Université de Montréal

Optimisation quadratique en variables binaires : Quelques résultats et techniques

par

Calvin Mbuntcha Wuntcha

Département de mathématiques et de statistique

Faculté des arts et des sciences

Thèse présentée à la Faculté des études supérieures

en vue de l'obtention du grade de

Philosophiæ Doctor (Ph.D.)

en Mathématiques

Orientation mathématiques appliquées

Mai 2009

Université de Montréal

Faculté des études supérieures

Cette thèse intitulée

Optimisation quadratique en variables binaires : Quelques résultats et techniques

présentée par

Calvin Mbuntcha Wuntcha

a été évaluée par un jury composé des personnes suivantes :

Hahn, Gena
(président-rapporteur)

Rosenberg, Ivo G.
(directeur de recherche)

Broer, Abraham
(membre du jury)

Hansen, Pierre
(examinateur externe)

Ferland, Jacques
(représentant du doyen de la FES)

Thèse acceptée le:
1er Septembre 2009

SOMMAIRE

Plusieurs problèmes en recherche opérationelle et analyse de décisions se ramènent à l'optimisation en variables binaires. Dans ce travail, nous traitons de l'optimisation quadratique en variables $0-1$ et sans contrainte. C'est l'optimisation des polynômes multilinéaires quadratiques sur les sommets de l'hypercube unité. Ce problème est NP-dur. Au chapitre 1, nous étendons un résultat de 1972 traitant des points de minimum dans les sommets et dans l'hypercube en entier. Nous proposons aussi une perturbation sur certaines fonctions discrètes de manière à conserver un unique minimum global au problème sans créer de nouveaux minima locaux. Nous déterminons ensuite une inégalité linéaire que vérifie tout point de minimum d'une fonction quadratique sur un ensemble ayant un centre de symétrie. Au chapitre 2, nous caractérisons l'ensemble de tous les minima $0-1$ pour chaque fonction quadratique f par des inégalités linéaires inspirées de celles citées au chapitre 1. Des résultats numériques sur le point s qui minimise la principale de ces fonctions linéaires sont donnés, montrant que sa valeur pour f est souvent assez petite. Au chapitre 3, nous étudions une heuristique à partir de certaines propriétés matricielles de changement d'origine du cube. Nous présentons les résutats de simulations numériques de cet algorithme pour certains problèmes tests bien connus, et nous faisons ressortir un impact non négligeable du point s. Au chapitre 4, nous entamons l'étude d' un algorithme de descente du gradient pour partir d'un point intérieur de l'hypercube et atteindre un point de minimum local dans les sommets de l'hypercube. Cet algorithme utilise les propriétés de la matrice associée, et les propriétés des fonctions multilinéaires sur l'hypercube et ses sommets ; quelques résultats expérimentaux préliminaires sont donnés pour cet algorithme particulier qui est encore en amélioration. Au chapitre 5, nous présentons une nouvelle classe de problèmes d'optimisation quadratique à variables binaires solubles en temps polynômial.

Mots-clés : Hypercube unité, NP-dur, polynôme multilinéaire, quadratique, temps polynomial, variables $0-1$, minimum local

SUMMARY

Many problems in operations research and related fields are connected to binary optimization. This work is about unconstrained quadratic binary optimisation that treats of the optimization of multilinear quadratic polynomials on the set of $0-1$ vectors. The problem is known to be NP-hard. In Chapter 1 we extend a result of 1972 on the optimal points of a multilinear polynomial on the $0-1$ vectors and on the unit hypercube. We also propose a pertubation on certains discretes functions such that a unique point of minimum is kept without creating new points of local minimum for the function. We also determine a linear inequality satisfied by every minimum point of a quadratic fonction on a set having a symmetric centre. In chapter 2, we caracterise the set of all $0-1$ points of minimum for a quadratic polynomial by a set of linear inequalities inspired by those cited in Chapter 1. The point of global minimum of the main linear function used in the caracterization is considered, and simulations with some well known test problems are presented, showing that the value of the point is often small. In chapter 3, we study a local search heuristic using the matricial properties of changing the origin of the unit hypercube by complementing some variables. Simulation results cary out on some known test problems are presented. In chapter 4, we present a gradient descent method for $0-1$ quadratic optimization; this method uses the particular properties of the associated matrix, and the property that optimizing a multilinear polynomial on the unit hypercube is equivalent to optimizing it on the $0-1$ vectors. Preliminary simulation results are reported

for small size problems. In chapter five, we present a new class of polynomially sovable quadratic $0-1$ optimization.

Keywords : Unit hypercube quadratic, binary optimization, multilinear polynomial, NP-hard, polynomially solvable.

TABLE DES MATIÈRES

Sommaire iii

Summary v

Liste des figures xi

Liste des tableaux xii

Dédicace xiv

Remerciements xv

Introduction 2

0.1. Une brève présentation de l'optimisation en variables binaires et du cas quadratique 2

0.2. Présentation de notre travail centré sur le cas quadratique sans contrainte 6

0.3. Contributions et collaborations des autres 9

Chapitre 1. Some remarks on non linear $0-1$ optimization 10

Résumé et Contributions 10

Abstract 11

Introduction 11

1.1. Quadratic functions having square variable terms with non positive coefficients.......... 13

1.2. Perturbation of discrete functions to keep a unique minimal point and not create new local minimal points.......... 20

1.3. Some properties of quadratic optimization on a centrally symmetric set.......... 23

1.4. Acknowledgements.......... 29

Chapitre 2. A characterization of minimal points of a quadratic $0-1$ function.......... 30

Résumé.......... 30

Abstract.......... 31

Introduction.......... 31

2.1. A set of linear functions and its properties.......... 32

2.2. First reduction of the Set L.......... 37

2.3. *A* second reduction of the set L and a bound on the number of $0-1$ minimum points.......... 40

2.4. Concluding remarks on the main linear function.......... 42

2.5. Acknowledgements.......... 43

Chapitre 3. Matricial properties of changing the origin and a heuristic for quadratic $0-1$ minimization.......... 45

Résumé et Contributions.......... 45

Abstract 46

Introduction 46

3.1. Effects and matricial properties of origin change. 48

3.2. Finding k-resistant points with some properties of the changing of the origin 54

3.3. Experimental results and ending remarks 59

Chapitre 4. Study of a gradient method in quadratic $0-1$ optimization 64

Résumé et Contributions 64

Abstract 65

Introduction 65

4.1. Description of the principle. 67

4.2. The basic algorithm and some observations. 73

4.3. Experimental results and ending remarks 75

Chapitre 5. A new polynomial case of quadaratic $0-1$ optimization related to a minimal eigenvalue 76

Résumé. 76

Abstract 76

Introduction 77

5.1. The new case related to the minimal eigenvalue and some appropriate points of the hypercube. 78

5.2. Ending remarks 83

Chapitre 6. Conclusion 84

Bibliographie 87

LISTE DES TABLEAUX

2.1 The values of s_p 44

3.1 2-r and other heuristics (values) 61

3.2 2-r and other heuristics (time) 62

3.3 1-r and 2-r (time and values) 63

4.1 Gradient descent 75

DÉDICACE

Je dédie cette thèse à quatre personnes spéciales de ma vie à qui je tiens à dire un mot personnel :

À mon papa le Rev. Salomon Wuntcha :

Tu as accepté Jésus-Christ dans ta vie, tu as décidé de le servir, et Dieu t'a rempli de sagesse et de bénédictions contagieuses. Je te dédie cette thèse, parce que je t'aime et je suis heureux de te ressembler. Je te la dédie aussi parce que je suis toujours impressionné et inspiré par ton retour aux études dans ta soixantaine, et l'obtention de ce diplôme que tu voulais.

À ma maman Sara T. Wuntcha :

Tu m'es chère, et en finissant cette thèse, je me souviens avec joie des premiers jours où tu m'accompagnais à l'école.

À mon épouse Aude-Sidoine K. Wuntcha :

Ma chérie, tu es une bonne épouse, je suis heureux, "I love you !"

À ma fille Priscilla-Blessed Wuntcha :

Tu es arrivée dans notre vie comme un ange, belle comme ta mère, bénie comme ton père et ton grand-père ! J'espère avoir plus de temps après cette thèse, pour jouer avec toi et te montrer le secret de la vie, dans une génération piégée.

REMERCIEMENTS

Je crois que la recherche scientifique est un effort qu'on fait pour découvrir une petite partie de la logique avec laquelle Dieu a créé et fait fonctionner les composantes concrètes ou abstraites de Son univers qui nous héberge. Je crois qu'on doit lui dire merci lorsqu'il nous accorde la grâce d'en saisir quelques principes. C'est pourquoi, je commence par bénir **Dieu, le Père de notre Seigneur Jésus-Christ**, qui m'a permis de faire ce travail. Je béni encore **le Seigneur Jésus-Christ** , par qui et pour qui toutes choses ont été créées, celles qui sont sur la terre et celles qui sont dans les cieux, les visibles et les invisibles, celles que nous connaissons et celles que nous cherchons à connaître.

Les êtres humains que Dieu a utilisé pour rendre possible plusieurs aspects de ce travail, méritent aussi toute ma gratitude. Le premier c'est **le Professeur Ivo Rosenberg** qui m'a proposé ce sujet et qui a dirigé les travaux de bout en bout, je lui dis un merci tout particulier. Je remercie **le professeur Endré Boros** pour sa généreuse collaboration marquée par l'invitation et l'attention qu'il m'a offerte dans le centre de recherches opérationnelles (RUTCOR) qu'il dirige aux États-Unis. Je remercie **le professeur Charles Audet** pour les discussions utiles qu'on a eu. Je remercie **l'Agence Universitaire de la Francophonie (A.U.F)** pour la bourse doctorale qui a soutenu la base de ce travail. Je remercie le **département de mathématiques (DMS)** pour tout son soutien, et je dis un merci particulier à **la professeure Marlène Frigon** pour ses conseils et son attention. Je remercie **mon épouse Aude-Sidoine** qui a accepté de faire tous les sacrifices que cette thèse nous a imposés, les nuits blanches qu'elle a passé n'étant qu'un des aspects. Je remercie **ma belle-mère** qui est venue prendre soin de notre fille, ce qui m'a permis de me concentrer pour finir. Je remercie **ma fille Priscilla** qui m'a souvent laissé aller à contre-coeur. Je **remercie Rodrigue Kéou** pour les nombreuses heures consacrées à m'aider dans les ajustements, et je remercie **son épouse** pour son soutien. Je remercie **Florence Wékapé** pour la relecture et les remarques. Je remercie **Emma Rosine Ngansop** pour son assistance très généreuse dans plusieurs détails. Je remercie **Etienne Koffi et son épouse** pour la relecture et l'aide aux simulations. Je remercie **Etienne Temgoua** pour sa contribution à la relecture. Je remercie mon collègue **Moussa Kounta** qui m'a remplacé dans ma tâche d'enseignement pendant une semaine déterminante. Je dis merci à **mes frères et soeurs de l'église CMCI de Montréal** qui ont parfois fait le sacrifice de laisser que j'aille méditer sur des théorèmes mathématiques. Je remercie **le pasteur Jean Ngankwé** qui m'a exhorté depuis des années à aller jusqu'au bout avec ce doctorat. À titre postume, je dis un merci particulier au **Professeur Zacharias T. Fomum**, un grand scientifique et grand serviteur de Dieu, qui m'a profondement influencé et encouragé avant d'achever sa course en mars dernier. Je remercie tous **les membres de ma famille, tous les amis et tous ceux** qui m'ont soutenu de près ou de loin.

INTRODUCTION

L'optimisation en variables binaires fait partie de la recherche opérationnelle et des disciplines connexes comme l'analyse des décisions. Elle est aussi connue sous les noms de : optimisation pseudo-booléènne, programmation en variables bivalentes, optimisation en variables $0-1$, etc.

0.1. Une brève présentation de l'optimisation en variables binaires et du cas quadratique

C'est vers la fin des années quarante que la théorie des algèbres de Boole trouve ses premières applications dans l'étude des circuits de commutation ; les éléments de ces circuits ayant régulierement deux états possibles, ce qui fait appel à l'algèbre de Boole à deux éléments. C'est dans son prolongement que l'optimisation pseudo-booléenne s'est developpée. En fait, il existe plusieurs problèmes de décisions dont la modélisation se réduit à l'optimisation d'une fonction réelle dont les variables n'ont que deux valeurs possibles. L'on trouve fréquemment de tels problèmes en recherche opérationnelle, dans la théorie des graphes, en mathématique combinatoire, en statistique mécanique (Sping Glasses), en sciences économiques, en finances, en informatique, dans les circuits, dans les réseaux de neurones, dans la théorie des jeux, etc. (voir [BHT 07],[BH 02], [HS 68]). C'est dans les années soixante que l'optimisation pseudo-booléènne prendra effectivement forme, et l'ouvrage Boolean methods in Operations Research and related Areas de Peter L. Hammer et S. Rudeanu en 1968 présente les bases essentielles de la théorie (voir [HS 68]) : Par exemple, ils y présentent des exemples précis de problèmes pratiques (en transport, théorie des jeux, etc.) qui se ramènent à l'optimisation pseudo-booléenne ; les fonctions pseudo-booléènnes et certaines de leurs propriétés y sont décrites. Une fonction pseudo-booléènne à n variables est une fonction f de $\{0,1\}^n$ vers $\mathbb{R}$, et il est établit qu'une telle fonction peut se représenter par un polynôme à n variables. Très souvent ces fonctions sont données sous la forme polynômiale, et comme $c^2 = c$ pour tout $c \in \{0,1\}$, l'on peut supposer qu'un tel polynôme est multilinéaire, i.e. que son expression ne possède pas de variable à une puissance plus grande que 1. Plus précisement, soient n variables $x_1, ..., x_n$, appelons monôme de longueur k tout produit de k variables par une constante, $0 \leq k \leq n$ en considérant que un monôme de longeur 0 est une constante. On peut dire qu'un polynôme multilinéaire à n variables $x_1, ..., x_n$ et de degré k est une combinaison linéaire de monômes dont la plus grande longeur est k. Exemples : parmi les polynomes à trois variables suivants : $f_1(x_1, x_2, x_3) = 2x_1x_2 - 3x_1x_3 + 4x_2x_3 - 2x_2 + x_1 + 3$; $f_2(x_1, x_2, x_3) = -x_2^2 + 2x_3 + x_1 - 2$, $f_3(x_1, x_2, x_3) = 5x_1x_2x_3 - 4x_1x_2 + x_3$, $f_4(x_1, x_2, x_3) = 2x_1 + x_2 - x_3$, $f_5(x_1, x_2, x_3) = -2x_2^3x_1 + 3x_1x_3 - 2$, f_1, f_3 et f_4 sont multilinéaires, f_1 est de degré 2, f_3 de degré 3, et f_4 de degré 1 ; alors que f_2 et f_5 ne sont pas multilinéaires. Habituellement, les polynômes de degré 1 sont dits linéaires, ceux de degré 2 sont dits quadratiques et ceux de degré 3 sont dits cubiques. Le problème d'optimisation revient à celui de trouver la valeur minimale c_0 de f sur $\{0,1\}^n$ et $w \in \{0,1\}^n$ telle que $f(w) = c_0$. L'ensemble $\{0,1\}^n$ est de cardinalité exponentielle 2^n, et il devient impossible en un temps raisonnable (même pour un ordinateur) de faire le calcul des valeurs de tous les 2^n points quand le nombre de variables n devient grand ($n = 50$, $n = 100$, $n = 500$, $n = 1000$, etc.). Ce problème d'optimiser un polynôme réel à n variables sur $\{0,1\}^n$ avec ou sans contraintes est rattaché à une classe de problèmes dits "NP-dur" : en fait, depuis l'identification de tels problèmes dans les années soixante-dix, la question de savoir s'il peut exister un algorithme permettant en général de résoudre un tel problème en temps polynômial reste une question ouverte (voir [GJ 79], [BH 02]). En 1972, I. G. Rosenberg a prouvé que l'optimisation d'un polynôme multilinéaire f sur $\{0,1\}^n$ revient à l'optimisation de f sur $[0,1]^n$ (voir [Ro 72]) ; ce passage de la condition $x_i \in \{0,1\}$ à $0 \leq x_i \leq 1$ pour tout $i = 1, 2, ...n$ appelé encore relaxation permet d'utiliser les notions et résultats de continuité, différentiabilité, convexité etc. qui ne sont pas possibles dans $\{0,1\}^n$. Le premier cas non trivial d'optimisation des polynômes multilinéaires sur $\{0,1\}^n$ est le cas quadratique. Il a été établi que le cas d'un polynôme de degré plus grand que 2 peut se ramener au cas quadratique par l'addition de variables supplémentaires dites "variables de pénalité", et ce passage au cas quadratique se fait en temps polynômial (voir [Ro 75]), ce qui justifie la place centrale que le cas quadratique occupe dans la théorie. On peut présenter un polynôme multilinéaire quadratique à n variables sous les formes

$$f(x) = \sum_{1 \leq i < j \leq n} a_{ij} x_i x_j + \sum_{i=1}^{n} b_i x_i + c = \frac{1}{2} x^t A x + b^t x + c. \tag{0.1.1}$$

où $A = [a_{ij}]$ est une matrice symétrique de taille $n \times n$, tous les éléments diagonaux étant nuls, $b = (b_1, ..., b_n)^t$ est un vecteur constant de $\mathbb{R}^n$, $x = (x_1, x_2, ..., x_n)^t$ est le vecteur des variables, c est une constante réelle, et t représente l'opération de transposition. Comme dans tous les autres cas, deux classes habituelles de minimisation se présentent dans le cas quadratique : La minimisation avec contraintes et la minimisation sans contrainte. La minimisation sans contrainte consiste à trouver la valeur minimale c_0 de f sur $\{0,1\}^n$ ainsi qu'un point $w \in \{0,1\}^n$ tel que $f(w) = c_0$, alors que la minimisation avec contrainte consiste à trouver le point de plus petite valeur de f sur une partie de $\{0,1\}^n$, cette partie (feasible region) étant souvent décrite par des égalités ou inégalités de la forme $g(x) \leq 0$, $h(x) = 0$, où g et h sont d'autres fonctions sur $\{0,1\}^n$. Par exemple, il peut s'agir de trouver un point $w = (w_1, ..., w_1) \in \{0,1\}^n$ où f prend sa plus petite valeur et tel que w satisfasse une contrainte de cardinalité comme : $x_1 + ... + x_2 = q$ où q est un entier positif.

Plusieurs travaux ont été faits dans l'optimisation quadratique avec ou sans contraites : i) Des résultats et techniques ont été établis et testés pour la recherche effective d'un point de minimum global ; ces techniques sont appelées méthodes exactes mais de nombreux problèmes issus des applications pratiques sont de large dimension (le nombre n de variables est très grand) et sont intraitables par ces méthodes (voir [BHT 07], [BH 02]). Plusieurs autres techniques appelées heuristiques ont aussi été developpées pour la recherche d'un point de valeur assez petite mais sans garantie d'obtenir le point de minimum global (voir [BH,07] , [MF 02],[Be 98]), et ces derniers donnent parfois des solutions assez satisfaisantes aux problèmes de petites et grandes dimensions. ii) Il a été prouvé que certaines classes de problèmes sont solubles en temps polynomial et des algorithmes ont été developpés pour les traiter ; c'est le cas par exemple des fonctions dites sousmodulaires (que Hansen dans [Ha 74] ramenait à la forme équivalente dite "positive-négative") i.e lorsqu'on a $a_{ij} \leq 0$ pour tout $1 \leq i, j \leq 1$ dans (0.1.1). Des versions de l'algorithme dévellopé dans les années soixante par Ford et Fulkerson ([FJF 62]) et des algorithmes plus récents comme celui de Schrijver([Sc 00]) permettent de résoudre cette classe de problèmes en temps polynomial. Malgrés les nombreux travaux et progrès déjà faits dans l'optimisation quadratique en variables binaires, plusieurs problèmes restent ouverts. Même si le plus célèbre reste la question de savoir si la complexité de ce problème est polynomiale ou non, de nombreuses autres questions se posent sur : i) trouver certaines propriétés de la fonction f et de ses multiples reformulations ; ii) trouver des algorithmes efficaces pour de nombreux problèmes ayant des caractéristiques différentes ; iii) détecter de nouvelles classes de problèmes solubles en temps polynômial. Nous nous sommes intéressés à certaines de ces questions.

0.2. Présentation de notre travail centré sur le cas quadratique sans contrainte

Notre travail est essentiellement centré sur le cas quadratique sans contrainte. Nous utilisons des approches discrètes et des approches continues pour élaborer des propriétés et des techniques pour l'optimisation quadratique en variables $0-1$. La plus grande partie du travail est théorique, mais quelques algorithmes et quelques resultats expérimentaux préliminaires sont présentés.

Dans le chapitre 1, nous établissons essentiellement trois résultats :

i) soit $i = 1, 2, ..., n$, soient des réels $c_i < d_i$ et soit

$$R = [c_1, d_1]\times[c_2, d_2]\times...\times[c_n, d_n] = \{(a_1, .., a_n) \in \mathbb{R}^n, c_1 \leq a_1 \leq d_1, ..., c_n \leq a_n \leq d_n\} \tag{0.2.1}$$

$$VR = \{c_1, d_1\} \times \{c_2, d_2\} \times ... \times \{c_n, d_n\} \tag{0.2.2}$$

Notons par $\min_R(f)$ et $\min_{VR} f$ les sous ensembles respectifs de R et VR où f prend sa valeur minimale. En 1972, Rosenberg a montré (voir [Ro 72]) que : $\min_{VR}(f) \subset \min_R(f)$; et si $(a_1, ..., a_n) \in \min_R(f)$ satisfait $c_i < a_i < d_i$ $1 \leq i \leq n$, alors $(a_1, ..., a_{i-1}, x, a_{i+1}, ..., a_n) \in \min_R(f)$ pour tout $c_i \leq x \leq d_i$. Dans la première section du chapitre, nous étendons ce résultat à une plus grande classe de polynômes (non forcément multilinéaires) et nous en présentons d'autres propriétés.

ii) Afin de prouver que le problème de décider si un polynôme multilinéaire quadratique a un unique point de minimum dans $\{0,1\}^n$ est NP-dur, Pardalos et Somesh dans [PS 92] considèrent la fonction

$$F(x) = 2^n f(x) + \sum_{i=1}^{n} 2^{i-1} x_i \tag{0.2.3}$$

Dans la deuxième section, nous utilisons une idée similaire, et pour toute fonction f sur un ensemble fini $S \subset \mathbb{R}^n$ de points à coordonnées entières non-négatives, nous construisons une fonction u telle que $f + u$ a un unique point de minimum global dans S lequel est aussi un point de minimum global pour f dans S.

iii) Dans la troisième section, nous donnons quelques propriétés sur l'optimisation générale d'un polynôme quadratique sur tout ensemble $S \subset \mathbb{R}^n$ ayant un centre de symétrie (sphères, polytopes réguliers, cubes, sommets d'un cube etc.). En particulier nous présentons une inégalité linéaire satisfaite par tous les points de minimum global d'un tel problème. Au chapitre 2, nous exploitons l'inégalité linéaire $p(x) \leq 0$ décrite dans la troisième section du chapitre 1 pour le cas particulier de $H = \{0,1\}^n$ et $C = [0,1]^n$ qui ont pour centre de symétrie $r = (\frac{1}{2}, \frac{1}{2}, ..., \frac{1}{2})$: soit M l'ensemble de tous les points de minimum global du polynôme quadratique f sur H (ou sur C). Nous montrons que $p(y) \leq 0$ pour tout $y \in M$ tout en constatant que l'hyperplan p défini par $p(x) = 0$ passe par r si $p(x)$ n'est pas constant. En partant de cette contrainte linéaire particulière qui s'impose à l'optimisation sans contrainte de f, nous présentons d'autres inégalités linéaires, de sorte que leur ensemble caractérise tous les points de minimum global de f sur H en ce sens que : un point $y \in H$ est point de minimum global pour f si et seulement si y satisfait toutes ces inégalités linéaires, i.e., M est exactement l'intersection de H et du polytope convexe créé par les inégalités et C. Par des arguments similaires, nous trouvons une borne sur la cardinalité de l'ensemble de tous les points de minimun de f en fonction de n : on établit que $| M | \leq 3 \times 2^{n-2}$. La famille des inégalités linéaires présentée est finie, mais peut être exponentielle. Toutefois, la principale fonction linéaire $p(x)$ (qui se construit facilement) présente un espoir algorithmique considérable : par exemple, quelques simulations numériques à la fin du chapitre montrent que son point de minimum global est très souvent de valeur assez petite . Nous le testons comme point de départ d'un algorithme dans le chapitre qui suit. .

Dans le chapitre 3, nous observons comment chaque sommet s de l'hypercube unité peut être transformé en l'origine du cube par l'application φ^s définie par $\varphi^s(x) = x^s$ avec $x_i^s = 1 - x_i$ si $s_i = 1$ et $x_i^s = x_i$ si $s_i = 0$ pour $i = 1, 2, ...n$. Par une approche matricielle, nous étudions plusieurs propriétés de φ^s et ses effets sur le polynôme quadratique multilinéaire $f(x) = \frac{1}{2}x^tAx + b^tx$ dont l'expression devient $f(\varphi^s(x)) = \frac{1}{2}x^tA^sx + (b^s)^tx + c_s$ et demeure multilinéaire dans le nouveau système d'origine s. Par exemple : i) soit $H(x, y) = \|x - y\|^2$ la distance de Hamming entre les points x et y de H (i.e., le nombre de coordonnées en lesquelles x et y diffèrent) ; on dit que x et y sont voisins si $H(x, y) = 1$. Chaque point de H a n voisins; soit I^s la matrice diagonale d'ordre n définie par $I^s_{ii} = 1$ si $s_i = 0$, et $I^s_{ii} = -1$ si $s_i = 0$. Nous remarquons que $\varphi^s(x) = s + I^sx$ et en tirons des conséquences. L'une d'elles est que le vecteur $b^s = I^s(As + b) + f(s)(1, ..., 1)^t$ contient exactement les valeurs des n voisins de s, qui de manière élémentaire seraient calculées par n opérations de la forme $f(y) = \frac{1}{2}y^tAy + b^ty$. Une conséquence semblable en découle pour les point y tels que $H(s, y) = 2$. Nous présentons dans son état préliminaire une heuristique pour la minimisation quadratique sans contrainte en variables $0 - 1$. Des simulations numériques sont faites avec des problèmes tests proposés par Beasley (voir [Be 98]). Des résultats numériques encourageants et des comparaisons sont présentés avec des tableaux en fin de chapitre.

Au chapitre 4 : Nous présentons l'idée d'une heuristique par la descente du gradient, tirant profit des propriétés matricielles de nos fonctions quadratiques et du résultat décrit dans [Ro 72] qui établit l'équivalence de leur optimisation sur tout l'hypercube unité et sur l'ensemble de ses sommets. Nous proposons un algorithme et des résultats préliminaires. Nous travaillons encore à améliorer cet algorithme prometteur qui doit cependant contourner certaines limites numériques.

Au chapitre 5, nous présentons un nouvelle classe polynomiale d'optimisation quadratique $0 - 1$. On peut trouver en temps polynomial deux sommets u et m de l'hypercube unité telle que $f(m) < f(u)$ et $f(m) \leq 0 \leq f(u)$. Soit φ^u l'application définie au chapitre 3 qui transforme u en l'origine du cube et soit $f(\varphi^u(x)) = \frac{1}{2}x^tA^ux + (b^u)^tx + c_u$. On considère la matrice symétrique $Q = [q_{ij}]$ avec $q_{ij} = \frac{1}{2}a^u_{ij}$ et $q_{ii} = b^u_{ii}$ pour tout $1 \leq i, j \leq n$ et on suppose que sa valeur propre $\lambda_{min(Q)} < 0$. Nous prouvons que si $\frac{f(m)-f(u)}{\lambda_{min(Q)}}$ est assez large, alors f se minimise en temps polynomial.

0.3. Contributions et collaborations des autres

Les chapitres de cette thèse sont des articles en cours de finalisation pour la soumission. Les articles sont écrits en anglais pour être soumis dans les journaux. Mais un bref résumé en francais ainsi que les mots clés sont présentés au début de chaque article. Le professeur Ivo G. Rosenberg (département de mathématiques et statistiques de l'université de Montréal) qui a dirigé cette thèse est co-auteur avec moi dans les articles constituants les chapitres 1, 3 et 4. Une présentation plus précise est faite des contributions du co-auteur dans chaque article concerné. Endré Boros, directeur du centre de recherche opérationelle RUTCOR (Rutgers university, New-Jersey, USA) nous a aidé dans les chapitres 1 et 2, et sa contribution est aussi signalée dans les remerciements (acknowledgments) de ces articles. Nous avons eu quelques conversations utiles avec d'autres chercheurs comme Charles Audet de l'École Polytechnique de Montréal que nous signalons dans dans les remerciements des chapitres 1 et 2. La permission requise du co-auteur (Rosenberg Ivo) pour inclure les articles dans cette thèse se trouve à la fin de la thèse.

Chapitre 1

SOME REMARKS ON NON LINEAR $0 - 1$ OPTIMIZATION

AUTEURS : Calvin M. Wuntcha, Ivo Rosenberg

RÉSUMÉ ET CONTRIBUTIONS

Dans ce travail, nous donnons quelques observations générales sur l'optimisation non linéaire en variables $0 - 1$. Dans la section 1, nous étendons un résultat de 1972 concernant l'optimisation des polynômes multilinéaires sur les hyper-rectangles de $\mathbb{R}^n$ à une plus grande classe de fonctions. Dans la section 2, nous utilisons une idée de Pardalos et Somesh pour associer à chaque fonction f sur certains sous-ensembles finis de points de $\mathbb{R}^n$, une autre fonction u telle que $f + u$ possède un unique point de minimum global qui soit aussi un point de minimum global de f, sans que de nouveaux minima locaux soit créés. Dans la section 3, nous considérons l'optimisation générale sans contrainte d'un polynôme quadratique sur un ensemble possédant un centre de symétrie; en particulier, nous trouvons une inégalité linéaire satisfaite par tous les points de minimum de f.

Mot clés : Polynôme multilinéaire, minimum local.

CONTRIBUTIONS : C'est Rosenberg Ivo qui a proposé le Lèmme 1.1.1 ainsi que l'exemple qui suit la Remarque 3, pour montrer que changer les valeurs diagonales de la matrice associée à f peut créer de nouveaux points de minimum local, et doit se faire en tenant compte de celà. C'est aussi lui qui a traité l'exemple 2 qui suit le Corollaire 1.3.1.

ABSTRACT

In this paper, we give some general observations on non linear $0-1$ optimization. In Section 1, we extend a result given in 1972 on the optimization of multilinear polynomials defined on a hyperrectangle of $\mathbb{R}^n$ to a more general class of functions. In Section 2, using an idea of Pardalos and Somesh, we find for any real function f on a finite set of non negative integer points in $\mathbb{R}^n$, a function u such that : $f+u$ has a unique minimum point which is also a minimum point of f, and $f+u$ has no new local minimal point. In Section 3, we consider the optimization of any quadratic function f on a centrally symmetric set : for such a set, we find a linear inequality to be satisfied by each minimal points of f.

Key word : Multilinear polynomial, local minimal point.

INTRODUCTION

Let f be an n-variable polynomial with real coefficients and $S \subset \mathbb{R}^n$. Call $a \in S$ an *S-minimal point* of f if $f(a) \leq f(s)$ for all $s \in S$. If f and S are clear from the context, we simply refer to it as a minimal point. A local minimal point of f in S (ordinary) is a point $a \in S$ such that there exists an open set $V \subset \mathbb{R}^n$ with $f(a) \leq f(x)$ for all $x \in S \cap V$. A $0-1$ **local minimal** point of f is a point $w \in \{0,1\}^n$ such that $f(w) \leq f(x)$ for all $x \in \{0,1\}^n$ differing from w in only one coordinate. A **monomial of length** k is a product of k distinct variables and a constant. A real polynomial $f(x_1, ..., x_n)$ of n variables is **multilinear** if it is a sum of monomials in $x_1, ..., x_n$. The degree of a multilinear polynomial is the largest length of its monomials. For example, $h(x) = 2x_1x_3 - x_2x_3 + x_1 - 3x_3 + 1$ is a second degree (quadratic) multilinear polynomial, while $f(x) = -x_2^2 + 2x_3 + x_1 - 2$ is not multilinear. Multilinear polynomials appear in the minimization of real n variable polynomials over $\{0,1\}^n$. Rosenberg proved in [Ro 75] that the minimization of a multilinear polynomial f of a degree greater than two over $\{0,1\}^n$ can be reduced to the quadratic case by adding penalty monomials in additional variables, and it is also known (see [BH 02]) that the reduction is done in polynomial time. For this reason, the quadratic case is central. Let f be a multilinear polynomial. For $i = 1, 2, ..., n$, let $c_i < d_i$ and set

$$R = [c_1, d_1] \times [c_2, d_2] \times ... \times [c_n, d_n] = \{(a_1, .., a_n) \in \mathbb{R}^n, c_1 \leq a_1 \leq d_1, ..., c_n \leq a_n \leq d_n\} \tag{1.0.1}$$

$$VR = \{c_1, d_1\} \times \{c_2, d_2\} \times ... \times \{c_n, d_n\}. \tag{1.0.2}$$

Let $\min_R(f)$ and $\min_{VR}(f)$ be respectively the subsets of R and VR where f takes its minimal value. We can notice that if R and VR are nonvoid sets, then $\min_R(f)$ and $\min_{VR}(f)$ are nonvoid sets (in fact, VR is finite, R is compact, and f is continuous). In [Ro 72] Rosenberg proved the following results : (i) $\min_{VR}(f) \subset \min_R(f)$, and (ii) if $(a_1, ..., a_n) \in \min_R f$ satisfies $c_i < a_i < d_i$ for some $1 \leq i \leq n$, then $(a_1, ..., a_{i-1}, x, a_{i+1}, ..., a_n) \in \min_R(f)$ for all $c_i \leq x \leq d_i$. Both statements are consequences of the fact that for all $1 \leq i \leq n$, and all reals $a_1, ..., a_{i-1}, a_{i+1}, ..., a_n$, the one-variable polynomial $h(x) = f(a_1, ..., a_{i-1}, x, a_{i+1}, ..., a_n)$ is linear. In our first section, we extend this result to a larger class of polynomials and we also prove some additional related properties.

To prove that the problem of deciding whether a given quadratic n variable multilinear polynomial f has a unique minimal point on $\{0,1\}^n$ is NP-hard, Pardalos and Somesh in [PS 92] consider the function

$$F(x) = 2^n f(x) + \sum_{i=1}^{n} 2^{i-1} x_i. \tag{1.0.3}$$

In the second section, we use a similar idea to find for any real polynomial f on a finite set of non negative integer points $S \subset \mathbb{R}^n$, a function u such that $f + u$ has a unique S-minimal point which is also a S-minimal point of f, without creating new local minimal points.

Many sets usually considered in optimization like disks, spheres, ellipsoids, rectangles and regular polytopes are centrally symmetric. In the third section, we discuss some aspects of quadratic optimization on such sets. In particular, for any quadratic function f on such a set S, we find a linear inequality to be satisfied by each minimal point of f on S.

1.1. Quadratic functions having square variable terms with non positive coefficients.

Recall that any quadratic polynomial f can be expressed as

$$f(x) = \frac{1}{2}x^t Ax + b^t x + c \tag{1.1.1}$$

where A is a nonzero real symmetric matrix , b is a constant vector on $\mathbb{R}^n$ (considered as a column vector), and $x = (x_1, ..., x_n)^t$ is the vector of the variables. For optimisation purposes, we shall often assume that $c = 0$, then the expansion of $f(x)$ is

$$f(x) = \sum_{1 \le i < j \le n} a_{ij} x_i x_j + \sum_{i=1}^{n} \frac{1}{2} a_{ii} x_i^2 + \sum_{i=1}^{n} b_i x_i \tag{1.1.2}$$

Clearly, f is multilinear if $a_{ii} = 0$ for $i = 1, 2, ..., n$. So, in the quadratic case, the results of Rosenberg in [Ro 72] hold when all the diagonal entries of the matrix A are zero; but we shall try to generalise these results both for the quadratic and the non quadratic case. First recall that the symmetric matrix A has n real eigenvalues $\lambda_1, \lambda_2, ..., \lambda_n$, not necessarily distinct (a well known result that can be found in most textbooks on linear algebra). We also recall the well known results of the following theorem (see [PSU 88] and [Sn 05]).

Denote by $\mathbf{0}$ the vector $(0,...,0)^t$.

Theorem 1.1.1. *For a given* $f(x) = \frac{1}{2}x^tAx + b^tx$

(1) *x is a local minimal point in $\mathbb{R}^n$ if and only if :*

(i) grad $f(x) = \mathit{0}$ i.e. $Ax + b = \mathit{0}$ and

(ii) all eigenvalues of A are non negative.

(2) *f is convex if and only if (ii) is satisfied; and strictly convex if and only if all eigenvalues of A are positive and in that case, A is nonsingular and $A^{-1}b$ is the unique local and global minimal point for f.*

(3) *If A has a negative eigenvalue, f has no local minimal point in $\mathbb{R}^n$. Thus, any local minimal point of f on any compact subset K of $\mathbb{R}^n$ is in the boundary of K.*

For R and VR defined in (1.0.1) and (1.0.2), and for the quadratic case, we have the following theorem which is an extension of the main result of [Ro 72] :

Theorem 1.1.2. *Let $f(x) = \frac{1}{2}x^tAx + b^tx$ be an n-variable quadratic polynomial.*

(1) *If the diagonal entries of A are non positive then* $\min_{VR}(f) \subset \min_R(f)$.

(2) *If all the diagonal entries of A are negative then* $\min_{VR}(f) = \min_R(f)$.

Proof : *(1)* We know from the results of [Ro 72] that the result is true if f is multilinear, i.e., if $a_{ii} = 0$ for all $i = 1, 2, ..., n$. From linear algebra (see[HJ 85]), recall that the trace of the matrix A is equal to the sum of its eigenvalues, i.e.,

$$\sum_{i=1}^{n} a_{ii} = \sum_{i=1}^{n} \lambda_i. \tag{1.1.3}$$

Suppose that the left part of the equality is negative, at least one eigenvalue is negative. From Theorem 1.1.1 it follows that the minimal points of f on R are on the boundary B of R. Let $x = (x_1, ..., x_n)$ be a point of minimum of f on R. Now $x \in B$ implies that x must satisfy $x_k = c_k$ or $x_k = d_k$ for some $1 \leq k \leq n$. Now, let us choose such a coordinate x_k and fix the value of x_k in (1.1.2) ; we obtain a new polynomial f_1. Notice that the matrix of f_1 is obtained from A by deleting the k^{th} row and k^{th} column. We repeat this reduction as long as the reduced matrix has some $a_{ii} < 0$. At the end of the reduction, the remaining polynomial is multilinear and a point of minimum is reached in VR according to the results of [Ro 72]. Hence $\min_{VR}(f) \subset \min_R(f)$.

(2) If $a_{ii} < 0$, for all $i = 1, 2, ..., n$, we proceed as in 1). Suppose a point of minimum $x = (x_1, ..., x_n)$ of f on R and recall that there are n negative diagonal entries. For a chosen negative diagonal entry, a variable x_k, $1 \leq k \leq n$, is fixed ($x_k = c_k$ or $x_k = d_k$), the k^{th} row and the k^{th} column are deleted from the matrix; and all the diagonal entries of the new matrix remain negative. So, the reduction continues till all the variables are fixed, and this implies that $x \in VR$. Hence, $\min_{VR}(f) = \min_R(f)$. □

Remark 1. *A p-dimensional face F of R can be defined by : $1 < j_1 < ... < j_{n-p} < n$ and $u_{j_i} \in \{c_{j_1}, d_{j_i}\}$ $(i = 1, 2, ...n - p)$.*
$F = \{(a_1, ..., a_n) \in R, \quad a_{j_i} = u_{j_i} \quad \forall i = 1, 2, ..., n - p\}$. All the vertices of a given face of R may be minimal points of f while other interior points of the face are not minimal. Certainly, this will only happen when the polynomial expression of f corresponding to that face has squared variable terms .

Example 1.1 : Consider $f(x_1, x_2) = x_1x_2 - x_2^2 - x_1$ on $[0, 1]^2$. By a direct verification, $(0, 1)$, $(1, 0)$, $(1, 1)$ are the minimal points with value -1. On the edge of $\{0, 1\}^2$ defined by $x_2 = 1$, the polynomial is constant with value -1 and so each point of that face is a minimal point. On the edge of $\{0, 1\}^2$ defined by $x_1 = 1$, the two vertices $(1, 0)$ and $(1, 1)$ are the only minimal points, since the polynomial on the edge is $-x_2^2 + x_2 - 1$ and has no other minimal point.

Recall that given any reals $c_1, ..., c_n$, the polynomial

$$h(x) = \sum_{i=1}^{n} c_i x_i^2 - \sum_{i=1}^{n} c_i x_i \tag{1.1.4}$$

is identically 0 on $\{0,1\}^n$ and $F(x) = f(x) + h(x)$ equals $f(x)$ in $\{0,1\}^n$. However, F does not always satisfy $\min_{\{0,1\}^n}(f) \subset \min_{[0,1]^n}(F)$.

Example 1.2 : Take $f(x_1, x_2) = x_1x_2 - x_2 - x_1$ and $h(x_1, x_2) = x_1^2 - x_1$. Then $F(x_1, x_2) = x_1x_2 - x_2 - x_1 + x_1^2 - x_1$, and we see that $f(x)$ and $F(x)$ take the value 0 at $(0,0)$ and the value -1 at $(1,0)$, $(0,1)$ and $(1,1)$. But the minimal value $\frac{-5}{4}$ of F is reached at $(\frac{1}{2}, 1)$.

Remark 2. *It follows from Theorem 1.1.2 that* $\min_{\{0,1\}^n}(f) \subset \min_{[0,1]^n}(F)$ *if* $c_i \leq 0$ *for all* $i = 1, 2, ..., n$ *and* $\min_{\{0,1\}^n}(f) = \min_{[0,1]^n}(F)$ *if* $c_i < 0$ *for all* $i = 1, 2, ..., n$.

Lemma 1.1.1. *Let A be a non zero symmetric matrix and let $f(x) = \frac{1}{2}x^tAx + b^tx$ be any quadratic function. Let A and the augmented matrix $[A : b]$ have the same rank, and let u be a solution of $Ax = -b$, then :*

(i) the translation $x = y + u$ of $\mathbb{R}^n$ transforms $f(x)$ into $g(y) = \frac{1}{2}y^tAy^t + f(u)$;

(ii) in particular, if A is nonsingular, the result of i) holds exactly for $u = -A^{-1}b$.

Proof : (i) Let $x = y + u$. Then

$f(x) = \frac{1}{2}(y+u)^tA(y+u) + b^t(y+u) = \frac{1}{2}y^tAy + \frac{1}{2}y^tAu + \frac{1}{2}u^tAy + \frac{1}{2}u^tAu + b^ty + b^tu$ $= \frac{1}{2}y^tAy + f(u) + \frac{1}{2}(y^tAu + u^tAy) + b^ty$. Here, A is a symmetric matrix and y^tAu is a 1×1 symmetric matrix, so $(y^tAu)^t = u^tAy$ and similarly, $y^tb = b^ty$. Hence, $\frac{1}{2}(y^tAu + u^tAy) + b^ty = y^tAu + y^tb = y^t(Au + b) = 0$ since $Au = -b$.

(ii) If A is nonsingular, $u = -A^{-1}b$ is the unique solution to $Ax = -b$. □

Remark 3. *a) It follows from Lemma 1.1.1 that the minimization of a multilinear quadratic function* $f(x) = \frac{1}{2}x^tAx + b^tx$ *over the unit hypercube when* A *is nonsingular, can be reduced to the minimization of the quadratic function* $q(y) = y^tAy$ *over the hypercube obtained by translating the unit hypercube with the map :* $x = y + u$. *In fact,* $\frac{1}{2}y^tAy^t + f(u)$ *and* y^tAy *have the same minimal points on a given hypercube.*

b) For a multiliear quadratic function f *(*A *has a zero diagonal), we consider the case where* A *is singular. Considering Remark 2 with the function* $h(x)$ *defined in 1.1.4 and the related non multilinear function* $F(x) = f(x) + h(x) = \frac{1}{2}x^tBx + d^tx$ *. (*B *is the Symmetric matrix related to* F *and* d *the related constant vector), we may replace* A *by a nonsingular symmetric matrix* B *and reduce the minimization of* $f(x)$ *over the unit hypercube to the minimization of* $q(y) = y^tBy$ *over a hypercube, where* $y = x - u$ *and* $u = -B^{-1}d$. *For example, let* $\lambda_1 \leq \lambda_2 \leq ... \leq \lambda_n$ *be the eigenvalues of* A. *Recall that* A *is a non zero matrix* $\Leftrightarrow$ A *has some non zero eigenvalue.* $Tr(A) = 0 = \lambda_1 + \lambda_2 + ... + \lambda_n$ *since* A *has a zero diagonal, and this implies that* $\lambda_1 < 0$ *and* $\lambda_n > 0$. *Now, consider the function* $h(x)$ *defined in 1.1.4 with* $c_i = -\frac{1}{2}(\lambda_n - 1)$ *for all* $i = 1, 2, ..., n$. *Notice that* $B = A - (\lambda_n + 1)I$ *where* I *is the Identity matrix of the same rank with* A, *and* d *is defined by* $d_i = b_i + \lambda_n + 1$. *But* A *is a symmetric matrix and can be expressed as* $A = PDP^t$ *where* P *is an orthogonal matrix and* D *a diagonal matrix having the eigenvalues of* A *on its diagonal. Hence* $B = A - (\lambda_n + 1)I = PDP^t - (\lambda_n + 1)PIP^t = P[D - (\lambda_n + 1)I]P^t$ *shows that no eigenvalue of* B *is zero, i.e.,* B *is non singular.*

Many such manipulations on the diagonal entries of A may be done. However, for minimization with some methods like the gradient descent methods using interior points, we may need to be careful (as we shall see in the next example), since the new function $F(x) = \frac{1}{2}x^tBx + d^tx$ may create some additional new local minima (in the sens of natural topology) on the unit hypercube, and those new local minima may 'attract' and stop the descent, which is not favorable. We demonstrate this in the following example on $[0,1]^4$.

Example 1.3 : Recall that a $0-1$ local minimal point for a multilinear polynomial f is a $0-1$ point w such that $f(w) \leq f(y)$ for all $0-1$ points y differing from w only in one coordinate $w_j \neq y_j$. Set

$$A = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix} \tag{1.1.5}$$

and $b = (-1,1,1,1)^t$. It can be verified that $det(A) = 0$ while

$$det \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} = 2 \tag{1.1.6}$$

The expansion $f(x) = x_1x_2 + x_1x_3 + x_2x_3 + x_2x_4 + x_3x_4 - x_1 + x_2 + x_3 + x_4$ shows that f satisfies : $f(0,0,0,0) = 0$ and $f(1,0,0,0) = -1$. Then, since f is linear on the edge defined by $x_2 = x_3 = x_4 = 0$ (the edge relating $(0,0,0,0)$ to $(1,0,0,0)$) it comes that $(0,0,0,0)$ is not local minimal point of f on $[0,1]^4$ with the natural topology. Now, set

$$F(x) = -2x_1^2 + 2x_1 + f(x) = \frac{1}{2}x^tBx + d^tx \tag{1.1.7}$$

where B differs from A only in $b_{11} = -1$ and $d = (1,1,1,1)^t$. Here, $det(B) = -2$ and B is non singular. We know that F and f agree on $\{0,1\}^4$. However $(0,0,0,0)$ is an ordinary local minimal point for F on $[0,1]^4$. To see it, notice that $F(x) = (-2x_1^2 + x_1) + (x_1x_2 + x_1x_3 + x_2x_3 + x_2x_4 + x_3x_4 + x_2 + x_3 + x_4)$. The second component of the sum is non negative on $[0,1]^4$. The first component $(-2x_1^2 + x_1) = x_1(1-2x_1)$ is also non negative for all $0 \leq x_1 \leq \frac{1}{2}$. Hence

$F(0,0,0,0) = 0 \leq F(x)$ for all $x \in [0,1]^4$ such that $0 \leq x_1 \leq \frac{1}{2}$. So, in the natural toplogy sens, $(0,0,0,0)$ is a local minimal point for F in $[0,1]^4$ (can stop a gradient descent minimization of F in the hypercube), but is not a local minimal point for f in $[0,1]^4$ (does not stop a gradient descent minimization of f in the hypercube). So, for new local minima not to be created on the unit hypercube when manipulating the diagonal elements of A, some care should be taken.

To conclude this section, let us examine the optimization of any polynomial of the form

$$F(x) = f(x) + q(x) \tag{1.1.8}$$

where $f(x)$ is a multilinear polynomial of any degree and

$$q(x) = \sum_{i=1}^{n} c_i x_i^2. \tag{1.1.9}$$

It is well known (see [PSU 88] and [Sn 05]) that for any n-variable function F having first and second derivatives at $x \in \mathbb{R}^n$, the point x is an ordinary local mimimal point of F on R if and only if :

(i) $grad\ F(x) = \mathbf{0}$ and

(ii) the hessian matrix $H_F = \left[\frac{\partial F}{\partial x_i \partial x_j}\right]_{1 \leq i,j \leq n}$ is positive semidefinite at x.

In fact, in the quadratic case, the matrix A is the hessian matrix. So with the notations of $F(x)$ as defined in (1.1.8), $q(x)$ and c_i, $i = 1, 2, ..., n$ as presented in 1.1.9, we have the following proposition which generalizes the results of [Ro 72] :

Proposition 1.1.1.

(1) *If* $c_i \leq 0$ *for* i=1,2,...,n ; *then* $\min_{VR}(F) \subset \min_R(F)$.

(2) *If* $c_i < 0$ *for* $i = 1, 2, ..., n$; *then* $\min_{VR}(F) = \min_R(F)$.

Proof : Concerning the trace of H_F observe that :

$$Tr(H_F) = \sum_{i=1}^{n} \frac{\partial F}{\partial x_i \partial x_i} = \sum_{i=1}^{n} c_i \leq 0 \tag{1.1.10}$$

The rest of the proof follows as the proof of Theorem 1.1.2. □

1.2. Perturbation of discrete functions to keep a unique minimal point and not create new local minimal points.

To prove that the problem of deciding when an n-variable multilinear quadratic polynomial f has a unique minimal point in $\{0,1\}^n$ is NP-hard, Pardalos and Somesh in [PSU 88] consider $f(x) = \frac{1}{2}x^t Ax + b^t x$, with integer coefficients. They prove that

$$F(x) = 2^n f(x) + \sum_{i=1}^{n} 2^{i-1} x_i \tag{1.2.1}$$

has a unique minimal point in $\{0,1\}^n$ which is also a minimal point of f. Then

$$g(x) = \frac{1}{2^n} F(x) = f(x) + \sum_{i=1}^{n} \frac{1}{2^{n+1-i}} x_i = f(x) + 2^{-n} \sum_{i=1}^{n} 2^{i-1} x_i \tag{1.2.2}$$

has the same property.

For a real-valued non constant function f defined on a finite set S. Set

$$\varepsilon_f = \min\{|f(x) - f(z)| : x, z \in S, f(x) \neq f(z)\} \tag{1.2.3}$$

i.e. ε_f is the smallest positive difference between the values of f on S. For example, if f has only integer values on S, then ε_f is a positive integer.

Definition 1. *We shall say that a function u defined on a finite set S is a* ***perturbator*** *for the function f on S, if max $\{| u(x) - u(y) |: x, y \in S\} < \varepsilon_f$.*

Let us study some properties of the perturbation $f + u$ of f :

Lemma 1.2.1. *If u is an injective perturbator for f, then $f+u$ is an injective function and hence $f+u$ has a unique minimal point in S.*

Proof : Let f be a function on the finite set S, let u be a perturbator for f on S and let $x, y \in S$. We shall consider the proof in two steps : (α) if $f(x) \neq f(y)$, and (β) if $f(x) = f(y)$.

(α) If $f(x) \neq f(y)$, we can suppose that $f(x) < f(y)$. Then

$u(x) - u(y) \;\leq \mid u(x) - u(y) \mid < \varepsilon_f \;\leq\; |f(x) - f(y)| = f(y) - f(x)$

showing that $f(x) + u(x) < f(y) + u(y)$.

(β) : Let $x \neq y$ and $f(x) = f(y)$. Since u is injective, we can suppose that $u(x) < u(y)$.

Then $f(x) + u(x) < f(y) + u(y)$.

So for all distinct points $x, y \in S$, $f(x) + u(x) \neq f(y) + u(y)$. This proves that $f+u$ is injective and hence it has a unique minimal point. □

We obtain the following consequence :

Corollary 1.2.1. *If f is a real non constant function on $\{0,1\}^n$, then*

$$u(x_1, ..., x_n) = \varepsilon_f 2^{-n} \sum_{i=1}^{n} 2^{i-1} x_i \tag{1.2.4}$$

is an injective perturbator for f on $S = \{0,1\}^n$ and hence $f+u$ has a unique minimal point on $\{0,1\}^n$.

Proof : It is well known that every $k \in \{0, 1, 2, ..., 2^{n-1}\}$ has a unique binary representation as $k = x_1 + 2x_2 + ... + 2^{n-1}x_n$ with $x_1, ..., x_n \in \{0,1\}^n$. Obviously, $v(x_1, ..., x_n) = x_1 + 2x_2 + ... + 2^{n-1}x_n$ is an injective map from $\{0,1\}^n$ onto $\{0, 1, 2, ..., 2^{n-1}\}$. Now, $u(x_1, ..., x_n) = \varepsilon_f 2^{-n} v(x_1, ..., x_n)$. Clearly, u is also injective and $\mid u(x) - u(y) \mid \leq \mid u(x) - u(0, ..., 0) \mid < \varepsilon_f 2^{-n} 2^n = \varepsilon_f$. Thus, u is an injective perturbator for f. By Lemma 1.2.1, $f+u$ has a unique minimum on

$\{0,1\}^n$. □

Let S be a finite set of non negative integer points in $\mathbb{R}^n$ and let $p_1, p_2, ..., p_n, ...$ be the natural encreasing sequence of prime numbers. Consider a real

$$m \geq max\left\{p_1^{y_1-x_1}p_2^{y_2-x_2}...p_n^{y_n-x_n} : (x_1,..,x_n),(y_1,...,y_n) \in S\right\} \tag{1.2.5}$$

and let ln be the natural logarithm function. Take $E > ln(m)$ (for example we can take $m = p_n^{nx_{max}}$, where x_{max} is the greatest possible coordinate value of the points in S and $E = nx_{max}ln(p_n) + 1$). We have the following result :

Theorem 1.2.1. *The function below is an injective perturbator for f :*

$$u(x_1, ..., x_n) = \varepsilon_f E^{-1} ln(p_1^{x_1}...p_n^{x_n}) = \varepsilon_f E^{-1} \sum_{i=1}^{n} x_i ln(p_i). \tag{1.2.6}$$

Proof : Let $x = (x_1, ..., x_n) \in S$ and $y = (y_1, ..., y_n) \in S$ be two distinct points. Then

$u(y) - u(x) = \varepsilon_f E^{-1}\left[ln(p_1^{y_1}...p_n^{y_n}) - ln(p_1^{x_1}...p_n^{x_n})\right]$

$=\varepsilon_f E^{-1} ln(p_1^{y_1-x_1}...p_n^{y_n-x_n}) = \varepsilon_f E^{-1} ln(p_1^{y_1-x_1}p_2^{y_2-x_2}...p_n^{y_n-x_n})$.

Here, $u(y)-u(x) = 0$ exactly if $p_1^{y_1-x_1}p_2^{y_2-x_2}...p_n^{y_n-x_n} = 1$, leading to $y_i - x_i = 0$ for all $i = 1, 2, ..., n$ (since the prime numbers are different) and this leads to $y = x$. So, for $x \neq y$, $u(x) \neq u(y)$ and hence u is injective. Now, for $u(y) > u(x)$, we get $|u(y) - u(x)| = u(y) - u(x) \leq \varepsilon_f E^{-1} ln(p_1^{y_1-x_1}p_2^{y_2-x_2}...p_n^{y_n-x_n}) < \varepsilon_f E^{-1} l(m) < \varepsilon_f E^{-1} E = \varepsilon_f$ proving that u is an injective perturbator for f. □

Theorem 1.2.2. *Let u be any injective perturbator for a non constant function f on S and M the set of all minimal points for f on S.*

(1) *The unique minimal point of $f + u$ on S is a minimal point of f on S. It is precisely the minimal point of u on M.*

(2) *For* $S = \{0,1\}^n$, *any* $0-1$ *local minimal point for* $f+u$ *is a* $0-1$ *local minimal point for* f.

Proof : *(1)* M is the set of all the S-minimal points of f. u is injective. Let w be the unique minimal point of u in M. For $y \in S$, we have $u(w) - u(y) \leq |u(w) - u(y)| < \varepsilon_f \leq f(y) - f(w)$. So $f(w) + u(w) < f(y) + u(y)$. This shows that w is the unique minimal point of $f + u$ on S.
(2) Let $S = \{0,1\}^n$ and $w \in S$. Suppose that w is not a $0-1$ local minimal point of f ; then there is a $0-1$ point $y \in S$ different from w by only one coordinate and such that $f(y) < f(w)$. So, $u(y) - u(w) \leq |u(w) - u(y)| < \varepsilon_f \leq f(w) - f(y)$; then $f(y) + u(y) < f(w) + u(w)$ shows that w is not a $0-1$ local minimal point for $f + u$. □
The claim 2) can be extended to the set S of all vertices of a given polytope.

1.3. Some properties of quadratic optimization on a centrally symmetric set.

Many sets like disks, spheres, ellipses, rectangles and regular polytopes are centrally symmetric and are often considered in optimization. We discuss some aspects of quadratic optimization on such sets.
In this section, let $S \subset \mathbb{R}^n$ be a nonvoid set with a symmetric center r which means that $x \in S \Rightarrow 2r - x \in S$, where the operations are of the real vector space $\mathbb{R}^n$. Thus, the self map $x \mapsto x' = 2r - x$ is an involution on S (i.e. $(x')' = x$ for all $x \in S$). In particular, x' is a permutation of S. Notice that r is the unique invariant point of the map (if $r \in S$) because $x' = 2r - x = x \Rightarrow r = x$. To simplify the notations, we set $e = 2r$. We shall assume that S is an n-dimensional set, i.e. the convex hull $Conv(S)$ of S is a n-dimentional set. We also assume that the set $\{f(x) : x \in S\}$ of $\mathbb{R}$ has a least element c_0 and set $M = \{x \in S : f(x) = c_0\}$.

If $f(x) = \frac{1}{2}x^tAx + b^tx$ is a quadratic function, we consider

$$f'(x) = f(x') = f(e - x). \tag{1.3.1}$$

We obtain the following :

Theorem 1.3.1. *Let $f(x) = \frac{1}{2}x^tAx + b^tx$ be a quadratic function and $p(x) = f(x) - f'(x)$. Then :*

(1) *$p(x) = d^tx - c$, where $d^t = e^tA + 2b^t$ and $c = f(e)$;*

(2) *$p(r) = 0$;*

(3) *$p(x)$ constant $\Leftrightarrow p(x) = 0$ for all $x \in \mathbb{R}^n \Leftrightarrow e^tA = -2b^t$;*

(4) *If p is not constant, then the hyperplane $P = \{x \in \mathbb{R}^n : p(x) = 0\}$ separates S in two nonvoid subsets $p_- = \{x \in S : p(x) \leq 0\}$ and $p_+ = \{x \in S : p(x) > 0\}$; and $M \subset p_-$.*

Proof : *(1)* $f'(x) = f(e - x) = \frac{1}{2}(e^t - x^t)A(e^t - x) + b^t(e - x)$. Using the fact that A is a symmetric matrix and x^tAe is symmetric 1×1 matrix, the expansion of $f'(x)$ yields

$f'(x) = f(x') = \frac{1}{2}x^tAx - e^tAx + \frac{1}{2}e^tAe - b^tx + b^te = \frac{1}{2}x^tAx - (e^tAx + b^tx) + f(e)$.

Therefore

$p(x) = f(x) - f(x') = (\frac{1}{2}x^tAx + b^tx) - \left[\frac{1}{2}x^tAx - (e^tAx + b^tx) + f(e)\right]$

$= (e^tA + 2b^t)x - f(e)$.

(2) From $r' = r$, we obtain $f(r) = f'(r)$ i.e. $p(r) = 0$.

(3) It follows from 1) and 2).

(4) Suppose that $p(x)$ is not constant. It is clear from the definition that p_+ and p_- are two disjoint sets whose union is S. We shall now prove that p_- and p_+ are nonvoid. Recall that S is a nonvoid n-dimensional set and the hyperplane P defined by $p(x) = 0$ is a convex $n - 1$ dimensional set. So, $S \subset P$ is not possible, otherwise we have the contradiction $Conv(S) \subset P$ by convexity.

So there exists $y \in S$ such that $p(y) \neq 0$. But $p(y') = f(y') - f((y')') = f(y') - f(y) = -(p(y) - p(y')) = -p(y)$. Hence, $y \in p_+ \Leftrightarrow y' \in p_-$. Thus p_- and p_+ are nonvoid. Finaly, suppose that $x \in M$ is a minimal point of f. Then, $f(x) \leq f'(x) \Leftrightarrow f(x) - f'(x) \leq 0 \Leftrightarrow x \in p_-$; so $M \subset p_-$. □

Corollary 1.3.1. *Let $r = (0, 0, ..., 0)$ be the symmetry center of the set S. Then :*

(1) *Then $p(x) = 2b^t x$;*

(2) *$b^t x \leq 0$ for all $x \in M$;*

(3) *If $M \cap P$ is finite then $|(M \cap P) \setminus \{r\}|$ is even.*

Proof : *(1)* and *(2)* follow from Theorem 1.3.1.

(3) Let $M \cap P$ be finite and $y \in M \cap P$. Then $f(y) - f(y') = p(y) = 0$ and $f(y) = f(y')$ is the minimal value of f on S. Here, $y = y' \Leftrightarrow y = r$ and the elements of $(M \cap P) \setminus \{r\}$ come in pairs. □

Example 1.4 : 1) Let $S = \{0,1\}^2$. Clearly S is centrally symmetric with $r = (\frac{1}{2}, \frac{1}{2})^t$. Let

$$f(x_1, x_2) = -x_1 x_2 + x_1 = \frac{1}{2}(x_1, x_2) \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + x_1. \qquad (1.3.2)$$

The minimal value of f on S is 0, and it is reached at $(0,0)$, $(0,1)$ and $(1,1)$. Next

$$p(x) = (1,1) \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + 2x_1 = x_1 - x_2. \qquad (1.3.3)$$

$p(0,0) = p(1,1) = 0$. Therefore $(0,0), (1,1) \in M \cap P$ and $(0,1) \in M \setminus P$.

2) Consider the following function :

$$f(x_1, x_2) = x_1^2 + x_2^2 - 2x_2 - x_1 = \frac{1}{2}(x_1, x_2)\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + (-1, -2)\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \tag{1.3.4}$$

on the unit disk $C \subset \mathbb{R}^2$. The center of symmetry of C is $e = (0,0)^t$; $p(x) = 2b^t \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = -2x_1 - 4x_2$. See that $f(x_1, x_2) = (x_1^2 - x_1) + (x_2^2 - 2x_2)$. In $\mathbb{R}^2$, f has a unique critical point (where the gradient is zero) at $(\frac{1}{2}, 1)$ outside C, and thus, it takes its least value on C on the unit circle. Now, set $x_1 = sin(\theta)$ and $x_2 = cos(\theta)$ and remark that $f(x_1, x_2) = g(cos(\theta), sin(\theta)) = 1 - cos(\theta) - 2sin(\theta)$ satisfies $g'(\theta) = sin(\theta) - 2cos(\theta)$ and its critical points satisfy $sin(\theta) = 2cos(\theta)$. Then $cos^2(\theta) + sin^2(\theta) = 1$ implies that $5cos^2(\theta) = 1$. So $(\frac{1}{\sqrt{5}}, \frac{2}{\sqrt{5}})$ and $(\frac{-1}{\sqrt{5}}, \frac{-2}{\sqrt{5}})$ are the two stationary points. Finally, $g(\frac{1}{\sqrt{5}}, \frac{2}{\sqrt{5}}) = 1 - \sqrt{5} < 0$ and $g((\frac{-1}{\sqrt{5}}, \frac{-2}{\sqrt{5}}) = 1 + \sqrt{5} > 0$. Thus $f(x_1, x_2)$ on C has its unique minimal point at $(\frac{1}{\sqrt{5}}, \frac{2}{\sqrt{5}})$ and $p(\frac{1}{\sqrt{5}}, \frac{2}{\sqrt{5}}) = \frac{-10}{\sqrt{5}} < 0$. (Notice from the two eigenvalues in the symmetric associated matrix that f is strictly convex on C).

3) For $x = (x_1, x_2, x_2) \in \mathbb{R}^3$, consider

$$f(x) = 2x_1x_2 + x_3^2 - 3x_1 + x_2 - 2x_3 = \frac{1}{2}(x_1, x_2, x_3)\begin{pmatrix} 0 & 2 & 0 \\ 2 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} + (-3, 1, -2)\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \tag{1.3.5}$$

and $S = [-1, 1]^3$ (f is not convex in this case, in fact 1, 2, and -2 are the eigenvalues). The origin is the symmetry center and as above, the linear part $-3x_1 + x_2 - 2x_3 = \frac{1}{2}p(x_1, x_2, x_3)$. We can write $f(x_1, x_2, x_3) = (2x_1 2x_2 - 3x_1 + x_2) + (x_3^2 - x_3)$. Given the shape of S, we can minimize the two components independently : $x_3^2 - x_3$ has its minimum at $x_3 = \frac{1}{2}$; the function $g(x_1, x_2) = (2x_1 2x_2 - 3x_1 + x_2)$ is linear in x_1 and x_2 ; $g(-1, y) = 3 - y$ shows that $4 = g(-1, -1) > g(-1, 1) = 2$.Similarly, $g(1, y) = -3 + 3y$ shows that

$g(1,-1) = -6 < 0 = g(1,1)$. Thus $(-1,1)$ is the unique minimum of g and $(-1,1,1)$ is the unique point of minimum of f. We have $p(1,-1,1) = -6 < 0$. Minimizing f on the unit sphere may be less easy; but we know at least in which half of the sphere the minimal points are.

We resume Corollary 1.3.1 for a $0-1$ quadratic minimization as follow :

Corollary 1.3.2. *Let $f(x) = \frac{1}{2}x^tAx + b^tx$ be any quadratic function, $S = \{0,1\}^n$ and let M denotes the set of the points of minimal value of f. Then :*

(1) *$p(x) = f(x) - f'(x) = d^tx - c$, where $d^t = (1,1,...,1)A + 2b^t$ and $c = f(1,1,...,1)$.*

(2) *p is constant $\Leftrightarrow (1,1,...,1)A = -2b^t \Leftrightarrow p(x) = 0$ for all $x \in \mathbb{R}^n$.*

(3) *If $p(x)$ is not constant, then $p(\frac{1}{2},\frac{1}{2},...,\frac{1}{2}) = 0$ and the hyperplane $P = \{x \in \mathbb{R}^n,\ p(x) = 0\}$ partitiones $[0,1]^n$ in two convex sets p_+ and p_- such that $M \subset p_-$ and $|M \cap P|$ is even.*

Many sets of $\mathbb{R}^n$ have $r = (\frac{1}{2},\frac{1}{2},...,\frac{1}{2})^t$ as their symmetry center. Some examples are : the polytopes $[0,1]^n, [-1,2]^n, [-1,2]^d \times [0,1]^{n-d}$ for an integer $1 < d < n$, the set of all vertices of each; any disk centered at $r = (\frac{1}{2},\frac{1}{2},...,\frac{1}{2})$. Let $f(x) = \frac{1}{2}x^tAx + b^tx$ where A is a symmetric matrix with a zero diagonal and $F(x) = f(x) + h(x)$, where $h(x) = \sum_{i=1}^n c_i(x_i^2 - c_i)$. We have

$$F(x) = f(x) + \frac{1}{2}(x^tDx^t - v^tx) \tag{1.3.6}$$

where D is a diagonal matrix with diagonal $v = (c_1,...,c_n)^t$. We can observe the following :

Corollary 1.3.3. *Let f and F be the polynomials defined in (1.3.6) and $S \subset \mathbb{R}^n$ a set with the symmetry center $r = (\frac{1}{2}, \frac{1}{2}, ..., \frac{1}{2})^t$.*
Then $p(x) = f(x) - f'(x) = F(x) - F'(x)$.

Proof : Recall that $e = 2r = (1, ..., 1)^t$. From Theorem 1.3.1 we have, on one hand, $f(x) = \frac{1}{2}x^tAx + b^tx \Rightarrow f(x) - f'(x) = (e^tA + 2b^t)x - f(e)$; and on the other, $F(x) = \frac{1}{2}x^tAx + \frac{1}{2}x^tDx + (b^t - \frac{1}{2}v^t)x = \frac{1}{2}x^tBx + l^tx$ with $B = A + D$ and $l = b - \frac{1}{2}v$, implies $F(x) - F'(x) = (e^tB + 2l^t)x - f(e) = (e^tA + e^tD + 2b^t - v^t)x - f(e)$. But $e^tD = (1, ..., 1)D = v$. Hence $F(x) - F'(x) = e^tA + 2b^t)x - f(e) = f(x) - f'(x) = p(x)$.□

Like $\{0, 1\}^n$, many finite sets of non negative integers are centrally symmetric (for example, a set of the form VR in (1.0.2), where c_i, d_i for $i = 1, ..., n$ are non negative integers, is centered at $r = \left[\frac{1}{2}(c_1 + d_1), \frac{1}{2}(c_2 + d_2), ..., \frac{1}{2}(c_n + d_n)\right]$). We have the following observation on such sets :

Corollary 1.3.4. *Let $S \subset \mathbb{R}^n$ be an n-dimensional finite set of non negative integers with symmetry center r such that $r \notin S$. Let $|S| = k$, $g(x) = f(x) + u(x)$ be an injective perturbation for f, $p(x) = g(x) - g(x')$, $P = \{x \in \mathbb{R}^n : p(x) = 0\}$, $p_- = \{x \in \mathbb{R}^n : p(x) < 0\}$ and $p_+ = \{x \in \mathbb{R}^n : p(x) > 0\}$. Then :*

(1) $P \cap S = \emptyset$;

(2) $|p_+ \cap S| = |p_- \cap S| = \frac{k}{2}$;

(3) *at least half of the points of minima of f belong to p_- ;*

(4) *the unique point of minimum of g belongs to p_-.*

Proof : *(1)* g is injective on S and r is the unique point satisfying $x = x'$. Since $r \notin S$, we have $g(x) \neq g(x')$ i.e. $g(x) - g(x') \neq 0$ for all $x \in S$.

(2) $p(x') = g(x') - g(x) = -(g(x) - g(x') = -p(x)$. By the same token, $x \in p_- \Leftrightarrow x' \in p_+$ for all $x \in S$, proving the claim.

(3) Suppose that a minimal point x of f on S belongs to p_+ i.e. $g(x) > g(x')$. Then $f(x') \geq f(x)$, and $f(x) + u(x) = g(x) > g(x') = f(x') + u(x')$. We obtain $u(x) - u(x') > f(x') - f(x) \geq 0$. This implies that $f(x') - f(x) = 0$; otherwise, $|u(x) - u(x')| \geq u(x) - u(x') > f(x') - f(x) \geq \varepsilon_f$ which is a contradiction with the definition of a perturbator. So $f(x') = f(x)$ and $g(x) \neq g(x')$. Hence, from the proof of 2) exactly one of x and x' belongs to p_-. This proves the claim.

(4) Let x be the unique minimal point of g on S. Then $g(x) < g(x')$ and $p(x) < 0$.

□

1.4. Acknowledgements

We acknowledge many fruitful discussion with Endré Boros about the linear inequality $p(x) \leq 0$ specially in the case of the unit hypercube; his remark that many such inequalities should exist, led us to generalize it to all centrally symmetric sets. We also thank him for efficient remarks on some initial idea that we had on the perturbation $f + u$. We also acknowledge some usefull discussion on linearization with Charles Audet of the École Polytechnique de Montréal.

Chapitre 2

A CHARACTERIZATION OF MINIMAL POINTS OF A QUADRATIC $0-1$ FUNCTION

AUTEUR : Calvin Wuntcha

RÉSUMÉ

Soit f un polynôme multilinéaire à n variables et soit M l'ensemble de tous les points de minimum global de f sur $\{0,1\}^n$. Déterminer M ou même un élément de M est un problème NP-dur bien connu en optimisation discrète. Nous construisons une famille L d'inégalités linéaires qui caractérisent M en ce sens que $C \cap \{0,1\}^n = M$ où C est le polytope convexe créé par les hyperplans de $\mathbb{R}^n$ associé aux inégalités. Nous réduisons L de deux manières en un sous-ensemble ayant la même propriété que L. Nous montrons aussi que $\mid M \mid \leq 3 \times 2^{n-2}$ pour tout polynôme quadratique sur $\{0,1\}^n$. Nous terminons par quelques remarques et résultats expérimentaux sur les valeurs du point de minimum global de l'une des fonctions linéaires (la principale) qui caractérisent M.

Mots-clés : Multilinéaire, NP-dur, polytope

Abstract

Let $f(x_1, x_2, ..., x_n)$ be a quadratic square-free real polynomial, let u be its minimal value on $\{0,1\}^n$, and let $M = \{x \in \{0,1\}^n : f(x) = u\}$. Finding M or even just one $x \in M$ is known as an NP-hard problem in discrete optimization. We construct a finite set L of linear inequalities defining a convex polytope C such that $C \cap \{0,1\}^n = M$. In two ways, we reduce the initial set L to a subset with the same property. We also show that $| M | \leq 3 \times 2^{n-2}$ for every $0-1$ quadratic function f.

Introduction

An n-ary pseudo-boolean function is a map f from $\{0,1\}^n$ into the reals. It is well known that f can be represented by a real square-free polynomial. Denote by $u = min(f)$ the least value of f on $\{0,1\}^n$ and set $M = \{x \in \{0,1\}^n : f(x) = u\}$. Generally, the following tasks are known to be difficult : (i) finding u ; (ii) finding some $x \in M$; (iii) finding the entire M. The case of f given by a quadratic square free polynomial

$$f(x) = \sum_{i<j} a_{ij} x_i x_j + \sum_{i=1}^{n} b_i x_i + c \qquad (2.0.1)$$

(where a_{ij}, b_i, $1 \leq i, j \leq n$, and c are reals, at least one $a_{ij} \neq 0$, and $x = (x_1, x_2, ..., x_n) \in \{0,1\}^n$ is the vector of variables) is the first non trivial case. Nevertheless, (i), (ii) and (iii) for a quadratic pseudo-boolean function (qpbf) are known to be NP-hard. With the help of many additional variables (see [Ro 75]), the above problems for a square free polynomial of degree greater than 2 can be reduced to the quadratic case. It is known (see [BH 02] and [Ro 72]) that optimizing (2.0.1) on the set $\{0,1\}^n$ of $0-1$ vectors can equivalently be done by optimizing the corresponding polynomial in the unit hypercube $[0,1]^n$. For evident optimization reasons, one may discard the constant c in the polynomial .

For any $i, j \in N = \{1, 2, ...n\}$ and for $x_i, x_j \in \{0, 1\}$, the following inequalities are known and evident :

$$\begin{aligned} 0 &\leq x_i x_j; \\ x_i + x_j - 1 &\leq x_i x_j; \\ x_i x_j &\leq x_i; \\ x_i x_j &\leq x_j \end{aligned} \tag{2.0.2}$$

For any $I \subseteq N$ define a self map of $[0,1]^n$ by $x \mapsto x^I$, by setting $x_i^I = 1 - x_i$ if $i \in I$ and $x_i^I = x_i$ otherwise. Clearly, x^I is an involution on the unit hypercube. When $I = N$, $x^I = x' = (1 - x_1, 1 - x_2, ..., 1 - x_n)$ is the complement of x, and $e = (\frac{1}{2}, \frac{1}{2}, ..., \frac{1}{2})$ is the unique fixed point of the map $x \rightarrow x'$.

In this spécial case where $I = N$, we remark that the quadratic parts of $f(x')$ and $f(x)$ are identical and hence, the function $p(x) = f(x) - f(x')$ is linear. It is obvious that if x^* is any $0-1$ minimal point of f, then $p(x^*) = f(x^*) - f((x^*)') \leq 0$. So, when p is not constant, $p(x) = 0$ defines a hyperplane in $\mathbb{R}^n$ such that all the minimal points of f are "under" it in the sense that they are all in the half-space defined by $p(x) \leq 0$. In addition, this hyperplane meets the hypercube; in fact, $e = (\frac{1}{2}, \frac{1}{2}, ..., \frac{1}{2}) = e'$, and so $p(e) = f(e) - f(e') = 0$ which means that p passes through the center of $[0,1]^n$.

In what follows, we characterize the set M of minimal points for a given qpbf f. We use the inequalities (2.0.2) to construct a set of inequalities similar to $p(x) \leq 0$ and such that the set M is exactly the set of $0 - 1$ vectors satisfying all the inequalities.

2.1. A SET OF LINEAR FUNCTIONS AND ITS PROPERTIES

We consider the qpbf f as defined in (2.0.1). For a given $I \subseteq N = \{1, 2, .., n\}$, define a new qpbf f^I by setting $f^I(x) = f(x) - f(x^I)$, with x^I defined in the introduction. Evidently, $f(x^I)$ is a quadratic polynomial obtained by replacing x_i

in $f(x)$ by $1-x_i$ for all $i \in I$ and so, f^I is a difference of two quadratic functions.

Example 2.1 : Let $n = 3$ and

$$f(x) = 3x_1x_2 - 4x_1x_3 + 3x_2x_3 + x_1 - 2x_2 + 3x_3\,;$$

For $I = \{1, 2\}$,

$f(x^I) = 3(1-x_1)(1-x_2) - 4(1-x_1)x_3 + 3(1-x_2)x_3 + (1-x_1) - 2(1-x_2) + 3x_3$

$= 3x_1x_2 - 3x_1 - 3x_2 + 3 + 4x_1x_3 - 4x_3 - 3x_2x_3 + 3x_3 + 1 - x_1 - 2 + 2x_2 + 3x_3$

$= 3x_1x_2 + 4x_1x_3 - 3x_2x_3 - 4x_1 - x_2 + 2x_3 + 2.$

Now,

$$f^I(x) = f(x) - f(x^I) = -8x_1x_3 + 6x_2x_3 + 5x_1 - x_2 + x_3 + 2\,;$$

For $I = N = \{1, 2, 3\}$, we have

$f(x^N) = f(x')$

$= 3(1-x_1)(1-x_2) - 4(1-x_1)(1-x_3) + 3(1-x_2)(1-x_3) - 2(1-x_2) + 1 - x_1 + 3(1-x_3)$

$= 3x_1x_2 - 3x_1 - 3x_2 - 4x_1x_3 + 4x_1 + 4x_3 + 3x_2x_3 - 3x_2 - 3x_3 - x_1 + 2x_2 - 3x_3 + 4$

$= 3x_1x_2 - 4x_1x_3 + 3x_2x_3 - 4x_2 - 2x_3 + 4$

So, $f^N(x) = p(x) = f(x) - f(x') = x_1 + 2x_2 + 5x_3 - 4.$

For any qpbf f with the corresponding set N, and for any $I \subseteq N$, we have :

Lemma 2.1.1. *$f^I(x^*) \leq 0$ for all $0-1$ global minimal point x^* of f.*

Proof : If x^* is a global minimal point of f, then $f(x^*) - f(y) \leq 0$ for all $0-1$ vectors y. Since $(x^*)^I$ is a $0-1$ vector, it follows that

$f^I(x^*) = f(x^*) - f((x^*)^I) \leq 0.$ □

Let

$$f^I(x) = \sum_{i<j} d_{ij}x_ix_j + \sum_{i=1}^{n} h_ix_i + k \tag{2.1.1}$$

where the coefficients d_{ij}, h_i $(1 \leq i, j \leq n)$ and k are reals. Using (2.0.2), construct a linear (or constant) *minorant* of f^I as follows :

If $d_{ij} > 0$, replace $d_{ij}x_ix_j$ in $f^I(x)$ either by 0 or by $d_{ij}(x_1 + x_2 - 1)$.

If $d_{ij} < 0$, replace the term $d_{ij}x_ix_j$ in $f^I(x)$ either by x_j or by x_i.

The linear part of $f^I(x)$ is left unchanged.

Clearly, the function l so obtained from $f^I(x)$ is linear (or constant) and satisfies $l(x) \leq f^I(x)$ for all $x \in \{0,1\}^n$.

Example 2.2 : Consider f as in Example 1.1, and set $I = \{1, 2\}$. Then

$$f^I(x) = -8x_1x_3 + 6x_2x_3 + 5x_1 - x_2 + x_3 + 2.$$

We construct the following linear minorants of $f^I(x)$:

1) We let vanish x_2x_3 and substitute x_1x_3 by x_3. Then

$$f^I(x) = -8x_3 + 5x_1 - x_2 + x_3 + 2 = 5x_1 - x_2 - 7x_3 + 2.$$

2) We let vanish x_2x_3 and substitute x_1x_3 by x_1. Then

$$f^I(x) = -8x_1 + 5x_1 - x_2 + x_3 + 2 = -3x_1 - x_2 + x_3 + 2.$$

3) We can substitute x_2x_3 by $x_2 + x_3 - 1$ and substitute x_1x_3 by x_1. Then

$f^I(x) = -8x_1 + 6(x_2 + x_3 - 1) + 5x_1 - x_2 + x_3 + 2 = -3x_1 + 5x_2 + 7x_3 - 4.$

4) We can substitute x_2x_3 by $x_2 + x_3 - 1$ and substitute x_1x_3 by x_3. Then

$f^I(x) = -8x_3 + 6(x_2 + x_3 - 1) + 5x_1 - x_2 + x_3 + 2 = 5x_1 + 5x_2 - 7x_3 - 4.$

The following result is evident.

Lemma 2.1.2. *Let f be an $n-$variable qpbf, $I \subseteq N = \{1, 2, .., n\}$ and let x^* be a $0-1$ minimal point of f. If l is a linear minorant of f^I (constructed above), then*

$$l(x^*) \leq 0.$$

Remark 4. *We are not interested in constant minorants, though they may effectively exist as we shall see in the example below.*

We denote by l_I the set of all linear (non constant) minorants of f^I. Clearly $| l_I | \leq 2^q$ where q is the number of quadratic terms in (2.1.1). Denote by L the union of all the l_I, with $I \subseteq N = \{1, 2, ..., n\}$. Then obviously, $| L | \leq 2^{q+n}$.

Before studying some properties of L, we must explain why we may assume L non empty (since we have rejected all the constant minorants). The following result is helpful :

Lemma 2.1.3. *If $p(x) = f(x) - f(x')$ is constant with value c, then*

(1) $c = 0$,

(2) x^* is a $0-1$ minimal point of f if and only if $(x^*)'$ is a minimal point of f.

Proof : Let $p(x) = f(x) - f(x') = c$ be a constant function. Then :

(1) $0 = p(\frac{1}{2}, \frac{1}{2}, ..., \frac{1}{2})$ as we saw in the introduction, and hence $c = 0$.

(2) In particular, $p(x)$ is constant implies that $p(x^*) = f(x^*) - f((x^*)') = 0$. □

So when $p(x)$ is constant, an immediate consequence is that for any fixed $i \in \mathbb{N}$; there exists a $0-1$ minimal point x satisfying $x_i = 0$ and there exists a $0-1$ minimal point y satisfying $y_i = 1$.

Example 2.3 : Let $f(x) = 2x_1x_2 - 4x_1x_3 + 2x_2x_3 + x_1 - 2x_2 + x_3$.

Then $f(x') = 2(1-x_1)(1-x_2) - 4(1-x_1)(1-x_3) + 2(1-x_2)(1-x_3) + (1-x_1) - 2(1-x_2) + (1-x_3) = 2x_1x_2 - 4x_1x_3 + 2x_2x_3 + x_1 - 2x_2 + x_3 = f(x)$. Then $p(x) = f(x) - f(x') = 0$, for all x.

By simple calculations, we have $f(0,0,0) = 0$, $f(1,0,0) = 1$, $f(0,1,0) = -2$, $f(0,0,1) = 1$, $f(1,1,0) = 1$, $f(1,0,1) = -2$, $f(0,1,1) = 1$, $f(1,1,1) = 0$.

It is clear that the two minimal points of f are $(0,1,0)$ and $(1,0,1)=(0,1,0)'$.

Remark 5. *In the case of* $p(x)=f(x)-f(x')$ *being the constant 0, we can fix* $x_i=0$ *for some* $i\in\mathbb{N}$ *and minimize* $f(x_1,..,x_{i-1},0,x_i,...,x_n)$*. This process can be repeated until* $p(x)$ *is non constant or the last qpbf* f *is linear. In any of these cases, finding the set of* $0-1$ *minimal points of the last function leads easily to the set of* $0-1$ *minimal points of the initial function. This allows us to assume that the qpbf* f *(which may be obtained from the initial one by the above process) is such that* $p(x)=f(x)-f(x')$ *is not constant. In that way the set* L *is always not empty since* p *is by construction the single element in* $l_N\subseteq L$

We can now study some properties of L. We start by the following separation result.

Theorem 2.1.1. ***(Separation)*** *Let* y *and* z *be two* $0-1$ *vectors such that* $f(y)>f(z)$*. Then, there exists* $l\in L$ *such that* $l(y)>0$ *and* $l(z)<0$*.*

Proof : Consider $I=\{i\in N: y_i\neq z_i\}$ and observe that $y^I=z$ and $z^I=y$. Consider the resulting

$$f^I(x)=\sum_{1\leq i<j\leq n} d_{ij}x_ix_j+\sum_{i=1}^{n} h_ix_i+k. \tag{2.1.2}$$

Leaving the linear part unchanged, we linearize $f^I(x)$ as follows :

(i) For $d_{ij}>0$: if $y_i=y_j=1$, replace x_ix_j in $f^I(x)$ by x_i+x_j-1, else replace it by 0. Notice that $d_{ij}y_iy_j=(y_i+y_j-1)d_{ij}=d_{ij}$ in the first case, and $d_{ij}y_iy_j=0$ in the second one.

(ii) For $d_{ij}<0$: if $y_i=0<y_j=1$, replace x_ix_j in $f^I(x)$ by x_i. If $y_i=y_j$, replace x_ix_j in $f^I(x)$ either by x_i or by x_j. Notice that $d_{ij}y_iy_j=0=d_{ij}y_i$ if $y_i=0<y_j=1$ and $d_{ij}y_iy_j=d_{ij}y_i=d_{ij}y_j$ if $y_i=y_j$.

We denote the resulting linear function by l^{yz}. Observe that $l^{yz}(y)=f^I(y)$.

On one hand,

$l^{yz}(y) = f^I(y) = f(y) - f(y^I) = f(y) - f(z) > 0$

since $f(y) > f(z)$. On the the other hand, l^{yz} is a linear minorant of f^I, and so $l^{yz}(z) \leq f^I(z) = f(z) - f(z^I) = f(z) - f(y) < 0$ since $f(y) > f(z)$. □

Definition : We say that :

1) The linear function l^{yz} ***separates*** y and z if $l^{yz}(y) > 0 > l^{yz}(z)$.

2) A set K of linear inequalities is ***complete*** for f if for every $x \in \{0,1\}^n$,
$k(x) \leq 0$ for all $k \in K \iff x$ is a minimal point of f.

Corollary 2.1.1. *Let y be a $0-1$ vector such that $f(y) > min(f) = u$, then $l(y) > 0$ for some $l \in L$.*

Proof : It follows from Theorem 2.1.1 for z a minimal point of f. □

Theorem 2.1.2. *(**Completeness**) The set L is complete for f.*

Proof : Let x be a minimal point of f and let $l \in L$. Then l is a minorant of f^I for some $I \subseteq N = \{1,2,..,n\}$. From Lemma 2.1.1 , $l(x) \leq f^I(x) \leq 0$. So $l(x) \leq 0$ for all $l \in L$. By contraposition, if x is not a minimal point of f, then from Corollary 2.1.1 there exists $l \in L$, such that $l(x) > 0$. □

Clearly, Theorem 2.1.2 states that the $0-1$ minimal points of f are exactly the $0-1$ vectors in the convex polytope $\{x \in R^n : l(x) \leq 0 \ \forall \ l \in L\} \cap [0,1]^n$.

2.2. First reduction of the Set L

Consider $f(x)$ as defined in (2.0.1) and $f^I(x)$ as in (2.1.1). For each of them, the number of quadratic terms is bounded by $\frac{1}{2}(n^2 - n)$.

For a given $I \subseteq N = \{1,2,..,n\}$, any $l \in l_I$ is obtained by choosing one of the two possible minorizations for each quadratic term in $f^I(x)$. And so $\mid l_I \mid$ is at most $2^{\frac{1}{2}(n^2-n)}$. Since there are 2^n subsets of N, it follows that $\mid L \mid$ is bounded by $2^{\frac{1}{2}(n^2+n)}$. We shall construct a complete subset of L, of cardinality at most $3 \times 2^{n-2}$. Recall that the linear function $p(x) = f(x) - f(x') \in L$. The following lemma is needed :

Lemma 2.2.1. *A hyperplane in a n-dimensional space, contains at most* 2^{n-1} *vertices of the unit hypercube.*

Proof : By induction on n . For $n \geq 2$, the result is obvious since a straight line encounters at most two vertices of a square.
Suppose that the statement holds for some $k \geq 2$. In a $(k+1)$-dimensional space, consider a hyperplane t defined by $t(x) = \sum_{i=1}^{k+1} t_i x_i + e = 0$, and consider the two hyperplanes of the k-dimensional space defined by

$$\sum_{i=1}^{k} t_i x_i + e = 0 \ , \ \sum_{i=1}^{k} t_i x_i + t_{k+1} + e = 0$$

(obtained by fixing $x_{k+1} = 0$ and $x_{k+1} = 1$, respectively). By the induction hypothesis, each contains at most 2^{k-1} vertices of $\{0,1\}^k$, and so t contains at most $2 \times 2^{k-1} = 2^k$ vertices of $\{0,1\}^{k+1}$. □

Lemma 2.2.2. *Let* p_- *denote the set of 0 - 1 vectors* x *with* $p(x) = f(x) - f(x') < 0$, p_0 *the set of 0 - 1 vectors with* $p(x) = f(x) - f(x') = 0$, *and* p_+ *the set of* $0-1$ *vectors with* $p(x) = f(x) - f(x') > 0$. *Then* $\mid p_- \cup p_0 \mid \leq 3 \times 2^{n-2}$.

Proof : Notice that $x \neq x'$ for any $0-1$ vector x, and

$$p(x) = f(x) - f(x') < 0 \Longleftrightarrow p(x') = f(x') - f(x) > 0. \qquad (2.2.1)$$

So $x \in p_- \iff x' \in p_+$ which implies that $\mid p_+ \mid = \mid p_- \mid$. It is also evident that $x \in p_0 \iff x' \in p_0$. Recall that p_- , p_0 , p_+ form a partition of $\{0,1\}^n$ and hence $\mid p_+ \mid + \mid p_- \mid + \mid p_0 \mid = 2^n$. Then, $2 \mid p_- \mid = 2^n - \mid p_0 \mid$. Now,

$$\mid p_- \cup p_0 \mid = \mid p_- \mid + \mid p_0 \mid = \tfrac{1}{2} \times (2^n - \mid p_0 \mid) + \mid p_0 \mid = (2^{n-1}) + \tfrac{1}{2} \times \mid p_0 \mid.$$

So by Lemma 2.2.1,

$$\mid p_- \cup p_0 \mid = (2^{n-1}) + \tfrac{1}{2} \times \mid p_0 \mid \leq (2^{n-1}) + \tfrac{1}{2} \times 2^{n-1} = 3 \times 2^{n-2}. \ \square$$

Set $m = \mid M \mid$ and consider a fixed $x^* \in M$. Knowing that all $0-1$ minimal points of f are in $p_- \cup p_0$, and using the separation presented in Theorem 2.1.1 and Corollary 2.1.1, we shall consider the following set

$$L_{x^*} = \{p\} \cup \left\{ l^{yx^*} : y \in p_- \cup p_0,\ f(y) > u \right\}. \tag{2.2.2}$$

Theorem 2.2.1. *Let $x^* \in M$, and f a qpbf then,*

(1) (1) The Set L_{x^*} is complete for f.

(2) (2) $\mid L_{x^*} \mid \leq 3 \times 2^n - m + 1$.

Proof :

(1) As $L_{x^*} \subseteq L$, from Lemma 2.1.2 , if x is a $0-1$ global minimal point of f, then $l(x) \leq 0$ for all $l \in L_{x^*} \subseteq L$ and all $x \in M$. Conversely, consider a $0-1$ vector x such that $l(x) \leq 0$ for all $l \in L_{x^*}$. Then clearly, $p \in L_{x^*}$, i.e. $x \in p_- \cup p_0$ and $f(x) > u$. It follows that $l^{xx^*} \in L_{x^*}$ by construction and $l^{xx^*}(x) > 0$ by Corollary 2.1.1, contradicting $l^{xx^*}(x) \leq 0$. Thus $f(x) = u$ and $x \in M$.

(2) By construction, L_{x^*} consists of p and all l^{yx^*}, $y \in (p_- \cup p_0) \setminus M$. Now, the result follows from Lemma 2.2.2. $\square$

Clearly, L_{x^*} depends on a $0-1$ minimal point x^* and $\mid L_{x^*} \mid$ can be exponential.

In what follows, we examine another subset of L and give a bound on the number of optimal points.

2.3. *A* SECOND REDUCTION OF THE SET L AND A BOUND ON THE NUMBER OF $0-1$ MINIMUM POINTS

Recall that we consider a qpbf is considered with the property that $p(x)$ is not constant. The set M stands for the set of its points. Using the separation tool, we build up an other complete subset of L inductively as follows :
Set $L_0 = \{p\}$. Suppose we have constructed L_i for some $i \geq 0$. Set

$$F_i = \{x \in \{0,1\}^n : l(x) \leq 0 \quad \forall\, l \in L_i\}. \tag{2.3.1}$$

If there are $x, y \in F_i$ with $f(y) > f(x)$, set $L_{i+1} = L_i \cup \{l^{yx}\}$, and notice from the separation theorem that $F_{i+1} \subsetneqq F_i$ (since $y \in F_i \setminus F_{i+1}$). When f is finally constant on F_i for some $i \geq 0$, the following is true :

Theorem 2.3.1. *If f is constant on F_{i_0} then,*
(1) $F_{i_0} = M$ (the set of minimal points of f) and L_i is complete.
(2) $|\ L_{i_0}\ | = |\ (p_- \cup p_0) \setminus M\ | \leq 3 \times 2^{n-2} - m + 1$.

Proof :
(1) Recall that

$$F_{i_0} = \{x \in \{0,1\}^n : l(x) \leq 0 \quad \forall\, l \in L_{i_0}\}\ ,\ L_{i_0} \subseteq L$$

and M is the set of minimal points of f. Theorem 2.1.2 implies that $M \subseteq F_{i_0}$. and so ,
(f is constant on F_{i_0}) $\Longleftrightarrow F_{i_0} = M$ and this also means that L_{i_0} is complete.
(2) From $L_0 = \{p\}$, it is clear that only the $0-1$ vectors of $p_- \cup p_0$ are involved in the construction of L_i for any $i \geq 0$. Since $f(y) > u$ for any $y \in (p_- \cup p_0) \setminus M$, there is a unique $x \in p_- \cup p_0$ such that l^{yx} was included in the construction of

L_{i+1}, separating y from the feasible region. Since the process stops when f is constant on $F_i = M$, the result follows. □

Next, we examine a bound on the cardinality of M.

Theorem 2.3.2. *For any non constant qpbf f in n variables the number of $0-1$ points of minimum is at most $3 \times 2^{n-2}$.*

Proof :

If $p(x)$ is not constant then the set M of $0-1$ points of minimum satisfies $M \subseteq p_{-} \cup p_0$, and thus $\mid M \mid \leq 3 \times 2^{n-2}$ by Lemma 2.2.2.

Let $p(x)$ be constant. Then $p(x) = p(x')$ for any $0-1$ point x, in particular for a point of minimum y and

$$y \in \{0,1\}^n, \, y \in \{x : \; x_1 = 0\} \Leftrightarrow y' \in \{x : \; x_1 = 1\}.$$

So the set of minimal points with $x_1 = 0$ and the set of minimal points with $x_1 = 1$ have the same cardinality. Then $\mid M \mid = 2 \mid \{x \in M : \; x_1 = 0\} \mid$. Hence we can fix $x_1 = 0$ in $f(x)$ and obtain the non constant function $f_1(x)$ of at most $n-1$ variables, having the property that $\mid M \mid = 2 \mid M_1 \mid$; where M_1 is the set of minimal points of $f_1(x)$. If $f_1(x)$ is quadratic and the associated $p_1(x)$ is also constant, then we continue the fixations by fixing to zero the variable of least index in $f_1(x)$. At this second step, we obtain for similar reasons, a non constant function $f_2(x)$ of at most $n-2$ variables, having the property that $\mid M_1 \mid = 2 \mid M_2 \mid$, i.e. $\mid M \mid = 2^2 \mid M_2 \mid$, where M_2 is the set of minimal points of $f_2(x)$. If the property also holds for $p_2(x)$ the fixation process can continue. In general, $\mid M \mid = 2^k \mid M_k \mid$ where M_k is the set of minimal points of $f_k(x)$ for any $k \geq 1$. But that process will always stop for some $k \leq n-1$, either because $f_k(x)$ is linear or because $f_k(x)$ is quadratic with $p_k(x)$ not constant. Recall that $f_k(x)$ cannot be constant, otherwise, f would be constant. Now, when the process stops, the $n-k$ variable function $f_k(x)$ is linear, and hence from Lemma 2.2.1, the equation $f_k(x) = m$ (minimal value) has at most 2^{n-k-1} solutions, and hence

$$| M |= 2^k | M_k | \leq 2^k 2^{n-k-1} = 2^{n-1} \leq 3 \times 2^{n-2}.$$

If $f_k(x)$ is quadratic with $p_k(x)$ not constant, then from Lemma 2.2.2, we have $| M_k |\leq 3 \times 2^{n-k-2}$, and hence $| M |\leq 2^k 3 \times 2^{n-k-2} = 3 \times 2^{n-2}$. □

2.4. Concluding remarks on the main linear function

Let $f(x) = \frac{1}{2}x^t Ax + a^t x$. From Chapter 1 (Theorem 1.3.1) we know that : i) the linear function $p(x) = f(x) - f(x') = (Ae + 2a)^t x - f(e) = b^t x + c$ where $e = (1, 1, ..., 1)^t$, $x' = e - x$, $b = Ae + a, c = -f(e)$. So $p(x)$ is easily constructed from f ; ii) the hyperplane P defined by $p(x) = 0$ passes through $(\frac{1}{2}, ..., \frac{1}{2})$ and there exist $0-1$ vectors y and w such that $p(y) > 0$ and $p(w) < 0$. So the point of minimum for $p(x)$ on the hypercube satisfies $p(x) < 0$. A minimal point s of $p(x)$ on the unit hypercube is a $0-1$ vector where the difference between $f(x)$ and $f(x')$ is the largest. Now s satisfies $p(x) < 0$ means that geometrically s is in the same half of the hypercube as all the points of global minimum of f. Such a point s is obtained by setting $s_i = 1$ if $b_i < 0$ and $s_i = 0$ if not. The following experiments in Tab 2.1 show that the value $f(s)$ of is often small (in absolute value, $f(s)$ is averagely 70 percent of the best known value), and this fact makes s a potential good starting point for some minimization heuristics. In Tab 2.1, the values are given in their absolute values (we can talk of maximal values) and the best known values that we refer to are given by Boros et al. in [BHT 07]. We indicate the percentage of $f(s)$ to the best value for each problem. The 50 test problems are from Beasley (a benchmark available at the OR-Library website, 10 problemes for each n, $n = 50$, $n = 100$, $n = 250$, $n = 500$, $n = 1000$).

2.5. ACKNOWLEDGEMENTS

We thank Endré Boros who greatly cooperated in this work. In particular we are indebted towards him for : 1) the main idea of generating other linear functions like $p(x)$ to characterize the $0-1$ points of minimum, 2) remarks on the separation theorem and the essential ideas of the proof, 3) the invitation at RUTCOR for a two week research visit during which some essential ideas were settled. We also thank I.G. Rosenberg of the University of Montreal, for reading and improving the presentation of this work.

n	Problem Instance	Best know f(*x)	f(Sp)	%
50	bae50-1	2,098	1011	48.18
50	bae50-2	3,702	2570	69.42
50	bae50-3	4,626	3442	74.40
50	bae50-4	3,544	2817	79.48
50	bae50-5	4,012	3244	80.85
50	bae50-6	3,693	3284	88.92
50	bae50-7	4,520	3586	79.33
50	bae50-8	4,216	3578	84.86
50	bae50-9	3,780	3018	79.84
50	bae50-10	3,507	3057	87.16
100	bae100-1	7,970	3537	44.37
100	bae100-2	11,036	8750	79.28
100	bae100-3	12,723	9642	75.78
100	bae100-4	10,368	7083	68.31
100	bae100-5	9,083	5684	62.57
100	bae100-6	10,210	6625	64.88
100	bae100-7	10,125	6697	66.14
100	bae100-8	11,435	8377	73.25
100	bae100-9	11,455	8139	71.05
100	bae100-10	12,565	9834	78.26
250	bae250-1	45,607	29879	65.51
250	bae250-2	44,810	31680	70.69
250	bae250-3	49,037	40125	81.82
250	bae250-4	41,274	29205	70.75
250	bae250-5	47,961	33089	68.99
250	bae250-6	41,014	26212	63.90
250	bae250-7	46,757	33794	72.27
250	bae250-8	35,726	19557	54.74
250	bae250-9	48,916	37309	76.27
250	bae250-10	40,442	26001	64.29

n	Problem Instance	Best know f(*x)	f(Sp)	%
500	bae500-1	116,586	77676	66.62
500	bae500-2	128,339	90486	70.50
500	bae500-3	130,812	93129	71.19
500	bae500-4	130,097	93592	71.94
500	bae500-5	125,487	86715	69.10
500	bae500-6	121,772	90097	73.98
500	bae500-7	122,201	81303	66.53
500	bae500-8	123,559	87162	70.54
500	bae500-9	120,798	80279	66.46
500	bae500-10	130,619	88833	68.00
1000	bae1000-1	371,438	268700	72.34
1000	bae1000-2	354,932	238799	67.28
1000	bae1000-3	371,236	254175	68.46
1000	bae1000-4	370,675	260370	70.24
1000	bae1000-5	352,760	256018	72.57
1000	bae1000-6	359,629	247009	68.68
1000	bae1000-7	371,193	272771	73.48
1000	bae1000-8	351,994	244346	69.41
1000	bae1000-9	349,337	247589	70.87
1000	bae1000-10	351,415	257314	73.22

TAB. 2.1. The values of s_p.

Chapitre 3

MATRICIAL PROPERTIES OF CHANGING THE ORIGIN AND A HEURISTIC FOR QUADRATIC $0-1$ MINIMIZATION

AUTEURS : Calvin Wuntcha, Ivo G. Rosenberg

RÉSUMÉ ET CONTRIBUTIONS

Plusieurs branches de la recherche opérationnelle et des disciplines connexes mènent à l'optimisation des fonctions quadratiques multilinéaires sur $\{0,1\}^n$. C'est un problème NP-dur bien connu et plusieurs heuristiques ont été developées pour trouver des solutions approximatives. Par un changement de certaines variables on peut changer un sommet donné de l'hypercube en nouvel origine tout en conservant l'hypercube, et on peut en tirer des informations pour l'optimisation. Nous étudions les effets et les propriétés matricielles de ces changements d'origine, et nous utilisons certaines de ces propriétés dans une heuristique que nous proposons pour la minimisation quadratique en variables $0-1$. Quelques résultats expérimentaux sont présentés.

Mots clés : Polynôme multilinéaire, problème NP-dur

CONTRIBUTIONS : L'idée d'exploiter ce changement de variable (bien connu) comme changement d'origine est de Ivo Rosenberg qui avait observé le resultat (c) du Théorème 3.1.1, et c'est aussi lui qui avait établi la première preuve du Théoreme 4.3.1 (i). Il a aussi reformulé et amélioré la présentation de ce travail.

ABSTRACT

Many branches of operations research and related fields lead to the problem of minimizing multilinear quadratic functions of n variables on the set of the $0-1$ vectors in $\mathbb{R}^n$ or, equivalently, on the whole unit hypercube. This problem is well known to be NP-hard, and many local search methods have been studied to address the problem. By changing some variables, we transform a given $0-1$ point into the new origin of the unit hypercube to gain information on the minimization of the function. We study matricial properties of this change of the origin on a given function and use it in a local search heuristic for quadratic $0-1$ minimization. Some experimental results are presented.

Key words : Multilinear function, NP-hard problem.

INTRODUCTION

Let f be a real quadratic polynomial of n variables and $x = (x_1, ..., x_n)^t \in \{0,1\}^n$ a vector of variables. The problem of finding $c_0 = min\,\{f(x), x \in \{0,1\}^n\}$ and $x^* \in \{0,1\}^n$ with $f(x^*) = c_0$ is NP-hard. Since $c^2 = c$ for all $c \in \{0,1\}$, we can assume that f is square free (multilinear), i.e., f is a linear combination of the monomials x_ix_j, x_i, and the constant 1 with $(1 \leq i,j \leq n)$. Thus f can be written in the form

$$f(x) = \sum_{1\leq i<j\leq n} a_{ij}x_ix_j + \sum_{i=1}^{n} b_ix_i + c. \qquad (3.0.1)$$

or equivalently

$$f(x) = \frac{1}{2}x^t Ax + b^t x + c \tag{3.0.2}$$

where a_{ij} is the $(i,j)-th$ entry of the real symmetric matrix $A = [a_{ij}]$ of order n with null diagonal, b_i is the $i-th$ entry of the vector $b \in \mathbb{R}^n$, and c is a real constant. For optimization, we can discard c and consider $f(x) = \frac{1}{2}x^t Ax + b^t x$. Let $x = (x_1, ..., x_n)^t \in \{0,1\}^n$ and $y = (y_1, ..., y_n)^t \in \{0,1\}^n$. The Hamming distance of x and y is

$$H(x,y) = \sum_{i=1}^{n} |x_i - y_i| = \sum_{i=1}^{n} (x_i - y_i)^2 \tag{3.0.3}$$

i.e., the number of different coordinates between x and y. As usual, $x, y \in \{0,1\}^n$ are neighbors if $H(x,y) = 1$. For $k \geq 1$, the $k - ball$ **centered** at y is $B_k(y) = \{x \in \{0,1\}^n : H(x,y) \leq k\}$. A $0-1$ point y is **a point of local minimum** for f if $f(y) \leq f(x)$ for all neighbors x of y, and a point of **global minimum** for f if $f(y) \leq f(x)$ for all $x \in \{0,1\}^n$. For $c \in \{0,1\}$, set $c' = 1-c$ (the negation of c) and consider the following maps : for $x = (x_1, ..., x_n)^t$, set $x' = (1-x_1, ..., 1-x_n)^t$; for a given integer $m \in \{1, 2, ..., n\}$ set $x^m = (1-x_1, 1-x_2, ..., 1-x_m, x_{m+1}, ..., x_n)^t$; for a given $0-1$ vector s, $\varphi^s(x) = x^s$ is defined by : $x_i^s = 1 - x_i$ if $s_i = 1$ and $x_i^s = x_i$ if $s_i = 0$. Clearly, φ^s preserves the unit hypercube and its vertices i.e., for all $x \in \{0,1\}^n$ we have that $\varphi^s(x) \in \{0,1\}^n$ and for all $x \in [0,1]^n$, also $\varphi^s(x) \in [0,1]^n$. It is also clear that φ^s changes s into the new origin of the hypercube, i.e. $\varphi^s(s) = \mathbf{0} = (0, ..., 0)^t$. Notice that $\varphi^{\mathbf{0}}$ is the identity map and $\mathbf{0}' = e = (1, ..., 1)$, $\varphi^e(x) = x'$ for all $x \in \{0,1\}^n$. In the first section, we study the effects and matricial properties of φ^s on $f(x)$. In the second section, we say that a $0-1$ point y is $\mathbf{k-resistant}$ (shortly y is $\mathbf{k-r}$) if :

1) $f(y) \leq 0$;

2) $f(y) \leq f(z)$ for all $z \in B_k(y)$ and all $z \in B_k(y')$;

3) $f(y) \leq f(z)$ for all $z \in B_k(y^m)$ and all $z \in B_k((y')^m)$, for all $m = 1, 2, ..., n$.

Notice from the condition 2) that for $k \geq 2$, any $k-r$ point y is a point of local minimum for f but the converse is not true in general; however, it is clear that a point of global minimum is $k-r$ for all $k = 1, 2, ...n$. We can also remark that if y is k-R then y is $(k-1)-r$ for all $k = 2, 3, ..., n$. It is reported in [PJ 92] that finding a point of local minimum for a given f is relatively simple. Then, an attempt to get closer to the global minimal value, is to find a local minimal point y and require it to be a $k-r$ point, having a method of switching to a point z with $f(z) < f(y)$ if y fails to be a $k-r$ point. But, knowing that the global minimization of (3.0.1) is NP-hard, exploring the $k-r$ points for a given f becomes difficult as k grows from 1 to n. The first non trivial cases of finding a $k-r$ point for $k=1$ and $k=2$ are our main preoccupation in the second section. We propose an algorithm based on the change of origin to find the $1-r$ and $2-r$ points. In the third section we present some preliminary experimental results of our algorithms finding $1-r$ and $2-r$ points.

3.1. Effects and matricial properties of origin change.

For an integer vector $w = (w_1, ..., w_n)^t$, denote by $\mathbf{w}:\mathbf{2}$ the vector $z = (z_1, ..., z_n)^t$ where $z_i = 1$ if w_i is odd and $z_i = 0$ if w_i is even. Let I be the $n \times n$ identity matrix and let I^s be the $n \times n$ diagonal matrix defined by : $I^s_{ii} = 1$ if $s_i = 0$, and $I^s_{ii} = -1$ if $s_i = 1$ for all $i = 1, 2, ..., n$. The results of the following lemma are evident :

Lemma 3.1.1. *For any $s, z \in \{0,1\}^n$,*

i) $(I^s)^{-1} = I^s = (I^s)^t$.

ii) $I^s I^z = I^z I^s$.

iii) $I^{s'} = -I^s$.

iv) I^s *is an orthogonal matrix.*

As usual, the composition $h \circ g$ of two self maps h and g of $\{0,1\}^n$ is defined by setting $(h \circ g)(x) = h(g(x))$ for all $x \in \{0,1\}^n$.

Lemma 3.1.2. *Let $s, z, s^1, s^2, ..., s^k$ be points of $\{0,1\}^n$. Then*

(a) $\varphi^s(x) = s + I^s x = (x+s) : 2$ *for all* $x \in \{0,1\}^n$.

(b) φ^s *is an involution of* $\{0,1\}^n$.

(c) i) $\varphi^s(z) = \varphi^z(s)$ *and ii)* $\varphi^s \circ \varphi^z = \varphi^z \circ \varphi^s$

(d) $\varphi^{s_1} \circ \varphi^{s_2} \circ ... \circ \varphi^{s_k} = \varphi^{(s_1+s_2+...+s_k):2}$

(e) $H(x,y) = k \Leftrightarrow H(\varphi^s(x), \varphi^s(y)) = k$ *for all* $k = 1,2,...,n$ *and hence* φ^s *preserves the "neighbors" relation.*

Proof : *(a)* For $x, s \in \{0,1\}^n$, recall that $\varphi^s(x) = x^s$ with $x_i^s = 1 - x_i$ if $s_i = 1$ and $x_i^s = x_i$ if $s_i = 0$. On one hand, we remark that $x_i^s = 0 \Leftrightarrow x_i = s_i$ and $x_i^s = 1 \Leftrightarrow x_i \neq s_i$ for all $i = 1,2,...,n$. Let $(s+x) : 2 = z$. Then we also have $z_i = 0 \Leftrightarrow x_i = s_i$ and $z_i = 1 \Leftrightarrow x_i \neq s_i$ for all $i = 1,2,...,n$ and hence $\varphi^s(x) = (x+s) : 2$. On the other hand, we remark from its definition that $I^s x = y$ where $y_i = -x_i$ if $s_i = 1$ and $y_i = x_i$ if $s_i = 0$ and hence $x^s = x + I^s x$.
(b) It is clear from its definition that $\varphi^s(x) \in \{0,1\}^n$ for all $x \in \{0,1\}^n$. Now,
$\varphi^s \circ \varphi^s(x) = \varphi^s(s + I^s x) = s + I^s(s + I^s x) = s + I^s s + I^s I^s x = \varphi^s(s) + x$
then $\varphi^s \circ \varphi^s(x) = (0,...,0)^t + x = x$.
(c) i) We have $\varphi^s(z) = (s+z) : 2 = (z+s) : 2 = \varphi^z(s)$,
ii) $\varphi^s(\varphi^z(x)) = \varphi^s(z + I^z x) = s + I^s(z + I^z x) = s + I^s z + I^s I^z x = \varphi^s(z) + I^s I^z x$.
From i) and by the commutativity of I^s and I^z, we have
$\varphi^s(z) + I^s I^z x = \varphi^z(s) + I^z I^s x = z + I^z s + I^z I^s x = z + I^z(s + I^s x) = \varphi^z(\varphi^s(x))$.
(d) From elementary arithmetic
$[x + (y+z) : 2] : 2 = (x+y+z) : 2 = [(y+x) : 2 + z] : 2$.
And so,

$\varphi^{s_1} o \varphi^{s_2}(x) = \varphi^{s_1}((s_2 + x) : 2) = [s_1 + (s_2 + x) : 2] : 2 = [(s_2 + s_1) : 2 + x] : 2$
$= \varphi^{(s_1+s_2):2}(x)$.

The remainder follows by an easy induction.

(e) Remark that for all $0-1$ vectors x and y,

$H(x, y) = k \Leftrightarrow$ (x and y differ exactly in k coordinates).

For a given $0-1$ vector s, the nonequal coordinates between $s+x$ and $s+y$ and between $(s+y) : 2$ and $(s+x) : 2$ are exactly the same as between x and y. Hence by *(a)* $H(x, y) = k \Leftrightarrow H(\varphi^s(x), \varphi^s(y)) = k$. From the particular case $k = 1$, it is clear that φ^s conserves neighbors. □

As usual, a real matrix or a real vector is *nonnegative* if all its entries are nonnegative. Let $f(x)$ be as given in (3.0.2), and for $s \in \{0, 1\}^n$, set $f^s(x) = f(\varphi^s(x))$. Recall that $e = (1, ..., 1)^t$ and notice that for any $0-1$ vector s, we have that $s + s' = e$. We have the following properties :

Theorem 3.1.1. *For any $0-1$ vector s,*

(a) $f^s(x) = \frac{1}{2}x^t A^s x + (b^s)^t x + c^s$ where $A^s = I^s A I^s$, $b^s = I^s(As + b)$ and $c = f(s)$.

(b) $A^{s'} = A^s$ and hence $f^s(x) - f^{s'}(x) = d^t x + c$ where $d = I^s(Ae + 2b)$ and $c = f(s) - f(s')$.

(c) s is a point of local minimum for f if and only if b^s is non negative.

(d) if $b^s = 0$, then s is a global minimal point of f if and only if A^s is non negative.

(e) i) φ^s conserves the points of global minimum and ii) φ^s conserves the points of local minimum.

Proof : Recall from Lemma (3.1.2) that $\varphi^s(x) = s + I^s x$.

(a) $f^s(x) = f(\varphi^s(x)) = \frac{1}{2}(s + I^s x)^t A(s + I^s x) + b^t(s + I^s x)$

$= \frac{1}{2}s^t As + \frac{1}{2}x^t(I^s)^t AI^s x + \frac{1}{2}s^t AI^s x + \frac{1}{2}x^t I^s As + b^t s + b^t I^s x.$

Recall that A and I^s are symmetric matrices and $x^t(I^s A)s$ is a 1×1 matrix and thus a symmetric matrix. Hence,

$f^s(x) = \frac{1}{2}x^t I^s AI^s x + s^t I^s Ax + b^t I^s x + f(s)$

$= \frac{1}{2}x^t[I^s AI^s]x + I^s(As + b)^t x + f(s).$

(b) From *(a)*,

$f^{s'}(x) = \frac{1}{2}x^t[I^{s'} AI^{s'}]x + I^{s'}(As' + b)^t x + f(s')$

$= \frac{1}{2}x^t[I^s AI^s]x - I^s(As' + b)^t x + f(s')$ since $I^{s'} = -I^s$. Hence, $f^s(x) - f(s')(x) = I^s[(As + b) + (As' + b)]^t x + f(s) - f(s') = I^s[A(s + s') + 2b]^t x + f(s) - f(s') = I^s(Ae + 2b)^t x + f(s) - f(s') = d^t x + c$.

(c) Recall that φ^s is an involution on $\{0,1\}^n$ and so $f^s(\varphi^s(y)) = f(\varphi^s(\varphi^s(x)) = f(y)$ for any $0-1$ vector y. In particular, $\varphi^s(s) = \mathbf{0} = (0, ..., 0)^t$ and from *(a)* $f^s(\mathbf{0}) = f(s)$. Suppose that y is a neighbor of s , i.e., for a unique $1 \leq j \leq n$, we have that $y_j = 1 - s_j$ and $y_i = s_i$ for all $i \neq j$, $i = 1, 2, ..., n$. From the proof of Lemma 3.1.2-(e) the self map φ^s preserves neighbors and if $\varphi^s(y) = z$, then $z_j = 1$ and $z_i = 0$ for all $i \neq j$, thus from *(a)* $f(y) = f^s(\varphi^s(y)) = f^s(z) = b^s_j + f(s)$. Recall that s is a point of local minimum if and only if $f(s) \leq f(y)$ for all neighbors y of s, hence, s is a point of local minimum if and only if $f^s(\mathbf{0}) = f(s) \leq b^s_j + f(s)$, i.e., $0 \leq b^s_j$ for any $j \in \{1, 2, ..., n\}$. The claim is proved.

(d) Suppose that $b^s = \mathbf{0}$. Then from *(a)*, $f^s(x) = \frac{1}{2}x^t A^s x + fs$. On one hand, we clearly have the fact that : A^s is non negative $\Rightarrow f(\varphi^s(x)) = f^s(x) \geq f(s)$ for all $x \in \{0,1\}^n$, i.e., s is point of global minimum for f. On the other hand, suppose that s is a point of global minimum for f, and let z be the $0-1$ vector defined by $z_j = 1$,$z_k = 1$ and $z_i = 0$ for $i = 1, 2, ..., n$; $i \neq j$ and $i \neq k$; then $f^s(z) = a^s_{ij} + f(s) = f(\varphi^s(z)) \geq f(s)$, i.e., $a^s_{ij} \geq 0$ where a^s_{ij} is the $(i, j) - th$ entry of A^s and hence, $a^s_{ij} \geq 0$ for all $1 \leq i, j \leq n$, i.e., A^s is nonnegative.

(e) i) φ^s is a permutation of $\{0,1\}^n$ and implies that f and $f^s = f^s(\varphi^s)$ have the

same set of values on $\{0,1\}^n$. Next, $f(x) = f^s(\varphi^s(x))$, hence x is a point of global minimum for $f \Leftrightarrow \varphi^s(x)$ is a point of global minimum for f^s.

ii) If x is a point of local minimum for f then $f(x) \leq f(y)$ for all neighbors y of x, i.e., $f^s(\varphi^s(x)) \leq f^s(\varphi^s(y))$ and since φ^s preserves neighbors, it follows that $f^s(\varphi^s(x)) \leq f^s(z)$ for all neighbors z of $\varphi^s(x)$. □

Remark 6. *Since I^s is a diagonal matrix, we notice that $A^s = I^s A I^s$ and A have the same diagonal entries, so A has a zero diagonal $\Rightarrow A^s$ has a zero diagonal. It is well known that the quadratic multilinear function $f(x) = \frac{1}{2}x^t Ax + b^t x$ is* ***submodular*** *if all the entries of A are nonpositve.*

Let $c \in \{0,1\}$; it can easily be verified that fixing a variable $x_k = c$ in $f(x)$ results in a new multilinear quadratic function and the new associated symmetric matrix is obtained by deleting the $k-th$ row and column of the symmetric matrix A associated to f. Let A^{s+} be defined by $a^s_{ij} = a^s_{ij}$ if $a^s_{ij} \geq 0$ and $a^s_{ij} = 0$ if $a^s_{ij} < 0$, and A^{s-} defined by $a^s_{ij} = a^s_{ij}$ if $a^s_{ij} \leq 0$ and $a^s_{ij} = 0$ if $a^s_{ij} > 0$. Notice that $f^s(x) = \frac{1}{2}x^t A^{s+}x + \frac{1}{2}x^t A^{s-}x + (b^s)^t x + f(s)$. Remark that f^s is a submodular function if A^{s+} is zero. The following theorem is a consequence of well known results :

Theorem 3.1.2.

(a) *If A^s and b^s are nonnegative, then s is a point of global minimum for f.*

(b) *if A^s is nonpositive, then a global minimal point for f can be found in polynomial time.*

(c) *A $0-1$ global minimal point w of $\frac{1}{2}x^t A^{s-}x + (b^s)^t x$ can be found in polynomial time and if $w^t A^{s+} w = 0$, then $\varphi^s(w)$ is a $0-1$ global minimal point for f.*

(d) *Let $1 \leq k < n$. If deleting k rows and the corresponding columns changes A^{s+} into the a zero matrix, then a $0-1$ global minimal point of f can be found by minimizing 2^k submodular functions.*

Proof : *(a)* If A^s and b^s are nonnegative, then the origin is a $0-1$ point of global minimum for f^s, i.e., s is a $0-1$ point of global minimum for f.

(b) and *(c)* are direct consequences of the well known fact (see [BH 02], [Sc 00], [Ha 74]) that a $0-1$ point of global minimum of a submodular function can be found in polynomial time, and the least value of $x^t A^{s+} x$ is 0.

(d) Deleting k rows and the corresponding columns is equivalent to the fixing of k variables in $f^s(x)$. Since each variable can take two values (0 or 1) it follows that there are 2^k possible ways of such fixing; and each fixing results in a submodular function. □

From *(d)*, it follows that f^s can be minimized in polynomial time if k is small (for example $k \leq 10$ for large dimensions) and if the corresponding variables (rows) can be identified in polynomial time.

Theorem 3.1.3. *For any $0-1$ vector s : i) the matrices A^s and A are similar and hence they have the same determinant, the same characteristic polynomial and the same eigenvalues. ii) A vector v is an eigenvector of A if and only if $I^s v$ is an eigenvector of A^s.*

Proof : i) Recall from Lemma 3.1.1 that $I^s = (I^s)^{-1}$ and I^s is orthogonal. $A^s = I^s A^s I^s = I^s A (I^s)^{-1}$ proving that A and A^s are similar and the remainder is a well known consequence. ii) A is a symmetric matrix , so A can be written in the form $A = PDP^t$ where P^t is an orthogonal matrix whose rows are the eigenvectors of A, and D a diagonal matrix whose diagonal consists of the eigenvalues of A (i.e. the eigenvalues of A^s). Now, $A^s = I^s A I^s = I^s PDP^t I^s$ and $P^t I^s$ is an orthogonal matrix whose row are the eigenvectors of A^s and the result follows. □

3.2. FINDING k-RESISTANT POINTS WITH SOME PROPERTIES OF THE CHANGING OF THE ORIGIN

For optimization without loss of generality, we consider $f(x) = \frac{1}{2}x^tAx + b^tx$ as in (3.0.2). Clearly, $f(\mathbf{0}) = 0$. Recall from the introduction that a $0-1$ point y is $\mathbf{k}-\mathbf{resistant}$ (shortly y is $\mathbf{k}-\mathbf{r}$) if :

1) $f(y) \leq 0$;

2) $f(y) \leq f(z)$ for all $z \in B_k(y)$ and all $z \in B_k(y')$;

3) $f(y) \leq f(z)$ for all $z \in B_k(y^m)$ and all $z \in B_k((y^m)')$, for all $m = 1, 2, ..., n$, where $y^m = (1-y_1, 1-y_2, ..., 1-y_m, y_{m+1}, ..., y_n)^t$. An immediate remark is that for $m = n$, $B_k(y^m) \cup B_k((y')') = B_k(y) \cup B_k(y')$ and the related requirement in 3) is exactly the requirement in 2) and this implies that for 3) it is sufficient to check the condition for $m = 1, 2, ..., n-1$. For $k \geq 1$, we can notice that a $k-r$ point s is a $0-1$ point of local minimum such that : $f(s) \leq 0$ (and preferably $f(s) < 0$),

$$f(s) \leq f(x), \forall x \in B_{k-s} = \bigcup \{[B_k(s^m) \cup B_k((s^m)')], \forall m = 1, 2, ..., n\} \quad (3.2.1)$$

We can expect in many cases that not all the $0-1$ points of local minimum will satisfy this property, while a point of global minimum always does. Like all heuristics, the ideal is to reach a point of global minimum. In a sense, we aim to find a $0-1$ local minimal point s which passes the test of "k-resistance" i.e., its value remains the best compared to the values of all points in all the k-balls centered at s^m and all the k-balls centered at $(s')^m$ for all $m = 1, 2, ..., n$. A clear geometric observation is that B_{k-s} is not condensed in one zone of the hypercube, and being spread around the hypercube may have some advantage in particular if f has a unique minimal point that may be far from s in the Hamming distance. We now describe our algorithm for $k = 1$, and $k = 2$. For a given point s, we use some properties of φ^s seen in the preceding section to minimize the cost of the investigations in B_{k-s}. When a point z with $f(z) < f(s)$ is found in B_{k-s}, s is replaced by z and the process starts at new, untill a $1-r$ (or a $2-r$) point s is found.

An algorithm to find a $1-r$ point :

We have $f(x) = \frac{1}{2}x^tAx + b^tx$. We know from the proof of Theorem 3.1.1 that : $f^s(x) = \frac{1}{2}x^tA^sx + (b^s)^tx + c^s$ for any $0-1$ vector s, with $c^s = f(s)$; and for all $i = 1, 2, ..., n$, $b_i^s + c^s = f(y)$ where y is the neighbor differing from s at the $i-th$ coordinate. Hence, set $e = (1, ..., 1)^t$, and notice that the vector $l^s = b^s + c^se$ represents exactly the set of values of the corresponding neighbors of s. We have $b^s = I^s(As + b)$ and $(b^s)' = -I^s(As' + b)$.

Recall that $e = (1, ..., 1)^t$. Thus, due to φ^s, the computation of $f(s)$, $f(s')$, $l^s = I^s(As+b)+f(s)e$ and $l^{s'} = -I^s(As'+b)+f(s)e$ gives us all the values of the points in $B_1(s) \cup B_1(s')$. This is clearly an advantage over computing the $2n + 2$ values of $f(x) = \frac{1}{2}x^tAx + b^tx$ for all $x \in B_1(s) \cup B_1(s')$. Now, for b^s we can find the pair $(min(b^s), ind(b^s))$ where $min(b^s)$ is the minimal entry of b^s, and $ind(b^s)$ the indice of the corresponding coordinate an the corresponding neighbor. We can then find the minimum $ms = (f(s), min(b^s) + f(s), f(s'), min((b^s)') + f(s')$ of f in $B_1(s) \cup B_1(s')$ and the corresponding $0-1$ point that we denote *best*. To avoid the dilemma of having many points of minimal value for f on $B_1(s) \cup B_1(s')$, we decide to choose the first according to the following sequence in case of equality : $(s, s', s^1, ..., s^n, (s')^1, ..., (s')^n)$ where s^i is the neighbor of s with value $b_i^s + f(s)$, $i = 1, ..., n$. Now proceed with a kind of "discrete descent" as follows. Recall that we know at least one $0-1$ point $\mathbf{0}$ satisfying $f(x) \leq 0$.

ALGORITHM $1-r$

Step-0 :

To directly satisfy the condition 1) For any $k-r$ take a starting $0-1$ point s, having the known smallest value $f(s) \leq 0$: Go to **Step-1**

Step-1 :

Find the pair $(ms, best)$: Go to **Step-2**

Step-2 :

If $best = s$, then s is the "best" point in $B_1(s) \cup B_1(s')$ so s satisfies condition 2) : Go to **Step-3**,

If $best = s'$, then s' is the "best" point in $B_1(s) \cup B_1(s') = B_1(s') \cup B_1((s')')$. So s' satisfies conditions 1) and 2). Replace s by s' and Go to **Step-3**.

If $best \neq s$ and $best \neq s'$, then replace s by $best$ and Go back to **Step-1**.

Step-3 : Register s in *potential*

1- While $m \leq n-2$.

If $f(s) \leq f(potential)$, replace m by $m+1$, and s by s^m and Go back to **Step-1**.

If $f(s) > f(potential)$, replace s by potential and Go back to **Step-1**.

2- If $m = n-1$, set $Solution = Potential$, return **Solution**, **STOP**

End.

Clearly, **ALGORITHM** $1-r$ finds a $1-r$ point. Notice that instead of starting the algorithm at $\mathbf{0}$, we will peferably start it at a $0-1$ point w having the smallest known value.

An algorithm to find a $2-r$ point :

We similarly describe an algorithm to find a $2-r$ point using some properties of the change of origin. A way of finding a $2-r$ point is to first find a $1-r$ point s and compare the value of s to the value of all $y \in B_{k-s}$ such that $H(s,y) = 2$ or $H(s^m, y) = 2$ or $H((s^m)'), y) = 2$, $m = 1, 2, ..., n$ knowing that as a $1-r$ point, s already has the best value for the remainder of B_{k-s}. Now, $H(s,y) = 2$ means that s and y differ exactly in two coordinates. Set $y_k = 1 - s_k$, $y_j = 1 - s_j$ and $y_i = s_i$ for all $i \neq j$ and $i \neq k, i = 1, 2, ..., n$. We have $f^s(x) = \frac{1}{2}x^t A^s x + (b^s)^t x + f(s)$; from Lemma 3.1.2-e $H(\varphi^s(s), \varphi^s(y)) = H(\mathbf{0}, \varphi^s(y)) = 2$ and from Theorem 3.1.1 $f(y) = a^s_{ij} + b^s_i + b^s_j + f(s)$. There are exactly $\frac{1}{2}(n^2 - n)$ points y such that

$H(s,y) = 2$ and the set $M_s = \{a^s_{ij} + b^s_i + b^s_j + f(s), 1 \leq i < j \leq n\}$ is the set of all $f(y)$ where $H(s,y) = 2$. From Théorem 3.1.1-b, we have $A^{s'} = A^s$ and hence $M_{s'} = \{a^s_{ij} + b^{s'}_i + b^{s'}_j + f(s'), 1 \leq j < i \leq n\}$. Define the $n \times n$ matrix M^s by : $m^s_{ij} = a^s_{ij} + b^s_i + b^s_j + f(s)$ for $1 \leq i < j \leq n$, $m^s_{ij} = a^s_{ij} + b^{s'}_i + b^{s'}_j + f(s')$ for $1 \leq j < i \leq n$ and $m^s_{ii} = 0$, $i = 1, 2, .., n$. The off-diagonal entries of M^s are clearly the values $f(x)$ such that $H(s,x) = 2$ or $H(s',x) = 2$.

Again, due to φ^s the computations $A^s = I^s A I^s$, $b^s = I^s(As+b)$, $b^{s'} = -I^s(As'+b)$, $f(s)$, $f(s')$ lead to the fact that the value $f(x)$ of each of the $n^2 - n$ points is obtained by the addition of four known reals which is clearly preferable to making $n^2 - n$ times the matricial computation $f(x) = \frac{1}{2}x^t Ax + b^t x$. Now, if s is a $1 - R$ point then $f(s) \leq 0$ and $min(M^s)$ denotes the smallest entry of M^s, $f(s) \leq min(M^s) \Leftrightarrow f(s) \leq f(x)$ for all $x \in B_2(s) \cup B_2(s')$, and similarly, for all $m = 1, 2, ...n - 1$, $f(s) \leq min(M^{s^m}) \Leftrightarrow f(s) \leq f(x)$ for all $x \in B_2(s^m) \cup B_2((s^m)')$. Finally, a $1 - r$ point s is a $2 - r$ point if and only if $f(s) \leq min(M^{s^m}) \Leftrightarrow f(s) \leq f(x)$ for all $x \in B_2(s^m) \cup B_2((s^m)')$ for $m = 1, 2, ..., n$. Recall that any entry of M^{s^m} corresponds to a well known point $B_2(s^m) \cup B_2((s^m)')$. To avoid the dilemma when many entries of M^{s^m} have the least value, priority will be given according to the indices (i,j) to : $i < j$ first, $i = j$ secondly, and $i > j$ finally ; and to the natural order $1, 2, ..., n$ on i and j. Now, we can describe the algorithm to find a $2 - R$ point as follows :

ALGORITHM $2 - r$

Step-1 : Start a $0 - 1$ point w with $f(w) \leq 0$ and find a $1 - r$ point s. Initialize $m = 1$: Go to **Step-2**.

Step-2 : Find the pair $(min(M^{s^m}), best)$ where *best* is the $0 - 1$ point associated to $min(M^{s^m})$:

if $f(s) \leq min(M^{s^m})$ Go to **Step-3** ;

if $f(s) > min(M^{s^m})$, replace w by s. Go back to **Step-1**.

Step-3 : While $f(s) \leq min(M^{s^m})$:

if $m \leq n-1$, replace m by $m+1$, and go back to **Step-2**.

if $m = n$, **stop**. s is a $2-R$ point.

END.

Clearly, the algorithm finds a $2-r$ point.

Now, let us say a few words on the complexity of the algorithms. Recall that the complexity of reaching a point of local minimum in general is still an open problem (see [PJ 92]). We suppose that all the coefficients of $f(x)$ are integers (or rationals that it principle can be replaced by integers by multiplying f with a possibly very large positive integer c). Suppose the integer L_f to be the best known lower bound to f (the sum of all the non positive coefficients in f is an example). To find a $k-r$ point ($k=1$, or $k=2$), the algorithm starts with a point s and tries to explore all the points of B_{k-s} (see (3.2.1)) : i) it stops and s is a $k-R$ point if a better point w (with $f(w) < f(s)$) is not found in B_{k-s}; ii) it starts all again with B_{k-w} if a "better point" w (with $f(w) < f(s)$) is found in B_{k-s}. Now, finding a better point means that the value of s has to decrease by at least 1 ($f(w) < f(s) \Rightarrow f(w) \leq f(s) - 1$); call it an "amelioration". Clearly, "ameliorations" can not happen more than $f(s) - L_f = C_s$ times before the end of the algorithm, otherwise, we should finally reach a point w with $f(w) < f(s) + L_f \leq L_f$ since $f(s) \leq 0$.

So, the algorithm does not consider more than C_s different sets of the type B_{k-s}. We have $B_{k-s} = \bigcup \{[B_k(s^m) \cup B_k((s^m)')], \forall m = 1, 2, ..., n\}$ For $= 1$, we have by a direct evaluation that $\mid B_1(s^m) \cup B_1((s^m)') \mid = 2\left[\binom{1}{n} + 1\right] = 2(n+1)$, hence $\mid B_{k-s} \mid \leq 2n(n+1) = 2n^2 + 2n$. So the number of points visited to find a $1-r$ point is bounded by $2C_s(n^2+n)$ and the computation of their values benefits the properties of φ^s.

For $k=2$, we have by a direct evaluation $\mid B_2(s^m) \cup B_2(s^{m\prime}) \mid = \left[2\left(\binom{1}{n} + 1\right) + 2\binom{2}{n}\right] = (2n + 2 + n^2 - n) = n^2 + n + 2$, hence $\mid B_{k-s} \mid \leq n^3 + n^2 + 2n$. Then the number

of points visited to find a $2-R$ point is $C_s(n^3+n^2+2n)$ and the computation of their values benefits from the properties of φ^s.
Therefore, if the size of C_s is polynomial in n then the two algorithms are polynomial.

3.3. Experimental results and ending remarks

Our experiments are done on some test problems proposed by Beasley (the benchmark on the website of OR-Library, for $n=50$, $n=100$, $n=250$, $n=500$, 10 problems for each dimension). On this set of 40 problems, we consider the algorithm with two different starting points : The origin denoted (Or), and the point s (called Sp) that minimises the linear function $p(x)=f(x)-f(x')$ already mentionned in chapter two. The results are encouraging (values and time) ; in fact $2-r$ has been able to reach 3 of the best known values starting from Sp and not from the origin, and also 6 best known values have been reached starting from Or and not from Sp proving that their combination is advantageous. $2-r$ has been able to reached 7 other best known values starting from each of the two points. In the below tables, we compare $2-r$ to three other algorithms : TS-B (Tabu Search- Beasley), SA-B (Simulated Annealing- Beasley) are the well known heuristics described in [Be 98]. The "Perturbation method" was proposed by Solayapan et al. in [SMP 08]. The "Best known value" that we report and refer to is the so far recorded best known value of the problem as reported by Boros et al. in [BHT 07]. The percentage column indicates the percentage of our solution to the best known general value ; for example 100 percent means that we have reached the best known value. The values are given in the maximization case (positive). The values are presented in Tab 3.1 and the running time are in Tab 3.2 . Though $2-r$ does not outperform the other algorithms in general, it has been able to perform better than the three on the problem denoted bea500-2 (see Tab 3.1) reaching the current best known value that the 3 other algorithms did not. We have marked in bold the values where $2-r$ reached the best known value. In Tab 3.3 we present the performance of the $1-r$ and the difference with the $2-r$. From the difference between $1-r$ and $2-r$, there is hope that $3-r$ will perform far better than $2-r$ in the values, but the running time will also increase considerably. Running the $2-r$ from many more different starting points may yield interesting results. We are currently working on those possibilities. Our simulations were done by the 7.7.0 version of MATLAB on a CPU Intel(R) Core(TM)2@ 2.66GHz ; RAM 2GB(memory)

Problem Instance	Values obtained in the maximization case by			Values obtained by 2-R(Or-Sp)	Best known values (general)	Percentage (%)
	TS-B	SA-B	P-method			
bae50-1	2098	2098	2098	**2098**	2098	100
bae50-2	3702	3702	3702	**3702 (Or)**	3702	100
bae50-3	4626	4626	4626	**4626**	4626	100
bae50-4	3544	3544	3544	**3544**	3544	100
bae50-5	4012	4012	4012	**4012**	4012	100
bae50-6	3693	3693	3693	**3693**	3693	100
bae50-7	4520	4520	4520	4510	4520	99.77
bae50-8	4216	4216	4216	**4216 (Sp)**	4216	100
bae50-9	3780	3780	3780	**3780 (Or)**	3780	100
bae50-10	3507	3507	3507	**3507**	3507	100
bae100-1	7970	7942	7904	7904 (Sp)	7970	99.17
bae100-2	11036	11036	11036	11026 (Or)	11036	99.90
bae100-3	12723	12723	12723	**12723 (Or)**	12723	100
bae100-4	10368	10368	10368	**10368 (Or)**	10368	100
bae100-5	9083	9083	9083	9042 (Sp)	9083	99.54
bae100-6	10210	10210	10122	10156 (Or)	10210	99.47
bae100-7	10125	10125	10098	10075 (Sp)	10125	99.50
bae100-8	11435	11435	11435	**11435**	11435	100
bae100-9	11435	11435	11455	**11455 (Sp)**	11455	100
bae100-10	12565	12565	12547	12547	12565	99.85
bae250-1	45607	45607	45579	**45607 (Or)**	45607	100
bae250-2	44810	44810	44502	44738 (Sp)	44810	99.83
bae250-3	49037	49037	49019	48949 (Or)	49037	99.82
bae250-4	41274	41274	41236	41106 (Sp)	41274	99.59
bae250-5	47961	47961	47948	47819 (Sp)	47961	99.70
bae250-6	41014	41014	40996	40771 (Sp)	41014	99.40
bae250-7	46757	46757	46757	**46757 (Or)**	46757	100
bae250-8	35726	35726	35666	35146 (Sp)	35726	98.37
bae250-9	48916	48916	48733	48523 (Or)	48916	99.19
bae250-10	40442	40442	40442	40220 (Sp)	40442	99.45
bae500-1	116586	116586	116452	115964 (Sp)	116586	99.46
bae500-2	128223	128204	128255	**128339 (Sp)**	128339	100
bae500-3	130812	130812	130812	130524 (Sp)	130812	99.77
bae500-4	130097	130077	130045	129728 (Or)	130097	99.71
bae500-5	125487	125315	125397	125069 (Sp)	125487	99.66
bae500-6	121719	121719	121118	120273 (Sp)	121772	98.76
bae500-7	122201	122201	122159	121484 (Or)	122201	99.41
bae500-8	123559	1233469	123421	122204 (Or)	123559	98.90
bae500-9	120798	120798	120616	119958 (Or)	120798	99.30
bae500-10	130619	130619	130608	129841 (Or)	130619	99.40

TAB. 3.1. 2-r and other heuristics (values).

Problem Instance	Total Time as Reported			2-R(Or-Sp) time
	TS-B	SA-B	Perturbation method	
bae**50**-1	14	19	0.65	0.091
bae50-2	16	20	0.56	0.112
bae50-3	17	21	1.3	0.12
bae50-4	16	21	1.05	0.115
bae50-5	16	20	1.8	0.088
bae50-6	16	22	1.1	0.107
bae50-7	17	22	1.7	0.079
bae50-8	17	22	1.7	0.095
bae50-9	17	22	1.6	0.104
bae50-10	17	21	1	0.104
bae**100**-1	34	31	1.4	0.397
bae100-2	35	34	1.3	0.405
bae100-3	37	34	3.1	0.39
bae100-4	33	33	1.5	0.212
bae100-5	36	33	2.3	0.48
bae100-6	36	34	1.3	0.394
bae100-7	36	32	1.1	0.413
bae100-8	36	31	3.5	0.423
bae100-9	35	35	1.3	0.39
bae100-10	38	36	2.9	0.421
bae**250**-1	238	226	2.4	7.667
bae250-2	239	226	4.4	7.85
bae250-3	254	240	4.1	7.039
bae250-4	234	218	2.7	7.723
bae250-5	245	232	4.2	47.763
bae250-6	240	221	3.3	8.204
bae250-7	250	232	2.5	7.374
bae250-8	225	212	4.8	8.563
bae250-9	246	229	2.7	7.409
bae250-10	235	218	3.5	7.452
bae**500**-1	956	1006	8.79	133.25
bae500-2	979	1009	8.6	120.92
bae500-3	987	1030	7	104.76
bae500-4	1003	10061	11.96	121.49
bae500-5	964	10030	13.1	1125.6
bae500-6	966	10028	12.6	116.37
bae500-7	952	10014	12.7	100.53
bae500-8	1006	10050	15.4	115.117
bae500-9	954	998	15.6	135.67
bae500-10	971	10012	13.3	111.595

TAB. 3.2. 2-r and other heuristics (time).

n	Problem Instance	1-R-or	2-R-or	Difference	Time 1-R-or	Time 2-R-or	Difference
50	bae50-1	851	2098	1247	0.004	0.0710	0.066
50	bae50-2	3678	3702	24	0.006	0.058	0.052
50	bae50-3	4430	4626	196	0.007	0.059	0.052
50	bae50-4	3544	3544	0	0.009	0.057	0.048
50	bae50-5	3908	4012	104	0.005	0.055	0.05
50	bae50-6	3664	3693	29	0.008	0.061	0.053
50	bae50-7	4510	4510	0	0.008	0.031	0.023
50	bae50-8	4160	4160	0	0.007	0.05	0.043
50	bae50-9	3780	3780	0	0.004	0.042	0.038
50	bae50-10	3505	3507	2	0.06	0.074	0.014
100	bae100-1	7737	7737	0	0.020	0.189	0.169
100	bae100-2	11026	11026	0	0.021	0.192	0.171
100	bae100-3	12665	12723	58	0.019	0.209	0.19
100	bae100-4	10362	10369	7	0.018	0.020	0.002
100	bae100-5	8487	8725	238	0.020	0.236	0.216
100	bae100-6	10132	10156	24	0.021	0.214	0.193
100	bae100-7	9969	9969	0	0.195	0.214	0.019
100	bae100-8	11299	11435	136	0.022	0.195	0.173
100	bae100-9	11256	11256	0	0.017	0.189	0.172
100	bae100-10	12389	12547	158	0.019	0.240	0.221
250	bae250-1	45158	45607	449	0.148	3.763	3.615
250	bae250-2	43147	44285	1138	0.144	4.011	3.867
250	bae250-3	48989	48989	0	0.140	3.510	3.37
250	bae250-4	40439	40927	488	0.174	4.159	3.985
250	bae250-5	47801	47801	0	0.143	3.558	3.415
250	bae250-6	39336	40193	857	0.177	3.930	3.753
250	bae250-7	46757	46757	0	0.144	3.537	3.303
250	bae250-8	34108	34768	660	0.166	3.954	3.788
250	bae250-9	48120	48388	268	0.144	3.889	3.745
250	bae250-10	39225	39225	0	0.1608	3.508	3.3472
500	bae500-1	114136	114286	150	4.438	53.601	49.163
500	bae500-2	127783	128339	556	3.939	69.440	65.501
500	bae500-3	129648	129948	300	3.0731	49.901	46.8279
500	bae500-4	128559	129729	1170	3.836	63.104	59.268
500	bae500-5	124109	124421	312	3.430	55.320	51.89
500	bae500-6	118933	120077	1144	3.250	52.659	49.409
500	bae500-7	121296	121484	188	3.515	50.291	46.776
500	bae500-8	121477	122204	727	2.845	56.511	53.666

TAB. 3.3. 1-r and 2-r (time and values)

Chapitre 4

STUDY OF A GRADIENT METHOD IN QUADRATIC $0-1$ OPTIMIZATION

AUTEURS : Calvin Wuntcha , Ivo G. Rosenberg

RÉSUMÉ ET CONTRIBUTIONS

En recherche opérationelle et dans les domaines connexes, l'importance de l'optimisation pseudo-booléenne est établie, et le problème est connu comme étant NP-dur. Les heuristiques pour trouver des solutions approximatives sont très utilisées faute d'algorithme exacte qui soit polynomial. Nous proposons un algorithme qui utilise la descente du gradient pour partir d'un point intérieur et atteindre un point de minimum local. Il utilise les propriétés de la matrice associée, et les propriétés des fonctions multilinéaires sur l'hypercube et ses sommets. Quelques résultats expérimentaux préliminaires sont donnés.

Mots clés : NP-dur, descente du gradient

CONTRIBUTIONS : L'idée et le principe sont venus de Ivo Rosenberg. Calvin Wuntcha a pu montrer que l'algorithme n'aboutissait pas toujours dans un point de minimum local, et Ivo Rosenberg a encore trouvé comment aller jusqu' à un point de minimum local. Les équations pour déterminer la face atteinte dans une descente sont de Calvin Wuntcha, ainsi que l'algorithme et les simulations préliminaires présentés.

Abstract

In operation researchs and related areas, the importance of quadratic pseudo-boolean optimization is well established, and the problem is known to be NP-hard. Local search heuristics are the frequently used methods to address the problem. In this paper, we study a gradient local search method. Starting from a point of the hypercube, we follow the curve of the stepest descent till we reach a face of the hypercube and we restrict the problem to the face and continue till we reach a $0-1$ point of local minimum. In the process, we take advantage of the properties of quadratic forms. Some experimental results for small size problems are reported.

Key words : NP-hard, steepest descent.

Introduction

An n-ary pseudo-boolean function is a map $f : \{0,1\}^n \rightarrow \mathbb{R}^n$. The problem is to find : i) the minimum value u of f on 2^n element set $\{0,1\}^n$, ii) a point $w \in \{0,1\}^n$ with $f(w) = u$. It is well known that f can be interpolated by an n variable real polynomial (i.e. such that $p(x) = f(x)$ for all $x \in \{0,1\}^n$). In applications f often comes in a polynomial form and thus in the sequel we assume that f is already an n-ary polynomial. As $0^2 = 0$ and $1^2 = 1$, for $m \geq 2$, we can replace x_i^m by x_i and assume that the polynomial f is square free, i.e., a linear combination of products of distinct variables. For example, $f(x_1, x_2, x_3) = 3x_1x_2x_3 - 2x_1x_2 + 3x_2x_3 - 7x_1 + 2x_2 - 4x_3$. The degree of f is the greatest number of variables in a product appearing in $f(x)$. The degree of f in the above example is 3 or also, f is cubic.

The first non trivial case is that of the degree 2. The n-ary polynomials of degree two are called quadratic. Thus a n-ary quadratic polynomial is a linear combination of the products $x_i x_j$, $(1 \leq i < j \leq n)$, of linear terms x_i $(i = 1, 2, ..., n)$, and a constant. We concentrate on quadratic f for the following reasons :

a) it is the first non trivial case ;

b) the minimization of a polynomial f of degree greater than 2 can be reduced to a quadratic one by introducing "penalty terms" which increase the number of variables with the same minimum value as f and ;

c) we can use some elementary results on quadratic forms. A quadratic n-variable square-free polynomial can be written in the form

$$f(x) = \frac{1}{2} x^t A x + b^t x + c \qquad (4.0.1)$$

where $A = [a_{ij}]$ is a non zero real symmetric matrix with zero diagonal ; i.e., such that $a_{ii} = 0$, $i = 1, 2, ..., n$; $x = (x_1, ..., x_n)^t \in \{0, 1\}^n$ and $a = (a_1, ..., a_n)^t \in \mathbb{R}^n$. We assume that $c = 0$ for optimization. The following polynomial expression of f is obtained by expanding (4.0.1) :

$$f(x) = \sum_{1 \leq i < j \leq n} a_{ij} x_i x_j + \sum_{i=1}^{n} a_i x_i \qquad (4.0.2)$$

The general problem of minimizing f on $\{0, 1\}^n$ is a well known NP-hard problem, although some classes of special polynomials are known to be solvable in polynomial time (see [AFLS 01]). Many local search heuristics have been constructed for this problem (see [BH 02], [Be 98], [BHT 07]). Gradient methods with interior points use the fact that the minimization of f in $\{0, 1\}^n$ is equivalent to the minimization of f on $[0, 1]^n$ (see [Ro 72]). The transition from $\{0, 1\}^n$ to $[0, 1]^n$ called relaxation means the replacement of the constraints $x_i \in \{0, 1\}$ by $0 \leq x_i \leq 1$ $(i = 1, 2, ..., n)$. This allows the use of the continuity and the differentiability of the polynomial function. Recall that a $0-1$ local minimun is a $0-1$ point y such that $f(y) \leq f(x)$ for all $0-1$ point x differing with y in only one coordinate.

The heuristic we propose starts from an interior point, and uses the properties of the quadratic function f and the properties of its gradient to first move to a $0-1$ point, and from there to a $0-1$ local minimum point.

4.1. Description of the principle.

Recall from (4.0.2) that the polynomial expression of f has no square or, equivalently the matrix A has zero diagonal. For $0 \leq k \leq n$, a k-dimensional face of $[0,1]^n$ given by $1 \leq i_1 < ... < i_{n-k} \leq n$ and $c_1, ..., c_{n-k} \in \{0,1\}$ is the set $\{(w_1, ..., w_n)^t \in [0,1]^n : w_{i_j} = c_j, \forall j = 1, ..., n-k\}$. Thus a 0-dimensional face of $[0,1]^n$ is the singleton $(c_1, ..., c_n)^t$; a 1-dimensional face of $[0,1]^n$ is the edge $\{(c_1, ...c_{i-1}, x_i, c_{i+1}, ..., c_n) : 0 \leq x_i \leq 1\}$ and a $(n-1)$-dimensional face of $[0,1]^n$ is the set $\{(x_1, ...x_{i-1}, c_1, x_{i+1}, ..., x_n) : 0 \leq x_1, ..., x_{i-1}, x_{i+1}, ..., x_n \leq 1\}$ called a facet of $[0,1]^n$.

Remark 7. *The following fact is shown in [Ro 72]. Let $f(x)$ be a square free nonlinear n-variable polynomial and let $w = (w_1, ..., w_n) \in [0,1]^n$ be a point of global minimum of f. If $0 < w_i < 1$ for some $1 \leq i \leq n$, then $g_i(x) = f(w_i, ..., w_{i-1}, x_i, w_{i+1}, ..., w_n)$ is constant. In particular $g_i(0) = f(w) = g_i(1)$ and consequently, f takes its minimum value on $[0,1]^n$; also on $\{0,1\}^n$. Indeed, as f is square-free, the function g_i is linear; i.e. there exist $\alpha, \beta \in \mathbb{R}$ such that $g_i(x_i) = \alpha x_i + \beta$ for all $x_i \in \mathbb{R}$. Suppose $\alpha > 0$, then $f(w) = g_i(w_i) > g_i(0) = f(w_1, ..., w_{i-1}, 0, w_{i+1}, ..., w_n)$ contrary to a point of global minimum. Similarly $\alpha < 0$ leads to $f(w) > f(w_1, w_{i-1}, 1, w_{i+1}, ..., w_n)$. Thus $\alpha = 0$ as required.*

Now, let denote by *grad* $f(x)$ the gradient of f at $x \in \mathbb{R}^n$. It is well known (see [Sn 05]) that $-grad\ f(x)$ is the vector in the direction of the steepest decrease of f at x. Starting from a point $q \in [0,1]^n$, denote by $\varphi^q(T)$ the curve $\in \mathbb{R}^n$ such that $\varphi^q(0) = q$ and such that for all $T \geq 0$, the tangent vector at $\varphi^q(T)$

is $-grad\ f(\varphi^q(T))$; we can then decrease (following the opposite direction of the gradient) till we reach some face of $[0,1]^n$ since we are sure to encounter no ordinary local minimal point on our way. In order to determine the curve $\varphi^q(T)$ of stepest descent in a simple way, we need some elementary linear algebra facts. For example :

i) The matrix A is a symmetric real matrix, and hence the eigenvalues $\lambda_1, \lambda_2, ..., \lambda_n$ are real.

ii) There exist a $n \times n$ matrix P such that $A = PDP^t$ where D is the diagonal matrix with $d_{ii} = \lambda_i$ $(i = 1, 2, ..., n)$ and $P = [v_1, v_2, ..., v_n]$ is the orthogonal matrix whose columns are the eigenvectors of A. The linear transformation $z = P^t x$ where $z = (z_1, ..., z_n)^t$ (i.e. a rotation around the origin) defines a new orthogonal coordinate system with the same origin.

iii) The transformation is angle and distance preserving.

iv) We then have $x = Pz$ and in the new system,

$$f(x) = \frac{1}{2}x^t Ax + b^t x = \frac{1}{2}x^t PDP^t x + b^t Pz = \frac{1}{2}z^t Dz + d^t z = \sum_{i=1}^{n}(\frac{1}{2}\lambda_i z_i^2 + d_i z_i) = F(z) \tag{4.1.1}$$

where $d^t = b^t P$.

The above fact can be found in numerous textbooks on linear algebra.

We may try to simplify (4.1.1) by a translation of the coordinates system. Considering the matrix A and the augmented matrix $[A : b]$; it is well known that the equation $Ax = -b$ has a solution u if and only if A and $[A : b]$ have the same rank. Let u be any solution of the equation $Ax = -b$. The translation $x = y + u$ of $\mathbb{R}^n$ transforms $f(x) = \frac{1}{2}x^t Ax + b^t x$ into

$$g(y) = \frac{1}{2}y^t Ay^t + f(u). \tag{4.1.2}$$

In fact, we have

$f(x) = \frac{1}{2}(y+u)^tA(y+u) + b^t(y+u) = \frac{1}{2}y^tAy + \frac{1}{2}(y^tAu + u^tAy) + b^ty + f(u)$.

Notice that y^tAu is a real (a 1×1 symmetric matrix) and A is also a symmetric matrix, so $(y^tAu)^t = u^tAy$ and similarly $y^tb = b^ty$. It then follows that

$\frac{1}{2}(y^tAu + u^tAy) + b^ty = y^tAu + y^tb = y^t(Au + b) = \mathbf{0}$. In particular, if A is non singular, the result holds exactly for the unique solution $u = -A^{-1}b$. So, if u is a solution to $Ax = -b$, we have

$$f(x) = g(y) = F(z) = \frac{1}{2}\sum_{i=1}^{n}\lambda_i z_i^2 + f(u). \tag{4.1.3}$$

with $y = x - u$ and $z = P^ty = P^t(x-u)$. We first use the translation and get (4.1.2) before using the rotation; but it is well known and easy to show that the rotation and the translation of $\mathbb{R}^n$ commute. Hence we have the same result by first applying the rotation followed by the translation. We return to the general case in (4.1.1) : for $1 \le i \le n$, the i-th partial derivative is

$$F_i = \lambda_i z_i + d_i. \tag{4.1.4}$$

For a given starting point q in the unit hypercube, let $r = P^tq$ be the corresponding point in the new system, and let $\gamma^r(T) = (\gamma_1^r(T), ..., \gamma_n^r(T))$, $(T \ge 0)$ the curve of steepest descent of F (in the new system) starting from r and following the opposite direction of the gradient. From (4.1.4) the function $\gamma_i^r(T)$ is the solution to the first-order linear equation

$$\frac{d}{dT}z_i(T) = -\lambda_i z_i(T) - d_i. \tag{4.1.5}$$

With the initial condition $z_i(0) = r_i$, we have two cases :

1) $\lambda_i = 0$ (the matrix A is singular) and the solution to (4.1.5) is

$$z_i(T) = -d_iT + r_i. \tag{4.1.6}$$

2) $\lambda_i \neq 0$ and the classical solution to (4.1.5) is

$$z_i(T) = (r_i + \frac{d_i}{\lambda_i})e^{-\lambda_i T} - \frac{d_i}{\lambda_i} \tag{4.1.7}$$

and can be found in calculus books treating first order linear differential equations. In the case where $Ax = -b$ has a solution u, it can be checked directly from (4.1.3) that (4.1.5) simplifies to $\frac{d}{dT}z_i(T) = -\lambda_i z_i$ and the solution simplifies to

$$z_i(T) = r_i e^{-\lambda_i T} \tag{4.1.8}$$

with $z = P^t(x-u)$ and $r = P^t(q-u)$ as mentioned above in that case. Let $H = [0,1]^n$ be the unit hypercube in the new coordinates z (i.e., H is an n-dimensional unit hypercube not necessarily in the basic position) and let $r \in H$ be an interior point, or a point such that the vector $-grad\ F(r)$ points into H (i.e. not outside H). According to (4.1.6) and (4.1.7) for all $i = 1, ..., n$. The i^{th} coordinate of the curve of steepest descent starting at r is constant, monotonic decreasing or monotonic increasing. In the case (4.1.7) it is of the form $c_i e^{\lambda_i T} + s_i$; and in the case (4.1.6) it is of the form $c_i T + s_i$.

Remark 8. *Since f is not constant and has no stationary point in $\mathbb{R}^n$, notice that the situation where $c_i = 0$ for all $i = 1, ..., n$ is impossible. We can then assume that there exist $c_i \neq 0$, and notice that $c_i T + s_i$ goes to ∞ or $-\infty$ when T grows, depending on $c_i > 0$ and $c_i < 0$. There are four cases for $c_i e^{\lambda_i T} + s_i$ according to the signs of c_i and λ_i. If $\lambda_i > 0$ then $c_i e^{\lambda_i T} + s_i$ goes to s_i as T increases and is monotonic decreasing (increasing) according to $c_i > 0$ or $c_i < 0$ (if this was the case for all $i = 1, ..., n$, then $s = (s_1, ..., s_n)$ will be a stationary point, but this situation can not happen since there exist at least one positive and at least one negative eigenvalue). If $\lambda_i < 0$ then $c_i e^{\lambda_i T} + s_i$ goes to ∞ or $-\infty$ depending on $c_i < 0$ and $c_i > 0$. It follows that there is $T_0 > 0$ such that $R = \gamma^r(T_0)$ is on the boundary of H and $-grad\ F(R)$ points inside or lies in the boundary. It could* *happen that $\gamma^r(T)$ reinters H for some $T \geq 0$ and this situation may repeat itself several times. Obviously, we should pick the greatest T for which $\gamma^r(T)$ exits H the last time.*

Let us suppose that $\gamma^r(T)$ enters the boundary from an interior point at T_0 for the last time. Let $\gamma^r(T_0) = R = (R_1, ..., R_n)$ and $-grad\ F(R) = (g_1, ..., g_n)$. We know that $\gamma^r(T)$ leaves $[0,1]^n$ towards the exterior.
Set $I = \{1 \leq i \leq n :\ R_i \in \{0,1\}, R_i = 1 \Rightarrow g_i > 0,\ R_i = 0 \Rightarrow g_i < 0\}$. From the hypothesis, $|I| = l \geq 1$. For notational simplicity suppose that $I = \{n-1+l, ..., n\}$ and define $F_I(z_1, ..., z_{n-l}) = F(z_1, ..., z_{n-l}, R_{n-l+1}, ..., R_n)$. Clearly, the polynomial F_I is at most quadratic. As we do not want to follow $\gamma^r(T)$ outside the hypercube $[0,1]^n$, we replace F by F_I and follow the curve $\gamma_I^{r^1}$ of steepest descent of F_I starting from $r^1 = (R_1, ..., R_{n-l})$. The normal procedure is to follow $\gamma_I^{r^1}$ till it reaches the boundary of $[0,1]^{n-l}$ (this approach was used in our program). However, some modifications can be considered later. Let us recall that the curve $\varphi_q(T)$ of the steepest descent in the initial space (the x coordinates) is related to $\gamma^r(T)$ by $\varphi^q(T) = P\gamma^r(T)$ (from the map $x = Pz = [v_1, ..., v_n]\, z$, recall that $q = Pr$) and each coordinate is defined in detail by

$$\varphi_i^q(T) = \sum_{j=1}^{n} v_{ij}\gamma_j^r(T) \tag{4.1.9}$$

Where $\gamma_i^r(T)$ is a one variable exponential or linear function according to the above analysis on $\gamma_i^r(T)$, $i = 1, ..., n$.

Hence, the curve φ^q (starting from q in the initial unit hypercube H) reaches a boundary of H for a solution T_1 of one or many of the $2n$ following equations :

$$\varphi_i^q(T) = 0\ ;\quad \varphi_i^q(T) = 1,\ i = 1, 2, ..., n. \tag{4.1.10}$$

We have the following simple result :

Lemma 4.1.1. *1) If $i \in N$ is such that for all $x_1, ..., x_{i-1}, x_{i+1}, ..., x_n \in [0,1]$*

$$a_i + \sum_{j \in N} a_{ij} x_j > 0 \tag{4.1.11}$$

then all points $(w_1, ..., w_n) \in \{0,1\}^n$ of local minimal of f have $w_i = 0$.

2) If $i \in N$ is such that for all $x_1, ..., x_{i-1}, x_{i+1}, ..., x_n \in [0,1]$

$$a_i + \sum_{j \in N} a_{ij} x_j < 0 \tag{4.1.12}$$

then all points $(w_1, ..., w_n) \in \{0,1\}^n$ of local minimal of f have $w_i = 1$.

Proof : 1) Let (4.1.11) holds and suppose by contraposition that a point of local minimum $(w_1, ..., w_n) \in \{0,1\}^n$ has $w_i = 0$. Recall that $f(w, ..., w_{i-1}, x_i, w_{i+1}, ..., w_n)$ is linear and

$$f(w, ..., w_{i-1}, 1, w_{i+1}, ..., w_n) - f(w) = a_i + \sum_{j \in N} a_{ij} w_j > 0 \tag{4.1.13}$$

shows that
$f(w, ..., w_{i-1}, x_i, w_{i+1}, ..., w_n)$ strictly decreases between $w = (w, ..., w_{i-1}, 1, w_{i+1}, ..., w_n)$ and $(w, ..., w_{i-1}, 0, w_{i+1}, ..., w_n)$ meaning that w is not a local minimal point.
2) With similar arguments. □

Remark 9. *Let P_i be the n-dimensional hyperplane*

$$\sum_{j \in N} a_{ij} x_j = -a_i. \tag{4.1.14}$$

We can notice that due to $a_{ii} = 0$, P_i is parallel to the axis x_i. The condition (4.1.10) states that the hypercube $[0,1]^n$ lies in the positive half-space defined by P_i and similarly (4.1.12) states that $[0,1]^n$ is situated in the negative half-plane and is disjoint from P_i.

4.2. THE BASIC ALGORITHM AND SOME OBSERVATIONS.

Recall that w is a point of local minimum if $f(w) \leq f(w^i)$ where $w^i_j = w_j$ if $i \neq j$ and $w^i_i = 1 - w_i$. A point of local maximum is a point of local minimun for $-f$. for all $i = 1, 2, ...n$, Let $q \in [0,1]^n$ be arbitrary. The restricted algorithm finds $w = (w_1, ..., w_n)$; $w\,]0,1[^n$ such that $f(w) \leq f(q)$ in polynomial time. We start with the curve of steepest descent $\varphi^q(T)$. There exist $T > 0$ such that $\varphi^q(T) = c = (c_1, ..., c_n)$ has at least one coordinate $c_j \in \{0,1\}$. Set

$$I = \{1 \leq i \leq n : \; c_i \in \{0,1\} \, , \; c_i = 1 \Rightarrow \varphi^q_i(T) > 0, c_i = 0 \Rightarrow \varphi^q_i(T) < 0\} \, .$$

The set I is nonvoid since $j \in I$. Let $I = \{i_1, i_2, ..., i_k\}$ where $1 \leq i_1 < i_2 <, ..., < i_k \leq n$ and let $N \setminus I = j_1, ..., j_{n-k}$ with $1 \leq j_1 < j_2 <, ..., < j_{n-k} \leq n$. Further define $f^1(x_{j_1}, ..., x_{j_{n-k}})$ as $f(y_1, ..., y_n)$ where $y_l = x_l$ if $l \in N \setminus I$ and $y_l = c_l$ if $l \in I$. In other words f^1 is the restriction of f to the face of $[0,1]^n$ defined by $x_i = c_i$ for all $i \in I$. Finally, let $c^1 = (c_{j_1}, ..., c_{j_{n-k}})$ Notice that $f^1(c^1) = f(c) \leq f(q)$. Now, we replace f by f^1 and apply the same procedure to f^1 with q replaced by c^1 ; instead of $-grad\ f$ we now use $-grad\ f^1$. Repeating this construction at most n times, we arrive to a constant f^p i.e $f^p = u \leq f(q)$. Now, we restore step by step the frozen variables till we finally find $w = (w_1, ..., w_n) \in \{0,1\}^n \leq f(q)$ with $f(w) = u \leq f(q)$. Clearly, the algorithm finds w in polynomial time. A vertex $w \in \{0,1\}^n$ reached by the restricted algorithm may not be a point of local minimum.

We can proceed as follow to reach a point of local minimum : Let w^i be the $0-1$ point differing with w only at the i-th coordinate, ie $w^i_i = 1 - w_i$ and $w^i_k = w_i$ for all $1 \leq k \neq n$. For all $i = 1, 2, ..., n$, if $f(w) < f(w^i)$ fix $x_i = w_i$ in $f(x)$ and if $f(w) \geq f(w^i)$ let the variable x_i free. There is at least one free variable and f is not constant since since w is not a point of local minimum. After the fixing, w is clearly is a point of local maximum for the function in the facet determine by the fixations. We chose a positive constant ϵ (example $\epsilon = \frac{1}{n^2}$) ; and set $q^*_j = w_j - \epsilon$

if $w_j = 1$, and $q_j^* = \epsilon$ if $w_j = 0$ for all j such that $f(w) < f(w^j)$ and we start a new descent from q^* in the corresponding facet with the restricted algorithm. The whole procedure is repeated and it stops when a point of local minimum is found. We can illustrate the algorithm as follow :

Gradient-descent algorithm

Start :

Go from q and follow the descent of the gradient till you reach a point q^1 in a face of the hypercube.

Verification and action :

For all $i = 1, 2, ...n$, if $q_i^1 = c \in \{0, 1\}$, then fix $x_i = c$ in the function and cancel the ith coordinate in q^1. If the new function becomes linear or constant, then go to a minimal $0 - 1$ vector of it and stops. If the new function is still quadratic, calculate the new gradient and use it to follow a new descent from q^1 to a second point q^2 in a face of the new cube.

Repeat :

Repeat the "**Verification and action**" till a $0 - 1$ vector w is found. Give all the coordinates of w

Check local minimum

If w is a point of local minimum **Stop**,

Else fix the appropriate variables of w, compute q^* and Go back to **Start**

End.

The algorithm ends in a point of local minimum. The complexity of finding a point of local minimum is still unknown, but as we saw about the complexity of $2 - r$ in the last chapter, the algorithm is polynomial if the coefficients of the given quadratic function are integers with a lower bound value polynomial in n (see the comment after presentation of a $2 - r$).

4.3. Experimental results and ending remarks

In the following table (Tab 4.1) we have tested the above algorithm on 120 small size problems ranging from $n = 10$ to $n = 25$. The problems were generated randomly using the 7.7.0 version of Matlab on a CPU Intel(R) Core(TM)2@ 2.66GHz; RAM 2GB(memory). Each function has integer coefficients uniformly generated in $[-2n, 2n]$. In average, the algorithm takes less than a second for $n \leq 25$ and its solution is the optimal value in more than 80 percent of the problems. These results are encouraging though there we still have numerical difficulties to handle when n becomes large. In fact, the equations to solve in order to find the face reached in the descent and the new starting point in that face are influenced by numerical limits. In fact we are working on those details to improve the algorithm.

n	Number of tests	Reached optimum	Percentage	Average time
10	30	28	93.33%	0.1458834
15	30	25	83.33%	0.2974911
20	30	26	86.66%	0.5375665
25	30	24	80.00%	0.930406

Tab. 4.1. Gradient descent.

Chapitre 5

A NEW POLYNOMIAL CASE OF QUADARATIC $0-1$ OPTIMIZATION RELATED TO A MINIMAL EIGENVALUE

AUTEUR : Calvin Wuntcha

Résumé

Même s'il est vrai que le problème d'optimisation quadratique sans contrainte est NP-dur, certaines classes sont connues comme étant solubles en temps polynomial. Dans ce travail nous présentons une classe de problèmes d'optimisation à variables binéaires solubles en temps polynomial. La condition repose sur la connaissance de deux sommets du hypecube dont les valeurs ont une certaine relation avec la valeur propre minimale d'une matrice reliée au problème par changement d'origine.

Mot-clés : NP-dur, temps polynomial

Abstract

Although the general unconstrained quadratic $0-1$ optimization problem is known to be NP-hard, some classes of the problem are known to be solvable in

polynomial time. We propose a new polynomially solvable class. The requirement for our case is related to the minimal eigenvalue of some associated matrix and the knowlege of some points of appropriate values in the unit hypercube.

Keywords : NP-hard, Polynomial time.

Introduction

An n-variable pseudo-boolean function is a map from $\{0,1\}^n$ into $\mathbb{R}$. It is known that it can be interpolated by an n-variable polynomial, and often it is already given in such form. Given that $c^2 = c$ for all $c \in \{0,1\}$, it can be assumed that the polynomial is square-free, i.e. it is a linear combination of products of distinct variables. The minimization problem is to find the minimum c_0 of f on $\{0,1\}^n$ and $w \in \{0,1\}^n$ such that $f(w) = c_0$. The first nontrivial case is when f is quadratic, i.e. f is of the form

$$f(x) = \sum_{1 \le i < j \le n} a_{ij} x_i x_j + \sum_{i=1}^{n} b_i x_i = \frac{1}{2} x^t A x + b^t x \tag{5.0.1}$$

where $A = [a_{ij}]$ is a real $n \times n$ symmetric matrix with zero diagonal, $b = (b_1, ..., b_n)^t$ a real column vector and $x = (x_1, x_2, ..., x_n)^t$ is the vector of variables where t denotes the transposition operator. The general minimization of f over $\{0,1\}^n$ is NP-hard (see [BH 02]). However, for some classes of functions, the problem is known to be solvable in polynomial time (see [BH 02], [AFLS 01] ,[CKM 06], [Al 01], [Sc 03]). The following classes are well known :

(a) The class of functions where $f(x)$ is submodular (i.e. all the entries of A are nonpositive). P. Hansen in [Ha 74] proved that the minimization of a function of that form can be brought to a positive-negative form (i.e. all entries of A are nonpositive and all entries of b are nonnegative) and it can be solved by the algorithm of Ford and Fulkerson (see [FF 62]). Although the classical and basic polynomial time algorithm for this class is the Ford and Fulkerson algorithm and its variants, there exist even strongly polynomial time algorithms recently developed, for example the one due to Schrijver(see [Sc 00]). But if a constraint is added (even a single cardinality constraint like $x_1 + ... + x_n = k$, $x_i \in \{0,1\}, i = 1, ..., n$; where k is any non negative integer) then minimizing a submodular function remains NP-hard (see [Qu 02]).

(b) If the underlying graph associated to the max-cut problem is series parallel, the problem is solvable in polynomial time (see [Ba 86]).

(c) Consider the matrix $Q = [q_{ij}]$ built from the above A and b by setting $q_{ij} = \frac{1}{2}a_{ij}$ if $i \neq j$ and $q_{ii} = b_i$. If Q is negative semidefinite with a fixed (small) rank, then the problem is solvable in polynomial time ([AFLS 01]). This case (c) is the subject of [AFLS 01].

In [CKM 06] two other cases where the initial matrix A has a fixed rank are considered.

We propose a new polynomial case related to the least eigenvalue of Q, possibly after the origin of the hypercube is changed by complementing some variables. We need to know $m, u \in \{0,1\}^n$ such that $f(m) \leq 0 \leq f(u)$. As in [AFLS 01], we do not discuss the complexity of finding the eigenvectors and eigenvalues of a given matrix, but we rather consider that they are given or easily known. In Section 1, we present this new polynomial case and functions belonging to that case. In Section 2, we briefly indicate some remarks that in some cases may help in trying to find the appropriate u and m.

5.1. THE NEW CASE RELATED TO THE MINIMAL EIGENVALUE AND SOME APPROPRIATE POINTS OF THE HYPERCUBE.

Recall that the Hamming distance between any $x, y \in \{0,1\}^n$ is $H(x,y) = \|x - y\|^2$; i.e. the number of coordinates where x and y differ. If $H(x,y) = 1$ then x and y are called **neighbors**, and it is obvious that each $0-1$ point has exactly n neighbors. For a given function f on $\{0,1\}^n$, a point w is a **point of local**

minimum (a **point of local maximum**) for f if $f(w) \leq f(y)$ ($f(w) \geq f(y)$) for all neighbors of w. A $0-1$ point w is a point of global minimum for f if $f(w) \leq f(y)$ for all $y \in \{0,1\}^n$. Several local search algorithms for $0-1$ quadratic optimization have been developed for $n \leq 5000$ (see[BHT 07], [BH 02], [PR 90], [HJM 00]). The methods presented [BHT 07] can be used to find in polynomial time a point of large negative value (or large positive value) when they exist, even though optimality is not guaranteed. In what follows, we consider that two points $m, u \in \{0,1\}^n$ such that $f(m) \leq 0 \leq f(u)$ were found in polynomial time. We do not require local optimality for the two points. The possibility to find such two points is easily explained by the following facts :

Consider $f(x)$ described in (5.0.1).

1) In the first and obvious case, we may take u to be the origin $(0, ..., 0)^t$.

2) Consider a $0-1$ point w of non negative value, for example $w = (0, ..., 0)$. If w is not a point of local minimum for f, then a neighbor y of w satisfies with $f(y) < f(w)$; so from y and $(0, ..., 0)$ we can get u and m such that $f(m) < 0 \leq f(u)$. A similar fact holds if w is not a point of local maximum for f.

3) If the origin $(0, ..., 0)$ is neither a point of local minimum nor a point of local maximum, then m and u with $f(m) < 0 = f(0, ..., 0) < f(u)$ are obviously found. If $(0, ..., 0)$ is both a point of local minimum and a point of local maximum, then one can easily check that $b_i = 0$ for all $i = 1, 2, ..., n$, i.e. $f(x) = \frac{1}{2}x^t Ax$ and since f is not constant, there exist $1 \leq i \neq j \leq n$ such that $a_{ij} \neq 0$ and it is evident that $a_{ij} = f(y)$ when $y_i = y_j = 1$, and $y_k = 0$ for all $k \in \{1, ..., n\} \setminus \{i, j\}$.

4) If $w \in [0,1]^n$ is not a $0-1$ point, there exist simple polynomial time algorithms (like IMPROVE, see [BHT 07]) using w to find a $0-1$ point s such that $f(w) \neq f(s)$.

Recall that any $n \times n$ real symmetric matrix Q has real eigenvalues $\lambda_1, ..., \lambda_n$ and further, there exist two matrices P and D such that $Q = PDP^t$, where

$P = [v^1, v^2, .., v^n]$ is an orthogonal matrix whose columns are normalized eigenvectors of Q, and $D = [d_{ij}]$ is the diagonal matrix whose diagonal is $\lambda_1, ..., \lambda_n$. We denote by $\lambda_{min(Q)}$ the minimal eigenvalue of a matrix Q. The following linear minorization is well known :

Lemma 5.1.1. *For any* $0-1$ *vector* x,

$$\lambda_{min(Q)} \sum_{i=1}^{n} x_i \leq x^t Q x.$$

Proof : Let $x \in \mathbb{R}^n$ and set $z = P^t x$. Then

$$x^t Q x = x^t P D P^t x = z^t D z = \sum_{i=1}^{n} \lambda_i z_i^2 \geq \lambda_{min(Q)} \sum_{i=1}^{n} z_i^2 = \lambda_{min(Q)} \|z\|^2 \quad (5.1.1)$$

where $\|.\|$ denotes the euclidean norm. From a well known property of orthogonal matrices, we have

$$\|z\|^2 = \|P^t x\|^2 = \|x\|^2 = \sum_{i=1}^{n} x_i^2; \quad (5.1.2)$$

hence,

$$\lambda_{min(Q)} \sum_{i=1}^{n} x_i^2 \leq x^t Q x \quad (5.1.3)$$

and the result follows from the fact that $x_i^2 = x_i$, $i = 1, 2, ..., n$ for any $0-1$ vector x. □

For a given $0-1$ vector $s = (s_1, ..., s_n)$, a selfmap l^s of $\{0,1\}^n$ is obtained by complementing the variables x_i for $s_i = 1$; i.e. : for $x = (x_1, ..., x_n)$ set $l^s(x) = y = (y_1, y_2, ..., y_n)$ where for $i = 1, 2, ...n$, the coordinate $y_i = 1 - x_i$ if $s_i = 1$ and $y_i = x_i$ if $s_i = 0$. Define $f^s(x)$ by setting $f^s(x) = f(l^s(x))$ for all $x \in \{0,1\}^n$. It can be verified that

$$f^s(x) = \frac{1}{2} x^t A^s x + (b^s)^t x + f(s) \quad (5.1.4)$$

where A^s and b^s are respectively a new symmetric matrix with zero diagonal and a new constant vector. Notice that l^s is an involution (i.e. l^s is its own inverse) and $f(x) = f^s(l^s(x))$; so f^s and f have the same set of values on $\{0,1\}^n$. From Chapter 4 (Theorem 3.1.1), the selfmap l^s preserves the neighborhood relation and preserves the local and global minima and maxima. From the construction, it is clear that : $l^s(s) = (0,0,..,0) = \mathbf{0}$ and the number of non zero coordinates of $l^s(x)$ is $\|l^s(x)\|^2 = H(s,x)$. Moreover, for $s = \mathbf{0}$ the map $l^{\mathbf{0}}$ is the identity map. We use f^u where u is our known $0-1$ point with nonnegative value. Recall that $f^u(x) = \frac{1}{2}x^tA^ux + (b^u)^tx + f(u)$ and set $Q = (q_{ij})$ with $q_{ij} = \frac{1}{2}a^u_{ij}$ for $i \neq j$ and $q_{ii} = b^u_i$, $1 \leq i,j, \leq n$. Set $F(x) = x^tQx + f(u)$. We know that F, f^u and f agree on $\{0,1\}^n$. We can see that $f(m) = f^u(l^u(m)) = F(l^u(m))$, and therefore every $0-1$ point y of global minimum of F (and hence of f) satisfies $F(y) - f(m) \leq 0$ and by Lemma 5.1.1

$$\lambda_{min(Q)} \sum_{i=1}^{n} y_i + f(u) - f(m) \leq y^tQy + f(u) - f(m) = F(y) - f(m) \leq 0. \quad (5.1.5)$$

This leads to :

Theorem 5.1.1. *Let y be a point of global minimum of f on $\{0,1\}^n$; $u,m \in \{0,1\}^n$ satisfy $f(m) \leq 0 \leq f(u)$ and $f(m) < f(u)$. Then $\lambda = \lambda_{min(Q)} < 0$ and*

$$\|l^u(y)\|^2 \geq \lambda^{-1}(f(m) - f(u)) \quad (5.1.6)$$

Proof : By contraposition, suppose that $\lambda \geq 0$, i.e. Q is positive semidefinite. Then the new origin u is a point of global minimum of F on $\mathbb{R}^n$ and hence a point of global minimum for f on $\{0,1\}^n$ leading to $f(u) \leq f(m)$. Thus $\lambda < 0$. Let y be a $0-1$ point of global minimum of f. As the images of f and f^u agree, clearly $w = l^u(y)$ is a point of global minimum of f^u. From Lemma 5.1.1 and $\|w\|^2 = w_1 + w_2 + ... + w_n$, we get $\lambda\|w\|^2 + f(u) - f(m) \leq 0$ and using $\lambda < 0$, we get $\|w\|^2 \geq \lambda^{-1}(f(m) - f(u))$. □

Remark 10. *Denote the right-hand side of (5.1.6) by c_{um} and set $k_{um} = \lceil c_{um} \rceil$, the least integer greater or equal to c_{um}. Now the number of $x \in \{0,1\}^n$ with $H(x,u) \geq k_{um}$ is*

$$\binom{n}{k} + \binom{n}{k+1} + \ldots \binom{n}{n} = \binom{n}{0} + \binom{n}{1} + \ldots \binom{n}{n-k}. \tag{5.1.7}$$

For k_{um} big enough (i.e. when $\lambda^{-1}(f(m) - f(u))$ is large) the number in (5.1.7) is polynomial. This leads to the fact that checking if one can get two such $0-1$ points m and u is of interest.

The following example shows some special cases.

Example 5.1 :

Let Q be a symmetric matrix such that among the eigenvectors corresponding to $\lambda = \lambda_{min(Q)} < 0$ there is a $0-1$ vector $m = (m_1, ..., m_n)$ with $m_1 + ... + m_n = k$. Then every point of global minimum of $F(x) = x^t Q x$ satisfies $\|x^*\|^2 \geq k$.

Proof : By definition m is an eigenvector of A and so $k > 0$. Set $Q = PDP^t$ where D is a diagonal matrix with the diagonal $(\lambda_1, \lambda_2, ..., \lambda_n)$ and $P = [v^1, v^2, ..., v^n]$ is an orthogonal matrix whose columns are the normalized eigenvectors of Q. Clearly we have $z_1 = v_1.m = \sqrt{k} v_1.v_1 = \sqrt{k}$ due to the orthogonality of P. By the same token for $i = 2, ..., n$ we get

$v^i.m = \sqrt{k} v^i.v_1 = 0$. By the proof of Lemma 5.1.1 we get

$$f(m) = m^t Q m = \sum_{i=1}^{n} z_i^2 = \lambda_{min(Q)} (\sqrt{k})^2 = \lambda_{min(Q)} k < 0. \tag{5.1.8}$$

Set $u = (0, ..., 0)^t$ and recall that l^u is the identity map. Theorem 5.1.1 yields that every point x^* of global minimum of f satisfies $\|x^*\|^2 \geq \dfrac{f(m)}{\lambda_{min(Q)}} = k$.

5.2. Ending remarks

Notice that from their construction (see the introduction) that many polynomial cases of $0-1$ quadratic minimization are related to special cases, for example the polynomial case treated in [AFLS 01] requires the matrix Q to be negative semidefinite and have a fixed rank d. An algorithm of complexity $O(d-1)$ is used to explore the points in a d dimensional image (zonotope) of the unit hypercube. Our approach requires $\lambda_{min(Q)} < 0$ and allows the matrix Q to have any rank. The problem of our approach is that to find the two appropriate $0-1$ points u, m may be neither easy nor possible. If Q is negative semidefinite, i.e. all λ_i are nonpositive, then we can choose $u = \mathbf{0}$ and try to find m with $f(m)$ as small as possible, and we may start the search of m from a $0-1$ vector w that minimizes $z_1^2 = (v_1.x)^2$ (where v_i is the eigenvector corresponding to $\lambda_{min(Q)}$ in P). In the other cases, different methods may be used to try to find the two points. We can also consider the following two cases of the initial function $f(x) = \frac{1}{2}x^tAx + b^tx$ where A is a symmetric matrix with a zero diagonal

Suppose that b is nonnegative. Then $b^tx \geq 0$. If we have $y \in [0,1]^n$ such that $\frac{1}{2}y^tAy > 0$, we may set $m = \mathbf{0}$ and find in polynomial time a $0-1$ vector u such that $f(u) = b^tu + \frac{1}{2}u^tAu^t \geq \frac{1}{2}u^tAu^t \geq \frac{1}{2}y^tAy$ where $\frac{f(m)-f(u)}{\lambda_{min(Q)}} = \frac{-f(u)}{\lambda_{min(Q)}}$ may be large enough. Similarly, one may try to find m if b is non positive.

Chapitre 6

CONCLUSION

Dans cette thèse, nous avons travaillé sur l'optimisation quadratique en variables binaires, tout en faisant certaines observations qui vont au-delà. Résumons notre apport et indiquons quelques pistes et préoccupations pour nos recherches futures :

Notre chapitre 1 est assez général dans le sens qu'il va parfois au-delà de l'optimisation en variables binaires. En même temps il a servi de veritable base aux autres chapitres. En particulier, nous avons étendu un résultat de 1972 concernant l'optimisation des polynômes multilinéaires sur les hyperrectangles de $\mathbb{R}^n$ à une plus grande classe de fonctions. Nous avons utilisé une idée de Pardálos et Somesh pour associer à chaque fonction f sur un sous ensemble fini de points entiers de $\mathbb{R}^n$, une autre fonction u telle que $f + u$ possède un unique point de minimum global qui soit aussi un point de minimum global de f. Nous avons trouvé une inégalité linéaire qui doit être vérifiée par tout point de minimum d'une fonction quadratique sur un ensemble ayant un centre de symétrie.

Au chapitre 2, nous avons montré que pour tout polynôme multilinéaire à optimiser sans contrainte sur $\{0,1\}^n$ est soumis à une contrainte linéaire naturelle $p(x) \leq 0$, et nous avons montré que plusieurs autres contraintes linéaires existent et forment avec $[0,1]^n$ un polytope convexe C tel que : un vecteur $0-1$ est dans C si et seulement si ce vecteur est un minimum global pour f sur $\{0,1\}^n$. Nous avons attiré l'attention sur le point s qui minimise $p(x)$ et dont la valeur $f(s)$ est souvent petite.

Au chapitre 3, nous avons considéré comme une fonction de changement d'origine une opération bien connue de complémentation de certaines variables ; et nous avons étudié les propriétés de matricielles de cette fonction. Nous avons utilisé certaines de ces propriétés pour construire une heuristique qui a donné des résultats considérables en un temps court et qui, après amélioration (par exemple le $2-r$ avec plusieurs points de départs, le $3-r$ ou le $4-r$), va certainement donner de meilleurs résultats.

Au chapitre 4 nous avons commencé l'étude d'une approche continue pour l'optimisation quadratique en variable $0-1$, l'approche de la descente la plus rapide. Nous avons construit un algorithme préliminaire utilisant des valeurs propres et le gradient de la fonction f, et utilisant aussi le fait que chaque frontière de l'hypercube unité est encore un hypercube sur lequel la restriction de f a essentiellement les mêmes propriétés. Cet algorithme de base qui a plusieurs améliorations possibles donne déjà des résultats assez encourageants sur les problèmes de petites dimensions. Mais, des réalités numériques restent à considérer pour l'appliquer à des problèmes de grandes dimensions : la manière de rechercher la frontière de l'hypercube atteinte par une descente partant d'un point intérieur, le temps positif auquel on l'atteint, et le point en lequel il est atteint. Théoriquement nous les avons établis, mais il nous faut optimiser ces résultats et contourner certaines limites numériques, et des pistes pour le faire sont en étude.

Au chapitre 5, nous avons proposé une nouvelle classe polynomiale de l'optimisation quadratique en variable $0-1$, et nous en avons indiqué des exemples.

Au titre des travaux et pistes futurs ; nous comptons examiner et exploiter encore plus les inégalités linéaires qui caractérisent l'ensemble des points de minimum de f sur les vecteurs $0-1$; par exemple, les effets algorithmiques de $p(x)$ pourraient être plus nombreux et nous comptons les examiner plus profondément.

Nous comptons examiner les cas $3 - r$ et $4 - r$ de l'heuristique présentée dans le chapitre 3 et qui, logiquement, devraient améliorer les valeurs atteintes par le $2 - r$. Toutefois, trouver plusieurs autres points de départ intéressants pour le $2 - r$ pourrait être plus efficace et moins couteux en temps de calcul qu'un $3 - r$, et nous comptons examiner cela. La descente la plus rapide est encore en cours d'optimisation et nous avons plusieurs pistes en étude pour rendre cet algorithme plus efficace.

BIBLIOGRAPHIE

[AFLS 01] ALLEMAND K., FUKUDA K., LIEBLING T.M., STEINER E., *A polynomial Case of Unconstrained Zero-one Quadratic Optimization*, Mathematical Programming, Ser. A 91, pp. 49-52, 2001.

[Al 01] ALLEMAND K., *Optimisation Quadratique en Variables Binaires : Heuristiques et Cas Polynomiaux*, Thèse, École Polytechnique Fédérale de Lausanne, no 2496, 2001.

[AS 86] ADAMS W. P., SHERALI H. D., *A Tigth Linearization and an Algorithm for Zero-one Quadratic Programming Problems*, Management Science, Vol. 32 (10), pp. 1274-1290, 1986.

[Ba 86] BARAHONA F., *A Solvable Case of Quadratic* $0-1$ *Programming* , Discrete Applied Mathematics, Vol. 13, pp. 23-26, 1986.

[Be 98] BEASLEY J. E., *Heuristic Algorithms for the Unconstrained Binary Quadratic Programming Problem*, http :mscmga.ms.ic.ac.uk/jeb/jeb.html, 1998.

[BEP 05] BILLIONNET A., ELLOUMI S., PLATEAU M.C., *Convex Quadratic Programming for Exact Solution of* $0-1$ *Quadratic Programs*, Technical Report 000, CEDRIC, 2005.

[BH 02] BOROS A., HAMMER P. L., *Pseudo-Boolean Optimization*, Discrete Applied Mathematics, Vol. 123, pp. 155-225, 2002.

[BHT 07] BOROS A., HAMMER P. L., TAVAREZ G., *Local Search Heuristics for Quadratic Unconstrained Binary Optimization (QUBO)*, Journal of Heuristics, Vol. 13, pp. 99-132, 2007.

[Ca 84] CARTER M. W., *The Indefinite Zero-one Quadratic Problem*, Discrete Applied Mathematics, Vol. 7, pp. 23-44, 1984.

[CKM 06] ÇELA E., KLINZ B., MEYER C., *Polynomially Solvable Cases of the Constant Rank Unconstrained Quadratic* $0-1$ *Programming Problem*, Journal of Combinatorial Optimization, Vol. 12, pp. 187-215, 2006.

[FFL 05] FERREZ J.-A., FUKUDA K., LIEBLING TH.M., *Solving the fixed rank convex quadratic maximization in binary variables by a parallel zonotope construction algorithm*, European Journal of Operation Research, Vol. 166, pp. 35-50, 2005.

[FJF 62] FORD L.R., JR., FULKERSON D.R., *Flows in Networks*, Princeton University Press, Princeton, 1962.

[GJ 79] GAREY M. R., JOHNSON D. S., *Computers and Intractability : A Guide to the Theory of NP-Completeness*, W.H. Freeman and Company, San Francisco, 1979

[Ha 74] HANSEN P., *Programmes mathématiques en variables* $0-1$, Thèse, Université Libre de Bruxelles, 1974.

[HDR 94] HAMMER P. L., DAVID J., RADER JR., *Efficient Methods for Solving Quadratic* $0-1$ *Knapsack Problems*, RUTCOR Research Report (RRR 40-94), Rutgers University, 1994.

[HH 92] HAMMER P.L., HOLZMAN R., *Approximation of Pseudo-Boolean Functions; Applications to Game Theory*, ZOR-Methods and Models of Operation Research, Vol. 36, pp. 3-21, 1992.

[HJ 85] HORN R. A., JOHNSON C. R., *Matrix Analysis*, Cambridge University Press, Cambridge, 1985.

[HJM 00] HANSEN P., JAUMARD B., MEYER C., *A simple Enumerative Algorithm for Unconstrained* $0-1$ *Quadratic Programming*, Cahier du GERAD, Université de Montréal, G-59, 2000.

[HR 70] HAMMER P.L., RUBIN A. A., *Some Remarks on Quadratic Programing with* $0-1$ *Variables*, R.I.R.O, Vol. 3, pp. 67-79, 1970.

[HS 68] HAMMER P. L., RUDEANU S., *Boolean Methods in Operation Research and Related Areas*, Springer-Verlag, New-York, 1968.

[KB 90] KALANTARI B., BAGCHI A, *An Algorithm for Quadratic Zero-one Programs*, Naval Research Logistics, Vol. 37, pp. 527-538, 1990.

[Ke 86] KEITH W. N., *Linear Algebra with Applications*, Third edition ; PWS Publishing Company, Boston, 1986.

[KR 87] KALANTARI B., ROSEN J. B., *An Algorithm for the Global Minimization of Linearly Constrained Concave Quadratic Functions*, Mathematics of Operations Research ; Vol. 12, pp. 544-561, 1987.

[Le 83] LEROUX P., *Algebre Lineaire, une Approche Matricielle*, Modulo, Montreal, 1983.

[MF 02] MERZ P., FREISLEBEN B., *Greedy and Local Search Heuristics for Unconstrained Binary Quadratic Programming*, Journal of Heuristics, Vol. 8, pp. 197-213, 2002.

[PR 90] PARDALOS, P., RODGERS G., *Computational Aspects of a Branch and Bound Algorithm for Quadratic* $0-1$ *Programming*, Computing, Vol. 45, pp. 131-144, 1990.

[PS 92] PARDALOS P. M., SOMESH J., *Complexity of Uniqueness and Local Search in Quadratic* $0-1$ *Programming.*, Operations Research Letters, Vol. 11, pp. 119-123, 1992.

[PSU 88] PERESSINI A. L., SULLIVIAN F. E., UHL J.J., *The Mathematics of non Linear Programming*, Springer-Verlag, New-York, 1988.

[Qu 02] QUEYRANNE M., *An Introduction to Submodular Functions and Optimization.*, Available at :
http ://www.ima.umn.edu/optimization/seminar/queyranne.pdf. 2002.

[Ro 72] ROSENBERG I. G., $0-1$ *Optimization and non Linear Programming*, Revue Francaise d'Automatique, Informatique et Recherche Opérationnelle, Vol. 2, pp. 95-97, 1972.

[Ro 75] ROSENBERG I. G., *Reduction of Bivalent Maximization to the Quadratic Case*, Cahiers du Centre d'Etudes de Recherche Opérationelle, Vol. 17, pp. 71-74, 1975.

[Sa 70] SAATY T. L., *Optimization in Integers and Related Extremal Problems*, Macgraw-Hill Book Company, New-York, 1970.

[Sc 00] SCHRIJVER A., *A Combinatorial Algorithm Minimizing Submodular Functions in Strongly Polynomial Time*, Journal of Combinatorial Theory, Vol. 80, pp. 346-355, 2000.

[Sc 03] SCHRIJVER A., *Combinatorial Optimization*, Polyhedra and Efficiency, Vol. B, Volume 24 of Algorithms and Combinatorics, Springer-Verlag, Berlin, 2003.

[Sn 05] SNYMAN J. A., *Practical Mathematical Optimization. An Introduction to Basic Optimization Theory and Classical and New Gradient-Based Algorithms*, Springer Science + Business Media, 2005.

[SMP 08] SOLAYAPAN M., MING K., POH K.L., *Perturbation Method for Solving Unconstrained Binary Quadratic Programming Problem*, Proceedings of World Academy of Science Engineering and Technology, Vol. 27, pp. 185-190, 2008.