nœuds explorés est plus facile à analyser à partir de la figure 3.2. Pourtant, dans certains cas particuliers, l’ajout d’inégalités de couvertures robustes peut n’apporter aucune amélioration dans le nombre de nœuds (voir le cas $\Gamma = 30$), ou même entraîner une augmentation de ce dernier (voir le cas $\Gamma = 40$). Cette observation, surprenante de prime abord, montre que l’ajout de coupes peut “perturber” le logiciel de résolution, certainement en modifiant la manière de brancher à un nœud de calcul.

L’ajout d’inégalités de couvertures robustes entraîne l’augmentation des temps de calcul. Le temps total est approximativement doublé si on le compare au cas standard où aucune inégalité n’est ajoutée (simple algorithme de *branch-and-bound*). Il faut observer que cette augmentation du temps de calcul n’est pas liée à la procédure de séparation, qui s’avère extrêmement rapide (voir les colonnes ST du tableau 3.2). Puisque le nombre global de nœuds de branchement est diminué, on en déduit que l’augmentation globale du temps de calcul est la conséquence de accroissement de la taille des programmes linéaires à résoudre à chaque nœud, du fait des inégalités ajoutées. Ces observations concernant le temps de calcul ne sont pas surprenantes : c’est un comportement couramment observé lorsqu’on utilise directement des inégalités de couverture pour la résolution d’un problème de sac-à-dos classique (voir plus haut). On peut donc toujours espérer que les inégalités de couvertures robustes proposées seront utiles pour diminuer le temps de résolution de gros problèmes appliqués, comme les inégalités de couvertures classiques le font effectivement pour des problèmes sans incertitude.

Comme souligné auparavant, l’impact plus ou moins positif de l’ajout d’inégalités de couvertures robustes dépend de la valeur de Γ . Pour notre ensemble d’instances, elles sont clairement plus efficaces quand $\Gamma \geq 50$. La même observation vaut lorsque les coupes sont étendues. Une explication de ce comportement peut venir du nombre d’éléments présents dans un remplissage optimal du sac-à-dos, qui dépend directement du rapport entre la valeur de la capacité

Γ	aucune inégalité		couvertures robustes				couvertures robustes étendues				
	NN	T	NC	NN	T	ST	NC	NE	NN	T	ST
10	113,5	0,05	20,6	111	0,13	0,00	17,5	17	111,5	0,13	0,01
20	163,9	0,06	17,4	153,2	0,16	0,01	14,5	14,3	152,6	0,16	0,01
30	138	0,05	11,4	138,1	0,14	0,01	9,3	8,8	137,1	0,14	0,01
40	108,4	0,04	9,9	111,2	0,12	0,00	9,7	9,5	113,1	0,12	0,00
50	128,2	0,05	28,1	120,5	0,15	0,00	25	23,7	114,9	0,14	0,01
60	89,4	0,04	24,4	68,3	0,11	0,01	23,4	21,5	66,8	0,1	0,01
70	97,6	0,04	56,4	73,6	0,15	0,01	52	41,6	69,6	0,14	0,01
80	96,3	0,04	56,4	63	0,14	0,02	51,5	40,4	54,2	0,13	0,01
90	95,4	0,04	57	63,4	0,14	0,01	49,6	39,3	52,8	0,12	0,01

TAB. 3.2 – Tests sur des instances de sac-à-dos robuste, $n = 100$.

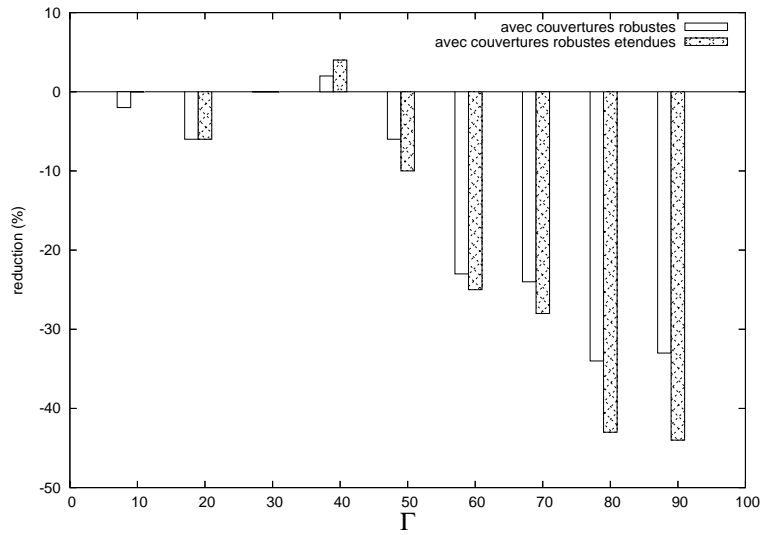


FIG. 3.2 – Impact de l'ajout d'inégalités sur le nombre de nœuds de calcul.

Γ	aucune inégalité		couvertures robustes				couvertures robustes étendues				
	NN	T	NC	NN	T	ST	NC	NE	NN	T	ST
10	53,5	0,03	8,1	34,2	0,05	0,03	7,3	3,7	34,1	0,05	0,03
20	56,8	0,03	3,1	35,2	0,05	0,03	3,1	2,3	35,1	0,05	0,03
30	86,1	0,04	1,5	66	0,07	0,04	1,5	0,9	65,7	0,07	0,04
40	59,2	0,04	0,8	51,3	0,07	0,03	0,8	0,5	51,3	0,07	0,03
50	122,1	0,06	3,2	106,7	0,12	0,07	3,2	1,6	106,7	0,12	0,07
60	92	0,05	6,2	74,1	0,09	0,06	6,4	4,2	89,1	0,11	0,06
70	74,3	0,04	16,5	75,2	0,11	0,06	15,5	9,8	76,5	0,11	0,06
80	72,3	0,03	27,6	36,3	0,09	0,04	25,7	18,9	32,8	0,08	0,04
90	74	0,03	29,8	35,9	0,09	0,05	25,5	20	29,5	0,08	0,03

TAB. 3.3 – Tests sur des instances de sac-à-dos robustes, $n = 100$, avec une capacité plus grande.

et les valeurs des poids. En effet, on rappelle que la capacité est choisie entre $1/3 \cdot \sum_{i \in I} \underline{w}_i$ et $2/3 \cdot \sum_{i \in I} \underline{w}_i$. Donc, en moyenne, $c \approx 1/2 \cdot \sum_{i \in I} \bar{w}_i$. Comme on considère des instances à $n = 100$ éléments on peut s'attendre à ce qu'un remplissage du sac implique environ $1/2 \cdot 100 = 50$ éléments. Quand $\Gamma \geq 50$, il est donc probable qu'en fait, tous les éléments sont traités comme s'ils prenaient uniquement leur poids maximal \bar{w}_i , ce qui n'est pas le cas pour des valeurs plus faibles du paramètre de robustesse.

Pour illustrer et confirmer cette intuition, une nouvelle série de tests a été effectuée. Les instances sont définies comme précédemment, à l'exception de la capacité des sac-à-dos, choisies entre $7/10 \cdot \sum_{i \in I} \underline{w}_i$ et $8/10 \cdot \sum_{i \in I} \underline{w}_i$. Les résultats apparaissent dans le tableau 3.3 et sur la figure 3.3. Comme pour la série de tests précédente, l'impact des coupes dépend de la valeur de Γ . Ici, $\Gamma = 70$ est la valeur critique du paramètre de robustesse, au lieu de $\Gamma = 50$ auparavant. Si l'on suit le raisonnement du paragraphe précédent, cette valeur est cohérente avec la nouvelle manière de définir la capacité c , entre 70% et 80% de $\sum_{i \in I} \underline{w}_i$. L'intuition développée plus haut sur le lien entre la valeur de la capacité (et le nombre d'éléments impliqués dans un remplissage) et l'efficacité des coupes semble donc correcte.

Par ailleurs, cette deuxième série de tests appelle un autre commentaire relatif aux extensions. Alors qu'elles sont toujours réalisées sur une proportion importante des inégalités générées (voir tableau 3.3, colonne NE), elles n'impliquent pas une réduction significative du nombre de nœuds de calcul lorsque Γ est inférieur à la valeur critique 70.

Finalement, en considérant à nouveau la première série de tests (avec capacité choisie entre $1/3 \cdot \sum_{i \in I} \underline{w}_i$ et $2/3 \cdot \sum_{i \in I} \underline{w}_i$), on a aussi évalué l'impact des coupes sur la relaxation continue du problème. La réduction de l'écart entre la valeur optimale (entière) du problème et la valeur de relaxation a été mesurée au nœud racine de l'arbre de branchement. Les inégalités de couvertures robustes sont séparées de manière exacte à l'aide du modèle (3.24) ; tant qu'on trouve des coupes violées, on les ajoute. Les coupes s'avèrent plus efficaces pour renforcer la relaxation linéaire quand $\Gamma < 30$ ou $\Gamma > 50$. Cette observation est cohérente avec les réflexions faites plus haut.

3.5 Conclusion

Dans ce chapitre, on s'est concentré sur l'utilisation du modèle de robustesse proposé par Bertsimas et Sim [12, 13] pour les problèmes linéaires en nombres entiers. On s'est particulièrement intéressé aux aspects polyédriques du modèle robuste. Dans une première partie, le cas de la programmation en nombres entiers généraux est étudié. Des inégalités très générales

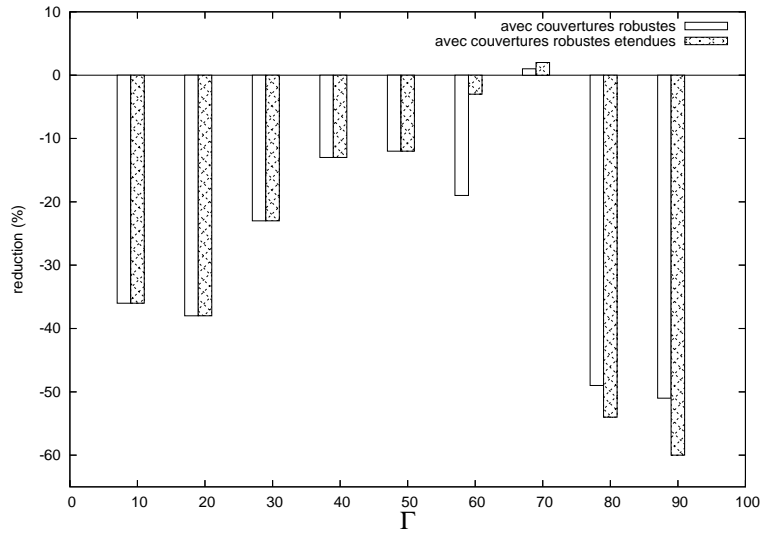


FIG. 3.3 – Impact de l’ajout de coupes sur le nombre de nœuds de calcul, pour une capacité plus grande.

Γ	couvertures robustes		couvertures robustes étendues		
	NC	réduction de l’écart	NC	NE	réduction de l’écart
10	2,6	8,31%	3,4	3,3	11,73%
20	0,8	2,72%	0,9	0,9	2,76%
30	0,6	0,65%	0,6	0,6	0,65%
40	0,3	0,09%	0,3	0,3	0,09%
50	0,5	1,05%	0,5	0,4	1,05%
60	3,4	8,44%	3,3	3,2	8,69%
70	7	12,02%	6,8	6,6	12,91%
80	7,8	17,19%	7,5	7,1	17,43%
90	7,8	17,19%	7,5	7,1	17,43%

TAB. 3.4 – Réduction de l’écart entre optimum et relaxation au nœud racine.

sont proposées, certaines définissant des facettes dans des cas particuliers. Dans un deuxième temps, l'étude s'est concentrée sur la programmation en nombres entiers 0-1, et en particulier sur le problème de sac-à-dos robuste. On a montré comment les inégalités de couvertures, très largement utilisées pour les problèmes sans incertitudes, peuvent être adaptées au contexte robuste. Sous certaines hypothèses, ces inégalités de couvertures robustes définissent des facettes du polyèdre associé aux solutions entières. Des algorithmes exacts et approchés sont donnés pour la séparation de ces inégalités. Des tests numériques montrent que de telles coupes ont un impact similaire dans le contexte robuste à celui des inégalités de couvertures dans le cas classique. Ces tests numériques ont aussi mis en évidence le lien existant entre la valeur du paramètre de robustesse Γ et l'impact des coupes ajoutées sur le processus de résolution.

Ces travaux montrent que l'approche robuste de Bertsimas et Sim se prête bien à des analyses polyédriques dans le cadre de problèmes en nombres entiers. En particulier, il semble que ce modèle assure la conservation de plusieurs propriétés du problème nominal correspondant. Il est donc probable qu'une étude spécifique des problèmes pratiques rencontrés mène rapidement à une ré-utilisation efficace des résultats classiques déjà disponibles.

Chapitre 4

Une approche robuste pour le problème de sac-à-dos sous contrainte probabiliste

Motivations : On a vu dans le chapitre précédent comment utiliser le modèle robuste de Bertsimas et Sim pour prendre en compte des données incertaines dans les programmes linéaires en nombres entiers. Le lien entre optimisation robuste et programmation sous contraintes probabilistes a par ailleurs été souligné dans le chapitre 2. Il est donc naturel de voir comment le modèle de Bertsimas et Sim pourrait être utilisé pour la résolution de problèmes combinatoires sous contraintes probabilistes. Ce chapitre traite de cette question pour le cas spécifique du sac-à-dos.

4.1 Travaux antérieurs sur la programmation sous contraintes probabilistes avec nombres entiers

La programmation sous contraintes probabilistes a été présentée au chapitre 1, partie 1.2.3 (voir e.g. [67, 45]). Selon cette approche, on cherche la meilleure solution réalisable avec une probabilité au moins $1 - \varepsilon$, où $\varepsilon \in]0, 1[$. L'intérêt et la difficulté de ce modèle ont déjà été discutés au chapitre 1. Il semble que relativement peu de travaux antérieurs ont abordé des problèmes combinatoires sous contraintes probabilistes. La complexité des contraintes probabilistes avec variables continues rend d'autant plus improbable la résolution exacte de tels problèmes avec des variables entières, sauf dans des cas très particuliers. Par exemple, [44] fournit une étude intéressante d'un problème d'arbre couvrant stochastique. On cherche à minimiser le coût maximal parmi les arêtes utilisées par l'arbre. Les coûts sont supposés incertains, et cette valeur du coût maximal d'arête est alors évaluée de manière à être satisfaite avec une probabilité au moins $\alpha \in [1/2, 1[$. Mais ce travail est restreint au cas spécifique de variables aléatoires indépendantes et normalement distribuées. Ces hypothèses sont d'ailleurs les plus fréquentes dans la littérature, parce qu'elles permettent la construction de modèles équivalents déterministes [75, 84, 16, 45] (voir aussi la partie 6.2.1).

Le lien entre optimisation robuste et contraintes probabilistes a été mis en évidence dans la partie 1.2.3. Par ailleurs, plusieurs publications récentes en ont tiré parti. Les modèles robustes présentés dans [8] et [13] sont accompagnés d'analyses probabilistes. Dans chaque approche, les auteurs montrent qu'une solution du problème robuste est réalisable avec une probabilité

que l'on sait borner. On peut donc *a priori* se servir de ces modèles pour obtenir des solutions réalisables pour des problèmes sous contraintes probabilistes. [24, 58] mettent en évidence ce lien entre modèles robustes et programmation sous contraintes probabilistes. [58] propose une autre approximation, complexe, des contraintes probabilistes ; elle améliore le travail antérieur de [8]. Un travail de même teneur est proposé par [7] pour le contexte des inégalités linéaires de matrices (*linear matrix inequalities*), qui généralisent notamment la programmation linéaire.

L'approche de [19, 57] est différente : les auteurs étudient la construction aléatoire d'un ensemble d'incertitude, pour lequel une solution robuste sera réalisable avec probabilité au moins $1 - \varepsilon$. Cette approche implique l'utilisation d'un nombre important de scénarios (échantillons) qui peut rendre le problème difficile à résoudre. De plus, [58] met en évidence que ces approches basées sur un échantillonnage sont en pratique trop conservatives ; le même type d'observation est effectué dans [7]. Noter encore que [30] s'appuie aussi sur des méthodes d'échantillonnage, mais de manière à prendre aussi en compte des incertitudes sur la distribution des variables aléatoires (contraintes probabilistes dites *ambiguës*).

En théorie, ces méthodologies robustes sont pour la plupart utilisables directement pour des problèmes en nombres entiers. Cependant, une première difficulté peut venir de la non-linéarité de certains problèmes robustes, conduisant à des problèmes non-linéaires en nombres entiers. Par ailleurs, utilisés tels qu'ils ont été proposés, ces modèles semblent trop conservatifs en pratique pour approcher de manière efficace un problèmes combinatoire sous contraintes probabilistes.

Quelques rares travaux ont traité explicitement, et de manière assez générale, de problèmes sous contraintes probabilistes avec nombres entiers. Le premier d'entre eux semble être [41] : dans cet article, l'auteur propose des approximations linéaires, intérieures et extérieures, à des problèmes linéaires mixtes sous contraintes probabilistes séparées. Des variables 0-1 sont explicitement prises en compte, et des processus de résolution optimale sont présentés. Malgré un souci évident de garder l'exposé aussi général que possible, l'ensemble de l'analyse suppose que, pour toute contrainte j , la distribution de $(\sum_{j_i} A_{ji}x_i - \sum_{j_i} \mathbb{E}[A_{ji}]x_i) / \sqrt{\mathbb{E}[(\sum_{j_i} A_{ji}x_i)^2]}$ est connue et ne dépend pas de x . C'est le cas si l'on suppose les variables aléatoires distribuées selon des lois normales, par exemple, ou si l'on peut légitimement appliquer un théorème de limite centrale (voir annexe B.2). Outre ce travail initial, plusieurs contributions se sont concentrées sur le cas d'un nombre fini de scénarios. Le problème peut s'écrire de manière équivalente à l'aide d'un PLNE (voir par exemple [40]). Ce modèle équivalent, quoique naturel, est de taille prohibitive. Toujours avec un nombre fini de scénarios, un algorithme de *branch-and-bound* est développé dans [10] quand seul le second membre des contraintes est stochastique. [80] semble être la seule publication proposant un algorithme de résolution pour des problèmes combinatoires sous contraintes probabilistes généraux. Les méthodes reposent sur des approches non-standard pour la programmation linéaire (cf. la théorie des bases de Gröbner). Malheureusement, sans plus d'amélioration, la méthode paraît restreinte à des problèmes de très petites tailles.

Finalement, mentionnons l'existence de travaux utilisant des méta-heuristiques pour la résolution de tels problèmes [1, 14]. Le principal inconvénient dans l'usage de ces algorithmes est leur paramétrage, qui requiert souvent beaucoup de tests préliminaires.

Notre but dans les deux chapitres qui viennent est de chercher à résoudre des problèmes sous contraintes probabilistes en utilisant une succession de problèmes robustes dont l'ensemble d'incertitudes augmente progressivement (cette idée a déjà été évoquée dans la partie 2.1). On essaiera d'abord de tirer partie des bonnes propriétés du modèle robuste de Bertsimas et Sim [13] pour le problème fondamental du sac-à-dos 0-1. Puis le même type d'algorithmes sera développé pour des programmes en nombres entiers plus généraux au chapitre 5.

4.2 Le problème de sac-à-dos sous contraintes probabilistes

Le problème de sac-à-dos classique a été introduit dans la partie 3.4. Il s'écrit :

$$\begin{aligned} \max & \quad \sum_{i \in I} p_i x_i \\ \text{s.c.} & \quad \sum_{i \in I} w_i x_i \leq c, \\ & \quad x \in \{0, 1\}^n, \end{aligned} \quad (4.1)$$

où $p \in \mathbb{R}_+^n$ est un vecteur de profits, $w \in \mathbb{R}_+^n$ un vecteur de poids, et $c > 0$ est la capacité du sac. Supposons à partir de maintenant que w est un vecteur de variables aléatoires réelles, on s'intéresse au problème CCKP (*chance-constrained knapsack problem*) :

$$\begin{aligned} \max & \quad \sum_{i \in I} p_i x_i \\ \text{s.c.} & \quad \mathbb{P}(\sum_{i \in I} w_i x_i \leq c) \geq 1 - \varepsilon, \\ & \quad x \in \{0, 1\}^n, \end{aligned} \quad (4.2)$$

avec $\varepsilon \in]0, 1[$.

Exemple : À titre d'illustration, on développe un exemple très simple dans \mathbb{R}^2 . Soit l'ensemble : $\mathcal{X} = \{x \in [0, 1]^2 \mid \mathbb{P}(w_1 x_1 + w_2 x_2 \leq 1) \geq 3/4\}$ (donc $\varepsilon = 1/4$). On suppose que $w_{1,2}$ est une variable aléatoire à valeurs dans $[0, 1]$. Soit f leur densité de probabilité : $f = \mathbb{1}_{[0,1]}$ (fonction caractéristique de $[0, 1]$).

Soit $(x_1, x_2) \in [0, 1]^2$. On note W_1 (resp. W_2) la variable aléatoire $x_1 w_1$ (resp. $x_2 w_2$). Chaque W_i a pour densité de probabilité $d_i = 1/x_i \cdot \mathbb{1}_{[0, x_i]}$. Alors :

$$\mathbb{P}(w_1 x_1 + w_2 x_2 \leq 1) = \mathbb{P}(W_1 + W_2 \leq 1) = \int_{t=0}^1 d_1(t) \cdot \mathbb{P}(W_2 \leq 1-t) dt = \int_{t=0}^{x_1} d_1(t) \cdot \mathbb{P}(W_2 \leq 1-t) dt.$$

$$\text{Des calculs faciles mènent à :} \quad \mathbb{P}(W_2 \leq 1-t) = \begin{cases} 1, & \text{si } t \leq 1-x_2 \\ (1-t)/x_2, & \text{si } t \geq 1-x_2 \end{cases}.$$

Si $x_1 + x_2 \leq 1 \Leftrightarrow 1-x_2 \geq x_1$, on a : $\mathbb{P}(w_1 x_1 + w_2 x_2 \leq 1) = 1$. Si $x_1 + x_2 > 1 \Leftrightarrow 1-x_2 < x_1$, on a :

$$\begin{aligned} \mathbb{P}(w_1 x_1 + w_2 x_2 \leq 1) &= \int_{t=0}^{1-x_2} d_1(t) dt + \int_{t=1-x_2}^{x_1} d_1(t) \cdot (1-t)/x_2 dt \\ &= 1/x_1 + 1/x_2 - 1/2 \cdot (x_2/x_1 + x_1/x_2) - 1/(2x_1 x_2). \end{aligned}$$

L'ensemble \mathcal{X} est représenté sur la figure 4.1.

Cet exemple très simple donne une idée de la difficulté d'aborder directement les problèmes sous contraintes probabilistes. C'est ce qui motive le recours à des méthodes heuristiques s'appuyant sur des modèles robustes simples. On rappelle que CCKP peut aussi s'écrire (cf. lemme 1.2.3) :

$$\max_{V: \mathbb{P}(V) \geq 1-\varepsilon} \left\{ \max_{x \in \{0,1\}^n} \{p \cdot x \mid \forall w \in V, w \cdot x \leq c\} \right\}.$$

Cette formulation s'appuie sur des problèmes d'optimisation robuste définis pour des ensembles d'incertitudes V . Ici, on s'appuiera sur le modèle robuste de Bertsimas et Sim, dont les avantages ont déjà été soulignés dans la partie 3.1. On suppose donc que chaque poids w_i est une variable aléatoire à valeurs dans $[\underline{w}_i, \bar{w}_i]$, $\underline{w}_i \geq 0$ et $\bar{w}_i > \underline{w}_i$. On introduit également, pour tout $i \in I$, la variable aléatoire normalisée $\eta_i : \eta_i = (w_i - \underline{w}_i)/(\bar{w}_i - \underline{w}_i) \in [0, 1]$.

Soit $\Gamma \in \{0, \dots, n\}$, on rappelle que le modèle de sac-à-dos robuste RKP (*Robust Knapsack Problem*) s'écrit (voir partie 3.4) :

$$\begin{aligned} \max & \quad \sum_{i \in I} p_i x_i \\ \text{s.c.} & \quad \sum_{i \in I \setminus S} \underline{w}_i x_i + \sum_{i \in S} \bar{w}_i x_i \leq c, \quad \forall S \subseteq I \text{ t.q. } |S| = \Gamma, \\ & \quad x \in \{0, 1\}^n. \end{aligned} \quad (4.3)$$

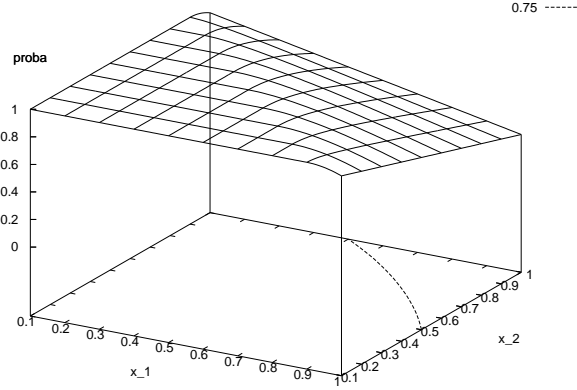


FIG. 4.1 – Probabilités des solutions $x \in [0, 1]^2$ pour l'exemple.

4.3 Lien entre RKP et CCKP

Pour tout $x \in \{0, 1\}^n$, on note : $I(x) = \{i \in I \mid x_i = 1\}$. Le résultat suivant généralise très légèrement un de ceux énoncés dans [13] :

Lemme 18 *Pour toute solution x de RKP : $\mathbb{P}(w.x \leq c) \geq \mathbb{P}(\sum_{i \in I} \eta_i \leq \Gamma)$.*

Preuve : Soit x une solution réalisable de RKP. Si $|I(x)| \leq \Gamma$, il existe $S \subseteq I$ de cardinalité $|S| = \Gamma$ tel que : $I(x) \subseteq S$. Alors : $\sum_{i \in I} \bar{w}_i x_i = \sum_{i \in I(x)} \bar{w}_i x_i = \sum_{i \in S} \bar{w}_i x_i + \sum_{i \notin S} \underline{w}_i x_i \leq c$ (puisque x est réalisable pour RKP). Ceci implique : $\mathbb{P}(w.x \leq c) = 1$, et donc : $\mathbb{P}(w.x \leq c) \geq \mathbb{P}(\sum_{i \in I} \eta_i \geq \Gamma)$.

Supposons maintenant que $|I(x)| \geq \Gamma + 1$. Par souci de clarté, on note : $\delta_i = \bar{w}_i - \underline{w}_i$. Pour tout ensemble $S \subseteq I(x)$ de cardinalité Γ , on a :

$$\begin{aligned}
 \mathbb{P}(w.x > c) &\leq \mathbb{P}\left(w.x > \sum_{i \in S} \bar{w}_i x_i + \sum_{i \notin S} \underline{w}_i x_i\right) \\
 &= \mathbb{P}\left(\sum_{i \in I(x)} w_i > \sum_{i \in S} \bar{w}_i + \sum_{i \in I(x) \setminus S} \underline{w}_i\right) \\
 &= \mathbb{P}\left(\sum_{i \in I(x)} \eta_i \delta_i > \sum_{i \in S} \delta_i\right) \\
 &= \mathbb{P}\left(\sum_{i \in I(x) \setminus S} \eta_i \delta_i > \sum_{i \in S} \delta_i (1 - \eta_i)\right) \\
 &\leq \mathbb{P}\left(\sum_{i \in I(x) \setminus S} \eta_i \delta_i > \min_{j \in S} \delta_j \cdot \sum_{i \in S} (1 - \eta_i)\right) \\
 &= \mathbb{P}\left(\sum_{i \in I(x) \setminus S} \eta_i [\delta_i / \min_{j \in S} \delta_j] + \sum_{i \in S} \eta_i > \Gamma\right)
 \end{aligned}$$

La première inégalité vient du fait que x est réalisable pour RKP. On peut maintenant choisir $S \subseteq I(x)$ de manière à satisfaire : $\forall i \in I(x) \setminus S, \delta_i \leq \min_{j \in S} \delta_j$. En ce cas : $\mathbb{P}(w.x > c) \leq \mathbb{P}\left(\sum_{i \in I(x)} \eta_i > \Gamma\right)$, et donc : $\mathbb{P}(w.x > c) \leq \mathbb{P}\left(\sum_{i \in I} \eta_i > \Gamma\right)$. \square

Cette observation est générale, puisqu'aucune hypothèse particulière n'a été réalisée sur la distribution probabiliste des poids. On rappelle maintenant ce résultat [42] :

Théorème 12 (Hoeffding, 1963) *Soient X_1, \dots, X_n des variables aléatoires réelles indépendantes telles que, pour tout $i \in I$: $\mathbb{P}(X_i \in [\alpha_i, \beta_i]) = 1$. Alors, en notant $\Theta = \sum_{i=1}^n X_i$:*

$$\forall \tau \geq 0, \mathbb{P}(\Theta \geq \mathbb{E}[\Theta] + \tau) \leq \exp\left(-\frac{2\tau^2}{\sum_{i=1}^n (\beta_i - \alpha_i)^2}\right)$$

En utilisant ce résultat, on montre que :

Proposition 13 *Supposons que les variables aléatoires $\{w_i\}_{i \in I}$ sont indépendantes et symétriquement distribuées. Si $\Gamma \geq 1/2 \cdot \left(n + \sqrt{-2n \ln(\varepsilon)}\right)$, alors une solution réalisable pour RKP est aussi réalisable pour CCKP.*

Preuve : Soit : $S = \sum_{i \in I} \eta_i$. On a : $\mathbb{P}(S \geq \Gamma) = \mathbb{P}(S \geq \mathbb{E}[S] + (\Gamma - \mathbb{E}[S])) = \mathbb{P}(S \geq \mathbb{E}[S] + (\Gamma - n/2))$, puisque $\mathbb{E}[S] = n/2$. D'après l'inégalité de Hoeffding (théorème 12), si $\Gamma \geq n/2$: $\mathbb{P}(S \geq \Gamma) \leq \exp(-2(\Gamma - n/2)^2/n)$. Supposons alors que : $\exp(-2(\Gamma - n/2)^2/n) < \varepsilon \Leftrightarrow \Gamma \geq n/2 + \sqrt{-n \ln(\varepsilon)}/2$. D'après le lemme 18, on obtient que toute solution réalisable pour RKP l'est aussi pour CCKP. \square

Ce résultat se déduit déjà du travail de [13]. De meilleures bornes de probabilités sont détaillées dans [13], mais elles sont plus complexes aussi.

On détaille maintenant d'autres observations qui relient RKP et CCKP. Pour tout $I' \subseteq I$ et $\Gamma \in \{0, \dots, |I'|\}$, on introduit le problème RKP(I', Γ) :

$$\begin{aligned} \max \quad & p \cdot x \\ \text{s.c.} \quad & \sum_{i \in S} \bar{w}_i x_i + \sum_{i \in I' \setminus S} \underline{w}_i x_i + \sum_{i \notin I'} \bar{w}_i x_i \leq c, \quad \forall S \subseteq I' \text{ t.q. } |S| = \Gamma, \\ & x \in \{0, 1\}^n. \end{aligned} \quad (4.4)$$

RKP(I', Γ) est une instance particulière du modèle robuste de Bertsimas et Sim, tous les poids de $I \setminus I'$ étant fixés à leur pire valeur, sans incertitude. Pour plus de précision, on notera dorénavant RKP(I, Γ) au lieu de RKP. Observer que pour toute paire de sous-ensembles I' et I'' de I , $I' \subseteq I''$ implique que toute solution réalisable pour RKP(I', Γ) l'est aussi pour RKP(I'', Γ). En particulier, toute solution réalisable de RKP(I', Γ) est aussi réalisable pour RKP(I, Γ). D'autre part, si $\Gamma' \leq \Gamma$, toute solution réalisable de RKP(I', Γ) l'est aussi pour RKP(I', Γ').

Lemme 19 *Soit $\Gamma \in \{0, \dots, n\}$. Supposons que x^* est une solution optimale de RKP(I, Γ) :*

- (i) *si $\Gamma \leq |I(x^*)|$, x^* est une solution optimale de RKP($I(x^*), \Gamma$) ;*
- (ii) *si $\Gamma \geq |I(x^*)|$, x^* est une solution optimale de RKP(I, n).*

Preuve : Supposons que $\Gamma \leq |I(x^*)|$. Puisque $I(x^*) \subseteq I$, toute solution réalisable de RKP($I(x^*), \Gamma$) l'est aussi pour RKP(I, Γ). Mais x^* est aussi une solution réalisable de RKP($I(x^*), \Gamma$). Comme x^* est une solution optimale de RKP(I, Γ), elle est optimale aussi pour RKP($I(x^*), \Gamma$).

Si $\Gamma \geq |I(x^*)|$, il existe $S \subseteq I$ de cardinal $|S| = \Gamma$ tel que $I(x^*) \subseteq S$. Alors : $\sum_{i \in I} \bar{w}_i x_i^* = \sum_{i \in I(x^*)} \bar{w}_i x_i^* = \sum_{i \in S} \bar{w}_i x_i^* + \sum_{i \notin S} \underline{w}_i x_i^* \leq c$. La dernière inégalité vient du fait que x^* est réalisable pour RKP(I, Γ). Ainsi, on montre que x^* est réalisable pour RKP(I, n). Mais comme $\Gamma \leq n$, RKP(I, Γ) est une relaxation de RKP(I, n) : x^* est optimale également pour RKP(I, n). (Observer que dans le cas où $\Gamma = |I(x^*)|$, les deux affirmations (i) et (ii) ont lieu : RKP($I(x^*), \Gamma$) et RKP(I, n) ont alors la même solution optimale.) \square

Le résultat suivant montre la pertinence du cadre robuste envisagé pour traiter CCKP quand toutes les variations $\{\bar{w}_i - \underline{w}_i\}_{i \in I}$ sont identiques :

Théorème 13 *Supposons que : $\forall i \in I, \bar{w}_i - \underline{w}_i = \delta > 0$, et que $\{\underline{w}_i/\delta\}_{i \in I}$ et c/δ sont des entiers. Alors il existe $I^* \subseteq I$ et $\Gamma^* \in \{0, \dots, |I^*|\}$ tels que toute solution optimale de RKP(I^*, Γ^*) est aussi optimale pour CCKP.*

Preuve : Soit x^* une solution optimale, on note $I^* = I(x^*)$. Puisque x^* est réalisable avec probabilité au moins $1 - \varepsilon$, on a : $\mathbb{P}(w.x^* \leq c) = \mathbb{P}(\sum_{i \in I^*} w_i \leq c) \geq 1 - \varepsilon$. On note $V = \{w \in W \mid \sum_{i \in I^*} w_i \leq c\}$, et on considère le problème :

$$\begin{aligned} \max \quad & p.x \\ \text{s.c.} \quad & w.x \leq c, \quad \forall w \in V, \\ & x \in \{0, 1\}^n. \end{aligned} \tag{4.5}$$

Montrons que toute solution optimale de (4.5) est optimale pour CCKP. Toute solution réalisable x de (4.5) est réalisable pour CCKP, puisque : $\mathbb{P}(w.x \leq c) \geq \mathbb{P}(w \in V) \geq 1 - \varepsilon$. Il suffit donc de montrer que (4.5) a la même solution optimale que CCKP. Ceci est clair, puisque par construction, la solution optimale x^* de CCKP est réalisable pour (4.5).

Notre but maintenant est de trouver $\Gamma^* \in \{0, \dots, |I^*|\}$ tel que (4.5) est équivalent à $\text{RKP}(I^*, \Gamma^*)$. Observer que : $w \in V \Leftrightarrow \sum_{i \in I^*} \eta_i \leq (c - \sum_{i \in I^*} \underline{w}_i) / \delta$. On définit alors : $\Gamma^* = \min \{|I^*|, (c - \sum_{i \in I^*} \underline{w}_i) / \delta\}$. On a : $w \in V \Leftrightarrow \sum_{i \in I^*} \eta_i \leq \Gamma^*$. On notera que $V \neq \emptyset$, ce qui assure $(c - \sum_{i \in I^*} \underline{w}_i) / \delta \geq 0$. D'autre part, d'après nos hypothèses, Γ^* est entier. Montrons que (4.5) est équivalent à $\text{RKP}(I^*, \Gamma^*)$.

On prouve que x est réalisable pour (4.5) si et seulement si x est réalisable pour $\text{RKP}(I^*, \Gamma^*)$. Soit x réalisable pour (4.5). Pour tout $S \subseteq I^*$ tel que $|S| = \Gamma^*$, on montre que : $\sum_{i \in S} \bar{w}_i x_i + \sum_{i \in I^* \setminus S} \underline{w}_i x_i + \sum_{i \notin I^*} \bar{w}_i x_i \leq c$. Soit $w(S)$ le vecteur dont les composantes sont \underline{w}_i si $i \in I^* \setminus S$, et \bar{w}_i sinon. On a : $\sum_{i \in I^*} w(S)_i = \sum_{i \in S} \bar{w}_i + \sum_{i \in I^* \setminus S} \underline{w}_i = \Gamma^* \delta + \sum_{i \in I^*} \underline{w}_i \leq c$, et en conséquence : $w(S) \in V$. Donc, puisque x est réalisable pour (4.5) : $w(S).x = \sum_{i \in S} \bar{w}_i x_i + \sum_{i \in I^* \setminus S} \underline{w}_i x_i + \sum_{i \notin I^*} \bar{w}_i x_i \leq c$. Ainsi, x est réalisable pour $\text{RKP}(I^*, \Gamma^*)$.

Réciproquement, soit x non-réalisable pour (4.5), c'est-à-dire : il existe $w \in V$ tel que $w.x > c$. Montrons qu'il existe $S \subseteq I^*$ avec $|S| = \Gamma^*$ tel que : $\sum_{i \in S} \bar{w}_i x_i + \sum_{i \in I^* \setminus S} \underline{w}_i x_i + \sum_{i \notin I^*} \bar{w}_i x_i = \sum_{i \in I^*} \underline{w}_i x_i + \sum_{i \in S} \delta x_i + \sum_{i \notin I^*} \bar{w}_i x_i > c$. Si $|I(x) \cap I^*| \geq \Gamma^*$ (cas (a)), on construit S tel que $|S| = \Gamma^*$ et $S \subseteq I(x) \cap I^*$. Si $|I(x) \cap I^*| < \Gamma^*$ (cas (b)), soit S une extension quelconque de $I(x) \cap I^*$ dans I^* , telle que $|S| = \Gamma^*$ (on rappelle que, par construction, $\Gamma^* \leq |I^*|$).

Si on est dans le cas (a), on a : $\sum_{i \in I^*} \underline{w}_i x_i + \sum_{i \in S} \delta x_i + \sum_{i \notin I^*} \bar{w}_i x_i = \sum_{i \in I^*} \underline{w}_i x_i + \delta \Gamma^* + \sum_{i \notin I^*} \bar{w}_i x_i \geq \sum_{i \in I^*} \underline{w}_i x_i + \sum_{i \in I^*} \eta_i \delta + \sum_{i \notin I^*} \bar{w}_i x_i \geq \sum_{i \in I^*} \underline{w}_i x_i + \sum_{i \in I^*} \eta_i \delta x_i + \sum_{i \notin I^*} \bar{w}_i x_i = w.x > c$. La première inégalité vient du fait que $w \in V$; pour la dernière inégalité, se rappeler que $w_i = \underline{w}_i + \eta_i \delta$. De la même manière, dans le cas (b), on a $\sum_{i \in I^*} \underline{w}_i x_i + \sum_{i \in S} \delta x_i + \sum_{i \notin I^*} \bar{w}_i x_i = \sum_{i \in I^*} \underline{w}_i x_i + \delta |I(x) \cap I^*| + \sum_{i \notin I^*} \bar{w}_i x_i \geq \sum_{i \in I^*} \underline{w}_i x_i + \sum_{i \in I(x) \cap I^*} \eta_i \delta x_i + \sum_{i \notin I^*} \bar{w}_i x_i = \sum_{i \in I^*} \underline{w}_i x_i + \sum_{i \in I^*} \eta_i \delta x_i + \sum_{i \notin I^*} \bar{w}_i x_i = w.x > c$.

Par contraposition, ceci prouve que toute solution de $\text{RKP}(I^*, \Gamma^*)$ est aussi réalisable pour (4.5). On a donc montré que x est réalisable pour (4.5) si et seulement si x est réalisable pour $\text{RKP}(I^*, \Gamma^*)$. En conséquence, puisque les deux problèmes ont la même fonction-objectif, ils sont équivalents. En particulier, ils admettent les mêmes solutions optimales. Puisque toute solution optimale de (4.5) est optimale pour CCKP, on a prouvé que toute solution optimale de $\text{RKP}(I^*, \Gamma^*)$ est aussi optimale pour CCKP. \square

Une des forces du théorème ci-dessus est qu'il ne requiert aucune hypothèse particulière sur la distribution des variables aléatoires. D'autre part, la condition d'intégrité de $\{\underline{w}_i / \delta\}_{i \in I}$ et de c / δ sont nécessaires pour assurer que Γ^* peut être choisi entier. Néanmoins, dans un contexte plus général, le paramètre Γ peut aussi ne pas être entier (voir la partie 3.1, et en particulier le modèle (3.4)). Dans ce cas, le théorème 13 se généralise directement en omettant les contraintes d'intégrité sur $\{\underline{w}_i / \delta\}_{i \in I}$ et c / δ .

Dans la suite de cette partie, on détaille deux cas particuliers pour lesquels le théorème précédent peut être renforcé.

Lemme 20 *Supposons que :*

- les poids et les profits satisfont : $i < j \Rightarrow \begin{cases} \underline{w}_i \leq \underline{w}_j \\ p_i \geq p_j \end{cases}$,
- pour tout $i \in I$, $\bar{w}_i - \underline{w}_i = \delta > 0$,
- les variables aléatoires $\{\eta_i\}_{i \in I}$ sont indépendantes et identiquement distribuées (i.i.d.).

Alors il existe une solution optimale x^* de CCKP telle que $I(x^*) = \{1, \dots, |I(x^*)|\}$.

Preuve : Soit x une solution optimale de CCKP, on construit x^* tel que $I(x^*) = \{1, \dots, |I(x)|\}$. Comme $|I(x^*)| = |I(x)|$, et puisque les profits sont triés en ordre décroissant : $p.x^* \geq p.x$. De plus, observer que par construction : $\sum_{i \in I(x^*)} \underline{w}_i \leq \sum_{i \in I(x)} \underline{w}_i$. Alors :

$$\begin{aligned}
\mathbb{P}(w.x^* \leq c) &= \mathbb{P}(\sum_{i \in I(x^*)} w_i \leq c) = \mathbb{P}(\delta \cdot \sum_{i \in I(x^*)} \eta_i \leq c - \sum_{i \in I(x^*)} \underline{w}_i) \\
&= \mathbb{P}(\delta \cdot \sum_{i \in I(x)} \eta_i \leq c - \sum_{i \in I(x^*)} \underline{w}_i) & (a) \\
&\geq \mathbb{P}(\delta \cdot \sum_{i \in I(x)} \eta_i \leq c - \sum_{i \in I(x)} \underline{w}_i) & (b) \\
&= \mathbb{P}(\sum_{i \in I(x)} w_i \leq c) = \mathbb{P}(w.x \leq c).
\end{aligned}$$

Comme les $\{\eta_i\}_{i \in I}$ sont supposés i.i.d., et puisque $|I(x^*)| = |I(x)|$, les variables aléatoires $\sum_{i \in I(x)} \eta_i$ et $\sum_{i \in I(x^*)} \eta_i$ sont aussi i.i.d. : ceci assure (a). (b) est une conséquence directe de : $\sum_{i \in I(x^*)} \underline{w}_i \leq \sum_{i \in I(x)} \underline{w}_i$. Comme x est réalisable pour CCKP, on a donc montré que x^* est aussi réalisable pour ce problème. On en déduit que x^* est une solution optimale de CCKP. \square

Théorème 14 *Supposons que :*

- les poids et les profits satisfont : $i < j \Rightarrow \begin{cases} \underline{w}_i \leq \underline{w}_j \\ p_i \geq p_j \end{cases}$,
- pour tout $i \in I$, $\bar{w}_i - \underline{w}_i = \delta > 0$,
- $\{\underline{w}_i/\delta\}_{i \in I}$ et c/δ sont des entiers,
- les variables aléatoires $\{\eta_i\}_{i \in I}$ sont i.i.d..

Alors il existe $\Gamma^* \in \{0, \dots, n\}$ tel que toute solution optimale de $RKP(I, \Gamma^*)$ est aussi optimale pour CCKP.

Preuve : D'après le théorème 13, on sait qu'il existe $I^* \subseteq I$ et $\Gamma^* \in \{0, \dots, |I^*|\}$ tels qu'une solution optimale de $RKP(I^*, \Gamma^*)$ est optimale également pour CCKP. De plus, d'après la preuve du théorème, on sait que I^* peut être choisi de manière à satisfaire $I^* = I(x^*)$ pour une solution optimale x^* de CCKP, et que Γ^* peut être aussi choisi tel que : $\Gamma^* = \min \{|I^*|, (c - \sum_{i \in I^*} \underline{w}_i) / \delta\}$. Sans perte de généralité, d'après le lemme 20, on considère : $I(x^*) = \{1, \dots, |I^*|\}$. Supposons d'abord que $\Gamma^* = |I^*|$: dans ce cas, les problèmes $RKP(I^*, \Gamma^*)$ et $RKP(I, n)$ sont équivalents.

Supposons maintenant que $\Gamma^* = (c - \sum_{i \in I^*} \underline{w}_i) / \delta < |I^*|$. On va montrer que toute solution optimale de $RKP(I, \Gamma^*)$ l'est aussi pour CCKP. Soit x' une solution optimale de $RKP(I, \Gamma^*)$. Si $|I(x')| \leq \Gamma^*$, x' est réalisable pour $RKP(I, n)$ et en conséquence : $\mathbb{P}(w.x' \leq c) = 1$. Donc dans ce cas, x' est réalisable pour CCKP. Si $|I(x')| \geq \Gamma^* + 1$, montrons que x' est une solution réalisable de CCKP. Si on suppose que : $\sum_{i \in I(x')} \underline{w}_i > \sum_{i \in I^*} \underline{w}_i$, alors pour tout ensemble $S \subseteq I(x')$ tel que $|S| = \Gamma^*$:

$$\sum_{i \in S} \bar{w}_i x'_i + \sum_{i \in I \setminus S} \underline{w}_i x'_i = \sum_{i \in I(x')} \underline{w}_i + \Gamma^* \delta > \sum_{i \in I^*} \underline{w}_i + \Gamma^* \delta = c - \Gamma^* \delta + \Gamma^* \delta = c.$$

Ceci est une contradiction avec la réalisabilité de x' pour $RKP(I, \Gamma^*)$. On en déduit que : $\sum_{i \in I(x')} \underline{w}_i \leq \sum_{i \in I^*} \underline{w}_i$. Puisque I^* est l'ensemble des $|I^*|$ plus faibles valeurs de $\{\underline{w}_i\}_{i \in I}$, cette dernière relation implique : $|I(x')| \leq |I^*|$. D'autre part, comme $I^* \subseteq I$, x^* est réalisable pour $RKP(I, \Gamma^*)$, et on a donc : $p.x^* \leq p.x'$, i.e. : $\sum_{i \in I^*} p_i \leq \sum_{i \in I(x')} p_i$. Puisque I^* est l'ensemble des $|I^*|$ plus grandes valeurs de $\{p_i\}_{i \in I}$, on obtient : $|I(x')| \geq |I^*|$. De plus, comme I^* contient

les $|I^*|$ poids les plus faibles, cette dernière inégalité implique : $\sum_{i \in I(x')} \underline{w}_i \geq \sum_{i \in I^*} \underline{w}_i$. En définitive : $|I(x')| = |I^*|$ et $\sum_{i \in I(x')} \underline{w}_i = \sum_{i \in I^*} \underline{w}_i$.

Maintenant, montrons que x' est réalisable pour CCKP :

$$\begin{aligned} \mathbb{P}(w.x' \leq c) &= \mathbb{P}(\delta \cdot \sum_{i \in I(x')} \eta_i \leq c - \sum_{i \in I(x')} \underline{w}_i) \\ &= \mathbb{P}(\delta \cdot \sum_{i \in I(x')} \eta_i \leq c - \sum_{i \in I(x^*)} \underline{w}_i) \quad (a) \\ &= \mathbb{P}(\delta \cdot \sum_{i \in I(x^*)} \eta_i \leq c - \sum_{i \in I(x^*)} \underline{w}_i) \quad (b) \\ &= \mathbb{P}(w.x^* \leq c). \end{aligned}$$

(a) vient de : $\sum_{i \in I(x')} \underline{w}_i = \sum_{i \in I^*} \underline{w}_i$. (b) est une conséquence de $|I(x')| = |I^*|$ et des hypothèses probabilistes sur les $\{\eta_i\}_{i \in I}$. On a donc prouvé que x' est une solution réalisable de CCKP. Finalement, comme on a vu que $p.x^* \leq p.x'$, x' est optimal pour CCKP. \square

Comme précédemment pour le théorème 13, la condition d'intégrité de c/δ et des \underline{w}_i/δ pour $i \in I$ n'est pas une contrainte en pratique, puisque le modèle robuste peut aussi être écrit avec Γ fractionnaire. On déduit deux résultats du théorème précédent :

Corollaire 3 *Supposons que :*

- pour tout $i \in I$, $\underline{w}_i = \omega > 0$ et $\bar{w}_i - \underline{w}_i = \delta > 0$,
- ω/δ et c/δ sont des entiers,
- les variables aléatoires $\{\eta_i\}_{i \in I}$ sont i.i.d..

Alors il existe $\Gamma^* \in \{0, \dots, n\}$ tel que toute solution optimale de $RKP(I, \Gamma^*)$ est optimale pour CCKP.

Corollaire 4 *Supposons que :*

- pour tout $i \in I$, $p_i = \rho > 0$ et $\bar{w}_i - \underline{w}_i = \delta > 0$,
- pour tout $i \in I$, \underline{w}_i/δ et c/δ sont entiers,
- les variables aléatoires $\{\eta_i\}_{i \in I}$ sont i.i.d..

Alors il existe $\Gamma^* \in \{0, \dots, n\}$ tel que toute solution optimale de $RKP(I, \Gamma^*)$ est optimale pour CCKP.

L'intérêt de ces résultats est théorique uniquement. En pratique, si les poids et les profits sont triés respectivement de manière décroissante et croissante, on sait résoudre CCKP facilement. Il suffit en effet de construire une solution x en lui incorporant successivement les éléments de plus fort profit (de plus faible poids), et de calculer à chaque fois la probabilité de réalisabilité de la solution. Cependant, il est fort probable qu'en pratique, les instances satisfaisant "presque" ces hypothèses seront bien approchées par un problème robuste $RKP(I, \Gamma)$.

4.4 Algorithme pour la résolution approchée de CCKP

Dans le résultat qui suit, on notera simplement RKP le problème $RKP(I, \Gamma)$. De plus, toutes les données $\{\underline{w}_i\}_{i \in I}$, $\{\bar{w}_i\}_{i \in I}$ et c seront supposées entières. L'algorithme de programmation dynamique classique pour le sac-à-dos peut être généralisé pour résoudre RKP :

Théorème 15 *RKP est NP-difficile au sens faible : il existe un algorithme pseudo-polynomial pour le résoudre en temps $O(n\Gamma c)$.*

Preuve : RKP est NP-difficile, puisque le cas $\Gamma = 0$ correspond au problème de sac-à-dos classique. Le fait que RKP peut être résolu en temps pseudo-polynomial peut être déduit du théorème 8, qui montre que RKP peut être résolu *via* $n + 1$ problèmes de sac-à-dos classique

(voir § 3.2). L'application directe de ce résultat garantit l'existence d'un algorithme $O(n^2c)$, puisqu'un problème de sac-à-dos classique est résolu par programmation dynamique en temps $O(nc)$. On donne ici une preuve qui améliore cette complexité, et qui précise l'algorithme de programmation dynamique dans le cas robuste.

On suppose sans perte de généralité que les éléments de I sont triés par variations croissantes : $i < j \Rightarrow (\bar{w}_i - \underline{w}_i) \leq (\bar{w}_j - \underline{w}_j)$. On note $\text{RKP}_k(\Gamma, b)$ le problème RKP écrit avec le paramètre de robustesse $\Gamma \in \{0, \dots, n\}$ et la capacité $b \in \{1, \dots, c\}$, en se restreignant aux seuls éléments $I_k = \{1, \dots, k\}$ pour $k \in I$. Soit $F_k(\Gamma, b)$ la valeur optimale de ce problème.

Pour tout $k \in I$, on suppose : $b < 0 \Rightarrow F_k(., b) = -\infty$. Pour $k = 1$, F_1 est défini comme suit :

$$\begin{aligned} \text{si } \Gamma = 0 : F_1(0, b) &= \begin{cases} 0 & \text{si } b < \underline{w}_1 \\ p_1 & \text{si } b \geq \underline{w}_1 \end{cases}, \\ \text{si } \Gamma \geq 1 : F_1(\Gamma, b) &= \begin{cases} 0 & \text{si } b < \bar{w}_1 \\ p_1 & \text{si } b \geq \bar{w}_1 \end{cases}. \end{aligned}$$

Soient $b \in \{1, \dots, c\}$ et $\Gamma \in \{1, \dots, n\}$. La valeur optimale de $\text{RKP}_k(\Gamma, b)$ peut être obtenue à partir de la formule de récurrence suivante :

$$F_k(\Gamma, b) = \max \{F_{k-1}(\Gamma, b), p_k + F_{k-1}(\Gamma - 1, b - \bar{w}_k)\}.$$

Montrons que $F_k(\Gamma, b)$ est la valeur optimale de $\text{RKP}_k(\Gamma, b)$. Le résultat est correct pour $k = 1$. Soit $k \geq 2$.

On montre d'abord que toute solution $x^{k-1} \in \{0, 1\}^{k-1}$ réalisable de $\text{RKP}_{k-1}(\Gamma, b)$ ou de $\text{RKP}_{k-1}(\Gamma - 1, b - \bar{w}_k)$ peut être étendue en une solution $x^k \in \{0, 1\}^k$ réalisable pour $\text{RKP}_k(\Gamma, b)$, en ajoutant un k^e élément à x^{k-1} . Si x^{k-1} est réalisable pour $\text{RKP}_{k-1}(\Gamma, b)$, alors il suffit de prendre $x_k^k = 0$ pour que x^k soit réalisable pour $\text{RKP}_k(\Gamma, b)$. Si x^{k-1} est une solution réalisable de $\text{RKP}_{k-1}(\Gamma - 1, b - \bar{w}_k)$, considérons le vecteur x^k tel que $x_k^k = 1$. On a : $\sum_{i \in I_k} \underline{w}_i x_i^k + \max\{\sum_{i \in S} (\bar{w}_i - \underline{w}_i) x_i^k : S \subseteq I_k, |S| \leq \Gamma\} = \sum_{i \in I_{k-1}} \underline{w}_i x_i^{k-1} + \bar{w}_k + \max\{\sum_{i \in S} (\bar{w}_i - \underline{w}_i) x_i^{k-1} : S \subseteq I_{k-1}, |S| \leq \Gamma - 1\}$. Ceci vient de l'ordre supposé sur les variations des poids. En observant de plus que : $\sum_{i \in I_{k-1}} \underline{w}_i x_i^{k-1} + \max\{\sum_{i \in S} (\bar{w}_i - \underline{w}_i) x_i^{k-1} : S \subseteq I_{k-1}, |S| \leq \Gamma - 1\} \leq b - \bar{w}_k$, alors on obtient que x^k est aussi réalisable pour $\text{RKP}_k(\Gamma, b)$.

On montre la réciproque : pour toute solution réalisable de $\text{RKP}_k(\Gamma, b)$, les $k - 1$ premières composantes donnent une solution réalisable pour $\text{RKP}_{k-1}(\Gamma, b)$ ou pour $\text{RKP}_{k-1}(\Gamma - 1, b - \bar{w}_k)$. En effet, soit x^k une solution réalisable de $\text{RKP}_k(\Gamma, b)$. Soit $x^{k-1} \in \{0, 1\}^{k-1}$ le vecteur composé des $k - 1$ premiers éléments de x^k . Deux cas se présentent :

(a) $x_k^k = 0$: alors x^{k-1} est réalisable pour $\text{RKP}_{k-1}(\Gamma, b)$.

(b) $x_k^k = 1$: alors on peut montrer par contraposition que x^{k-1} est une solution réalisable de $\text{RKP}_{k-1}(\Gamma - 1, b - \bar{w}_k)$. Supposons qu'il existe $S \subseteq I_{k-1}$ tel que $|S| = \Gamma - 1$ et : $\sum_{i \in I_{k-1}} \underline{w}_i x_i^{k-1} + \sum_{i \in S} (\bar{w}_i - \underline{w}_i) x_i^{k-1} > b - \bar{w}_k$. Alors, x^k ne peut pas être réalisable pour $\text{RKP}_k(\Gamma, b)$ puisque pour $S' = S \cup \{k\}$, on obtient : $\sum_{i \in I_k} \underline{w}_i x_i^k + \sum_{i \in S'} (\bar{w}_i - \underline{w}_i) x_i^k = \sum_{i \in I_{k-1}} \underline{w}_i x_i^{k-1} + \sum_{i \in S} (\bar{w}_i - \underline{w}_i) x_i^{k-1} + \bar{w}_k > b$.

Ainsi, on a montré que l'ensemble des vecteurs composés des $k - 1$ premières composantes de solutions réalisables x^k de $\text{RKP}_k(\Gamma, b)$ est égal à l'union de l'ensemble des solutions respectives de $\text{RKP}_{k-1}(\Gamma, b)$ et de $\text{RKP}_{k-1}(\Gamma - 1, b - \bar{w}_k)$. En définitive : $F_k(\Gamma, b) = \max \{F_{k-1}(\Gamma, b), p_k + F_{k-1}(\Gamma - 1, b - \bar{w}_k)\}$.

Finalement, observer que la valeur $F_n(\Gamma, c)$ peut être calculée en temps pseudo-polynomial $O(n\Gamma c)$. \square

On précise les caractéristiques de la solution dans les cas particuliers des poids ou des profits uniformes.

Lemme 21 *Supposons que pour tout $i \in I$, $\underline{w}_i = \omega > 0$ et $\bar{w}_i - \underline{w}_i = \delta > 0$. Alors toute solution x^* optimale pour $RKP(I, \Gamma)$ vérifie :*

$$\sum_{i \in I} x_i^* = \begin{cases} \min \{n, \lfloor (c - \Gamma\delta)/\omega \rfloor\}, & \text{si } \Gamma(\omega + \delta) \leq c, \\ \lfloor c/(\omega + \delta) \rfloor, & \text{sinon.} \end{cases}$$

Preuve : Soit x^* une solution optimale de $RKP(I, \Gamma)$, x^* maximise la somme de ses composantes $(\sum_{i \in I} x_i)$. Supposons que $\Gamma(\omega + \delta) \leq c$. Ceci signifie que x^* comporte au moins Γ éléments : $\sum_{i \in I} x_i^* \geq \Gamma$. Alors x^* est en fait une solution optimale du problème : $\max\{p \cdot x \mid \sum_{i \in I} \omega x_i \leq c - \Gamma\delta\}$.

Si au contraire $\Gamma(\omega + \delta) > c$, une solution réalisable de $RKP(I, \Gamma)$ ne peut pas comporter plus de $\Gamma - 1$ éléments. Alors x^* est en fait solution optimale de : $\max\{p \cdot x \mid \sum_{i \in I} (\omega + \delta)x_i \leq c\}$. Ceci conclut la preuve. \square

Ainsi, la solution de $RKP(I, \Gamma)$ peut être construite en temps linéaire en considérant les éléments de plus fort profit. Le même type de résultat est disponible quand les profits sont uniformes :

Lemme 22 *Supposons que pour tout $i \in I$, $p_i = \rho > 0$ et $\bar{w}_i - \underline{w}_i = \delta > 0$. Soit x' une solution optimale de :*

$$\begin{aligned} \max & \sum_{i \in I} x_i \\ \text{s.c.} & \sum_{i \in I} \underline{w}_i x_i \leq c - \Gamma\delta, \\ & x \in \{0, 1\}^n, \end{aligned} \tag{4.6}$$

et soit x'' une solution optimale de :

$$\begin{aligned} \max & \sum_{i \in I} x_i \\ \text{s.c.} & \sum_{i \in I} \bar{w}_i x_i \leq c, \\ & x \in \{0, 1\}^n. \end{aligned} \tag{4.7}$$

Si $c \geq \Gamma\delta$ et $\sum_{i \in I} x_i' \geq \Gamma$, x' est optimal pour $RKP(I, \Gamma)$. Sinon, x'' est optimal pour $RKP(I, \Gamma)$.

Remarquer que, du fait des hypothèses réalisées, les profits n'interviennent plus explicitement dans les formulations. Par ailleurs, les deux problèmes (4.6) et (4.7) sont résolus en temps linéaire. De plus, observer que (4.6) n'a pas de solution si $c - \Gamma\delta < 0$.

L'idée de l'algorithme suivant est de faire augmenter progressivement la valeur du paramètre de robustesse Γ et de résoudre les problèmes robustes correspondants jusqu'à obtenir une solution réalisable de CCKP.

Algorithme HA_CCKP

Étape 1 : Soit $k = 0$.

Étape 2 : Résoudre $RKP(I, k)$. Soit $x^{(k)}$ la solution obtenue.

Étape 3 : Soit $I' = I(x^{(k)})$. Calculer Γ' , la plus grande valeur de $\{0, \dots, |I'|\}$ telle que $x^{(k)}$ est réalisable pour $RKP(I', \Gamma')$.

Étape 4 : Calculer (ou borner) $\mathbb{P}(\sum_{i \in I'} w_i \leq c)$. Si $\mathbb{P}(\sum_{i \in I'} w_i \leq c) \geq 1 - \varepsilon$, STOP.

Étape 5 : Poser $k \leftarrow \Gamma' + 1$ et retourner à l'étape 2.

La valeur Γ' requise à l'étape 3 peut être calculée en temps linéaire. En effet, elle est solution du problème :

$$\begin{aligned} \max & \Gamma \\ \text{s.c.} & \sum_{i \in I'} \underline{w}_i + \max_{S \subseteq I': |S| = \Gamma} \sum_{i \in S} (\bar{w}_i - \underline{w}_i) \leq c \\ & \Gamma \in \{0, \dots, n\} \end{aligned}$$

Il est suffisant de considérer $S = \phi$ et d'augmenter cet ensemble progressivement en y ajoutant les indices correspondant aux plus grandes variations $\bar{w}_i - \underline{w}_i$. Dès que $\sum_{i \in S} \bar{w}_i - \underline{w}_i > c - \sum_{i \in I'} \bar{w}_i$, on a : $\Gamma' = |S| - 1$.

À l'étape 4, observer que : $\mathbb{P}(\sum_{i \in I'} w_i \leq c) = \mathbb{P}(w \cdot x^{(k)} \leq c)$. On assure donc que l'algorithme ne s'arrête qu'avec une solution réalisable pour CCKP. De plus, comme $x^{(k)}$ est une solution réalisable de RKP(I', k) (cf. lemme 19), on a $\Gamma' \geq k$. L'indice k croît strictement à chaque itération, ce qui assure la convergence : en effet, quand $k = n$, la probabilité de réalisabilité de $x^{(n)}$ est égale à 1 et l'algorithme s'arrête lors du test de l'étape 4. Ainsi :

Lemme 23 *L'algorithme HA_CCKP fournit une solution réalisable pour CCKP en résolvant au plus $n + 1$ problèmes de sac-à-dos robustes.*

Enfin, à l'étape 5, on passe directement à la valeur $\Gamma' + 1$ sans examiner les problèmes RKP(I, l) pour $k+1 \leq l \leq \Gamma'$. En effet, la solution $x^{(k)}$ est réalisable pour le problème RKP(I, Γ'), puisqu'elle est réalisable pour RKP(I', Γ').

Pour le calcul de probabilité de l'étape 4, on peut par exemple utiliser l'inégalité de Hoeffding (voir théorème 12), ou toute autre approximation. Dans certains cas, il est possible de calculer exactement la probabilité de réalisabilité d'une solution. Il est clair que la qualité du calcul de probabilité a un impact direct sur la qualité de la solution finale obtenue ; on détaillera ultérieurement, dans la partie 5.4.1, quelques techniques utilisables. Enfin, on peut noter que si à un moment on a $|I'| = \Gamma'$, alors la probabilité de réalisabilité de la solution courante est 1 et l'algorithme s'arrête (cf. lemme 19).

En conséquence du lemme 23, on a :

Proposition 14 *L'algorithme HA_CCKP s'arrête en temps pseudo-polynomial si l'on utilise l'algorithme de programmation dynamique décrit dans la preuve du théorème 15.*

Finalement, d'après le théorème 14, l'algorithme HA_CCKP conduit à une solution optimale si les poids et les profits sont triés respectivement en ordre croissant et décroissant. Il est donc probable que cette propriété sera encore vérifiée si les données satisfont "presque" ces hypothèses.

On remarque que les propriétés de l'algorithme HA_CCKP sont très proches de l'algorithme glouton utilisé classiquement pour résoudre les problèmes de sac-à-dos de manière heuristique (voir e.g. [52]). Cette méthode se présente donc comme un équivalent stochastique de l'algorithme glouton.

On termine avec un exemple illustratif de la manière dont l'algorithme fonctionne.

Exemple : Considérons le problème suivant :

$$\begin{aligned} \max \quad & \sum_{i=1}^{10} i x_i \\ \text{s.c.} \quad & \sum_{i=1}^{10} w_i x_i \leq 80, \\ & x \in \{0, 1\}^{10}, \end{aligned}$$

où les poids w_i sont tous uniformément distribués sur $[8, 12]$, i.e. : $\underline{w}_i = 8$ et $\bar{w}_i = 12$ pour tout $i \in I = \{1, \dots, 10\}$, et $\delta = 4$. Soit $\varepsilon = 0,9$, ce qui signifie qu'on cherche le meilleur remplissage du sac réalisable avec probabilité 0,9. Sous les hypothèses probabilistes effectuées, les probabilités $\mathbb{P}(\sum_{i \in I'} w_i \leq c)$ sont connues de manière analytique pour tout $I' \subseteq I$ (voir les formules de la partie 6.4.3). Dans la suite, on note e^i le vecteur tel que $e_i^i = 1$ et $e_j^i = 0$ pour $j \neq i$.

- k=0 : D'après le lemme 21, on sait que : $\sum_{i \in I} x_i^{(0)} = \lfloor 80/8 \rfloor = 10$, et donc : $x^{(0)} = \sum_{i=1}^{10} e^i$.
On a : $\Gamma' = \lfloor (80 - 10 * 8)/4 \rfloor = 0$, et on calcule : $\mathbb{P}(\sum_{i=1}^{10} w_i \leq 80) = 0$
(le seul événement possible est $w = \underline{w}$).
- k=1 : D'après le lemme 21 : $\sum_{i \in I} x_i^{(1)} = \lfloor (80 - 4)/8 \rfloor = 9$, et donc : $x^{(1)} = \sum_{i=2}^{10} e^i$.
On a : $\Gamma' = \lfloor (80 - 9 * 8)/4 \rfloor = 2$, et on calcule : $\mathbb{P}(\sum_{i=2}^{10} w_i \leq 80) = 0.001... < 0.9$.
- k=3 : Comme précédemment, on a : $\sum_{i \in I} x_i^{(3)} = \lfloor (80 - 12)/8 \rfloor = 8$, et donc : $x^{(3)} = \sum_{i=3}^{10} e^i$.
On a : $\Gamma' = \lfloor (80 - 8 * 8)/4 \rfloor = 4$, et : $\mathbb{P}(\sum_{i=3}^{10} w_i \leq 80) = 0,5 < 0,9$.
- k=5 : De la même manière : $\sum_{i \in I} x_i^{(5)} = \lfloor (80 - 20)/8 \rfloor = 7$, et : $x^{(5)} = \sum_{i=4}^{10} e^i$.
On calcule : $\mathbb{P}(\sum_{i=4}^{10} w_i \leq 80) = 0.99... > 0.9$.
STOP : $x^{(5)}$ est réalisable, et donc optimale ici.

Observer que si l'on avait requis une réalisabilité de 100%, on aurait dû mettre moins d'éléments dans le sac. En effet, si l'on considère que tous les poids prennent leurs pires valeurs, on peut placer au plus $\lfloor 80/12 \rfloor = 6$ éléments dans le sac.

4.5 Conclusion

On s'est intéressé dans cette partie au problème de sac-à-dos sous contrainte probabiliste. Le modèle robuste de Bertsimas et Sim a été utilisé comme outil de base pour approcher le problème sous contrainte probabiliste. Plusieurs liens théoriques entre les deux problèmes ont été établis. Puis un algorithme a été proposé pour résoudre le problème sous contrainte probabiliste de manière heuristique. Cet algorithme fournit une solution réalisable du problème en temps pseudo-polynomial. Par ailleurs, les propriétés de l'algorithme sont comparables à celles de l'algorithme glouton couramment utilisé pour le problème de sac-à-dos classique. On peut donc considérer qu'on a là une sorte d'équivalent probabiliste de cette méthode déterministe.

Des tests numériques de cet algorithme seront présentés dans le chapitre suivant, au paragraphe 5.5.2 (méthode dite "M2"). Cette approche sera alors comparée avec d'autres.

Chapitre 5

Généralisation aux problèmes combinatoires sous contraintes probabilistes

Motivations : Les idées du précédent chapitre vont maintenant être étendues à des problèmes linéaires en nombres 0-1 généraux. Ce sera l'occasion d'introduire un modèle robuste original, proche de celui de Bertsimas et Sim. Outre de bonnes propriétés théoriques en rapport avec la programmation sous contraintes probabilistes, ce nouveau modèle peut être résolu très efficacement de manière approchée. Il constitue donc un composant bien adapté pour des heuristiques de résolution de problèmes combinatoires sous contraintes probabilistes.

5.1 Modèles mathématiques

On considère le programme linéaire en nombres 0-1 suivant (ILP, pour *Integer Linear Program*) :

$$\begin{aligned} \max \quad & cx \\ \text{s.c.} \quad & Ax \leq b, \\ & x \in \{0, 1\}^n. \end{aligned} \tag{5.1}$$

On suppose que certains coefficients de la matrice sont incertains. Plus précisément, $I_u(j) \subseteq I$ désigne l'ensemble des coefficients de la contrainte $j \in J$ qui sont incertains ; les coefficients de $I \setminus I_u(j)$ sont supposés connus de manière exacte. Pour tout $j \in J$ et pour tout $i \in I_u(j)$, A_{ji} appartient à un intervalle $[\underline{A}_{ji}, \overline{A}_{ji}]$, avec $\overline{A}_{ji} > \underline{A}_{ji}$. On notera $\delta_{ji} = \overline{A}_{ji} - \underline{A}_{ji}$. (On rappelle que considérer que seuls des coefficients de A sont incertains englobe les cas où des coefficients de b et/ou de c sont incertains.)

Soit \mathcal{A} l'ensemble des matrices réelles A satisfaisant la description ci-dessus.

Chaque coefficient incertain est associé à une variable aléatoire, aussi notée A_{ji} , de support $[\underline{A}_{ji}, \overline{A}_{ji}]$. On note : $\eta_{ji} = (A_{ji} - \underline{A}_{ji})/\delta_{ji}$, qui est une variable aléatoire de support $[0, 1]$. On introduit alors le problème linéaire en nombres entiers sous contrainte probabiliste (CCILP,

pour *Chance-Constrained Integer Linear Program*) :

$$\begin{aligned} & \max && cx \\ & \text{s.c.} && \mathbb{P}(Ax \leq b) \geq 1 - \varepsilon, \\ & && x \in \{0, 1\}^n. \end{aligned} \quad (5.2)$$

On introduit également un problème robuste, paramétré par $\gamma \in [0, 1]^m$ et noté RILP(γ) (pour *Robust Integer Linear Program*) :

$$\begin{aligned} & \max && cx \\ & \text{s.c.} && \sum_{i \in I_u(j)} \underline{A}_{ji} x_i + \Delta_j(x) + \sum_{i \notin I_u(j)} A_{ji} x_i \leq b_j, \quad \forall j \in J, \\ & && x \in \{0, 1\}^n, \end{aligned} \quad (5.3)$$

avec :

$$\Delta_j(x) = \min \left\{ \sum_{i \in I_u(j)} \delta_{ji} x_i, \gamma_j \cdot \sum_{i \in I_u(j)} \delta_{ji} \right\}.$$

RILP(γ) sera souvent noté plus simplement RILP. $\Delta_j(x)$ est le terme de protection de la contrainte $j \in J$. Si $\gamma_j = 0$, ce terme de protection est nul : seul le meilleur scénario \underline{A}_j est pris en compte pour la contrainte j . Au contraire, si $\gamma_j = 1$, le scénario le plus défavorable, \bar{A}_j , est pris en compte.

Diverses extensions du cadre présenté sont possibles.

Problèmes mixtes en nombres entiers : Par souci de simplicité, tout ce chapitre est écrit pour des problèmes purs en nombres 0-1. Pourtant, tous les résultats s'étendent immédiatement aux problèmes mixtes de la forme suivante :

$$\begin{aligned} & \max && cx \\ & \text{s.c.} && Ax \leq b, \\ & && x_i \in \{0, 1\}, \quad \forall i \in I_1, \\ & && x_i \geq 0, \quad \forall i \in I_2, \end{aligned}$$

si l'on suppose que pour toute ligne j , l'incertitude ne concerne que des composantes entières de x , c'est-à-dire : $I_u(j) \subseteq I_1$.

S'il existe j tel que $I_u(j) \not\subseteq I_1$, le modèle robuste proposé a encore du sens si, pour tout $i \in I_u(j) \setminus I_1$: $x_i \in [0, 1]$. Plusieurs résultats présentés par la suite restent valides, en particulier ceux relatifs à la complexité ou aux calculs de probabilités. D'autres peuvent être adaptés (c'est le cas par exemple pour la reformulation linéaire de la partie 5.3.2). Pourtant, on perd dans ce cas la relation établie avec la programmation sous contraintes probabilistes (cf. partie 5.2), et les algorithmes décrits dans la partie 5.4.3 ne peuvent plus être utilisés directement.

Variabes en nombres entiers généraux : Soit $i \in I$, supposons que x_i prend ses valeurs dans \mathbb{N} , et qu'il existe $j \in J$ tel que $i \in I_u(j)$. Pour simplifier, on suppose que $x_l \in \{0, 1\}$ quand $l \neq i$. Sans perte de généralité, on peut supposer que x_i est borné : $x_i \in \{0, \dots, p\}$, $p \in \mathbb{N}$. On peut décomposer x_i à l'aide de variables 0-1 : $x_i = \sum_{k=1}^p x_i^k$, où $x_i^k \in \{0, 1\}$. On retrouve alors le cadre initial, et le terme de protection de la contrainte j s'écrit :

$$\begin{aligned} \Delta_j(x) &= \min \left\{ \sum_{l \in I_u(j) \setminus \{i\}} \delta_{jl} x_l + \delta_{ji} \left(\sum_{k=1}^p x_i^k \right), \gamma_j \cdot \left(\sum_{l \in I_u(j) \setminus \{i\}} \delta_{jl} + p \delta_{ji} \right) \right\} \\ &= \min \left\{ \sum_{l \in I_u(j)} \delta_{jl} x_l, \gamma_j \cdot \left(\sum_{l \in I_u(j) \setminus \{i\}} \delta_{jl} + p \delta_{ji} \right) \right\}. \end{aligned}$$

On en déduit la forme générale du terme de protection :

$$\Delta_j(x) = \min \left\{ \sum_{i \in I_u(j)} \delta_{ji} x_i, \gamma_j \cdot \sum_{i \in I_u(j)} \delta_{ji} \cdot \max\{x_i\} \right\}$$

Cette formule a encore du sens si on considère $x_i \geq 0$ pour certains indices $i \in I_u(j)$.

Contraintes probabilistes séparées : Le modèle (5.2) considère une contrainte de probabilité jointe sur l'ensemble du problème. Pourtant, certaines applications nécessitent la prise en compte de contraintes séparées :

$$\begin{aligned} & \max && cx \\ & \text{s.c.} && \mathbb{P}(A_j x \leq b_j) \geq 1 - \varepsilon_j, \quad \forall j \in J, \\ & && x \in \{0, 1\}^n. \end{aligned} \tag{5.4}$$

Encore dans un souci de simplicité, tous les résultats seront énoncés et prouvés pour le seul cas des contraintes de probabilité jointes. Mais tous se généralisent aisément au cas des contraintes séparées.

5.2 Lien entre RILP et CCILP

On montre ici un résultat similaire au théorème 13 énoncé pour le cas du problème de sac-à-dos, avec le modèle robuste de Bertsimas et Sim. Ce théorème se trouve en fait étendu et généralisé grâce au nouveau modèle robuste (5.3) introduit.

Soient $I' \subseteq I$ et $\gamma \in [0, 1]^m$. On note $\text{RILP}(I', \gamma)$ le problème :

$$\begin{aligned} & \max && cx \\ & \text{s.c.} && \sum_{i \in I_u(j) \cap I'} \underline{A}_{ji} x_i + \Delta_j(I', x) + \sum_{i \in I_u(j) \setminus I'} \bar{A}_{ji} x_i \\ & && \quad + \sum_{i \notin I_u(j)} A_{ji} x_i \leq b_j, \quad \forall j \in J, \\ & && x \in \{0, 1\}^n, \end{aligned} \tag{5.5}$$

où : $\Delta_j(I', x) = \min \left\{ \sum_{i \in I_u(j) \cap I'} \delta_{ji} x_i, \gamma_j \cdot \sum_{i \in I_u(j) \cap I'} \delta_{ji} \right\}$. Dans ce problème $\text{RILP}(I', \gamma)$, tous les coefficients $\{A_{ji}\}_{j \in J, i \notin I'}$ sont supposés prendre leur valeur maximale. Ainsi, pour la contrainte $j \in J$, le paramètre de robustesse γ_j a un impact sur les seuls coefficients $\{A_{ji}\}_{i \in I'}$. Noter que $\text{RILP}(\gamma)$ et $\text{RILP}(I, \gamma)$ désignent le même problème.

Théorème 16 *Il existe $I^* \subseteq I$ et $\gamma^* \in [0, 1]^m$ tels que toute solution optimale de $\text{RILP}(I^*, \gamma^*)$ est une solution optimale de CCILP.*

Preuve : Soit x^* une solution optimale de CCILP, on note : $I^* = \{i \in I \mid x_i^* = 1\}$. Pour simplifier les notations, on note aussi : $I^*(j) = I^* \cap I_u(j)$ pour tout $j \in J$. x^* est réalisable avec probabilité au moins $1 - \varepsilon$, i.e. : $\mathbb{P}(Ax^* \leq b) \geq 1 - \varepsilon$. Notons $\mathcal{A}^* = \{A \in \mathcal{A} \mid Ax^* \leq b\}$, et considérons le problème suivant :

$$\begin{aligned} & \max && cx \\ & \text{s.c.} && Ax \leq b, \quad \forall A \in \mathcal{A}^*, \\ & && x \in \{0, 1\}^n. \end{aligned} \tag{5.6}$$

Montrons d'abord que toute solution optimale x' de (5.6) est aussi optimale pour CCILP. On a : $\mathbb{P}(Ax' \leq b) \geq \mathbb{P}(A \in \mathcal{A}^*) = \mathbb{P}(Ax^* \leq b) \geq 1 - \varepsilon$. Ainsi, x' est réalisable pour CCILP, ce qui

implique : $cx^* \geq cx'$. Mais comme x^* est réalisable pour (5.6), on a également : $cx^* \leq cx'$. Donc x' est une solution réalisable de CCILP de même coût que x^* : x' est optimale pour CCILP.

Notre but maintenant est de trouver des paramètres $(\gamma_j^*)_{j \in J}$ tels que RILP(I^*, γ^*) est équivalent à (5.6). Soit $j \in J$, on a : $A_j x^* \leq b_j \Leftrightarrow \sum_{i \in I_u(j)} A_{ji} x_i^* \leq b_j - \sum_{i \notin I_u(j)} A_{ji} x_i^* \Leftrightarrow \sum_{i \in I_u(j)} \delta_{ji} \eta_{ji} x_i^* \leq b_j - \sum_{i \notin I_u(j)} A_{ji} x_i^* - \sum_{i \in I_u(j)} \underline{A}_{ji} x_i^*$. On note :

$$\gamma_j^* = \min \left\{ 1, \left(b_j - \sum_{i \notin I_u(j)} A_{ji} x_i^* - \sum_{i \in I_u(j)} \underline{A}_{ji} x_i^* \right) / \sum_{i \in I^*(j)} \delta_{ji} \right\}.$$

Observer que puisque $\mathcal{A}^* \neq \phi$, $\gamma_j^* \geq 0$. Alors : (a) $A \in \mathcal{A}^* \Leftrightarrow \forall j \in J, \sum_{i \in I^*(j)} \delta_{ji} \eta_{ji} \leq \gamma_j^* \sum_{i \in I^*(j)} \delta_{ji}$.

Prouvons que x est une solution réalisable de (5.6) si et seulement si x est réalisable pour RILP(I^*, γ^*). Soit x une solution réalisable de (5.6). Soit $j \in J$. Si $\sum_{i \in I^*(j)} \delta_{ji} x_i \leq \gamma_j^* \sum_{i \in I^*(j)} \delta_{ji}$, soit A_j défini par : $\eta_{ji} = x_i$ si $i \in I^*(j)$, et $\eta_{ji} = 1$ si $i \in I_u(j) \setminus I^*$. On a : $\sum_{i \in I^*(j)} \delta_{ji} \eta_{ji} = \sum_{i \in I^*(j)} \delta_{ji} x_i \leq \gamma_j^* \sum_{i \in I^*(j)} \delta_{ji}$.

Si au contraire $\sum_{i \in I^*(j)} \delta_{ji} x_i > \gamma_j^* \sum_{i \in I^*(j)} \delta_{ji}$, soit A_j défini par :

$$\begin{aligned} - \text{ si } i \in I^*(j) & : \eta_{ji} = x_i \cdot \gamma_j^* \cdot \sum_{k \in I^*(j)} \delta_{jk} / \sum_{k \in I^*(j)} \delta_{jk} x_k, \\ - \text{ si } i \in I_u(j) \setminus I^* & : \eta_{ji} = 1. \end{aligned}$$

Comme $\sum_{i \in I^*(j)} \delta_{ji} x_i > \gamma_j^* \sum_{i \in I^*(j)} \delta_{ji}$, on a pour tout $i \in I^*(j)$: $\eta_{ji} \leq x_i \leq 1$. Donc les composantes de η_j sont bien dans $[0, 1]$. De plus :

$$\begin{aligned} \sum_{i \in I^*(j)} \delta_{ji} \eta_{ji} & = \sum_{i \in I^*(j)} \delta_{ji} x_i \cdot \gamma_j^* \cdot \sum_{k \in I^*(j)} \delta_{jk} / \sum_{k \in I^*(j)} \delta_{jk} x_k \\ & = \gamma_j^* \cdot \sum_{k \in I^*(j)} \delta_{jk}. \end{aligned}$$

Il est donc montré que, pour tout $j \in J$: $\sum_{i \in I^*(j)} \delta_{ji} \eta_{ji} \leq \gamma_j^* \sum_{i \in I^*(j)} \delta_{ji}$.

D'après (a), on a donc construit une matrice A qui appartient à \mathcal{A}^* . Par ailleurs, pour $j \in J$, on peut vérifier que dans chacun des deux cas évoqués : $A_j x = \sum_{i \in I^*(j)} \underline{A}_{ji} x_i + \Delta_j(I^*, x) + \sum_{i \in I_u(j) \setminus I^*} \bar{A}_{ji} x_i + \sum_{i \notin I_u(j)} A_{ji} x_i$. Comme x est réalisable pour (5.6), on a : $A_j x \leq b_j$, ce qui implique que x est réalisable pour RILP(I^*, γ^*).

Réciproquement, soit x une solution réalisable de RILP(I^*, γ^*). Soit un quelconque $A \in \mathcal{A}^*$, on a pour tout $j \in J$: $A_j x = \sum_{i \in I_u(j)} (\underline{A}_{ji} + \delta_{ji} \eta_{ji}) x_i + \sum_{i \notin I_u(j)} A_{ji} x_i \leq \sum_{i \in I^*(j)} \underline{A}_{ji} x_i + \sum_{i \in I^*(j)} \delta_{ji} \eta_{ji} x_i + \sum_{i \in I_u(j) \setminus I^*} \bar{A}_{ji} x_i + \sum_{i \notin I_u(j)} A_{ji} x_i$. Puisque $A \in \mathcal{A}^*$, l'affirmation : $\sum_{i \in I^*(j)} \delta_{ji} \eta_{ji} \leq \gamma_j^* \sum_{i \in I^*(j)} \delta_{ji}$ est correcte. Ceci implique que, puisque $x_i \leq 1$ pour tout $i \in I^*(j)$: $\sum_{i \in I^*(j)} \delta_{ji} \eta_{ji} x_i \leq \gamma_j^* \sum_{i \in I^*(j)} \delta_{ji}$. D'autre part, puisque $\eta_{ji} \leq 1$ pour tout $i \in I^*(j)$: $\sum_{i \in I^*(j)} \delta_{ji} \eta_{ji} x_i \leq \sum_{i \in I^*(j)} \delta_{ji} x_i$. Il en résulte que : $\sum_{i \in I^*(j)} \delta_{ji} \eta_{ji} x_i \leq \Delta_j(I^*, x)$. On obtient donc : $A_j x \leq \sum_{i \in I^*(j)} \underline{A}_{ji} x_i + \Delta_j(I^*, x) + \sum_{i \in I_u(j) \setminus I^*} \bar{A}_{ji} x_i + \sum_{i \notin I_u(j)} A_{ji} x_i$. Comme x est réalisable pour RILP(I^*, γ^*), on obtient : $A_j x \leq b_j$. Ceci prouve que x est réalisable pour (5.6).

Finalement, on a montré que x est réalisable pour (5.6) si et seulement si x est réalisable pour RILP(I^*, γ^*). Comme les deux problèmes ont la même fonction-objectif, ils sont complètement équivalents. On en conclut que toute solution optimale de RILP(I^*, γ^*) est optimale pour CCILP. \square

Ceci établit un lien théorique entre CCILP et le modèle robuste RILP. Noter que ce résultat ne nécessite aucune hypothèse probabiliste d'aucune sorte.

5.3 Complexité théorique et pratique de RILP

5.3.1 Relations de complexité entre RILP et ILP

Si le problème nominal ILP est NP-difficile, alors il est clair que sa version robuste RILP l'est aussi (considérer $\gamma_j = 0$ pour tout $j \in J$). Des liens plus précis entre la complexité de ILP et celle de RILP peuvent être établis. Ci-après, m' désigne le nombre de contraintes $j \in J$ telles que $I_u(j) \neq \emptyset$.

Théorème 17 *Une solution optimale de RILP peut être obtenue par la résolution de $2^{m'}$ problèmes nominaux ILP.*

Preuve : Dans cette preuve, par souci de simplicité, on suppose que les ensembles $\{I_u(j)\}_{j \in J}$ sont tous non-vides. En conséquence : $m' = m$. Soit V la valeur optimale de RILP. Soit $\alpha \in \{0, 1\}^m$. Pour tout $j \in J$, on utilise dans le cadre de cette preuve la notation suivante : $\Delta_j(x, \alpha_j) = (1 - \alpha_j) \cdot \left(\sum_{i \in I_u(j)} \delta_{ji} x_i \right) + \alpha_j \cdot \left(\gamma_j \sum_{i \in I_u(j)} \delta_{ji} \right)$. On introduit alors :

$$V(\alpha) = \max_{\substack{cx \\ \text{s.c. } \sum_{i \in I_u(j)} A_{ji} x_i + \Delta_j(x, \alpha_j) + \sum_{i \notin I_u(j)} A_{ji} x_i \leq b_j, \quad \forall j \in J, \\ x \in \{0, 1\}^n.}}$$

Pour tout $\alpha \in \{0, 1\}^m$, une solution réalisable du problème ci-dessus est réalisable pour RILP, puisque $\Delta_j(x, \alpha_j) \geq \Delta_j(x)$. Il s'ensuit que : $\max\{V(\alpha) \mid \alpha \in \{0, 1\}^m\} \leq V$. De plus, il existe nécessairement $\alpha^* \in \{0, 1\}^m$ tel que $V \leq V(\alpha^*)$. En effet, étant donnée une solution optimale x^* de RILP, il est suffisant de considérer : $\alpha_j^* = 1 \Leftrightarrow \sum_{i \in I_u(j)} \delta_{ji} x_i^* \geq \gamma_j \sum_{i \in I_u(j)} \delta_{ji}$. Dans ce cas, x^* est solution réalisable du problème correspondant à $V(\alpha^*)$.

Ceci montre que : $V = \max\{V(\alpha) \mid \alpha \in \{0, 1\}^m\}$, ce qui achève la démonstration. \square

Si le nombre m' de contraintes sujettes à incertitudes n'a pas d'impact sur la complexité du problème ILP, *i.e.* si toute instance de ILP a un nombre fixe de contraintes incertaines, alors la version robuste RILP peut être résolue *via* un nombre constant $2^{m'}$ de problèmes ILP. Le problème de sac-à-dos est un exemple rentrant dans cette catégorie : sa version robuste sera résolue par seulement deux problèmes de sac-à-dos classique. La même observation peut être faite dans le cas où seuls les coefficients de l'objectif sont sujets à incertitude : 2 problèmes nominaux suffisent pour résoudre le problème robuste.

On notera au passage que la complexité obtenue est bien meilleure que celle du problème de Bertsimas et Sim, cf. [12] et le théorème 8 de la partie 3.2 (on était conduit à la résolution de $(n + 1)^m$ problèmes nominaux).

Le théorème 17 n'assure pas que l'éventuelle polynomialité du problème nominal sera conservée pour le problème robuste, puisqu'un nombre exponentiel 2^m de problèmes doit *a priori* être résolu. Cependant, dans certains cas spécifiques, ceci peut être nettement amélioré.

Cas des variations constantes :

Proposition 15 *Supposons que pour tout $j \in J$, $I_u(j) = I_u \subseteq I$, et pour tout $i \in I_u$, $\delta_{ji} = \delta_j$. Alors une solution optimale de RILP peut être obtenue par la résolution de $m' + 1$ problèmes nominaux ILP.*

Preuve : On continue la preuve précédente (avec $m = m'$), où il était montré que : $V = \max\{V(\alpha) \mid \alpha \in \{0, 1\}^m\}$. Comme précédemment, on note x^* une solution optimale de RILP, et $\alpha^* \in \{0, 1\}^m$ tel que : $\alpha_j^* = 1 \Leftrightarrow \sum_{i \in I_u(j)} \delta_{ji} x_i^* \geq \gamma_j \sum_{i \in I_u(j)} \delta_{ji}$. On sait que $V = V(\alpha^*)$.

On suppose maintenant que pour tout $j \in J$, $I_u(j) = I_u$ et toutes les variations vérifient : $\delta_{ji} = \delta_j$ pour $i \in I_u$. Sans perte de généralité, on suppose les indices triés de sorte que : $j < k \Rightarrow \gamma_j \leq \gamma_k$. Soit $j \in J$. Considérons le cas où $\sum_{i \in I_u} \delta_j x_i^* < \gamma_j \sum_{i \in I_u} \delta_j$, i.e. : $\alpha_j^* = 0$. Ceci implique que pour tout $k \geq j$: $\sum_{i \in I_u} x_i^* < \gamma_k |I_u|$, et donc : $\sum_{i \in I_u} \delta_k x_i^* < \gamma_k \sum_{i \in I_u} \delta_k$. On a alors : $\alpha_j^* = 0 \Rightarrow \forall k \geq j, \alpha_k^* = 0$.

On en déduit : $V = \max \{V(\alpha) \mid \alpha \in \{0, 1\}^m, j \leq k \Rightarrow \alpha_j \geq \alpha_k\}$. En conséquence, une solution optimale de RILP peut être obtenue *via* la résolution de $m + 1$ problèmes nominaux ILP, en conservant la meilleure des solutions calculées. \square

Dans ce dernier cas, si ILP est de complexité polynomiale en temps, on garde la même propriété pour RILP. Remarquer que la proposition 15 s'applique en particulier quand seuls les membres de droite d'un problème mixte en nombres entiers sont sujets à incertitude.

Cas des variations proportionnelles : Un autre cas mérite quelques remarques supplémentaires.

Supposons que pour tout $j \in J$, $I_u(j) \in \{\phi, I\}$, et dans le cas où $I_u(j) = I$, il existe $\kappa_j > 0$ tel que : $\forall i \in I, \delta_{ji} = \kappa_j \underline{A}_{ji}$. Tous les coefficients d'une ligne sont donc considérés incertains, et la variation d'un coefficient est proportionnelle à sa valeur minimale. Dans ce cas de figure, en pratique, RILP peut être résolu avec bien moins de $2^{m'}$ problèmes nominaux. En effet, on montre que plusieurs problèmes $V(\alpha)$ n'ont pas besoin d'être considérés.

Soit $\alpha \in \{0, 1\}^{m'}$, et soit $x(\alpha)$ la solution associée à $V(\alpha)$. Supposons que pour un quelconque $j \in J$, on a : $\alpha_j = 0$ et $\sum_{i \in I} \underline{A}_{ji} x(\alpha)_i + \gamma_j \sum_{i \in I} \delta_{ji} > b_j$. Considérons maintenant α' égal à α à l'exception de : $\alpha'_j = 1$, et soit $x(\alpha')$ la solution associée. Puisque $x(\alpha')$ est réalisable pour $V(\alpha')$, on a : $\sum_{i \in I} \underline{A}_{ji} x(\alpha')_i + \gamma_j \sum_{i \in I} \delta_{ji} \leq b_j$. De là, d'après nos hypothèses : $\sum_{i \in I} \underline{A}_{ji} [x(\alpha')_i - x(\alpha)_i] < 0$.

Supposons maintenant que $x(\alpha')$ n'est pas réalisable pour le problème $V(\alpha)$: ceci signifie que $\sum_{i \in I} \underline{A}_{ji} x(\alpha')_i + \sum_{i \in I} \delta_{ji} x(\alpha')_i > b_j$. La réalisabilité de $x(\alpha)$ pour $V(\alpha)$ entraîne que : $\sum_{i \in I} \underline{A}_{ji} x(\alpha)_i + \sum_{i \in I} \delta_{ji} x(\alpha)_i > \sum_{i \in I} \underline{A}_{ji} x(\alpha')_i + \sum_{i \in I} \delta_{ji} x(\alpha')_i$. Puisque les variations sont proportionnelles aux valeurs minimales, on obtient : $(1 + \kappa_j) \sum_{i \in I} \underline{A}_{ji} [x(\alpha')_i - x(\alpha)_i] > 0$. C'est une contradiction.

On en déduit que $x(\alpha')$ est réalisable pour $V(\alpha)$, ce qui implique : $V(\alpha') \leq V(\alpha)$. Donc il est inutile de s'intéresser au problème $V(\alpha')$. Un raisonnement similaire peut être produit quand $\alpha_j = 1$ et $\sum_{i \in I} \underline{A}_{ji} x(\alpha)_i + \sum_{i \in I} \delta_{ji} x(\alpha)_i > b_j$. En considérant α' égal à α à l'exception de : $\alpha'_j = 0$, on montre que $V(\alpha') \leq V(\alpha)$.

5.3.2 Une formulation linéaire en nombres entiers de RILP

La relaxation continue de RILP telle qu'écrite en (5.3) n'est pas linéaire, elle n'est pas même convexe :

Lemme 24 Soit $j \in J$, $x \mapsto \Delta_j(x)$ est concave sur $[0, 1]^n$.

Preuve : Supposons que $x = \lambda x^1 + (1 - \lambda)x^2$ pour un $\lambda \in [0, 1]$, on a : $\sum_{i \in I_u(j)} x_i = \lambda \sum_{i \in I_u(j)} x_i^1 + (1 - \lambda) \sum_{i \in I_u(j)} x_i^2$. Alors :

$$\begin{aligned} & \min \left\{ \sum_{i \in I_u(j)} \delta_{ji} x_i, \gamma_j \cdot \sum_{i \in I_u(j)} \delta_{ji} \right\} \\ &= \min \left\{ \lambda \sum_{i \in I_u(j)} \delta_{ji} x_i^1 + (1 - \lambda) \sum_{i \in I_u(j)} \delta_{ji} x_i^2, \lambda \gamma_j \cdot \sum_{i \in I_u(j)} \delta_{ji} + (1 - \lambda) \gamma_j \cdot \sum_{i \in I_u(j)} \delta_{ji} \right\} \\ &\geq \lambda \min \left\{ \sum_{i \in I_u(j)} \delta_{ji} x_i^1, \gamma_j \cdot \sum_{i \in I_u(j)} \delta_{ji} \right\} + (1 - \lambda) \min \left\{ \sum_{i \in I_u(j)} \delta_{ji} x_i^2, \gamma_j \cdot \sum_{i \in I_u(j)} \delta_{ji} \right\}. \end{aligned}$$

Ceci montre que : $\Delta_j(x) \geq \Delta_j(x^1) + \Delta_j(x^2)$. \square

Bien que ce lemme n'assure pas la concavité stricte de Δ_j en général, c'est ce qu'on a le plus souvent, comme l'illustre l'exemple très simple qui suit.

Exemple : On considère $n = 1$, $x^1 = 0$, $x^2 = 1$, $\delta = 1$ et $\lambda = 1/2$. On a donc $x = \lambda x^1 + (1 - \lambda)x^2 = 1/2$. Alors : $\min \{\delta x, \gamma \cdot \delta\} = \min\{1/2, \gamma\}$, et :

$$\lambda \min \{\delta x^1, \gamma \cdot \delta\} + (1 - \lambda) \min \{\delta x^2, \gamma \cdot \delta\} = 0 + 1/2 \cdot \min\{1, \gamma\} = \gamma/2$$

Alors, sauf dans le cas où $\gamma = 1$: $\min \{\delta x, \gamma \cdot \delta\} > \lambda \min \{\delta x^1, \gamma \cdot \delta\} + (1 - \lambda) \min \{\delta x^2, \gamma \cdot \delta\}$.

On donne cette reformulation équivalente de RILP :

Proposition 16 *RILP peut s'écrire :*

$$\begin{aligned} \max \quad & cx \\ \text{s.c.} \quad & \sum_{i \in I_u(j)} A_{ji} x_i + \left(\gamma_j \cdot \sum_{i \in I_u(j)} \delta_{ji} \right) \cdot z_j \\ & + \sum_{i \in I_u(j)} \delta_{ji} y_{ji} + \sum_{i \notin I_u(j)} A_{ji} x_i \leq b_j, \quad \forall j \in J, \\ & z_j + y_{ji} \geq x_i, \quad \forall j \in J, i \in I_u(j), \\ & x_i \in \{0, 1\}, \quad \forall i \in I, \\ & z_j \geq 0, y_{ji} \geq 0, \quad \forall j \in J, i \in I_u(j). \end{aligned} \quad (5.7)$$

Preuve : Pour tout $x \in \{0, 1\}^n$, il est facile de voir que quel que soit $j \in J$:

$$\Delta_j(x) = \min \left\{ \left(\gamma_j \cdot \sum_{i \in I_u(j)} \delta_{ji} \right) \cdot z_j + \left(\sum_{i \in I_u(j)} \delta_{ji} x_i \right) \cdot (1 - z_j) \mid z_j \in [0, 1] \right\}.$$

Ceci peut être linéarisé de manière classique en introduisant $y_{ji} = x_i \cdot (1 - z_j) = \max\{0, x_i - z_j\}$:

$$\begin{aligned} \Delta_j(x) = \min \quad & \left(\gamma_j \cdot \sum_{i \in I_u(j)} \delta_{ji} \right) \cdot z_j + \sum_{i \in I_u(j)} \delta_{ji} y_{ji} \\ \text{s.c.} \quad & z_j + y_{ji} \geq x_i, \quad \forall i \in I_u(j), \\ & y_{ji} \geq 0, \quad \forall i \in I_u(j), \\ & z_j \in [0, 1]. \end{aligned}$$

Finalement, les contraintes $z_j \leq 1$ peuvent être omises. En effet, toute solution optimale (x, y, z) de (5.7) peut être modifiée en (x, y', z') satisfaisant : $z'_j + y'_{ji} = x_i$, pour tous $j \in J$ et $i \in I_u(j)$. Comme $x_i \leq 1$ pour $i \in I$ et $y \geq 0$, ceci implique que $z'_j \leq 1$ pour tout $j \in J$. Ainsi, la contrainte $z_j \leq 1$ est inutile. \square

Ainsi, en relâchant les contraintes d'intégrité de (5.7), on obtient une relaxation linéaire de RILP ; on la note RILP-L.

Remarque : Les contraintes d'intégrité sur z ne peuvent être relâchées dans (5.7) que parce que x est un vecteur 0-1. Si le problème avait comporté des variables $x_i \in [0, 1]$ associées à au moins un coefficient incertain A_{ji} , la reformulation exacte du problème robuste aurait nécessité de conserver la contrainte $z_j \in \{0, 1\}$.

La formulation (5.7) peut être directement comparée avec celle du modèle de Bertsimas et Sim, qui s'écrit (cf. (3.4), partie 3.1) :

$$\begin{aligned} \max \quad & cx \\ \text{s.c.} \quad & \sum_{i \in I_u(j)} A_{ji} x_i + \Gamma_j z_j + \sum_{i \in I_u(j)} \delta_{ji} y_{ji} + \sum_{i \notin I_u(j)} A_{ji} x_i \leq b_j, \quad \forall j \in J \\ & z_j + \delta_{ji} y_{ji} \geq \delta_{ji} x_i, \quad \forall j \in J, i \in I_u(j) \\ & x_i \in \{0, 1\}, \quad \forall i \in I \\ & z_j \geq 0, y_{ji} \geq 0, \quad \forall j \in J, i \in I_u(j) \end{aligned} \quad (5.8)$$

avec les paramètres de robustesse $\Gamma_j \in [0, |I_u(j)|]$ pour $j \in J$. En particulier, si toutes les variations δ_{ji} ont la même valeur δ , alors ce modèle (5.8) est complètement équivalent à RILP : il suffit de considérer $\Gamma_j = \gamma_j |I_u(j)|$ pour le voir. Ceci explique notamment les relations entre les théorèmes 16 et 13 (voir partie 4.3).

Le modèle robuste (5.8) sera appelé par la suite RILP₂(Γ).

5.3.3 Description des ensembles d'incertitudes

En s'appuyant sur l'analyse rapportée dans la partie 2.4, on décrit l'ensemble des événements auxquels une solution de RILP est robuste. En particulier, on pourra ainsi se faire une idée plus intuitive des similarités et différences entre RILP et RILP₂. Intéressons-nous à une seule contrainte $j \in J$, et supposons que $I_u(j) = I$.

On note : $C_j = \gamma_j \sum_{i \in I} \delta_{ji} > 0$. On définit, pour tout $v \in \mathbb{R}^n$:

$$\begin{aligned} \|v\|_R = \min \quad & C_j \cdot \lambda + \sum_{i \in I} \mu_i \\ \text{s.c.} \quad & \delta_{ji} \cdot \lambda + \mu_i \geq |v_i|, \quad \forall i \in I, \\ & \lambda, \mu \geq 0. \end{aligned}$$

En utilisant les notations introduites au § 2.4 :

Lemme 25 *Pour tout $x \in \{0, 1\}^n$: $\Delta_j(x) = \|\delta_j \otimes x\|_R$.*

Preuve : Soit $x \in \{0, 1\}^n$. On a :

$$\begin{aligned} \|\delta_j \otimes x\|_R = \min \quad & C_j \cdot \lambda + \sum_{i \in I} \mu_i \\ \text{s.c.} \quad & \delta_{ji} \cdot \lambda + \mu_i \geq \delta_{ji} x_i, \quad \forall i \in I, \\ & \lambda, \mu \geq 0. \end{aligned}$$

Soit (λ, μ) une solution optimale de ce problème. On a : $\lambda = \max_{i \in I} ([\delta_{ji} x_i - \mu_i] / \delta_{ji})^+ \leq 1$. De la même manière, on a pour tout $i \in I$: $\mu_i = \delta_{ji} (x_i - \lambda)^+$. Si $x_i = 1$, comme $\lambda \leq 1$: $\mu_i = \delta_{ji} (1 - \lambda)$. Si $x_i = 0$, comme $\lambda \geq 0$: $\mu_i = 0$. Donc, pour tout $i \in I$: $\mu_i = (1 - \lambda) x_i$. On en déduit que la valeur de l'objectif s'écrit : $C_j \lambda + \sum_{i \in I} \delta_{ji} (1 - \lambda) x_i = \lambda C_j + (1 - \lambda) \cdot \sum_{i \in I} \delta_{ji} x_i$. La valeur minimale de cette quantité pour $\lambda \in [0, 1]$ est exactement $\Delta_j(x)$. \square

Lemme 26 *$\|\cdot\|_R$ est une norme.*

Preuve : On vérifie facilement que : $\|u\|_R = 0 \Leftrightarrow u = 0$. Soit un scalaire quelconque $t \in \mathbb{R}$. On voit que (λ, μ) est solution réalisable pour le problème $\|u\|_R$ si, et seulement si, $(|t| \cdot \lambda, |t| \cdot \mu)$ est réalisable pour le problème $\|tu\|_R$. On en déduit : $\|tu\|_R = |t| \cdot \|u\|_R$. Reste à montrer l'inégalité triangulaire. On a :

$$\begin{aligned} \|u + v\|_R = \min \quad & C_j \cdot \lambda + \sum_{i \in I} \mu_i \\ \text{s.c.} \quad & \delta_{ji} \cdot \lambda + \mu_i \geq |u_i + v_i|, \quad \forall i \in I, \\ & \lambda, \mu \geq 0. \end{aligned}$$

Soient (λ^1, μ^1) et (λ^2, μ^2) les solutions optimales respectives des problèmes $\|u\|_R$ et $\|v\|_R$. On a : $\|u\|_R + \|v\|_R = C_j \cdot (\lambda^1 + \lambda^2) + \sum_{i \in I} (\mu_i^1 + \mu_i^2)$. De plus, remarquer que : $\forall i \in I, \delta_{ji} \cdot (\lambda^1 + \lambda^2) + (\mu_i^1 + \mu_i^2) \geq |u_i| + |v_i|$, donc aussi : $\forall i \in I, \delta_{ji} \cdot (\lambda^1 + \lambda^2) + (\mu_i^1 + \mu_i^2) \geq |u_i + v_i|$. On en déduit que $(\lambda^1 + \lambda^2, \mu^1 + \mu^2)$ est une solution réalisable du problème $\|u + v\|_R$. Ceci implique : $\|u + v\|_R \leq C_j \cdot (\lambda^1 + \lambda^2) + \sum_{i \in I} (\mu_i^1 + \mu_i^2) = \|u\|_R + \|v\|_R$. \square

Lemme 27 *La norme duale de $\|\cdot\|_R$ est définie par : $\|v\|_R^* = \max \left\{ 1/C_j \cdot \|\delta_j \otimes v\|_1, \|v\|_\infty \right\}$.*

Preuve : On a par définition :

$$\|v\|_R^* = \max_{e: \|e\|_R \leq 1} v \cdot e = \max_{\text{s.c.}} \begin{cases} \sum_{i \in I} v_i e_i \\ C_j \lambda + \sum_{i \in I} \mu_i \leq 1, \\ \delta_{ji} \cdot \lambda + \mu_i \geq |e_i|, \quad \forall i \in I, \\ \lambda \geq 0, \mu \geq 0, e \in \mathbb{R}^n. \end{cases}$$

Remarquer qu'il existe une solution optimale de ce problème telle que : $\delta_{ji} \cdot \lambda + \mu_i = |e_i|$ pour tout $i \in I$. On supposera cette propriété vérifiée par la suite.

Montrons qu'il existe une solution optimale de ce problème qui vérifie : $\lambda = 0$ ou $\mu = 0$. En effet, supposons que pour tout $i \in I$, $\delta_{ji} \geq C_j$. Alors, s'il existe $k \in I$ tel que $\mu_k > 0$, on construit la solution (λ', μ', e') telle que :

- μ' est égal à μ sauf pour la composante k : $\mu'_k = 0$;
- e' est égal à e à l'exception de : $|e'_k| = |e_k| + (\delta_{jk}/C_j - 1)\mu_k$;
- $\lambda' = \lambda + \mu_k/C_j$.

(λ', μ', e') est une solution réalisable telle que : $e' \cdot v = e \cdot v + |(\delta_{jk}/C_j - 1)\mu_k v_k| \geq e \cdot v$. De plus, $\mu'_k = 0$. Ceci montre qu'il existe une solution optimale (λ, μ) telle que $\mu = 0$. Si au contraire il existe $k \in I$ tel que $\delta_{jk} < C_j$, on pose :

- $\lambda' = 0$;
- μ' est égal à μ , excepté : $\mu'_k = \mu_k + C_j \lambda$;
- e' est égal à e à l'exception de : $|e'_k| = |e_k| + (C_j - \delta_{jk})\lambda$.

(λ', μ', e') est une solution réalisable telle que : $e' \cdot v = e \cdot v + |(C_j - \delta_{jk})\lambda v_k| > e \cdot v$. Dans ce cas, il existe donc une solution optimale (λ, μ) telle que $\lambda = 0$.

Si $\lambda = 0$, on a : $\|v\|_R^* = \|v\|_\infty$. Si $\mu = 0$, on a : $\|v\|_R^* = 1/C_j \cdot \|\delta_j \otimes v\|_1$. On en déduit : $\|v\|_R^* = \max \{ \|v\|_\infty, 1/C_j \cdot \|\delta_j \otimes v\|_1 \}$. \square

D'après le théorème 6, on a :

$$\underline{A}_j x + \Delta_j(x) \leq b_j \Leftrightarrow \left\{ \forall A_j \text{ t.q. } \|(A_j - \underline{A}_j) \otimes \delta_j\|_R^* \leq 1 : A_j x \leq b_j \right\}.$$

Observer que : $\|(A_j - \underline{A}_j) \otimes \delta_j\|_R^* \leq 1 \Leftrightarrow \begin{cases} \|(A_j - \underline{A}_j) \otimes \delta_j\|_\infty \leq 1 \\ \|A_j - \underline{A}_j\|_1 \leq C_j \end{cases}$. Sur la figure 5.1, on donne une illustration de la forme des ensembles d'incertitude dans un espace à 3 dimensions ($n = 3$). L'ensemble de gauche correspond à l'ensemble d'incertitude du modèle de RILP₂ pour $\Gamma_j = 1, 5$; celui de droite représente l'ensemble d'incertitude de le modèle robuste RILP pour $\gamma_j = 0, 5$. Il s'agit dans les deux cas d'un sous-ensemble du pavé des événements possibles, délimité par un plan dont l'inclinaison change suivant le modèle de robustesse.

5.4 Algorithmes efficaces pour CCILP

5.4.1 Calcul de la probabilité de réalisabilité d'une solution

Une des difficultés fondamentales pour la résolution de problèmes sous contraintes probabilistes réside dans l'évaluation de la probabilité $\mathbb{P}(Ax \leq b)$ pour une solution x donnée. Dans certains cas particuliers, on connaît des formules analytiques qui permettent de calculer exactement les probabilités voulues (voir par exemple la partie 6.4.3). Mais la plupart du temps, une évaluation exacte requiert des intégrations multi-dimensionnelles compliquées numériquement ; de telles approches directes sont souvent trop difficiles si l'on souhaite un processus de résolution raisonnablement rapide. Ceci motive l'approximation de $\mathbb{P}(Ax \leq b)$, qui peut être effectuée de différentes manières. On présente ci-après trois approches, qui sont utilisables en fonction des hypothèses probabilistes effectuées et/ou du degré de précision voulu.

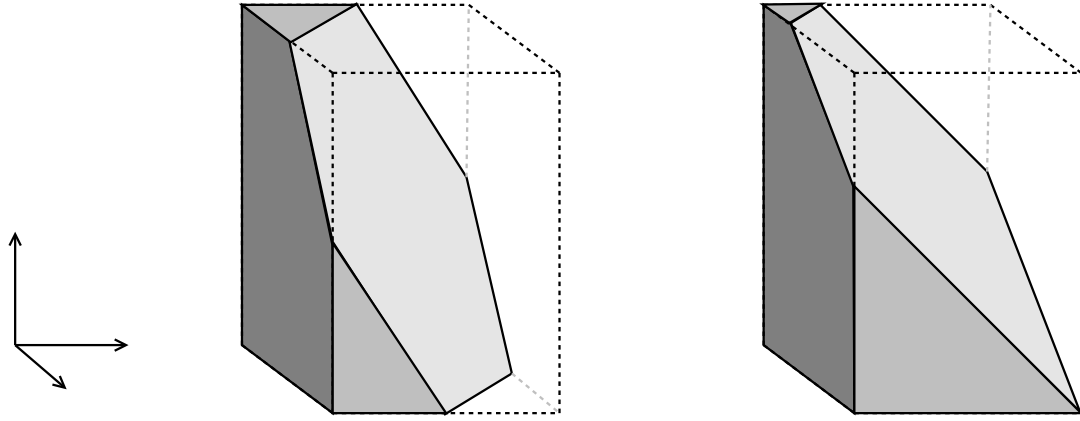


FIG. 5.1 – Représentation d’ensembles d’incertitudes pour le modèle de Bertsimas et Sim (à gauche), et pour le modèle robuste (5.3) à droite.

Inégalités de probabilités : Une première idée couramment évoquée dans la littérature est de s’appuyer sur des bornes de probabilités pour garantir une approximation sûre de la solution obtenue (voir e.g. [16, 82]). Les plus connues de ces inégalités sont sans doute celle de Markov et celle de Chebychev (voir tout livre traitant des probabilités) :

Théorème 18 (Inégalités de Markov et de Chebychev)

(i) Soit X une variable aléatoire à valeurs positives, pour tout réel $a > 0$: $\mathbb{P}(X \geq a) \leq \mathbb{E}[X]/a$ (inégalité de Markov).

(ii) Soit X une variable aléatoire d’espérance μ et de variance σ^2 finies, alors pour tout $a > 0$: $\mathbb{P}(|X - \mu| \geq a) \leq \sigma^2/a$ (inégalité de Chebychev).

Leur caractère extrêmement général rend ces inégalités intéressantes en théorie. En pratique, elles peuvent s’avérer très faibles. On dispose d’ailleurs souvent d’un peu plus d’information probabiliste. Il est souvent pertinent d’utiliser l’inégalité de Hoeffding, déjà présentée auparavant (voir théorème 12, § 4.3), et dont on déduit par exemple :

Lemme 28 Soit $x \in \mathbb{R}_+^n$. Soit $j \in J$, on note : $\mu_j(x) = \mathbb{E} \left[\sum_{i \in I_u(j)} A_{ji} x_i \right]$. Supposons que : $b_j \geq \mu_j(x) + \sum_{i \notin I_u(j)} A_{ji} x_i$, et que les variables aléatoires $\{A_{ji}\}_{i \in I_u(j)}$ sont indépendantes :

$$\mathbb{P}(A_j x \geq b_j) \leq \exp \left(- \frac{2d_j(x)^2}{\sum_{i \in I_u(j)} \delta_{ji}^2 x_i^2} \right) \quad (5.9)$$

où : $d_j(x) = b_j - \sum_{i \notin I_u(j)} A_{ji} x_i - \mu_j(x)$.

Dans le cas de contraintes de probabilités jointes, il s’agit d’évaluer $\mathbb{P}(Ax \leq b) = \mathbb{P}(\forall j \in J, A_j x \leq b_j)$. Par exemple, si on suppose l’indépendance de toutes les variables aléatoires $\{A_{ji}\}_{j \in J, i \in I_u(j)}$ du problème, on a : $\mathbb{P}(Ax \leq b) = \prod_{j \in J} \mathbb{P}(A_j x \leq b_j)$.

L’usage de telles bornes analytiques $B(x) \leq \mathbb{P}(Ax \leq b)$ signifie que le programme CCILP est en fait approché par le problème suivant :

$$\begin{aligned} \max \quad & cx \\ \text{s.c.} \quad & B(x) \geq 1 - \varepsilon \\ & x_i \in \{0, 1\}, \quad \forall i \in I \end{aligned} \quad (5.10)$$

Ce problème est le plus souvent non-linéaire (sauf dans le cas de l'inégalité de Markov), et donc en général très difficile à résoudre en pratique. Mais dans certains cas, ce problème non-linéaire en nombres entiers peut être linéarisé. C'est le cas, par exemple, si l'on utilise l'inégalité (5.9) pour un problème à contraintes de probabilités séparées (cf. modèle (5.4)). En effet, sous la condition $\mathbb{E}[A_j x] \leq b_j$, la contrainte $\mathbb{P}(A_j x \leq b_j) \geq 1 - \varepsilon_j$ devient : $\exp\left(-\frac{2d_j(x)^2}{\sum_{i \in I_u(j)} \delta_{ji}^2 x_i^2}\right) \leq \varepsilon_j \Leftrightarrow -2d_j(x)^2 \leq \ln(\varepsilon_j) \cdot \sum_{i \in I_u(j)} \delta_{ji}^2 x_i^2$. Puisque toutes les variables sont 0-1, on a : $x_i^2 = x_i$ pour tout $i \in I_u(j)$, et on peut introduire selon le procédé classique des variables $y_{kl} = x_k x_l$ pour tout $(k, l) \in I_u(j)^2$ tel que $k < l$. Il faut alors aussi ajouter les contraintes : $y_{kl} \leq x_k$ et $y_{kl} \leq x_l$.

Par ailleurs, il est nécessaire d'ajouter des contraintes qui définissent le domaine de validité de l'inégalité (5.9) : $\mu_j(x) + \sum_{i \notin I_u(j)} A_{ji} x_i \leq b_j$ pour tout $j \in J$. La linéarisation réalisée implique l'ajout de $n(n-1)/2$ nouvelles variables continues, ainsi que $n(n-1)$ nouvelles contraintes. On peut alors écrire l'approximation suivante du problème (5.4) :

$$\begin{aligned}
& \max && cx \\
& \text{s.c.} && \sum_{i \in I} \left(4b_j \mathbb{E}[A_{ji}] - 2\mathbb{E}[A_{ji}]^2 - \ln(\varepsilon_j) \delta_{ji}^2 \right) \cdot x_i \\
& && -2 \sum_{(k,l) \in I^2, i \neq k} \mathbb{E}[A_{jk}] \cdot \mathbb{E}[A_{jl}] \cdot y_{kl} \leq 2b_j^2, \quad \forall j \in J, \\
& && \sum_{i \in I} \mathbb{E}[A_{ji}] x_i \leq b_j, \quad \forall j \in J, \\
& && y_{kl} \leq x_k, \quad \forall (k, l) \in I^2, k \neq l, \\
& && y_{kl} \leq x_l, \quad \forall (k, l) \in I^2, k \neq l, \\
& && x \in \{0, 1\}^n, y \in \mathbb{R},
\end{aligned} \tag{5.11}$$

avec la convention naturelle : $\delta_{ji} = 0$ si $i \notin I_u(j)$. Un tel modèle sera appelé "approximation de Hoeffding" d'un problème avec contraintes de probabilités séparées. Cette approche est particulièrement appropriée quand on dispose de très peu d'information sur la distribution des variables aléatoires du problème.

Méthodes d'échantillonnage : Si les distributions en probabilité sont connues, les techniques de type Monte-Carlo [68] peuvent donner de bonnes estimations de $\mathbb{P}(Ax \leq b)$. En effet, d'après la loi faible des grands nombres, la taille de l'échantillon peut être calculée de manière à garantir une qualité contrôlée pour l'évaluation de $\mathbb{P}(Ax \leq b)$. Plus précisément :

Lemme 29 *Soit $q \in \mathbb{N}$. Soit Θ une variable aléatoire à valeurs dans \mathbb{R}^q , et soit $t \in \mathbb{R}^q$. On note : $\Psi = \mathbb{P}(\Theta \leq t)$. Soit μ_1, \dots, μ_k un échantillon aléatoire de la densité de probabilité $\mathbb{1}_{\{\Theta \leq t\}}$, on note : $\bar{\mu} = \sum_{p=1}^k \mu_p / k$. De plus, on introduit $d = |\bar{\mu} - 1 + \varepsilon|$:*

$$\begin{aligned}
& - \text{si } \Psi > 1 - \varepsilon, \text{ on a : } \Pr\left(\bar{\mu} < 1 - \varepsilon\right) \leq \frac{1}{4kd^2}; \\
& - \text{si } \Psi < 1 - \varepsilon, \text{ on a : } \Pr\left(\bar{\mu} > 1 - \varepsilon\right) \leq \frac{1}{4kd^2}.
\end{aligned}$$

Preuve : Notons : $T = \mathbb{1}_{\{\Theta \leq t\}}$, on a : $\Psi = \mathbb{E}[T]$. Supposons que $\Psi > 1 - \varepsilon$, alors : $\Pr\left(\bar{\mu} < 1 - \varepsilon\right) \leq \Pr\left(|\bar{\mu} - \Psi| > d\right)$. D'après la loi faible des grands nombres, on obtient : $\Pr\left(\bar{\mu} < 1 - \varepsilon\right) \leq \sigma^2 / (kd^2)$, où σ^2 est la variance de T . De plus, on a : $\sigma^2 = \mathbb{E}[T^2] - \mathbb{E}[T]^2 = \mathbb{E}[T^2] - \Psi^2$. Puisque T est à valeurs dans $\{0, 1\}$, on obtient en définitive : $\sigma^2 = \mathbb{E}[T] - \Psi^2 = \Psi(1 - \Psi) \leq 1/4$.

Le cas où $\Psi < 1 - \varepsilon$ est complètement similaire. \square

Ainsi, si la taille k de l'échantillon est suffisamment importante, $\bar{\mu}$ et $\mathbb{P}(Ax \leq b)$ sont ou bien tous deux supérieurs à $1 - \varepsilon$, ou bien tous deux inférieurs, avec une forte probabilité. Cette

approche comporte plusieurs avantages pratiques : elle est très facile à mettre en œuvre, et elle permet de prendre aisément en compte des hypothèses probabilistes variées (corrélations par exemple). Évidemment, son principal inconvénient est d'imposer un nombre d'échantillons très important si l'on veut une garantie assez élevée de bien évaluer la probabilité, ce qui se traduit par une durée longue de simulation.

Approximations gaussiennes : Enfin, la probabilité $\mathbb{P}(Ax \leq b)$ peut être évaluée de manière analytique si les variables aléatoires A_{ji} sont toutes supposées suivre des lois normales [46, 16, 45]. Dans notre cas, on a supposé que chaque variable aléatoire A_{ji} appartenait à un intervalle $[\underline{A}_{ji}, \overline{A}_{ji}]$. Une manière d'approcher ce comportement est de supposer une distribution normale telle que la probabilité $\mathbb{P}(A_{ji} \notin [\underline{A}_{ji}, \overline{A}_{ji}])$ est "petite". Cette approche, séduisante car elle rend des calculs exacts possibles, semble souvent très contraignante d'un point de vue modélisation. Dans certains cas, pourtant, le recours au théorème de la limite centrale donne de bonnes raisons d'effectuer des approximations gaussiennes (voir annexe B.2).

Plus généralement, ce traitement analytique du problème est possible dès que la distribution de probabilité des variables aléatoires est *stable*, c'est-à-dire : si X_1 et X_2 sont des variables aléatoires chacune distribuée selon une distribution stable (de paramètres différents pour chaque variable), alors $X_1 + X_2$ est aussi distribuée selon cette loi, avec des paramètres différents bien sûr (voir [45] pour plus de détails). Typiquement, si $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ (loi normale de moyenne μ_1 et de variance σ_1^2) et si $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$, alors $X_1 + X_2 \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$. La loi de Cauchy est un autre exemple de distribution stable.

Enfin, notons que [20] a proposé d'intéressantes extensions de ce cas de figure gaussien à une famille plus générale de distributions dites "radiales". Les auteurs montrent que pour ces distributions, les contraintes probabilistes séparées peuvent s'écrire comme des contraintes coniques convexes de second ordre. Dans notre cadre 0-1, de telles contraintes pourraient être linéarisées comme décrit plus haut.

5.4.2 Difficultés et restrictions

Dans ce qui suit, l'accent est mis sur la simplicité de mise en œuvre et le passage à l'échelle de la méthode de résolution. Notre but est de trouver une bonne solution de CCILP à partir de la famille de problèmes robustes RILP. D'après le théorème 16, trouver la solution optimale de CCILP impliquerait de considérer tous les problèmes $\text{RILP}(I', \gamma)$, pour tous les sous-ensembles $I' \subseteq I$ et pour tout $\gamma \in [0, 1]^m$. Ceci serait plus coûteux encore que d'énumérer les points de $\{0, 1\}^n$. Pour conserver une complexité algorithmique acceptable, notre recherche se limitera aux problèmes $\text{RILP}(I, \gamma)$, aussi notés $\text{RILP}(\gamma)$. On pourrait imaginer des améliorations consistant à tester différents sous-ensembles I' , sans bien sûr chercher à tous les énumérer. Ceci rendrait toutefois les temps de calcul moins attractifs.

On détaille ci-après les limitations de l'approche proposée.

Sous-optimalité de la solution trouvée : Malheureusement, la restriction ci-dessus peut conduire à manquer la solution optimale du programme sous contrainte probabiliste, ce qu'illustre l'exemple suivant.

Exemple : Considérer le problème de sac-à-dos suivant :

$$\begin{aligned} \max \quad & 2x_1 + 2x_2 + 3x_3 \\ \text{s.c.} \quad & w_1x_1 + w_2x_2 + w_3x_3 \leq 2, 5, \\ & x \in \{0, 1\}^3, \end{aligned}$$

où les poids w_i sont supposés indépendants et uniformément distribués sur leurs intervalles respectifs d'existence : $w_1 \in [0, 5; 1, 5]$, $w_2 \in [0, 5; 1, 5]$ et $w_3 \in [1, 6; 2, 6]$. Supposons que nous voulions résoudre le problème sous contrainte probabiliste correspondant avec $\varepsilon = 0, 1$.

En énumérant toutes les solutions et en calculant leurs probabilités respectives de réalisabilité, on trouve que la solution optimale est : $x^* = (0, 0, 1)$, de valeur associée $V^* = 3$ et de probabilité de réalisabilité $P^* = 0, 9$. Maintenant, écrivons le problème robuste pour $\gamma \in [0, 1]$:

$$\begin{aligned} \max & && 2x_1 + 2x_2 + 3x_3 \\ \text{s.c.} & && 0, 5x_1 + 0, 5x_2 + 1, 6x_3 + \min\{x_1 + x_2 + x_3, 3\gamma\} \leq 2, 5, \\ & && x \in \{0, 1\}^3. \end{aligned}$$

On peut vérifier que :

- si $3\gamma \leq 1, 5$, la solution robuste est $(1, 1, 0)$, de valeur $V = 4$ et de probabilité $P = 7/8 < 0, 9$ (voir l'annexe B.1 pour un exemple de calcul proche),
- si $3\gamma > 1, 5$, la solution robuste est $(1, 0, 0)$, de valeur $V = 2$ et de probabilité $P = 1$.

Ainsi, la solution x^* n'est optimale pour aucun des problèmes robustes considérés. Pourtant, observer que x^* est bien une solution optimale du problème RILP(I', γ) avec $I' = \{3\}$ et $\gamma = 0$.

Non-monotonie de la probabilité de réalisabilité : Il aurait été utile de caractériser le comportement de la probabilité de réalisabilité $\mathbb{P}(Ax \leq b)$. Plus précisément, pour $\gamma \in [0, 1]^m$, on note $x(\gamma)$ une solution optimale de RILP(γ). Il est clair que si $\gamma = (1, \dots, 1)$, alors $\mathbb{P}(Ax(\gamma) \leq b) = 1$. Il aurait semblé naturel que lorsque $\gamma \leq \gamma'$, alors $\mathbb{P}(Ax(\gamma) \leq b) \geq \mathbb{P}(Ax(\gamma') \leq b)$. Malheureusement, ce n'est pas toujours le cas, comme montré dans l'exemple suivant. Ceci rend les méthodes de recherche basées sur la dichotomie peu sûres.

Exemple : Considérer le problème de sac-à-dos suivant :

$$\begin{aligned} \max & && 2x_1 + 2x_2 + 3x_3 \\ \text{s.c.} & && w_1x_1 + w_2x_2 + w_3x_3 \leq 3, 5, \\ & && x \in \{0, 1\}^3, \end{aligned}$$

où les poids w_i sont supposés indépendants et uniformément distribués sur leurs intervalles respectifs d'existence : $w_1 \in [1, 2]$, $w_2 \in [1, 2]$ et $w_3 \in [1, 5; 5, 5]$. Le problème robuste associé à $\gamma \in [0, 1]$ est :

$$\begin{aligned} \max & && 2x_1 + 2x_2 + 3x_3 \\ \text{s.c.} & && x_1 + x_2 + 1, 5x_3 + \min\{x_1 + x_2 + 4x_3, 6\gamma\} \leq 3, 5, \\ & && x \in \{0, 1\}^3. \end{aligned}$$

Les solutions robustes peuvent être décrites pour toutes les valeurs de γ :

- si $\gamma = 0$, la solution robuste est $(1, 1, 1)$, de valeur $V = 7$ et de probabilité $P = 0$;
- si $0 < 6\gamma \leq 1$, la solution robuste est $(1, 0, 1)$, de valeur $V = 5$ et de probabilité $P = 1/8$;
- si $1 < 6\gamma \leq 1, 5$, la solution robuste est $(1, 1, 0)$, de valeur $V = 4$ et de probabilité $P = 7/8$;
- si $1, 5 < 6\gamma \leq 2$, la solution robuste est $(0, 0, 1)$, de valeur $V = 3$ et de probabilité $P = 1/2$;
- si $2 < 6\gamma$, la solution robuste est $(0, 0, 0)$, de valeur $V = 0$ et de probabilité $P = 1$.

Cette énumération montre que la probabilité de réalisabilité n'est pas croissante avec γ , même pour ce cas simple de problème de sac-à-dos.

5.4.3 Algorithmes

Malgré ces inconvénients, l'approche robuste proposée demeure intéressante en raison de son très faible coût algorithmique. On détaille maintenant quelques implémentations possibles. On suppose dans cette partie que \underline{A} est une matrice d'entiers. Remarquons pour commencer :

Lemme 30 Soit x^* une solution optimale de $RILP(\gamma)$. x^* est aussi optimal pour $RILP(\gamma')$, où pour tout $j \in J$:

- si $\sum_{i \in I_u(j)} \delta_{ji} x_i^* \leq \gamma_j \sum_{i \in I_u(j)} \delta_{ji} : \gamma_j' = 1$,
- sinon : $\gamma_j' = \left(b_j - \sum_{i \in I_u(j)} \underline{A}_{ji} x_i^* - \sum_{i \notin I_u(j)} \underline{A}_{ji} x_i^* \right) / \sum_{i \in I_u(j)} \delta_{ji}$.

Preuve : Le fait que x^* est réalisable pour $RILP(\gamma')$ est immédiat. De plus, puisque $\gamma' \geq \gamma$, $RILP(\gamma)$ est une relaxation de $RILP(\gamma')$. Donc x^* est optimal pour $RILP(\gamma')$. \square

Ce résultat montre qu'il est inutile de tester les paramètres de robustesse situés entre γ et γ' , puisqu'ils ont tous deux des solutions optimales en commun.

Lemme 31 Soit $\gamma \in [0, 1]^m$, et soit γ' calculé comme décrit dans le lemme 30. Considérons γ'' égal à γ' , à l'exception de : $\gamma_j'' = \min \left\{ 1, \gamma_j' + 1 / \sum_{i \in I_u(j)} \delta_{ji} \right\}$ pour un indice $j \in J$ quelconque tel que $\gamma_j' < 1$. Soit x'' une solution optimale de $RILP(\gamma'')$. Pour tout $g \in [0, 1]^m$ tel que $g \neq \gamma'$ et $\gamma' \leq g \leq \gamma''$, x'' est une solution optimale de $RILP(g)$.

Preuve : Soit x une solution optimale associée à γ , on définit γ' à partir du lemme 30 : $\sum_{i \in I} \underline{A}_{ji} x_i + \gamma_j' \sum_{i \in I_u(j)} \delta_{ji} = b_j$. Considérons un quelconque $g \in [0, 1]^m$ tel que $\gamma' \leq g \leq \gamma''$ et $g \notin \{\gamma', \gamma''\}$. Notons \hat{x} une solution optimale de $RILP(g)$. Montrons que \hat{x} est une solution réalisable de $RILP(\gamma'')$. Si on a : $\sum_{i \in I} \underline{A}_{ji} \hat{x}_i + \sum_{i \in I_u(j)} \delta_{ji} \hat{x}_i \leq b_j$, alors c'est clair puisque : $\min \left\{ \gamma_j'' \sum_{i \in I_u(j)} \delta_{ji}, \sum_{i \in I_u(j)} \delta_{ji} \hat{x}_i \right\} \leq \sum_{i \in I_u(j)} \delta_{ji} \hat{x}_i$. Supposons maintenant que l'on ait seulement : $\sum_{i \in I} \underline{A}_{ji} \hat{x}_i + g_j \sum_{i \in I_u(j)} \delta_{ji} \leq b_j$. Ceci implique : $\sum_{i \in I} \underline{A}_{ji} \hat{x}_i - \sum_{i \in I} \underline{A}_{ji} x_i \leq (\gamma_j' - g_j) \sum_{i \in I_u(j)} \delta_{ji} < 0$. Comme tous les coefficients $\{\underline{A}_{ji}\}_{i \in I}$ ont été supposés entiers, on a en fait : $\sum_{i \in I} \underline{A}_{ji} \hat{x}_i - \sum_{i \in I} \underline{A}_{ji} x_i \leq -1$, ou de manière équivalente : $\sum_{i \in I} \underline{A}_{ji} \hat{x}_i + 1 \leq \sum_{i \in I} \underline{A}_{ji} x_i$. Alors : $\sum_{i \in I} \underline{A}_{ji} \hat{x}_i + \gamma_j' \sum_{i \in I_u(j)} \delta_{ji} + 1 \leq \sum_{i \in I} \underline{A}_{ji} x_i + \gamma_j' \sum_{i \in I_u(j)} \delta_{ji} = b_j$. En conséquence : $\sum_{i \in I} \underline{A}_{ji} \hat{x}_i + \gamma_j'' \sum_{i \in I_u(j)} \delta_{ji} \leq b_j$, et donc \hat{x} est une solution réalisable de $RILP(\gamma'')$.

Comme \hat{x} est solution optimale de $RILP(g)$ et $\gamma'' \geq g$, la réalisabilité de \hat{x} pour $RILP(\gamma'')$ implique son optimalité pour ce même problème. On a donc montré que $RILP(\gamma'')$ et $RILP(g)$ ont la même valeur optimale. De plus, x'' est réalisable pour $RILP(g)$, puisque $g \leq \gamma''$. Ainsi, x'' est optimal pour $RILP(g)$. \square

Ceci montre qu'il est également inutile de tester les paramètres robustes "entre" γ' et γ'' . Les lemmes 30 et 31 sont utilisés pour concevoir l'algorithme suivant :

Algorithme rapide pour CCILP

Étape 0 : Poser $\gamma = 0$.

Étape 1 : Résoudre $RILP(\gamma)$; soit x la solution optimale trouvée.

Étape 2 : Calculer (ou borner) $\mathbb{P}(Ax \leq b)$.

Étape 3 : Si $\mathbb{P}(Ax \leq b) < 1 - \varepsilon$:

calculer $\gamma' \geq \gamma$ comme décrit dans le lemme 30, et mettre à jour $\gamma \leftarrow \gamma'$;

choisir j tel que : $j \in \arg \min_{k \in J: \gamma_k' < 1} \left\{ b_k - \sum_{i \in I} \mathbb{E}[A_{ki}] x_k \right\}$;

poser $\gamma_j \leftarrow \gamma_j + 1 / \sum_{i \in I_u(j)} \delta_{ji}$ et aller à l'étape 1.

Étape 4 : Si $\mathbb{P}(Ax \leq b) \geq 1 - \varepsilon$: STOP.

L'idée est de considérer une suite croissante de paramètres de robustesse γ . À chaque itération, le problème robuste est résolu (étape 1). On évalue alors la probabilité de réalisabilité

de la solution calculée (étape 2). Si la solution est réalisable, alors on s'arrête (étape 4). Sinon, γ est augmenté selon la procédure décrite à l'étape 3. Cette étape garantit en particulier que l'algorithme s'arrête en temps fini avec une solution réalisable de CCILP.

Concernant cette étape 3, quelques explications supplémentaires sont utiles. L'indice j est choisi de manière à augmenter efficacement la probabilité de la solution. Dans l'implémentation proposée, on décide de renforcer le terme de protection de la contrainte dont l'écart entre le membre de droite et l'espérance du membre de gauche est le plus faible. Ce choix est motivé par le fait que cet écart est impliqué dans la plupart des inégalités de probabilités (cf. par exemple les inégalités de Markov ou de Hoeffding). Mais d'autres critères pourraient s'avérer tout aussi pertinents. Si par exemple toutes les données A et b sont positives, en s'inspirant de l'inégalité de Markov, on pourrait prendre : $j = \arg \min_{k \in J} \{1 - (\sum_{i \in I} \mathbb{E}[A_{ki}]x_k)/b_k\} = \arg \max_{k \in J} \{(\sum_{i \in I} \mathbb{E}[A_{ki}]x_k)/b_k\}$. Dans les tests numériques que nous avons réalisés, on a observé que le type de critère utilisé n'a pas d'impact significatif sur les résultats obtenus.

L'algorithme rapide décrit ci-dessus peut s'avérer encore trop difficile en pratique, du fait de la résolution exacte de RILP, par exemple avec le modèle linéaire (5.7). Une idée naturelle est de remplacer le terme de protection $\Delta_j(x)$ par une constante : $\Delta'_j(x) = \gamma_j \cdot \sum_{i \in I_u(j)} \delta_{ji}$. Cette dernière démarche implique simplement une modification du membre de droite b . Afin de respecter l'esprit du terme de protection $\Delta_j(x)$, quand la protection constante devient "trop grande", il est judicieux de remplacer la contrainte j par le pire cas : $\bar{A}_j x \leq b_j$. Ceci évite de contraindre inutilement le problème (voir l'annexe C.1 pour un modèle robuste simpliste lié à cette remarque). Ces idées sont utilisées dans l'algorithme suivant :

Algorithme rapide alternatif pour CCILP

Étape 0 : Poser $\gamma = 0$. On note RP le problème suivant :

$$\begin{aligned} \max \quad & cx \\ \text{s.c.} \quad & \sum_{i \in I_u(j)} \underline{A}_{ji} x_i + \gamma_j \cdot \sum_{i \in I_u(j)} \delta_{ji} + \sum_{i \notin I_u(j)} A_{ji} x_i \leq b_j, \quad \forall j \in J \\ & x \in \{0, 1\}^n \end{aligned}$$

Étape 1 : Résoudre RP ; soit x la solution optimale correspondante.

Étape 2 : Calculer $\gamma' \geq \gamma$ comme décrit dans le lemme 30.

Étape 3 : Pour tout $j \in J$ tel que $\gamma'_j = 1$, remplacer la contrainte j de RP par :

$$\sum_{i \in I_u(j)} \bar{A}_{ji} x_i + \sum_{i \notin I_u(j)} A_{ji} x_i \leq b_j.$$

Si RP a été modifié, aller à l'étape 1.

Étape 4 : Calculer (ou borner) $\mathbb{P}(Ax \leq b)$.

Si $\mathbb{P}(Ax \leq b) \geq 1 - \varepsilon$: STOP.

Étape 5 : Poser $\gamma \leftarrow \gamma'$.

Choisir j tel que : $j = \arg \min_{k \in J: \gamma'_k < 1} \{b_k - \sum_{i \in I} \mathbb{E}[A_{ki}]x_i\}$.

Poser $\gamma_j \leftarrow \gamma_j + 1/\sum_{i \in I_u(j)} \delta_{ji}$ et mettre à jour le problème RP.

Aller à l'étape 1.

À l'étape 5, observer qu'il existe forcément $k \in J$ tel que $\gamma'_k < 1$. En effet, si $\gamma'_k = 1$ pour tout k , alors $\mathbb{P}(Ax \leq b) = 1$ et on sort à l'étape 4.

Dans cet algorithme alternatif, à chaque nouvelle valeur du paramètre de robustesse, seul un problème nominal ILP est résolu (étape 1). La simplicité de l'algorithme apparaît ainsi clairement. De plus, l'algorithme est très facile à mettre en œuvre, aucun paramétrage n'étant nécessaire. Ceci est un avantage substantiel par rapport aux méta-heuristiques, pour lesquelles un temps non-négligeable doit souvent être consacré au paramétrage correct de l'algorithme.

5.4.4 Bornes supérieures pour CCILP

Jusqu'ici, on a décrit des heuristiques permettant d'obtenir des solutions réalisables pour CCILP. Pour être capable d'évaluer la qualité de ces solutions, on a besoin d'une borne supérieure de la valeur optimale de CCILP. La résolution du problème nominal correspondant au scénario le plus favorable, \underline{A} , fournit déjà une borne évidente. On peut l'améliorer à l'aide du résultat suivant :

Proposition 17 *Pour tout $j \in J$, on introduit :*

- $p_j \geq 0$ tel que, pour toute solution optimale x^* de CCILP : $p_j \leq \sum_{i \in I_u(j)} x_i^*$;
- $I'_j \subseteq I_u(j)$ tel que $|I'_j| = p_j$ et : $\max\{\delta_{ji} \mid i \in I'_j\} \leq \min\{\delta_{ji} \mid i \in I_u(j) \setminus I'_j\}$.

Supposons que les $\{\eta_{ji}\}_{i \in I_u(j)}$ sont indépendantes et identiquement distribuées (i.i.d.). Considérons alors $\gamma \in [0, 1]^m$ tel que, pour tout $j \in J$:

$$\mathbb{P} \left(\sum_{i \in I'_j} \delta_{ji} \eta_{ji} \leq \gamma_j \sum_{i \in I_u(j)} \delta_{ji} \right) < 1 - \varepsilon.$$

Toute solution optimale x^ de CCILP satisfait, pour tout $j \in J$:*

$$\sum_{i \in I} \underline{A}_{ji} x_i^* + \gamma_j \sum_{i \in I_u(j)} \delta_{ji} \leq b_j.$$

Preuve : Soit x^* une solution optimale de CCILP. On montre par contradiction que x^* satisfait les inégalités ci-dessus. Supposons qu'il existe $j \in J$ tel que : $\sum_{i \in I} \underline{A}_{ji} x_i^* + \gamma_j \sum_{i \in I_u(j)} \delta_{ji} > b_j$. Alors :

$$\begin{aligned} \mathbb{P}(A_j x^* \leq b_j) &\leq \mathbb{P} \left(A_j x^* \leq \underline{A}_j x^* + \gamma_j \sum_{i \in I_u(j)} \delta_{ji} \right) \\ &= \mathbb{P} \left(\sum_{i \in I_u(j)} \delta_{ji} \eta_{ji} x_i^* \leq \gamma_j \sum_{i \in I_u(j)} \delta_{ji} \right) \\ &\leq \mathbb{P} \left(\sum_{i \in I'_j} \delta_{ji} \eta_{ji} \leq \gamma_j \sum_{i \in I_u(j)} \delta_{ji} \right). \end{aligned}$$

La dernière inégalité vient du fait que les $\{\eta_{ji}\}_{i \in I_u(j)}$ sont i.i.d., ainsi que de la définition de I'_j . Observer alors que : $\mathbb{P}(A_j x^* \leq b_j) \leq \mathbb{P}(A_j x^* \leq b_j) < 1 - \varepsilon$ (contradiction). \square

Lorsque p_j est connu, l'ensemble I'_j est constitué des indices $i \in I_u(j)$ correspondant aux éléments de plus faibles variations δ_{ji} . γ_j peut alors être trouvé par dichotomie. On a donc montré que RILP(γ) est une relaxation de CCILP si γ satisfait les conditions de la proposition ci-dessus. Noter que plus γ_j est élevé, plus la borne supérieure associée est intéressante.

Souvent, des paramètres $\{p_j\}_{j \in J}$ peuvent être calculés assez facilement. À titre d'illustration, on donne le résultat suivant (on rappelle que $\bar{A}_{ji}^+ = \max\{\bar{A}_{ji}, 0\}$).

Lemme 32 *Supposons que $c > 0$, et que pour tout $j \in J$: $I_u(j) = I_u \subseteq I$. On pose :*

$$p = \min_{j \in J} \left\{ \max \left\{ k \mid \max_{K \subseteq I_u, |K|=k} \sum_{i \in K} \bar{A}_{ji}^+ \leq b_j - \sum_{i \notin I_u} \bar{A}_{ji}^+ \right\} \right\}.$$

Alors toute solution optimale x^ de CCILP satisfait : $\sum_{i \in I_u} x_i^* \geq p$.*

Preuve : On raisonne par contraposition. Soit une solution x de CCILP, supposons que : $\sum_{i \in I_u} x_i < p$. On note $K = \{i \in I_u \mid x_i = 1\}$. Il existe forcément $l \in I_u$ tel que $x_l = 0$ (sinon $|K| = |I_u| \geq p$), et on note $K' = K \cup \{l\}$. Soit $j \in J$ quelconque. On a : $\sum_{i \in K'} \bar{A}_{ji} \leq \sum_{i \in K'} \bar{A}_{ji}^+$.

De plus, comme $|K'| \leq p : \sum_{i \in K'} \bar{A}_{ji}^+ \leq b_j - \sum_{i \notin I_u} \bar{A}_{ji}^+$. Donc : $\sum_{i \in K'} \bar{A}_{ji} \leq b_j - \sum_{i \notin I_u} \bar{A}_{ji}^+$. Soit x' égal à x , sauf pour la composante l : $x'_l = 1$. On a : $\bar{A}_j x' \leq \sum_{i \in K'} \bar{A}_{ji} + \sum_{i \notin I_u} \bar{A}_{ji}^+ \leq b_j$.

Ceci étant vrai pour tout $j \in J$, on en déduit : $\mathbb{P}(Ax' \leq b) = 1$. x' est donc une solution réalisable de CCILP. De plus, comme $c > 0$ et que x' comprend un élément de plus que x : $cx' > cx$. Donc x n'est pas une solution optimale de CCILP. \square

Dans le contexte du lemme 32, le calcul de p est très simple en pratique. Pour chaque ligne j , il s'agit de déterminer le nombre maximum d'éléments qu'on peut considérer, en prenant d'abord ceux dont les poids $\{\bar{A}_{ji}\}_{i \in I_u(j)}$ sont les plus grands. La complexité temporelle du calcul de p est donc celle d'un tri des poids $\{\bar{A}_{ji}\}_{i \in I_u(j)}$ pour chaque ligne $j \in J$, soit $O(mn \log(n))$.

Le lemme 32 peut également s'appliquer dans le cas où l'on suppose seulement que $c_i \neq 0$ pour tout $i \in I$, et que $I_u(j) = I_u$ pour tout $j \in J$. En effet, si $c_i < 0$ pour un $i \in I$ donné, alors il suffit de ré-écrire le problème avec une nouvelle variable $x'_i = 1 - x_i$ pour retrouver les hypothèses du lemme 32.

Si l'on travaille avec des contraintes probabilistes séparées (cf. (5.4)), la proposition 17 peut facilement être adaptée. Il suffit de considérer γ tel que, pour tout $j \in J$:

$$\mathbb{P}\left(\sum_{i \in I'_j} \delta_{ji} \eta_{ji} \leq \gamma_j \sum_{i \in I_u(j)} \delta_{ji}\right) < 1 - \varepsilon_j.$$

Dans ce dernier cas, et si seulement un coefficient par ligne est sujet à incertitude, la borne obtenue à partir de la proposition 17 est la valeur optimale de CCILP (pour peu bien sûr que γ soit choisi assez grand). Ce cas de figure correspond par exemple au cas où seul le membre de droite b est incertain.

La proposition 17 permet par ailleurs d'accélérer les algorithmes présentés dans la partie 5.4.3. En effet, au lieu de commencer avec $\gamma = 0$ à l'étape initiale, on peut considérer directement un paramètre γ satisfaisant des critères de la proposition 17. Ce processus préalable permet d'éviter des itérations inutiles dans les algorithmes. En pratique, la mise en œuvre de la proposition 17 peut se faire à l'aide du lemme 32.

Finalement, si l'on s'appuie sur des bornes analytiques pour évaluer les probabilités (inégalité de Hoeffding par exemple), le problème déterministe équivalent à CCILP qui en résulte peut fournir des bornes supérieures par le biais de techniques classiques, telles que la relaxation continue ou la relaxation lagrangienne par exemple.

Quelques résultats supplémentaires sur la caractérisation de bornes supérieures sont donnés dans l'annexe C.3.

5.4.5 Cas particulier du sac-à-dos

C'est une des applications les plus simples, puisqu'on a une seule contrainte :

$$\begin{aligned} \max \quad & \sum_{i \in I} p_i x_i \\ \text{s.c.} \quad & \sum_{i \in I} w_i x_i \leq c, \\ & x \in \{0, 1\}^n. \end{aligned}$$

$c > 0$ est la capacité du sac. Pour tout $i \in I$, $p_i > 0$ est le profit associé au fait de prendre l'élément i , et $w_i > 0$ est le poids de cet élément. Pour simplifier la présentation, on suppose que tous les poids sont incertains : $\forall i \in I, w_i \in [\underline{w}_i, \underline{w}_i + \delta_i]$, $\underline{w}_i \in \mathbb{N}$ et $\delta_i > 0$. Les résultats s'étendent immédiatement lorsque seule une partie des poids est sujette à incertitude.

Les algorithmes heuristiques de la partie 5.4.3 peuvent être adaptés pour ce problème particulier. RKP(γ) désigne la version robuste du problème de sac-à-dos (*Robust Knapsack Problem*)

avec poids incertains pour $\gamma \in [0, 1]$, et CCKP désigne sa version sous contrainte probabiliste. D'après le théorème 17, résoudre RKP(γ) ne requiert la résolution que de 2 problèmes de sac-à-dos classiques, à savoir :

$$(a) \max \{p.x \mid \bar{w}.x \leq c, x \in \{0, 1\}^n\}$$

$$\text{et : } (b) \max \{p.x \mid \underline{w}.x + \gamma \sum_{i \in I} \delta_i \leq c, x \in \{0, 1\}^n\}.$$

La valeur optimale de RKP(γ) est obtenue comme la meilleure solution parmi celle de (a) et celle de (b). Le problème (a) peut être résolu une fois pour toutes au début de l'algorithme : ainsi, à chaque nouvelle valeur de γ , seul le problème (b) doit être résolu. Ainsi, l'algorithme rapide de la partie 5.4.3 devient simplement ici :

Algorithme pour le sac-à-dos sous contrainte probabiliste (CCKA)

Étape 0 : Résoudre le problème (a), et soit \bar{x} la solution optimale trouvée. Poser $\gamma = 0$.

Étape 1 : Résoudre (b), soit x la solution optimale correspondante.

Étape 2 : Si $p.x \leq p.\bar{x}$: poser $x = \bar{x}$ et STOP.

Étape 3 : Calculer ou borner $\mathbb{P}(wx \leq c)$.

Étape 4 : Si $\mathbb{P}(wx \leq c) < 1 - \varepsilon$:
calculer $\gamma' \geq \gamma$ comme décrit dans le lemme 30 ;
poser $\gamma = \gamma' + 1 / \sum_{i \in I} \delta_i$ et aller à l'étape 1.

Étape 5 : Si $\mathbb{P}(wx \leq c) \geq 1 - \varepsilon$: STOP.

Noter que si l'on sort à l'étape 2, on a une solution x telle que $\mathbb{P}(wx \leq c) = 1$.

Certains résultats donnés dans la partie 4.3 peuvent être adaptés ici, montrant ainsi que l'algorithme ci-dessus fournit une solution optimale à CCKP dans certains cas particuliers. Soit $x \in \{0, 1\}^n$, on dit que x est décroissant si : $i < j \Rightarrow x_i \geq x_j$. Ceci signifie que x peut s'écrire de la manière suivante : $(1, 1, \dots, 1, 0, \dots, 0)$.

Proposition 18 *Supposons que les poids et les profits satisfont :*

$$i < j \Rightarrow \begin{cases} p_i > p_j \\ \underline{w}_i \leq \underline{w}_j \\ \delta_i \leq \delta_j \end{cases} . \quad (5.12)$$

De plus, supposons que les variables aléatoires $\{\eta_i\}_{i \in I}$ sont i.i.d.. Alors il existe $\gamma^ \in [0, 1]$ tel que toute solution optimale de RKP(γ^*) est optimale pour CCKP.*

Preuve : Montrons d'abord qu'il existe une solution optimale x^* de CCKP qui est décroissante. Soit x une solution optimale de CCKP qui n'est pas décroissante, on construit x^* décroissant tel que : $\|x\|_1 = \|x^*\|_1$ (rappel : $\|x\|_1 = \sum_{i \in I} |x_i|$). D'après les propriétés des valeurs de profits, on a : $p.x^* \geq p.x$. De plus : $\mathbb{P}(wx^* \leq c) = \mathbb{P}(\sum_{i \in I} (\underline{w}_i + \delta_i \eta_i) x_i^* \leq c)$. Puisque les poids minimaux sont triés en ordre croissant, on a : $\sum_{i \in I} \underline{w}_i x_i \geq \sum_{i \in I} \underline{w}_i x_i^*$. Donc : $\mathbb{P}(wx^* \leq c) \geq \mathbb{P}(\sum_{i \in I} \underline{w}_i x_i + \sum_{i \in I} \delta_i \eta_i x_i^* \leq c)$. De la même manière, d'après l'hypothèse sur les variations $\{\delta_i\}_{i \in I}$ et les variables aléatoires $\{\eta_i\}_{i \in I}$, on voit que : $\mathbb{P}(\sum_{i \in I} \delta_i \eta_i x_i^* \leq c - \sum_{i \in I} \underline{w}_i x_i) \geq \mathbb{P}(\sum_{i \in I} \delta_i \eta_i x_i \leq c - \sum_{i \in I} \underline{w}_i x_i)$. En définitive : $\mathbb{P}(wx^* \leq c) \geq \mathbb{P}(wx \leq c) \geq 1 - \varepsilon$. Ceci montre que x^* est une solution réalisable de CCKP de valeur objectif au moins aussi bonne que celui de x : x^* est une solution optimale de CCKP.

D'après la preuve du théorème 16, x^* peut être utilisé pour construire $I^* \subseteq I$ et $\gamma^* \in [0, 1]$ tels que toute solution optimale de RKP(I^*, γ^*) est aussi optimale pour CCKP. On sait aussi que x^* est une solution optimale de RKP(I^*, γ^*).

Soit x' une solution optimale de $\text{RKP}(\gamma^*)$. x' est optimal soit pour : $\max \{p.x \mid \bar{w}.x \leq c, x \in \{0, 1\}^n\}$, soit pour : $\max \{p.x \mid \underline{w}.x \leq c - \gamma^* \sum_{i \in I} \delta_i, x \in \{0, 1\}^n\}$. Dans chaque cas, d'après les hypothèses sur le tri des profits et de poids, x' optimal implique que x' est décroissant. Comme $\text{RKP}(\gamma^*)$ est une relaxation de $\text{RKP}(I^*, \gamma^*)$, on a : $px' \geq px^*$. Ceci implique notamment que : $\|x'\|_1 \geq \|x^*\|_1$, puisque x^* est décroissant et que les profits sont en ordre décroissant. Par ailleurs, si x' satisfait : $\bar{w}.x' \leq c$, alors $\mathbb{P}(wx' \leq c) = 1$, et x' est une solution réalisable, donc optimale, de CCKP. Si au contraire : $\bar{w}.x' > c$, alors on a forcément : $\underline{w}.x' + \gamma^* \sum_{i \in I} \delta_i \leq c$. Rappelons que γ^* satisfait : $\underline{w}.x^* + \gamma^* \sum_{i \in I} \delta_i = c$. On obtient donc : $\underline{w}.x' \leq \underline{w}.x^*$. Comme x^* est décroissant et que les poids minimaux $\{\underline{w}_i\}_{i \in I}$ sont triés en ordre croissant, on obtient : $\|x'\|_1 \leq \|x^*\|_1$. Ainsi, on a prouvé que : $\|x'\|_1 = \|x^*\|_1$. Comme x' et x^* sont tous deux décroissants, on a : $x' = x^*$. Donc x' est une solution optimale de CCKP.

On a donc montré que toute solution optimale de $\text{RKP}(\gamma^*)$ est optimale aussi pour CCKP. \square

Corollaire 5 *Supposons que les conditions (5.12) sont satisfaites et que les $\{\eta_i\}_{i \in I}$ sont i.i.d.. Soit $\gamma \in [0, 1]$, la solution optimale x^* de $\text{RKP}(\gamma)$ est une solution optimale de CCKP avec $\varepsilon = \mathbb{P}(wx^* \geq c)$.*

Preuve : Soit $\varepsilon = \mathbb{P}(wx^* \geq c)$. D'après la proposition 18, il existe γ' tel que toute solution optimale x' de $\text{RKP}(\gamma')$ est optimale pour CCKP. Par construction, x^* est réalisable pour CCKP. Puisque les vecteurs x' et x^* sont tous deux décroissants, on a : $x' \geq x^*$ (sinon, $px' < px^*$). Si $x' \neq x^*$, on aurait : $\mathbb{P}(wx' \leq c) < \mathbb{P}(wx^* \leq c) = 1 - \varepsilon$. Donc $x' = x^*$, et en conséquence x^* est optimal pour CCKP. \square

Ainsi, sous les conditions (5.12), l'algorithme CCKA fournit une solution optimale de CCKP. Ceci n'est pas d'un intérêt pratique majeur, puisque sous les conditions (5.12), CCKP se résout facilement en ajoutant successivement dans le sac les éléments de plus fort profit, en arrêtant quand la condition de probabilité est satisfaite. Néanmoins, les résultats ci-dessus montrent la pertinence de l'algorithme de résolution, au moins pour des instances satisfaisant "presque" les conditions (5.12). Comme déjà évoqué dans la partie 4.4, ces propriétés de l'algorithme CCKA rappellent celles de l'algorithme glouton pour le sac-à-dos classique (voir e.g. [52]) : on peut le considérer comme un équivalent stochastique de l'algorithme glouton.

5.5 Tests numériques

Le cadre d'étude proposé est pertinent pour une large gamme de problèmes, comme par exemple :

- le problème de sac-à-dos 0-1 avec poids et/ou profits incertains, ainsi que ses extensions classiques (sac-à-dos multi-dimensionnel et sac-à-dos multiple) ;
- le problème de plus courts chemins avec poids incertains sur les arêtes ;
- les problèmes de multiflots mono-routés avec demandes incertaines ;
- certains problèmes de localisation.

On donne ci-après les résultats de tests effectués sur des problèmes de sac-à-dos classique et multi-dimensionnel, avec poids incertains. Les programmes informatiques ont été écrit en C++ avec Cplex 10.0. Tous les tests ont été réalisés sur un ordinateur équipé d'un processeur Intel(R) Xeon(TM) 2,8 GHz avec 2 Go de RAM.

5.5.1 Description des instances et des méthodes de résolution

Deux types d'instances ont été générées. Dans la première famille d'instances, toutes les variations δ_{ji} sont égales à 10% du poids minimal de l'objet i (variations dites "proportionnelles" dans la suite). Dans le second type d'instances, les variations δ_{ji} sont choisies aléatoirement dans un intervalle. Ceci signifie que, dans ce cas, les variations ne sont pas corrélées à la valeur minimale des poids (variations dites "non-corrélées" par la suite).

Pour chaque type de problème, deux séries distinctes de tests ont été réalisées. Chaque série est caractérisée par la méthode utilisée pour évaluer les probabilités de réalisabilité : dans la première, on utilise des méthodes d'échantillonnage de type Monte-Carlo ; dans la seconde, on utilise la borne de Hoeffding (voir la partie 5.4.1).

Pour les problèmes de sac-à-dos multi-dimensionnels, les cas des contraintes de probabilités jointes et séparées ont été testés (cf. modèle (5.4)). Dans le cas des contraintes séparées, plusieurs méthodes de résolutions sont comparées :

- **Pire cas** : cette valeur est celle obtenue en considérant le pire scénario possible, quand tous les coefficients A_{ji} prennent leur plus grande valeur \bar{A}_{ji} .
- **M1** : Cette approche fait usage des seules idées décrites dans [13], où le modèle RILP_2 a été introduit (cf. modèle (5.8)). Les auteurs mettent en évidence le fait que, étant donné un paramètre Γ , la probabilité de réalisabilité d'une solution x de $\text{RILP}_2(\Gamma)$ peut être bornée par : $\forall j \in J, \mathbb{P}(A_j x \leq b_j) \geq \mathbb{P}\left(\sum_{i \in I_u(j)} \eta_{ji} \leq \Gamma_j\right)$. Ainsi, pour tout $j \in J$, on cherche le plus petit Γ_j satisfaisant : $\mathbb{P}\left(\sum_{i \in I_u(j)} \eta_{ji} \leq \Gamma_j\right) \geq 1 - \varepsilon_j$. On résout ensuite le problème $\text{RILP}_2(\Gamma)$. Observer qu'avec cette méthode, seul un problème robuste a besoin d'être résolu.
- **M2** : Une deuxième méthode de résolution, notée M2, est une adaptation directe de l'algorithme proposé pour le sac-à-dos sous contrainte de probabilité au chapitre 4 (§ 4.4). En partant de $\Gamma = 0$, une suite croissante de vecteurs de robustesse Γ est générée. À chaque nouveau Γ , le problème correspondant $\text{RILP}_2(\Gamma)$ est résolu. Le choix du prochain vecteur Γ s'appuie sur des idées très semblables à celles détaillées dans la partie 5.4.3 pour l'algorithme rapide : à partir du Γ actuel, on construit un nouveau vecteur $\Gamma' \geq \Gamma$ tel que $\text{RILP}_2(\Gamma')$ et $\text{RILP}_2(\Gamma)$ ont la même valeur optimale. Puis la plus petite des composantes $\{\Gamma_j\}_{j \in J}$ est augmentée de $1/\max_{i \in I_u(j)} \delta_{ji}$. (Pour plus de détails techniques, on renvoie à la partie 4.4.)
- **AFA** : Cette approche correspond à l'algorithme rapide alternatif de la partie 5.4.3. Dans le cas du sac-à-dos, cette approche s'appuie sur l'algorithme CCKA décrit dans la partie 5.4.5.
- **Optimum** : Quand cela est possible, on a aussi calculé la solution optimale du problème sous contraintes probabilistes. Pour les modèles utilisant l'inégalité de Hoeffding, dans le cas de probabilités séparées, la solution optimale a été obtenue grâce au programme linéaire en nombres entiers (5.11) (présenté dans la partie 5.4.1) ; cette approche sera nommée HA. Dans les autres cas, quand on s'appuie sur des méthodes d'échantillonnage ou quand on considère des contraintes de probabilités jointes, l'optimum est obtenu par énumération de toutes les solutions. Évidemment, ceci n'est possible que pour des instances de petites tailles.
- **Borne sup.** : Quand on a recours aux méthodes d'échantillonnage pour évaluer les probabilités, une borne supérieure à la valeur optimale de CCILP est calculée selon la proposition 17. Puisqu'on teste ici des problèmes de sac-à-dos multi-dimensionnels, chaque valeur p_j de la proposition est calculée en remplissant progressivement le sac j avec les plus forts poids \bar{A}_{ji} d'abord (cf. lemme 32).
- **Meilleur cas** : Enfin, cette valeur correspond à la valeur de la solution optimale pour le

scénario le plus optimiste, quand tous les coefficients A_{ji} prennent leur valeur minimale \underline{A}_{ji} .

Les approches “pire cas”, “M1”, “M2”, “AFA” et “optimum” fournissent toutes des solutions réalisables au problème sous contraintes probabilistes (bornes inférieures). Au contraire, “borne sup.” et “meilleur cas” donnent des bornes supérieures de la valeur optimale.

Quand on travaille avec des contraintes de probabilités jointes, la méthode M1 n’est pas utilisable. Par ailleurs, comme l’approximation de Hoeffding (cf. partie 5.4.1) n’est pas linéarisable dans le cas des contraintes jointes, elle n’a pas été implémentée ici.

Les tableaux de résultats comportent un certain nombre de colonnes dont le contenu est maintenant explicité. La colonne “valeur” donne la valeur de la solution finale renvoyée par l’algorithme, et “amélioration” mesure l’écart (en %) par rapport au pire cas, qui sert ici de référence. Pour le problème CCKP, on reporte dans la colonne “proba.” la probabilité de réalisabilité estimée pour la solution finale obtenue. Quand on s’appuie sur l’inégalité de Hoeffding, la colonne “temps (sec.) ou écart (%)” donne la durée totale d’exécution de l’algorithme, ou, pour l’algorithme HA, s’il n’a pas convergé dans la limite de temps imposée (30 minutes), l’écart prouvé à l’optimum. Lorsqu’on utilise la méthode d’échantillonnage, on indique le temps total d’exécution de l’algorithme (“TT”), le temps d’évaluation des probabilités (“TE”) et le temps de résolution des PLNE (“TC”).

5.5.2 Le problème de sac-à-dos

On s’intéresse ici au problème de sac-à-dos :

$$\begin{aligned} \max \quad & \sum_{i \in I} p_i x_i \\ \text{s.c.} \quad & \sum_{i \in I} w_i x_i \leq c, \\ & x \in \{0, 1\}^n. \end{aligned}$$

$c > 0$ est la capacité du sac. Pour tout $i \in I$, $p_i > 0$ est le profit associé au fait de prendre l’élément i , et $w_i > 0$ est le poids de cet élément. Tous les poids sont supposés incertains : $\forall i \in I, w_i \in [\underline{w}_i, \underline{w}_i + \delta_i]$, $\underline{w}_i \in \mathbb{N}$ et $\delta_i > 0$.

Pour construire les instances de test, les poids minimaux $\{\underline{w}_i\}_{i \in I}$ et les profits $\{p_i\}_{i \in I}$ sont des entiers choisis aléatoirement dans $[100, 1000]$. La capacité c est un entier choisi aléatoirement entre $\sum_{i \in I} \underline{w}_i / 3$ et $2 \sum_{i \in I} \underline{w}_i / 3$. Dans le cas des “variations proportionnelles”, on pose : $\forall i \in I, \delta_i = 0, 1 \underline{w}_i$ (δ_i est éventuellement arrondi pour être entier). Dans le cas des “variations non-corrélées”, les $\{\delta_i\}_{i \in I}$ sont choisis aléatoirement dans $[10, 100]$.

Quand on utilise des méthodes d’échantillonnage pour évaluer les probabilités de réalisabilité, les variables aléatoires $\{w_i\}_{i \in I}$ sont supposées indépendantes et uniformément distribuées sur les intervalles de définition. On s’appuie sur le lemme 29 pour contrôler la probabilité de se tromper dans l’estimation de la réalisabilité de la solution. À chaque simulation de Monte-Carlo, on s’arrête si cette probabilité d’erreur est inférieure à 10^{-4} . Mais il peut arriver que l’obtention de cette garantie nécessite un temps de simulation très long (c’est le cas quand la quantité d du lemme 29 est très petite). C’est pour cela qu’on borne le temps de chaque évaluation à 5 secondes maximum. On retient la probabilité d’erreur sur l’estimation la borne alors disponible (cf. lemme 29), sauf si celle-ci est supérieure à 0,5. Dans ce dernier cas, comme l’estimation est vraiment trop litigieuse, on supposera que la solution correspondante n’est pas réalisable.

Ces tests numériques ont l’intérêt de montrer que les simulations de Monte-Carlo peuvent en pratique être utilisées. Elles ont l’avantage de permettre la prise en compte d’hypothèses probabilistes aussi diverses que voulues, au niveau des distributions, des corrélations, etc. Malgré

la simplicité des hypothèses considérées ici (variables aléatoires uniformément distribuées et indépendantes), les tests réalisés ont valeur d'illustration quant à la maniabilité de ce genre d'approches.

Quand on utilise l'inégalité de Hoeffding (cf. (5.9)), les variables aléatoires $\{w_i\}_{i \in I}$ sont supposées indépendantes et centrées : $\forall i \in I, \mathbb{E}[w_i] = \underline{w}_i + \delta_i/2$. Dans tous les cas, le but est de trouver un remplissage du sac réalisable avec une probabilité d'au moins 0,9, ce qui signifie que $\varepsilon = 0, 1$.

Résultats avec méthode d'échantillonnage : Des instances avec $n = 25$, $n = 100$ et $n = 200$ ont été considérées ; pour chaque taille, 50 instances ont été testées. La solution optimale, obtenue par énumération des solutions possibles, n'est donnée que pour $n = 25$. On peut tirer du tableau 5.1 de grandes tendances. On note d'abord que M1 ne conduit qu'à une amélioration très faible par rapport au pire cas. Bien que très rapide, cet algorithme fournit des solutions bien moins intéressantes que M2 et CCKA. Ces deux derniers algorithmes fournissent des solutions de qualité comparable. CCKA est sensiblement plus rapide que M2. Ce gain en temps de calcul global s'accroît avec la taille des instances, atteignant 25% en moyenne pour $n = 200$. Cette différence est d'autant plus importante si l'on regarde seulement la colonne TC, qui correspond au temps de résolution des PLNE. En effet, l'essentiel du temps est consacré à l'estimation de la probabilité (colonne TE). Finalement, on constate que les solutions obtenues grâce à CCKA ou M2 sont très proches de l'optimum dans le cas $n = 25$ (voir le tableau 5.3).

Le tableau 5.3 donne le nombre d'instances où CCKA est plus ou moins efficace que M2. Il apparaît nettement que CCKA est plus souvent meilleur que M2 dans le cas des variations proportionnelles ; c'est le contraire dans le cas des variations non-corrélées. On peut expliquer cette observation par l'analyse des termes de protection relatifs à chaque méthode. Pour RILP, $\Delta(x) = \min \{ \gamma \cdot \sum_{i \in I} \delta_i, \sum_{i \in I} \delta_i x_i \}$; pour RILP₂, $\Delta(x)$ est remplacé par $\Delta_2(x) = \max \{ \sum_{i \in S} \delta_i x_i : S \subseteq I, |S| = \Gamma \}$ (lorsque Γ est entier). On voit que $\Delta_2(x)$ pénalisera toujours les solutions impliquant des coefficients i dont les variations δ_i sont grandes. Ainsi, par exemple, si on a deux éléments, i et j , tels que $p_i = p_j$, $\underline{w}_i = \underline{w}_j$ et $\delta_i < \delta_j$, RILP₂ "préfèrera" l'élément i à l'élément j . À l'inverse, RILP ne fera pas de différence entre eux. En effet, le terme de protection Δ ne prend en compte les variations qu'au travers de leur somme globale $\sum_{i \in I} \delta_i$, puisque la plupart du temps : $\Delta(x) = \gamma \cdot \sum_{i \in I} \delta_i$. Ces réflexions peuvent expliquer la meilleure capacité de RILP₂ à sélectionner de bonnes solutions quand les variations ne sont pas corrélées aux valeurs minimales des poids.

Finalement, remarquer que la borne supérieure calculée à partir de la proposition 17 améliore la borne fournie par le meilleur cas. Cependant, elle reste de qualité médiocre, comme on s'en rend compte avec les tests pour $n = 25$. L'obtention de meilleures bornes sur la valeur de CCKP est nécessaire, et semble devoir requérir des algorithmes plus sophistiqués.

Résultats avec l'inégalité de Hoeffding : Comme précédemment, cinquante instances ont été testées pour chaque taille $n = 50$, $n = 100$ et $n = 200$. D'après le tableau 5.4, on voit que les principales observations détaillées ci-dessus restent valides : M1 produit des solutions de qualité très médiocre ; M2 et CCKA produisent des solutions comparables, mais CCKA est beaucoup plus rapide que M2. Ici, contrairement au cas des méthodes d'échantillonnage, le calcul des probabilités est immédiat : la diminution du temps de résolution des PLNE se répercute donc directement sur le temps de calcul global. Ainsi, l'algorithme CCKA permet un gain de l'ordre de 85% en temps de calcul par rapport à M2, en moyenne pour $n = 200$.

Pour $n = 50$, une comparaison avec la solution optimale est donnée, grâce au modèle linéaire en nombres entiers (5.11) (HA). Le tableau 5.6 montre que l'optimum est trouvé par les algorithmes heuristiques dans une très large majorité des cas. Pour $n = 100$, HA a été résolu avec

TAB. 5.1 – CCKP : résultats moyens sur 50 instances, quand on utilise la méthode d'échantillonnage.

			valeur	amélio- ration	proba.	erreur (.10 ⁻²)	TE (sec.)	TC (sec.)	TT (sec.)
variations proportionnelles	n = 25	pire cas	9562,28	0	1	-	-	0,01	0,01
		M1	9567,84	0,06%	1,00	0,01	0,17	0,01	0,18
		M2	9756,14	2,03%	0,97	0,03	4,05	0,07	4,12
		CCKA	9756,14	2,03%	0,97	0,03	4,40	0,04	4,44
		optimum	9756,14	2,03%	0,97	0,02	103,57	11,13	114,70
		borne sup.	10011,54	4,70%	-	2,35	23,01	0,00	23,01
		meilleur cas	10082,06	5,44%	-	-	-	0,01	0,01
	n=100	pire cas	38904,14	0	1	-	-	0,05	0,05
		M1	38962,96	0,15%	1,00	0,01	2,43	0,06	2,49
		M2	39774,8	2,24%	0,936	0,0019	22,91	3,75	26,66
		CCKA	39776,24	2,24%	0,93	1,27	23,21	0,79	24
		borne sup.	40640,38	4,46%	-	4,81	33,05	0,01	33,06
		meilleur cas	40935,88	5,22%	-	-	-	0,01	0,01
	n = 200	pire cas	79590,46	0,00%	1	-	-	0,18	0,18
		M1	79739,42	0,19%	1,00	0,01	3,4	0,26	3,67
		M2	81445,96	2,33%	0,92	3,9	71,62	36,01	107,63
		CCKA	81449,44	2,34%	0,92	6,1	76,86	5,48	82,34
		borne sup.	83140,62	4,46%	-	12,6	41,48	0,02	41,50
meilleur cas		83740,24	5,21%	-	-	-	0,02	0,02	
variations non-corrélées	n = 25	pire cas	9676,98	0	1	-	-	0,01	0,01
		M1	9692,28	0,16%	1	0,01	0,19	0,01	0,2
		M2	9912,6	2,43%	0,97	0,02	5,36	0,09	5,45
		CCKA	9914,36	2,45%	0,97	0,04	5,28	0,04	5,32
		optimum	9915,44	2,46%	0,96	0,03	104,59	10,99	115,58
		borne sup.	10213,16	5,54%	-	3,2	24,27	0,00	24,27
		meilleur cas	10307,42	6,51%	-	-	-	0,00	0,00
	n = 100	pire cas	38656,12	0	1	-	-	0,06	0,06
		M1	38721,16	0,17%	1,00	0,01	2,14	0,07	2,21
		M2	39742,18	2,81%	0,92	1,32	30,00	4,86	34,86
		CCKA	39739,12	2,80%	0,92	1,08	28,48	1,12	29,6
		borne sup.	40856,64	5,69%	-	7,05	33,87	0,00	33,87
		meilleur cas	41172,88	6,51%	-	-	-	0,01	0,01
	n = 200	pire cas	75902,72	0	1	-	-	0,15	0,15
		M1	75998,22	0,13%	1,00	0,01	2,92	0,08	3,00
		M2	78196,46	3,02%	0,92	4,2	88,96	39,79	128,75
		CCKA	78189,46	3,01%	0,92	3,7	88,45	6,85	95,3
		borne sup.	80470,04	6,02%	-	0,1289	40,37	0,02	40,39
meilleur cas		81065,22	6,80%	-	-	-	0,02	0,02	

TAB. 5.2 – CCKP : nombre d’instances où CCKA donne une solution meilleure ou moins bonne que M2, quand on utilise la méthode d’échantillonnage.

		CCKA > M2	CCKA = M2	CCKA < M2
variations proportionnelles	$n = 25$	0%	100%	0%
	$n = 100$	10%	90%	0%
	$n = 200$	34%	62%	4%
variations non-corrélées	$n = 25$	2%	96%	2%
	$n = 100$	2%	78%	20%
	$n = 200$	4%	54%	42%

TAB. 5.3 – CCKP : nombre d’instances pour lesquelles CCKA ou M2 fournit une solution optimale, quand on utilise la méthode d’échantillonnage.

	CCKA=optimum	M2=optimum
variations proportionnelles, $n = 25$	100%	100%
variations non-corrélées, $n = 25$	96%	94%

Cplex avec une limite de temps de 30 minutes ; comme la très large majorité des instances n’a pas été résolue à l’optimum, l’écart moyen prouvé avec l’optimum est indiqué dans la dernière colonne du tableau 5.4. Il apparaît que les solutions trouvées par ce dernier procédé sont de moins bonne qualité que celles produites par M2 et CCKA en seulement quelques secondes. Comme HA était déjà très difficile pour $n = 100$, ce modèle n’a pas été testé pour $n = 200$.

Ainsi, les solutions obtenues grâce à M2 et CCKA sont de très bonne qualité. Une comparaison plus fine des deux méthodes peut être faite, notamment à partir du tableau 5.5. Quand on travaille avec des variations proportionnelles, les solutions de CCKA sont plus souvent meilleures que celles de M2. C’est le contraire qui se produit quand on travaille avec des variations non-corrélées. L’explication de cette observation est la même que celle déjà donnée pour le cas des méthodes d’échantillonnage.

Remarque : Noter que la probabilité de réalisabilité indiquée dans les tableaux pour la méthode M1 est toujours “1,00”. En fait, cette probabilité n’est pas exactement égale à 1, mais elle en est trop proche par rapport à la précision numérique des résultats donnés.

5.5.3 Problèmes de sac-à-dos multi-dimensionnels

On s’intéresse à la généralisation suivante du problème de sac-à-dos, dite sac-à-dos multi-dimensionnel (MKP) :

$$\begin{aligned} \max \quad & \sum_{i \in I} p_i x_i \\ \text{s.c.} \quad & \sum_{i \in I} w_{ji} x_i \leq c_j, \quad \forall j \in J, \\ & x \in \{0, 1\}^n. \end{aligned}$$

On présente des résultats seulement pour $m = 5$ (5 dimensions). Toutes les données c , w_{ji} et δ_{ji} sont des entiers strictement positifs. Comme précédemment, tous les poids sont supposés incertains, avec des variations associées qui sont ou bien proportionnelles, ou bien non-corrélées. Les données sont générées aléatoirement selon les mêmes règles que celles décrites plus haut pour les instances de sac-à-dos. On conserve par ailleurs les mêmes hypothèses probabilistes : avec les méthodes d’échantillonnage, les variables aléatoires $\{w_{ji}\}_{j \in J, i \in I}$ sont supposées indépendantes

TAB. 5.4 – CCKP : Résultats moyens sur 50 instances, quand on utilise l'inégalité de Hoeffding pour évaluer les probabilités.

			valeur	amélioration	proba.	temps (sec.) ou écart (%)
variations proportionnelles	$n = 50$	pire cas	19990,13	-	1	0,01 sec.
		M1	19991,64	0,01%	1,00	0,02 sec.
		M2	20268,33	1,39%	0,94	0,41 sec.
		CCKA	20268,32	1,39%	0,94	0,12 sec.
		HA	20268,54	1,39%	0,94	50,03 sec.
	$n = 100$	pire cas	38939,2	-	1	0,05 sec.
		M1	38954,94	0,04%	1,00	0,05 sec.
		M2	39614,8	1,74%	0,92	4,15 sec.
		CCKA	39617,11	1,74%	0,92	0,79 sec.
		HA	39611,01	1,73%	0,92	0,66%
	$n = 200$	pire cas	79704,63	-	1	0,16 sec.
		M1	79768,07	0,08%	1,00	0,21 sec.
M2		81262,11	1,95%	0,91	37,66 sec.	
CCKA		81264,83	1,96%	0,91	4,92 sec.	
variations non-corrélées	$n = 50$	pire cas	19428,92	-	1	0,02 sec.
		M1	19432,77	0,02%	1,00	0,02 sec.
		M2	19766,52	1,74%	0,94	0,57 sec.
		CCKA	19764,66	1,73%	0,94	0,18 sec.
		HA	19767,16	1,74%	0,94	218,25 sec.
	$n = 100$	pire cas	38537,29	-	1	0,07 sec.
		M1	38552,72	0,04%	1,00	0,07 sec.
		M2	39357,63	2,13%	0,92	5,63 sec.
		CCKA	39352,04	2,11%	0,92	1,13 sec.
		HA	39351,06	2,11%	0,92	0,94%
	$n = 200$	pire cas	77473,05	-	1	0,16 sec.
		M1	77518,12	0,06%	1,00	0,23 sec.
M2		79394,54	2,48%	0,91	40,15 sec.	
CCKA		79378,69	2,46%	0,91	5,97 sec.	

TAB. 5.5 – CCKP : nombre d'instances où CCKA donne de meilleurs résultats que M2, en utilisant l'inégalité de Hoeffding.

		CCKA > M2	CCKA = M2	CCKA < M2
variations proportionnelles	$n = 50$	1%	98%	1%
	$n = 100$	14%	84%	2%
	$n = 200$	26%	72%	2%
variations non-corrélées	$n = 50$	0%	90%	10%
	$n = 100$	2%	76%	22%
	$n = 200$	6%	37%	57%

TAB. 5.6 – CCKP : nombre d’instances où CCKA ou M2 fournit une solution optimale, en utilisant l’inégalité de Hoeffding.

	CCKA=optimum	M2=optimum
variations proportionnelles, $n = 50$	99%	99%
variations non-corrélées, $n = 50$	86%	95%

et uniformément distribuées sur les intervalles de distribution ; avec l’inégalité de Hoeffding, les variables aléatoires sont indépendantes et centrées.

Deux problèmes sous contraintes probabilistes sont considérés. Le premier, noté CCMKP₁, est à contraintes de probabilités séparées :

$$\begin{aligned} & \max \\ \text{s.c.} \quad & \mathbb{P} \left(\sum_{i \in I} w_{ji} x_i \leq c_j \right) \geq 1 - \varepsilon_j, \quad \forall j \in J, \\ & x \in \{0, 1\}^n. \end{aligned}$$

Le deuxième, noté CCMKP₂, contient une contrainte de probabilité jointe :

$$\begin{aligned} & \max \\ \text{s.c.} \quad & \mathbb{P} \left(\forall j \in J : \sum_{i \in I} w_{ji} x_i \leq c_j \right) \geq 1 - \varepsilon, \\ & x \in \{0, 1\}^n. \end{aligned}$$

On résout ces problèmes pour $\varepsilon_j = 0, 1$, pour tout $j \in J$, quand on considère CCMKP₁, ou bien $\varepsilon = 0, 1$ pour CCMKP₂.

Pour CCMKP₁, comme pour le sac-à-dos uni-dimensionnel, l’approximation de Hoeffding (HA) correspondante est écrite sous forme de programme linéaire en nombres entiers (cf. modèle (5.11)). Les résultats obtenus par nos heuristiques sont comparés à ceux produits par HA. Pour CCMKP₂, on rappelle qu’il n’y a pas de modèle linéaire en nombres entiers équivalent. Il faudrait résoudre un problème non-linéaire en nombres entiers, ce qui est en général très difficile : ceci dépasse le cadre de ce travail. Par ailleurs, se souvenir que pour CCMKP₂, la méthode M1 ne peut pas être utilisée.

Des tests ont été faits pour des tailles entre $n = 15$ et $n = 100$, avec cinquante instances pour chaque taille. Les résultats moyens sont présentés dans les tableaux 5.7 à 5.10. Comme pour le problème de sac-à-dos, M1 n’apporte quasiment pas d’amélioration par rapport au pire cas, tandis que M2 et AFA donnent des solutions de qualités très proches. Pourtant, il apparaît que les solutions réalisables obtenues, tout comme les bornes supérieures calculées, restent assez éloignées de la valeur optimale, bien qu’elles améliorent significativement le pire cas ou le meilleur cas. Noter que pour la résolution de HA, un temps limite de 30 minutes a été imposé. Ainsi, quand les instances ne sont pas résolues à l’optimum, l’écart prouvé à l’optimum est indiqué (en %).

En ce qui concerne les temps de calcul, AFA se révèle bien plus rapide pour la résolution des PLNE (cf. colonne TC pour la méthode d’échantillonnage, et le temps total pour l’inégalité de Hoeffding). Cependant, lorsqu’on s’appuie sur la méthode d’échantillonnage, le temps total de calcul est souvent plus important pour AFA que pour M2. Les temps d’évaluation de la probabilité augmentent lorsqu’on traite de solutions dont la probabilité de réalisabilité est proche de $1 - \varepsilon$; AFA semble tomber plus souvent sur de telles solutions.

On donne les performances comparées de AFA et de M2, en terme de solution trouvée, dans les tableaux 5.11 et 5.12. Contrairement au cas du sac-à-dos simple, il est difficile de dégager une tendance pour dire lequel des deux algorithmes parvient le mieux à résoudre CCMKP₁ et CCMKP₂. Compte tenu des performances en temps de calcul, il semble donc judicieux d’encourager le recours à AFA.

TAB. 5.7 – CCMKP₁ : résultats moyens sur 50 instances, quand on utilise la méthode d'échantillonnage.

			valeur	amélioration	erreur (.10 ⁻²)	TE (sec.)	TC (sec.)	TT (sec.)
variations proportionnelles	n = 15	pire cas	3936,4	0	-	-	0,02	0,02
		M1	3936,4	0,00%	0	0,00	0,03	0,03
		M2	4003,6	1,71%	0,01	0,94	0,08	1,02
		AFA	4003,6	1,71%	0,01	1,48	0,03	1,51
		optimum	4048,9	2,86%	0,01	651,55	0,06	651,62
		borne sup.	4193,7	6,54%	0,21	97,31	0,01	97,32
		meilleur cas	4234,7	7,58%	-	-	0,01	0,01
	n = 50	pire cas	16357	0	-	-	0,14	0,14
		M1	16357	0,00%	0,00	0,00	0,15	0,15
		M2	16679,32	1,97%	0,04	14,95	2,46	17,41
		AFA	16681,18	1,98%	0,04	25,76	0,62	26,38
		borne sup.	17299	5,76%	4,19	147,51	0,02	147,53
		meilleur cas	17440,48	6,62%	-	-	0,03	0,03
	n = 100	pire cas	32983,52	0	-	-	0,83	0,83
		M1	32984,78	0,00%	0,00	0,00	1,01	1,01
		M2	33627,22	1,95%	0,98	60,23	35,31	95,54
		AFA	33605,2	1,88%	2,11	112,01	5,97	117,98
		borne sup.	34862,12	5,70%	9,58	181,70	0,07	181,77
meilleur cas		35137,84	6,53%	-	-	0,06	0,06	
variations non-corrélées	n = 15	pire cas	4426,7	0	-	-	0,02	0,02
		M1	4426,7	0,00%	0,00	0	0,03	0,03
		M2	4501	1,68%	0,04	2,83	0,08	2,91
		AFA	4498,6	1,62%	0,04	4,14	0,03	4,17
		optimum	4544,6	2,66%	0,02	353,41	0,07	353,48
		borne sup.	4711,2	6,43%	0,62	100,03	0,01	100,04
		meilleur cas	4820,4	8,89%	-	-	0,00	0,00
	n = 50	pire cas	15951,44	0	-	-	0,14	0,14
		M1	15951,44	0,00%	0,00	0,00	0,13	0,13
		M2	16330,44	2,38%	0,75	16,07	2,52	18,59
		AFA	16325,26	2,34%	0,45	25,26	0,68	25,94
		borne sup.	17039,58	6,82%	4,97	149,11	0,02	149,13
		meilleur cas	17195,88	7,80%	-	-	0,02	0,02
	n = 100	pire cas	31576,94	0	-	-	0,75	0,75
		M1	31576,94	0,00%	0,00	0,00	0,54	0,54
		M2	32347,52	2,44%	0,94	54,61	41,41	96,03
		AFA	32359,46	2,48%	0,52	96,06	7,53	103,59
		borne sup.	33751,52	6,89%	8,10	178,00	0,08	178,07
meilleur cas		34026,3	7,76%	-	-	0,07	0,07	

TAB. 5.8 – CCMKP₁ : résultats moyens sur 50 instances, obtenus avec l'inégalité de Hoeffding.

			valeur	amélioration	temps (sec.) ou écart (%)
variations proportionnelles	$n = 50$	pire cas	16451,21	0	0,15
		M1	16451,21	0,00%	0,16
		M2	16640,05	1,15%	2,33
		AFA	16638,28	1,14%	0,62
		HA	16717,26	1,62%	117,3
	$n = 100$	pire cas	16336,75	0	0,28
		M1	16336,75	0,00%	0,30
		M2	16563,59	1,39%	16,78
		AFA	16566,81	1,41%	2,79
		HA	16649,65	1,92%	0,56%
variations non-corrélées	$n = 50$	pire cas	16160,2	0	0,14
		M1	16160,2	0,00%	0,15
		M2	16362,68	1,25%	2,80
		AFA	16356,13	1,21%	0,83
		HA	16454,76	1,82%	291,07
	$n = 100$	pire cas	16456,5	0	0,37
		M1	16456,5	0,00%	0,37
		M2	16739,52	1,72%	20,84
		AFA	16739,7	1,72%	3,48
		HA	16830,41	2,27%	0,69%

5.5.4 Synthèse

On peut tirer trois principaux enseignements des tests numériques précédents. Le premier concerne la souplesse et la facilité de mise en œuvre des approches itératives développées dans ce chapitre et le précédent. Elles permettent en pratique d'utiliser des méthodes d'évaluation de probabilité variées, comme par exemple les simulations de Monte-Carlo. Ceci signifie la très grande richesse des hypothèses probabilistes qui peuvent concrètement être prises en compte (distributions particulières, corrélations, etc.).

Le deuxième enseignement concerne la pertinence du modèle robuste RILP introduit dans ce chapitre, et exploité dans l'algorithme dit "AFA". Le recours à cette approche de la robustesse peut conduire à résoudre simplement des problèmes de complexité équivalente au problème nominal sans incertitudes. Les résultats obtenus sont souvent de qualité comparable à ceux correspondant à l'utilisation du modèle de Bertsimas et Sim, tout en consommant beaucoup moins de temps dans la résolution de PLNE.

Enfin, les limites de ces approches heuristiques sont également mises en évidence. Les solutions produites peuvent rester assez éloignées de l'optimum. L'évaluation de l'écart à la valeur optimale reste difficile.

5.6 Conclusion

Un nouveau modèle robuste a été proposé. Sa complexité pratique et théorique reste souvent acceptable. Des liens théoriques ont été mis en évidence entre cette modélisation robuste et la programmation sous contraintes probabilistes avec variables 0-1. Puis des algorithmes

TAB. 5.9 – CCMKP₂ : résultats moyens sur 50 instances, quand on utilise la méthode d'échantillonnage.

			valeur	amélio- ration	erreur (.10 ⁻²)	TE (sec.)	TC (sec.)	TT (sec.)	
variations proportionnelles	<i>n</i> = 15	pire cas	4121,9	0	-	-	0,02	0,02	
		M2	4164	1,02%	0,02	3,95	0,11	4,06	
		AFA	4164	1,02%	0,02	4,72	0,04	4,76	
		optimum	4174,8	1,28%	0,00	54,09	0,29	54,38	
		borne sup.	4353,5	5,62%	0,17	96,53	0,01	96,54	
		meilleur cas	4396,4	6,66%	-	-	0,01	0,01	
	<i>n</i> = 50	pire cas	16147,56	0	-	-	0,15	0,15	
		M2	16451,7	1,88%	0,47	10,12	2,43	12,55	
		AFA	16455,3	1,91%	0,54	14,05	0,62	14,67	
		borne sup.	17075,54	5,75%	5,2	149,92	0,02	149,94	
		meilleur cas	17217,4	6,63%	-	-	0,02	0,02	
	<i>n</i> = 100	pire cas	33426	0	-	-	0,61	0,61	
		M2	34063,3	1,91%	1,08	19,143	30,59	49,74	
		AFA	34056,76	1,89%	1,82	32,691	4,97	37,66	
		borne sup.	35289,32	5,57%	9,25	181,224	0,07	181,29	
		meilleur cas	35560,94	6,39%	-	0,00	0,07	0,07	
	variations non-corrélées	<i>n</i> = 15	pire cas	4510,9	0	-	-	0,01	0,01
			M2	4534,1	0,51%	0,01	3,72	0,08	3,79
AFA			4534,1	0,51%	0,01	5,07	0,03	5,1	
optimum			4542,7	0,70%	0,01	56,51	0,22	56,72	
borne sup.			4778,8	5,94%	1,95	99,48	0,00	99,49	
meilleur cas			4864,3	7,83%	-	-	0,01	0,01	
<i>n</i> = 50		pire cas	15852,78	0	-	-	0,13	0,13	
		M2	16222,34	2,33%	0,83	10,48	2,94	13,43	
		AFA	16231,66	2,39%	0,68	15,78	0,73	16,5	
		borne sup.	16975,22	7,08%	2,55	149,56	0,02	149,59	
		meilleur cas	17115,42	7,96%	-	-	0,03	0,03	
<i>n</i> = 100		pire cas	32827,8	0	-	-	0,81	0,81	
		M2	33613,62	2,39%	0,96	23,29	38,59	61,87	
		AFA	33618,36	2,41%	1,04	36,54	7	43,54	
		borne sup.	35019,84	6,68%	8,93	179,05	0,07	179,12	
		meilleur cas	35298,84	7,53%	-	-	0,1	0,1	

TAB. 5.10 – CCMKP₂ : résultats moyens sur 50 instances, en utilisant l'inégalité de Hoeffding.

			valeur	amélioration	proba.	temps (sec.)
variations proportionnelles	$n = 20$	pire cas	6186,07	-	1	0,03
		M2	6202,62	0,27%	0,99	0,15
		AFA	6201,99	0,26%	0,99	0,06
		optimum	6211,11	0,40%	0,98	3,69
	$n = 100$	pire cas	33200,16	-	1	1,00
		M2	33649,11	1,35%	0,94	51,35
		AFA	33651,99	1,36%	0,94	8,49
		optimum	33651,99	1,36%	0,94	8,49
variations non-corrélées	$n = 20$	pire cas	6213,74	-	1	0,03
		M2	6244,21	0,49%	0,98	0,16
		AFA	6237,87	0,39%	0,99	0,07
		optimum	6260,27	0,75%	0,98	3,67
	$n = 100$	pire cas	32506,93	-	1	1,08
		M2	33043,85	1,65%	0,93	68,87
		AFA	33051,84	1,68%	0,94	12,84
		optimum	33051,84	1,68%	0,94	12,84

TAB. 5.11 – CCMKP : nombre d'instances où AFA donne de meilleurs résultats que M2, avec la méthode d'échantillonnage.

			AFA > M2	AFA = M2	AFA < M2
CCMKP ₁	variations proportionnelles	$n = 15$	0%	100%	0%
		$n = 50$	10%	82%	8%
		$n = 100$	26%	46%	28%
	variations non-corrélées	$n = 15$	0%	98%	2%
		$n = 50$	8%	82%	10%
		$n = 100$	26%	36%	38%
CCMKP ₂	variations proportionnelles	$n = 15$	0%	100%	0%
		$n = 50$	18%	76%	6%
		$n = 100$	20%	60%	20%
	variations non-corrélées	$n = 15$	0%	100%	0%
		$n = 50$	16%	70%	14%
		$n = 100$	36%	38%	26%

TAB. 5.12 – CCMKP : nombre d'instances où AFA donne de meilleurs résultats que M2, quand on utilise l'inégalité de Hoeffding.

			AFA > M2	AFA = M2	AFA < M2
CCMKP ₁	variations proportionnelles	$n = 50$	6%	88%	6%
		$n = 100$	24%	60%	16%
	variations non-corrélées	$n = 50$	7%	78%	15%
		$n = 100$	24%	48%	28%
CCMKP ₂	variations proportionnelles	$n = 20$	0%	99%	1%
		$n = 100$	18%	69%	13%
	variations non-corrélées	$n = 20$	2%	89%	9%
		$n = 100$	19%	49%	32%

heuristiques, basés sur le modèle robuste proposé, ont été définis pour obtenir des solutions réalisables à des problèmes combinatoires sous contraintes probabilistes. À notre connaissance, c'est la première fois que de tels algorithmes sont développés de manière à traiter des problèmes combinatoires sous contraintes probabilistes très généraux. De nombreux tests numériques ont été effectués sur des problèmes de sac-à-dos multi-dimensionnels. On met ainsi en évidence la facilité de mise en œuvre des algorithmes en pratique, leurs temps de calcul restreints et la qualité raisonnable des solutions produites. Ces méthodes semblent ainsi bien adaptées pour traiter de manière efficace des problèmes de grandes tailles.

Ces résultats numériques montrent aussi que des améliorations sont possibles, voire nécessaires. Les solutions obtenues pour des sac-à-dos multi-dimensionnels améliorent grandement le pire cas, mais restent relativement éloignées de l'optimum. De plus, on a défini une borne supérieure pour évaluer la qualité des résultats, mais elle se révèle très faible en pratique. L'efficacité en temps de calcul des algorithmes proposés les rend attractifs pour des problèmes de grandes tailles, mais une résolution optimale pour des tailles plus faibles serait bienvenue.

Deuxième partie

Résolution optimale de problèmes
combinatoires sous contraintes
probabilistes

Chapitre 6

Résolution optimale de problèmes combinatoires sous contraintes probabilistes

Motivations : Des approches robustes ont été développées pour résoudre de manière approchée des problèmes combinatoires sous contraintes probabilistes, en s'appuyant sur l'optimisation robuste. Cependant, peu de garanties ont été données sur la qualité des résultats de ces algorithmes. Il est donc naturel de s'intéresser maintenant à la résolution optimale de programmes linéaires en nombres entiers sous contraintes probabilistes.

Dans ce chapitre, notre but est de définir un algorithme de résolution optimale de problèmes combinatoires sous contraintes probabilistes. Après avoir précisé le cadre général de cette étude dans la partie 6.1, on définit une description linéaire du problème traité (partie 6.2). Ce modèle s'appuie sur un scénario dit “basique”, dont les propriétés sont étudiées au paragraphe 6.3. Les aspects algorithmiques sont abordés dans la partie 6.4, et des tests numériques sont enfin rapportés dans la partie 6.5.

6.1 Notations et préliminaires

6.1.1 Cadre principal

On rappelle que $I = \{1, \dots, n\}$ et $J = \{1, \dots, m\}$ ($n, m \in \mathbb{N}$). Considérons une matrice aléatoire A , dont les composantes $\{A_{ji}\}_{j \in J, i \in I}$ sont des variables aléatoires à valeurs réelles, et un vecteur $b \in \mathbb{R}^m$. Pour chaque ligne $j \in J$, certains coefficients peuvent être connus avec certitude ; ceci correspond à des variables aléatoires dont le support est réduit à une unique valeur. On note $I_u(j) \subseteq I$ l'ensemble des indices des coefficients dont le support n'est pas réduit à une seule valeur : ce sont les coefficients réellement incertains du problème pour la ligne j .

On s'intéresse à l'ensemble suivant :

$$X = \left\{ x \in \{0, 1\}^n \mid \forall j \in J, \mathbb{P}(A_j x \leq b_j) \geq 1 - \varepsilon_j \right\}.$$

Pour tout $j \in J$, $\varepsilon_j \in [0, 1[$ est la tolérance de non-réalisabilité de la ligne j . Observer que les contraintes linéaires classiques sont bien prises en compte dans ce modèle, puisque l'on peut avoir $I_u(j) = \emptyset$ et $\varepsilon_j = 0$ pour certains $j \in J$. De plus, comme déjà remarqué précédemment

(voir le début du chapitre 3), considérer que seuls les coefficients de A sont incertains permet de prendre en compte les cas où le second membre b serait également incertain.

Soit $j \in J$, on utilisera la notation suivante : $X_j = \{x \in \{0, 1\}^n \mid \mathbb{P}(A_j x \leq b_j) \geq 1 - \varepsilon_j\}$. On a bien sûr : $X = \bigcap_{j \in J} X_j$.

6.1.2 Extensions de l'analyse

Contraintes de probabilités jointes : X correspond à des contraintes de probabilités *séparées* pour chaque ligne $j \in J$. Pourtant, les contraintes de probabilités *jointes* sont aussi très importantes pour nombre d'applications, où l'on cherche une solution globalement réalisable avec une probabilité au moins $1 - \varepsilon$. L'ensemble des solutions d'un tel problème est :

$$X' = \{x \in \{0, 1\}^n \mid \mathbb{P}(Ax \leq b) \geq 1 - \varepsilon\}.$$

Lemme 33 Si $\varepsilon_j = \varepsilon$ pour tout $j \in J$, alors : $X' \subseteq X$.

Ainsi, les caractérisations données pour X peuvent servir pour définir des relaxations de X' .

Programmation mixte en nombres entiers 0-1 : Par souci de simplicité, ce chapitre est écrit entièrement dans un contexte de variables 0-1. Pourtant, tous les résultats et algorithmes décrits sont directement adaptables pour des ensembles faisant intervenir des variables continues, s'ils sont de la forme suivante :

$$X'' = \{x \geq 0 \mid \forall i \in N, x_i \in \{0, 1\}, \text{ et } \forall j \in J, \mathbb{P}(A_j x \leq b_j) \geq 1 - \varepsilon_j\},$$

où $N \subseteq I$ et $\bigcup_{j \in J} I_u(j) \subseteq N$.

Variables en nombres entiers généraux : Supposons qu'il existe $i \in \bigcup_{j \in J} I_u(j)$ tel que x_i soit une variable à valeurs entières : $x_i \in \mathbb{N}$. On suppose sans perte de généralité que x_i est borné : il existe $p \in \mathbb{N}^*$ tel que $x_i \leq 2^p$. On peut alors décomposer x_i en base 2 : $x_i = \sum_{k=0}^p 2^k . x_i^k$, avec $x_i^k \in \{0, 1\}$ pour tout k . On se ramène ainsi au cadre de ce chapitre.

Programmes non-linéaire : L'approche décrite peut être adaptée à la programmation non-linéaire sous contraintes probabilistes. Pour plus de détails, on se reportera au paragraphe 6.2.3.

6.2 Description de X à l'aide d'inégalités linéaires

6.2.1 Modèles déterministes équivalents

Dans certains cas, des modèles déterministes équivalents classiques peuvent être utilisés (cas des distributions stables, cf. partie 5.4.1). Supposons par exemple que tous les coefficients incertains sont indépendants et normalement distribués : pour tout $j \in J$, pour tout $i \in I_u(j)$, $A_{ji} \sim \mathcal{N}(\mu_{ji}, \sigma_{ji}^2)$. Dans ce cas, il est bien connu que (voir [75, 84, 66, 16, 45]) :

$$X_j = \left\{ x \in \{0, 1\}^n \mid \sum_{i \in I_u(j)} \mu_{ji} x_i + \sum_{i \notin I_u(j)} A_{ji} x_i + \Phi^{-1}(1 - \varepsilon_j) \sqrt{\sum_{i \in I_u(j)} \sigma_{ji}^2 x_i^2} \leq b_j \right\},$$

où Φ est la fonction de répartition de la loi normale standard $\mathcal{N}(0, 1)$. Noter que de tels modèles déterministes équivalents existent aussi si les variables aléatoires sont des gaussiennes corrélées

(on renvoie à [84] ; voir aussi le paragraphe 8.3.3 pour une mise en œuvre dans le cas d'une application).

Pour simplifier les écritures, on note : $\mu_{ji} = A_{ji}$ pour $i \notin I_u(j)$. L'inégalité $j \in J$ donnée ci-dessus peut être transformée comme suit :

$$\begin{aligned} & \sum_{i \in I} \mu_{ji} x_i + \Phi^{-1}(1 - \varepsilon_j) \sqrt{\sum_{i \in I_u(j)} \sigma_{ji}^2 x_i^2} \leq b_j \quad (a) \\ \Leftrightarrow & \begin{cases} \sum_{i \in I} \mu_{ji} x_i \leq b_j, \\ [\Phi^{-1}(1 - \varepsilon_j)]^2 \cdot \sum_{i \in I_u(j)} \sigma_{ji}^2 x_i^2 \leq (b_j - \sum_{i \in I} \mu_{ji} x_i)^2. \end{cases} \end{aligned}$$

Puisque $x \in \{0, 1\}^n$, cette dernière expression peut être linéarisée grâce à des procédés classiques. Alors (a) est équivalent à :

$$\begin{aligned} & \begin{cases} \sum_{i \in I} \mu_{ji} x_i \leq b_j, \\ [\Phi^{-1}(1 - \varepsilon_j)]^2 \cdot \sum_{i \in I_u(j)} \sigma_{ji}^2 x_i \leq b_j^2 + \sum_{i \in I} \mu_{ji} (\mu_{ji} - 2b_j) x_i + \sum_{i \neq k} \mu_{ji} \mu_{jk} x_i x_k \end{cases} \\ \Leftrightarrow & \exists y \geq 0 \text{ t.q. } \begin{cases} \sum_{i \in I} \mu_{ji} x_i \leq b_j, \\ [\Phi^{-1}(1 - \varepsilon_j)]^2 \cdot \sum_{i \in I_u(j)} \sigma_{ji}^2 x_i \leq b_j^2 + \sum_{i \in I} \mu_{ji} (\mu_{ji} - 2b_j) x_i + \sum_{i \neq k} \mu_{ji} \mu_{jk} y_{ik}, \\ y_{ik} \leq x_i \text{ et } y_{ik} \leq x_k, \forall (i, k) \in I^2. \end{cases} \end{aligned}$$

Ainsi, sous les hypothèses gaussiennes formulées, X_j peut être représenté à l'aide d'un nombre polynomial d'inégalités linéaires. [20] a donné récemment d'intéressantes extensions à ce cadre gaussien classique, en considérant une classe plus générale de distributions dites "radiales". Les auteurs prouvent que, dans ce contexte, la contrainte probabiliste de X_j peut s'écrire de manière équivalente comme une contrainte conique convexe du second ordre. Comme on travaille avec des variables 0-1, de telles contraintes pourraient être linéarisées selon la démarche déjà utilisée ci-dessus.

Bien que le cadre gaussien puisse paraître très restrictif à première vue, il peut fournir dans bien des cas une approximation satisfaisante d'hypothèses probabilistes diverses. Ceci est une conséquence du théorème de la limite centrale et de ses variantes (voir annexe B.2). Pourtant, dans l'étude présente, notre but est de travailler avec des hypothèses aussi générales que possible. De telles approximations gaussiennes ne seront discutées qu'au niveau des tests numériques (paragraphe 6.5).

6.2.2 Inégalités linéaires pour caractériser X

Définition 8 Soit $j \in J$. $\underline{A}_j \in \mathbb{R}^n$ est un scénario basique si pour tout $i \in I \setminus I_u(j)$, $\underline{A}_{ji} = A_{ji}$, et si :

$$\forall I_1 \subseteq I_u(j), I_1 \neq \emptyset : \mathbb{P} \left(\sum_{i \in I_1} (A_{ji} - \underline{A}_{ji}) \leq 0 \right) < 1 - \varepsilon_j. \quad (6.1)$$

Dans ce qui suit, un tel scénario basique est supposé connu. On verra plus loin (paragraphe 6.3) que ceci n'est restrictif ni en théorie, ni en pratique. La propriété principale de ce scénario basique est que tout point $x \in X_j$ doit être réalisable pour ce scénario. Ceci est l'objet du résultat suivant.

Lemme 34 Pour tout $j \in J$, si \underline{A}_j est un scénario basique, alors l'inégalité suivante est valide pour X_j :

$$\underline{A}_j x \leq b_j. \quad (6.2)$$

Preuve : Supposons qu'il existe $x \in X_j$ tel que $\underline{A}_j x > b_j$. Si $x = 0$, alors le fait que $x \in X_j$ implique que $b_j \geq 0$, et on ne peut pas avoir $\underline{A}_j x > b_j$. En conséquence, $x \neq 0$. Observer que : $\mathbb{P}(A_j x \leq \underline{A}_j x) \geq \mathbb{P}(A_j x \leq b_j) \geq 1 - \varepsilon_j$. Considérons : $I_1 = \{i \in I_u(j) | x_i = 1\}$, on a : $\mathbb{P}(\sum_{i \in I_1} (A_{ji} - \underline{A}_{ji}) \leq 0) = \mathbb{P}(A_j x \leq \underline{A}_j x) \geq 1 - \varepsilon_j$. En conséquence, \underline{A}_j n'est pas un scénario basique. \square

Les inégalités (6.2) seront aussi dites *basiques*. Soit $j \in J$. Pour tout $i \in I_u(j)$, on note : $\eta_{ji} = A_{ji} - \underline{A}_{ji}$. Soit $\bar{x} \in \{0, 1\}^n$, observer que : $\sum_{i \in I_u(j)} A_{ji} \bar{x}_i = \sum_{i \in I_u(j)} (\underline{A}_{ji} + \eta_{ji}) \bar{x}_i$. Si l'on pose : $I_1 = \{i \in I_u(j) | \bar{x}_i = 1\}$, cette dernière expression peut s'écrire : $\sum_{i \in I_u(j)} A_{ji} \bar{x}_i = \sum_{i \in I_u(j)} \underline{A}_{ji} \bar{x}_i + \sum_{i \in I_1} \eta_{ji}$. Il en découle que :

$$\bar{x} \in X_j \Leftrightarrow \mathbb{P}\left(\sum_{i \in I_1} \eta_{ji} \leq b_j - \underline{A}_j \cdot \bar{x}\right) \geq 1 - \varepsilon_j.$$

Cette observation motive la définition de la quantité suivante, pour tout $I_1 \subseteq I_u(j)$:

$$D_j(I_1) = \sup \left\{ d \mid \mathbb{P}\left(\sum_{i \in I_1} \eta_{ji} \leq d\right) < 1 - \varepsilon_j \right\}. \quad (6.3)$$

Cette grandeur satisfait les propriétés suivantes :

Lemme 35 Soient $j \in J$ et $I_1 \subseteq I_u(j)$:

- (i) si $I_1 = \phi$, alors $D_j(I_1) = 0$;
- (ii) $D_j(I_1) > 0$;
- (iii) $\mathbb{P}(\sum_{i \in I_1} \eta_{ji} \leq D_j(I_1)) \geq 1 - \varepsilon_j$.

Preuve : Pour prouver (i), il suffit d'observer que si $I_1 = \phi$, alors $\sum_{i \in I_1} \eta_{ji} = 0$. Dans ce cas, pour tout $d < 0$, $\mathbb{P}(\sum_{i \in I_1} \eta_{ji} \leq d) = 0$, et pour tout $d \geq 0$, $\mathbb{P}(\sum_{i \in I_1} \eta_{ji} \leq d) = 1$. Passons maintenant au point (iii). Supposons que : $\mathbb{P}(\sum_{i \in I_1} \eta_{ji} \leq D_j(I_1)) < 1 - \varepsilon_j$. Puisqu'une fonction de répartition est toujours continue à droite, il existe $\tau > 0$ tel que : $\mathbb{P}(\sum_{i \in I_1} \eta_{ji} \leq D_j(I_1) + \tau) < 1 - \varepsilon_j$. Ceci est en contradiction avec la définition (6.3). (ii) est une conséquence directe de la définition (6.1) et de (iii). \square

On peut aussi remarquer que :

Lemme 36 Soit $j \in J$. Supposons que les $\{\eta_{ji}\}_{i \in I_u(j)}$ sont indépendants et identiquement distribués (i.i.d.). Alors, pour tous sous-ensembles $I_1 \subseteq I_u(j)$ et $I_2 \subseteq I_u(j)$, non-vides, tels que $|I_1| = |I_2|$: $D_j(I_1) = D_j(I_2)$.

Ce fait évident aura des conséquences pratiques importantes. Maintenant, ces coefficients $D_j(I_1)$ peuvent être utilisés pour construire des inégalités valides pour X_j :

Lemme 37 Soit $j \in J$. Pour tout $I_1 \subseteq I_u(j)$, soit $I_0 = I_u(j) \setminus I_1$:

$$\sum_{i \in I_1} [\underline{A}_{ji} + D_j(I_1)] x_i + \sum_{i \in I_0} [\underline{A}_{ji} - D_j(I_1)] x_i + \sum_{i \notin I_u(j)} A_{ji} x_i \leq b_j + (|I_1| - 1) \cdot D_j(I_1) \quad (6.4)$$

est une inégalité valide pour X_j .

Preuve : Lorsque $I_1 = \phi$, on obtient l'inégalité basique (6.2). Supposons maintenant que $I_1 \neq \phi$. Observer d'abord que (6.4) peut s'écrire de manière équivalente :

$$\sum_{i \in I} \underline{A}_{ji} x_i \leq b_j + \left(\sum_{i \in I_0} x_i + |I_1| - \sum_{i \in I_1} x_i - 1 \right) \cdot D_j(I_1).$$

Soit $d < D_j(I_1)$, on montre que l'inégalité suivante est valide pour X_j :

$$(a) \quad \sum_{i \in I} \underline{A}_{ji} x_i \leq b_j + \left(\sum_{i \in I_0} x_i + |I_1| - \sum_{i \in I_1} x_i - 1 \right) \cdot d.$$

Soit $x \in X_j$, supposons par contradiction que x viole l'inégalité (a). Considérons d'abord le cas où $\{i \in I_u(j) | x_i = 1\} = I_1$. On a alors : $\sum_{i \in I} \underline{A}_{ji} x_i > b_j - d$. Donc :

$$\mathbb{P}(A_j x \leq b_j) \leq \mathbb{P}(A_j x \leq \underline{A}_j x + d) = \mathbb{P}\left(\sum_{i \in I_u(j)} \eta_{ji} x_i \leq d \right) = \mathbb{P}\left(\sum_{i \in I_1} \eta_{ji} \leq d \right).$$

Puisque $d < D_j(I_1)$, d'après la définition de $D_j(I_1)$: $\mathbb{P}(\sum_{i \in I_1} \eta_{ji} \leq d) < 1 - \varepsilon_j$, ce qui implique : $\mathbb{P}(A_j x \leq b_j) < 1 - \varepsilon_j$. Ceci est en contradiction avec $x \in X_j$.

Considérons maintenant le cas où $\{i \in I_u(j) | x_i = 1\} \neq I_1$. Alors : $\sum_{i \in I_0} x_i + |I_1| - \sum_{i \in I_1} x_i \geq 1$, ce qui implique : $\sum_{i \in I} \underline{A}_{ji} x_i > b_j$ (contradiction, cf. lemme 34).

On a donc prouvé que l'inégalité (a) est valide pour X_j , quel que soit $d < D_j(I_1)$. On en déduit que (a) est valide également pour $D_j(I_1)$ lui-même. \square

Dans le résultat suivant, on montre que les inégalités (6.4) suffisent à décrire exactement l'ensemble des points 0-1 de X :

Proposition 19

$$X = \left\{ x \in \{0, 1\}^n \mid x \text{ satisfait toutes les inégalités (6.4)} \right\}.$$

Preuve : L'inclusion \subseteq vient du lemme 37. Pour l'inclusion inverse, considérer $x \in \{0, 1\}^n$ tel que $x \notin X$: il existe $j \in J$ tel que $\mathbb{P}(A_j x \leq b_j) < 1 - \varepsilon_j$. Supposons que x satisfait toutes les inégalités (6.4). On a en particulier : $\underline{A}_j x + D_j(I_1) \leq b_j$, avec $I_1 = \{i \in I_u(j) | x_i = 1\}$. On en déduit que : $\mathbb{P}(A_j x \leq \underline{A}_j x + D_j(I_1)) < 1 - \varepsilon_j$. Ceci est équivalent à : $\mathbb{P}(\eta_j x \leq D_j(I_1)) = \mathbb{P}(\sum_{i \in I_1} \eta_{ji} \leq D_j(I_1)) < 1 - \varepsilon_j$, ce qui est une contradiction avec le point (iii) du lemme 35. \square

Quelques remarques méritent d'être faites. D'abord, d'après la proposition ci-dessus, l'ensemble X peut être décrit à l'aide de $\sum_{j \in J} 2^{|I_u(j)|}$ inégalités linéaires. Cependant, il est clair que l'on n'a jamais besoin de plus de $2^{|I_u|}$ inégalités pour décrire X , où $I_u = \bigcup_{j \in J} I_u(j)$. En effet, pour tout $x \in \{0, 1\}^n$ tel que $x \notin X$, il suffit d'une seule inégalité pour séparer x de X . Le choix de cette inégalité n'implique que les composantes $\{x_i\}_{i \in I_u}$.

Ensuite, d'après la preuve du lemme 37, lorsque $I_1 \neq \phi$, on voit que l'inégalité (6.4) demeure valide pour X si l'on considère $d \leq D_j(I_1)$ au lieu de $D_j(I_1)$. Pourtant, si $\{i \in I_u(j) | x_i = 1\} = I_1$, l'inégalité écrite pour $D_j(I_1)$ est plus forte qu'une autre écrite avec $d < D_j(I_1)$. Si $|I_1| - \sum_{i \in I_1} x_i - 1 \geq 0$, l'inégalité (6.4) est dominée par l'inégalité basique $\underline{A}_j x \leq b_j$, que l'on considère $D_j(I_1)$ ou $d < D_j(I_1)$.

6.2.3 Idée de généralisation à un contexte non-linéaire

Cette partie constitue une parenthèse consacrée à la généralisation des idées exposées dans un contexte non-linéaire. Considérons l'ensemble suivant :

$$X = \left\{ x \in \{0, 1\}^n \mid \forall j \in J, \mathbb{P}(g_j(x, \xi) \leq 0) \geq 1 - \varepsilon_j \right\}.$$

ξ est une variable aléatoire prenant ses valeurs dans \mathbb{R}^p ($p \in \mathbb{N}$). On suppose que pour tout $j \in J$, il existe $\underline{\xi}_j \in \mathbb{R}^p$ tel que, pour tout $x \in \{0, 1\}^n$: $\mathbb{P}\left(g_j(x, \xi) \leq g_j(x, \underline{\xi}_j)\right) < 1 - \varepsilon_j$.

Les principaux résultats de la partie 6.2.2 peuvent maintenant être transposés dans ce contexte non-linéaire.

Lemme 38 *Pour tout $j \in J$, l'inégalité $g_j(x, \underline{\xi}_j) \leq 0$ est valide pour X .*

Preuve : Soit $x \in X$, supposons que $g_j(x, \underline{\xi}_j) > 0$. Alors : $\mathbb{P}(g_j(x, \xi) \leq 0) \leq \mathbb{P}\left(g_j(x, \xi) \leq g_j(x, \underline{\xi}_j)\right)$. De plus, d'après la définition de $\underline{\xi}_j$: $\mathbb{P}\left(g_j(x, \xi) \leq g_j(x, \underline{\xi}_j)\right) < 1 - \varepsilon_j$ (contradiction). \square

Pour tout $I_1 \subseteq I$, on note $x(I_1) \in \{0, 1\}^n$ le vecteur caractéristique de I_1 , et :

$$D_j(I_1) = \sup \left\{ d \mid \mathbb{P}\left(g_j(x(I_1), \xi) \leq g_j(x(I_1), \underline{\xi}_j) + d\right) < 1 - \varepsilon_j \right\}.$$

Comme dans le lemme 35, on a $D_j(I_1) \geq 0$. De plus, il est aisé de montrer que : $\mathbb{P}\left(g_j(x(I_1), \xi) \leq g_j(x(I_1), \underline{\xi}_j) + D_j(I_1)\right) \geq 1 - \varepsilon_j$. Par ailleurs, l'inégalité suivante est valide pour X :

$$g_j(x, \underline{\xi}_j) + D_j(I_1) \cdot \left(\sum_{i \in I_1} x_i - \sum_{i \notin I_1} x_i \right) \leq (|I_1| - 1) \cdot D_j(I_1). \quad (6.5)$$

En effet, lorsque $x \neq x(I_1)$, l'inégalité est dominée par $g_j(x, \underline{\xi}_j) \leq 0$. Si $x = x(I_1)$, l'inégalité (6.5) devient : $g_j(x, \underline{\xi}_j) + D_j(I_1) \leq 0$, dont on montre qu'elle est valide, compte tenu de nos hypothèses. Il suffit de reprendre le raisonnement de la preuve du lemme 37 : soit d tel que $\mathbb{P}(g_j(x, \xi) \leq g_j(x, \underline{\xi}_j) + d) < 1 - \varepsilon_j$, si $g_j(x, \underline{\xi}_j) + d > 0$, alors $\mathbb{P}(g_j(x, \xi) \leq 0) \leq \mathbb{P}(g_j(x, \xi) \leq g_j(x, \underline{\xi}_j) + d) < 1 - \varepsilon_j$, et donc $x \notin X$.

Proposition 20

$$X = \left\{ x \in \{0, 1\}^n \mid x \text{ satisfait les inégalités (6.5)} \right\}.$$

Preuve : On a déjà dit que l'inclusion \subseteq était correcte. Considérons maintenant $x \notin X$: il existe $j \in J$ tel que $\mathbb{P}(g_j(x, \xi) \leq 0) < 1 - \varepsilon_j$. Supposons que x satisfait toutes les inégalités (6.5), alors en particulier : $g_j(x, \underline{\xi}_j) + D_j(I_1) \leq 0$, avec $I_1 = \{i \in I \mid x_i = 1\}$. On en déduit : $\mathbb{P}(g_j(x, \xi) \leq g_j(x, \underline{\xi}_j) + D_j(I_1)) \leq \mathbb{P}(g_j(x, \xi) \leq 0) < 1 - \varepsilon_j$. Ceci est en contradiction avec la définition de $D_j(I_1)$. \square

6.3 Étude des scénarios basiques

6.3.1 Structure de l'ensemble des scénarios basiques

Soit $j \in J$, on note \mathcal{B}_j l'ensemble des scénarios basiques, *i.e.* l'ensemble des vecteurs satisfaisant la propriété (6.1). Il est clair que tout élément dans la fermeture $cl(\mathcal{B}_j)$ peut en pratique être utilisé comme scénario basique pour construire des inégalités (6.2) et (6.4). Illustrons ce fait à l'aide des inégalités (6.2) : soit $\underline{A}_j \in cl(\mathcal{B}_j)$. Il existe une suite $(\underline{A}_j^{(k)})_{k \in \mathbb{N}} \subset \mathcal{B}_j$ telle que $\underline{A}_j^{(k)} \rightarrow \underline{A}_j$ lorsque $k \rightarrow \infty$. Soit $x \in X_j$ quelconque, on sait que pour tout $k \in \mathbb{N}$: $\underline{A}_j^{(k)} \cdot x \leq b_j$. En passant à la limite, on a : $\underline{A}_j \cdot x \leq b_j$. Ceci est vrai pour tout $x \in X_j$: l'inégalité (6.2) est donc valide pour X_j si l'on prend comme scénario basique $\underline{A}_j \in cl(\mathcal{B}_j)$.

Lemme 39 Soit $j \in J$, $cl(\mathcal{B}_j)$ est un polyèdre.

Preuve : Pour tout $I_1 \subseteq I_u(j)$ non-vidé, on introduit la variable aléatoire : $S(I_1) = \sum_{i \in I_1} A_{ji}$. Alors \underline{A}_j est un scénario basique si, et seulement si :

$$\forall I_1 \subseteq I_u(j), I_1 \neq \emptyset : \sum_{i \in I_1} \underline{A}_{ji} < \sup \{s \mid \mathbb{P}(S(I_1) \leq s) < 1 - \varepsilon_j\}. \quad (6.6)$$

Donc l'ensemble $cl(\mathcal{B}_j)$ est défini à l'aide d'un nombre fini d'inégalités linéaires. \square

Un scénario de $cl(\mathcal{B}_j)$ est dit *extrême* s'il ne peut pas s'exprimer comme combinaison linéaire d'autres scénarios de $cl(\mathcal{B}_j)$.

Lemme 40 La description linéaire de X_j obtenue en considérant tous les scénarios de $cl(\mathcal{B}_j)$ est équivalente à celle obtenue en ne considérant que les scénarios extrêmes.

Preuve : Considérons un scénario basique non-extrême $\underline{A}_j \in cl(\mathcal{B}_j)$, il existe deux autres scénarios basiques \underline{A}_j^1 et \underline{A}_j^2 tels que : $\underline{A}_j = \lambda \underline{A}_j^1 + (1 - \lambda) \underline{A}_j^2$, avec $\lambda \in]0, 1[$. Soit $I_1 \subseteq I_u(j)$ non-vidé. Observer que :

$$D_j(I_1) + \sum_{i \in I_1} \underline{A}_{ji} = D_j^1(I_1) + \sum_{i \in I_1} \underline{A}_{ji}^1 = D_j^2(I_1) + \sum_{i \in I_1} \underline{A}_{ji}^2. \quad (6.7)$$

En effet, toutes ces expressions sont égales à : $\sup \{s \mid \mathbb{P}(\sum_{i \in I_1} A_{ji} \leq s) < 1 - \varepsilon_j\}$. On en déduit que : $D_j(I_1) = \lambda D_j^1(I_1) + (1 - \lambda) D_j^2(I_1)$. Il est donc clair que l'inégalité (6.4) construite pour \underline{A}_j est une combinaison linéaire de celles correspondant à \underline{A}_j^1 et \underline{A}_j^2 . \square

Malheureusement, les scénarios extrêmes sont le plus souvent très nombreux et difficiles à calculer.

6.3.2 Construction de “bons” scénarios basiques

Le résultat suivant permet d'obtenir un premier scénario basique \underline{A}_j :

Lemme 41 Soit $j \in J$, on note : $n_j = |I_u(j)|$. Soit \underline{A}_j défini comme suit :

(i) si aucune hypothèse probabiliste particulière n'est effectuée, alors pour tout $i \in I_u(j)$: $\mathbb{P}(A_{ji} \leq \underline{A}_{ji}) < (1 - \varepsilon_j)/n_j$;

(ii) si les variables aléatoires $\{A_{ji}\}_{i \in I_u(j)}$ sont indépendantes, alors pour tout $i \in I_u(j)$: $\mathbb{P}(A_{ji} \leq \underline{A}_{ji}) < 1 - \varepsilon_j^{1/n_j}$;

(iii) si les $\{A_{ji}\}_{i \in I_u(j)}$ sont indépendants et symétriquement distribués, et si $\varepsilon_j < 0,5$, alors pour tout $i \in I_u(j)$: $\underline{A}_{ji} = \mathbb{E}[A_{ji}]$.

Alors \underline{A}_j est un scénario basique, i.e. il satisfait (6.1).

Preuve : Soit $I_1 \subseteq I_u(j)$. On a :

$$\mathbb{P}\left(\sum_{i \in I_1} (A_{ji} - \underline{A}_{ji}) \leq 0\right) \leq \mathbb{P}\left(\exists i \in I_1 \text{ t.q. } A_{ji} \leq \underline{A}_{ji}\right) \leq \mathbb{P}\left(\exists i \in I_u(j) \text{ t.q. } A_{ji} \leq \underline{A}_{ji}\right).$$

Si aucune hypothèse probabiliste particulière n'est effectuée, on a toujours : $\mathbb{P}(\exists i \in I_u(j) \text{ t.q. } A_{ji} \leq \underline{A}_{ji}) \leq \sum_{i \in I_u(j)} \mathbb{P}(A_{ji} \leq \underline{A}_{ji})$. En supposant que \underline{A}_j est défini selon (i), on a : $\mathbb{P}(\sum_{i \in I_1} (A_{ji} - \underline{A}_{ji}) \leq 0) < n_j \cdot (1 - \varepsilon_j)/n_j = 1 - \varepsilon_j$.

Supposons maintenant que les variables aléatoires sont indépendantes. Observer que : $\mathbb{P}(\exists i \in I_u(j) \text{ t.q. } A_{ji} \leq \underline{A}_{ji}) = 1 - \mathbb{P}(\forall i \in I_u(j), A_{ji} > \underline{A}_{ji}) = 1 - \prod_{i \in I_u(j)} \mathbb{P}(A_{ji} > \underline{A}_{ji})$. D'après (ii), on a : $\mathbb{P}(\sum_{i \in I_1} (A_{ji} - \underline{A}_{ji}) \leq 0) < 1 - \varepsilon_j$.

Finalement, supposons que les variables aléatoires sont indépendantes et symétriquement distribuées. Dans ce cas, la variable aléatoire $S = \sum_{i \in I_1} A_{ji}$ est également symétriquement distribuée. Ceci implique que : $\mathbb{P}(S \leq \mathbb{E}[S]) = 0,5$. Alors, d'après (iii) : $\mathbb{P}(\sum_{i \in I_1} (A_{ji} - \underline{A}_{ji}) \leq 0) = \mathbb{P}(\sum_{i \in I_1} A_{ji} \leq \sum_{i \in I_1} \underline{A}_{ji}) = \mathbb{P}(S \leq \mathbb{E}[S]) = 0,5$. Ceci prouve que (6.1) est vérifié si $\varepsilon_j < 0,5$. \square

La description linéaire de X donnée à la proposition 19 repose sur un scénario basique \underline{A}_j qui n'est pas unique, et qui peut être construit de différentes manières. On montre ci-après que le choix de ce scénario basique a un impact direct sur la force des inégalités (6.2) et (6.4) :

Lemme 42 *Soit $j \in J$, considérons deux scénarios basiques \underline{A}_j^1 et \underline{A}_j^2 tels que $\underline{A}_j^1 \leq \underline{A}_j^2$. Alors la relaxation continue de X_j obtenue à partir des inégalités (6.2) et (6.4) avec \underline{A}_j^2 est plus forte que celle associée à \underline{A}_j^1 .*

Preuve : Considérons d'abord les inégalités basiques (6.2) : puisque $\underline{A}_j^1 \leq \underline{A}_j^2$, il est clair que $\underline{A}_j^2 x \leq b_j$ domine $\underline{A}_j^1 x \leq b_j$. Maintenant, considérons les inégalités (6.4). Soit $I_1 \subseteq I_u(j)$, $I_1 \neq \emptyset$. Soient $D_j^1(I_1)$, $D_j^2(I_1)$, associés respectivement à \underline{A}_j^1 , \underline{A}_j^2 . En se souvenant de l'observation (6.7), on a : $D_j^1(I_1) = D_j^2(I_1) + \sum_{i \in I_1} (\underline{A}_{ji}^2 - \underline{A}_{ji}^1)$. Soit $x \in [0, 1]^n$, on introduit les notations suivantes (comme précédemment, $I_0 = I_u(j) \setminus I_1$) :

$$\begin{cases} Q(x) = \sum_{i \in I_1} x_i - \sum_{i \in I_0} x_i - |I_1| + 1, \\ \Gamma_j^1(x) = \sum_{i \in I} \underline{A}_{ji}^1 x_i + Q(x) \cdot D_j^1(I_1), \\ \Gamma_j^2(x) = \sum_{i \in I} \underline{A}_{ji}^2 x_i + Q(x) \cdot D_j^2(I_1). \end{cases}$$

Alors :

$$\begin{aligned} \Gamma_j^1(x) &= \sum_{i \in I} \underline{A}_{ji}^1 x_i + Q(x) \cdot [D_j^2(I_1) + \sum_{i \in I_1} (\underline{A}_{ji}^2 - \underline{A}_{ji}^1)] \\ &= \Gamma_j^2(x) + \sum_{i \in I} (\underline{A}_{ji}^1 - \underline{A}_{ji}^2) x_i + Q(x) \cdot \sum_{i \in I_1} (\underline{A}_{ji}^2 - \underline{A}_{ji}^1) \\ &= \Gamma_j^2(x) + \sum_{i \in I_1} (\underline{A}_{ji}^1 - \underline{A}_{ji}^2) (x_i - Q(x)) + \sum_{i \in I_0} (\underline{A}_{ji}^1 - \underline{A}_{ji}^2) x_i. \end{aligned}$$

Supposons qu'il existe $i \in I_1$ tel que $x_i < Q(x)$. Dans ce cas, on aurait : $\sum_{k \in I_1 \setminus \{i\}} x_k - \sum_{i \in I_0} x_i + 1 - |I_1| > 0$, i.e. : $\sum_{k \in I_1 \setminus \{i\}} x_k - \sum_{i \in I_0} x_i > |I_1| - 1$. Puisque $x_k \leq 1$ pour tout $k \in I_1$, ce n'est pas possible. Donc on a prouvé que : $\forall i \in I_1, x_i \geq Q(x)$.

Finalement, comme $\underline{A}_j^1 \leq \underline{A}_j^2$ et $x \geq 0$, on obtient : $\Gamma_j^1(x) \leq \Gamma_j^2(x)$. On en déduit que pour tout $x \in [0, 1]^n$: $\Gamma_j^2(x) \leq b_j \Rightarrow \Gamma_j^1(x) \leq b_j$. Ceci montre que la relaxation continue de X_j obtenue à partir de \underline{A}_j^2 est plus forte que celle issue de \underline{A}_j^1 . \square

On a donc intérêt à choisir un scénario basique aussi grand que possible. Un scénario basique $\underline{A}_j \in cl(\mathcal{B}_j)$ est dit *non-dominé* si : $\forall A_j \in cl(\mathcal{B}_j), A_j \geq \underline{A}_j \Rightarrow A_j = \underline{A}_j$. Étant donné un scénario basique dominé, on peut chercher à l'améliorer. On définit, pour tout $v \in \mathbb{R}_+^n$:

$$\lambda^* = \min_{I_1 \subseteq I_u(j), I_1 \neq \emptyset} \left\{ \frac{D_j(I_1)}{\sum_{i \in I_1} v_i} \right\},$$

avec la convention : $\sum_{i \in I_1} v_i = 0 \Rightarrow D_j(I_1) / \sum_{i \in I_1} v_i = +\infty$.

Lemme 43 *Soit $j \in J$. Soient un scénario basique \underline{A}_j et un vecteur $v \in \mathbb{R}_+^n$. Alors : $\lambda^* = \max \{ \lambda \in \mathbb{R} \mid \underline{A}_j + \lambda \cdot v \in cl(\mathcal{B}_j) \}$. De plus, si : $I_u(j) \in \arg \min_{I_1 \subseteq I_u(j)} \{ D_j(I_1) / \sum_{i \in I_1} v_i \}$, alors $\underline{A}_j + \lambda^* \cdot v$ est non-dominé.*

Preuve : D'après la preuve du lemme 39, chaque sous-ensemble I_1 est associé à un hyperplan dont l'équation est : $\sum_{i \in I_1} a_{ji} = \sup \{s \mid \mathbb{P}(\sum_{i \in I_1} A_{ji} \leq s) < 1 - \varepsilon_j\} = D_j(I_1) + \sum_{i \in I_1} \underline{A}_{ji}$ (cf. (6.6)). Notons \mathcal{L} la droite des scénarios égaux à $\underline{A}_j + \lambda.v$ pour $\lambda \in \mathbb{R}$. Soit $I_1 \subseteq I_u(j)$ non-vide tel que $\sum_{i \in I_1} v_i > 0$, on note : $\lambda(I_1) = D_j(I_1) / \sum_{i \in I_1} v_i \geq 0$. Observer que $\underline{A}'_j = \underline{A}_j + \lambda(I_1).v$ est l'intersection de \mathcal{L} avec le plan associé à I_1 , puisque : $\sum_{i \in I_1} \underline{A}'_{ji} = D_j(I_1) + \sum_{i \in I_1} \underline{A}_{ji}$. On en déduit que $\lambda^* = \min_{I_1 \subseteq I_u(j)} \lambda(I_1)$ correspond à un scénario dans $cl(\mathcal{B}_j)$. Le fait que λ^* ne peut pas être augmenté est clair : on violerait alors une inégalité valide pour $cl(\mathcal{B}_j)$.

Supposons maintenant que $I_u(j) \in \arg \min \{D_j(I_1) / \sum_{i \in I_1} v_i\}$. Ceci implique que : $\sum_{i \in I_u(j)} \underline{A}'_{ji} = D_j(I_u(j)) + \sum_{i \in I_u(j)} \underline{A}_{ji}$. Alors, aucun des coefficients $\{\underline{A}'_{ji}\}_{i \in I_u(j)}$ ne peut être augmenté sans violer cette inégalité, qui est valide pour $cl(\mathcal{B}_j)$. \square

Calculer λ^* ne peut en général être fait qu'en temps exponentiel. Pour remédier à cette difficulté pratique, on propose un autre coefficient $\lambda \leq \lambda^*$ plus facile à obtenir. Souvent en pratique, étant donné une direction v d'amélioration, la quantité suivante est facile à calculer pour tout $k \in \{1, \dots, |I_u(j)|\}$:

$$\lambda(k) = \frac{\min \{D_j(I_1) \mid I_1 \subseteq I_u(j), |I_1| = k\}}{\max \{\sum_{i \in I_1} v_i \mid I_1 \subseteq I_u(j), |I_1| = k\}}. \quad (6.8)$$

Il est clair alors que : $\min_k \lambda(k) \leq \lambda^*$. Cette valeur peut être utilisée pour améliorer un scénario basique initial dans la direction v .

Lemme 44 Soit $j \in J$, supposons que les $\{\eta_{ji}\}_{i \in I_u(j)}$ sont i.i.d. : $\lambda^* = \min_k \lambda(k)$.

En effet, dans ce contexte de variables aléatoires i.i.d., $D_j(I_1)$ dépend seulement du cardinal $|I_1|$ (cf. lemme 36) : $\lambda(k) = \min_{I_1: |I_1|=k} \{D_j(I_1) / \sum_{i \in I_1} v_i\}$. Le numérateur de $\lambda(k)$ est obtenu directement par le calcul d'une valeur D_j ; ainsi, du fait du dénominateur de $\{\lambda(k)\}_k$, λ^* peut être calculé en temps $O(n \log(n))$ (complexité du tri des composantes $\{v_i\}_{i \in I_u(j)}$).

Lemme 45 Soit $j \in J$, supposons que les $\{\eta_{ji}\}_{i \in I_u(j)}$ sont i.i.d.. Considérons la direction v telle que $v_i = 1$ pour tout $i \in I_u(j)$, et $v_i = 0$ ailleurs. Alors, $\underline{A}_j^* = \underline{A}_j + \lambda^*.v$ est un scénario non-dominé de $cl(\mathcal{B}_j)$.

Preuve : Le fait que $\underline{A}_j^* \in cl(\mathcal{B}_j)$ vient du lemme précédent. Supposons que ce scénario est dominé. Il existe $k \in I_u(j)$ et $\underline{A}'_j \in cl(\mathcal{B}_j)$ tels que : $\underline{A}'_{jk} > \underline{A}_{jk}^*$ et $\underline{A}'_{ji} = \underline{A}_{ji}^*$ pour tout $i \neq k$. L'idée de la preuve est simple : puisque les variables aléatoires sont i.i.d., et comme $v_i = 1$ pour tout $i \in I_u(j)$, toutes les données sont symétriques, toutes les directions jouent le même rôle. En conséquence, une domination dans la direction k implique une domination dans toute direction $l \neq k, l \in I_u(j)$. En considérant une combinaison convexe des $|I_u(j)|$ scénarios ainsi obtenus, on obtient une domination dans la direction v , ce qui est contradictoire avec la définition de λ^* .

Détaillons maintenant ces idées. Dans cette preuve, on note $U \equiv V$ si les deux variables aléatoires U et V sont identiquement distribuées. Considérons un quelconque $l \in I_u(j) \setminus \{k\}$. On définit un autre scénario $\underline{A}_j^{(l)}$ égal à \underline{A}_j^* , sauf : $\underline{A}_{jl}^{(l)} = \underline{A}_{jl}^* + \underline{A}'_{jk} - \underline{A}_{jk}^*$. Ainsi : $\underline{A}_{jl}^{(l)} > \underline{A}_{jl}^*$. Notre but est de montrer que $\underline{A}_j^{(l)} \in cl(\mathcal{B}_j)$. Soit $I_1 \subseteq I_u(j)$, on s'intéresse à la variable aléatoire : $S^{(l)}(I_1) = \sum_{i \in I_1} A_{ji} - \underline{A}_{ji}^{(l)}$. De la même manière, on note : $S'(I_1) = \sum_{i \in I_1} A_{ji} - \underline{A}_{ji}^*$. Montrons que $S^{(l)}(I_1) \equiv S'(I_1)$ pour tout $I_1 \subseteq I_u(j)$ non-vide. Si $l \notin I_1$ et $k \notin I_1$, c'est clair. Supposons que $l \in I_1$ et $k \in I_1$. On a : $S^{(l)}(I_1) = \sum_{i \in I_1 \setminus \{k, l\}} (A_{ji} - \underline{A}_{ji}^*) + A_{jl} - \underline{A}_{jl}^{(l)} + A_{jk} - \underline{A}_{jk}^{(l)}$. De plus : $A_{jl} - \underline{A}_{jl}^{(l)} + A_{jk} - \underline{A}_{jk}^{(l)} = A_{jl} - \underline{A}_{jl}^* - \underline{A}'_{jk} + \underline{A}_{jk}^* + A_{jk} - \underline{A}_{jk}^* = A_{jl} - \underline{A}_{jl}^* + A_{jk} - \underline{A}'_{jk}$. En conséquence : $S^{(l)}(I_1) = S'(I_1)$.

Supposons maintenant que $l \in I_1$ et $k \notin I_1$. On a : $S^{(l)}(I_1) = \sum_{i \in I_1 \setminus \{l\}} (A_{ji} - \underline{A}'_{ji}) + A_{jl} - \underline{A}_{jl}^{(l)}$. Observer que, puisque η_{jl} et η_{jk} sont i.i.d. : $A_{jl} - \underline{A}_{jl}^{(l)} = \eta_{jl} + \underline{A}_{jl} - \underline{A}_{jl}^{(l)} \equiv \eta_{jk} + \underline{A}_{jl} - \underline{A}_{jk}^* - \underline{A}'_{jk} + \underline{A}_{jk}^* = \eta_{jk} - \underline{A}'_{jk} - \lambda^* + \underline{A}_{jk}^* = \eta_{jk} - \underline{A}'_{jk} + \underline{A}_{jk} = A_{jk} - \underline{A}'_{jk}$. Ainsi, il est prouvé que : $S^{(l)}(I_1) \equiv S'((I_1 \setminus \{l\}) \cup \{k\})$. De plus, comme les variables aléatoires $\{\eta_{ji}\}_{i \in I_u(j)}$ sont i.i.d., et comme $|I_1| = |(I_1 \setminus \{l\}) \cup \{k\}|$, on a : $S'((I_1 \setminus \{l\}) \cup \{k\}) \equiv S'(I_1)$.

Reste le cas où $l \notin I_1$ et $k \in I_1$. On a : $S^{(l)}(I_1) = \sum_{i \in I_1 \setminus \{k\}} (A_{ji} - \underline{A}'_{ji}) + A_{jk} - \underline{A}_{jk}^{(l)}$. Puisque η_{jl} et η_{jk} sont i.i.d., on a aussi : $A_{jk} - \underline{A}_{jk}^{(l)} = \eta_{jk} + \underline{A}_{jk} - \underline{A}_{jk}^{(l)} \equiv \eta_{jl} + \underline{A}_{jk} - \underline{A}_{jk}^* = A_{jl} - \underline{A}_{jl} - \lambda^* = A_{jl} - \underline{A}_{jl}^* = A_{jl} - \underline{A}'_{jl}$. Donc : $S^{(l)}(I_1) \equiv S'((I_1 \setminus \{k\}) \cup \{l\})$. Comme auparavant, puisque $|I_1| = |(I_1 \setminus \{k\}) \cup \{l\}|$, on a : $S'((I_1 \setminus \{k\}) \cup \{l\}) \equiv S'(I_1)$.

On a donc prouvé que pour tout $I_1 \subseteq I_u(j)$ non-vidé : $S^{(l)}(I_1) \equiv S'(I_1)$. Il en résulte que, comme $\underline{A}'_j \in cl(\mathcal{B}_j)$, on a aussi : $\underline{A}_j^{(l)} \in cl(\mathcal{B}_j)$ (cf. propriété (6.1)). Ceci est vrai pour tout $l \in I_u(j) \setminus \{k\}$. Considérons alors la combinaison convexe suivante des scénarios correspondants : $\underline{A}_j'' = 1/n_j \cdot (\underline{A}'_j + \sum_{l \in I_u(j) \setminus \{k\}} \underline{A}_j^{(l)})$, avec $n_j = |I_u(j)|$. On a :

$$\underline{A}_j'' = \underline{A}_j^* + \frac{\underline{A}'_{jk} - \underline{A}_{jk}^*}{n_j} \cdot v.$$

Donc : $\underline{A}_j'' = \underline{A}_j + \lambda'' \cdot v$, avec $\lambda'' > \lambda^*$. Ceci est en contradiction avec la définition de λ^* . \square

Dans nos tests numériques, on a observé que même quand les $\{\eta_{ji}\}_{i \in I_u(j)}$ ne sont pas i.i.d., la direction v proposée dans le lemme ci-dessus est efficace si l'on travaille avec l'approximation (6.8). En effet, utiliser des directions avec des composantes moins "équilibrées" pénalise beaucoup $\lambda(1)$, qui a alors tendance à être la valeur minimale recherchée : $\min_k \lambda(k) = \lambda(1)$.

6.3.3 Illustration sur un exemple simple

Dans cette partie, on s'intéresse à l'ensemble $X = \left\{ x \in \{0, 1\}^n \mid \mathbb{P}(a_1 x_1 + a_2 x_2 \leq 2) \geq 4/5 \right\}$, où a_1 et a_2 sont deux variables aléatoires indépendantes et uniformément distribuées respectivement sur $[0, 1]$ et sur $[1, 2]$.

Le scénario basique le plus évident pour ce problème est $\underline{a} = (0, 1)$, composé des valeurs minimales de chaque variable aléatoire. L'inégalité basique correspondante est : $0 \cdot x_1 + 1 \cdot x_2 = x_2 \leq 2$. Noter qu'avec ce scénario basique \underline{a} , $\eta_1 = a_1 - \underline{a}_1$ et $\eta_2 = a_2 - \underline{a}_2$ sont i.i.d.. On énumère les inégalités (6.4) pour X :

- si $I_1 = \{1\}$: on a : $\mathbb{P}(\eta_1 \leq d) < 4/5 \Leftrightarrow d < 4/5$. En conséquence : $D(I_1) = 4/5$. L'inégalité valide correspondante est alors : $(0 + 4/5) \cdot x_1 + (1 - 4/5) \cdot x_2 \leq 2$, *i.e.* : $4x_1 + x_2 \leq 10$.
- Si $I_1 = \{2\}$: on a : $\mathbb{P}(\eta_2 \leq d) < 4/5 \Leftrightarrow d < 4/5$. Comme auparavant, on dérive l'inégalité valide correspondante : $(0 - 4/5) \cdot x_1 + (1 + 4/5) \cdot x_2 \leq 2$, *i.e.* : $-4x_1 + 9x_2 \leq 10$.
- Si $I_1 = \{1, 2\}$: on a, après calculs (voir par exemple l'annexe B.1) : $\mathbb{P}(\eta_1 + \eta_2 \leq d) < 4/5 \Leftrightarrow d < 2 - \sqrt{2/5}$. Alors : $D(I_1) = 2 - \sqrt{2/5} \geq 1,36$. L'inégalité valide correspondante est : $(0 + D(I_1)) \cdot x_1 + (1 + D(I_1)) \cdot x_2 \leq 2 + (2 - 1) \cdot D(I_1)$, *i.e.* : $1,36x_1 + 2,36x_2 \leq 3,36$.

On remarque que les deux premières inégalités sont impliquées par les contraintes de domaine $x_{1,2} \leq 1$: elles sont inutiles en pratique.

Impact du scénario basique : Comme les variables aléatoires sont indépendantes et symétriquement distribuées, d'après le lemme 41, on peut définir le scénario basique comme : $\underline{a}' = (1/2, 3/2)$. L'inégalité basique correspondante est : $1/2 \cdot x_1 + 3/2 \cdot x_2 \leq 2$ (ou de manière équivalente : $x_1 + 3x_2 \leq 4$). Comme précédemment, on donne les inégalités (6.4) :

- si $I_1 = \{1\}$: on a : $\mathbb{P}(\eta_1 \leq d) < 4/5 \Leftrightarrow d < 4/5 - 1/2 = 3/10$. En conséquence : $D(I_1) = 3/10$, et l'inégalité valide correspondante est : $4/5.x_1 + 6/5.x_2 \leq 2$, *i.e.* : $4x_1 + 6x_2 \leq 10$.
- Si $I_1 = \{2\}$: de la même manière : $\mathbb{P}(\eta_2 \leq d) < 4/5 \Leftrightarrow d < 3/10$. L'inégalité valide correspondante est : $1/5.x_1 + 9/5.x_2 \leq 2$, *i.e.* : $x_1 + 9x_2 \leq 10$.
- Si $I_1 = \{1, 2\}$: après calculs, on a : $\mathbb{P}(\eta_1 + \eta_2 \leq d) < 4/5 \Leftrightarrow d < 1 - \sqrt{2/5}$. Alors : $D(I_1) = 1 - \sqrt{2/5} \geq 0,36$, et on obtient : $0,86.x_1 + 1,86.x_2 \leq 2,36$.

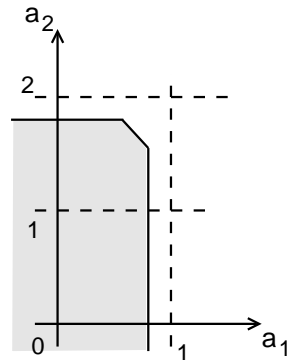
Comme prévu, les inégalités obtenues sont plus fortes que celles dérivées du scénario basique \underline{a} ; en effet, $\underline{a}' \geq \underline{a}$.

Description de l'ensemble des scénarios basiques : On donne une description complète des scénarios basiques pour ce problème. Observer que :

$$\begin{cases} \mathbb{P}(a_1 \leq \alpha) < 4/5 \Leftrightarrow \alpha < 4/5, \\ \mathbb{P}(a_2 \leq \alpha) < 4/5 \Leftrightarrow \alpha < 9/5, \\ \mathbb{P}(a_1 + a_2 \leq \alpha) < 4/5 \Leftrightarrow \alpha < 3 - \sqrt{2/5}. \end{cases}$$

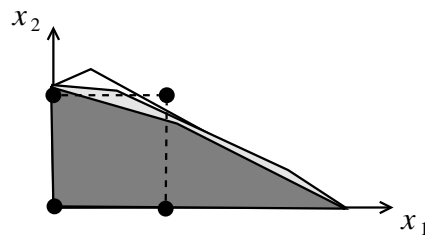
Alors l'ensemble des scénarios basiques est :

$$cl(\mathcal{B}) = \left\{ \tilde{a} \in \mathbb{R}^2 \mid \tilde{a}_1 \leq 4/5, \tilde{a}_2 \leq 9/5 \text{ et } \tilde{a}_1 + \tilde{a}_2 \leq 3 - \sqrt{2/5} \right\}.$$



Finalement, on utilise le lemme 43 pour améliorer le scénario basique \underline{a} dans la direction $v = (1, 1)$ (cf. lemme 45). Après des calculs faciles, on obtient : $\lambda^* = 1 - 1/\sqrt{10}$, et le nouveau scénario basique est donc : $\underline{a}'' = (1 - 1/\sqrt{10}, 2 - 1/\sqrt{10})$.

Les trois relaxations linéaires obtenues respectivement à partir des scénarios basiques \underline{a} , \underline{a}' et \underline{a}'' sont représentés sur la figure suivante.



6.4 Algorithme de résolution

Considérons le problème d'optimisation suivant, appelé CCILP (pour *Chance-Constrained Integer Linear Problem*) :

$$\max \left\{ cx \mid x \in X \right\},$$

où $c \in \mathbb{R}^n$. La partie précédente a fourni une relaxation linéaire de X qui, en théorie, peut être utilisée pour mettre en œuvre un algorithme de *branch-and-bound*. La difficulté vient du nombre exponentiel des inégalités (6.4). Elles ne peuvent pas toutes être présentes dans la formulation du problème et doivent être générées dynamiquement au cours du processus de résolution.

6.4.1 Séparation des inégalités (6.4)

Dans cette partie, soit $\tilde{x} \in [0, 1]^n$ (fractionnaire). Soit $j \in J$, pour tout sous-ensemble $I_1 \subseteq I_u(j)$ non-vide, $I_0 = I_u(j) \setminus I_1$, on note :

$$\begin{aligned} \Delta_j(I_1) = & b_j + \left(|I_1| - 1 \right) \cdot D_j(I_1) \\ & - \left\{ \sum_{i \in I_1} [A_{ji} + D_j(I_1)] \tilde{x}_i + \sum_{i \in I_0} [A_{ji} - D_j(I_1)] \tilde{x}_i + \sum_{i \notin I_u(j)} A_{ji} \tilde{x}_i \right\}. \end{aligned} \quad (6.9)$$

$\Delta_j(I_1)$ est la valeur d'écart de l'inégalité (6.4) associée à j et I_1 pour \tilde{x} . Le but est de trouver I_1 correspondant à la plus petite valeur d'écart possible, donc à l'inégalité la plus violée pour \tilde{x} . En général, ceci ne peut pas être fait directement à partir de la formule ci-dessus, puisqu'une énumération complète des sous-ensembles n'est pas envisageable. On va donc tâcher de contourner cette difficulté.

Il faut d'abord se souvenir qu'une inégalité associée à I_1 est directement liée au point $x(I_1)$, le vecteur caractéristique de I_1 (voir la preuve du lemme 37). Ainsi, il est naturel de chercher I_1 tel que $x(I_1)$ est "proche" de \tilde{x} . On en déduit que les indices $i \in I_u(j)$ tels que $\tilde{x}_i = 1$ devraient apparaître dans I_1 .

Décrivons maintenant une manière de prendre en compte les composantes fractionnaires de \tilde{x} . On propose d'étendre I_1 de manière itérative. Considérons $I_1 \subset I_u(j)$ fixé, on définit $I'_1 = I_1 \cup \{k\}$ pour un $k \in I_u(j) \setminus I_1$. On a :

$$\begin{aligned} \Delta_j(I'_1) - \Delta_j(I_1) &= (|I_1| - 1) \cdot [D_j(I'_1) - D_j(I_1)] + D_j(I'_1) \\ &\quad - [D_j(I'_1) - D_j(I_1)] \cdot \left(\sum_{i \in I_1} \tilde{x}_i - \sum_{i \in I_0 \setminus \{k\}} \tilde{x}_i \right) - [D_j(I'_1) + D_j(I_1)] \cdot \tilde{x}_k \\ &= [D_j(I'_1) - D_j(I_1)] \cdot \left(|I_1| - 1 - \sum_{i \in I_1} \tilde{x}_i + \sum_{i \in I_0 \setminus \{k\}} \tilde{x}_i \right) \\ &\quad + D_j(I'_1) - [D_j(I'_1) + D_j(I_1)] \cdot \tilde{x}_k. \end{aligned}$$

Lemme 46 On a :

- (i) si $D_j(I'_1) \geq D_j(I_1)$: $\Delta_j(I'_1) - \Delta_j(I_1) \geq D_j(I_1) - [D_j(I'_1) + D_j(I_1)] \cdot \tilde{x}_k$,
- (ii) si $D_j(I'_1) \leq D_j(I_1)$: $\Delta_j(I'_1) - \Delta_j(I_1) \leq D_j(I_1) - [D_j(I'_1) + D_j(I_1)] \cdot \tilde{x}_k$,
- (iii) si : $\forall i \in I_1, \tilde{x}_i = 1$, et : $\forall i \in I_0 \setminus \{k\}, \tilde{x}_i = 0$, alors : $\Delta_j(I'_1) - \Delta_j(I_1) = D_j(I_1) - [D_j(I'_1) + D_j(I_1)] \cdot \tilde{x}_k$.

Preuve : (i) et (ii) viennent du fait que : $|I_1| - 1 - \sum_{i \in I_1} \tilde{x}_i + \sum_{i \in I_0 \setminus \{k\}} \tilde{x}_i \geq -1$. (iii) est évident. \square

Comme $D_j \geq 0$ (cf. lemme 35), les points (i) et (ii) du lemme ci-dessus soulignent l'intérêt qu'il y a à considérer d'abord les composantes \tilde{x}_k "assez grandes". D'une part, dans le cas (i),

prendre \tilde{x}_k “trop petit” implique que $\Delta_j(I'_1) \geq \Delta_j(I_1)$: inclure k dans I_1 n’améliore pas la coupe. D’autre part, dans le cas (ii), plus \tilde{x}_k est grand, plus la valeur d’écart est améliorée. En s’appuyant sur ces observations, on propose la construction gloutonne suivante de I_1 :

Construction gloutonne de I_1

- Étape 1 :** Trier les valeurs $\{\tilde{x}_i\}_{i \in I_u(j)}$ en ordre décroissant.
Soit $\varphi(r)$ l’indice du r^e élément de la liste triée.
- Étape 2 :** Soient $I_1 = \{i \in I_u(j) | \tilde{x}_i = 1\}$ et $r = |I_1| + 1$.
- Étape 3 :** Poser $I'_1 = I_1 \cup \{\varphi(r)\}$, et calculer $D_j(I'_1)$.
- Étape 4 :** Calculer $\Delta_j(I'_1) - \Delta_j(I_1)$.
Si $\Delta_j(I'_1) < \Delta_j(I_1)$, poser $I_1 = I'_1$.
- Étape 5 :** Si $r = |I_u(j)|$, STOP.
Mettre à jour $r \leftarrow r + 1$ et aller à l’étape 3.
-

Cet algorithme construit I_1 de manière itérative en considérant d’abord les indices correspondant aux valeurs les plus grandes de $\{\tilde{x}_i\}_{i \in I_u(j)}$. Quand un indice améliore la valeur d’écart, il est ajouté dans I_1 . L’ensemble I_1 finalement obtenu correspond à une inégalité (6.4).

Lemme 47 *Soit $j \in J$. Si $\tilde{x}_i \in \{0, 1\}$ pour tout $i \in I_u(j)$, alors la construction gloutonne de I_1 conduit à l’inégalité (6.4) la plus violée.*

Preuve : On rappelle que : $\Delta_j(I_1) = b_j - \underline{A}_j \tilde{x} + (|I_1| - 1 - \sum_{i \in I_1} \tilde{x}_i + \sum_{i \in I_0} \tilde{x}_i) \cdot D_j(I_1)$. L’inégalité la plus violée est obtenue avec $I_1 = \{i \in I_u(j) | \tilde{x}_i = 1\}$ (voir par exemple la preuve du lemme 37). Ceci correspond à l’ensemble initial construit à l’étape 2. Cet ensemble ne sera pas étendu après-coup, puisque la coupe correspondante est la plus violée. (Pour le voir, on peut aussi considérer $k \in I_0$: $\tilde{x}_k = 0$ et $\Delta_j(I'_1) - \Delta_j(I_1) = D_j(I_1) \geq 0$, avec $I'_1 = I_1 \cup \{k\}$.) \square

Cette observation signifie que, même si le processus de séparation décrit ne génère pas l’inégalité la plus violée en général, c’est le cas quand on a un point entier. En conséquence, l’utilisation de cet algorithme glouton dans un algorithme de *branch-and-cut* assure que la solution finale retenue est bien optimale. En effet, toute solution non-réalisable sera coupée.

Séparation exacte dans un cas spécial : Dans ce paragraphe, on considère le cas particulier où les $\{\eta_{ji}\}_{i \in I_u(j)}$ sont i.i.d.. On rappelle que sous cette hypothèse, $D_j(I_1)$ dépend seulement du cardinal $|I_1|$ (cf. lemme 36). Alors, pour chaque $k \leq |I_u(j)| = n_j$, résoudre le problème suivant est facile :

$$\min \{ \Delta_j(I_1) \mid I_1 \subseteq I_u(j), |I_1| = k \}.$$

Dans ce problème, puisque le cardinal de I_1 est fixé à la valeur k , $D_j(I_1)$ est une constante. Donc la solution I_1^* optimale est constituée des k indices correspondant aux plus grandes valeurs de $\{\tilde{x}_i\}_{i \in I_u(j)}$. En effet, d’après l’équation (6.9) : $\Delta_j(I_1) = b_j - \underline{A}_j \cdot \tilde{x} + (k - 1 - \sum_{i \in I_1} \tilde{x}_i + \sum_{i \in I_0} \tilde{x}_i) \cdot D_j(I_1)$. Obtenir l’inégalité la plus violée s’effectue alors simplement par énumération des valeurs de cardinal possibles, de $k = 1$ à $k = n_j$.

D’un point de vue complexité, cette séparation est réalisable en temps $O(n_j \log(n_j))$ (complexité du tri des $\{\tilde{x}_i\}_{i \in I_u(j)}$). Noter que les n_j valeurs de D_j peuvent être calculées une fois pour toutes au début. Des algorithmes relatifs seront décrits dans la partie suivante.

Enfin, ce cas particulier valide l’orientation prise avec l’heuristique gloutonne pour le cas général. En effet, le rôle joué par les composantes les plus grandes de \tilde{x} est bien mis en évidence ici.

6.4.2 Calcul de $D_j(I_1)$

Le processus de séparation décrit plus haut nécessite le calcul de $D_j(I_1)$ (cf. étape 4 de l'algorithme glouton). On souhaite pouvoir effectuer une recherche efficace de cette valeur par dichotomie. Ceci requiert des bornes inférieure et supérieure sur $D_j(I_1)$. Une borne inférieure est déjà donnée par le lemme 35, où l'on a montré que $D_j(I_1) > 0$. On donne maintenant quelques résultats permettant de définir une majoration de $D_j(I_1)$.

On a défini auparavant un scénario basique, qui peut être vu comme un scénario "minorant". De manière symétrique, on introduit un scénario "majorant", noté \bar{A} , et qui satisfait, pour $j \in J$:

$$\forall I_1 \subseteq I_u(j), I_1 \neq \emptyset : \mathbb{P} \left(\sum_{i \in I_1} (A_{ji} - \bar{A}_{ji}) \leq 0 \right) \geq 1 - \varepsilon_j. \quad (6.10)$$

Comme pour le scénario basique, quelques règles simples permettent de construire un tel scénario \bar{A} :

Lemme 48 Soit $j \in J$, on note : $n_j = |I_u(j)|$. Soit \bar{A}_j défini comme suit :

(i) si aucune hypothèse probabiliste particulière n'est effectuée, alors pour tout $i \in I_u(j)$: $\mathbb{P}(A_{ji} \leq \bar{A}_{ji}) \geq 1 - \varepsilon_j/n_j$;

(ii) si les variables aléatoires $\{A_{ji}\}_{i \in I_u(j)}$ sont indépendantes, alors pour tout $i \in I_u(j)$: $\mathbb{P}(A_{ji} \leq \bar{A}_{ji}) \geq (1 - \varepsilon_j)^{1/n_j}$.

Alors \bar{A}_j satisfait (6.10).

Preuve : Soit $I_1 \subseteq I_u(j)$. On a :

$$\begin{aligned} \mathbb{P} \left(\sum_{i \in I_1} (A_{ji} - \bar{A}_{ji}) \leq 0 \right) &\geq \mathbb{P} \left(\forall i \in I_u(j), A_{ji} \leq \bar{A}_{ji} \right) \\ &= 1 - \mathbb{P} \left(\exists i \in I_u(j) \text{ t.q. } A_{ji} > \bar{A}_{ji} \right). \end{aligned}$$

Si aucune hypothèse n'est effectuée, on a toujours : $\mathbb{P}(\exists i \in I_u(j) \text{ t.q. } A_{ji} > \bar{A}_{ji}) \leq \sum_{i \in I_u(j)} \mathbb{P}(A_{ji} > \bar{A}_{ji})$. En supposant (i), on a : $\mathbb{P}(\sum_{i \in I_1} (A_{ji} - \bar{A}_{ji}) \leq 0) \geq 1 - n_j \cdot \varepsilon_j/n_j = 1 - \varepsilon_j$.

Supposons que les variables aléatoires sont indépendantes. Alors : $\mathbb{P}(\forall i \in I_u(j), A_{ji} \leq \bar{A}_{ji}) = \prod_{i \in I_u(j)} \mathbb{P}(A_{ji} \leq \bar{A}_{ji})$. On déduit de (ii) : $\mathbb{P}(\sum_{i \in I_1} (A_{ji} - \bar{A}_{ji}) \leq 0) \geq 1 - \varepsilon_j$. \square

On peut maintenant énoncer le résultat suivant, qui définit une borne supérieure pour $D_j(I_1)$:

Lemme 49 Soient $j \in J$ et $I_1 \subseteq I_u(j)$ non-vide : $D_j(I_1) \leq \sum_{i \in I_1} (\bar{A}_{ji} - \underline{A}_{ji})$.

Preuve : Observer que : $\mathbb{P} \left(\sum_{i \in I_1} \eta_{ji} \leq \sum_{i \in I_1} (\bar{A}_{ji} - \underline{A}_{ji}) \right) = \mathbb{P} \left(\sum_{i \in I_1} A_{ji} \leq \sum_{i \in I_1} \bar{A}_{ji} \right) \geq 1 - \varepsilon_j$, par définition de \bar{A}_{ji} . Ceci prouve que $D_j(I_1) \leq \sum_{i \in I_1} (\bar{A}_{ji} - \underline{A}_{ji})$. \square

On peut alors définir un algorithme de recherche de $D_j(I_1)$ par dichotomie :

Algorithme de calcul de $D_j(I_1)$

Étape 0 : Poser $D_{min} = 0$ et $D_{max} = \sum_{i \in I_1} (\overline{A_{ji}} - \underline{A_{ji}})$.
Soit $\delta > 0$ la précision voulue sur le résultat.

Étape 1 : Si $D_{max} - D_{min} < \delta$, poser $D = D_{min}$ et STOP.

Étape 2 : Poser $D = (D_{min} + D_{max})/2$, et calculer la probabilité :
 $\Psi = \mathbb{P}(\sum_{i \in I_1} \eta_{ji} \leq D)$.

Étape 3 : Si $\Psi < 1 - \varepsilon_j$, poser $D_{min} = D$ et aller à l'étape 1.
Sinon, poser $D_{max} = D$ et aller à l'étape 1.

La quantité D finale correspond au terme voulu, avec une précision $\delta > 0$.

Pré-calcul des termes $D_j(I_1)$: On a déjà dit que lorsque les $\{\eta_{ji}\}_{i \in I_u(j)}$ sont i.i.d., les termes $D_j(I_1)$ dépendent uniquement de la cardinalité $n_1 = |I_1|$ (cf. lemme 36). En conséquence, ces coefficients sont peu nombreux et peuvent facilement être tous pré-calculés. Cette démarche peut s'étendre au cas où les $\{\eta_{ji}\}_{i \in I_u(j)}$ sont indépendants et distribués selon un petit nombre de distributions différentes. Cette intuition est détaillée maintenant.

Supposons que les $\{\eta_{ji}\}_{i \in I_u(j)}$ sont tous indépendants, et qu'il existe une partition $\{U_1, \dots, U_K\}$ de $I_u(j)$ telle que, pour tout $k \in \{1, \dots, K\}$, les $\{\eta_{ji}\}_{i \in U_k}$ sont identiquement distribués.

Pour comprendre notre démarche, imaginons que l'on considère deux sous-ensembles non-vides de $I_u(j)$, I_1 et I'_1 , de même cardinal $c : |I_1| = |I'_1| = c$. La question est de savoir si $D_j(I_1) = D_j(I'_1)$, et donc s'il est vraiment nécessaire de calculer ces deux termes de protection, ou si le calcul d'un seul suffit. Il est clair que si, pour chaque classe $k \in \{1, \dots, K\}$, les ensembles I_1 et I'_1 impliquent le même nombre d'éléments de U_k , alors $D_j(I_1) = D_j(I'_1)$. Il s'agit donc de compter le nombre d'agencements différents de c éléments choisis dans les classes $\{U_1, \dots, U_K\}$; on note $N_K(c)$ ce nombre. Il faudra donc calculer $N_K(c)$ termes de protection, pour chaque cardinal c possible, et pas plus.

Plus formellement, $N_K(c)$ est défini comme le nombre de solutions (u_1, \dots, u_K) du problème suivant :

$$\begin{cases} \sum_{k=1}^K u_k = c, \\ u_k \leq |U_k|, \forall k \in \{1, \dots, K\}, \\ u_k \in \mathbb{N}, \forall k \in \{1, \dots, K\}. \end{cases}$$

Observer que l'on a :

- si $c > \sum_{k=1}^K |U_k|$ ou $c < 0 : N_K(c) = 0$;
- $N_K(0) = 1$ et $N_K(1) = K$;
- $N_K(|I_u(j)|) = 1$.

Lemme 50 Pour tout $c \in \mathbb{Z} : N_1(c) = \mathbb{1}_{0 \leq c \leq |U_1|}$. Si $K \geq 2$, pour tout $c \in \mathbb{Z} :$

$$N_K(c) = \sum_{t=0}^{|U_K|} N_{K-1}(c-t).$$

Preuve : Il suffit de voir que si l'on considère t éléments de l'ensemble U_K ($u_K = t$), alors il y reste $N_{K-1}(c-t)$ solutions différentes possibles. \square

$N_K(c)$ peut donc être calculé par récurrence. Il y a en définitive un total de $\sum_{c=1}^{|I_u(j)|} N_K(c)$ termes de protection différents à pré-calculer pour le problème.

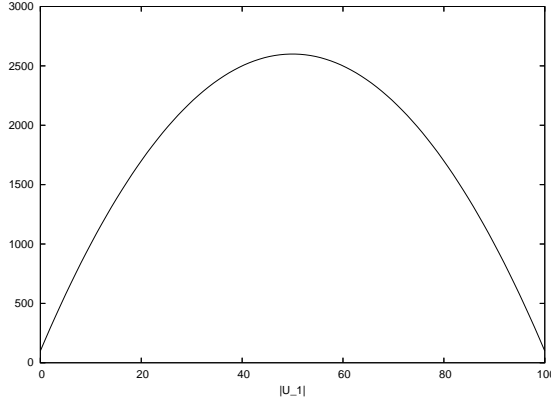


FIG. 6.1 – Nombre de termes de protection à pré-calculer, quand $|I_u(j)| = 100$ et $K = 2$, en fonction de $|U_1|$.

À titre d'illustration, considérer le cas où $K = 2$. Pour tout $c \in \mathbf{Z}$, $c \leq |I_u(j)|$:

$$N_2(c) = \sum_{t=0}^{|U_2|} N_1(c-t) = \sum_{t=0}^{|U_2|} \mathbb{1}_{0 \leq c-t \leq |U_1|} = \sum_{t=(c-|U_1|)^+}^{\min\{c, |U_2|\}} 1 = \min\{c, |U_2|\} - (c - |U_1|)^+ + 1 .$$

Supposons que $|U_1| \leq |U_2|$. Alors le nombre total de termes de protection à pré-calculer est :

$$\begin{aligned} \sum_{c=1}^{|I_u(j)|} N_2(c) &= \sum_{c=1}^{|U_1|} (c+1) + \sum_{c=|U_1|+1}^{|U_2|} (|U_1|+1) + \sum_{c=|U_2|+1}^{|I_u(j)|} (|I_u(j)| - c + 1) \\ &= |U_1| \cdot (|U_1| + 3)/2 + (|U_1| + 1) \cdot (|U_2| - |U_1|) + |U_1| \cdot (|U_1| + 1)/2 . \end{aligned}$$

Ce nombre est représenté sur la figure 6.1 pour $|I_u(j)| = 100$.

6.4.3 Calcul des probabilités

Le calcul de probabilités est un élément algorithmique central. Il est requis par exemple pour tester si une solution x est réalisable ou non (*i.e.* pour vérifier si $\mathbb{P}(A_j x \leq b_j) \geq 1 - \varepsilon_j$). Le calcul de probabilités est nécessaire également pour calculer les coefficients $D_j(I_1)$: l'algorithme du paragraphe 6.4.2 montre que l'on doit être capable d'évaluer $\mathbb{P}(\sum_{i \in I_1} \eta_{ji} \leq D)$ pour un D donné (voir l'algorithme de calcul de $D_j(I_1)$). Plusieurs approches peuvent être envisagées, dont certaines ont déjà été présentées dans la partie 5.4.1, à savoir l'utilisation de bornes analytiques, les méthode d'échantillonnage (Monte-Carlo) et les approximations gaussiennes (utilisation du théorème de la limite centrale).

$D_j(I_1)$ est défini à partir d'une somme de variables aléatoires : $\sum_{i \in I_1} \eta_{ji}$. Dans certains cas particuliers, la fonction de répartition d'une telle somme est connue de manière analytique. C'est le cas, par exemple, si les $\{\eta_{ji}\}_{i \in I_u(j)}$ sont indépendantes et uniformément distribuées sur $[0,1]$. En notant $n_1 = |I_1|$, des résultats classiques de calcul de probabilités (voir e.g. [36]) permettent d'obtenir :

$$\begin{aligned} \mathbb{P}\left(\sum_{i=1}^{n_1} \eta_{ji} \leq d\right) &= \frac{1}{n_1!} \sum_{k=0}^{\lfloor d \rfloor - 1} \sum_{l=0}^k (-1)^l \binom{n_1}{l} \left[(k+1-l)^{n_1} - (k-l)^{n_1} \right] \\ &\quad + \frac{1}{n_1!} \sum_{l=0}^{\lfloor d \rfloor} (-1)^l \binom{n_1}{l} \left[(d-l)^{n_1} - (\lfloor d \rfloor - l)^{n_1} \right]. \end{aligned} \tag{6.11}$$

Ce résultat peut être utilisé comme base pour obtenir des caractérisations exactes pour de nombreuses distributions différentes. Évidemment, la formule s'adapte directement dans le cas où les variables aléatoires sont indépendantes et uniformément distribuées sur un intervalle quelconque $[l, u]$, $l < u$. De plus, (6.11) permet la description exacte de toute distribution qui peut s'obtenir comme la somme de variables uniformément distribuées sur un intervalle. Par exemple, supposons que les $\{\eta_{ji}\}_{i \in I_u(j)}$ sont indépendantes et suivent toutes une distribution triangulaire symétrique sur $[0, 2]$. Observer que chaque η_{ji} est en fait la somme de deux variables aléatoires indépendantes et uniformément distribuées sur $[0, 1]$. Ainsi, la fonction de répartition de $\sum_{i \in I_1} \eta_{ji}$ s'obtient directement à partir de (6.11) en remplaçant n_1 par $2n_1$.

D'autres résultats analytiques sur les sommes de variables aléatoires existent et peuvent être utilisés dans notre contexte. Par exemple, on renvoie à [18] pour la fonction de répartition d'une somme de variables aléatoires indépendantes, uniformément distribuées, mais non-identiques. En pratique, ces formules exactes sont très difficiles à traiter numériquement quand n_1 est grand (problèmes de précision et d'instabilités numériques). Quand de tels problèmes surviennent, le recours au théorème de la limite centrale peut fournir des approximations très précises. Ce résultat classique assure que, lorsque les $\{\eta_{ji}\}_{i \in I_u(j)}$ sont i.i.d. (voir annexe B.2) :

$$\lim_{n_1 \rightarrow \infty} \mathbb{P} \left(\frac{\sum_{i=1}^{n_1} \eta_{ji} - n_1 \mu}{\sigma \sqrt{n_1}} \leq d \right) = \Phi(d),$$

où μ et σ sont respectivement la moyenne et l'écart-type de η_{ji} , et Φ désigne la fonction de répartition de la loi normale standard $\mathcal{N}(0, 1)$. Ceci signifie que la fonction de répartition de $\mathcal{N}(n_1 \mu, n_1 \sigma^2)$ fournit une approximation de celle de $\sum_{i=1}^{n_1} \eta_{ji}$. On rappelle que des généralisations du théorème de la limite centrale existent lorsque les $\{\eta_{ji}\}_{i \in I_u(j)}$ ne sont pas identiquement distribués, voir annexe B.2.

6.4.4 Contraintes de probabilités jointes

Comme montré dans la partie 6.1.2, X est une relaxation de X' si $\varepsilon_j = \varepsilon$ pour tout $j \in J$. On peut donc utiliser l'algorithme de *branch-and-cut* décrit auparavant pour résoudre : $\max \{cx \mid \mathbb{P}(Ax \leq b) \geq 1 - \varepsilon\}$. Il s'agit simplement de retirer les points de $X \setminus X'$ de l'ensemble des solutions. Plus précisément, lorsqu'au cours de l'algorithme on obtient une solution \bar{x} entière, il faut s'assurer qu'elle appartient à X . Ceci revient à calculer la probabilité $\mathbb{P}(A\bar{x} \leq b)$: si cette quantité est supérieure à $1 - \varepsilon$, \bar{x} est une solution admissible. Dans le cas contraire, il faut retirer \bar{x} de l'ensemble des solutions. Ceci peut être fait très simplement en ignorant cette solution dans l'algorithme de résolution. Une autre manière de faire est de séparer \bar{x} de l'ensemble X par une inégalité : en effet, une solution $x \in \{0, 1\}^n$ est différente de \bar{x} si, et seulement si : $\sum_{i \in I} (x_i - \bar{x}_i)^2 \geq 1$. Puisque toutes les composantes sont 0-1, ceci s'écrit de manière équivalente : $\sum_{i \in I} (x_i + \bar{x}_i - 2\bar{x}_i x_i) \geq 1$. L'ajout de cette inégalité dans le modèle assure que \bar{x} est retiré de l'ensemble des solutions.

6.5 Tests numériques

L'algorithme de *branch-and-cut* présenté plus haut a été codé en C++ en utilisant Cplex 10.0 (*i.e.*, on s'appuie sur l'algorithme de *branch-and-bound* de Cplex, et on ajoute des inégalités aux noeuds de branchement). Les procédures de générations automatiques de coupes de Cplex n'ont pas été utilisées. Les tests présentés dans cette partie sont réalisés sur un ordinateur équipé d'un processeur Intel(R) Xeon(TM) 2,8 GHz, avec 2 Go de RAM. Les temps indiqués dans les tableaux sont toujours exprimés en secondes. La durée limite d'un test est de 15 minutes

(900 secondes); dans les tableaux, le sigle “t.l.” indique que cette limite a été atteinte, et que l’algorithme s’est donc arrêté sans avoir trouvé l’optimum.

Par souci de simplicité, les expériences présentées se concentrent sur des cas où les variables aléatoires η_{ji} sont i.i.d.. On rappelle néanmoins que des algorithmes similaires peuvent être utilisés dans des contextes beaucoup plus généraux.

6.5.1 Problèmes de sac-à-dos multi-dimensionnels

Des tests numériques sont effectués sur des instances du problème de sac-à-dos multi-dimensionnel :

$$\max \left\{ \sum_{i \in I} p_i x_i \mid \mathbb{P} \left(\sum_{i \in I} w_{ji} x_i \leq c_j \right) \geq 1 - \varepsilon_j, \forall j \in J \right\}.$$

Pour tout $j \in J$, $c_j > 0$ est la capacité du sac j . Pour chaque $i \in I$, $p_i > 0$ est le profit lié au fait de sélectionner l’élément i . Le poids w_{ji} est une variable aléatoire uniformément distribuée sur un intervalle $[\underline{w}_{ji}, \overline{w}_{ji}]$, avec $\underline{w}_{ji} > 0$ et $\overline{w}_{ji} = \underline{w}_{ji} + \delta$ ($\delta > 0$). Avec cette dernière hypothèse, les variables aléatoires $\{\eta_{ji}\}_{(i,j) \in I \times J}$ seront i.i.d.. Dans ce contexte, on a montré dans la partie 6.4.1 que la séparation des inégalités (6.4) peut être réalisée efficacement.

On pose $\varepsilon_j = 0,1$ pour tout $j \in J$. On cherche ainsi la meilleure solution qui satisfait chaque contrainte de sac-à-dos avec une probabilité au moins 0,9. Les données p_i et \underline{w}_{ji} sont choisies aléatoirement dans $[100, 1000]$, et $\delta = 20$. c_j est choisi aléatoirement dans l’intervalle $[1/3 \cdot \sum_{i \in I} \underline{w}_{ji}, 2/3 \cdot \sum_{i \in I} \underline{w}_{ji}]$.

Comme cela a été expliqué dans la partie 6.3, le scénario basique choisi peut avoir en théorie un impact sur l’efficacité de l’algorithme de résolution. Trois scénarios basiques sont testés :

- B1 : le premier est constitué des valeurs minimales de chaque poids : $\{\underline{w}_{ji}\}_{(i,j) \in I \times J}$;
- B2 : en s’appuyant sur le point (iii) du lemme 41, le deuxième scénario basique est construit à l’aide des valeurs moyennes $\{(\underline{w}_{ji} + \overline{w}_{ji})/2\}_{(i,j) \in I \times J}$;
- B3 : enfin, le dernier scénario basique testé correspond à une amélioration de B2 dans la direction $v = (1, \dots, 1)$, en s’appuyant sur (6.8) (cf. lemme 43). Noter que d’après le lemme 45, le scénario basique obtenu est non-dominé, puisque les $\{\eta_{ji}\}_{(i,j) \in I \times J}$ sont i.i.d..

Les termes de protection $D_j(I_1)$ sont pré-calculés puisqu’ils ne dépendent que du cardinal $|I_1|$ (cf. lemme 36). On utilise la formule exacte (6.11). Cependant, des instabilités numériques sont observées pour $n_1 = |I_1| \geq 70$. Pour pallier cette difficulté, et en s’inspirant du théorème de la limite centrale, on a utilisé la distribution normale $\mathcal{N}(n_1 \delta / 2, n_1 \delta^2 / 12)$ (cf. partie 6.4.3). Une comparaison numérique de valeurs obtenues *via* la formule analytique exacte et *via* l’approximation gaussienne (nommée TLC, pour “théorème de la limite centrale”) a été réalisée, voir la figure 6.2. L’erreur absolue de l’approximation gaussienne est faible (moins de 0,1 lorsque $n_1 \geq 50$). L’erreur relative de cette approximation converge très rapidement vers 0. On a donc utilisé l’approximation gaussienne pour calculer les termes de protection quand $n_1 \geq 50$. Les instabilités numériques observées sont illustrées sur la figure 6.3.

Plusieurs stratégies de génération de coupes ont été testées. On a observé qu’ajouter systématiquement toutes les inégalités violées pouvait détériorer le temps de calcul, même si le nombre de nœuds de branchement explorés est significativement réduit. Il a semblé plus efficace de n’ajouter à chaque nœud qu’au plus une inégalité violée par contrainte j du problème. La seule exception réside dans le nœud racine, pour lequel les inégalités violées sont ajoutées tant qu’on en trouve. Cette stratégie s’applique à chacun des tests présentés ici.

Les résultats sont donnés dans les tableaux 6.1-6.5 pour des problèmes de $n = 200$ éléments, avec $m = 1$ ou $m = 5$ dimensions. Lorsque le scénario naïf B1 est utilisé, aucune des instances n’est résolue à l’optimum dans le temps de résolution imparti (15 minutes), voir le tableau 6.1. De plus, l’écart prouvé avec la valeur optimale reste grand (1,46% en moyenne), et on observe

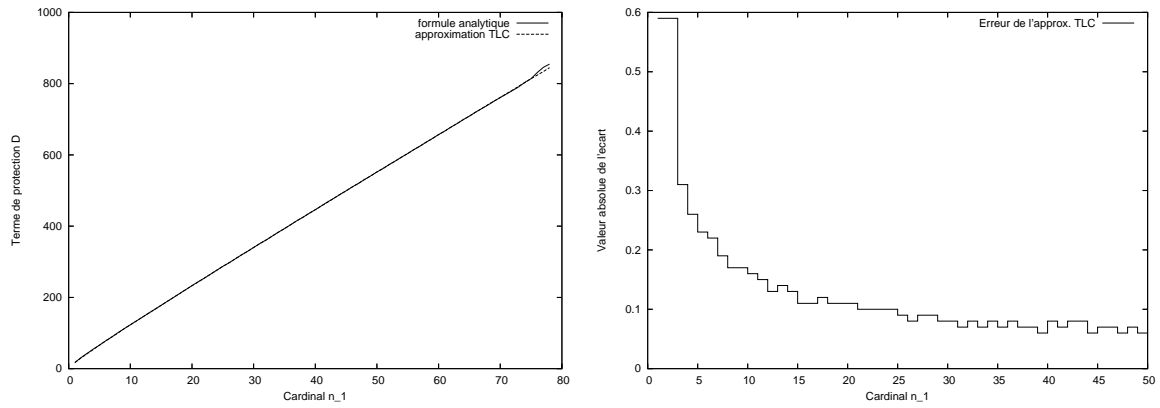


FIG. 6.2 – Distributions uniformes : comparaison des calculs exact et approché des termes $D_j(I_1)$, en fonction de $n_1 = |I_1|$.

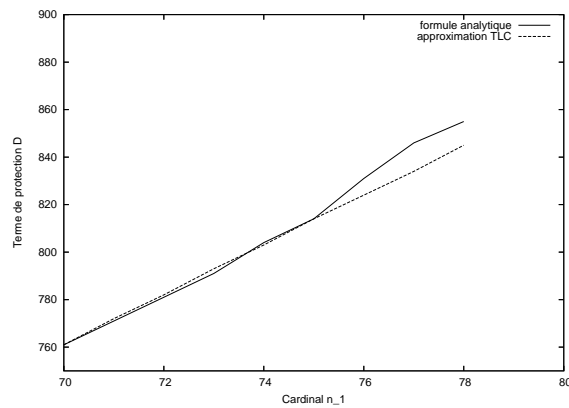


FIG. 6.3 – Distributions uniformes : illustration des instabilités numériques résultant de l'usage de la formule analytique exacte.

TAB. 6.1 – Résultats pour 10 instances de sac-à-dos, avec $n = 200$ ($m = 1$), en utilisant le scénario basique B1.

instance	nb de nœuds	nb d'inégalités	valeur	écart / optimum	temps (sec.)
0	131398	35719	89190	1,25%	t.l.
1	110874	29329	86879	1,22%	t.l.
2	103431	27887	89336	1,23%	t.l.
3	86143	23047	75059	1,59%	t.l.
4	83101	22596	73426	1,38%	t.l.
5	88401	22786	99069	1,29%	t.l.
6	84831	23108	92005	1,71%	t.l.
7	25431	8088	64992	2,17%	t.l.
8	68401	20507	70874	1,44%	t.l.
9	96051	24762	87241	1,29%	t.l.
moyenne	87806,2	23782,9	82807,1	1,46%	t.l.

TAB. 6.2 – Résultats pour 10 instances de sac-à-dos, avec $n = 200$ ($m = 1$), en utilisant le scénario basique B2.

instance	nb de nœuds	nb d'inégalités	valeur	écart / optimum	temps (sec.)
0	1521	125	89345	0	5,85
1	26228	1767	86879	0	91,5
2	6224	419	89383	0	22,75
3	22361	8816	75232	0,06%	t.l.
4	22700	8791	73446	0,07%	t.l.
5	10339	546	99313	0	40,18
6	610	74	92647	0	3,2
7	5592	2123	65483	0	75,71
8	20663	8643	70880	0,08%	t.l.
9	1189	117	87333	0	6,98
moyenne	11742,7	3142,1	82994,1	0,02%	307,14

en fait que dans la plupart des cas, la solution courante n'est pas optimale (comparer avec les résultats du tableau 6.3, où les mêmes instances sont résolues à l'optimum). Si l'on considère le scénario basique B2, pour $m = 1$, la résolution est grandement améliorée (voir tableau 6.2). 7 des 10 instances sont résolues à l'optimum en moins de 2 minutes. Cependant, 3 des instances ne sont pas complètement résolues dans le temps limite, même si l'écart avec la valeur optimale est faible. Finalement, la comparaison avec le tableau 6.5 montre la grande amélioration résultant de l'utilisation du scénario basique B3. Le temps maximal de résolution est ici ramené à moins de 3 secondes, et toutes les instances sont résolues à l'optimum. Lorsque $m = 5$, la comparaison des scénarios basiques B2 et B3 conduit aux mêmes observations, cf. tableaux 6.4 et 6.5. Les temps de résolution avec le scénario basique B3 restent raisonnables (moins de 80 secondes).

Pour les instances de sac-à-dos classique ($m = 1$), dans le cas où le scénario basique B3 est utilisé, l'impact des inégalités ajoutées au nœud racine est étudié. On s'intéresse à la réduction de l'écart entre la valeur de la relaxation linéaire et la valeur de l'optimum. Les résultats sont donnés dans le tableau 6.6. Comme on pouvait s'y attendre, l'impact des inégalités ajoutées reste modeste, puisque l'écart à l'optimum est réduit de seulement 5,72% en moyenne.

TAB. 6.3 – Résultats pour 10 instances de sac-à-dos, avec $n = 200$ ($m = 1$), en utilisant le scénario basique B3.

instance	nb de nœuds	nb d'inégalités	valeur	temps (sec.)
0	202	8	89345	0,81
1	781	22	86879	2,56
2	274	12	89383	1,03
3	776	29	75254	2,79
4	207	4	73485	0,61
5	176	15	99313	0,74
6	24	0	92647	0,09
7	455	11	65483	1,89
8	363	28	70916	1,38
9	42	1	87333	0,17
moyenne	330	13	83003,8	1,21

TAB. 6.4 – Résultats pour 10 instances de sac-à-dos multi-dimensionnels, avec $n = 200$ et $m = 5$, en utilisant le scénario basique B2.

instance	nb de nœuds	nb d'inégalités	valeur	écart / optimum	temps (sec.)
0	27661	4784	81353	0,08%	t.l.
1	13094	9776	63987	0,11%	t.l.
2	11301	8381	61986	0,20%	t.l.
3	12369	8965	67729	0,05%	t.l.
4	14039	6360	64082	0,12%	t.l.
5	11497	8580	70711	0,12%	t.l.
6	12301	7186	61663	0,16%	t.l.
7	11301	6741	68490	0,15%	t.l.
8	10667	8640	74675	0,11%	t.l.
9	11198	9550	62930	0,07%	t.l.
moyenne	13542,8	7896,3	67760,6	0,12%	t.l.

TAB. 6.5 – Résultats pour 10 instances de sac-à-dos multi-dimensionnels, avec $n = 200$ et $m = 5$, en utilisant le scénario basique B3.

instance	nb de nœuds	nb d'inégalités	valeur	temps (sec.)
0	3349	14	81353	41,03
1	1020	0	64022	12,87
2	6296	42	62018	80,53
3	299	1	67745	3,49
4	329	8	64127	5,28
5	2719	18	70731	37,07
6	1266	24	61708	17,38
7	3250	27	68520	41,07
8	1046	0	74706	12,57
9	1333	11	62930	16,74
moyenne	2090,7	14,5	67786	26,8

TAB. 6.6 – Impact des coupes ajoutées au nœud racine, pour $n = 200$ et $m = 1$.

instance	nb d'inég.	valeur optimale	valeur init. de la relaxation	valeur finale de la relaxation	réduction de l'écart
0	2	89345	89400,2	89395,2	9,14%
1	2	86879	86926,6	86925	3,43%
2	2	89383	89431,6	89428,6	6,29%
3	2	75254	75330,4	75326,1	5,55%
4	2	73485	73528,7	73527,6	2,36%
5	2	99313	99343,9	99341,6	7,27%
6	2	92647	92681,6	92678,4	9,19%
7	2	65483	65564,9	65558,9	7,23%
8	2	70916	70958,9	70957,5	3,25%
9	2	87333	87362,9	87361,5	4,62%
moyenne	2	81726,75	81775,36	81772,53	5,72%

6.5.2 Un problème de localisation avec contrainte de capacité

On considère dans cette dernière partie un problème de localisation sous contraintes probabilistes. Ce problème a des applications dans de nombreux contextes. Soient I l'ensemble des clients, et J l'ensemble des sites pouvant accueillir un équipement. Chaque client $i \in I$ a une demande d_i , et la capacité d'un équipement est notée C . On suppose que chaque client doit être assigné à un seul site équipé. Le coût de transport entre le site j et le client i est noté a_{ij} , et enfin b_j désigne le coût d'installation d'un équipement au site j .

Chaque demande d_i est supposée incertaine : d_i est donc une variable aléatoire. On introduit alors le problème de localisation avec contrainte probabiliste de capacité (PCFLP, pour *Probabilistic Capacitated Facility Location Problem*) :

$$\begin{aligned}
 \min \quad & \sum_{i \in I} \sum_{j \in J} a_{ij} x_{ij} + \sum_{j \in J} b_j y_j \\
 \text{s.c.} \quad & \sum_{j \in J} x_{ij} = 1, \quad \forall i \in I, \\
 & \mathbb{P} \left(\sum_{i \in I} d_i x_{ij} \leq C y_j \right) \geq 1 - \varepsilon, \quad \forall j \in J, \\
 & x_{ij} \in \{0, 1\}, y_j \in \{0, 1\}, \quad \forall i \in I, j \in J.
 \end{aligned}$$

$\varepsilon \in [0, 1[$ est la tolérance sur le dépassement de la capacité de chaque site $j \in J$. Les algorithmes implémentés utilisent le scénario basique amélioré grâce au lemme 43, dans la direction $v = (1, \dots, 1)$.

Distributions triangulaires : Dans ce premier exemple, on suppose que les $\{d_i\}_{i \in I}$ sont des variables aléatoires indépendantes. Pour chaque $i \in I$, d_i a une distribution triangulaire symétrique sur l'intervalle $[\underline{d}_i, \bar{d}_i]$ ($\underline{d}_i \geq 0$). Plus précisément, la densité de probabilité est définie par :

$$f_i(x) = \begin{cases} 0, & \text{if } x \leq \underline{d}_i \text{ or } x \geq \bar{d}_i, \\ \alpha_i \cdot (x - \underline{d}_i), & \text{if } \underline{d}_i \leq x \leq (\underline{d}_i + \bar{d}_i)/2, \\ \alpha_i \cdot (\bar{d}_i - x), & \text{if } (\underline{d}_i + \bar{d}_i)/2 \leq x \leq \bar{d}_i, \end{cases}$$

où : $\alpha_i = 2/(\bar{d}_i - \underline{d}_i)^2$.

Les données sont générées aléatoirement : chaque demande minimale \underline{d}_i est choisie entière entre 50 et 500, chaque coût de lien est choisi entre 10 et 100, et chaque coût d'installation entre 200 et 300. L'ampleur de l'incertitude est la même pour toutes les demandes : $\bar{d}_i - \underline{d}_i = 30$,

TAB. 6.7 – Distributions triangulaires : résultats pour le problème de localisation PCFLP.

Taille		Algorithme exact				Modèle gaussien approché		
n	m	nb de nœuds	nb d'inég.	temps (sec.)	valeur	nb de nœuds	temps (sec.)	valeur
20	10	22	26	0,19	2152	14288	t.l. (2,89%)	2162
30	10	3	18	0,23	2369	2419	504,23	2369
50	10	72	54	1,41	2555	442	398,5	2555
50	15	246	111	6,72	3112	500	t.l. (1,26%)	3112
50	20	13872	4130	t.l. (5,59%)	3956	209	t.l. (8,05%)	3965
100	20	5772	1810	t.l. (1,33%)	5005	aucune solution trouvée		
100	30	2819	1591	t.l. (3,86%)	6422	aucune solution trouvée		

pour tout $i \in I$. La capacité d'un équipement est définie ainsi : $C = \lceil 1,5 * \sum_{i \in I} \bar{d}_i / m \rceil$. Enfin, on considère $\varepsilon = 0,1$.

L'usage d'hypothèses gaussiennes pour les problèmes sous contraintes probabilistes est souvent motivée dans la littérature à l'aide du théorème de la limite centrale. On construit un modèle gaussien approché de notre problème, comme décrit dans la partie 6.2.1. Après calculs et simplifications, on obtient :

$$\begin{aligned}
 \min \quad & \sum_{i \in I} \sum_{j \in J} a_{ij} x_{ij} + \sum_{j \in J} b_j y_j \\
 \text{s.c.} \quad & \sum_{j \in J} x_{ij} = 1, & \forall i \in I, \\
 & \sum_{i \in I} \mathbb{E}[d_i] x_{ij} \leq C y_j, & \forall j \in J, \\
 & \sum_{i \in I} \left(\Phi^{-1}(1 - \varepsilon)^2 \sigma^2 - \mathbb{E}[d_i]^2 \right) \cdot x_i + 2 \sum_{i \in I} \mathbb{E}[d_i] C \cdot v_{ji} \\
 & \quad - 2 \sum_{i < k} \mathbb{E}[d_i] \mathbb{E}[d_k] \cdot u_{jik} \leq C^2 y_j, & \forall j \in J, \\
 & u_{jik} \leq x_{ij} \text{ et } u_{jik} \leq x_{kj}, & \forall j \in J, (i, k) \in I^2, i < k, \\
 & v_{ji} \geq y_j + x_{ij} - 1, & \forall j \in J, i \in I, \\
 & x_{ij} \in \{0, 1\}, y_j \in \{0, 1\}, & \forall i \in I, j \in J, \\
 & u, v \geq 0.
 \end{aligned} \tag{6.12}$$

où σ^2 est la variance d'une demande. (Sous nos hypothèses : $\sigma^2 = \delta^2 / 24 = 75 / 2$.)

Les résultats apparaissent dans le tableau 6.7. Notre algorithme résout les instances de manière satisfaisante : l'optimum est atteint jusqu'à la taille 50×15 ; ensuite, même si l'optimalité de la solution courante n'est pas prouvée, l'écart avec l'optimum est faible. Remarquons que les tailles d'instances examinées sont suffisamment petites pour que l'algorithme dit "exact" soit en effet précis : toutes les probabilités sont calculées à partir de la formule analytique, sans recours à des approximations gaussiennes. Ainsi, les résultats du tableau 6.7 montrent que l'approximation gaussienne est très proche du problème exact. Pourtant, cette approximation est à l'usage beaucoup plus difficile à résoudre : les temps de résolution sont incomparablement plus longs. Il apparaît que ce modèle rend même la relaxation linéaire difficile à résoudre : celle-ci n'a pas été résolue au nœud racine dans le temps imparti pour les instances 100×20 et 100×30 .

Distributions de Bernoulli : On suppose pour finir que chaque demande d_i peut prendre seulement deux valeurs, \underline{d}_i et \bar{d}_i , avec probabilités respectives $1 - p$ et $p \in]0, 1[$. De plus, supposons que l'écart entre ces deux valeurs possibles est le même pour toute demande : pour tout $i \in I$, $\bar{d}_i - \underline{d}_i = \delta > 0$. On note $\eta_i = d_i - \underline{d}_i$. Pour tout $I_1 \subseteq I$ ($n_1 = |I_1|$), $\sum_{i \in I_1} \eta_i$ suit une

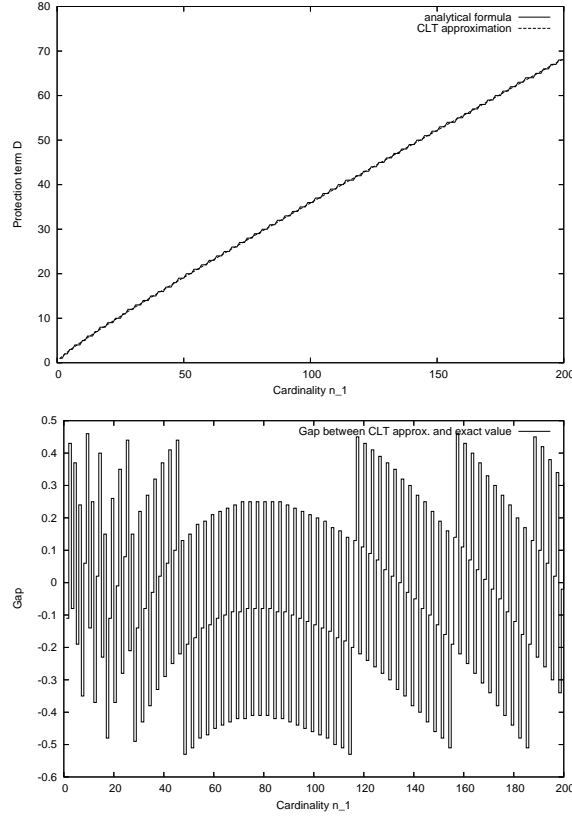


FIG. 6.4 – Distributions de Bernoulli : comparaison du calcul exact des termes $D_j(I_1)$ avec l’approximation dérivée du théorème de la limite centrale, en fonction de $n_1 = |I_1|$.

loi binomiale :

$$\forall k \in \{0, \dots, n_1\}, \mathbb{P} \left(\sum_{i \in I_1} \eta_i = k\delta \right) = \binom{n_1}{k} p^k (1-p)^{n_1-k}.$$

Noter que, dès que $p < 1 - \varepsilon$, le scénario \underline{d} satisfait la propriété (6.1) : c’est un scénario basique. Toutes les données \underline{d}_i , \bar{d}_i et C ont été générées comme dans le précédent paragraphe. On a donc : $\delta = 30$. On pose de plus : $p = 0,3$. (La variance associée est $\sigma^2 = \delta^2 p(1-p) = 189$.)

Comme précédemment, l’approximation gaussienne a aussi été testée dans ce cas, avec le modèle (6.12). Pourtant, la figure 6.4 illustre la difficulté d’approcher une distribution discrète avec une gaussienne (continue). Les résultats sont donnés dans le tableau 6.8. Les performances observées pour notre algorithme sont similaires à celles du paragraphe précédent (avec distributions triangulaires). Le modèle gaussien reste beaucoup plus difficile à résoudre. De plus, les solutions obtenues *via* cette approximation ne sont pas toujours optimales pour le problème initial considéré (voir les instances 30×10 et 50×10). Ceci est une motivation supplémentaire pour résoudre les problèmes sous contraintes probabilistes en tenant compte des hypothèses probabilistes de manière précise.

6.6 Conclusion

Dans ce chapitre, la résolution optimale de problèmes linéaires en nombres entiers 0-1 sous contraintes probabilistes a été étudiée. L’ensemble des solutions admissibles a été caractérisé par

TAB. 6.8 – Distributions de Bernoulli : résultats pour le problème de localisation probabiliste PCFLP.

Taille		Algorithme exact				Modèle gaussien approché		
n	m	nb de nœuds	nb d'inég.	temps (sec.)	valeur	nb de nœuds	temps (sec.)	valeur
20	10	172	54	0,68	2144	14345	t.l. (1,49%)	2152
30	10	11	21	0,22	2368	811	175,11	2369
50	10	56	42	0,95	2551	404	358,36	2555
50	15	6985	1069	171	3108	460	t.l. (3,60%)	3112
50	20	14501	4214	t.l. (5,64%)	3952	86	t.l. (9,95%)	4050

un nombre exponentiel d'inégalités linéaires. La séparation de ces inégalités a ensuite été étudiée, et un algorithme de résolution par *branch-and-cut* a été défini. L'ensemble de l'analyse s'appuie sur un scénario, dit basique, dont la définition a un impact fort sur la qualité de la relaxation linéaire définie. Des propriétés théoriques et des algorithmes sont détaillés afin d'obtenir des scénarios basiques efficaces. D'un point de vue général, un tel scénario basique devrait trouver usage quelle que soit la méthode de résolution envisagée pour un problème combinatoire sous contraintes probabilistes.

Des tests numériques ont été réalisés sur des instances du problème de sac-à-dos multi-dimensionnel, ainsi que sur des instances d'un problème de localisation. Les résultats obtenus confirment l'impact fort des différents composants de l'algorithme de résolution. Des instances de tailles moyennes sont résolues à l'optimum en quelques secondes. On observe de plus la très bonne capacité de cette approche à fournir rapidement des solutions réalisables de très bonne qualité. Par ailleurs, il est classique dans la littérature d'avoir recours à des hypothèses gaussiennes : dans ce contexte, on sait écrire des modèles déterministes équivalents au problème sous contraintes probabilistes. Une comparaison numérique avec notre approche montre cependant que le modèle obtenu est beaucoup plus dur à résoudre. En pratique, cette démarche classique met plus de temps à fournir des solutions réalisables, celles-ci sont de coût plus élevé, et l'écart théorique prouvé à l'optimum est également plus important. Enfin, le recours à des hypothèses gaussiennes est souvent motivé à l'aide du théorème de la limite centrale. Nos tests montrent qu'une telle approximation peut être imparfaite, et peut mener à des solutions différentes des optimums réels, obtenus par notre approche.

Troisième partie

Applications

Chapitre 7

Localisation de points de mesure dans un réseau

Motivations : Dans ce chapitre, on présente une première application des approches développées auparavant. On traite un problème de placement de points de mesure de trafic dans un réseau, pour lequel une modélisation avec contrainte probabiliste est particulièrement adaptée. La méthodologie présentée a fait l'objet d'un dépôt de brevet auprès de l'INPI, sous le titre "Procédé de sélection de points de collecte d'informations relatives au trafic à travers des liens d'un réseau" (dépôt 0753862).

7.1 Introduction

Les opérateurs de télécommunications ont besoin d'observations de trafic pour effectuer une gestion efficace de leurs réseaux. Une des données les plus classiquement utilisées est la quantité de trafic sur chaque lien du réseau, connue grâce au protocole *Simple Network Management Protocole* (SNMP). Depuis quelques années, des outils plus sophistiqués sont apparus pour permettre une analyse plus en détail du trafic dans les réseaux IP (Internet Protocole). C'est en particulier le cas de la technologie NetFlow, développée par Cisco [25] (des solutions équivalentes existent chez d'autres équipementiers, comme par exemple Juniper [73]). En utilisant NetFlow sur des interfaces de routeurs, les gestionnaires du réseau ont accès à des caractérisations détaillées du trafic, avec notamment les origines et destinations des flux observés. Cet aspect est essentiel pour l'estimation de matrices de trafic source-destination [3, 54, 89, 63, 39, 78].

Une telle description du trafic est essentielle pour de nombreux aspects, tels la supervision de réseau en temps réel, la planification du réseau (dimensionnement), des analyses de sécurité, l'ingénierie de trafic, et même des questions de facturation entre opérateurs. C'est la raison pour laquelle la plupart des opérateurs de télécommunications ont adopté ce genre de solutions de mesure. Pourtant, leur usage présente aussi quelques inconvénients [26, 88]. D'abord, comme observé par Zang et Nucci dans [88], la mise en place de cette technologie est coûteuse : des équipements et des logiciels doivent être achetés et maintenus, ce qui correspond à des CAPEX et à des OPEX (*capital expenditures* et *operational expenditures*). De plus, la collecte de l'information mesurée a un impact sur la charge du réseau. En effet, une observation assez précise du trafic consomme des ressources en processeur et en mémoire sur les routeurs. Par ailleurs, les données mesurées doivent être envoyées à un (ou plusieurs) serveur centralisé, qui doit aussi être acheté. Ces considérations motivent la limitation du déploiement ou de l'usage de NetFlow dans

le réseau. La question qui s'impose alors naturellement est la suivante : comment sélectionner un ensemble d'interfaces sur lesquelles NetFlow doit être activé ?

Zang et Nucci ont abordé cette question dans le contexte du déploiement de NetFlow dans un réseau [88]. À l'aide de programmes linéaires en nombres entiers, ils développent une méthodologie pour optimiser les investissements par une sélection des routeurs et des interfaces sur lesquels NetFlow doit être activé. Leur approche repose sur des données en trafic statiques et supposées connues avec précision. Le travail présenté dans ce chapitre se propose de prendre en compte la variabilité du trafic dans le temps. L'objectif est de sélectionner les interfaces sur lesquelles NetFlow doit être activé, tout en assurant qu'une proportion donnée du trafic global est effectivement observée. L'idée est de rendre la décision robuste aux variations de trafic qui surviendront inévitablement. La solution proposée permet de garantir une probabilité d'observation de la quantité voulue du trafic, malgré la variabilité du trafic.

Pour aborder ce problème, on s'appuie sur les modèles et techniques développés dans les chapitres 4 et 5. Des tests numériques sont effectués à partir de données collectées sur un réseau IP international de France Télécom. On montre ainsi la pertinence de l'approche théorique développée. De plus, la rapidité d'exécution de la méthode la rend utilisable dans un contexte dynamique, où la sélection des points de mesure à utiliser serait régulièrement mise à jour.

7.2 Description du problème

7.2.1 Notations et modèles mathématiques

Soit R l'ensemble des routeurs du réseau. À chaque routeur $r \in R$, on associe l'ensemble I_r de ses interfaces d'entrée. Noter que seules les interfaces externes d'entrée sont considérées pour éviter une redondance dans les mesures. On suppose que les mesures SNMP sont disponibles pour toutes les interfaces de $\bigcup_{r \in R} I_r$, et on note $t_{ri} \geq 0$ la quantité de trafic sur l'interface i du routeur r .

La variable x_{ri} prend la valeur 1 si NetFlow doit être activé sur l'interface i du routeur r , 0 sinon. Comme dans [88], notre contrainte est de sélectionner suffisamment d'interfaces pour observer une proportion au moins $\alpha \in [0, 1]$ du trafic total dans le réseau. Ceci s'écrit mathématiquement :

$$\sum_{(r,i) \in RI} t_{ri} x_{ri} \geq \alpha \sum_{(r,i) \in RI} t_{ri}, \quad (7.1)$$

où RI désigne l'ensemble de toutes les interfaces du réseau : $RI = \{(r, i) | r \in R, i \in I_r\}$. On note $n = |RI|$ le nombre total d'interfaces. À un instant donné, les données en trafic $\{t_{ri}\}_{(r,i) \in RI}$ sont supposées connues grâce aux mesures SNMP.

La difficulté principale vient de la satisfaction des contraintes (7.1) alors que le trafic varie au cours du temps. En effet, la sélection des interfaces est faite à un instant donné, mais ne peut être remise en cause trop souvent. On souhaite que la décision prise assure des résultats aussi stables que possible.

On doit donc considérer que les données t_{ri} sont des variables aléatoires. On cherche une sélection x d'interfaces qui satisfait :

$$\mathbb{P} \left(\sum_{(r,i) \in RI} t_{ri} x_{ri} \geq \alpha \sum_{(r,i) \in RI} t_{ri} \right) \geq 1 - \varepsilon, \quad (7.2)$$

où $\varepsilon \in [0, 1[$ est une tolérance fixée par le gestionnaire de réseau. L'équation (7.2) signifie qu'une sélection x d'interfaces assurera la mesure d'une proportion α du trafic total avec une probabilité garantie de $1 - \varepsilon$.

Cette contrainte probabiliste doit être insérée dans le problème de décision global. Pour rendre l'analyse aussi générale que possible, on note $f(x)$ le coût associé à la sélection x . Ce coût peut inclure différents objectifs, économiques (investissements) ou de qualité de service par exemple (impact sur la charge du réseau, sur la performance des routeurs, etc.). De plus, des contraintes autres que (7.2) peuvent en pratique vouloir être appliquées à une décision x : on les note simplement $g(x) \geq 0$. Ainsi, le problème de décision global s'écrit :

$$\begin{aligned} \min \quad & f(x) \\ \text{s.c.} \quad & g(x) \geq 0, \\ & \mathbb{P} \left(\sum_{(r,i) \in RI} t_{ri}(x_{ri} - \alpha) \geq 0 \right) \geq 1 - \varepsilon, \\ & x \in \{0, 1\}^n. \end{aligned} \tag{7.3}$$

Un exemple de fonctions f et g sera présenté dans la partie numérique (partie 7.4). Observer que si $\alpha = 1$, on a $x_{ri} \leq \alpha$ pour tout (r, i) : ceci signifie que NetFlow doit être activé sur toutes les interfaces. À partir de maintenant, on suppose que $\alpha \in]0, 1[$.

7.2.2 Approche pessimiste

Les valeurs de trafic t_{ri} varient au cours du temps, on doit donc caractériser leur aléas. On s'appuie pour cela sur des valeurs de trafic observées par le passé, qui serviront à caractériser ce qui se passera dans l'avenir. On suppose que chaque variable aléatoire t_{ri} prend ses valeurs dans un intervalle $[\underline{t}_{ri}, \bar{t}_{ri}]$, \underline{t}_{ri} (resp. \bar{t}_{ri}) étant la valeur maximale (resp. minimale) de trafic observée sur cette interface par le passé (d'après les mesures SNMP).

L'inégalité (7.1) peut être écrite de manière équivalente : $F(x, t) = \sum_{(r,i) \in RI} t_{ri}(x_{ri} - \alpha) \geq 0$. On souhaite dans un premier temps assurer que $F(x, t)$ est *toujours* positive, ce qui revient à considérer $\varepsilon = 0$ dans l'équation probabiliste (7.2). Alors, pour chaque $(r, i) \in RI$, il faut déterminer la valeur minimale que $t_{ri}(x_{ri} - \alpha)$ peut atteindre. Cette valeur minimale est différente selon que $x_{ri} = 1$ ou que $x_{ri} = 0$. Si $x_{ri} = 1$, alors : $t_{ri}(x_{ri} - \alpha) \geq (1 - \alpha)\underline{t}_{ri}$ (on rappelle que $\alpha < 1$). Si au contraire $x_{ri} = 0$, alors : $t_{ri}(x_{ri} - \alpha) \geq -\alpha\bar{t}_{ri}$ (en effet, $\alpha > 0$). Ceci implique : $t_{ri}(x_{ri} - \alpha) \geq x_{ri} \cdot (1 - \alpha)\underline{t}_{ri} - (1 - x_{ri}) \cdot \alpha\bar{t}_{ri}$. On déduit de cette analyse que l'inégalité suivante est toujours vraie :

$$\sum_{(r,i) \in RI} t_{ri}(x_{ri} - \alpha) \geq \sum_{(r,i) \in RI} x_{ri}(1 - \alpha)\underline{t}_{ri} - (1 - x_{ri})\alpha\bar{t}_{ri}.$$

Ainsi, une approche pessimiste à notre problème de sélection d'interfaces est :

$$\begin{aligned} \min \quad & f(x) \\ \text{s.c.} \quad & g(x) \geq 0, \\ & \sum_{(r,i) \in RI} x_{ri} [\alpha\bar{t}_{ri} + (1 - \alpha)\underline{t}_{ri}] \geq \alpha \sum_{(r,i) \in RI} \bar{t}_{ri}, \\ & x \in \{0, 1\}^n. \end{aligned} \tag{7.4}$$

Cette approche a l'avantage de fournir une solution extrêmement robuste, même lorsqu'aucune information probabiliste n'est disponible. Cependant, la solution calculée est valide quand, pour chaque interface sélectionnée, le trafic est minimal, et pour chaque interface délaissée, le trafic est maximal. Une telle chose n'arrive jamais en pratique. C'est ce qui motive le fait de considérer $\varepsilon > 0$.

7.2.3 Construction des intervalles de trafic

L'intervalle $[\underline{t}_{ri}, \bar{t}_{ri}]$ associé à l'interface i du routeur r est déduit de l'historique des mesures SNMP. Pour être pertinentes, il faut que ces données couvrent une période de temps assez

longue. En effet, supposons par exemple que l'on ne dispose de données que pour le trafic entre 7h00 et 8h00 du matin. Dans ce cas, \underline{t}_{ri} et \bar{t}_{ri} correspondent respectivement à la plus faible et à la plus forte valeur de trafic observée. Mais ces valeurs seront certainement très différentes de celles entre 8h00 et 9h00.

Si l'on dispose de mesures sur une journée, les choses sont plus stables. En effet, on sait que le trafic a tendance à se ressembler d'un jour sur l'autre (au moins en dehors du week-end). Il est ainsi justifié de s'appuyer sur les mesures de lundi, par exemple, pour caractériser le trafic jusqu'au vendredi suivant. Des raisonnements similaires sont valides à l'échelle de semaines ou de mois. Il suffit donc de déterminer une période de temps dont les mesures seront assez représentatives pour caractériser les périodes de temps à venir. Bien sûr, la sélection d'interfaces risque de devoir être remise en cause au bout d'un certain temps.

7.3 Un algorithme efficace

L'algorithme proposé est une adaptation directe de celui proposé au chapitre 4 pour le problème de sac-à-dos probabiliste. Il faut pour cela définir le modèle robuste de Bertsimas et Sim [13] associé à notre problème. Ceci n'est pas totalement immédiat. Soit un paramètre entier $\Gamma \in \{0, \dots, |RI|\}$: il indique qu'on autorise au plus Γ valeurs de trafic à prendre leur valeur la plus défavorable, tandis que les $n - \Gamma$ autres prennent leur valeur la plus favorable. Ainsi, plus la valeur de Γ est élevée, plus la probabilité de réalisabilité de la solution est grande, et plus le coût de la solution est important. Ici, les "pire" et "meilleure" valeurs d'un trafic t_{ri} dépendent de la solution x considérée.

Plus précisément, en se reportant à l'analyse de la partie 7.2.2, on sait que les inégalités suivantes sont toujours vraies :

$$\begin{cases} t_{ri}(x_{ri} - \alpha) \geq x_{ri}(1 - \alpha)\underline{t}_{ri} - (1 - x_{ri})\alpha\bar{t}_{ri} & (i) \\ t_{ri}(x_{ri} - \alpha) \leq x_{ri}(1 - \alpha)\bar{t}_{ri} - (1 - x_{ri})\alpha\underline{t}_{ri} & (ii) \end{cases}.$$

On cherche donc la meilleure solution réalisable même lorsque Γ quantités $\{t_{ri}(x_{ri} - \alpha)\}_{(r,i) \in RI}$ prennent leur plus faibles valeurs (i), tandis que les $n - \Gamma$ autres prennent leurs plus fortes valeurs (ii). On observe que si $\Gamma = n$, on retrouve l'approche pessimiste de la partie 7.2.2.

La solution x doit donc satisfaire :

$$\forall S \subseteq RI \text{ t.q. } |S| = \Gamma : \sum_{(r,i) \in S} [x_{ri}(1 - \alpha)\underline{t}_{ri} - (1 - x_{ri})\alpha\bar{t}_{ri}] + \sum_{(r,i) \notin S} [x_{ri}(1 - \alpha)\bar{t}_{ri} - (1 - x_{ri})\alpha\underline{t}_{ri}] \geq 0. \quad (7.5)$$

7.3.1 Forme polynomiale du problème robuste

En suivant l'analyse de Bertsimas et Sim, la formulation (7.5), de taille exponentielle, peut être ramenée à une autre plus compacte (taille polynomiale), en s'appuyant sur la dualité en programmation linéaire.

Soit $\delta_{ri} = \bar{t}_{ri} - \underline{t}_{ri}$ la variation maximale de trafic sur l'interface i du routeur r . On introduit :

$$F(x) = \min_{S \subseteq RI, |S| = \Gamma} \left\{ \alpha \sum_{(r,i) \in S} \delta_{ri}(x_{ri} - 1) + (1 - \alpha) \sum_{(r,i) \notin S} \delta_{ri}x_{ri} \right\}.$$

On a alors : (7.5) $\Leftrightarrow \sum_{(r,i) \in RI} \underline{t}_{ri}(x_{ri} - \alpha) + F(x) \geq 0$. Par ailleurs, pour une solution x^* fixée :

$$\begin{aligned} F(x^*) &= \min \quad \alpha \sum_{(r,i) \in RI} \delta_{ri}(x_{ri}^* - 1)z_{ri} + (1 - \alpha) \sum_{(r,i) \in RI} \delta_{ri}x_{ri}^*(1 - z_{ri}) \\ &\text{s.c.} \quad \sum_{(r,i) \in RI} z_{ri} = \Gamma, \\ &\quad z_{ri} \in \{0, 1\}, \forall (r, i) \in RI \\ &= \min \quad \alpha \sum_{(r,i) \in RI} \delta_{ri}(x_{ri}^* - 1)z_{ri} + (1 - \alpha) \sum_{(r,i) \in RI} \delta_{ri}x_{ri}^*(1 - z_{ri}) \\ &\text{s.c.} \quad \sum_{(r,i) \in RI} z_{ri} \leq \Gamma, \\ &\quad z_{ri} \in \{0, 1\}, \forall (r, i) \in RI. \end{aligned}$$

L'objectif peut s'écrire, à une constante additive $C(x^*) = (1 - \alpha) \sum_{(r,i) \in RI} \delta_{ri}x_{ri}^*$ près :

$$\alpha \sum_{(r,i) \in RI} \delta_{ri}(x_{ri}^* - 1)z_{ri} - (1 - \alpha) \sum_{(r,i) \in RI} \delta_{ri}x_{ri}^*z_{ri} = \sum_{(r,i) \in RI} \delta_{ri}z_{ri}(2\alpha x_{ri}^* - \alpha - x_{ri}^*).$$

Les coefficients de cet objectif sont tous négatifs, puisque $x_{ri}^* \in \{0, 1\}$. On ré-écrit $\tilde{F}(x^*) = F(x^*) - C(x^*)$ sous forme d'une maximisation de l'opposé, et on voit aisément que les contraintes d'intégrité peuvent être relâchées :

$$\begin{aligned} -\tilde{F}(x^*) &= \max \quad \sum_{(r,i) \in RI} \delta_{ri}z_{ri}(\alpha + x_{ri}^* - 2\alpha x_{ri}^*) \\ &\text{s.c.} \quad \sum_{(r,i) \in RI} z_{ri} \leq \Gamma, \\ &\quad 0 \leq z_{ri} \leq 1, \forall (r, i) \in RI. \end{aligned}$$

Par dualité, on a aussi :

$$\begin{aligned} -\tilde{F}(x^*) &= \min \quad \Gamma u + \sum_{(r,i) \in RI} v_{ri} \\ &\text{s.c.} \quad u + v_{ri} \geq \delta_{ri}(\alpha + x_{ri}^* - 2\alpha x_{ri}^*), \quad \forall (r, i) \in RI, \\ &\quad u, v \geq 0. \end{aligned}$$

Le problème robuste s'écrit :

$$\begin{aligned} \min \quad & f(x) & \min \quad & f(x) \\ \text{s.c.} \quad & g(x) \geq 0, & \text{s.c.} \quad & g(x) \geq 0, \\ & (7.5), & & \sum_{(r,i) \in RI} \underline{t}_{ri}(x_{ri} - \alpha) + C(x) + \tilde{F}(x) \geq 0, \\ & x \in \{0, 1\}^n & \Leftrightarrow & x \in \{0, 1\}^n \end{aligned}$$

En injectant la formulation duale de $\tilde{F}(x)$ dans cette formulation, on obtient un nouveau modèle équivalent à $\min_{x \in \{0, 1\}^n} \{f(x) \mid g(x) \geq 0 \text{ et } (7.5)\}$:

$$\begin{aligned} \min & \quad f(x) \\ \text{s.c.} & \quad g(x) \geq 0, \\ & \quad \sum_{(r,i) \in RI} \underline{t}_{ri}(x_{ri} - \alpha) + (1 - \alpha) \sum_{(r,i) \in RI} \delta_{ri}x_{ri} - \left(\Gamma u + \sum_{(r,i) \in RI} v_{ri} \right) \geq 0, \\ & \quad u + v_{ri} \geq \delta_{ri}(\alpha + x_{ri} - 2\alpha x_{ri}), \quad \forall (r, i) \in RI, \\ & \quad u, v \geq 0, \\ & \quad x \in \{0, 1\}^n, \end{aligned}$$

qui peut encore s'écrire :

$$\begin{aligned} \min & \quad f(x) \\ \text{s.c.} & \quad g(x) \geq 0, \\ & \quad \sum_{(r,i) \in RI} [\underline{t}_{ri} + (1 - \alpha)\delta_i]x_{ri} - \Gamma u - \sum_{(r,i) \in RI} v_{ri} \geq \alpha \sum_{(r,i) \in RI} \underline{t}_{ri}, \\ & \quad u + v_{ri} \geq \delta_{ri}(\alpha + x_{ri} - 2\alpha x_{ri}), \quad \forall (r, i) \in RI, \\ & \quad u, v \geq 0, \\ & \quad x \in \{0, 1\}^n. \end{aligned} \tag{7.6}$$

Cette dernière formulation est de taille polynomiale. De plus, noter qu'elle est en fait utilisable pour $\Gamma \in [0, n]$ non-nécessairement entier. On appelle ce problème robuste RNMSp(Γ) (*Robust NetFlow Measurement Selection Problem*).

7.3.2 Algorithme de résolution

L'algorithme consiste en la résolution d'une suite de problèmes robustes RNMSp(Γ) pour des valeurs croissantes de Γ (cf. chapitre 4). On commence avec $\Gamma = 0$, puis on augmente progressivement ce paramètre de robustesse. Pour chaque nouvelle valeur, le problème robuste correspondant est résolu. Ainsi, la probabilité de réalisabilité des solutions calculées tend à augmenter. On arrête dès qu'une solution satisfait la contrainte probabiliste (7.2).

Heuristique de résolution de (7.3)

Étape 0 : Soit $\Gamma = 0$, et soit $d > 0$ arbitrairement petit.

Étape 1 : Résoudre RNMSp(Γ), et soit x la solution obtenue.

Étape 2 : Vérifier si la contrainte (7.2) est satisfaite.
Si oui, STOP.

Étape 3 : Calculer $\Gamma(x)$ comme décrit plus bas.

Étape 4 : Mettre à jour le paramètre de robustesse : $\Gamma \leftarrow \Gamma(x) + d$.
Aller à l'étape 1.

Les performances de cet algorithme peuvent être grandement améliorées si l'on choisit judicieusement le point de départ. Ceci peut être fait facilement, en cherchant par exemple par dichotomie une valeur Γ telle que la solution du problème RNMSp(Γ) a une probabilité de réalisabilité assez élevée (mais pas trop proche de $1 - \varepsilon$). Pour ce qui est du paramètre d , il est indispensable pour garantir que la mise à jour du paramètre Γ , à l'étape 4, conduira à une nouvelle solution x . On renvoie à l'analyse de la partie 5.4.3 pour des indications sur la manière de définir d ; on peut par exemple prendre $d = 1 / \max_{(r,i) \in RI} \delta_{ri}$ (choix particulièrement justifié quand les données sont toutes entières, cf. algorithme M2 du § 5.5.1). Noter que plus d est grand, plus la convergence de l'algorithme est rapide, mais plus on risque de dégrader la qualité de la solution obtenue.

Soit x une solution obtenue à l'étape 1, $\Gamma(x)$ est le plus grand paramètre tel que x est une solution réalisable du problème robuste associé (on a bien sûr $\Gamma(x) \geq \Gamma$). Pour le calcul de $\Gamma(x)$, il faut remarquer que l'équation (7.5) s'écrit de manière équivalente (lorsque Γ est entier) :

$$(i) \quad \sum_{(r,i) \in RI} [\bar{t}_{ri}(1 - \alpha)x_{ri} - \underline{t}_{ri}\alpha(1 - x_{ri})] \geq \max_{|S|=\Gamma} \sum_{(r,i) \in S} |x_{ri} - \alpha| \cdot \delta_{ri}.$$

Il s'agit donc d'abord de trouver le plus grand Γ' entier satisfaisant cette relation (i). Si (i) est vérifiée à l'égalité pour Γ' , alors $\Gamma(x) = \Gamma'$. Sinon, il faut ajouter une partie fractionnaire à Γ' . L'algorithme suivant détaille cette procédure de construction de $\Gamma(x)$.

Algorithme de calcul de $\Gamma(x)$

Étape 0 : Soit $Q = \sum_{(r,i) \in RI} [\bar{t}_{ri}(1 - \alpha)x_{ri} - \underline{t}_{ri}\alpha(1 - x_{ri})]$, et $\Gamma(x) = 0$.

Étape 1 : Trier les valeurs $\{|x_{ri} - \alpha| \cdot \delta_{ri}\}_{(r,i) \in RI}$ par ordre décroissant.
On note $\{v_k\}_{1 \leq k \leq n}$ la liste ordonnée obtenue.

Étape 2 : Pour $k = 1$ à n :
- si $Q \geq v_k$: $Q \leftarrow Q - v_k$, $\Gamma(x) \leftarrow \Gamma(x) + 1$;
- sinon : $\Gamma(x) \leftarrow \Gamma(x) + Q/v_k$; STOP.

Cet algorithme s'arrête en temps $O(n \log(n))$ (complexité du tri de l'étape 1).

7.3.3 Hypothèses probabilistes

L'étape 2 de l'heuristique de résolution ci-dessus nécessite le calcul de la probabilité de réalisabilité d'une solution. On propose de s'appuyer sur l'inégalité de Hoeffding (voir théorème 12); des approches plus fines seraient envisageables, voir par exemple le paragraphe 7.5. La moyenne du trafic $\mathbb{E}[t_{ri}]$ est supposée connue pour chaque interface (r, i) . Comme pour les bornes supérieures et inférieures (\underline{t}_{ri} et \bar{t}_{ri}), cette moyenne est déduite facilement de l'historique des mesures SNMP sur l'interface considérée. On déduit directement de l'inégalité de Hoeffding le résultat suivant.

Lemme 51 *Supposons que les variables aléatoires $\{t_{ri}\}_{(r,i) \in RI}$ sont indépendantes. Soit \bar{x} une sélection d'interfaces sur lesquelles NetFlow est activé, on note : $T(\bar{x}) = \sum_{(r,i) \in RI} (\alpha - \bar{x}_{ri}) t_{ri}$. Alors, si $\mathbb{E}[T(\bar{x})] < 0$:*

$$\mathbb{P} \left(\sum_{(r,i) \in RI} t_{ri} (\bar{x}_{ri} - \alpha) \geq 0 \right) \geq 1 - \exp \left(- \frac{2\mathbb{E}[T(\bar{x})]^2}{\sum_{(r,i) \in RI} (\alpha - \bar{x}_{ri})^2 \delta_{ri}^2} \right). \quad (7.7)$$

Ce résultat s'appuie sur deux principales hypothèses. La première est l'indépendance des variables aléatoires $\{t_{ri}\}_{(r,i) \in RI}$. Ceci peut ne pas être vérifié en pratique. Certaines dépendances fortes peuvent éventuellement être incluses dans le modèle proposé. Cependant, on peut espérer qu'en pratique, même si l'hypothèse d'indépendance n'est pas vérifiée, le fait que l'inégalité de Hoeffding soit très générale (*i.e.* ne fasse aucune hypothèse sur la distribution des variables aléatoires) contrebalance cette imperfection de modélisation. C'est ce qu'on observera dans notre cas d'étude (voir aussi l'annexe B.3).

La deuxième hypothèse essentielle est : $\mathbb{E}[T(\bar{x})] < 0$. Mais observer que $\mathbb{E}[T(\bar{x})] \geq 0$ signifie : $\mathbb{E} \left[\sum_{(r,i) \in RI} \alpha t_{ri} \right] \geq \mathbb{E} \left[\sum_{(r,i) \in RI} \bar{x}_{ri} t_{ri} \right]$. Dans ce cas, le trafic à mesurer serait, en moyenne, supérieur au trafic moyen observé sur les interfaces sélectionnées dans \bar{x} . Un tel comportement en moyenne aurait peu de sens en pratique.

7.4 Tests sur un réseau IP de France Télécom

7.4.1 Problème spécifique traité

Plusieurs fonctions objectifs peuvent être envisagées, et la plupart d'entre elles sont liées au nombre de routeurs et/ou d'interfaces sur lesquels NetFlow doit être activé. Les interfaces sont déjà caractérisées par la variable binaire x . Pour modéliser les routeurs, on introduit une nouvelle variable y : pour tout routeur $r \in R$, $y_r = 1$ si NetFlow est activé sur au moins une de ses interfaces, et 0 sinon. y satisfait les équations suivantes :

$$\begin{cases} \forall (r, i) \in RI, y_r \geq x_{ri} \\ \forall r \in R, y_r \in \{0, 1\} \end{cases} .$$

Ces équations remplacent " $g(x) \geq 0$ " dans le modèle général (7.3). On considère ici un objectif très simple, qui est la minimisation du nombre de routeurs à solliciter :

$$\min \sum_{r \in R} y_r .$$

Si par exemple on voulait associer des coûts aux routeurs et aux interfaces, il suffirait de considérer l'objectif : $\min \sum_{r \in R} \{A_r y_r + \sum_{i \in I_r} B_{ri} x_{ri}\}$, où A et B représentent l'ensemble des coûts (OPEX et CAPEX) associés. Beaucoup d'objectifs sont possibles, et dépendent du contexte opérationnel.

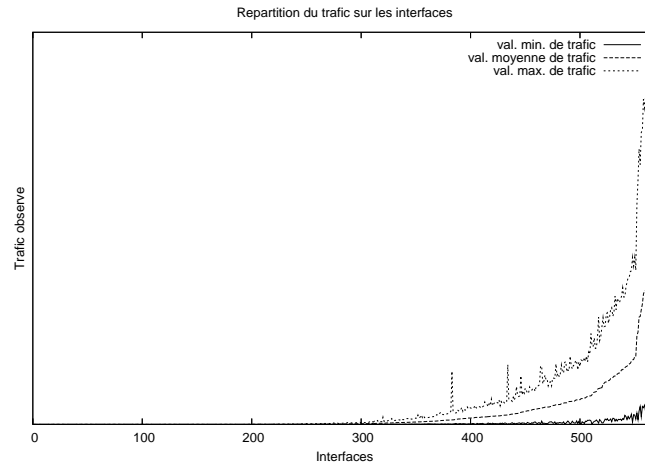


FIG. 7.1 – Distribution des valeurs de trafic sur les interfaces, le 2 octobre 2006.

7.4.2 Données SNMP

Des données SNMP ont été récupérées sur toutes les interfaces d'un réseau IP international de France Télécom. Ce réseau est constitué de 108 routeurs et de 556 interfaces d'entrée. Les données ont été collectées pendant une semaine, entre le 2 et le 8 octobre 2006. Toutes les 2 heures, la quantité de trafic de chaque interface est mesurée pendant 10 minutes. Ainsi, pour chaque journée, on a 12 mesures par interface du réseau.

On s'appuie sur les données du premier jour pour décider de l'emplacement des points de mesure. Pour chaque interface, on déduit des 12 mesures disponibles les valeurs de trafic minimale, maximale et moyenne. La répartition du trafic sur les interfaces du réseau est représentée sur la figure 7.1 (les interfaces sont triées par ordre croissant de valeurs moyennes de trafic). Une simple observation montre que plus de la moitié des interfaces ont un trafic négligeable, comparé au trafic total. Ceci est une motivation supplémentaire pour ne pas activer NetFlow partout.

On a déjà dit que l'hypothèse d'indépendance des trafics n'est pas toujours vraie en pratique. On a calculé les corrélations entre les trafics des différentes interfaces pour nos données. On rappelle que la corrélation de 2 variables aléatoires X et Y est définie par la covariance $cov(X, Y)$, divisée par le produit des écarts-types $\sigma_X \cdot \sigma_Y$. Le critère de tri des corrélations est le suivant (cf. [27]) :

type	valeur de la corrélation c
absente	$ c < 0,1$
faible	$0,1 \leq c < 0,3$
moyenne	$0,3 \leq c < 0,5$
forte	$0,5 \leq c $

Avec nos données, on obtient :

	nombre de paires d'interfaces
non-corrélées	1,1%
faiblement corrélées	5,3%
moyennement corrélées	38,5%
fortement corrélées	55,1%

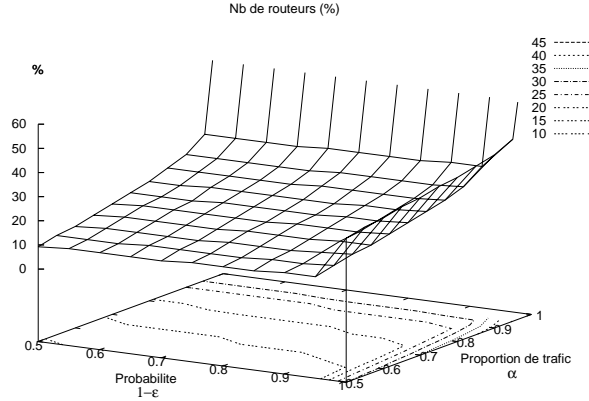


FIG. 7.2 – Nombre de routeurs (en %) où NetFlow doit être activé.

L’hypothèse d’indépendance est donc très largement invalidée dans notre cas. Pourtant, on va montrer que l’approche présentée est pertinente en pratique, ce qui confirmera les espérances exprimées à la suite du lemme 51.

7.4.3 Résultats et analyses

Les tests numériques ont tous été réalisés sur un ordinateur équipé d’un processeur Intel(R) Xeon(TM) à 2,8GHz, avec 3Go de RAM. Pour résoudre les problèmes robustes RNMSp, on a utilisé Cplex 10.0 [43]. L’algorithme de résolution de la partie 7.3.2 a été implémenté avec $d = 0, 1$.

La figure 7.2 donne le pourcentage calculé de routeurs sur lequel NetFlow doit être activé, en fonction de la proportion α de trafic et de la probabilité $p = 1 - \varepsilon$ visée. Pour toute paire (α, p) , l’algorithme de résolution tourne en moins de 15 secondes. Comme on s’y attendait, un petit nombre de routeurs suffit à observer une grande quantité de trafic. Par exemple, observer $\alpha = 90\%$ du trafic total avec une probabilité au moins $p = 0,95$ ne nécessite que 26,9% des routeurs (*i.e.* 29 routeurs sur les 108 du réseau).

Noter de plus que le paramètre α a plus d’impact que la probabilité p sur le nombre de routeurs à sélectionner, au moins quand $p \leq 0,95$. En d’autres termes, étant donnée une proportion de trafic α voulue, garantir que l’on observe cette quantité de trafic avec une probabilité 0,5 ou 0,95 ne représente pas une différence significative du nombre de routeurs concernés. Cette observation est importante : elle signifie la robustesse de la solution calculée aux différentes hypothèses probabilistes qu’on pourrait faire.

Pour valider l’intérêt de cette approche probabiliste, il convient de la comparer avec ce qui serait la démarche classique devant un tel problème. Il serait naturel d’effectuer la sélection des interfaces sur la base des seules valeurs moyennes de trafic. D’un point de vue théorique, on garantit donc uniquement : $\mathbb{E} \left[\sum_{(r,i) \in RI} t_{ri} x_{ri} \right] \geq \alpha \cdot \mathbb{E} \left[\sum_{(r,i) \in RI} t_{ri} \right]$. En pratique, on observe que la solution ainsi obtenue est significativement différente de celles résultant de l’approche probabiliste, comme le montre la figure 7.3. Plus précisément, les décisions issues de cette approche en moyenne semblent très optimistes : elles conduisent à sélectionner un nombre de routeurs inférieur même à ce que l’approche probabiliste donne pour une probabilité $p = 0,5$ (très faible donc). Il y a de fortes chances pour que la décision ainsi prise soit invalide la plupart du temps, du fait des variations de trafic.

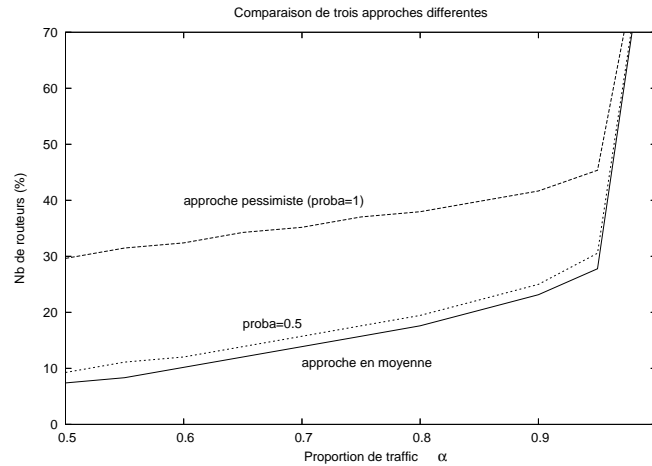


FIG. 7.3 – Comparaison de trois approches différentes.

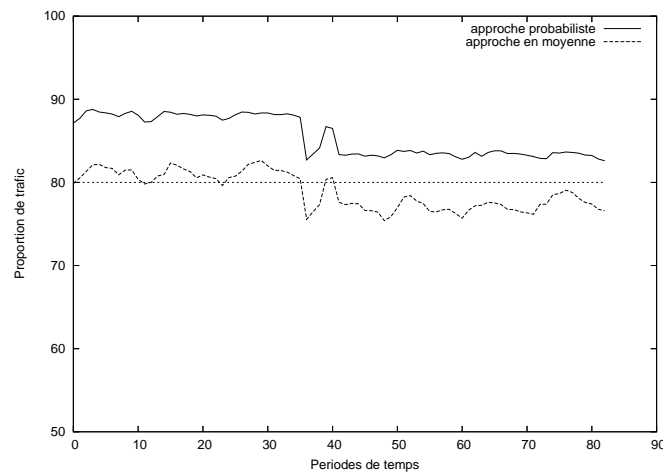


FIG. 7.4 – Trafic observé pendant une semaine, selon la sélection des routeurs considérée.

Au-delà de cette observation des solutions calculées, il faut s’assurer que l’objectif visé est bien atteint : les points de mesure sélectionnés assurent-ils l’observation du trafic voulu ? La solution calculée à partir des données du premier jour est ensuite évaluée sur les 6 jours suivants. On cherche à observer $\alpha = 80\%$ du trafic avec une probabilité d’au moins $p=0,9$ (*i.e.* $\varepsilon = 0,1$). On compare à nouveau l’approche probabiliste et l’approche en moyenne : la première requiert 21,3% des routeurs (donc 23 routeurs), contre 17,6% pour la seconde (donc 19 routeurs). Les quantités de trafic mesurées sont représentées sur la figure 7.4. Ce graphique montre que l’approche probabiliste apporte une robustesse évidente à la solution. En particulier, au troisième jour (5 octobre), une modification du trafic apparaît qui invalide complètement l’approche en moyenne. Elle est au contraire bien “absorbée” par la solution probabiliste.

Une observation plus précise du nombre de périodes où l’on a effectivement observé 80% du trafic total montre que la décision en moyenne est insatisfaisante aussi dès les premiers jours. Le tableau 7.1 indique par exemple que, même pour le premier jour, le placement effectué avec une analyse en moyenne ne satisfait pas nos exigences : on observe la quantité de trafic désirée sur seulement 83% des périodes de temps, alors que notre objectif était de 90% (puisque $p=0,9$).

TAB. 7.1 – Nombre de périodes où plus de $\alpha = 80\%$ du trafic est observé.

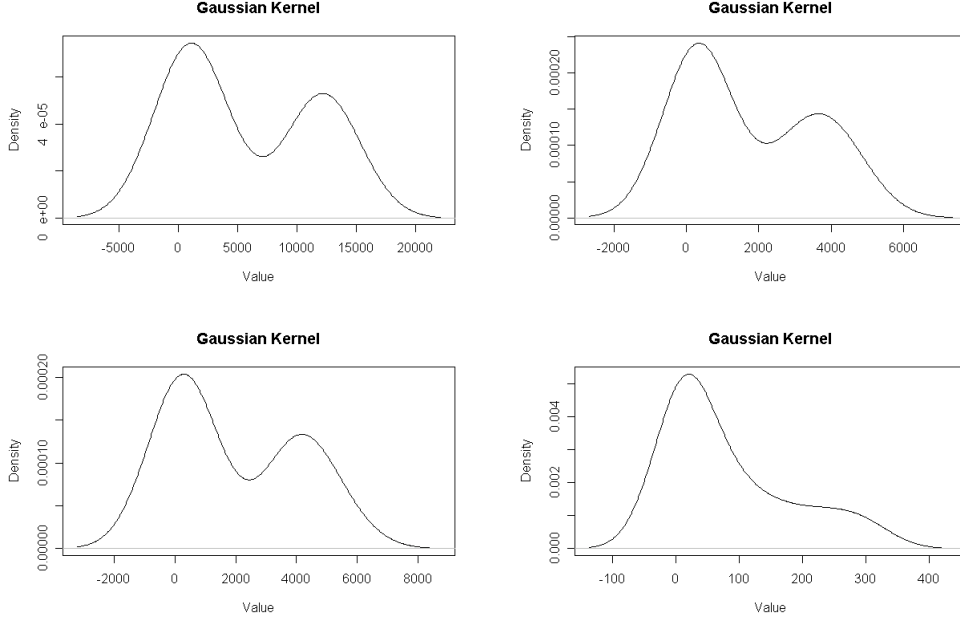
		j0	j+1	j+2	j+3	j+4	j+5	j+6
approche probabiliste	nb total de périodes	12	12	12	12	12	12	12
	périodes bien mesurées	12	12	12	12	12	12	12
	proportion	100%	100%	100%	100%	100%	100%	100%
approche en moyenne	nb total de périodes	12	12	12	12	12	12	12
	périodes bien mesurées	10	11	12	2	0	0	0
	proportion	83%	92%	100%	17%	0%	0%	0%

À partir du troisième jour, comme déjà remarqué, cette stratégie ne permet presque jamais l'observation des 80% de trafic souhaités.

7.5 Pistes pour une analyse empirique et la simulation de Monte-Carlo

On s'est appuyé jusqu'ici sur l'inégalité de Hoeffding pour évaluer les probabilités. Cette inégalité est très générale en ce qu'elle ne suppose presque rien sur la distribution des variables aléatoires. En revanche, elle rend difficile la prise en compte de corrélations entre ces variables aléatoires. Une autre approche, plus fine, consisterait à s'appuyer sur une description empirique plus spécifique de la distribution du trafic sur une interface. De plus, la connaissance de distributions de probabilité permettrait d'effectuer des simulations de Monte-Carlo de notre système. Ceci pourrait notamment permettre de valider la solution retenue (validation réalisée en haut sur la base de mesures réelles qui étaient disponibles sur une période de temps assez longue, à savoir une semaine).

On décrit maintenant comment mettre en œuvre de telles démarches. Grâce à l'outil statistique en ligne développé par P. Wessa [86], on fait une estimation de la densité de probabilité des trafics (effectuée ici selon la méthode du noyau, avec un noyau gaussien ; voir par exemple [83]). On observe que pour notre instance, le trafic sur les interfaces est généralement bimodal. À titre d'illustration, on donne les quatre densités suivantes, arbitrairement choisies :



On peut alors chercher à modéliser de manière analytique ces distributions, afin de pouvoir effectuer des simulations. Ces simulations peuvent être un moyen d'évaluer la probabilité de réalisabilité d'une solution (cf. méthode de Monte-Carlo).

On suppose connaître les lois marginales de chaque trafic $t_{r,i}$, déduites de l'observation des données (cf. méthode du noyau par exemple). Dans notre cas, compte tenu des observations effectuées, on suppose que la loi marginale de t_k (trafic sur l'interface k) est la somme de deux gaussiennes :

$$f_k(u) = \frac{a}{\sigma_1 \sqrt{2\pi}} \exp\left(-\frac{(u - \mu_1)^2}{\sigma_1^2}\right) + \frac{b}{\sigma_2 \sqrt{2\pi}} \exp\left(-\frac{(u - \mu_2)^2}{\sigma_2^2}\right),$$

où a et b sont à déterminer. Comme f est une densité de probabilité, on doit avoir : $a + b = 1$. On note T le nombre de périodes de mesures, et $t_k^{(p)}$ le trafic mesuré sur l'interface k lors de la période $p \in \{1, \dots, T\}$. La moyenne de t_k est : $\mathbb{E}[t_k] = a\mu_1 + b\mu_2$. On pose par ailleurs : $\mathbb{E}[t_k] = 1/T \cdot \sum_{p=1}^T t_k^{(p)}$, μ_1 la moyenne des valeurs de $\{t_k^{(p)}\}_{1 \leq p \leq T}$ inférieures à $\mathbb{E}[t_k]$ et μ_2 la moyenne des valeurs de $\{t_k^{(p)}\}_{1 \leq p \leq T}$ supérieures à $\mathbb{E}[t_k]$. On peut alors en déduire les valeurs de a et de b .

Par ailleurs, la variance de t_k est : $\mathbb{E}[t_k^2] - \mathbb{E}[t_k]^2 = a^2\sigma_1^2 + b^2\sigma_2^2$. Cette valeur est connue à partir des mesures : $\mathbb{E}[t_k^2] - \mathbb{E}[t_k]^2 = 1/T \cdot \sum_{p=1}^T (t_k^{(p)} - \mathbb{E}[t_k])^2$. Pour déterminer σ_1 et σ_2 , on évalue le rapport σ_1/σ_2 en calculant les variances respectives des mesures au-dessus et en-dessous de la moyenne $\mathbb{E}[t_k]$.

On simule alors un vecteur de trafic t avec des méthodes de Monte-Carlo séquentielles [29] : t_1 est choisi selon $f(t_1)$, t_2 selon $f(t_2|t_1)$, ..., t_k selon $f(t_k|t_1, \dots, t_{k-1})$, etc. On suppose une dépendance linéaire simple entre t_k et $\{t_l\}_{l \leq k-1}$:

$$t_k = \nu_k \cdot \sum_{l \leq k-1} t_l,$$

où ν_k est supposée normalement distribuée. Noter que cette relation linéaire peut être restreinte

aux seules valeurs de trafics $\{t_l\}_{l \leq k-1}$ fortement corrélées à t_k . On a de manière empirique :

$$\begin{cases} \mathbb{E}[\nu_k] = 1/T \cdot \sum_{p=1}^T \frac{t_k^{(p)}}{\sum_{l \leq k-1} t_l^{(p)}} \\ \text{var}(\nu_k) = 1/T \cdot \sum_{p=1}^T \left(\frac{t_k^{(p)}}{\sum_{l \leq k-1} t_l^{(p)}} - \mathbb{E}[\nu_k] \right)^2 \end{cases} \cdot$$

Sous ces hypothèses, on obtient :

$$f(t_k | t_1, \dots, t_{k-1}) = \frac{1}{\sigma_k \sqrt{2\pi}} \exp\left(-\frac{(t_k - \mu_k^2)^2}{\sigma_k^2}\right)$$

avec :

$$\begin{cases} \mu_k = \mathbb{E}[\nu_k] \cdot \left(\sum_{l \leq k-1} t_l \right) \\ \sigma_k^2 = \text{var}(\nu_k) \cdot \left(\sum_{l \leq k-1} t_l \right)^2 \end{cases} \cdot$$

7.6 Conclusion

Le problème de sélection de points de mesure NetFlow dans un réseau a été décrit et modélisé de manière à tenir compte des variations de trafic au cours du temps. Le modèle probabiliste développé est résolu à l'aide d'une heuristique s'appuyant sur la résolution d'une succession de problèmes robustes. Tant l'analyse théorique que les tests pratiques réalisés montrent l'efficacité de la démarche proposée, et sa nécessité par rapport à des approches plus naïves. La méthodologie développée est très générale, et peut s'appliquer indifféremment avec d'autres technologies de mesure, et dans d'autres contextes (réseaux électriques, routiers, etc.).

Chapitre 8

Dimensionnement de réseaux d'accès

Motivations : Dans ce dernier chapitre, on aborde une problématique de placement de fonctions de concentration dans un réseau de télécommunications. Il s'agit d'un problème typique de planification d'investissements reposant sur des prévisions de trafic, nécessairement incertaines. Une modélisation avec contraintes probabilistes est proposée. Le problème obtenu est trop difficile pour être résolu de manière exacte, d'où le recours à un algorithme simple de recuit simulé qui permet d'obtenir facilement des solutions réalisables.

8.1 Introduction

La planification d'investissements dans les réseaux est une tâche incontournable et récurrente pour un opérateur de télécommunications. Elle s'appuie sur les perspectives d'évolution des services, du nombre de clients et de leurs usages, et/ou du volume global de trafic. Il s'agit alors de décider du dimensionnement adéquat qui permettra de répondre aux évolutions du marché tout en étant aussi compétitif que possible. Il est généralement admis que les enjeux économiques se situent de nos jours plutôt au niveau des réseaux d'accès que des réseaux de cœur. Ces derniers sont en effet souvent déjà largement équipés.

Dans le présent travail, on se propose d'étudier la localisation et le dimensionnement de sites de raccordements, qui donnent aux sites clients un accès au réseau-cœur tout en effectuant alors une concentration des trafics. Ce problème, central, a déjà été largement traité dans la littérature, et le nombre de modèles proposés est important ; on renvoie à [35] pour un aperçu de différentes approches dans ce domaine. Dans le présent chapitre, on s'appuiera sur le modèle classique de localisation de *hub*. En effet, il a l'avantage de coupler le dimensionnement côté cœur et côté accès pour les sites de concentration. Cette famille de modèles a déjà été utilisée dans des contextes de télécommunications, où le dimensionnement modulaire des sites et/ou des liens joue un rôle particulier (voir par exemple [22, 49, 87, 70]).

Pour effectuer une planification de réseau pertinente, il est nécessaire de tenir compte des incertitudes inévitablement liées aux prévisions. Les modèles de localisation en contexte incertain abondent dans la littérature ; on en trouvera une vue d'ensemble dans [77]. Beaucoup d'entre eux font appel aux techniques d'optimisation robuste (couramment utilisées aussi pour la synthèse et le dimensionnement de réseaux, voir par exemple [51, 61, 50, 60, 62]). Parmi les modèles stochastiques de localisation, on trouve quelques usages plutôt rares de contraintes probabilistes (par exemple [76], ou encore [21], cité dans [77]).

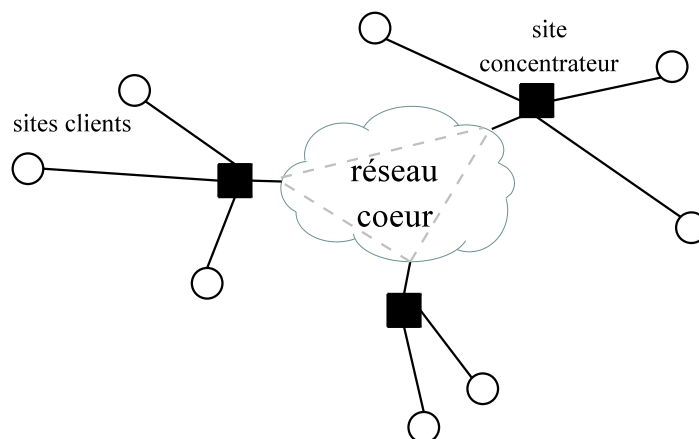
Dans ce chapitre, des modèles mathématiques avec contraintes de probabilités sont développés

en s'appuyant principalement sur des hypothèses gaussiennes. Les problèmes obtenus sont trop difficiles pour être résolus de manière optimale pour des instances réalistes de plusieurs centaines de sites. On définit donc un algorithme simple de recuit simulé qui permet d'obtenir en un temps raisonnable des solutions réalisables. À partir d'une instance construite à l'aide de données réelles, on analyse le gain économique que la prise en compte d'une incertitude sur les trafics peut entraîner.

8.2 Problème traité et modèles mathématiques

8.2.1 Modèle déterministe de base

Le modèle présenté est une extension du modèle classique de localisation de *hub*. On note I l'ensemble des nœuds du réseau. Tous les nœuds sont supposés être des *sites clients*, c'est-à-dire source ou destination de trafic. Le trafic entre la source $i \in I$ et la destination $j \in I$ est noté t_{ij} . Certains des nœuds du réseau doivent être sélectionnés pour être des *sites concentrateurs*. Chaque site client doit être rattaché à un site concentrateur, ou être lui-même un site concentrateur. Ceci se modélise à l'aide d'une variable binaire x_{ik} qui vaut 1 si le client i est relié au site concentrateur k , 0 sinon. Un nœud i est choisi comme site concentrateur si $x_{ii} = 1$. Les sites concentrateurs sont reliés entre eux par un réseau cœur qui n'est pas connu précisément.



Équipements : Les sites concentrateurs doivent être équipés de manière à pouvoir écouler le trafic voulu. Pour cela, on dispose de cartes d'accès (*access*) et de cartes de cœur (*backbone*), caractérisées de la manière suivante :

- une carte d'accès comporte N_a ports de capacité unitaire K_a ,
- une carte de cœur comporte N_b ports de capacité unitaire K_b .

Avec le souci de simplifier l'étude, on ne modélise pas les châssis nécessaires en pratique. Ceci est motivé par le fait que, souvent, le coût des châssis est négligeable devant celui des cartes.

La variable y_k^a (resp. y_k^b) désigne le nombre de cartes d'accès (resp. de cœur) installées sur le site concentrateur k .

Coûts : On note A_{ik} le coût de raccordement du site client i au site concentrateur k . Ce coût est supposé forfaitaire, mais nos modèles peuvent très facilement prendre en compte un coût linéaire en fonction du débit entre i et k . B_{kl} est le coût d'écoulement du trafic entre les sites concentrateurs k et l dans le réseau de cœur ; ce coût est supposé proportionnel à la quantité de trafic écoulee entre k et l . Enfin, pour un site concentrateur k , c_k^a est le coût d'une carte d'accès

et c_k^b est le coût d'une carte de cœur.

L'objectif est de minimiser le coût global :

$$\min \sum_{(i,k) \in I^2} A_{ik} x_{ik} + \sum_{k \in I} (c_k^a y_k^a + c_k^b y_k^b) + \sum_{(k,l) \in I^2} B_{kl} z_{kl},$$

où z_{kl} est le trafic allant du site concentrateur k au site concentrateur l dans le réseau cœur. On détaille maintenant les contraintes du problème :

- contraintes d'affectation : $\begin{cases} \sum_{k \in I} x_{ik} = 1, & \forall i \in I \\ x_{ik} \leq x_{kk}, & \forall (i,k) \in I^2 \end{cases}$
- contraintes de dimensionnement d'un site concentrateur $k \in I$:
côté accès : $N_a y_k^a \geq \sum_{i \in I} \max \left\{ \left\lceil \sum_{j \in I} t_{ij} / K_a \right\rceil, \left\lceil \sum_{j \in I} t_{ji} / K_a \right\rceil \right\} \cdot x_{ik}$
côté cœur : $\begin{cases} K_b r_k \geq \sum_{i \in I} \left(\sum_{j \in J} t_{ij} \cdot [1 - x_{jk}] \right) \cdot x_{ik} \\ K_b r_k \geq \sum_{i \in I} \left(\sum_{j \in J} t_{ji} \cdot [1 - x_{jk}] \right) \cdot x_{ik} \\ N_b y_k^b \geq r_k. \end{cases}$

On suppose ainsi un dimensionnement symétrique, effectué pour le maximum des trafics entrant et sortant du site concentrateur. La variable intermédiaire r_k désigne le nombre de ports nécessaires côté cœur.

Le problème global s'écrit donc :

$$\min \sum_{(i,k) \in I^2} A_{ik} x_{ik} + \sum_{k \in I} (c_k^a y_k^a + c_k^b y_k^b) + \sum_{(k,l) \in I^2} B_{kl} z_{kl} \quad (8.1)$$

$$\text{s.c.} \quad \sum_{k \in I} x_{ik} = 1, \quad \forall i \in I, \quad (8.2)$$

$$x_{ik} \leq x_{kk}, \quad \forall (i,k) \in I^2, \quad (8.3)$$

$$N_a y_k^a \geq \sum_{i \in I} \max \left\{ \left\lceil \sum_{j \in I} t_{ij} / K_a \right\rceil, \left\lceil \sum_{j \in I} t_{ji} / K_a \right\rceil \right\} \cdot x_{ik}, \quad \forall k \in I, \quad (8.4)$$

$$K_b r_k \geq \sum_{i \in I} \left(\sum_{j \in J} t_{ij} \cdot [1 - x_{jk}] \right) \cdot x_{ik}, \quad \forall k \in I, \quad (8.5)$$

$$K_b r_k \geq \sum_{i \in I} \left(\sum_{j \in J} t_{ji} \cdot [1 - x_{jk}] \right) \cdot x_{ik}, \quad \forall k \in I, \quad (8.6)$$

$$N_b y_k^b \geq r_k, \quad \forall k \in I, \quad (8.7)$$

$$z_{kl} = \sum_{(i,j) \in I^2} t_{ij} x_{ik} x_{jl}, \quad \forall (k,l) \in I^2, \quad (8.8)$$

$$x_{ik} \in \{0, 1\}, y_k^a \in \mathbb{N}, y_k^b \in \mathbb{N}, r_k \in \mathbb{N}, z_{kl} \geq 0, \quad \forall (i,k,l) \in I^3. \quad (8.9)$$

Le dimensionnement des sites concentrateurs et des liens de cœur fait intervenir des contraintes quadratiques. On pourrait linéariser ce modèle classiquement, en introduisant des variables $u_{ikjl} = x_{ik} x_{jl}$, associées aux contraintes $u_{ikjl} \geq x_{ik} + x_{jl} - 1$ et $u_{ikjl} \geq 0$. Mais le nombre très important de variables et de contraintes requis par une telle approche la rend presque impossible à mettre en œuvre.

On remarque que les contraintes de dimensionnement des cartes côté cœur peuvent aussi

s'écrire à l'aide des variables z_{kl} , ce qui conduit au problème global :

$$\begin{aligned}
\min \quad & \sum_{(i,k) \in I^2} A_{ik} x_{ik} + \sum_{k \in I} (c_k^a y_k^a + c_k^b y_k^b) + \sum_{(k,l) \in I^2} B_{kl} z_{kl} \\
\text{s.c.} \quad & \sum_{k \in I} x_{ik} = 1, & \forall i \in I, \\
& x_{ik} \leq x_{kk}, & \forall (i,k) \in I^2, \\
N_a y_k^a \geq \sum_{i \in I} \max \left\{ \left\lceil \frac{\sum_{j \in I} t_{ij}}{K_a} \right\rceil, \left\lceil \frac{\sum_{j \in I} t_{ji}}{K_a} \right\rceil \right\} \cdot x_{ik}, & \forall k \in I, \\
K_b r_k \geq \sum_{l \in I \setminus \{k\}} z_{kl}, & \forall k \in I, \\
K_b r_k \geq \sum_{l \in I \setminus \{k\}} z_{lk}, & \forall k \in I, \\
N_b y_k^b \geq r_k, & \forall k \in I, \\
z_{kl} = \sum_{(i,j) \in I^2} t_{ij} x_{ik} x_{jl}, & \forall (k,l) \in I^2, \\
x_{ik} \in \{0, 1\}, y_k^a \in \mathbb{N}, y_k^b \in \mathbb{N}, r_k \in \mathbb{N}, z_{kl} \geq 0, & \forall (i,k,l) \in I^3.
\end{aligned} \tag{8.10}$$

On distingue les formulations (8.1)-(8.9) et (8.10), car si la deuxième est plus simple à écrire et à mettre en œuvre, elle ne sera pas applicable lorsqu'on introduira des incertitudes dans le modèle.

Remarque : Pour les contraintes (8.8), on peut remarquer que (cf. [49]) :

$$\sum_{(i,j) \in I^2} t_{ij} x_{ik} x_{jl} = \max_{D \subseteq I^2} \sum_{(i,j) \in D} t_{ij} (x_{ik} + x_{jl} - 1).$$

Ainsi, les inégalités (8.8) peuvent être remplacées de manière équivalente par :

$$z_{kl} \geq \sum_{(i,j) \in D} t_{ij} (x_{ik} + x_{jl} - 1), \quad \forall D \subseteq I^2. \tag{8.11}$$

Ces inégalités linéaires sont en nombre exponentiel, mais on sait les séparer en temps polynomial $O(n^2)$. On évite ainsi l'usage des variables-produits u_{ikjl} .

En fait, ce problème apparaît trop compliqué pour être traité *via* la programmation mathématique pour des instances réelles de grandes tailles. Quelques tests réalisés sur nos instances montrent que même la résolution du relâché continu du problème est très difficile.

8.2.2 Modélisation des incertitudes

Ce problème de planification doit le plus souvent s'appuyer sur des données incertaines. Les coûts d'équipements ont tendance à décroître dans le temps, il est difficile de prédire leur valeur exacte au moment de l'investissement. Les coûts de raccordement et de transport du trafic sont souvent très complexes à évaluer. Enfin, les données en trafic s'appuient sur des prévisions statistiques dont la fiabilité est connue pour être faible.

Les coûts et les trafics sont maintenant considérés comme étant des variables aléatoires. Pour prendre en compte la variabilité sur les coûts, on cherche à minimiser le coût moyen de la solution :

$$\min \mathbb{E} \left[\sum_{(i,k) \in I^2} A_{ik} x_{ik} + \sum_{k \in I} (c_k^a y_k^a + c_k^b y_k^b) + \sum_{(k,l) \in I^2} B_{kl} z_{kl} \right].$$

Par linéarité de l'espérance, cet objectif devient en fait :

$$\min \sum_{(i,k) \in I^2} \mathbb{E}[A_{ik}] x_{ik} + \sum_{k \in I} (\mathbb{E}[c_k^a] y_k^a + \mathbb{E}[c_k^b] y_k^b) + \sum_{(k,l) \in I^2} \mathbb{E}[B_{kl} z_{kl}].$$

En effet, comme les trafics sont incertains, la variable z l'est aussi (cf. équation (8.8) : z modélise le trafic sur le cœur du réseau). Ceci dit, on sait que les coûts B et les trafics z

sont stochastiquement indépendants : $\mathbb{E}[B_{kl}z_{kl}] = \mathbb{E}[B_{kl}].\mathbb{E}[z_{kl}]$. Ainsi, il suffit pour les coûts de connaître leur valeur moyenne.

On souhaite que le dimensionnement décidé soit réalisable avec une probabilité $1 - \varepsilon$ que l'on peut contrôler. On garantit ainsi une bonne robustesse du réseau, essentielle pour assurer une qualité de service satisfaisante. Les inégalités (8.4)-(8.6) sont ainsi remplacées par :

$$\mathbb{P} \left(\begin{array}{l} N_a y_k^a \geq \sum_{i \in I} \max \left\{ \left\lceil \sum_{j \in I} t_{ij} / K_a \right\rceil, \left\lceil \sum_{j \in I} t_{ji} / K_a \right\rceil \right\} .x_{ik}, \quad \forall k \in I, \\ K_b r_k \geq \sum_{i \in I} \left(\sum_{j \in J} t_{ij} \cdot [1 - x_{jk}] \right) .x_{ik}, \quad \forall k \in I, \\ K_b r_k \geq \sum_{i \in I} \left(\sum_{j \in J} t_{ji} \cdot [1 - x_{jk}] \right) .x_{ik}, \quad \forall k \in I \end{array} \right) \geq 1 - \varepsilon. \quad (8.12)$$

Observer que cette contrainte se concentre sur la caractérisation du dimensionnement des sites concentrateurs. On aurait pu imposer des garanties au niveau des raccordements des clients en raisonnant sur le nombre de ports à attribuer à chaque client plutôt que sur le nombre de cartes côté accès. On privilégie ici une approche plus macroscopique.

La contrainte de probabilité jointe (8.12) est très difficile à traiter, et elle n'est d'ailleurs pas forcément la plus pertinente en pratique. Il peut être judicieux de distinguer les garanties souhaitées pour le dimensionnement côté cœur et côté accès :

$$\begin{aligned} \mathbb{P} \left(N_a y_k^a \geq \sum_{i \in I} \max \left\{ \left\lceil \sum_{j \in I} t_{ij} / K_a \right\rceil, \left\lceil \sum_{j \in I} t_{ji} / K_a \right\rceil \right\} .x_{ik}, \quad \forall k \in I \right) &\geq 1 - \varepsilon_a, \\ \mathbb{P} \left(\begin{array}{l} K_b r_k \geq \sum_{(i,j) \in I^2} t_{ij} \cdot [1 - x_{jk}] .x_{ik}, \quad \forall k \in I, \\ K_b r_k \geq \sum_{(i,j) \in I^2} t_{ji} \cdot [1 - x_{jk}] .x_{ik}, \quad \forall k \in I \end{array} \right) &\geq 1 - \varepsilon_b. \end{aligned} \quad (8.13)$$

Même sous cette forme (8.13), le programme stochastique demeure très difficile à étudier (contraintes probabilistes jointes). On s'intéresse dans cette partie à un modèle avec contraintes probabilistes complètement séparées, qui est une relaxation de (8.13) :

$$\mathbb{P} \left(N_a y_k^a \geq \sum_{i \in I} \max \left\{ \left\lceil \sum_{j \in I} t_{ij} / K_a \right\rceil, \left\lceil \sum_{j \in I} t_{ji} / K_a \right\rceil \right\} .x_{ik} \right) \geq 1 - \varepsilon_a, \quad \forall k \in I, \quad (8.14)$$

$$\mathbb{P} \left(K_b r_k \geq \sum_{(i,j) \in I^2} t_{ij} \cdot [1 - x_{jk}] .x_{ik} \right) \geq 1 - \varepsilon_b, \quad \forall k \in I, \quad (8.15)$$

$$\mathbb{P} \left(K_b r_k \geq \sum_{(i,j) \in I^2} t_{ji} \cdot [1 - x_{jk}] .x_{ik} \right) \geq 1 - \varepsilon_b, \quad \forall k \in I. \quad (8.16)$$

En séparant complètement les contraintes probabilistes comme on vient de le faire, on contrôle les garanties site par site.

8.3 Hypothèses probabilistes

8.3.1 Présentation d'un cadre gaussien

On remarque déjà que le calcul du coût d'une solution requiert la connaissance des moyennes pour tous les trafics $\{t_{ij}\}_{(i,j) \in I^2}$, puisque :

$$\mathbb{E}[z_{kl}] = \sum_{(i,j) \in I^2} \mathbb{E}[t_{ij}] x_{ik} x_{jl}.$$

Pour transformer le problème stochastique, on propose un cadre probabiliste gaussien. Les répartitions normales peuvent paraître naturelles pour modéliser les incertitudes dans notre contexte. On peut se référer au théorème de la limite centrale (en particulier le théorème 20 donné dans l'annexe B.2), qui permet de penser que les lois associées aux sommes de variables

aléatoires de notre problème sont bien approchées par des gaussiennes. On note, pour tout $(i, j) \in I^2$, μ_{ij} la moyenne de t_{ij} et σ_{ij} son écart-type. On supposera alors, pour tout $k \in I$:

$$\sum_{(i,j) \in I^2} t_{ij} \cdot (1 - x_{jk}) \cdot x_{ik} \sim \mathcal{N} \left(\sum_{(i,j) \in I^2} \mu_{ij} \cdot (1 - x_{jk}) \cdot x_{ik}, \sum_{(i,j) \in I^2} \sigma_{ij}^2 \cdot (1 - x_{jk}) \cdot x_{ik} \right).$$

Cette hypothèse n'est justifiée par le théorème 20 que si les trafics $\{t_{ij}\}_{(i,j) \in I^2}$ sont tous indépendants entre eux, et que le nombre de clients reliés à un site est "assez grand". De la même manière, on a :

$$\sum_{(i,j) \in I^2} t_{ji} \cdot (1 - x_{jk}) \cdot x_{ik} \sim \mathcal{N} \left(\sum_{(i,j) \in I^2} \mu_{ji} \cdot (1 - x_{jk}) \cdot x_{ik}, \sum_{(i,j) \in I^2} \sigma_{ji}^2 \cdot (1 - x_{jk}) \cdot x_{ik} \right).$$

Sous ces hypothèses, les contraintes (8.15) et (8.16) du dimensionnement côté cœur ont des équivalents déterministes classiques, qui seront explicités plus loin.

Pour le dimensionnement côté accès, on introduit les variables aléatoires $\{s_i\}_{i \in I}$ définies par :

$$s_i = \max \left\{ \left\lceil \sum_{j \in I} t_{ij} / K_a \right\rceil, \left\lceil \sum_{j \in I} t_{ji} / K_a \right\rceil \right\}.$$

s_i représente le nombre de ports requis par le site client i . On a, pour tout $q \in \mathbb{N}^*$:

$$\mathbb{P}(s_i \leq q) = \mathbb{P} \left(\sum_{j \in I} t_{ij} \leq q \cdot K_a \text{ et } \sum_{j \in I} t_{ji} \leq q \cdot K_a \right). \quad (8.17)$$

Il est alors possible de pré-calculer la distribution discrète de chaque s_i . Cependant, l'introduction directe de telles distributions discrètes dans le modèle le rend très difficile à traiter.

Pour pallier cette difficulté, on peut supposer que la somme $\sum_{i \in I} s_i x_{ik}$ (pour tout site concentrateur k) suit une loi normale $\mathcal{N}(\mu_k^a(x), (\sigma_k^a(x))^2)$, avec (cf. théorème 20) :

$$\begin{cases} \mu_k^a(x) = \sum_{i \in I} \mathbb{E}[s_i] \cdot x_{ik}, \\ (\sigma_k^a(x))^2 = \sum_{i \in I} \text{var}(s_i) \cdot x_{ik}. \end{cases}$$

Cette approche est bien sûr approximative, mais elle est facile à mettre en œuvre. Ce modèle gaussien est d'autant plus justifié que le nombre de sites clients raccordés à k est important. Pour tout $i \in I$, les valeurs $\mathbb{E}[s_i]$ et $\text{var}(s_i)$ peuvent en pratique être évaluées en s'appuyant sur la formule (8.17). Supposons en effet que les trafics entrants et sortants d'un site $i \in I$ soient stochastiquement indépendants, on a :

$$\mathbb{P}(s_i \leq q) = \mathbb{P} \left(\sum_{j \in I} t_{ij} \leq q \cdot K_a \right) \cdot \mathbb{P} \left(\sum_{j \in I} t_{ji} \leq q \cdot K_a \right).$$

Conformément aux hypothèses déjà effectuées, on a : $\sum_{j \in I} t_{ij} \sim \mathcal{N} \left(\sum_{j \in I} \mu_{ij}, \sum_{j \in I} \sigma_{ij}^2 \right)$. Ceci signifie :

$$\mathbb{P} \left(\sum_{j \in I} t_{ij} \leq q \cdot K_a \right) = \Phi \left(\frac{qK_a - \sum_{j \in I} \mu_{ij}}{\sqrt{\sum_{j \in I} \sigma_{ij}^2}} \right),$$

où Φ désigne la fonction de répartition de la loi normale standard. En appliquant le même raisonnement pour les trafics entrants, on obtient :

$$\mathbb{P}(s_i \leq q) = \Phi \left(\frac{qK_a - \sum_{j \in I} \mu_{ij}}{\sqrt{\sum_{j \in J} \sigma_{ij}^2}} \right) \cdot \Phi \left(\frac{qK_a - \sum_{j \in I} \mu_{ji}}{\sqrt{\sum_{j \in J} \sigma_{ji}^2}} \right).$$

De la fonction de répartition de s_i , on déduit aisément la distribution de s_i , *i.e.* les valeurs $\mathbb{P}(s_i = q)$ pour $q \in \mathbb{N}$. Une bonne approximation de cette loi consiste à ne considérer que $q \geq 0$, et à arrêter les évaluations lorsque $\mathbb{P}(s_i \leq q)$ est “suffisamment” proche de 1 (ce dernier critère étant laissé à l’appréciation du décideur). On peut alors en déduire la moyenne et la variance de s_i .

Remarque 1 : Une autre approche pourrait être d’assurer que le nombre de ports $r'_i \in \mathbb{N}$ requis pour chaque client i satisfait une contrainte de probabilité : $\mathbb{P}(r'_i \geq s_i) \geq 1 - \varepsilon_a$. On peut alors utiliser la caractérisation probabiliste exacte (8.17) pour calculer r'_i . On obtient alors une approximation “sûre” de la contrainte (8.14). Les $\{r'_i\}_{i \in I}$ peuvent être pré-calculés, faisant des contraintes probabilistes (8.14) des contraintes de dimensionnement classiques : $\forall k \in I, N_a y_k^a \geq \sum_{i \in I} r'_i x_{ik}$. Cette garantie contrôlée au niveau de chaque site client peut parfois être pertinente. Cependant, elle peut paraître très contraignante au regard des incertitudes de modélisation et de données.

Remarque 2 : Il faut mentionner aussi un cas fréquent en pratique, où chaque client requiert un et un seul port en accès. Une telle situation simplifie considérablement l’analyse et annule les problèmes de modélisation évoqués.

8.3.2 Modèle déterministe équivalent

On suppose dans cette partie que les trafics sont stochastiquement indépendants. On commence par traiter le cas des contraintes (8.14), qui correspondent au dimensionnement des sites côté accès. Soit $k \in I$, on sait que :

$$\mathbb{P} \left(N_a y_k^a \geq \sum_{i \in I} s_i \cdot x_{ik} \right) \geq 1 - \varepsilon_a \Leftrightarrow \sum_{i \in I} \mathbb{E}[s_i] x_{ik} + \Phi^{-1}(1 - \varepsilon_a) \cdot \sqrt{\sum_{i \in I} \text{var}(s_i) x_{ik}} \leq N_a y_k^a,$$

où Φ désigne la fonction de répartition de la distribution normale standard $\mathcal{N}(0, 1)$ (cf. § 6.2.1). De la même manière, les contraintes (8.15) et (8.16) de dimensionnement côté cœur peuvent s’écrire, pour tout $k \in I$:

$$\Leftrightarrow \begin{cases} \mathbb{P} \left(K_b r_k \geq \sum_{(i,j) \in I^2} t_{ij} \cdot (1 - x_{jk}) \cdot x_{ik} \right) \geq 1 - \varepsilon_b \\ \mathbb{P} \left(K_b r_k \geq \sum_{(i,j) \in I^2} t_{ji} \cdot (1 - x_{jk}) \cdot x_{ik} \right) \geq 1 - \varepsilon_b \\ \sum_{(i,j) \in I^2} \mu_{ij} (1 - x_{jk}) x_{ik} + \Phi^{-1}(1 - \varepsilon_b) \cdot \sqrt{\sum_{(i,j) \in I^2} \sigma_{ij}^2 (1 - x_{jk}) x_{ik}} \leq K_b r_k \\ \sum_{(i,j) \in I^2} \mu_{ji} (1 - x_{jk}) x_{ik} + \Phi^{-1}(1 - \varepsilon_b) \cdot \sqrt{\sum_{(i,j) \in I^2} \sigma_{ji}^2 (1 - x_{jk}) x_{ik}} \leq K_b r_k \end{cases}.$$

Ces différentes équations déterministes équivalentes semblent très difficiles à prendre en compte. Elles sont non-linéaires, et ne peuvent pas être linéarisées simplement. La résolution exacte du problème en nombres entiers résultant semble hors de portée des techniques actuelles de programmation mathématique.

8.3.3 Prise en compte de corrélations

On montre dans ce paragraphe comment transformer la contrainte probabiliste (8.15), avec $k \in I$ fixé, pour intégrer des corrélations. Le même travail peut être réalisé pour les équations (8.14) et (8.16).

Covariances connues : Si l'on se place dans le cadre gaussien, et que les covariances sont connues entre les trafics, alors le modèle déterministe équivalent s'écrit (voir e.g. [84]) :

$$\sum_{(i,j) \in I^2} \mu_{ij}(1-x_{jk})x_{ik} + \Phi^{-1}(1-\varepsilon_b) \cdot \sqrt{\sum_{(i,j) \in I^2} \sum_{(a,b) \in I^2} \sigma_{ij,ab}(1-x_{jk})x_{ik} \cdot (1-x_{bk})x_{ak}} \leq K_b r_k, \quad (8.18)$$

où $\sigma_{ij,ab}$ désigne la covariance entre t_{ij} et t_{ab} .

Corrélations fortes : Si l'on sait que les trafics seront très fortement corrélés entre eux, on peut par exemple s'appuyer sur l'hypothèse suivante :

$$\forall (i,j) \in I^2, t_{ij} = \mu_{ij} + \sigma_{ij} \cdot \varphi,$$

où φ est une variable aléatoire centrée (par exemple, $\varphi \sim \mathcal{N}(0,1)$). L'analyse est alors plus simple que précédemment. Considérons par exemple le cas de la contrainte (8.15) pour $k \in I$ donné :

$$\begin{aligned} & \mathbb{P} \left(K_b r_k \geq \sum_{(i,j) \in I^2} t_{ij} \cdot (1-x_{jk})x_{ik} \right) \geq 1 - \varepsilon_b \\ \Leftrightarrow & \mathbb{P} \left(\varphi \leq \frac{K_b r_k - \sum_{(i,j) \in I^2} \mu_{ij} \cdot (1-x_{jk})x_{ik}}{\sum_{(i,j) \in I^2} \sigma_{ij} (1-x_{jk})x_{ik}} \right) \geq 1 - \varepsilon_b. \end{aligned}$$

Si l'on note F la fonction de répartition de φ , et F^{-1} son inverse défini par :

$$F^{-1}(p) = \min\{x | F(x) \geq p\},$$

on a :

$$\frac{K_b r_k - \sum_{(i,j) \in I^2} \mu_{ij} \cdot (1-x_{jk})x_{ik}}{\sum_{(i,j) \in I^2} \sigma_{ij} (1-x_{jk})x_{ik}} \geq F^{-1}(1 - \varepsilon_b).$$

8.3.4 Exemple d'un autre contexte probabiliste : l'inégalité de Hoeffding

On a détaillé jusqu'ici un cadre d'étude gaussien. D'autres approches sont bien sûr possibles, et peuvent être plus pertinentes selon les cas pratiques à traiter. On peut par exemple s'appuyer sur l'inégalité de Hoeffding (cf. théorème 12). Les seules données nécessaires sont alors, pour chaque trafic t_{ij} , ses valeurs minimale, maximale, et moyenne. Les nombres moyens de ports d'accès requis pour chaque client i , $\mathbb{E}[s_i]$, peuvent être spécifiés directement (les valeurs minimale et maximale sont déduites des hypothèses sur les trafics). On notera que :

$$\max \left\{ \mathbb{E} \left[\sum_{i \in I} t_{ij} \right] / K_a, \mathbb{E} \left[\sum_{i \in I} t_{ji} \right] / K_a \right\} \leq \mathbb{E}[s_i] \leq 1 + \max \left\{ \mathbb{E} \left[\sum_{i \in I} t_{ij} \right] / K_a, \mathbb{E} \left[\sum_{i \in I} t_{ji} \right] / K_a \right\}.$$

Cette modélisation peut s'avérer plus facile à mettre en œuvre. En effet, avoir une intuition des valeurs extrémales de trafic est souvent plus naturel que de déterminer des variances. En revanche, la principale hypothèse du théorème de Hoeffding, à savoir l'indépendance des variables aléatoires, n'est pas satisfaite la plupart du temps en pratique. On peut cependant espérer que ce biais soit compensé par la quasi-absence d'hypothèses quant aux distributions

de ces variables aléatoires (cf. par exemple les observations des parties 7.4.2 et 7.4.3, et aussi l'annexe B.3).

Des développements complètement similaires à ceux de la partie 8.3.2 sont possibles ici. À titre d'exemple, on se concentre sur la contrainte de dimensionnement (8.14), à $k \in I$ donné :

$$\mathbb{P} \left(N_a y_k^a \geq \sum_{i \in I} s_i \cdot x_{ik} \right) \geq 1 - \varepsilon_a.$$

Comme les trafics sont supposés bornés, le nombre de ports requis pour un client i l'est également : $\underline{s}_i \leq s_i \leq \bar{s}_i$. Pour pouvoir utiliser l'inégalité de Hoeffding, on impose (cf. théorème 12) :

$$N_a y_k^a \geq \sum_{i \in I} \mathbb{E}[s_i] \cdot x_{ik}.$$

On a alors, en notant $\delta_i = \bar{s}_i - \underline{s}_i$:

$$\mathbb{P} \left(N_a y_k^a \leq \sum_{i \in I} s_i \cdot x_{ik} \right) \leq \exp \left(\frac{-2 (N_a y_k^a - \sum_{i \in I} \mathbb{E}[s_i] \cdot x_{ik})^2}{\sum_{i \in I} \delta_i^2 \cdot x_{ik}} \right).$$

L'inégalité probabiliste (8.14) devient alors :

$$\begin{aligned} -2 (N_a y_k^a - \sum_{i \in I} \mathbb{E}[s_i] \cdot x_{ik})^2 &\leq \ln(\varepsilon_a) \cdot \sum_{i \in I} \delta_i^2 \cdot x_{ik} \\ \Leftrightarrow N_a y_k^a - \sum_{i \in I} \mathbb{E}[s_i] \cdot x_{ik} &\geq \sqrt{-\ln(\varepsilon_a)/2} \cdot \sqrt{\sum_{i \in I} \delta_i^2 \cdot x_{ik}}. \end{aligned}$$

8.4 Relaxation du problème probabiliste pour le cas gaussien

À titre d'illustration, on décrit comment les idées de la partie 6.3.2 pourraient être utilisées ici pour construire une relaxation du problème sous contraintes probabilistes. Vue la difficulté de résolution du problème initial, même sans incertitudes, les modèles décrits dans ce paragraphe n'ont pas été utilisés numériquement.

On construit un scénario de trafics \underline{t} , aussi grand que possible, tel que la valeur de la solution optimale associée est une borne inférieure sur la valeur du problème probabiliste considéré. Comme les variables aléatoires sont toutes normalement distribuées, leur distribution est symétrique, et donc $\mathbb{E}[t]$ constitue un scénario minorant (ou scénario basique selon la terminologie du chapitre 6 ; cf. lemme 41).

On se concentre dans un premier temps sur la contrainte probabiliste (8.15), pour $k \in I$ fixé. On veut calculer la quantité, pour $\mathcal{I}_1 \subseteq I^2$:

$$D(\mathcal{I}_1) = \sup \left\{ d \mid \mathbb{P} \left(\sum_{(i,j) \in \mathcal{I}_1} t_{ij} - \mathbb{E}[t_{ij}] \leq d \right) < 1 - \varepsilon_b \right\}.$$

Noter que cette expression ne dépend pas de k . Pour tout $(i,j) \in \mathcal{I}_1$, la variable aléatoire $\eta_{ij} = t_{ij} - \mathbb{E}[t_{ij}]$ est centrée, de variance σ_{ij}^2 . D'après nos hypothèses probabilistes : $\sum_{(i,j) \in \mathcal{I}_1} \eta_{ij} \sim \mathcal{N} \left(0, \sum_{(i,j) \in \mathcal{I}_1} \sigma_{ij}^2 \right)$. On note $F_{\mathcal{I}_1}(d)$ la fonction de répartition de $\sum_{(i,j) \in \mathcal{I}_1} \eta_{ij}$, et alors :

$$D(\mathcal{I}_1) = F_{\mathcal{I}_1}^{-1}(1 - \varepsilon_b).$$

($F_{\mathcal{I}_1}$ est strictement croissante et continue, donc son inverse est définie classiquement, sans ambiguïté.)

On définit alors, pour tout $c \in \{1, \dots, n^2\}$ (cf. partie 6.3.2, formule (6.8) avec $v = (1, \dots, 1)$) :

$$\lambda(c) = \frac{\min \{D(\mathcal{I}_1) \mid \mathcal{I}_1 \subseteq I^2, |\mathcal{I}_1| = c\}}{c} = 1/c \cdot \min \{F_{\mathcal{I}_1}^{-1}(1 - \varepsilon_b) \mid \mathcal{I}_1 \subseteq I^2, |\mathcal{I}_1| = c\}.$$

$\lambda(c)$ se calcule aisément. En effet, le minimum est atteint pour un ensemble \mathcal{I}_1 de cardinal c et tel que la variance $\sum_{i \in \mathcal{I}_1} \sigma_{ij}^2$ est minimale. On construit donc \mathcal{I}_1 avec les c paires (i, j) de plus petites variances σ_{ij}^2 . Enfin, $F_{\mathcal{I}_1}^{-1}$ est l'inverse d'une fonction de répartition gaussienne : cette fonction est disponible dans la plupart des bibliothèques informatiques relatives aux probabilités (voir par exemple la GSL [31]).

Le nouveau scénario minorant pour la contrainte (8.15) est alors défini par :

$$\forall (i, j) \in I^2 : \underline{t}_{ij} = \mu_{ij} + \min_{1 \leq c \leq n} \lambda(c).$$

La contrainte probabiliste (8.15), pour un quelconque $k \in I$, peut donc être remplacée par l'inégalité :

$$\sum_{(i,j) \in I} \underline{t}_{ij} \cdot (1 - x_{jk}) x_{ik} \leq K_b r_k.$$

On a alors une relaxation du problème probabiliste initial.

On effectue les mêmes opérations pour les équations (8.14) et (8.16). Pour (8.14), on note \underline{s}_i le nombre minimum de cartes requis à l'accès pour chaque site i , calculé selon le procédé utilisé ci-dessus. Les trafics minimaux calculés pour (8.15) sont notés $\underline{t}_{ij}^{(1)}$, ceux obtenus pour (8.16) sont notés $\underline{t}_{ij}^{(2)}$. Observer que si les données sont symétriques, *i.e.* si t_{ij} et t_{ji} suivent la même loi, alors $\underline{t}_{ij}^{(1)} = \underline{t}_{ij}^{(2)}$.

On peut construire à partir de \underline{s} , $\underline{t}_{ij}^{(1)}$ et $\underline{t}_{ij}^{(2)}$ une relaxation du problème probabiliste :

$$\begin{aligned} \min \quad & \sum_{(i,k) \in I^2} A_{ik} x_{ik} + \sum_{k \in I} (c_k^a y_k^a + c_k^b y_k^b) + \sum_{(k,l) \in I^2} B_{kl} z_{kl} \\ \text{s.c.} \quad & \sum_{k \in I} x_{ik} = 1, & \forall i \in I, \\ & x_{ik} \leq x_{kk}, & \forall (i, k) \in I^2, \\ & N_a y_k^a \geq \sum_{i \in I} \underline{s}_i \cdot x_{ik}, & \forall k \in I, \\ K_b r_k \geq & \sum_{i \in I} \left(\sum_{j \in J} \underline{t}_{ij}^{(1)} \cdot [1 - x_{jk}] \right) \cdot x_{ik}, & \forall k \in I, \\ K_b r_k \geq & \sum_{i \in I} \left(\sum_{j \in J} \underline{t}_{ji}^{(2)} \cdot [1 - x_{jk}] \right) \cdot x_{ik}, & \forall k \in I, \\ & N_b y_k^b \geq r_k, & \forall k \in I, \\ & z_{kl} = \sum_{(i,j) \in I^2} \mathbb{E}[t]_{ij} x_{ik} x_{jl}, & \forall (k, l) \in I^2, \\ x_{ik} \in \{0, 1\}, & y_k^a \in \mathbb{N}, y_k^b \in \mathbb{N}, r_k \in \mathbb{N}, z_{kl} \geq 0, & \forall (i, k, l) \in I^3. \end{aligned}$$

Si l'on veut se ramener à un problème nominal de la forme (8.10), il faut considérer le scénario moyen $\mu = \mathbb{E}[t]$ (remarquer que $\underline{t}^{(1)} \geq \mu$ et $\underline{t}^{(2)} \geq \mu$). On a alors la relaxation suivante du problème probabiliste, plus faible que la précédente, mais moins difficile à résoudre :

$$\begin{aligned} \min \quad & \sum_{(i,k) \in I^2} A_{ik} x_{ik} + \sum_{k \in I} (c_k^a y_k^a + c_k^b y_k^b) + \sum_{(k,l) \in I^2} B_{kl} z_{kl} \\ \text{s.c.} \quad & \sum_{k \in I} x_{ik} = 1, & \forall i \in I, \\ & x_{ik} \leq x_{kk}, & \forall (i, k) \in I^2, \\ & N_a y_k^a \geq \underline{s}_i \cdot x_{ik}, & \forall k \in I, \\ K_b r_k \geq & \sum_{l \in I \setminus \{k\}} z_{kl}, & \forall k \in I, \\ K_b r_k \geq & \sum_{l \in I \setminus \{k\}} z_{lk}, & \forall k \in I, \\ & N_b y_k^b \geq r_k, & \forall k \in I, \\ & z_{kl} = \sum_{(i,j) \in I^2} \mu_{ij} x_{ik} x_{jl}, & \forall (k, l) \in I^2, \\ x_{ik} \in \{0, 1\}, & y_k^a \in \mathbb{N}, y_k^b \in \mathbb{N}, r_k \in \mathbb{N}, z_{kl} \geq 0, & \forall (i, k, l) \in I^3. \end{aligned}$$

8.5 Tests numériques

8.5.1 Algorithme de résolution

Les programmes mathématiques formulés jusqu'ici semblent tous hors de portée des techniques actuelles de résolution, même pour la forme déterministe initiale (8.10). Pour proposer tout-de-même des solutions réalisables, on s'appuie sur un recuit simulé. Cette méta-heuristique a l'avantage d'être très simple à mettre en œuvre ; l'inconvénient d'une telle approche est bien sûr l'absence totale de garanties quant à la qualité du résultat obtenu.

On définit le *voisinage* d'une solution courante. À partir des raccordements x , on peut calculer le coût global d'une solution, en utilisant les équations données auparavant. La définition du voisinage se fait donc à partir de ces seules variables x . On dira que x' est une solution voisine de x si ces deux solutions ne diffèrent que par un seul raccordement ; c'est-à-dire, il existe un unique site client i dont le site de concentration sera différent dans les solutions x et x' .

À chaque itération du recuit simulé, on choisit aléatoirement une solution voisine de la solution courante. Si le coût correspondant à cette nouvelle solution est meilleur que celui de la solution courante, alors on accepte ce déplacement et on change de solution courante. Si le nouveau coût est au contraire supérieur au coût courant, le déplacement n'est accepté qu'avec une probabilité égale à $\exp(-|\Delta c|/T)$, où T désigne la *température* et Δc est la différence entre nouveau coût et coût actuel (critère de Metropolis).

La température est initialisée de manière à garantir un taux d'acceptation des déplacements de l'ordre de 90% (voir [4] pour une étude relative à cette question de l'initialisation de la température). On effectue des paliers de 5000 itérations. D'un palier à l'autre, la température est multipliée par 0,9. L'algorithme s'arrête lorsqu'il y a eu moins de 5% de déplacements acceptés au cours d'un palier.

Noter que beaucoup d'heuristiques peuvent être envisagées. On renvoie à [22] pour la comparaison de différentes approches pour un problème de localisation de *hub* avec contraintes de capacité.

8.5.2 Description des données

On travaille sur la topologie d'un réseau Gigabit Ethernet de France Télécom comportant 276 sites clients. Parmi ceux-ci, 85 sont éligibles pour devenir sites de concentration. On compte 42556 liens d'accès possibles pour relier les sites clients aux sites concentrateurs. Les données en trafic sont construites à partir de prévisions réelles effectuées pour des services résidentiels (internet, voix sur IP, vidéo à la demande, télévision). On effectue une mise à l'échelle de ces trafics pour faciliter l'analyse : la demande maximale à un trafic de 10, et la valeur moyenne des demandes est 1 (voir la figure 8.1 pour une représentation graphique de la répartition des demandes). Le trafic total issu d'un site a une valeur maximale d'environ 910, et une valeur moyenne d'environ 280. Les trafics sont symétriques : pour tout $(i, j) \in I^2$, $t_{ij} = t_{ji}$.

Les coûts de raccordement des sites clients aux sites concentrateurs sont directement issus de données réelles. On suppose ici que les coûts de transport sur le réseau cœur sont nuls. On suppose un rapport de 4 entre les capacités des ports d'accès et des ports de cœur : $K_b = 4.K_a$. Une carte d'accès comporte $N_a = 8$ ports, contre $N_b = 2$ ports pour une carte de cœur. Le coût d'une carte de cœur est 0,6 fois celui d'une carte d'accès. On ajuste les valeurs de manière à assurer un équilibre réaliste entre coûts de commutation et coûts d'accès.

Chaque prévision de trafic t_{ij} entre les sites i et j est initialement un scalaire, noté \bar{t}_{ij} . On le transforme en une variable aléatoire de distribution gaussienne. La distribution sera caractérisée par une moyenne $\mu_{ij} \leq \bar{t}_{ij}$ et un écart-type $\sigma_{ij} = (\bar{t}_{ij} - \mu_{ij})/3$. Cette dernière définition assure

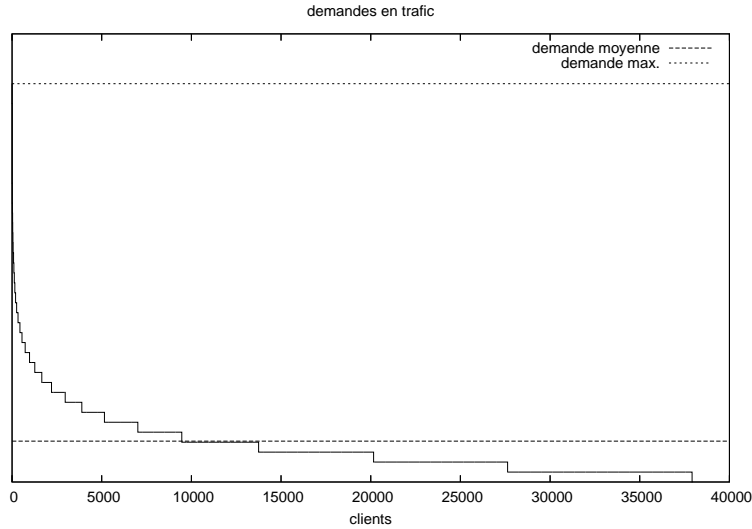


FIG. 8.1 – Répartition des demandes par volume décroissant.

que la probabilité que t_{ij} excède la valeur prévue, \bar{t}_{ij} , est très faible : $\mathbb{P}(t_{ij} \geq \bar{t}_{ij}) < 0,0015$. On testera différentes moyennes de trafic : $\mu_{ij} = \alpha \cdot \bar{t}_{ij}$, avec $\alpha \in \{50\%, 60\%, 70\%, 80\%, 90\%\}$. On désignera avec $\alpha = 100\%$ le dimensionnement pour la valeur prévue \bar{t} . On représente sur la figure 8.2 ces différentes répartitions.

8.5.3 Résultats obtenus

Les résultats obtenus sont récapitulés dans le tableau 8.1. Quelques données sont représentées graphiquement sur la figure 8.3. Comme on pouvait s’y attendre, la prise en compte d’incertitudes tend à faire baisser le coût du réseau. C’est tout particulièrement le cas pour le dimensionnement des sites de concentration (cf. évolutions du coût des cartes de cœur, par exemple), ce qui est normal : il est directement lié à la quantité de trafic écoulee. Le coût global diminue de 5% pour une hypothèse de $\alpha = 80\%$, de 15% lorsque $\alpha = 50\%$. On rappelle que dans toutes les solutions trouvées, on garantit que les valeurs prévisionnelles du trafic seront satisfaites avec une probabilité de l’ordre de 0,9.

L’observation la plus surprenante concerne le cas $\alpha = 90\%$, valeur pour laquelle le coût global de la solution est supérieur à celui obtenu à partir des simples valeurs prévisionnelles ($\alpha = 100\%$). Une analyse plus détaillée montre que ceci est en partie dû à la modélisation stochastique des ports d’accès. La somme des ports d’accès sur un site k , soit $\sum_{i \in I} s_i x_{ik}$, a été approchée par une loi normale (voir la partie 8.3.1). Or cette loi continue et symétriquement distribuée autour de sa moyenne rend parfois assez mal compte de la forme de la distribution du nombre de ports requis, et peut conduire dans certains cas à le sur-évaluer. Cet effet indésirable est notamment lié au nombre trop faible de clients raccordés à un même site de concentration (cf. le théorème de la limite centrale). Il peut donc être utile en pratique d’être attentif aux hypothèses de modélisation stochastique du nombre de ports requis pour les sites clients.

Ceci étant dit, le fait que la résolution ne soit pas optimale est aussi une des raisons de cette observation. On a été capable de trouver une autre solution, pour $\alpha = 90\%$, d’un coût total de 207398 ; ce coût reste supérieur à celui de la solution pour $\alpha = 100\%$, mais en est très proche.

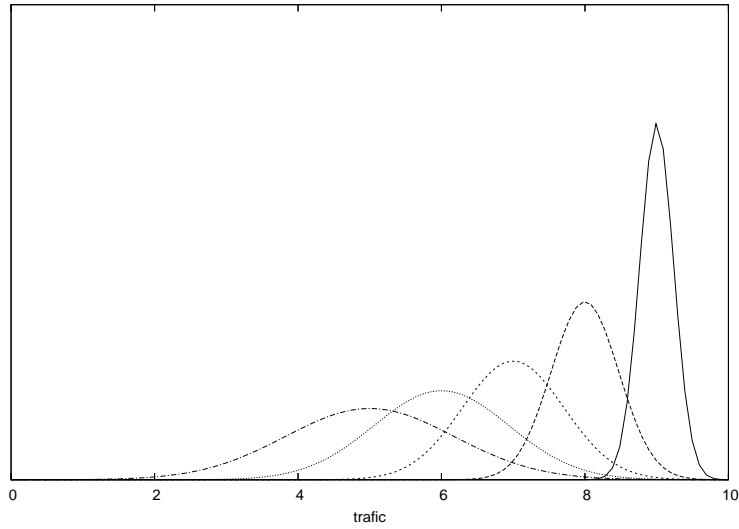


FIG. 8.2 – Transformation d'une prévision de trafic $\bar{t} = 10$ en distributions gaussiennes.

TAB. 8.1 – Récapitulatif des résultats obtenus.

α		50%	60%	70%	80%	90%	100%
coût	liens accès	105398	110804	112549	109250	123899	112478
	cartes accès	46000	51000	52000	56000	58000	60000
	cartes cœur	26400	28800	30000	32400	33600	34800
	total	177798	190604	194549	197650	215499	207278
équipements	nb de sites	43	48	50	53	53	54
	nb de cartes accès	46	51	52	56	58	60
	nb de cartes cœur	44	48	50	54	56	58
temps (sec.)		1893	1863	1773	1880	1551	1559

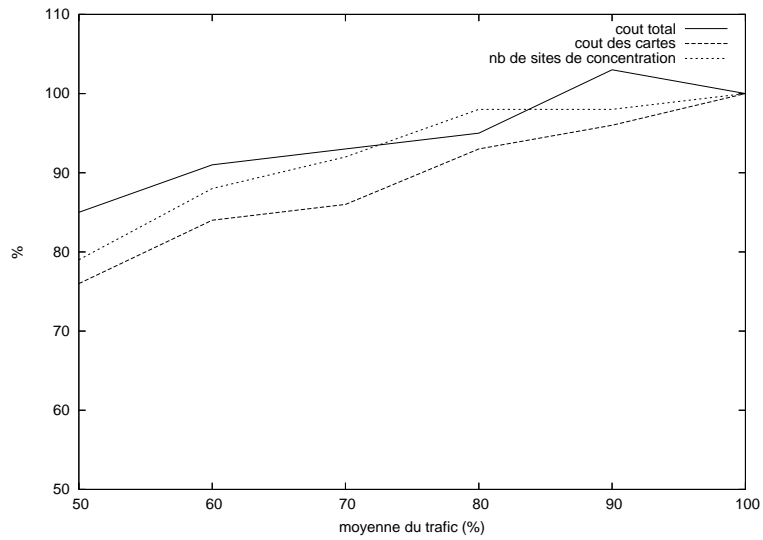


FIG. 8.3 – Comparaisons des résultats, en %, selon la valeur de α (moyenne du trafic).

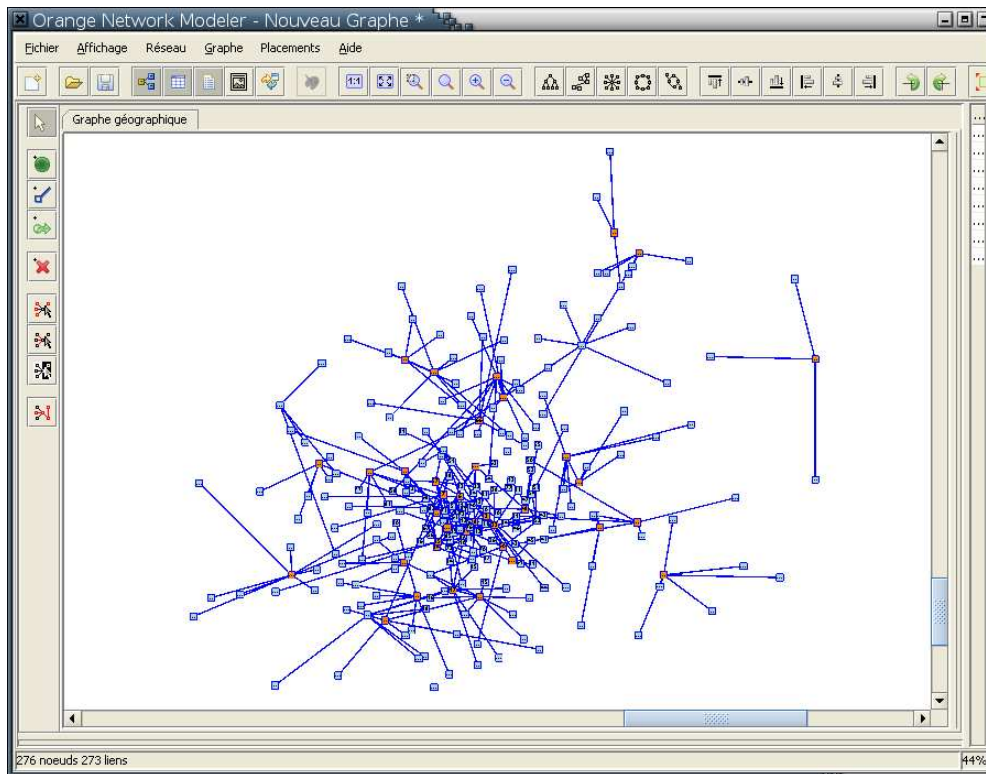


FIG. 8.4 – Capture d’écran de l’outil *Network Modeler*, utilisé pour visualiser les résultats.

8.6 Conclusion

On a étudié un problème de localisation de sites de concentration dans un réseau de télécommunications. Le modèle utilisé, très proche des modèles de localisation de *hub*, permet de prendre en compte les trafics côté accès et côté cœur pour le dimensionnement des sites et le calcul du coût final. Le modèle est adapté de manière à prendre en compte des incertitudes sur les trafics entre sites clients. On s’est principalement intéressé à des hypothèses gaussiennes pour modéliser l’incertitude dans le réseau.

Le problème traité, même dans sa version sans incertitudes, est trop difficile pour être abordé de manière exacte pour des instances réalistes de quelques centaines de sites. Pour permettre néanmoins une prise de décision, on produit des solutions réalisables à l’aide d’un algorithme simple de recuit simulé. Des tests numériques sont effectués à partir de données réalistes. Le dimensionnement obtenu garantit que les prévisions de trafic seront satisfaites avec une probabilité d’environ 0,9. L’économie effectuée par rapport au dimensionnement nominal prévu dépend de la valeur moyenne des trafics. Sur l’instance utilisée ici, on gagne 15% du coût global du réseau si la moyenne des trafics est égale à la moitié de leur valeur nominale prévue.

Chapitre 9

Conclusions et perspectives

Le travail mené dans cette thèse a concerné l'optimisation combinatoire avec données incertaines. Un des thèmes centraux a été le rapport entre optimisation robuste et programmation sous contraintes probabilistes. L'optimisation robuste est depuis assez longtemps utilisée pour prendre en compte des incertitudes dans le cas de problèmes avec variables de décision entières. Or les approches polyédriques ont permis d'améliorer grandement la résolution de programmes mathématiques en nombres entiers. On s'est donc attaché à dégager quelques propriétés polyédriques d'un modèle robuste proposé par Bertsimas et Sim, jugé particulièrement pertinent pour de nombreux contextes appliqués. Ce travail ouvre la voie à de nombreuses études polyédriques de problèmes particuliers. On peut légitimement penser que la résolution de problèmes robustes s'en trouvera facilitée, et avec elle la prise de décision avec données incertaines.

En prenant appui sur l'optimisation robuste, connue pour être relativement aisée à résoudre, des algorithmes de résolution heuristique de problèmes sous contraintes probabilistes ont été élaborés. Les programmes mathématiques avec contraintes probabilistes sont connus pour être difficiles à traiter, et ce constat n'est pas évidemment arrangé par l'introduction de variables de décision entières. Les heuristiques reposent sur la résolution successive d'une série de problèmes robustes. Cette approche devrait permettre de s'intéresser à des problèmes concrets relativement complexes. L'efficacité algorithmique de la démarche se trouve encore améliorée si l'on fait usage du modèle robuste extrêmement simple que l'on a proposé. La principale question qui demeure concerne la qualité des solutions produites. Quelques réflexions théoriques ont permis d'établir l'optimalité de la démarche pour certains problèmes très simples. Mais il semble probable que pour des problèmes plus compliqués, même si l'optimalité n'est pas garantie, on puisse quantifier l'écart à l'optimum. Dit autrement, il serait intéressant de déterminer les contextes dans lesquels les heuristiques itératives proposées sont des algorithmes d'approximation.

En parallèle de ces approches heuristiques s'appuyant sur l'optimisation robuste, on a élaboré un cadre algorithmique nouveau pour la résolution optimale de problèmes combinatoires sous contraintes probabilistes. Telle quelle, l'approche s'avère efficace pour des problèmes de tailles moyennes, si l'on fait des hypothèses quelque peu restrictives sur les distributions des variables aléatoires. En dehors de ces hypothèses, l'algorithme s'applique, mais de manière beaucoup moins satisfaisante. Il est clair que le travail effectué n'est qu'un début, et mériterait d'être affiné. Par exemple, une réflexion devrait être menée sur les moyens de déterminer de bons scénarios basiques pour des contextes quelconques. La séparation des inégalités linéaires caractérisant le problème doit aussi être améliorée pour ce cadre général. Par ailleurs, l'étude a été réalisée pour des variables 0-1. On a dit qu'elle s'appliquait aussi aux cas de variables entières générales, mais ceci reste assez théorique, puisque le nombre de variables 0-1 nécessitées serait alors considérable. La prise en compte spécifique de variables entières générales s'impose.

Un des sous-produits intéressants de cette approche optimale est l'obtention facilitée de bornes sur le problème, grâce notamment au scénario basique. De telles bornes sont essentielles pour évaluer la qualité des approches heuristiques. Si un travail théorique sur l'obtention de meilleures bornes est bien sûr souhaitable, il serait utile dans un premier temps de réaliser des études numériques des différentes approches heuristiques existant. Ceci serait du plus grand intérêt pour la pratique de l'aide à la décision, puisqu'on peut espérer donner de bonnes intuitions de la qualité des différents algorithmes envisageables.

Enfin, et pour revenir au champ d'application de cette thèse, à savoir les télécommunications, il nous semble que beaucoup reste à faire dans ce domaine. En pratique, la prise en compte d'incertitudes dans les analyses reste souvent très rudimentaire, et même la littérature scientifique demeure relativement restreinte pour ce champ d'applications. Les problèmes de planification de réseaux, en particulier, mériteraient d'être maintenant revus à la lumière des avancées théoriques et algorithmiques proposées dans cette thèse ou ailleurs.

Publications

On a mis en gras le titre des publications en rapport direct avec le travail de cette thèse. Sauf mention expresse, tous les articles ci-dessous sont déjà publiés, ou acceptés pour publication.

Revues

É. Gourdin et O. Klopfenstein, Multi-period Capacitated Location with Modular Equipments, *Computers & Operations Research*, 35 : 661-682, 2008.

O. Klopfenstein, Rerouting Tunnels for MPLS Network Resource Optimization, *European Journal of Operational Research*, 188(1) : 293-312, 2008.

D. Nace, L.N. Doan, O. Klopfenstein et A. Bashllari, Max-Min Fairness in multi-commodity flows, *Computers & Operations Research*, 35(2) : 557-573, 2008.

O. Klopfenstein et D. Nace, **A robust approach to the chance-constrained knapsack problem**, à paraître dans *Operations Research Letters*.

O. Klopfenstein, **Solving chance-constrained combinatorial problems to optimality**, à paraître dans *Computational Optimization and Applications*.

Soumission en cours :

O. Klopfenstein, **Tractable algorithms for chance-constrained combinatorial problems**, soumis à *RAIRO-OR*.

Conférences internationales avec comité de lecture

O. Klopfenstein. Sequential Rerouting of Demands for Network resource Optimization. *Proc. of the International Network Optimization Conference (INOC)*, 2005 (Lisbonne, Portugal).

D. Nace et O. Klopfenstein. On the lexicographically minimum loaded networks. *Proc. of the International Network Optimization Conference (INOC)*, 2005 (Lisbonne, Portugal).

O. Klopfenstein. A Randomized Rounding Heuristic to Reroute Tunnels in MPLS Networks. *Proc. of the International Workshop on Design of Reliable Communication Networks (DRCN)*, 2005 (île d'Ischia, Italie).

É. Gourdin et O. Klopfenstein, Comparison of different QoS-oriented Objectives for Multicommodity Flow Routing Optimization, *Proc. of the International Conference on Telecommunications (ICT)*, 2006 (île de Madère, Portugal).

O. Klopfenstein et S. Mamy. Choosing Weights for IP Network Dimensioning Optimization. *Proc. of the International Symposium on Computers and Communications (ISCC)*, 2006 (Cagliari, Italie).

A. Bashllari, D. Nace, É. Gourdin et O. Klopfenstein. The MMF rerouting computation problem. *Proc. of the International Network Optimization Conference (INOC)*, 2007 (Spa, Belgique).

A. Bashllari, D. Nace, É. Gourdin et O. Klopfenstein. A simple robust routing scheme. *Proc. of the International Workshop on Design of Reliable Communication Networks (DRCN)*, 2007 (La Rochelle, France).

O. Klopfenstein. Robust pre-provisioning of local protection resources in MPLS networks. *Proc. of the International Workshop on Design of Reliable Communication Networks (DRCN)*, 2007 (La Rochelle, France).

M. Bouhtou et O. Klopfenstein. **Robust Optimization for Selecting NetFlow Points of Measurement in an IP Network**. *Proc. of Globecom 2007*, 2007 (Washington DC, USA).

O. Klopfenstein, **Access network dimensioning with uncertain traffic forecasts**, *Proc. of the International Telecommunications Network Strategy and Planning Symposium (Networks'08)*, 2008 (Budapest, Hongrie).

O. Klopfenstein, D. Nace, **The robust flight level assignment problem**, *Proc. of the International Conference on Research in Air Transportation (ICRAT)*, 2008 (Fairfax, Virginia, USA).

Annexes

Annexe A

Amélioration des résultats de la partie 3.3.2

On rappelle qu'une fonction $F : \mathbb{R}^m \rightarrow \mathbb{R}$ est dite *super-additive* si :

$$\forall (a_1, a_2) \in \mathbb{R}^m \times \mathbb{R}^n, F(a_1 + a_2) \geq F(a_1) + F(a_2).$$

Observer en particulier que $\lfloor \cdot \rfloor : \mathbb{R} \rightarrow \mathbb{R}$ est super-additive. On rappelle ci-dessous quelques faits très classiques concernant l'usage de fonctions super-additives pour améliorer des inégalités valides en programmation en nombres entiers.

Proposition 21 *Soit $F : \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction décroissante et super-additive. Tout point entier $x \in \mathbb{Z}^n$ qui vérifie $Ax \leq b$, avec $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, vérifie également : $\sum_{i \in I} F(A_i)x_i \leq F(b)$ (où A_i désigne la colonne i de la matrice A).*

Preuve : Comme F est croissante, on a : $F(Ax) = F(\sum_{i \in I} A_i x_i) \leq F(b)$. Comme de plus F est super-additive : $\sum_{i \in I} F(A_i x_i) \leq F(Ax)$. Finalement, puisque $x \in \mathbb{Z}^n$, la super-additivité de F entraîne aussi : $\sum_{i \in I} F(A_i)x_i \leq F(b)$. \square

Une fonction super-additive souvent utilisée pour renforcer des inégalités valides est :

$$\begin{aligned} F_\rho : \mathbb{R} &\rightarrow \mathbb{R} \\ a &\mapsto \lfloor a \rfloor + (a - \lfloor a \rfloor - \rho)^+ / (1 - \rho) \end{aligned}$$

pour $\rho \in]0, 1[$.

Lemme 52 F_ρ est super-additive et croissante sur \mathbb{R} .

On renvoie à [56] pour la preuve. Ainsi, étant donnée une inégalité valide pour le programme en nombres entiers : $\sum_{i \in I} a_i x_i \geq b$, on peut en déduire une autre inégalité valide : $\sum_{i \in I} F_\rho(a_i)x_i \geq F_\rho(b)$. On voit facilement que pour $\rho \geq b - \lfloor b \rfloor$, cette dernière inégalité est plus forte que celle obtenue en arrondissant les coefficients à l'entier inférieur. La plus forte de ces inégalités est obtenue pour : $\rho = b - \lfloor b \rfloor$.

Annexe B

Notes sur le calcul de probabilités

B.1 Distribution de la somme de deux variables aléatoires

À titre d'illustration, on détaille ici un calcul de probabilité pour une somme de deux variables aléatoires, qui correspond au dernier exemple de la partie 5.4.2, quand $0 < 6\gamma \leq 1$. Le but est d'évaluer la probabilité $\mathbb{P}(w_1 + w_3 \leq 3,5)$, où w_1 et w_3 sont uniformément distribués sur, respectivement, $[1; 2]$ et $[1,5; 5,5]$. On note leurs densités de probabilité respectives f_1 et f_3 .

La densité f_{1+3} de la variable aléatoire $w_1 + w_3$ est obtenue par convolution de f_1 et f_3 (cf. e.g. [36]) : $f_{1+3}(t) = \int_{-\infty}^{\infty} f_1(u)f_3(t-u)du = \int_1^2 f_3(t-u)du$. Si $t-u \notin [1,5; 5,5]$, alors $f_3(t-u) = 0$; sinon, $f_3(t-u) = 1/4$. Plusieurs cas doivent être distingués :

- si $t < 2,5$, alors pour tout $u \in [1, 2]$, $t-u < 1,5$: $f_{1+3}(t) = 0$.
- Si $2,5 \leq t < 3,5$: $f_{1+3}(t) = \int_1^2 f_3(t-u)du = \int_1^{t-1,5} 1/4du = (t-2,5)/4$.
- Si $3,5 \leq t < 6,5$: $f_{1+3}(t) = \int_1^2 f_3(t-u)du = \int_1^2 1/4du = 1/4$.
- Si $6,5 \leq t < 7,5$: $f_{1+3}(t) = \int_1^2 f_3(t-u)du = \int_{t-5,5}^2 1/4du = (7,5-t)/4$.
- Si $t \geq 7,5$, alors pour tout $u \in [1, 2]$, $t-u \geq 5,5$: $f_{1+3}(t) = 0$.

Donc : $\mathbb{P}(w_1 + w_3 \leq 3,5) = \int_{-\infty}^{3,5} f_{1+3}(t)dt = \dots = 1/8$.

B.2 Différents théorèmes de la limite centrale

On note Φ la fonction de répartition de la loi normale standard $\mathcal{N}(0, 1)$. On donne d'abord le théorème limite le plus célèbre et le plus couramment utilisé.

Théorème 19 Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires indépendantes et identiquement distribuées. On suppose l'existence d'une moyenne μ et d'une variance σ^2 finies de ces variables aléatoires. Si l'on note $S_n = \sum_{i=0}^n X_i$ et $Z_n = (S_n - n\mu)/(\sigma\sqrt{n})$, alors Z_n converge en distribution vers la loi normale standard :

$$\forall z \in \mathbb{R}, \lim_{n \rightarrow \infty} \mathbb{P}(Z_n \leq z) = \Phi(z).$$

Une telle convergence est également vraie, sous certaines conditions, quand on suppose seulement l'indépendance des variables aléatoires (voir [71]) :

Théorème 20 Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires indépendantes. On suppose que chaque X_n a une moyenne μ_n et une variance σ_n^2 finies. Si :

(i) les variables $\{X_n\}_{n \in \mathbb{N}}$ sont uniformément bornées, i.e. : il existe un réel M tel que : $\forall n \in \mathbb{N}, \mathbb{P}(|X_n| < M) = 1$;

(ii) $\sum_{n \geq 0} \sigma_n^2 = \infty$,

alors la variable aléatoire $Z_n = \sum_{i=0}^n (X_i - \mu_i) / \sqrt{\sum_{i=0}^n \sigma_i^2}$ converge en distribution vers la loi normale standard :

$$\forall z \in \mathbb{R}, \lim_{n \rightarrow \infty} \mathbb{P}(Z_n \leq z) = \Phi(z).$$

D'après [85], cette version est également valide :

Théorème 21 Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires indépendantes. On suppose que chaque X_n a une moyenne μ_n et une variance σ_n^2 finies. Notons : $s_n^2 = \sum_{i=0}^n \sigma_i^2$, et $r_n^3 = \sum_{i=0}^n \mathbb{E}[|X_i - \mu_i|^3]$. Si :

(i) pour tout $n \in \mathbb{N}$, le troisième moment r_n^3 est fini ;

(ii) $\lim_{n \rightarrow \infty} r_n/s_n = 0$ (condition de Lyapunov) ;

alors la variable aléatoire $Z_n = \sum_{i=0}^n (X_i - \mu_i) / \sqrt{\sum_{i=0}^n \sigma_i^2}$ converge en distribution vers la loi normale standard :

$$\forall z \in \mathbb{R}, \lim_{n \rightarrow \infty} \mathbb{P}(Z_n \leq z) = \Phi(z).$$

Il existe également des résultats de convergence pour certains types de corrélations entre les variables aléatoires (pour les martingales, pour les suites de variables aléatoires m -dépendantes,...).

B.3 Validité de l'inégalité de Hoeffding pour des v.a. corrélées

L'inégalité de Hoeffding requiert normalement l'indépendance des variables aléatoires pour être valide (cf. théorème 12). En pratique, cette hypothèse est souvent difficile à garantir. On montre ici que dans certains cas de corrélations fortes, cette inégalité est encore applicable, pour peu que l'on effectue des hypothèses sur la distribution des variables aléatoires.

On considère une famille $\{X_i\}_{i \in I}$ de variables aléatoires telles que, pour tout i : $X_i = \mu_i + \sigma_i \cdot \varphi$, où φ est une variable aléatoire à support dans $[0,1]$. On suppose que $\sum_{i \in I} \sigma_i \neq 0$: ce cas trivial implique que $\sum_{i \in I} X_i$ est une constante. La quantité qui nous intéresse est, pour tout $\delta > 0$:

$$B(\delta) = \exp\left(\frac{-2\delta^2}{\sum_{i \in I} \sigma_i^2}\right).$$

L'inégalité de Hoeffding est valide lorsque : $\mathbb{P}\left(\sum_{i \in I} X_i \geq \mathbb{E}\left[\sum_{i \in I} X_i\right] + \delta\right) \leq B(\delta)$.

Pour tout $\delta > 0$, on note :

$$P(\delta) = \mathbb{P}\left(\sum_{i \in I} X_i \geq \mathbb{E}\left[\sum_{i \in I} X_i\right] + \delta\right) = \mathbb{P}\left(\varphi \geq \mathbb{E}[\varphi] + \delta / \sum_{i \in I} \sigma_i\right).$$

On cherche des conditions sous lesquelles $P(\delta) \leq B(\delta)$, pour tout $\delta > 0$.

Soit : $\delta^* = (1 - \mathbb{E}[\varphi]) \cdot \sum_{i \in I} \sigma_i$. φ est à support dans $[0,1]$, donc on sait déjà que pour tout $\delta > \delta^*$: $P(\delta) = 0$, ce qui assure : $P(\delta) \leq B(\delta)$. Par ailleurs, $B(0) = 1$ donc : $P(0) \leq B(0)$.

Lemme 53 Soit $\{X_i\}_{i \in I}$ une famille de variables aléatoires telles que, pour tout $i \in I$: $X_i = \mu_i + \sigma_i \cdot \varphi$, où φ est une variable aléatoire à support dans $[0,1]$. Supposons par ailleurs qu'il existe $k \in \mathbb{N}^*$ tel que :

(i) la densité de φ est dominée sur $[\mathbb{E}[\varphi], 1]$ par la courbe $x \mapsto \alpha(1-x)^k$, avec $\alpha > 0$;

(ii) $\alpha \leq (k+1)/(1 - \mathbb{E}[\varphi])^{k+1}$;

$$(iii) \mathbb{E}[\varphi] \geq 1 - \frac{\sqrt{(k+1) \sum_{i \in I} \sigma_i^2}}{\sum_{i \in I} \sigma_i}.$$

Alors l'inégalité de Hoeffding est valide :

$$\forall \delta > 0, \mathbb{P} \left(\sum_{i \in I} X_i \geq \mathbb{E} \left[\sum_{i \in I} X_i \right] + \delta \right) \leq \exp \left(\frac{-2\delta^2}{\sum_{i \in I} \sigma_i^2} \right).$$

Preuve : D'après l'hypothèse (i), on a :

$$\begin{aligned} P(\delta) &\leq \int_{x=\mathbb{E}[\varphi]+\delta/\sum_{i \in I} \sigma_i}^1 \alpha(1-x)^k dx \\ &= \frac{\alpha}{\sum_{i \in I} \sigma_i} \cdot \int_{t=\delta}^{\delta^*} \left(1 - \mathbb{E}[\varphi] - t / \sum_{i \in I} \sigma_i \right)^k dt \\ &= \frac{\alpha}{\left(\sum_{i \in I} \sigma_i \right)^{k+1}} \cdot \int_{t=\delta}^{\delta^*} (\delta^* - t)^k dt \\ &= \frac{\alpha}{(k+1) \left(\sum_{i \in I} \sigma_i \right)^{k+1}} \cdot (\delta^* - \delta)^{k+1}. \end{aligned}$$

On note : $g(x) = B(x) - (\delta^* - x)^{k+1} \cdot \alpha / \left[(k+1) \left(\sum_{i \in I} \sigma_i \right)^{k+1} \right]$. On veut assurer que cette fonction est positive sur l'intervalle $[0, \delta^*]$. On remarque déjà que $g(\delta^*) = B(\delta^*) \geq 0$. Par ailleurs, avec l'hypothèse (ii) : $g(0) = 1 - \alpha \delta^{*k+1} / \left[(k+1) \left(\sum_{i \in I} \sigma_i \right)^{k+1} \right] = 1 - \alpha(1 - \mathbb{E}[\varphi])^{k+1} / (k+1) \geq 0$.

D'autre part :

$$g'(x) = B'(x) + \frac{\alpha(\delta^* - x)^k}{\left(\sum_{i \in I} \sigma_i \right)^{k+1}} = -\frac{4xB(x)}{\sum_{i \in I} \sigma_i^2} + \frac{\alpha(\delta^* - x)^k}{\left(\sum_{i \in I} \sigma_i \right)^{k+1}}.$$

Assurer que g est positive sur $[0, \delta^*]$ revient à garantir que $g(\hat{x}) \geq 0$ pour tout extremum de cet intervalle. Soit donc \hat{x} un extremum de g sur $]0, \delta^*[$, en utilisant le fait que $g'(\hat{x}) = 0$, on obtient :

$$B(\hat{x}) = \frac{\sum_{i \in I} \sigma_i^2}{\left(\sum_{i \in I} \sigma_i \right)^{k+1}} \cdot \frac{\alpha(\delta^* - \hat{x})^k}{4\hat{x}}.$$

On en déduit :

$$g(\hat{x}) \geq 0 \Leftrightarrow \hat{x}^2 - \delta^* \hat{x} + \frac{k+1}{4} \cdot \sum_{i \in I} \sigma_i^2 \geq 0.$$

L'hypothèse (iii) assure alors que le discriminant de ce polynôme du second degré en \hat{x} est négatif, et donc qu'on a toujours $g(\hat{x}) \geq 0$ pour un extremum \hat{x} .

On a donc montré que tous les extremums de g sur $]0, \delta^*[$ sont positifs. Comme de plus $g(0) \geq 0$ et $g(\delta^*) \geq 0$, on a prouvé que g est positive sur l'intervalle $[0, \delta^*]$. \square

La condition (ii) est en fait peu restrictive. En effet, $\alpha = (k+1)/(1 - \mathbb{E}[\varphi])^{k+1}$ implique : $\int_{x=\mathbb{E}[\varphi]}^1 \alpha(1-x)^k dx = 1$. Dit autrement, pour cette valeur de α , la fonction majorant la densité de φ sur $[\mathbb{E}[\varphi], 1]$ a une intégrale égale à 1 sur cet intervalle. Or en pratique, on aura généralement plutôt $\mathbb{P}(\varphi \geq \mathbb{E}[\varphi]) \approx 0.5$.

Exemple : Supposons $k = 1$, et considérons φ telle que : $\mathbb{E}[\varphi] = 1/2$, et supposons que pour tout $i \in I$, $\sigma_i = \sigma > 0$. La condition (ii) se traduit : $\alpha \leq 8$. Quant à la condition (iii), elle devient ici : $1/2 \geq 1 - \sqrt{2/n}$, soit : $n \leq 8$.

Si on suppose maintenant $k = 2$ (décroissance quadratique de la distribution de φ sur $[\mathbb{E}[\varphi], 1]$), on obtient : $\alpha \leq 24$ et $n \leq 12$.

Annexe C

Annexes du chapitre 5

C.1 Note sur un modèle robuste simpliste

Introduisons le modèle robuste suivant, noté $\text{RILP}_3(\gamma)$:

$$\begin{aligned} & \max && cx \\ \text{s.c.} & && \sum_{i \in I_u(j)} \underline{A}_{ji} x_i + \gamma_j \cdot \sum_{i \in I_u(j)} \delta_{ji} + \sum_{i \notin I_u(j)} \underline{A}_{ji} x_i \leq b_j, \quad \forall j \in J, \\ & && x \in \{0, 1\}^n \quad \forall i \in I. \end{aligned} \quad (\text{C.1})$$

RILP_3 semble être un modèle très naturel, à la fois proche du modèle RILP introduit au chapitre 5, et plus simple que ce dernier. En particulier, $\text{RILP}_3(\gamma)$ n'est rien d'autre qu'une instance du problème nominal ILP (5.1). Enfin, noter que l'algorithme alternatif de la partie 5.4.3 utilise ce modèle robuste RILP_3 pour approcher RILP .

Proposition 22 *Supposons que pour tout $j \in J$, $I_u(j) = I$, et qu'il existe $\kappa_j > 0$ tel que : $\forall i \in I, \delta_{ji} = \kappa_j \underline{A}_{ji}$. Pour tout $\gamma \in [0, 1]^m$, il existe $\gamma' \in [0, 1]^m$ tel que toute solution optimale de $\text{RILP}_3(\gamma')$ est aussi optimale pour $\text{RILP}(\gamma)$.*

Preuve : Soit x^* une solution optimale de $\text{RILP}(\gamma)$. On définit γ' par, pour tout $j \in J$: $\gamma'_j = (b_j - \sum_{i \in I} \underline{A}_{ji} x_i^*) / \sum_{i \in I} \delta_{ji}$. Observer que : $\gamma'_j \sum_{i \in I} \delta_{ji} \geq \Delta_j(x^*)$. Ceci implique que, quand $\Delta_j(x^*) = \gamma_j \cdot \sum_{i \in I} \delta_{ji}$, on a $\gamma'_j \geq \gamma_j$. En conséquence, $\gamma'_j < \gamma_j$ seulement si $\Delta_j(x^*) = \sum_{i \in I} \delta_{ji} x_i^*$.

Soit x' une solution optimale de $\text{RILP}_3(\gamma')$. Par construction, x^* est une solution réalisable de $\text{RILP}_3(\gamma')$, et donc : $cx^* \leq cx'$. D'autre part, montrons par contradiction que x' est réalisable pour $\text{RILP}(\gamma)$. Supposons que ce n'est pas le cas : il existe $j \in J$ tel que $\Delta_j(x') > \gamma'_j \cdot \sum_{i \in I} \delta_{ji}$. De manière équivalente : $\sum_{i \in I} \delta_{ji} x'_i > \gamma'_j \cdot \sum_{i \in I} \delta_{ji}$ et $\gamma_j \sum_{i \in I} \delta_{ji} > \gamma'_j \cdot \sum_{i \in I} \delta_{ji}$. Cette dernière condition signifie que $\gamma'_j < \gamma_j$, ce qui n'arrive que si $\Delta_j(x^*) = \sum_{i \in I} \delta_{ji} x_i^*$. Par définition de γ'_j , on a aussi : $\Delta_j(x^*) \leq \gamma'_j \cdot \sum_{i \in I} \delta_{ji}$. On en déduit : $\sum_{i \in I} \delta_{ji} x'_i > \sum_{i \in I} \delta_{ji} x_i^*$.

Si $\delta_j = \kappa_j \underline{A}_j$, on obtient : $\sum_{i \in I} \underline{A}_{ji} x'_i > \sum_{i \in I} \underline{A}_{ji} x_i^*$. Mais comme x' est réalisable pour $\text{RILP}_3(\gamma')$: $\sum_{i \in I} \underline{A}_{ji} x'_i \leq b_j - \gamma'_j \cdot \sum_{i \in I} \delta_{ji}$. De plus, par définition de γ'_j : $\sum_{i \in I} \underline{A}_{ji} x_i^* = b_j - \gamma_j \sum_{i \in I} \delta_{ji}$ (contradiction).

Ainsi, on a montré que x' est réalisable pour $\text{RILP}(\gamma)$. Ceci assure que x' est une solution optimale de $\text{RILP}(\gamma)$. \square

Malgré ce résultat, une instance de RILP ne peut pas en général être remplacée par une instance de RILP_3 , comme l'exemple suivant le montre.

Exemple : Considérons le problème :

$$\begin{aligned} \max \quad & 2x_1 + 2x_2 + 3x_3 \\ \text{s.c.} \quad & w_1x_1 + w_2x_2 + w_3x_3 \leq 3 \\ & x \in \{0, 1\}^3 \end{aligned}$$

où les poids w_i sont supposés indépendants et uniformément distribués sur leurs intervalles respectifs d'existence : $w_1 \in [1, 2]$, $w_2 \in [1, 2]$ et $w_3 \in [2, 3]$. Supposons que l'on souhaite résoudre le problème sous contrainte probabiliste associé avec $\varepsilon = 0$, *i.e.* on veut trouver la solution correspondant au pire scénario. Il est facile de voir que la solution cherchée est $(0, 0, 1)$, avec pour valeur objectif $V = 3$.

On énumère les différentes solutions associées au modèle RILP₃, selon la valeur de γ :

- si $\gamma = 0$, alors la solution optimale de RILP₃(γ) est $(0, 1, 1)$, de valeur $V = 5$.
- Si $0 < \gamma \leq 1/3$, la solution optimale de RILP₃(γ) est $(1, 1, 0)$, de valeur $V = 4$.
- Si $\gamma > 1/3$, la solution optimale de RILP₃(γ) est $(0, 0, 0)$, de valeur $V = 0$.

On voit que la solution attendue ne peut pas être obtenue en utilisant le modèle RILP₃. Pourtant, elle est clairement solution optimale du modèle RILP avec $\gamma = 1$.

C.2 Propriétés de la valeur de RILP(γ) en fonction de γ

On s'intéresse à la fonction V qui à $\gamma \in [0, 1]^m$ associe la valeur du problème RILP(γ). On sait que V est "croissante" sur $[0, 1]^m$, dans le sens où si $\gamma \geq \gamma'$, alors $V(\gamma) \geq V(\gamma')$, puisque RILP(γ') est une relaxation de RILP(γ). Par contre, aucune propriété de type convexité ne peut être garantie sur cette fonction-valeur V dans le cas général, comme le montre l'exemple suivant.

Exemple : Soit $\tilde{V}(\gamma)$ la valeur optimale de RILP-L(γ), la relaxation linéaire de RILP(γ) (cf. partie 5.3.2). On construit une instance dans \mathbb{R}^2 où V n'est ni convexe, ni concave :

$$\begin{aligned} \max \quad & x_1 + x_2 \\ \text{s.c.} \quad & a_1x_1 + x_2 \leq 1, \\ & x_1 + a_2x_2 \leq 1, \\ & x \in [0, 1]^2. \end{aligned}$$

Les coefficients a_1 et a_2 sont supposés tous deux incertains : ce sont deux variables aléatoires uniformément distribuées sur l'intervalle $[0, 2]$. Le programme robuste associé est :

$$\begin{aligned} \max \quad & x_1 + x_2 \\ \text{s.c.} \quad & 2\gamma_1z_1 + 2y_1 + x_2 \leq 1, \\ & 2\gamma_2z_2 + 2y_2 + x_1 \leq 1, \\ & z_1 + y_1 \geq x_1, \\ & z_2 + y_2 \geq x_2, \\ & x \in [0, 1]^2, z \geq 0, y \geq 0. \end{aligned}$$

Il est facile de voir qu'il existe une solution optimale de ce problème telle que $z_i = x_i$ et $y_i = 0$ pour $i \in \{1, 2\}$. En conséquence :

$$\begin{aligned} \tilde{V}(\gamma) = \max \quad & x_1 + x_2 \\ \text{s.c.} \quad & 2\gamma_1x_1 + x_2 \leq 1, \\ & x_1 + 2\gamma_2x_2 \leq 1, \\ & x \in [0, 1]^2. \end{aligned}$$

Considérons maintenant deux familles de problèmes, la première telle que $\gamma_1 = \gamma_2$, la deuxième telle que $\gamma_1 + \gamma_2 = 1/4$:

- si $\gamma_1 = \gamma_2$, on voit que la solution optimale de $\tilde{V}(\gamma)$ est le point d'intersection de $\{x : 2\gamma_1 x_1 + x_2 = 1\}$ et $\{x : x_1 + 2\gamma_2 x_2 = 1\}$. En conséquence : $\tilde{V}(\gamma) = 2/(1 + 2\gamma_1)$, qui est convexe sur $[0,1]$.
- Considérons maintenant les cas où $\gamma_1 + \gamma_2 = 1/4$. On peut vérifier qu'ici aussi, la solution optimale de $V(\gamma)$ est le point d'intersection de $\{x : 2\gamma_1 x_1 + x_2 = 1\}$ et $\{x : x_1 + 2\gamma_2 x_2 = 1\}$. Donc : $\tilde{V}(\gamma) = 3/[2.(1 - \gamma_1 + 4\gamma_1^2)]$, qui est concave sur $[0,1/4]$.

On a donc extrait deux portions de $\gamma \mapsto \tilde{V}(\gamma)$ ayant des propriétés de convexité différentes : cette fonction n'est ni convexe, ni concave.

C.3 Bornes supérieures pour CCILP

Dans la partie 5.4.4, quelques résultats ont été donnés pour calculer des bornes supérieures au problème CCILP. Plus généralement en fait, on s'intéresse aux inégalités valides pour les solutions optimales de CCILP. En effet, de telles inégalités permettent de construire des relaxations de CCILP. On présente ci-après quelques inégalités supplémentaires qui sont valides pour CCILP. Noter que tous ces résultats s'adaptent immédiatement au cas des contraintes probabilistes séparées en remplaçant ε par ε_j .

Lemme 54 *Soit $j \in J$. Supposons que les variables aléatoires $\{A_{ji}\}_{i \in I_u(j)}$ sont indépendantes. Pour tout $i \in I_u(j)$, soit $a_{ji} \in [\underline{A}_{ji}, \bar{A}_{ji}]$ tel que : $\mathbb{P}(A_{ji} \leq a_{ji}) < 1 - \varepsilon^{1/|I_u(j)|}$. Alors toute solution réalisable x de CCILP satisfait :*

$$\sum_{i \in I_u(j)} a_{ji} x_i + \sum_{i \notin I_u(j)} A_{ji} x_i \leq b_j \quad (\text{C.2})$$

Preuve : Soit x une solution réalisable de CCILP. Par contradiction, supposons que : $\sum_{i \in I_u(j)} a_{ji} x_i + \sum_{i \notin I_u(j)} A_{ji} x_i > b_j$. Ceci implique qu'il existe $i \in I_u(j)$ tel que $x_i = 1$. En effet, si $x_i = 0$ pour tout $i \in I_u(j)$, comme $\varepsilon < 1$, la réalisabilité de x pour CCILP entraîne : $\sum_{i \notin I_u(j)} A_{ji} x_i \leq b_j$, et donc x vérifierait : $\sum_{i \in I_u(j)} a_{ji} x_i + \sum_{i \notin I_u(j)} A_{ji} x_i \leq b_j$, ce qui est contraire à notre hypothèse.

On a : $\mathbb{P}(A_j x \leq b_j) \leq \mathbb{P}(\sum_{i \in I_u(j)} A_{ji} x_i \leq \sum_{i \in I_u(j)} a_{ji} x_i)$. On a montré qu'il existe $i \in I_u(j)$ tel que $x_i = 1$, donc : $\mathbb{P}(\sum_{i \in I_u(j)} A_{ji} x_i \leq \sum_{i \in I_u(j)} a_{ji} x_i) \leq 1 - \mathbb{P}(\forall i \in I_u(j) \text{ t.q. } x_i = 1 : A_{ji} > a_{ji}) \leq 1 - \mathbb{P}(\forall i \in I_u(j) : A_{ji} \geq a_{ji})$. D'après l'indépendance des variables aléatoires, on obtient : $\mathbb{P}(A_j x \leq b_j) \leq 1 - \prod_{i \in I_u(j)} \mathbb{P}(A_{ji} \geq a_{ji}) < 1 - \varepsilon$.

Finalement, on a : $\mathbb{P}(Ax \leq b) \leq \mathbb{P}(A_j x \leq b_j) < 1 - \varepsilon$. Ceci est une contradiction avec la réalisabilité de x pour CCILP. \square

Ce résultat peut en pratique être renforcé en se focalisant sur les solutions optimales de CCILP. En effet, ε^{1/n_j} peut être considéré au lieu de $\varepsilon^{1/|I_u(j)|}$ pour choisir les nombres $\{a_{ji}\}_{i \in I_u(j)}$, si n_j satisfait pour toute solution optimale x^* de CCILP : $n_j \geq \sum_{i \in I_u(j)} x_i^*$. Alors x^* vérifie l'inégalité (C.2). Un tel n_j peut par exemple être obtenu facilement en "remplissant" la contrainte j avec les plus petites valeurs de $\{\underline{A}_{ji}\}_{i \in I_u(j)}$ d'abord (voir les idées du lemme 32, partie 5.4.4).

Lemme 55 *Soit $j \in J$. Pour tout $i \in I_u(j)$, soit $v_{ji} \in [0, 1]$ tel que : $\mathbb{P}(\eta_{ji} \leq v_{ji}) < 1 - \varepsilon$. Alors toute solution réalisable x de CCILP satisfait :*

$$\sum_{i \in I_u(j)} \underline{A}_{ji} x_i + \max_{i \in I_u(j)} \{\delta_{ji} v_{ji} x_i\} + \sum_{i \notin I_u(j)} A_{ji} x_i \leq b_j. \quad (\text{C.3})$$

Preuve : Soit x une solution réalisable de CCILP. Par contradiction, supposons que : $\sum_{i \in I_u(j)} \underline{A}_{ji}x_i + \max_{i \in I_u(j)} \{\delta_{ji}v_{ji}x_i\} + \sum_{i \notin I_u(j)} A_{ji}x_i > b_j$. Ceci implique qu'il existe $i \in I_u(j)$ tel que $x_i = 1$. En effet, si $x_i = 0$ pour tout $i \in I_u(j)$, comme $\varepsilon < 1$, la réalisabilité de x pour CCILP signifie que $\sum_{i \notin I_u(j)} A_{ji}x_i \leq b_j$, et x satisferait donc : $\sum_{i \in I_u(j)} \underline{A}_{ji}x_i + \max_{i \in I_u(j)} \{\delta_{ji}v_{ji}x_i\} + \sum_{i \notin I_u(j)} A_{ji}x_i \leq b_j$.

Soit $k \in I_u(j)$ tel que : $\delta_{jk}v_{jk}x_k = \max_{i \in I_u(j)} \{\delta_{ji}v_{ji}x_i\}$ (d'après les considérations précédentes, on a forcément $x_k = 1$). On a :

$$\begin{aligned} \mathbb{P}(A_jx \leq b_j) &\leq \mathbb{P}\left(\sum_{i \in I_u(j)} A_{ji}x_i \leq \sum_{i \in I_u(j)} \underline{A}_{ji}x_i + \delta_{jk}v_{jk}\right) \\ &= \mathbb{P}\left(\sum_{i \in I_u(j)} \delta_{ji}\eta_{ji}x_i \leq \delta_{jk}v_{jk}\right) \\ &\leq \mathbb{P}(\delta_{jk}\eta_{jk} \leq \delta_{jk}v_{jk}) \\ &< 1 - \varepsilon. \end{aligned}$$

Observer que : $\mathbb{P}(Ax \leq b) \leq \mathbb{P}(A_jx \leq b_j) < 1 - \varepsilon$ (contradiction). \square

Malgré leur intérêt théorique, les inégalités (C.2) et (C.3) n'ont pas permis d'améliorer significativement les bornes obtenues à partir de la proposition 17.

Annexe D

Optimisation avec contraintes probabilistes et VaR/CVaR

D.1 Concepts de VaR et de CVaR

Le concept de VaR (*Value at Risk*) est très couramment utilisé en gestion du risque. Si Y est une variable aléatoire modélisant la perte possible, la VaR est définie comme suit :

$$VaR_\varepsilon(Y) = \inf \{y \mid \mathbb{P}(Y \leq y) \geq 1 - \varepsilon\}.$$

$VaR_\varepsilon(Y)$ est donc la perte encourue avec un niveau de risque $\varepsilon \in [0, 1]$. $VaR_\varepsilon(Y)$ est décroissant avec ε .

Une autre mesure du risque, souvent jugée plus pertinente que la VaR, est la CVaR (*Conditional Value at Risk*, aussi appelée *expected shortfall*), qui est définie par :

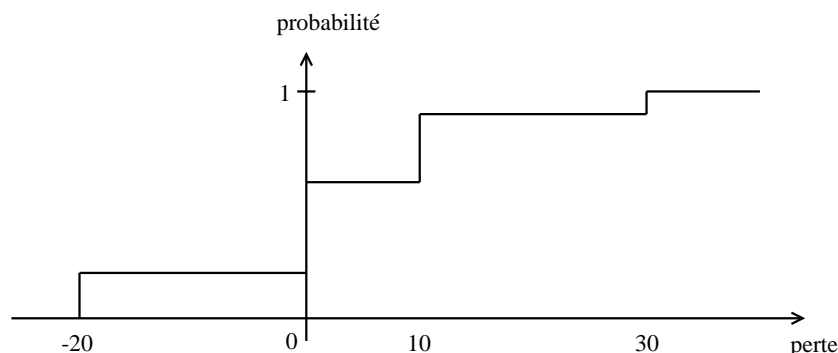
$$CVaR_\varepsilon(Y) = \mathbb{E}[Y \mid Y \geq VaR_\varepsilon(Y)].$$

C'est donc l'espérance de la perte encourue avec un niveau de risque ε . Il est clair que $CVaR_\varepsilon(Y) \geq VaR_\varepsilon(Y)$.

Exemple : On considère les scénarios suivants :

scénario	probabilité d'occurrence	perte associée
1	20%	-20
2	40%	0
3	30%	10
4	10%	30

Le scénario 1 correspond à un gain, les scénarios 3 et 4 à des pertes. Le scénario 2 est neutre. On peut représenter la fonction de répartition de la perte Y :



Voici des exemples de mesures de risques :

ε	$VaR_\varepsilon(Y)$	$CVaR_\varepsilon(Y)$
5%	30	30
10%	10	15
20%	10	15
40%	0	7,5

On détaille le calcul pour $\varepsilon = 10\%$. On a : $VaR_{0,1}(Y) = \inf \{y \mid \mathbb{P}(Y \leq y) \geq 0,9\} = 10$. Alors : $CVaR_{0,1}(Y) = \mathbb{E}[Y \mid Y \geq 10] = (0,3 * 10 + 0,1 * 30) / (0,3 + 0,1) = 15$.

D.2 Lien avec les contraintes probabilistes

Le rapport de VaR avec les contraintes probabilistes est immédiat. Dans une perspective d'optimisation, la perte Y dépend en fait d'une décision x ; on note donc $Y(x)$. Par exemple, la minimisation du risque $VaR_\varepsilon(Y(x))$ s'écrit alors simplement :

$$VaR_\varepsilon = \min_{\substack{y \\ \text{s.c. } \mathbb{P}(Y(x) \leq y) \geq 1 - \varepsilon, \\ x \in \mathcal{X}, y \in \mathbb{R},}} y$$

où \mathcal{X} est l'ensemble des décisions possibles.

L'écriture de la minimisation de $CVaR_\varepsilon(Y(x))$ s'avère plus compliquée. On détaille successivement deux cas particuliers.

Hypothèse de croissance stricte sur les fonctions de répartition : On se place d'abord dans le cas où, pour tout $x \in \mathcal{X}$, la fonction de répartition $y \mapsto \mathbb{P}(Y(x) \leq y)$ est continue et strictement croissante sur le support de $Y(x)$. On remarque alors que, si $\varepsilon \in]0, 1[$:

$$VaR_\varepsilon = \max_{\substack{y \\ \text{s.c. } \mathbb{P}(Y(x) \leq y) \leq 1 - \varepsilon, \\ x \in \mathcal{X}, y \in \mathbb{R}.}} y$$

En notant $f_{Y(x)}$ la densité de probabilité de la variable aléatoire $Y(x)$, on a :

$$CVaR_\varepsilon = \min_{\substack{z \\ \text{s.c. } \mathbb{P}(Y(x) \leq y) \leq 1 - \varepsilon, \\ \int_{t \geq y} t f_{Y(x)}(t) dt \leq z \cdot \mathbb{P}(Y(x) \geq y), \\ x \in \mathcal{X}, y \in \mathbb{R}, z \in \mathbb{R},}} z$$

La continuité de $y \mapsto \mathbb{P}(Y(x) \leq y)$ permet d'écrire, de manière équivalente :

$$CVaR_\varepsilon = \min_{\substack{z \\ \text{s.c. } \mathbb{P}(Y(x) \leq y) \leq 1 - \varepsilon, \\ \int_{t \geq y} t f_{Y(x)}(t) dt \leq \varepsilon \cdot z, \\ x \in \mathcal{X}, y \in \mathbb{R}, z \in \mathbb{R}.}} z$$

La fonction $(x, y) \mapsto \int_{t \geq y} t f_{Y(x)}(t) dt$ est le plus souvent impossible à écrire de manière analytique.

Nombre fini de scénarios : On suppose maintenant qu'il existe un nombre fini N de scénario : pour tout $x \in \mathcal{X}$, $Y(x) \in \{y_1(x), \dots, y_N(x)\}$. Chaque scénario y_k est associé à une probabilité π_k de réalisation ($\sum_{k=1}^N \pi_k = 1$). On introduit, pour tout $x \in \mathcal{X}$:

$$\begin{aligned} VaR_\varepsilon(x) = \min & \quad u \\ \text{s.c.} & \quad \sum_{k=1}^N \pi_k \cdot \mathbb{1}_{y_k(x) \leq u} \geq 1 - \varepsilon, \\ & \quad u \in \mathbb{R}. \end{aligned}$$

Alors la minimisation de CVaR s'écrit :

$$\begin{aligned} CVaR_\varepsilon = \min & \quad z \\ \text{s.c.} & \quad \sum_{k=1}^N \pi_k \cdot y_k(x) \cdot \mathbb{1}_{\{y_k(x) \geq VaR_\varepsilon(x)\}} \leq z \cdot \sum_{k=1}^N \pi_k \cdot \mathbb{1}_{\{y_k(x) \geq VaR_\varepsilon(x)\}}, \\ & \quad x \in \mathcal{X}, z \in \mathbb{R}, \end{aligned}$$

soit :

$$\begin{aligned} CVaR_\varepsilon = \min & \quad z \\ \text{s.c.} & \quad \sum_{k=1}^N \pi_k \cdot (y_k(x) - z) \cdot \mathbb{1}_{\{y_k(x) \geq VaR_\varepsilon(x)\}} \leq 0, \\ & \quad x \in \mathcal{X}, z \in \mathbb{R}. \end{aligned}$$

Dans certains cas particuliers (simplistes ?), cette analyse peut être simplifiée. Supposons par exemple que l'on ait $y_1(x) \leq \dots \leq y_N(x)$, quel que soit $x \in \mathcal{X}$. Alors il suffit de déterminer le plus petit entier q tel que : $\sum_{k=1}^q \pi_k \geq 1 - \varepsilon$. On a : $VaR_\varepsilon(x) = y_q(x)$, et :

$$\begin{aligned} CVaR_\varepsilon = \min & \quad z \\ \text{s.c.} & \quad \sum_{k=q}^N \pi_k \cdot (y_k(x) - z) \leq 0, \\ & \quad x \in \mathcal{X}, z \in \mathbb{R}, \end{aligned}$$

soit, de manière équivalente :

$$CVaR_\varepsilon = \frac{1}{\sum_{k=q}^N \pi_k} \cdot \min_{x \in \mathcal{X}} \sum_{k=q}^N \pi_k \cdot y_k(x).$$

Cette dernière mesure est à rapprocher de l'étude présentée dans [15].

Bibliographie

- [1] R. Aringheri. A tabu search algorithm for solving chance-constrained programs. Note del polo 73, DTI - University of Milano, 2005. Disponible à <http://www.crema.unimi.it/Biblioteca/SchedaNota.asp?Nota=92>.
- [2] A. Atamtürk. Strong formulations of robust mixed 0-1 programming. *Mathematical Programming*, 108(2-3) :235–250, 2006.
- [3] J.S. Bedo and V. Martin. Reconstitution de matrices de trafic à partir de mesures netflow dans otip, specification. Tech. report nt/divrd/trq/05/007, France Télécom R&D, 2005.
- [4] W. Ben-Ameur. Computing the initial temperature of simulated annealing. *Computational Optimization and Applications*, 29 :367–383, 2004.
- [5] W. Ben-Ameur and H. Kerivin. Offres de réseaux privés virtuels flexibles. technical report nt/ftird/7358, France Télécom R&D, 2001.
- [6] W. Ben-Ameur and H. Kerivin. Routing of uncertain demands. *Optimization and Engineering*, 6(3) :283–313, 2005.
- [7] A. Ben-Tal and A. Nemirovski. On safe tractable approximations of chance constrained linear matrix inequalities. submitted to *Mathematics of Operations Research*.
- [8] A. Ben-Tal and A. Nemirovski. Robust solutions of linear programming problems contaminated with uncertain data. *Mathematical Programming (Ser. A)*, 88 :411–424, 2000.
- [9] A. Ben-Tal and A. Nemirovski. Robust optimization - methodology and applications. *Mathematical Programming*, B92 :453–480, 2002.
- [10] P. Beraldi and A. Ruszczyński. A branch and bound method for stochastic integer problems under probabilistic constraints. *Optimization Methods and Software*, 17 :359–382, 2002.
- [11] D. Bertsimas, D. Pachamanova, and M. Sim. Robust linear optimization under general norms. *Operations Research Letters*, 32(6) :510–516, 2004.
- [12] D. Bertsimas and M. Sim. Robust discrete optimization and network flows. *Mathematical Programming (Ser. B)*, 98 :49–71, 2003.
- [13] D. Bertsimas and M. Sim. The price of robustness. *Operations Research*, 52(1) :35–53, 2004.
- [14] L. Bianchi, M. Dorigo, L.M. Gambardella, and W.J. Gutjahr. Metaheuristics in stochastic combinatorial optimization : a survey. technical report idsia-08-06, IDSIA, 2006. Disponible à www.idsia.ch/idsiareport/IDSIA-08-06.pdf.
- [15] D. Bienstock. Experiments in robust portfolio optimization. Disponible sur www.optimization-online.org, 2007.
- [16] J.R. Birge and F. Louveaux. *Introduction to Stochastic Programming*. Springer-Verlag, 1997.
- [17] J.F. Bonnans and A. Shapiro. *Perturbation Analysis of Optimization Problems*. Springer-Verlag, 2000.

- [18] D.M. Bradley and R.C. Gupta. On the distribution of the sum of non-identically distributed uniform random variables. *Annals of the Institute of Statistical Mathematics*, 54(3) :689–700, 2002.
- [19] G. Calafiore and M.C. Campi. Uncertain convex programs : randomized solutions and confidence levels. *Mathematical Programming (Ser. A)*, 102 :25–46, 2005.
- [20] G. Calafiore and L. El Ghaoui. Distributionally robust chance-constrained linear programs with applications. *Journal of Optimization Theory and Applications*, 130(1) :1–22, 2006.
- [21] R. Carbone. Public facilities location under stochastic demand. *INFOR*, 12 :261–270, 1974.
- [22] G. Carello, F. Della Croce, M. Ghirardi, and R. Tadei. Solving the hub location problem in telecommunication network design : A local search approach. *Networks*, 44(2) :94–105, 2004.
- [23] A. Charnes and W.W. Cooper. Chance-constrained programming. *Management Science*, 5 :73–79, 1959.
- [24] X. Chen, M. Sim, and P. Sun. A robust optimization perspective to stochastic models. *Operations Research*, 2007. A paraître.
- [25] Cisco white papers. Disponibles à www.cisco.com/en/US/products/ps6601/prod_white_papers_list.html.
- [26] Cisco netflow or rmon2 : Considerations for enterprises, 2003. NetQoS white paper. Disponible à www.cisco.com/warp/public/732/partnerpgm/docs/netqos_netflow_rmon2.pdf.
- [27] J. Cohen, P. Cohen, S.G. West, and L.S. Aiken. *Applied multiple regression/correlation analysis for the behavioral sciences*. (3rd ed.) Hillsdale, NJ : Lawrence Erlbaum Associates, 2003.
- [28] G.B. Dantzig. Linear programming under uncertainty. *Management Science*, 1 :179–206, 1955.
- [29] A. Doucet, N. de Freitas, and N. Gordon. *Sequential Monte Carlo Methods in Practice*. Springer Verlag, 2001.
- [30] E. Erdoğan and G. Iyengar. Ambiguous chance constrained problems and robust optimization. *Mathematical Programming (Ser. B)*, 107(1-2) :37–61, 2006.
- [31] M. Galassi et al. Gnu scientific library reference manual. (2nd Ed.), ISBN 0954161734. <http://www.gnu.org/software/gsl/>.
- [32] V. Gabrel et M. Minoux. A scheme for exact separation of extended cover inequalities and application to multidimensional knapsack problems. *Operations Research Letters*, 30 :252–264, 2002.
- [33] L. El Ghaoui, F. Oustry, and H. Lebret. Robust solutions to uncertain semidefinite programs. *SIAM Journal of Optimization*, 9(1) :33–52, 1998.
- [34] J.C. Gilbert. Optimisation différentiable : théorie et algorithmes, 1999. Polycopié du cours donné à l’ENSTA.
- [35] E. Gourdin, M. Labbé, and H. Yaman. Telecommunication and location. In Z. Drezner and H.W. Hamacher, editors, *Facility Location : Applications and Theory*, pages 275–305. Springer, 2002.
- [36] C.M. Grinstead and J.L. Snell. *Introduction to Probability*. American Mathematical Society, 1997. 2^e édition révisée.
- [37] Z. Gu, G.L. Nemhauser, and M.W.P. Savelsbergh. Lifted cover inequalities for 0-1 integer programs : computation. *INFORMS Journal on Computing*, 10 :427–437, 1998.
- [38] Z. Gu, G.L. Nemhauser, and M.W.P. Savelsbergh. Lifted cover inequalities for 0-1 integer programs : Complexity. *INFORMS Journal on Computing*, 11(1) :117–123, 1999.

- [39] A. Gunnar, M. Johansson, and T. Telkamp. Traffic matrix estimation on a large ip backbone - a comparison on real data. In *Proc. of ACM Internet Measurement Conference*, 2004.
- [40] W.K. Klein Haneveld and M.H. van der Vlerk. Stochastic integer programming : general models and algorithms. *Annals of Operations Research*, 85 :39–57, 1999.
- [41] F.S. Hillier. Chance-constrained programming with 0-1 or bounded continuous decision variables. *Management Science*, 14(1) :34–57, 1967.
- [42] W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58 :13–30, 1963.
- [43] Ilog Cplex : www.ilog.com/products/cplex/.
- [44] H. Ishii and S. Shiode. Chance constrained bottleneck spanning tree problem. *Annals of Operations Research*, 56(1) :177–187, 1995.
- [45] P. Kall and J. Mayer. *Stochastic Linear Programming*. Springer, 2005.
- [46] P. Kall and S.W. Wallace. *Stochastic Programming*. Wiley, Chichester, 1994.
- [47] H. Kellerer, U. Pferschy, and D. Pisinger. *Knapsack Problems*. Springer, 2004.
- [48] P. Kouvelis and G. Yu. *Robust discrete optimization and its applications*. Kluwer Academic Publishers, 1997.
- [49] M. Labbé, H. Yaman, and E. Gourdin. A branch and cut algorithm for hub location problems with single assignment. *Mathematical Programming*, 102(2) :371–405, 2005.
- [50] C. Lemaréchal, A. Ouorou, and G. Petrou. Robust network design in telecommunications. In *Proc. of International Network Optimization Conference (INOC'05)*, volume 1, pages 258–264, 2005.
- [51] A. Lissner, A. Ouorou, J.-P. Vial, and J. Gondzio. Capacity planning under uncertain demand in telecommunication networks. Note technique, France Télécom R&D, 1999. Disponible à www.hec.unige.ch/recherches_publications/cahiers/1995-1999/99.13.pdf.
- [52] S. Martello and P. Toth. *Knapsack Problems : Algorithms and Computer Implementations*. Wiley, 1990.
- [53] M. Minoux. *Programmation mathématique : théorie et algorithmes*. Dunod, 1983.
- [54] O. Montanuy. Traffic matrices obtained from mpls and ip counters. Rapport technique ref. yourdoc-05/10/021, France Télécom R&D, 2005.
- [55] J.M. Mulvey, R.J. Vanderbei, and S. A. Zenios. Robust optimization of large-scale system. *Operations Research*, 43(2) :264–281, 1995.
- [56] G.L. Nemhauser and L.A. Wolsey. *Integer and Combinatorial Optimization*. John Wiley & Sons, 1999.
- [57] A. Nemirovski and A. Shapiro. Scenario approximations of chance constraints. In G. Calafiore and F. Dabbene, editors, *Probabilistic and Randomized Methods for Design under Uncertainty*, pages 3–48. Springer, London, 2005.
- [58] A. Nemirovski and A. Shapiro. Convex approximations of chance constrained programs. *SIAM Journal on Optimization*, 17 :969–996, 2006.
- [59] Y. Nikulin. Robustness in combinatorial optimization and scheduling theory : An annotated bibliography. Disponible à www.optimization-online.org/DB_FILE/2004/11/995.pdf, 2004.
- [60] A. Ouorou. Robust capacity assignment in telecommunications. *Computational Management Science*, 3(4) :285–305, 2006.
- [61] A. Ouorou and H. Sechet. Robustesse des réseaux face à l'incertitude de la demande : modélisation et méthodes de résolution. Note technique nt/ft&d/7294, France Télécom R&D, 2001.

- [62] A. Ouorou and J.-Ph. Vial. A model for robust capacity planning for telecommunications networks under demand uncertainty. In *Proc. of Design of Reliable Communication Networks (DRCN'07)*, 2007.
- [63] K. Papagiannaki, N. Taft, and A. Lakhina. A distributed approach to measure ip traffic matrices. In *Proc. of ACM Internet Measurement Conference (IMC)*, pages 161–174, 2004.
- [64] M.C. Pinar. A note on robust 0-1 optimization with uncertain cost coefficients. *4OR, Quarterly Journal of the Belgian, French and Italian Operations Research Societies*, 2 :309–316, 2004.
- [65] D. Pisinger. Where are the hard knapsack problems? *Computers and Operations Research*, 32(9) :2271–2284, 2005.
- [66] A. Prékopa. *Stochastic Programming*. Kluwer Academic Publishers, Dordrecht, Boston, 1995.
- [67] A. Prékopa. Probabilistic programming. In A. Ruszczynski and A. Shapiro, editors, *Stochastic Programming*, volume 10 of *Handbooks in OR&MS*. Elsevier, 2003.
- [68] C.P. Robert and G. Casella. *Monte Carlo Statistical Methods*. Springer, 2002.
- [69] R.T. Rockafellar. *Convex Analysis*. Princeton University Press, 1970. Ré-impression : 1997.
- [70] I. Rodríguez-Martína and J.J. Salazar-González. Solving a capacitated hub location problem. *European Journal of Operational Research*, 184(2) :468–479, 2008.
- [71] S.M. Ross. *Initiation aux probabilités*. Presses Polytechniques Romandes, 1987.
- [72] N.V. Sahinidis. Optimization under uncertainty : state-of-the-art and opportunities. *Computers & Chemical Engineering*, 28 :971–983, 2004.
- [73] C. Semeria and H. Gredler. Juniper networks solutions for network accounting. Juniper white paper [online], 2001. Disponible à www.juniper.net/solutions/literature/white_papers/200010.pdf.
- [74] Y. Seppälä. On accurate linear approximations for chance-constrained programming. *Journal of the Operational Research Society*, 39(7) :693–694, 1988.
- [75] Y. Seppälä and T. Orpana. Experimental study on the efficiency and accuracy of a chance-constrained programming algorithm. *European Journal of Operational Research*, 16 :345–357, 1984.
- [76] S. Shiode and Z. Drezner. A competitive facility location problem on a tree network with stochastic weights. *European Journal of Operational Research*, 149(1) :47–52, 2003.
- [77] L.V. Snyder. Facility location under uncertainty : A review. Technical report #04t-015, Dept. of Industrial & Systems Engineering, Lehigh University, 2004. Disponible à www.lehigh.edu/~lvs2/Papers/stochloc.pdf.
- [78] A. Soule, A. Nucci, R. Cruz, E. Leonardi, and N. Taft. How to identify and estimate the largest traffic matrix elements in a dynamic environment. In *Proc. of ACM Sigmetrics*, pages 73–84, 2004.
- [79] A.L. Soyster. Convex programming with set-inclusive constraints and applications to inexact linear programming. *Operations Research*, 21(5) :1154–1157, 1973.
- [80] S.R. Tayur, R.R. Thomas, and N.R. Natraj. An algebraic geometry algorithm for scheduling in the presence of setups and correlated demands. *Mathematical Programming*, 69(3) :369–401, 1995.
- [81] S. van Hoesel and A. Wagelmans. On the complexity of postoptimality analysis of 0/1 programs. *Discrete Applied Mathematics*, 91(1-3) :251–263, 1999.

- [82] M.R. Wagner. Stochastic 0-1 linear programming under limited distributional information. *Operations Research Letters*, 36(2) :150–156, 2008.
- [83] L. Wasserman. *All of Statistics : A Concise Course in Statistical Inference*. Springer Texts in Statistics, 2005.
- [84] A. Weintraub and J. Vera. A cutting plane approach for chance constrained linear programs. *Operations Research*, 39(5) :776–785, 1991.
- [85] E.W. Weisstein. Lyapunov condition. From MathWorld—A Wolfram Web Resource. <http://mathworld.wolfram.com/LyapunovCondition.html>.
- [86] P. Wessa. Free statistics software. Office for Research Development and Education, 2007. Version 1.1.21-r4, URL <http://www.wessa.net/>.
- [87] H. Yaman and G. Carello. Solving the hub location problem with modular link capacities. *Computers and Operations Research*, 32 :3227–3245, 2005.
- [88] H. Zang and A. Nucci. Optimal netflow deployment in ip networks. In *Proc. of the 19th International Teletraffic Congress (ITC)*, 2005.
- [89] Q. Zhao, Z. Ge, J. Wang, and J.J. Xu. Robust traffic matrix estimation with imperfect information : Making use of multiple data sources. In *Proc. of the ACM SIGMETRICS/IFIP WG 7.3 Performance*, pages 133–144, 2006.