

REVUE FRANÇAISE D'INFORMATIQUE ET DE RECHERCHE OPÉRATIONNELLE, SÉRIE ROUGE

P. HUARD

Tour d'horizon : programmation non linéaire

Revue française d'informatique et de recherche opérationnelle, série rouge, tome 5, n^o 1 (1971), p. 3-48.

<http://www.numdam.org/item?id=M2AN_1971__5_1_3_0>

© AFCET, 1971, tous droits réservés.

L'accès aux archives de la revue « Revue française d'informatique et de recherche opérationnelle, série rouge » implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/legal.php>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
<http://www.numdam.org/>

TOUR D'HORIZON : PROGRAMMATION NON LINEAIRE ⁽¹⁾

par P. HUARD ⁽²⁾

Résumé. — *Tour d'horizon sur les méthodes de résolution des programmes mathématiques non linéaires. Cet article ne concerne que les méthodes générales, c'est-à-dire non spécifiques à des problèmes particuliers, comme les programmes quadratiques, fractionnaires, etc... Il se situe essentiellement sur le plan des idées de base des méthodes plutôt que sur le plan théorique, et donne quelques indications sur l'évolution de ces idées.*

I. INTRODUCTION

La littérature concernant les méthodes de résolution de programmes mathématiques non linéaires était pratiquement inexistante avant 1950, mais s'est développée rapidement depuis 1955. Parmi la centaine d'articles sur lesquels nous nous sommes appuyés pour rédiger cette analyse, une quinzaine ont été publiés entre 1950 et 1960 et le reste après 1960. Ce rythme qui va en s'accroissant a été renforcé depuis peu par les articles concernant le contrôle optimal, c'est-à-dire des problèmes d'optimisation définis dans des espaces plus généraux que R^n .

Il est bien entendu nécessaire, dans le cadre d'un article, de faire un choix. Notre tour d'horizon ne concerne que les problèmes définis dans R^n que nous présentons systématiquement sous la forme

$$\begin{array}{c} \text{Maximiser } f(x) \text{ sous les conditions} \\ x \in A \end{array}$$

(1) Exposé fait aux Journées Nationales d'Analyse Numérique, A.F.C.E.T. Bordeaux, 25-27 septembre 1969.

(2) Conseiller scientifique, E.D.F., direction des études et recherches, service Informatique et Mathématiques Appliquées.

où A est un ensemble de R^n , dont la définition est parfois précisée, quand cela est nécessaire : $A = \{ x \in R^n \mid g_i(x) \geq 0, i = 1, 2, \dots, m \}$, les fonctions f et g_i étant généralement différentiables et très souvent concaves. Nous n'envisageons que les méthodes générales, concernant les problèmes non linéaires, à l'exclusion des méthodes spécialement conçues pour des cas particuliers, comme les programmes quadratiques, fractionnaires, etc... A ce sujet, il semble que beaucoup de problèmes de structures particulières sont traités efficacement par des méthodes générales, que l'on adapte, au niveau des calculs, au problème envisagé.

Beaucoup de méthodes proposées jusqu'ici pour la résolution des problèmes avec contraintes conduisent à remplacer le problème donné par une suite de problèmes sans contraintes, dont les résolutions utilisent des méthodes classiques. Nous supposons ces méthodes connues du lecteur, en particulier les méthodes dites de gradients, et nous donnons quelques rappels très succincts quand cela est nécessaire.

Ce tour d'horizon se situe essentiellement sur le plan des idées de base des méthodes, plutôt que sur le plan théorique. Aussi ne donnons-nous guère de précisions sur les hypothèses assurant la convergence des méthodes, ni de démonstrations. Seules quelques méthodes très générales sont décrites de façon moins « littéraire ».

Toutes ces méthodes sont regroupées par familles, le critère choisi étant davantage l'idée de base, le schéma concret de la méthode, plutôt que ses caractéristiques théoriques. Ce procédé est évidemment artificiel, la même méthode pouvant être illustrée par des schémas différents. Ce procédé est assez pratique cependant pour évoquer l'histoire de ces méthodes et l'évolution des idées qui les ont provoquées.

Un tour d'horizon qui se veut utile devrait porter un jugement sur la valeur de chaque méthode décrite. Malheureusement, ce n'est pas si simple : tout d'abord, qu'est-ce qu'une bonne méthode ? Est-ce une méthode rapide sur le plan des calculs ? précise dans ses résultats ? simple à programmer et à mettre en œuvre ? ou bien d'application universelle, pouvant traiter des problèmes de natures variées ? Toutes ces qualités, on s'en rendra compte, n'étant pas toujours compatibles entre elles.

Et si l'on prend une qualité précise pour critère, la rapidité par exemple, le jugement n'est pas encore aisé, car le résultat dépend pour beaucoup de l'écriture du code. Le temps de calcul est souvent divisé par quelques unités après une simple réorganisation des calculs, sans modifier le principe de la méthode.

On pourrait évidemment reproduire l'opinion des auteurs eux-mêmes. Mais ce procédé conduirait sans doute le lecteur de ce tour d'horizon à un optimisme exagéré quand à la programmation non linéaire...

Par contre, nous essayons de préciser, quand c'est possible, si la convergence est établie ou non. Bien qu'ici encore il ne faille pas juger trop vite :

l'exemple classique de la méthode simpliciale, non munie de procédés anti-cyclage, mais satisfaisant ses nombreux utilisateurs, suffit à le rappeler. De même certaines méthodes, dont la convergence théorique est établie, ne convergent pas pratiquement, du fait de la production d'erreurs d'arrondis trop importantes, et inhérentes au principe même de la méthode.

Enfin, nous nous sommes abstenus de signaler trop nettement les antériorités d'invention, nous contentant de préciser les dates de publication, sans nous immiscer dans ces délicates questions de paternité. Notons que les références citées étant souvent des articles de revue, il peut y avoir un décalage d'une année ou deux entre « l'invention » et sa publication.

Nous avons essayé de rédiger un tour d'horizon assez exhaustif... dans le cadre de nos connaissances. Il y aura donc beaucoup d'omissions involontaires. Nous nous sommes appuyés sur divers tours d'horizons antérieurs, dont nous donnons une liste à la fin de cet article. Notre vœu est que ce travail puisse servir de base de départ à d'autres auteurs, et leur permettre d'écrire plus facilement un article plus complet.

1. METHODES DE PENALISATIONS EXTERIEURES

Afin de pouvoir utiliser directement les algorithmes de maximisations dans R^n pour la résolution des problèmes avec contraintes, les méthodes de pénalisations extérieures maximisent la fonction économique $f(x)$ du problème donné tant que la solution courante envisagée est réalisable, et modifient cette fonction dès que la solution sort du domaine. Cette modification conditionnelle de la fonction objective est obtenue en lui retranchant une fonction $p(x)$, de valeur nulle en tout point du domaine, et positive en dehors. Cette fonction $p(x)$, qui se présente donc comme un coût de non satisfaction des contraintes, d'où le nom de pénalisation, peut prendre des valeurs de plus en plus grandes lorsqu'on s'éloigne, en un certain sens, du domaine, et contribue à rapprocher la solution courante de ce dernier. (Mais cette propriété n'est pas nécessaire théoriquement, comme on le verra plus loin.)

La maximisation sans contraintes de $f(x) - p(x)$ n'est qu'une représentation approchée du problème donné, et ne fournira en général qu'un compromis entre la satisfaction des contraintes et la maximisation de $f(x)$. La solution ainsi obtenue sera d'autant meilleure que la pénalisation sera forte, et il semble intuitif que la solution optimale du problème approché puisse tendre vers la solution optimale $\hat{x}$ du problème d'origine, lorsque l'on augmente indéfiniment les valeurs de la pénalisation.

Théoriquement, on est conduit à envisager une suite infinie de fonctions de pénalisations $p_k(x)$, $k = 1, 2, \dots$, dont la valeur augmente indéfiniment avec k , en tout point x non réalisable. Pour chaque valeur de k , on maximise $f(x) - p_k(x)$ sans contraintes, et la solution optimale $\hat{x}^k$ ainsi obtenue converge,

sous des hypothèses assez larges, vers $\hat{x}$. Il est évident que sur le plan pratique pour une tolérance numérique donnée, on ne résout qu'un nombre fini de tels problèmes approchés.

Cette idée semble avoir été introduite par Courant [1.4] en 1943, pour des problèmes d'élasticité sous contraintes (problèmes variationnels avec conditions aux limites). Transposée pour un programme mathématique avec contraintes en égalités $g_i(x) = 0$, $i = 1, 2, \dots, m$, son procédé revient à prendre une pénalisation proportionnelle à $\sum_{i=1}^m (g_i(x))^2$.

La théorie de la convergence fut étudiée beaucoup plus tard, notamment en 1957 par Rubin et Ungar [1.11], et en 1962 par Butler et Martin [1.2], dans des contextes plus généraux.

Afin d'appliquer ce procédé de pénalisations aux problèmes avec contraintes en inégalité, de type $g_i(x) \geq 0$, $i = 1, 2, \dots, m$, Motzkin [1.7] en 1952 introduit des variables d'écart y_i telles que $g_i(x) - y_i = 0$, avec la pénalisation $\sum_{i=1}^m e^{-\lambda_i y_i}$, où les paramètres λ_i sont choisis positifs. Dans les problèmes de maximisation sans contraintes qui en découlent, il faut cependant tenir compte de la condition de signe sur les y_i . Cette difficulté est éliminée avec Camp en 1955, qui remplace les conditions $g_i(x) \geq 0$ par les équations $g_i(x) - (y_i)^2 = 0$.

En 1955 également, et indépendamment de ces travaux antérieurs, Ablow et Brigham [1.1] proposent les deux fonctions de pénalisation suivantes, pour des contraintes en inégalité $g_i(x) \geq 0$, $i = 1, 2, \dots, m$.

$$1) \quad -r \sum_{i=1}^m s_i \min \{ 0, g_i(x) \}$$

$$2) \quad r \sum_{i=1}^m s_i (\min \{ 0, g_i(x) \})^2$$

où les s_i sont des coefficients positifs donnés, et r un paramètre pouvant être pris aussi grand que l'on veut. Ces auteurs résolvent le problème posé, non par une suite discrète de maximisations sans contraintes, mais en suivant la trajectoire continue du point $x(r)$, solution optimale de la maximisation sans contraintes relative à r . Cette trajectoire est obtenue par calcul analogique, et en faisant tendre la valeur du paramètre r vers l'infini, $x(r) \rightarrow \hat{x}$.

Dans le cas de fonctions g_i continuellement différentiables, la pénalisation n° 2 a l'avantage d'avoir des dérivées continues à la frontière du domaine des solutions réalisables, ce qui n'est pas le cas pour la pénalisation n° 1. Cette pénalisation n° 2 a été étudiée (indépendamment de ces auteurs, ainsi que des travaux antérieurs) en 1962 par Pietrzykowski [1.8], qui établit la théorie de la convergence, dans le cadre de la suite discrète de maximisations sans contraintes.

Des classes de fonctions de pénalisations plus générales sont ensuite étudiées sur le plan théorique, notamment en 1967 par Zangwill [1.12], Fiacco [1.6], Eremin [1.5], en 1969 par Polak [1.9]. Citons également les travaux de Roode [1.10] sur les fonctions de Lagrange généralisées (que nous retrouverons au chapitre concernant les méthodes Lagrangiennes) présentant des liens très étroits avec les méthodes de pénalisations. Par ailleurs, le théorème de Zangwill établit que si dans le problème donné, la fonction objective est concave, et le domaine convexe, avec un intérieur non vide, on peut remplacer ce problème par une unique maximisation sans contraintes, en choisissant convenablement la fonction de pénalisation.

En résumé, nous donnons ci-dessous une théorie générale des pénalisations extérieures, que nous empruntons à Polak [1.9]. Le problème à résoudre se présentant sous la forme suivante :

$$\boxed{\begin{array}{c} \text{Maximiser } f(x) \text{ sous les conditions} \\ x \in A \subset R^n \end{array}} \quad (P)$$

avec $f: R^n \rightarrow R$. On appelle suite de fonctions de pénalisations extérieures pour l'ensemble A toute suite de fonctions continues $p_k: R^n \rightarrow R$, $k = 1, 2, \dots$ telle que, $\forall k$:

- (i) $p_k(x) = 0 \Leftrightarrow x \in A$
- (ii) $p_k(x) > 0 \Leftrightarrow x \notin A$
- (iii) $p_{k+1}(x) > p_k(x), \quad \forall x \notin A$
- (iv) $p_k(x) \rightarrow +\infty$ quand $k \rightarrow +\infty, \forall x \notin A$ fixé.

A chaque valeur de k , on associe le problème de maximisation sans contraintes (P_k) suivant :

$$\boxed{\begin{array}{c} \text{Maximiser } \varphi_k(x) \\ x \in R^n \end{array}} \quad (P_k) \quad k = 1, 2, \dots$$

avec $\varphi_k = f - p_k$.

En supposant que A est un ensemble fermé, que f est une fonction continue, atteignant son maximum sur A en un point $\hat{x}$, et que, $\forall k$, φ_k atteint son maximum dans R^n en un point x^k , on peut énoncer les résultats suivants :

1) Évolution des problèmes (P_k)

$$\begin{aligned} \varphi_k(x^k) &> \varphi_{k+1}(x^{k+1}) \geq f(\hat{x}), \quad \forall k: x^{k+1} \notin A \\ f(x^k) &> \varphi_k(x^k) \geq f(\hat{x}), \quad \forall k: x^k \notin A \end{aligned}$$

2) Convergence de la méthode

Si la suite infinie des points $\bar{x}^k$ admet un point d'accumulation, ce point est solution optimale du problème (P) donné.

3) Exemples de fonctions de pénalisations

Si $A = \{x \mid g_i(x) \geq 0, i = 1, 2, \dots, m\}$, les fonctions g_i étant continues, on peut prendre

$$p_k(x) = r_k \sum_{i=1}^m (-\min \{g_i(x), 0\})^\beta$$

avec $\beta \geq 1$ et $0 < r_1 < r_2 < \dots, r_k \rightarrow +\infty$ avec k .

Si $A = \{x \mid g_i(x) = 0, i = 1, 2, \dots, m\}$, et si g représente le vecteur de composantes g_i , on peut prendre

$$p_k(x) = r_k \|g(x)\|^\beta$$

avec $\beta \geq 1$ et $0 < r_1 < r_2 < \dots, r_k \rightarrow +\infty$ avec k .

Si les g_i sont des fonctions concaves, les deux pénalisations générales proposées ci-dessus sont convexes.

4) Combinaison de pénalisations

Soient B et C deux ensembles de R^n , q_k et $r_k, k = 1, 2, \dots$ deux suites de fonctions de pénalisations extérieures, relatives à B et C respectivement. Dans ces conditions, $q_k + r_k$ donne une pénalisation pour $B \cap C$ et $\min \{q_k, r_k\}$ une pénalisation pour $B \cup C$.

5) Algorithme de pénalisation partielle

Si le domaine des solutions réalisables est de la forme $A = B \cap C$ et si l'on dispose d'une suite de pénalisations $q_k, k = 1, 2, \dots$ relative à B , le problème (P) peut être résolu de façon analogue en résolvant la suite de problèmes (P'_k) :

Maximiser $f(x) - q_k(x)$ sous les conditions $x \in C$	(P'_k)
--	----------

En d'autres termes, on peut utiliser le principe des pénalisations extérieures de façon partielle, de façon à ne supprimer qu'une partie des contraintes. Cela peut être intéressant si, par exemple, la contrainte $x \in C$ représente des bornes sur les composantes de x : on sait en effet qu'il est pratiquement aussi facile de maximiser une fonction sur un pavé de R^n que dans R^n tout entier.

2. METHODES INTERIEURES REALISABLES

Nous distinguerons *a priori* trois catégories de méthodes :

- 1) pénalisations « intérieures » avec paramètre,
- 2) pénalisations « intérieures » sans paramètre, désignées sous le nom de « méthode des centres »,
- 3) déplacement suivant des directions réalisables.

En fait, ces trois catégories de méthodes relèvent d'un même procédé : passer d'une solution réalisable à une autre solution réalisable en faisant un compromis entre l'amélioration rapide de la fonction économique et le souci de ne pas se rapprocher trop vite de la frontière du domaine.

2.1. Méthodes de pénalisations « intérieures »

L'idée de base est analogue à celle des pénalisations extérieures, en ce sens que l'on remplace la résolution du problème donné (P) (problème de maximisation avec contraintes dont le domaine a un intérieur supposé non vide) par celle d'une suite de problèmes (P_k) sans contraintes. Les fonctions objectives $\varphi_k(x)$ de ces derniers sont obtenues en retranchant de $f(x)$, fonction objective de (P), une pénalisation $p_k(x)$ définie sur l'intérieur du domaine des solutions réalisables, dont la valeur tend vers $+\infty$ quand on s'approche de la frontière de ce domaine. Cette fonction p_k pénalise donc toute solution réalisable, avec pour critère la proximité de la frontière, et non plus le fait d'être non réalisable.

Les problèmes (P_k), qui consistent à maximiser $\varphi_k(x) = f(x) - p_k(x)$ sur A , représentent un compromis entre la maximisation de f et l'éloignement de la frontière de A . Leurs solutions optimales sont des points intérieurs à A , ce qui fait que, pour la résolution numérique, ce sont des problèmes pratiquement sans contraintes. Ces solutions sont donc toujours réalisables, et d'autant meilleures pour f que la pénalisation p_k est faible. En faisant tendre ces pénalisations vers zéro, les solutions optimales des problèmes (P_k) tendent, sous certaines hypothèses, vers une solution optimale $\hat{x}$ de (P).

La première méthode intérieure avec pénalité semble être due à Frisch [2.11], en 1955, avec son double « gradient » ou « potentiel logarithmique ». La théorie exposée était approximative, la méthode étant proposée dans un but de calculs pratiques. Elle utilisait pour pénalisations la somme des logarithmes des premiers membres des contraintes, ces derniers se présentant sous la forme

$$g_i(x) \geq 0, \quad i = 1, 2, \dots, m.$$

Plus précisément Frisch posait

$$p_k(x) = -r_k \sum_{i=1}^m \text{Log}(g_i(x)),$$

avec $0 < r_{k+1} < r_k$, $r_k \rightarrow 0$ quand $k \rightarrow +\infty$. Elle est reprise plus tard par Peuchot [2.22], en 1960, par Parisot [2.21] en 1961, puis par Pomentale [2.24] en 1965. Peuchot « tronque » le domaine, à chaque itération, par une contrainte $f(x) \geq C^{\text{te}}$, un peu « en dessous » de la dernière solution $\hat{x}^k$ trouvée, ceci afin de « recentrer » le gradient. Cette proposition pratique est consécutive à des difficultés rencontrées dans la résolution numérique des problèmes sans contraintes (P_k), les fonctions à maximiser comportant des logarithmes, dont les valeurs (ainsi que celles de leurs dérivées de tous ordres) deviennent infinies quand on se rapproche de la frontière. Or, les solutions $\hat{x}^k$ tendent naturellement vers la frontière si $\hat{x}$ s'y trouve (ce qui est le cas général). Les méthodes de gradients utilisées s'adaptent mal à des variations rapides de la valeur de la fonction à maximiser et de ses dérivées.

Indépendamment des travaux de Frisch et de ses successeurs, Rosenbrock [2.25] en 1960, faisait intervenir une pénalisation très différente, agissant uniquement dans une zone de largeur ε voisine de la frontière.

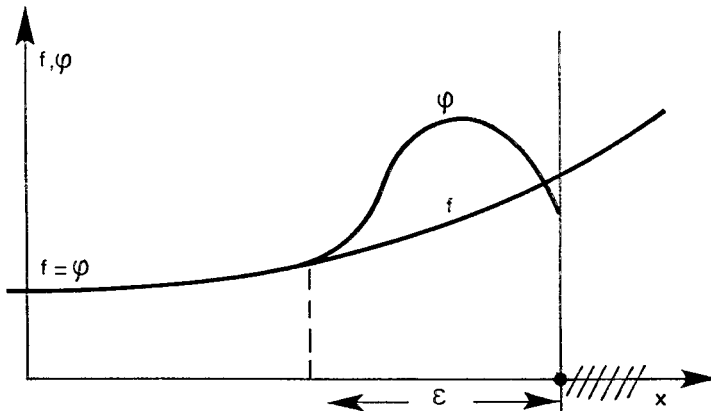


Figure 1

Dans cette bande frontière, f est remplacée par une fonction φ admettant un maximum à l'intérieur de la bande. Différentes améliorations pratiques ont été apportées par Wood [2.29] et Box [2.1] en 1965, afin d'éviter des oscillations dans la résolution du problème sans contraintes.

Indépendamment de Frisch également, Carroll [2.3] en 1959 utilise pour pénalisation dans le cas où $A = \{x \mid g_i(x) \geq 0, i = 1, 2, \dots, m\}$, la fonction $r_k \sum_{i=1}^m 1/g_i(x)$, mais sans établir de théorie. Cette idée est reprise en 1964 par

Fiacco et McCormick [2.7], qui apportent de nombreux résultats théoriques sur la convergence et sur des relations fort intéressantes avec la dualité. Ils réalisent également des codes efficaces. Leurs travaux sont ensuite étendus à des espaces topologiques par Stong [2.26] en 1965.

En étudiant ces travaux, Lootsma [2.19] réunit les méthodes de pénalisations intérieures et extérieures, et propose un algorithme mixte, où simultanément une partie des contraintes est remplacée par des pénalisations extérieures et l'autre par des pénalisations intérieures. Lootsma étudie également l'extrapolation des solutions optimales relatives aux problèmes (P_k) , question très importante dans l'accélération des calculs [2.18]. Enfin, il établit les relations entre les méthodes de pénalisations utilisant un paramètre (paramètre faisant varier la valeur de la pénalisation de $p_k(x)$ à $p_{k+1}(x)$ pour un point x fixé) avec les méthodes sans paramètre envisagées plus loin.

Dans le cadre des méthodes de Fiacco et McCormick, Fletcher et McCann [2.9] améliorent la résolution numérique des problèmes (P_k) en utilisant et en adaptant le calcul de la matrice des dérivées secondes (Hessien) à l'aide d'une méthode de Davidon.

La méthode originale proposée par Goldstein et Kripke [2.12] en 1964 se démarque nettement des méthodes précédentes, bien qu'elle relève du procédé des pénalisations. Elle permet de traiter tout domaine convexe borné dont on peut définir les plans tangents (plans d'appui), et prend en compte un nombre infini de contraintes $g_i(x) \geq 0$, $i = 1, 2, \dots$

Le domaine A étant défini comme l'intersection de « tranches » d'espaces D soit :

$$A = \bigcap_{h=1}^{\infty} D_h$$

avec

$$D_h = \{ x \in R^n \mid -1 - 1/h \leq g_i(x) \leq 1 + 1/h, i = 1, 2, \dots \}$$

Le problème (P_k) est défini à chaque étape k par :

$$\boxed{\begin{array}{c} \text{Maximiser } f(x) - \frac{1}{k^2} \sum_{i=1}^k (g_i(x))^{2k} \\ x \in R^n \end{array}} \quad (P_k)$$

Il est à remarquer ici l'utilisation d'ensembles D_h un peu « gonflées » par rapport à A , et l'utilisation progressive des fonctions des contraintes $g_i(x)$. De plus, la solution $\bar{x}$ de P_k n'est pas obligatoirement dans A . Cette utilisation de domaines plus grands que A , tendant vers A (en un certain sens) au cours des itérations, se retrouve chez Fiacco [2.6] et est reprise par Huard [2.16] dans le cadre des méthodes de centres, étudiées plus loin.

Citons encore, comme méthode de pénalisations un peu spéciale, celle de Frehel [2.10], qui utilise la pénalité classique $\sum_{i=1}^m \text{Log } g_i(x)$, puis en modifie le gradient dans la résolution du problème sans contraintes (P_k) . Au lieu du gradient résultant $\nabla f(x) - \sum_{i=1}^m \nabla g_i(x)/g_i(x)$, il utilise le gradient $\nabla f(x) - u(x)$, où $u(x)$ est la projection du gradient logarithmique sur le plan orthogonal à $\nabla f(x)$

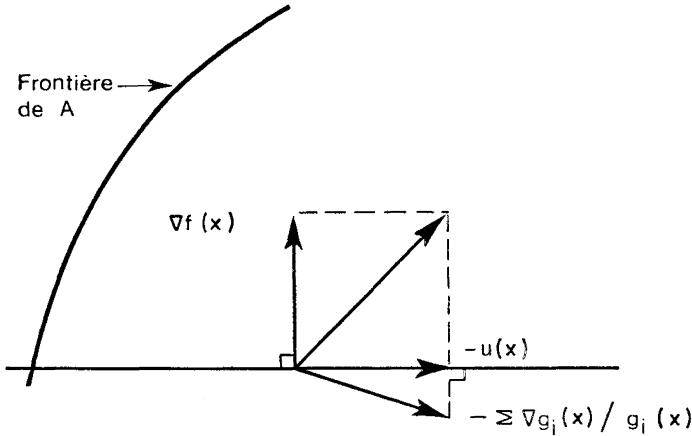


Figure 2

Pour résumer, nous donnons ci-dessous la présentation générale des pénalisations intérieures de Polak [2.23].

Le problème à résoudre se présentant sous la forme :

$$\begin{array}{c} \text{Maximiser } f(x) \text{ sous les conditions} \\ x \in A \subset R^n \end{array} \quad (P)$$

avec $\overset{0}{A} \neq \emptyset$, $A = \overline{\overset{0}{A}}$ (A est égal à la fermeture de son intérieur), $f: R^n \rightarrow R$, on appelle fonction de pénalisation intérieure pour l'ensemble A toute suite de fonctions continues $p_k: R^n \rightarrow R$, $k = 1, 2, \dots$ telles que, $\forall k$:

- (i) $0 < p_k(x) < +\infty$, $\forall x \in \overset{0}{A}$
- (ii) $\forall \{ \overset{h}{x} \in \overset{0}{A} \mid h = 1, 2, \dots \}$ telle que $\overset{h}{x} \rightarrow \bar{x} \in \text{Fr}(A)$ quand $h \rightarrow +\infty$, on a $p_k(\overset{h}{x}) \rightarrow +\infty$ quand $h \rightarrow +\infty$.
- (iii) $p_{k+1}(x) < p_k(x)$, $\forall x \in \overset{0}{A}$
- (iv) $p_k(x) \rightarrow 0$ quand $k \rightarrow +\infty$, $\forall x \in \overset{0}{A}$

A chaque valeur de k , on associe le problème (P_k) suivant :

Maximiser $\varphi_k(x)$ sous les conditions

$$x \in \overset{0}{A}$$

$(P_k) \quad k = 1, 2, \dots$

avec $\varphi_k = f - p_k$

En supposant que f est continue, atteignant son maximum sur A en un point $\hat{x} \in A$, que $\{x \in A \mid f(x) \geq \lambda\}$ est borné, $\forall \lambda$, et que $\forall k$, φ_k atteint son maximum sur $\overset{0}{A}$ en un point $\overset{k}{x}$, on peut énoncer les résultats suivants :

1) *Évolution des Problèmes (P_k)*

$$\varphi_k(\overset{k}{x}) \leq \varphi_{k+1}(\overset{k+1}{x}) \leq f(\hat{x}), \forall k$$

2) *Convergence de la méthode*

Tout point d'accumulation de la suite des points $\overset{k}{x}$ est solution optimale du problème (P) donné.

2.2. Méthode des centres

Il s'agit d'une famille de méthodes semblables à celles des pénalisations, décrite précédemment, et conduisant à des problèmes (P_k) pratiquement sans contraintes. La différence réside principalement dans le fait que la diminution des pénalisations n'est pas réalisée arbitrairement, en modifiant un paramètre, mais est déterminée automatiquement à l'étape k par la solution du problème précédent (P_{k-1}) .

On utilise pour cela une famille de fonctions définies pour chaque sous-ensemble ou « tronçon » $E(\lambda) = A \cap \{x \mid f(x) \geq \lambda\}$, pour toute valeur réelle de λ . En d'autres termes, on considère une fonction numérique $d(x, \lambda)$, où x est la variable et λ un paramètre scalaire continue, et dont le domaine de définition varie avec λ . On dit que d est une F -distance si elle est nulle en tout point de la frontière de tout tronçon, positive à l'intérieur, et si sa valeur $d(x, \lambda)$ en un point fixé x décroît quand λ croît (c'est-à-dire quand le tronçon $E(\lambda)$ diminue par inclusion).

Pour un tronçon donné $E(\lambda)$, on appelle « centre » tout point qui maximise $d(x, \lambda)$ sur $E(\lambda)$. Si l'intérieur de $E(\lambda)$ n'est pas vide, ce centre est un point intérieur, car sur la frontière, d a toujours une valeur nulle. Ce qui veut dire que le calcul d'un centre se ramène pratiquement à la maximisation de $d(x, \lambda)$ sans contraintes (λ fixé).

L'algorithme de la méthode des centres revient alors à résoudre les problèmes (P_k) suivants :

Maximiser $d(x, \lambda_k)$ sous les conditions $x \in E(\lambda_k)$	$(P_k) \quad k = 1, 2, \dots$
---	-------------------------------

avec $\lambda_k = f(\hat{x}^{k-1})$, $\hat{x}^{k-1}$ étant la solution optimale de (P_{k-1}) c'est-à-dire le centre

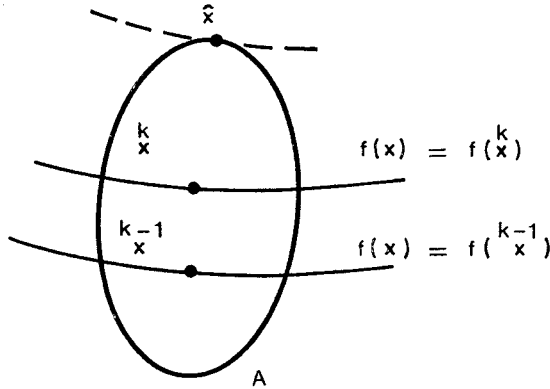


Figure 3

de $E(\lambda_{k-1})$. La condition $x \in E(\lambda_k)$ peut être pratiquement supprimée, d'après la remarque précédente.

On suppose, bien entendu, que l'intérieur de A n'est pas vide. Sous des hypothèses très larges (comme par exemple $Fr(A^0) = Fr(A)$, qui exclut l'existence de solutions réalisables isolées), on obtient, quelle que soit la F -distance choisie, une suite infinie de centre $\hat{x}^k$, convergeant vers la solution optimale $\hat{x}$ du problème posé, la suite des valeurs correspondantes $f(\hat{x}^k)$ étant strictement croissante. De plus il n'est pas nécessaire de maximiser exactement $d(x, \lambda_k)$ sur $E(\lambda_k)$, à condition que l'erreur ϵ_k ainsi commise tende vers zéro quand $k \rightarrow +\infty$.

Exemples de F -distance

Si A est défini par $g_i(x) \geq 0$, $i = 1, 2, \dots, m$, on peut prendre

$$d(x, \lambda) = (f(x) - \lambda)^\alpha \prod_{i=1}^m g_i(x)$$

où le paramètre $\alpha \geq 1$ permet d'accélérer la convergence. On peut prendre également :

$$d(x, \lambda) = \min \{ f(x) - \lambda, \quad g_i(x) \mid i = 1, 2, \dots, m \}$$

Plus généralement, si $A = B \cap C$ et si d_B et d_C sont deux F -distances respectivement pour B et pour C , alors $d_B \cdot d_C$ et $\min \{ d_B, d_C \}$ sont des F -distances pour A .

L'hypothèse faite sur l'intérieur de $A \overset{0}{\neq} \emptyset$ exclut *a priori* les contraintes en égalité. On peut cependant adapter la méthode à ce cas, en la généralisant de la façon suivante :

Soit $A = B \cap C$, avec $B \overset{0}{\neq} \emptyset$, et $d(x, \lambda)$ une F -distance relative aux tronçons de B seulement. On considère les problèmes (P_k) suivants :

$$\boxed{\begin{array}{l} \text{Maximiser } d(x, \lambda_k) \text{ sous les conditions} \\ x \in E(\lambda_k), x \in C \end{array}} \quad (P_k) \quad k = 1, 2, \dots$$

avec $\lambda_k = f(\overset{k-1}{x})$, $\overset{k-1}{x}$ étant la solution optimale de (P_{k-1})

$$| E(\lambda_k) = B \cap \{ x \mid f(x) \geq \lambda_k \}$$

Sous des hypothèses très larges, la suite infinie des points $\overset{k}{x}$ converge vers $\hat{x}$, maximisant f sur $B \cap C$. Dans les problèmes (P_k) la contrainte $x \in E(\lambda_k)$ est pratiquement inutile, du fait que $\overset{k}{x}$ est un point intérieur à $E(\lambda_k)$. Seule demeure la contrainte $x \in C$.

Une autre généralisation possible est le remplacement, à l'étape k , du domaine A par un ensemble plus grand A_k , à condition que la suite A_k diminue par inclusion [2.16]. On retrouve alors une méthode du type pénalisation extérieure. L'idée d'approcher le domaine par des ensembles plus grands, mais diminuant au cours des itérations, avait déjà été utilisée en particulier par Golstein et Kripke [2.12], Fiacco [2.5], Kelley [4.21].

La méthode des centres, dont la théorie est introduite en 1963 [2.14] apparaît donc comme une méthode très voisine des méthodes de pénalisations intérieures. Si une F -distance n'est pas une pénalisation au sens strict, elle est douée de propriétés analogues : il suffit, pour s'en convaincre, de remplacer une pénalisation — $p(x)$ par $e^{-p(x)}$, ce qui ne change pas le résultat pour une maximisation, car e^y est une fonction croissante de y . On pourrait donc supprimer le paramètre dans les méthodes de pénalisations intérieures, en considérant des pénalisations $p_k(x)$ définies sur les tronçons $A \overset{0}{\cap} \{ x \mid f(x) > f(\overset{k-1}{x}) \}$, et non plus sur l'ensemble constant A . Peuchot [2.22] avait déjà utilisé ce procédé pour des raisons de calcul numérique. Réciproquement, on pourrait

supprimer les troncatures dans une méthode de centres en introduisant un paramètre de la manière suivante :

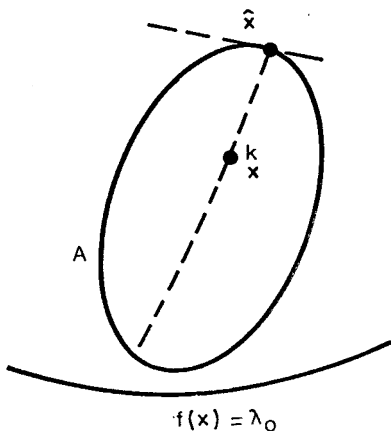


Figure 4

Soient $\lambda_0 < f(\hat{x})$

$d'(x, \lambda)$ une F -distance relative à A

k un paramètre

En posant

$$d_k(x) = \min \left\{ d'(x, \lambda_0), \frac{1}{k} (f(x) - \lambda_0) \right\}$$

$\hat{x}^k$: solution intérieure maximisant $d_k(x)$ sur A
le point $\hat{x}^k$ converge, sous certaines hypothèses, vers la solution optimale $\hat{x}$ quand $k \rightarrow +\infty$.

Ces relations entre méthodes de pénalisations et de centres sont étudiées en 1967 par Fiacco [2.6] et en 1968 par Lootsma

[2.20], et l'extrapolation de la ligne des centres par Faure en 1966, Trémolières [2.28] et Lootsma [2.18] en 1968.

La méthode des centres a été généralisée en 1966 à des espaces topologiques, par Bui-Trong-Lieu et Huard [2.2]. Par ailleurs, la linéarisation des calculs au niveau de la recherche d'un centre conduit à une séquence de programmes linéaires de taille constante et de structure analogue à celle du problème d'origine. Une telle variante (Huard [2.15]) est décrite plus loin avec les procédés dits de linéarisation.

Signalons enfin une méthode originale, dite « des ricochets », proposée en 1966 par Greenstadt [2.13] utilisant les notions de centres et de troncature. On « suit » le gradient jusqu'à ce que l'on rencontre la frontière du domaine. Là, on retourne à l'intérieur du domaine, en demeurant dans le plan tangent à l'équipotentielle économique, et en suivant la direction t faisant le plus grand angle avec la frontière. On s'arrête sur cette demi-droite, au milieu du domaine et l'on recommence le processus. Cette idée a été adaptée, dans le contexte de la méthode des centres, par Trémolières.

2.3. Méthode des directions réalisables

Il s'agit d'une méthode itérative, présentant plusieurs variantes, proposée par Zoutendijk [2.30] en 1960. On génère une suite de solutions réalisables $\hat{x}^k$, le passage de $\hat{x}^k$ à $\hat{x}^{k+1}$ se faisant suivant une direction $\hat{y}^k$ à la fois améliorante et réalisable, c'est-à-dire ayant les propriétés suivantes :

$$(i) \quad \nabla f(\hat{x}^k) \cdot \hat{y}^k > 0$$

(ii) $\exists \theta_M > 0 :$

$$^k x + \theta y^k \in A, \quad \forall \theta \in [0, \theta_M]$$

On choisit alors pour ^{k+1}x le point qui maximise f sur le segment réalisable de la demi-droite d'extrémité $^k x$ et de direction y^k . Dans le cas de contraintes

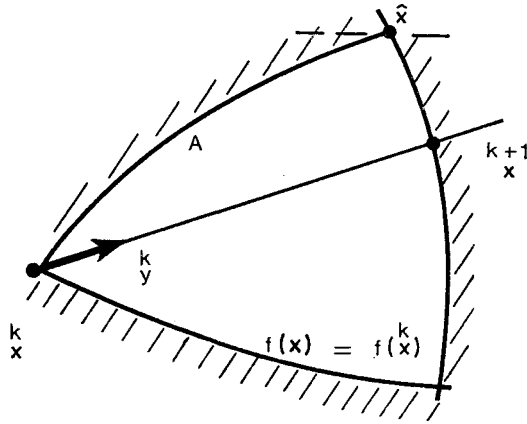


Figure 5

non linéaires, si l'on veut être sûr que ce segment réalisable ne soit pas réduit à un point, il faut prendre une direction y^k « intérieure » à A , ou plus précisément, une direction intérieure au cône tangent en x^k à l'intersection de A et de $\{x \mid f(x) \geq f(x^k)\}$. Si A est défini par des inégalités $g_i(x) \geq 0$, $i = 1, 2, \dots, m$, y^k est solution du programme linéaire suivant, où les inconnues sont y (à n composantes) et μ (scalaire) :

Maximiser μ sous les conditions

$$\nabla f(x^k) \cdot y \geq \mu$$

$$\nabla g_i(x^k) \cdot y \geq \mu, \quad \forall i \in E(x^k) = \{i \mid g_i(x^k) = 0\}$$

$$-1 \leq y_j \leq +1, \quad \forall j = 1, 2, \dots, n$$

La solution optimale y donnée par ce programme linéaire est telle que le plus petit des angles formés avec les normales aux contraintes actives et le gradient de la fonction économique, évalués en x^k , est maximum.

Tel qu'il vient d'être décrit, l'algorithme ne converge pas toujours. Il peut se produire le phénomène désigné sous le nom évocateur de « zig-zag », et qui

se traduit par la convergence de la suite des $\hat{x}^k$ vers un point différent de $\hat{x}$, après une suite infinie d'oscillations des $\hat{x}^k$ entre deux ou plusieurs contraintes, la longueur des segments réalisables correspondants tendant vers zéro. Divers procédés « anti-zig-zag » ont été proposés par Zoutendijk. On peut, par exemple, remplacer la définition précédente de $E(\hat{x}^k)$ (ensemble des contraintes actives en $\hat{x}^k$) par la suivante :

$$E(\hat{x}^k) = \{ i \mid g_i(\hat{x}^k) < \varepsilon \}$$

c'est-à-dire par l'ensemble des contraintes actives à ε près, ε étant une constante arbitraire positive. Chaque fois que l'algorithme s'arrête, on diminue la valeur de ε , et en faisant tendre ainsi ε vers zéro, l'algorithme converge vers $\hat{x}$.

Les points $\hat{x}^k$ fournis par la méthode des directions réalisables ne sont pas obligatoirement des solutions intérieures, car lors de la maximisation sur le segment réalisable, la solution obtenue $\hat{x}^{k+1}$ peut se trouver à l'autre extrémité du segment, c'est-à-dire sur la frontière de A . Seules les directions de déplacement sont intérieures, et c'est en ce sens que nous avons classé cette méthode dans la famille des méthodes intérieures réalisables. On remarquera que la notion de direction centrale remplace ici la notion de solution centrale de la méthode des centres.

Lorsque les contraintes sont linéaires, il n'est plus nécessaire de prendre des directions intérieures. On obtient alors un cas limite de la méthode, qui peut être préférablement classé avec les méthodes de projection (chapitre suivant).

La méthode de Zoutendijk, qui ne semble pas avoir attiré beaucoup d'attention au début de sa publication, a par la suite été reprise par différents auteurs, comme Topkis et Veinott [2.27] en 1967 et Polak [2.23] en 1969. Elle a été étendue à des espaces plus généraux, pour résoudre des problèmes de control optimal (Dem'Janov et Rubinov [2.5] en 1965). Zoutendijk [2.31] propose lui-même une modification de sa méthode en 1966, qui en fait une véritable méthode de solutions intérieures : c'est la méthode « Modified Feasible Directions » ou M.F.D., qui ne nécessite plus de procédés anti-zig-zag. Le successeur $\hat{x}^{k+1}$ de $\hat{x}^k$ est pris au milieu du segment réalisable.

3. METHODES DE FRONTIERE OU DE PROJECTIONS

Une tentative assez naturelle pour adapter les méthodes de maximisation dans R^n aux problèmes avec contraintes est la projection du déplacement sur la frontière du domaine, afin de rendre ce déplacement réalisable, tout en modifiant le moins possible la méthode d'origine et conserver ainsi le maximum de son efficacité.

Divers procédés, relevant de près ou de loin de cette idée, ont été proposés. Nous décrivons trois groupes, s'appuyant sur des méthodes de maximisations sans contraintes différentes, ainsi qu'une quatrième catégorie utilisant des changements de variables.

3.1. Adaptation des méthodes de gradients

Pour une méthode de gradient, il s'agit de projeter ce gradient sur le domaine, pratiquement sur sa frontière (la convexité du domaine arrange bien les choses). On chemine alors sur la frontière, selon la plus forte pente « relative », c'est-à-dire permise par les contraintes. Si ces dernières sont linéaires, le domaine est un polyèdre convexe, et la projection est dans ce cas facilement calculable. Rosen [3.14] en 1960 propose une telle méthode sous le nom de méthode du gradient projeté. Cet article important contient la théorie, ainsi que de nombreuses considérations pratiques, concernant le calcul de la projection du gradient, la détermination des facettes intéressées par cette projection, et la maximisation sur une droite. Cette méthode donne lieu à un code efficace. Un problème de taille importante, représentant des investissements à long terme d'Électricité de France, a été résolu facilement en 1962 avec ce code.

Il est à noter un court article de Abraham [3.1], publié en 1961, et indépendamment, semble-t-il, du travail de Rosen. Il s'agit de l'adaptation, aux cas des contraintes linéaires, de la méthode particulièrement simple qui consiste à choisir les directions de déplacements suivant les différents vecteurs d'une base de R^n . Ici, la base est changée chaque fois que cela est nécessaire, de telle sorte qu'on ne puisse quitter le polyèdre lorsqu'on se déplace suivant les vecteurs de la base.

Tout de suite après le cas linéaire, Rosen [3.15] adapte sa méthode au cas des contraintes non linéaires. Son procédé consiste à linéariser les contraintes afin de pouvoir utiliser la méthode du gradient projeté. Lorsque le déplacement sur la variété linéaire tangente éloigne trop la solution courante $\bar{x}^k$ du domaine, on revient sur ce dernier suivant une direction normale à la frontière. D'où l'expression d'« hémistiches » employée dans la littérature pour désigner ces déplacements (on pourrait dire aussi « en dents de scie »). Le procédé des dents de scie est également utilisé pour d'autres méthodes, comme on le verra plus loin. Un article de Roberts et Lyvers [3.13] en présente une variante en 1961, adaptable à diverses méthodes de gradients.

Mais d'autres procédés sont envisagés pour suivre la frontière dans le cas non linéaire. Bigg [3.2] utilise en 1963 la méthode de plus forte pente à l'intérieur du domaine. Quand la solution $\bar{x}^k$ sort du domaine, un sous-ensemble N de contraintes n'est plus vérifié. On présume alors que la solution optimale cherchée se trouve sur la variété non linéaire $\{ x \mid g_i(x) = 0, i \in N \}$ c'est-à-dire

l'intersection des contraintes non satisfaites par $\overset{k}{x}$. On utilise la méthode de Lagrange pour résoudre le problème :

Maximiser $f(x)$ sous les conditions

$$g_i(x) = 0, \quad i \in N$$

La solution trouvée remplace $\overset{k}{x}$ si elle est réalisable. Sinon on raccourcit le déplacement.

Par ailleurs, Kalfon, Ribière et Sogno [3.10] proposent en 1968 une méthode permettant de ne pas sortir du domaine, en calculant simultanément le déplacement tangentiel et le déplacement normal à la frontière. En fait, ils combinent la méthode des directions réalisables et celle du gradient projeté. En utilisant un procédé de triangularisation dans le calcul de la matrice de projection, ces auteurs obtiennent un code très performant.

Citons également Tanabe [3.18] qui résoud directement le problème de la trajectoire sur la frontière, constamment tangente à la projection du gradient, en utilisant une méthode de Runge et Kutta.

3.2. Adaptation de la méthode de Davidon

On sait que si $\hat{x}$ maximise dans R^n une fonction quadratique définie négative, cette solution peut être obtenue par la relation

$$\hat{x} = x - H^{-1} \nabla f(x)$$

où H est la matrice (constante) des dérivées secondes de f , et x un point quelconque de R^n . Si f n'est pas quadratique, cette relation peut être utilisée sous la forme itérative

$$\overset{k+1}{x} = \overset{k}{x} - (H(\overset{k}{x}))^{-1} \nabla f(\overset{k}{x})$$

où la matrice des dérivées secondes $H(x)$ n'est plus constante. Pour les problèmes de grandes tailles, le calcul de $H(x)$, puis de son inverse, est prohibitif, et Davidon [3.4] avait proposé en 1959 la méthode suivante : on se donne une approximation de $(H(x))^{-1}$, et on corrige cette approximation à chaque étape, à l'aide d'une formule tenant compte des valeurs du gradient en $\overset{k}{x}$ et $\overset{k+1}{x}$. Dans le cas d'une fonction quadratique de n variables, l'algorithme fournit la solution exacte en n itérations. Dans ce même article, Davidon suggérait l'adaptation de sa méthode au cas de contraintes en égalités linéaires, en plongeant en quelque sorte le problème dans la variété linéaire définie par ces équations. La modification correspondante des relations de passage entre une approximation de l'inverse de $H(x)$ et la suivante est étudiée également en 1966 par Goldfarb [3.8].

Dans le cas des contraintes en inégalités linéaires, on retrouve un procédé analogue à celui utilisé par Bigg, et qui consiste à remplacer les contraintes

d'inégalités actives par des égalités. Ce procédé, qui pose un problème déjà rencontré dans la méthode de Rosen (abandon ultérieur des contraintes qui ne sont plus actives) est étudié en 1966 par Goldfarb [3.8], en 1968 par Murtagh et Sargent [3.11], dans le cas des contraintes linéaires, et par Davies [3.5] dans le cas des contraintes non linéaires, en le combinant au procédé des « hémistiches ».

3.3. Adaptation de la méthode de Rosenbrock

La méthode proposée en 1960 par Rosenbrock [3.16] pour maximiser une fonction dans R^n , et désignée sous le nom de « recherche directe », repose sur une approximation quadratique de la fonction et conduit de proche en proche à déterminer les axes principaux de la fonction quadratique. Elle utilise les déplacements précédents pour construire le nouveau système de directions de déplacements orthogonales. Son adaptation aux contraintes linéaires est faite en 1964 par Davies et Swann [3.17], [3.6], qui modifient le procédé de construction de la nouvelle base orthogonale comme suit : si les k contraintes actives au point considéré forment, par leur intersection, une variété linéaire de dimension $n - k$, la nouvelle base orthogonale construite est de dimension $n - k$, et forme une base pour cette variété linéaire (ce résultat est obtenu en combinant le procédé classique d'orthogonalisation de Schmidt avec la projection sur une variété linéaire). Dans ces conditions, les déplacements ultérieurs peuvent demeurer réalisables.

3.4. Projection et changement de variables

Une difficulté commune à toutes les méthodes de projection ou de frontière est de déterminer un cheminement sur une partie de la frontière, constituée par l'intersection de plusieurs contraintes actives au point considéré. Un changement de variables peut théoriquement supprimer cette difficulté.

Un point x de R^n est défini naturellement par ses composantes $(x_1, x_2, \dots, x_n)$. Il peut être également défini par n valeurs $y_1, y_2, \dots, y_n$, fonctions de x , données par :

$$y_i = g_i(x), \quad i = 1, 2, \dots, n$$

Sous certaines hypothèses classiques, la connaissance des y_i permet de déterminer les composantes de x , et la valeur de la fonction économique $f(x)$ est alors donnée par une fonction $h(y)$.

En particulier, pour certains indices i , on peut avoir $g_i(x) \equiv x_i$

Si l'on désire définir un cheminement sur une variété non linéaire intersection de k surfaces d'équations $g_i(x) = 0$, $i = 1, 2, \dots, k$, on peut effectuer un changement de variables, remplaçant $(x_1, x_2, \dots, x_n)$ par $(y_1, y_2, \dots, y_k, x_{k+1}, \dots, x_n)$, avec $y_i = g_i(x)$ et imposer la valeur nulle à $y_1, y_2, \dots, y_k$: sous cette condition,

n'importe quelles valeurs attribuées à $x_{k+1}, \dots, x_n$ donneront un point de la variété non linéaire envisagée. Si l'on veut maximiser $f(x)$ sur cette variété

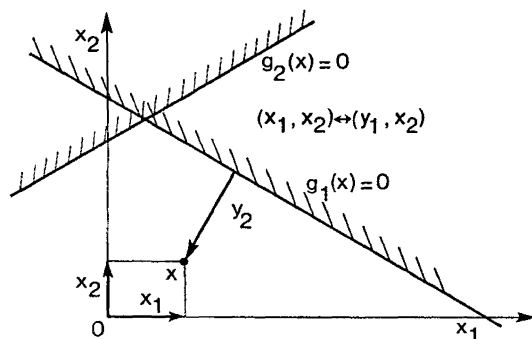


Figure 6

le déplacement sera choisi de façon à maximiser $h(y)$, plus précisément $h(0, 0, \dots, 0, x_{k+1}, \dots, x_n)$, à l'aide d'une méthode de maximisation sans contraintes (dans R^{n-k} en l'occurrence).

Si le problème est de maximiser $f(x)$ sous des conditions d'inégalités $g_i(x) \geq 0$, $i = 1, 2, \dots, m$, il suffit de prendre pour variété non linéaire, envisagée précédemment, l'intersection des contraintes devenues actives, lors du cheminement de la solution courante. Après le changement de variables indiqué, on obtient un problème dans R^{n-k} (k étant le nombre de contraintes actives) avec une difficulté supplémentaire : il est nécessaire, au cours du cheminement ultérieur, de surveiller le signe des fonctions $g_i(x)$, $i = k+1, \dots, m$ afin de toujours satisfaire aux conditions d'inégalités $g_i(x) \geq 0$. Si l'une de ces fonctions s'annule et tend à devenir négative, l'ensemble des contraintes actives n'est plus le même, et l'on doit effectuer un nouveau changement de variables.

Si certaines contraintes $g_i(x) \geq 0$ sont de simples conditions de bornes (inférieures ou supérieures) sur les variables x_j , il n'est pas nécessaire de les faire intervenir dans les changements de variables, car les méthodes de maximisation sans contraintes, du type « gradient », s'adaptent sans difficulté à ces contraintes simples, par annulation des composantes du déplacement relatives aux variables atteignant leurs bornes : il s'agit en fait d'une projection de la direction de déplacement sur la frontière du pavé représentant le domaine des solutions réalisables.

Sur le plan pratique des calculs, la méthode des changements de variables n'est guère utilisable directement, en dehors du cas des contraintes linéaires, du fait de la difficulté d'exprimer les fonctions inverses $x_i(y)$.

Si les contraintes sont linéaires, les calculs relatifs aux changements de variables peuvent se ramener pratiquement à des changements de bases (utilisés dans la méthode simpliciale) après avoir présenté les contraintes sous forme

standard (équations linéaires et variables non négatives). Cette méthode proposée par Wolfe [3.19] en 1962, est étudiée théoriquement et pratiquement par différents auteurs, et donne lieu à des codes performants : Faure, Guigou et Huard [3.7] en 1963, Ribière [3.12] en 1968. Il est à noter que cet algorithme nécessite en toute rigueur des procédés anti-zig-zag (non utilisés dans la pratique) pour assurer la convergence, comme dans la méthode des directions réalisables de Zoutendijk, ainsi que le montre l'exemple simple à trois variables décrit par Wolfe [3.20]. L'adaptation à des contraintes non linéaires est faite en 1966 par Carpentier et Abadie [3.3] en ajoutant à la méthode du gradient réduit l'idée des dents de scie de Rosen. Des codes performants furent écrits par Niederlander puis par Guigou [3.9].

4. METHODES DE LINEARISATIONS

Le succès de la méthode simpliciale dans la résolution des programmes linéaires, sa technique de calcul sans cesse améliorée un peu partout dans le monde, et la diffusion de codes performants pouvant traiter des problèmes de plus en plus grands, ont provoqué la naissance de nombreuses méthodes dites de linéarisation, dont le principe commun est de ramener la résolution du problème non linéaire donné à celle d'une suite de programmes linéaires d'approximation.

Qu'entend-on par linéarisation? Étant donné une fonction non linéaire dérivable $h(x)$, un premier procédé consiste à lui substituer la fonction affine (linéaire plus constante) de variable x et de paramètre y :

$$h'(x, y) = h(y) + \nabla h(y) \cdot (x - y)$$

Le point y est le point de linéarisation, et on a l'égalité entre les deux fonctions pour $x = y$. Cette approximation h' , qui ne sera généralement valable que pour une utilisation au voisinage de y , revient à remplacer le graphe de $h(x)$ par son plan tangent au point $[y, h(y)]$. D'où le nom d'approximation tangentielle. Un autre procédé, conduisant à un changement de variables, peut être utilisé dans le cas d'un programme du type : Maximiser $f(x)$ sous les conditions $g_i(x) \geq 0, i = 1, 2 \dots m$, lorsque les fonctions sont concaves. Il s'appuie sur la notion de barycentre, et nous désignerons les méthodes qui l'utilisent sous le nom de méthodes barycentriques.

4.1. Méthodes tangentielles

L'approximation d'un programme non linéaire par un programme linéaire, obtenu au moyen d'une linéarisation tangentielle unique, peut conduire à des erreurs non prévisibles, sans grand rapport avec la proximité du point de linéarisation y et de la solution optimale $\hat{x}$. On peut lire à ce sujet l'article de Baumol et Bushnell [4.2].

Si l'on désire adapter la méthode classique de Newton, utilisée dans la résolution des systèmes d'équations non linéaires, aux programmes mathématiques, l'idée la plus simple qui vient à l'esprit est de procéder comme suit : on linéarise la fonction économique $f(x)$ et les premiers nombres des contraintes $g_i(x) \geq 0, i = 1, 2, \dots, m$, en un point de linéarisation arbitraire $\overset{0}{x}$, puis on résout le programme linéaire ainsi obtenu. On recommence alors, en remplaçant $\overset{0}{x}$ par la solution optimale $\overset{1}{x}$ du programme linéaire. Et ainsi de suite.

En fait, un tel procédé ne converge pas, et peut donner lieu à des oscillations importantes, comme le montre l'exemple élémentaire suivant : le domaine est un trapèze (contraintes linéaires) et les équipotentielles de la fonction

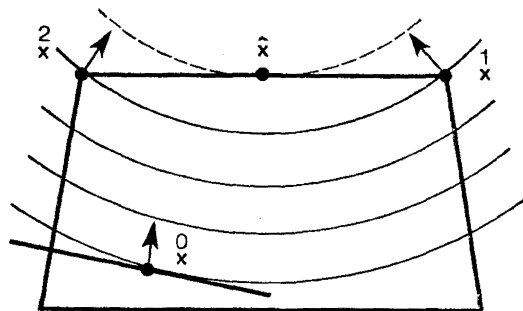


Figure 7

économique des cercles concentriques. En prenant $\overset{0}{x}$ pour point de départ, on obtient successivement $\overset{1}{x}, \overset{2}{x}$, puis alternativement ces deux derniers points indéfiniment.

4.1.1. Limitation des déplacements

Un remède simple est apporté par Griffith et Stewart [4.14] en 1961. Puisque les approximations linéaires des fonctions f et g_i peuvent devenir trop grossières quand on s'éloigne du point de linéarisation $\overset{k}{x}$, on limite l'utilisation du programme linéaire approché (P_k) correspondant à $\overset{k}{x}$, à un voisinage limité de $\overset{k}{x}$, par exemple un petit cube C_k de dimension ρ et centré en $\overset{k}{x}$. En d'autres termes, on ajoute aux contraintes de (P_k) la condition $x \in C_k$, qui s'exprime par des contraintes linéaires : le résultat est encore un programme linéaire. Lorsque l'algorithme s'arrête ou cycle, on diminue ρ , ce qui permet de continuer le processus.

Quand $\rho \rightarrow 0$, la suite des points $\overset{k}{x}$ converge vers $\hat{x}$. Il n'est pas donné de démonstration de convergence dans l'article cité.

4.1.2. Points de linéarisations multiples

Une autre possibilité pour limiter l'erreur apportée par la linéarisation, lorsqu'on s'éloigne trop du point de linéarisation est de multiplier ces points, en en mettant un peu partout : on tient compte simultanément de toutes les contraintes linéaires ainsi obtenues, et pour la valeur de la fonction économique, en un point x , on prend le minimum des valeurs en x des fonctions affines correspondant à ces points. (Ce qui revient à remplacer le graphe de f par une surface polyédrique, tangente aux points de linéarisation.)

Mais ce procédé devient vite prohibitif dans R^n dès que n dépasse quelques unités. Si l'on dispose les points de linéarisation y en une « grille » de R^n (ou plus exactement une portion de grille), le nombre de ces points est de k^n , k étant le nombre de valeurs discrètes retenues pour chaque composante de y . Et 2^{10} vaut déjà 1 024!

D'où l'intérêt du procédé proposé par Cheney et Golstein [4.6] en 1959, où l'on prend pour nouveau point de linéarisation la solution optimale du programme linéaire précédent (comme dans l'adaptation brutale de la méthode de Newton, envisagée au début de 4.1) mais en conservant tous les anciens points de linéarisation déjà créés au cours des itérations précédentes. En fait, l'article de Cheney et Goldstein est une théorie très complète de la convergence, et contient six algorithmes correspondant à différents types de problèmes. Nous donnons ci-dessous quelques précisions sur un schéma simplifié de l'un d'entre eux.

Le problème posé (P) étant : Maximiser $f(x)$ sous les conditions $g_i(x) \geq 0$, $i = 1, 2, \dots, m$, avec f et g_i concaves et continuellement différentiables, on représente chacune de ces fonctions par l'enveloppe inférieure de toutes les fonctions affines majorantes (propriété des fonctions concaves). En posant :

$$f'(x, y) = f(y) + \nabla f(y) \cdot (x - y)$$

on peut écrire $f(x) = \inf \{ f'(x, y) \mid y \in R^n \}$

De même, on a $g_i(x) = \inf \{ g'_i(x, y) \mid y \in R^n \}$, $\forall i = 1, 2, \dots, m$

Étant donné un ensemble $C \subset R^n$, compact, contenant la solution optimale $\hat{x}$ de (P), on peut alors remplacer (P) par le programme linéaire infini (nombre infini de contraintes) (P') suivant :

Maximiser μ , sous les conditions :

$$f'(x, y) \geq \mu, \quad \forall y \in C$$

$$g'_i(x, y) \geq 0, \quad \forall y \in C, \forall i = 1, 2, \dots, m$$

(P')

(P'), qui est équivalent à (P), a pour variables μ et x , le point y jouant le rôle d'un indice des contraintes, au même titre que l'indice i .

La méthode consiste alors à remplacer (P') par un programme linéaire fini approché (P'_k) , obtenu en remplaçant C par un sous-ensemble fini C_k de C . Soit $C_k = \{y^1, y^2, \dots, y^k\} \subset C$

En désignant par y^{k+1} la solution optimale de (P'_k) , on agrandit C_k du point y^{k+1} , soit :

$$C_{k+1} = C_k \cup \{y^{k+1}\}$$

Et ainsi de suite. Quand $k \rightarrow +\infty$ et sous des hypothèses assez larges, la suite des y^k converge vers la solution optimale $\hat{x}$.

La même méthode est proposée un peu plus tard, en 1960, par Kelley [4.21], avec des détails pratiques : en remarquant que la solution optimale y^{k+1} du programme linéaire (P'_k) est non réalisable pour le programme linéaire suivant (P'_{k+1}) , l'auteur propose d'utiliser la méthode Duale-Simpliciale, bien adaptée à cette situation. Cette méthode est également proposée en 1963 par Hartley et Hocking [4.16], qui disposent leurs points de linéarisation suivant une grille. Envisageant une grille très fine, conduisant à un programme linéaire approché avec un très grand nombre de contraintes, ils traitent en réalité le dual, c'est-à-dire un programme linéaire dont le nombre de contraintes est normal, mais dont le nombre de variables est pratiquement infini (à chaque point de la grille correspond m variables du dual). C'est pourquoi dans la résolution de ce dual par la méthode simpliciale ils utilisent le procédé déjà classique du « principe de décomposition » de Dantzig et Wolfe [4.9], grâce auquel les colonnes hors-bases du tableau simplicial n'ont pas besoin d'être calculées.

L'algorithme n° V de Cheney et Golstein est généralisé au cas des espaces infinis (au lieu de R^n) par Descoux, en 1963.

En 1966, Kunzi [4.22] propose de résoudre les programmes linéaires d'approximation par la méthode Duoplex, décrite antérieurement dans [4.23], et particulièrement adaptée aux programmes linéaires à grands nombres de contraintes.

On retrouve en 1968, proposée par Kaplan [4.20], une méthode très semblable à celle de Cheney et Golstein décrite plus haut, avec une légère variante : le nouveau point de linéarisation, n'est pas pris égal à la solution optimale y^{k+1} de (P'_k) , mais plus près du domaine de (P) , plus précisément à l'intersection de la frontière de (P) avec le segment $[x^*, y^{k+1}]$, x^* étant un point fixe intérieur au domaine.

4.1.3. Méthode de Frank et Wolfe

Une méthode qui, en un certain sens, pourrait se caractériser par la limitation des déplacements et se ranger en 4.1.1, est proposée en 1956 par Frank et Wolfe [4.12] pour la résolution des programmes à fonction économique

quadratique et à contraintes linéaires. En fait, elle s'applique sans modification de principe aux fonctions économiques différentiables. Le problème posé étant :

$$\begin{array}{l} \text{Maximiser } f(x) \text{ sous les conditions} \\ Ax \geq a \end{array} \quad (P)$$

On considère la suite de points $\hat{x}^k$ et de programmes linéaires (P_k) correspondants :

$$\begin{array}{l} \text{Maximiser } g^k \cdot x \text{ sous les conditions} \\ Ax \geq a \end{array} \quad (P_k) \quad k = 1, 2, \dots$$

où $g^k = \nabla f(\hat{x}^k)$. En désignant par $\hat{z}^k$ la solution optimale de (P_k) , on détermine $\hat{x}^{k+1}$ par :

$$\hat{x}^{k+1} : f(\hat{x}^{k+1}) = \max \{ f(x) \mid x \in [\hat{x}^k, \hat{z}^k] \}$$

La suite des $\hat{x}^k$ converge vers la solution optimale $\hat{x}$ de (P)

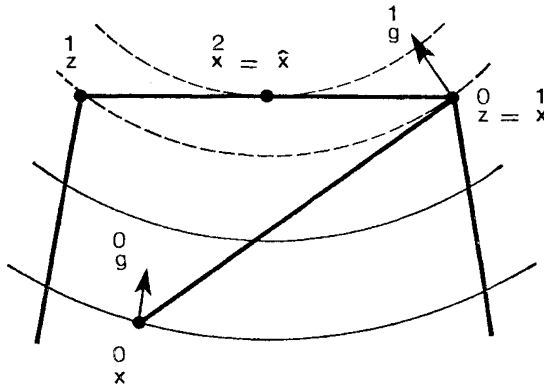


Figure 8

En appliquant cette méthode à l'exemple donné au début de 4.1, on voit que la solution optimale $\hat{x}$ est atteinte à la deuxième itération. Mais, en général, l'algorithme est infini.

On remarquera que si on applique cette méthode au cas particulier où le domaine est R^n tout entier, on retrouve la méthode du gradient (ou de plus forte pente) pour la maximisation d'une fonction dans R^n . On sait que la

convergence de cette méthode est très lente si on n'ajoute pas des procédés d'accélération comme la diagonalisation.

On peut penser que le même problème doit se présenter dans le cas normal avec contraintes. En 1968, la vitesse de convergence est étudiée par Canon et Cullum [4.4].

Une question naturelle se pose : peut-on adapter cette méthode aux cas des contraintes non linéaires, en linéarisant les contraintes comme il en est fait de la fonction économique et en déterminant $\hat{x}^{k+1}$ non plus sur le segment $[\hat{x}^k, \hat{z}^k]$, mais sur sa partie réalisable? La réponse est négative, comme on peut le vérifier sur l'exemple suivant :

Maximiser $x_1 + x_2$ sous les conditions :

$$4x_1^2 + 4x_2^2 \leq 1$$

$$-1 \leq x_1 \leq 1$$

$$-1 \leq x_2 \leq 1$$

En prenant $\hat{x}^0 = (0, -1/2)$, on obtient $\hat{x}^1 = \hat{x}^2 = \hat{x}^3 = \dots$ sans atteindre $\hat{x} = \left(\frac{1}{2\sqrt{2}}, \frac{1}{2\sqrt{2}} \right)$

La méthode de Franck et Wolfe est généralisée plus tard par de nombreux auteurs à des espaces plus généraux que R^n , pour des problèmes de contrôle optimal. Par exemple Valadier [4.26] en 1965, Poljak [4.24], Hougazeau [4.17], Gilbert [4.13] en 1966, Dem'Yanov et Rubinov [4.10] en 1967, Canon et Cullum [4.4], Cea [4.5], Auslender et Brodeau [4.1] en 1968...

4.1.4. Méthode des centres linéarisés [4.18]

Ainsi qu'il avait été signalé en 2.2, dans le cadre des méthodes intérieures, la linéarisation de la méthode des centres au niveau du calcul des centres conduit à résoudre une suite de programmes linéaires (P_k) , dont les solutions optimales $\hat{z}^k$ permettent de construire une suite de points $\hat{x}^{k+1}$ convergeant vers la solution optimale $\hat{x}$ du problème non linéaire (P) . Ces programmes linéaires, dont la taille est constante, ont une contrainte et une variable de plus que (P) . La convergence est ici assurée, malgré le nombre constant des contraintes linéaires des problèmes (P_k) , du fait que la linéarisation est utilisée avec une méthode intérieure, pour calculer une solution centrale, et non une solution sur la frontière. Ce qui élimine en quelque sorte les oscillations ou basculement des solutions courantes obtenues par une linéarisation directe de (P) , ces solutions étant des solutions optimales de programmes linéaires, et par suite toujours localisés en des points extrêmes du polyèdre.

Pour exposer le schéma de la méthode des centres linéarisée, il est commode de supposer tout d'abord que le problème posé (P) est linéaire. Soit :

Maximiser $f \cdot x$ sous les conditions $Ax \geq a$	(P)
--	-------

(P) est équivalent au programme linéaire (P') suivant, où l'on introduit une variable supplémentaire λ et une contrainte supplémentaire :

Maximiser λ sous les conditions $Ax \geq a$ $f \cdot x \geq \lambda$	(P')
--	--------

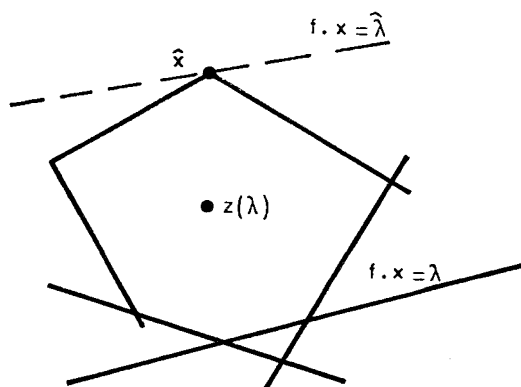


Figure 9

Pour une valeur fixée de λ , on obtient une solution « bien à l'intérieur » du domaine $\{x \mid Ax \geq a, f \cdot x \geq \lambda\}$ (un centre au sens de la méthode des centres) en résolvant :

$$\max_{x \in R^n} (\min \{ A_i x - a_i, f \cdot x - \lambda \mid i = 1, 2, \dots, m \})$$

ou ce qui revient au même, en résolvant le programme linéaire suivant, de variable (x, μ) , λ fixé :

Maximiser μ sous les conditions $A_i x - a_i \geq \mu, i = 1, 2, \dots, m$ $f \cdot x - \lambda \geq \mu$	$(P(\lambda))$
---	----------------

Soit $(z(\lambda), \bar{\mu}(\lambda))$ la solution optimale de $(P(\lambda))$ dépendant de la valeur choisie pour λ . Le point $z(\lambda)$ est le centre du tronçon correspondant et $\bar{\mu}(\lambda)$ représente la « valeur optimale » de $(P(\lambda))$. On voit que l'on peut résoudre (P') en faisant croître la valeur du paramètre λ du programme linéaire $(P(\lambda))$: $\bar{\mu}(\lambda)$ tend vers zéro quand λ tend vers $\lambda = f \cdot \hat{x}$. Ce qui peut se faire pratiquement, par exemple, en utilisant la méthode bien classique dite « méthode simpliciale avec paramétrisation du second membre ».

Dans le cas où le problème posé (P) est non linéaire, sous la forme :

$$\begin{array}{l} \text{Maximiser } f(x) \text{ sous les conditions} \\ g_i(x) \geq 0, i = 1, 2, \dots, m \end{array} \quad (P)$$

Le problème équivalent (P') s'écrit :

$$\begin{array}{l} \text{Maximiser } \lambda \text{ sous les conditions} \\ g_i(x) \geq 0, i = 1, 2, \dots, m \\ f(x) \geq \lambda \end{array} \quad (P')$$

On introduit alors les programmes linéaires $(P'(\lambda, y))$ dépendant à la fois de λ et du point de linéarisation choisi y :

$$\begin{array}{l} \text{Maximiser } \mu \text{ sous les conditions :} \\ g'_i(x, y) - \lambda \geq \mu, \quad i = 1, 2, \dots, m \\ f'(x, y) - \lambda \geq \mu \end{array} \quad (P'(\lambda, y))$$

où $f'(x, y)$ et $g'_i(x, y)$ sont les fonctions linéarisant $f(x)$ et $g_i(x)$ au point y .

L'algorithme se déroule ainsi : étant donnée une solution $\bar{x}^k$ (réalisable pour (P)), on résout $(P'(\lambda_k, \bar{x}^k))$, avec $\lambda_k = f(\bar{x}^k)$. Soit $(\bar{z}^k, \bar{\mu}^k)$, la solution optimale de ce programme linéaire. On détermine $\bar{x}^{k+1}$ en résolvant :

$$\bar{x}^{k+1} : d_k(\bar{x}^{k+1}) = \max \{ d_k(x) \mid x \in [\bar{x}^k, \bar{z}^k] \}$$

en posant $d_k(x) = \min \{ f(x) - \lambda_k, g_i(x) \mid i = 1, 2, \dots, m \}$.

Et ainsi de suite. Sous des hypothèses assez générales, les $\bar{x}^k$ convergent vers $\hat{x}$. Cette convergence est assurée du fait que cet algorithme est une méthode de centres. On montre en effet que les $\bar{x}^k$ sont des centres approchés à ϵ_k près, et que la suite des ϵ_k tend vers zéro quand $k \rightarrow +\infty$.

On notera une analogie de procédé entre cette méthode et celles de Zoutendijk (directions réalisables), de Franck et Wolfe, et de Rosen (gradient projeté) : entre chaque linéarisation, conduisant à la résolution d'un programme linéaire, on maximise une fonction non linéaire sur un segment (maximisation à une variable). En fait, la méthode des centres linéarisée peut être présentée sous une forme plus générale, permettant de retrouver ces méthodes comme cas particulier [4.19]. De ce point de vue, la méthode des directions réalisables et celle des centres linéarisés apparaissent comme deux variantes d'une même méthode, ne se différenciant que par la valeur choisie pour un paramètre scalaire ; il s'agit du paramètre ϵ défini en 2.3.

4.2. Méthodes barycentriques

Une voie différente des méthodes tangentielles consiste à approcher les graphes des fonctions considérées supposées concaves, par des surfaces polyédriques inscrites (c'est-à-dire, dont les sommets sont situés sur le graphe). Étant donnée une fonction concave $h(x)$, et un ensemble fini de points de R^n , $X^1, X^2, \dots, X^k$, appelés générateurs, les sommets du graphe approché (défini dans R^{n+1}) auront pour coordonnées

$$Z^1 = (X^1, h(X^1)), Z^2 = (X^2, h(X^2)), \dots, Z^k = (X^k, h(X^k)).$$

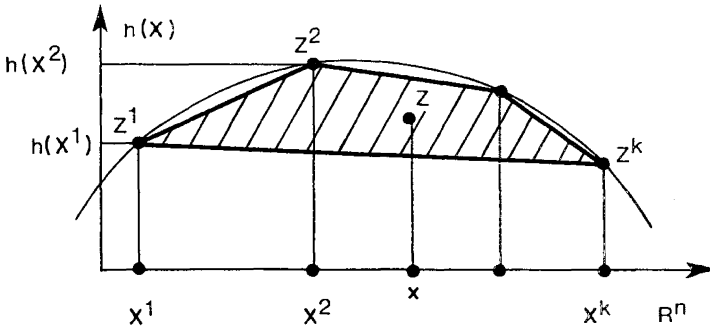


Figure 10

Un point x étant considéré comme un barycentre des générateurs X^j c'est-à-dire s'exprimant sous la forme :

$$x = \sum_{j=1}^k \lambda_j X_j \quad \text{avec} \quad \sum_{j=1}^k \lambda_j = 1, \lambda_j \geq 0, \quad j = 1, 2, \dots, k$$

on considère le point $z \in R^{n+1}$ qui lui est associé par :

$$z = (x, y), \quad \text{avec} \quad y = \sum_{j=1}^k \lambda_j h(X^j)$$

C'est le barycentre des sommets Z^j du graphe approché, correspond aux mêmes coefficients λ_j ayant servi à la définition de x . On a toujours $y \leq h(x)$. L'ensemble de tous les points z que l'on peut obtenir de cette manière est un polyèdre, enveloppe convexe des k points Z^j .

Ainsi on peut remplacer une fonction $h(x)$ par une fonction linéaire $h'(\lambda)$, de variables λ_j , $j = 1, 2, \dots, k$, en ajoutant les conditions linéaires $\sum \lambda_j = 1$ et $\lambda \geq 0$. Que donne un tel changement de variables pour un programme mathématique? Chaque contrainte $g_i(x) \geq 0$ est remplacée par $g'_i(\lambda) \geq 0$, et l'on doit ajouter les conditions précédentes sur les λ_j . Les fonctions g_i étant concaves, il en résulte d'après une propriété classique des fonctions concaves que le nouveau domaine ainsi défini (dans R^n) est plus petit. Par ailleurs, la valeur $f(x)$ de la fonction économique est remplacé par une valeur inférieure ou égale $f'(\lambda)$. Dans le cas de la maximisation, le point z , associé à x dans la figure précédente, sera toujours situé à l'optimum sur le graphe polyédrique, et non à l'intérieur de l'enveloppe convexe des points Z^j .

Ce changement de variables conduit donc à remplacer le programme « concave » (P) donné par un programme linéaire (P') approché par défaut. En 1961, Hartley [4.15] donne une application de cette idée au cas particulier des fonctions à variables séparées, c'est-à-dire de la forme :

$$g_i(x) = \sum_{j=1}^n g_{ij}(x_j), \quad i = 1, 2, \dots, m$$

ce qui permet d'approcher les fonctions à une seule variable $g_{ij}(x_j)$ indépendamment les unes des autres, et par suite d'obtenir un programme linéaire (P') dont la taille peut-être acceptable : le nombre de variables de (P') est égale à $k \times m$, si m est le nombre de contraintes et k le nombre de générateurs utilisé pour chaque fonction $g_{ij}(x_j)$. Il n'en est plus de même si les fonctions $g_i(x)$ ne sont plus à variables séparées, car on est alors conduit à utiliser des générateurs formant une « grille » de R^n , comme pour le procédé de linéarisation de Hartley et Hocking en 4.1.2, ce qui est prohibitif.

C'est pourquoi Dantzig [4.7] propose en 1959 (antérieurement à l'article de Hartley cité plus haut), pour les problèmes à contraintes linéaires et à fonction concave, puis un peu plus tard pour les contraintes non linéaires concaves, sous le nom de méthode des colonnes [4.8], un procédé de détermination automatique des générateurs, devenant de mieux en mieux placés, si l'on peut dire, au cours des itérations. Le principe de cette méthode est une extension du « Principe de décomposition » des programmes linéaires, publié par Dantzig et Wolfe en 1958 [4.9]. On trouvera également la théorie de ces méthodes et d'une variante non linéaire utilisant les coordonnées barycentriques dans [4.3].

La méthode des colonnes de Dantzig peut se résumer ainsi :

Un certain nombre de générateurs

$$X^j \in R^n, \quad j \in K = \{1, 2, \dots, k\}$$

étant donnés, on remplace le problème donné (P) par le programme linéaire approché (P'_k) :

Maximiser $\sum_{j \in K} \lambda_j f(X^j)$ sous les conditions

$$\sum_{j \in K} \lambda_j g_i(X^j) \geq 0, \quad i = 1, 2, \dots, m$$

$$\sum_{j \in K} \lambda_j = 1$$

$$\lambda_j \geq 0, \quad j \in K$$

(P'_k)

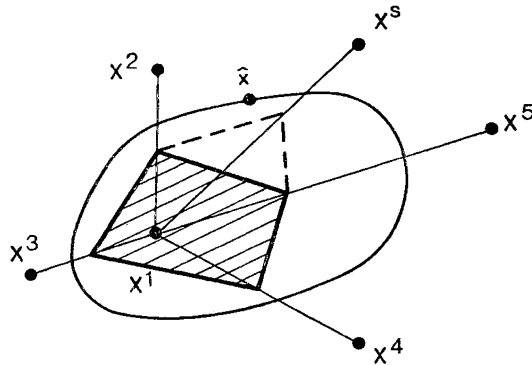


Figure 11

les $\lambda_j, j \in K$ étant les nouvelles variables telles que :

$$x = \sum_{j \in K} \lambda_j X^j$$

Le domaine D_k de (P'_k) est inclus dans le domaine de (P) . La résolution de (P'_k) , par la méthode simpliciale par exemple, fournit une solution optimale $\bar{\lambda}(K)$, correspondant à un point $\bar{x}(K)$, ainsi que la valeur des multiplicateurs de Kuhn et Tucker associés aux contraintes de (P'_k) d'indices $i = 1, 2, \dots, m$, soit le vecteur $\bar{u}(K)$. On résout alors le problème non linéaire, mais sans contraintes, suivant :

Maximiser $f(x) + \sum_{i \in K} \bar{u}^i(K) \cdot g_i(x)$

(Q_k)

par une méthode de gradient par exemple. Soit X^{k+1} la solution optimale de (Q_k) , supposée à distance finie. Cette solution fournit un nouveau générateur pour la définition du programme linéaire approché, et la matrice des contraintes de celui-ci augmente d'une « colonne » supplémentaire $\left[g(X^{k+1}) \right]$, d'où le nom de la méthode.

En définitive, on a à résoudre une suite (théoriquement infinie) de programmes linéaires, dont le nombre de variable augmente régulièrement d'une unité à chaque itération et dont les solutions, sous des hypothèses larges, convergent vers $\hat{x}$. Mais si on se contente, dans l'optimisation du problème auxiliaire (Q_k) , de rechercher la solution X^{k+1} uniquement sur l'ensemble fini des points d'une grille, comme dans la méthode de Hartley et Hocking, décrite en 4.1.2, la suite des (P'_k) est finie, et fournit une solution optimale approchée de (P) . En fait, la méthode des colonnes et celle de Cheney et Goldstein sont, en un certain sens, duales l'une de l'autre.

On pourra aussi comparer la méthode des colonnes à celle de Uzawa [5.6], décrite au chapitre suivant (méthodes Lagrangiennes) qui utilisent les mêmes problèmes auxiliaires (Q_k) , mais qui déterminent les valeurs des coefficients $\bar{u}^i(K)$ de façon différente. On peut dire que la méthode des colonnes linéarise la méthode de Uzawa au niveau des calculs des $\bar{u}^i(K)$.

5. METHODES LAGRANGIENNES

Dans le cas des programmes convexes, une méthode utilisant la fonction de Lagrange $\psi(x, u) = f(x) + \sum_{i=1}^m u^i g_i(x)$ a été introduite en 1958 par Uzawa [5.6]. La théorie repose essentiellement sur les notions de col (point-selle) et de dualité. Considérons le problème posé (P) et son dual (D) :

Maximiser $f(x)$ $g(x) \geq 0$	(P)	Minimiser $f(x) + u \cdot g(x)$ $\nabla f(x) + u \cdot \nabla g(x) = 0$ $u \geq 0$	(D)
-----------------------------------	-----	--	-----

où les vecteurs u et $g(x)$ sont indicés pareillement, et où $\nabla g(x)$ est une matrice dont l'élément (i, j) est la dérivée partielle de $g_i(x)$ par rapport à x_j . Les fonctions f et g_i sont supposées concaves et différentiables.

La méthode de Uzawa revient à résoudre (D) , dont la solution optimale $(\hat{x}, \hat{u})$ fournit, sous des hypothèses convenables (par exemple f strictement concave), la solution optimale $\hat{x}$ de (P) , de la manière suivante :

A toute valeur $u \geq 0$ fixée, on associe la solution x qui vérifie la contrainte $\nabla f(x) + u \nabla g(x) = 0$, ce qui peut s'obtenir pratiquement en résolvant le problème auxiliaire $Q(u)$:

$$\boxed{\begin{array}{l} \text{Maximiser } f(x) + u \cdot g(x) \\ x \in R^n \end{array}} \quad (Q(u) \quad (u \geq 0 \text{ fixé}))$$

L'équivalence du problème $Q(u)$ et de l'équation $\nabla f(x) + u \cdot \nabla g(x) = 0$ provient du fait que la fonction de Lagrange $\psi(x, u)$ est concave et différentiable par rapport à x . Soit $\bar{x}(u)$ la solution optimale de $Q(u)$, solution unique si $\psi(x, u)$ est strictement concave en x . (C'est le cas si f est strictement concave), que nous supposons à distance finie. Le problème (D) est donc équivalent au problème (D') , de variable u :

$$\boxed{\begin{array}{l} \text{Minimiser } \theta(u) \text{ sous les conditions :} \\ u \geq 0 \end{array}} \quad (D')$$

en posant $\theta(u) = \psi(\bar{x}(u), u)$. Par ailleurs, on montre, sous des hypothèses assez larges, que l'on a :

$$\nabla \theta(u) = g(\bar{x}(u))$$

On voit donc que, pour tout $u \geq 0$ fixé, on peut calculer les valeurs $\theta(u)$ et $\nabla \theta(u)$ en résolvant le problème auxiliaire (sans contraintes) $Q(u)$. On a alors la possibilité de résoudre (D') au moyen d'une méthode de gradient classique, adaptée au cas des variables bornées. La solution optimale $\hat{u}$ de (D') fournit $\bar{x}(\hat{u})$, et l'on a $\hat{x} = \bar{x}(\hat{u})$.

Cette méthode fut retrouvée ou étudiée par d'autres auteurs, en particulier Huard [5.4] en 1963, Falk et Thrall [5.3] en 1965, Falk [5.2], en 1967.

Une généralisation de la fonction de Lagrange est donnée par Roode [5.5] en 1968, formant une classe de fonctions qui, substituées à la fonction de Lagrange classique dans la méthode de Uzawa, donnent autant de variantes.

Si l'on se reporte aux méthodes de pénalisations exposées au chapitre 1, on voit que la méthode de Uzawa peut être associée à ces méthodes sur le plan du déroulement des calculs, avec ceci en plus : entre chaque résolution du problème auxiliaire sans contraintes $Q(u)$, la modification du paramètre u est déterminée automatiquement par une itération de (D') . En particulier, le calcul du $\nabla \theta(u)$ indique dans quelle direction doit être déplacé le point u . La fonction de Lagrange n'est pas une pénalisation au sens défini chapitre 1, mais cependant elle pénalise la non satisfaction des contraintes $g_i(x) \geq 0$, avec des coûts $u^i \geq 0$ différents pour chaque contrainte.

Il est clair que la théorie de la méthode de Uzawa s'accommode fort bien du problème plus compliqué suivant :

Maximiser $f(x)$ sous les contraintes $g(x) \geq 0$ $x \in C$	(P) avec $C \subset R^n$ convexe
---	------------------------------------

Dans ce cas, les problèmes auxiliaires $Q(u)$ deviennent

Maximiser $f(x) + u \cdot g(x)$ sous les conditions $x \in C$	$Q(u)$
--	--------

et ne sont plus des maximisations sans contraintes.

En 1967, Butz [5.1] modifie la méthode de Uzawa en remplaçant le terme $u \cdot g(x)$ de la fonction de Lagrange par $\lambda \cdot h(x)$, où λ est un scalaire,

$$h(x) = \sum_{i=1}^m p_i s_i(x) g_i(x)$$

avec $s_i(x) = 1$ si $g_i(x) \leq 0$, $s_i(x) = 0$ sinon, les p_i étant des constantes ou « poids ». La théorie de la méthode reste valable ici, et la maximisation de $\theta(u)$, avec $u \geq 0$ est remplacée par la maximisation d'une fonction d'une seule variable λ . On peut remarquer que $h(x)$ est une « vraie » pénalisation, selon la définition du chapitre 1, si l'on donne à λ une suite de valeurs λ_k croissant indéfiniment. Mais cette méthode de pénalisations est ici sans paramètre, avec modification automatique de la valeur de λ .

6. RECHERCHES PONCTUELLES OU PAR TRONCATURES

Lorsque l'on désire maximiser une fonction concave d'une seule variable, on peut utiliser une méthode qui, à défaut d'être rapide, est sûre, et s'adapte à toutes sortes de fonctions : c'est le procédé de recherche par points ou par dichotomies. On calcule les valeurs de la fonction f pour une suite de valeurs x^k croissantes de la variable, tant que les valeurs correspondantes de f sont croissantes, et l'on s'arrête au premier point x^k pour lequel $f(x^k) \leq f(x^{k-1})$. L'optimum $\hat{x}$ est alors localisé dans l'intervalle $[x^{k-2}, x^k]$. En calculant les valeurs de f pour $x^{k+1} \in]x^{k-2}, x^{k-1}[$ et pour $x^{k+2} \in]x^{k-1}, x^k[$ on diminue l'intervalle de confiance qui devient $[x^{k-2}, x^{k+1}]$, $[x^{k+1}, x^{k+2}]$ ou $[x^{k-2}, x^k]$ selon que le maximum des cinq

valeurs $f(\hat{x}^{k-2})$ à $f(\hat{x}^{k+2})$ est $f(\hat{x}^{k+1})$, $f(\hat{x}^{k-1})$ ou $f(\hat{x}^{k+2})$. Et ainsi de suite, à chaque étape les intervalles de confiance sont diminués. Par exemple, si les subdivisions se font aux milieux des segments, l'intervalle de confiance est divisé à chaque fois par deux. On peut aussi raffiner la méthode en utilisant des subdivisions basées sur les suites de Fibonacci.

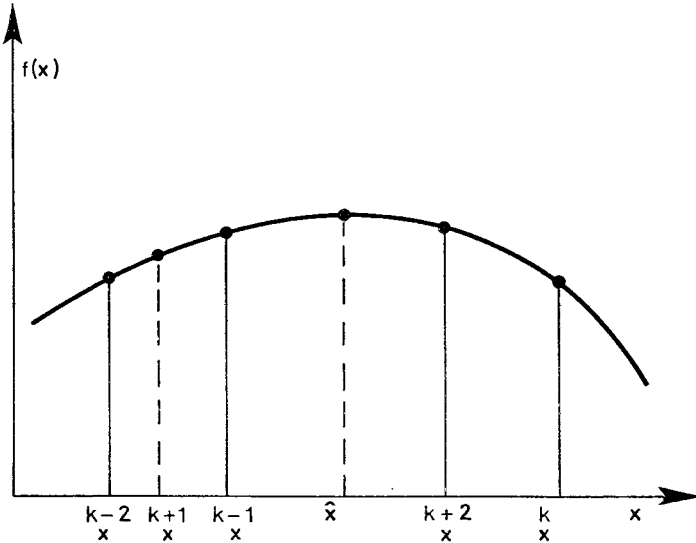


Figure 12

Si l'on cherche à étendre cette méthode au cas d'une fonction de plusieurs variables, deux voies naturelles assez différentes se présentent, selon que cette méthode est considérée comme une recherche par points, ou bien comme une suite de dichotomies ou troncatures, réduisant à chaque étape la portion d'espace (ici un segment) contenant la solution optimale.

6.1. Recherches par explorations ponctuelles

Étant donné un point $\hat{x}^k \in R^n$, on peut chercher un point $\hat{x}^{k+1}$, donnant, à la fonction f que l'on veut maximiser une valeur supérieure, parmi un ensemble fini de points « voisins » de $\hat{x}^k$, puis recommencer. Un tel procédé a l'avantage de n'utiliser que des valeurs de f et non de ses dérivées. C'est de la définition des points « voisins » que dépendra l'efficacité de ce procédé. Il n'est pas question, sans préciser davantage, de parler de méthode et surtout de convergence.

On peut penser choisir ces points voisins au hasard, selon une certaine loi de probabilité. Si celle loi est uniforme, le procédé ainsi obtenu est en moyenne le plus long. On peut lire à ce sujet les articles de Brooks [6.2] et

de Mc Arthur [6.5]. On peut donc considérer une telle recherche aléatoire comme une unité de mesure pour l'efficacité des méthodes ponctuelles.

Si l'on veut utiliser les données du problème pour orienter la recherche ponctuelle et obtenir ainsi un procédé plus évolué, on peut par exemple tenir compte des points déjà examinés, afin d'effectuer une approximation, fut-elle grossière, du gradient, ou plus généralement d'une direction de déplacement permettant d'améliorer $f(x)$. En 1959, Hookes et Jeeves [6.3] proposent un tel procédé, où ils alternent une exploration systématique suivant les axes de coordonnées, pris les uns après les autres (le pas de déplacement ρ étant fixé), avec un déplacement défini par la résultante des déplacements précédents. En fait, ce schéma doit être augmenté d'un certain nombre de contrôles, dont résulte le choix des explorations ultérieures. Lorsque le processus se bloque, on peut diminuer le pas de déplacement ρ . Cette méthode est appliquée par Wood [6.13] en 1960.

On pourrait également évoquer ici d'autres méthodes comme celle de Rosenbrock citée au chapitre 3, qui n'utilisent pas apparemment les dérivées de la fonction f et ne maximisent pas (de façon continue) sur les droites de déplacement. En fait, ces méthodes calculent de façon approchée, sous forme de différences finies, les dérivées premières et souvent secondes de f . Nous n'en parlerons donc pas ici.

En 1960, Wegner [6.12] choisit pour points d'exploration les sommets du polyèdre (non linéaire) des solutions réalisables. La technique de calcul est évidemment plus complexe. Une fois déterminée la facette non linéaire contenant la solution optimale $\hat{x}$, au moyen de ses sommets, il est nécessaire d'interpoler entre ces points en utilisant une surface du second degré.

Dans la méthode de Wegner, la détermination d'un sommet consiste à résoudre un système d'équations non linéaires, représentant les contraintes passant par ce sommet. Ce calcul serait fort cher si les sommets n'étaient calculés de proche en proche : quand on résout un système non linéaire, on connaît la solution d'un système « voisin », c'est-à-dire ne différant que par une équation, comme pour la méthode simpliciale dans le cas des programmes linéaires. La méthode de Wegner peut être considérée aussi comme une adaptation par linéarisation tangentielle du procédé d'exploration de la méthode simpliciale.

Une méthode d'exploration ponctuelle pour les problèmes de maximisation sans contraintes, de technique différente de celle de Hookes et Jeeves, est proposée en 1962 par Spendley, Heat et Himsworth [6.11]. Baptisée méthode du « Simplexe » (ne pas confondre avec méthode Simplex ou Simpliciale), elle consiste à explorer l'espace R^n aux sommets d'une famille de simplexes réguliers élémentaires, de dimension n , donc à $n + 1$ sommets chacun. (Pour une famille de cubes à n dimensions, il faudrait envisager 2^n sommets par cube).

Puis en 1965, Nelder et Mead [6.7] modifient cette méthode en utilisant des simplexes irréguliers, dont la forme est adaptée à la configuration des équipotentielles de f . En désignant par $\bar{x}^1, \bar{x}^2, \dots, \bar{x}^{n+1}$, les sommets d'un simplexe, supposés classés par ordre des valeurs croissantes de $f(x)$, on détermine $\bar{x}^{n+2}$ sur la droite reliant $\bar{x}^1$ (le plus mauvais point) au barycentre c des n autres points, selon la relation :

$$\bar{x}^{n+2} - \bar{x}^1 = k(c - \bar{x}^1)$$

la valeur du coefficient k étant choisie parmi des valeurs simples, par exemple : 0,5; 1,5; 2, 3, selon le classement de $f(\bar{x}^{n+2})$ parmi les valeurs $f(\bar{x}^1), f(\bar{x}^2), \dots, f(\bar{x}^{n+1})$. On obtient ainsi différents cas de modifications du simplexe initial, en remplaçant $\bar{x}^1$ par $\bar{x}^{n+2}$, et correspondant aux figures ci-dessous :

La convergence d'une telle méthode n'est évidemment pas assurée mais son intérêt réside dans sa simplicité pour dégrossir rapidement un problème. Le critère d'arrêt proposé par Nelder et Mead repose sur la valeur de l'écart quadratique moyen des valeurs de f aux sommets du simplexe.

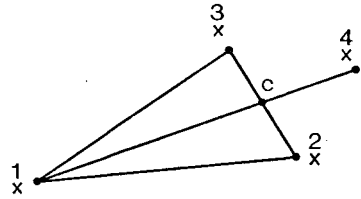


Figure 13

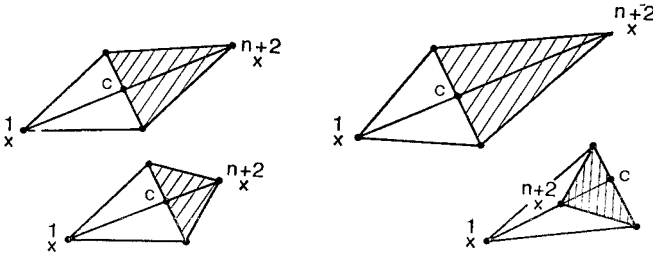


Figure 14

Cette méthode est reprise en 1968 par Spendley [6.10] dans le cas particulier d'une fonction quadratique, pour laquelle la comparaison de ses valeurs en trois points $\bar{x}^1, c$ et $\bar{x}^{n+2}$ se prête bien aux formules d'interpolation.

L'adaptation directe de la méthode aux problèmes avec contraintes est *a priori* possible, en aplatisant les simplexes de telle sorte que les sommets nouvellement créés, violant les contraintes, soient ramenés sur la frontière. Ce procédé qui avait été proposé par Spendley a l'inconvénient d'aplatir les simplexes sans pouvoir les « regonfler » ensuite : le procédé tend à la longue à n'explorer que des variétés linéaires de R^n . C'est pourquoi Box [6.1] propose en 1965 de remplacer les simplexes par des ensembles de points plus nom-

breux ($2n$ au lieu de $n + 1$), qu'il nomme des « complexes ». Une procédure de construction du complexe de départ est donnée par Mitchell et Kaplan [6.6] en 1968.

6.2. Recherches par troncatures

L'idée d'explorer le domaine des solutions réalisables, afin d'y localiser le maximum de la fonction économique, en partitionnant ce domaine en sous-ensembles de plus en plus petits n'a pas produit autant de littérature que les recherches ponctuelles.

Newman [6.8] propose en 1960 pour la maximisation sans contraintes de fonctions quasi-concaves (dont les équipotentielles sont enveloppes de leurs plans tangents) une méthode par dichotomie : la solution optimale $\hat{x}$ étant localisée dans un pavé $A \subset R^n$, on prend un point de départ quelconque $\overset{0}{x} \in A$, et on considère le demi-espace D_0 défini par :

$$D_0 = \{ x \mid \nabla f(\overset{0}{x}) \cdot (x - \overset{0}{x}) \geq 0 \}$$

On détermine alors le centre de gravité $\overset{1}{x}$ de $D_0 \cap A$, et on recommence avec $\overset{1}{x}$ au lieu de $\overset{0}{x}$, et $D_0 \cap A$ au lieu de A . D'une façon générale, étant donnés des points $\overset{0}{x}, \overset{1}{x}, \dots, \overset{k}{x}$ déjà trouvés, et les demi-espaces correspondants $D_0, D_1, \dots, D_k$, on prend pour $\overset{k+1}{x}$, le centre de gravité de

$$D_0 \cap D_1 \cap \dots \cap D_k \cap A,$$

et pour D_{k+1} le demi-espace défini par $\nabla f(\overset{k+1}{x}) \cdot (x - \overset{k+1}{x}) \geq 0$.

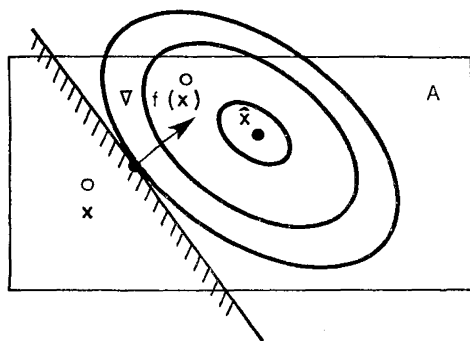


Figure 15

Cette méthode est proposée de nouveau en 1965 par Levin [6.4], et adaptée aux problèmes avec contraintes linéaires : cette adaptation ne pose évidemment pas de difficultés de principe, car le pavé initial A de Newman est vite transformé en un polyèdre quelconque par les dichotomies successives.

Mais la recherche du centre de gravité d'un polyèdre à n dimensions est pratiquement impossible, dès que n dépasse quelques unités. On peut remarquer toutefois que la qualité de centre de gravité n'est pas nécessaire à la méthode et qu'il suffit que les points $\bar{x}^k$ qu'elle génère soient des centres ou des ε -centres au sens de la méthode des centres, décrite en 2.2. En effet, on obtient un tel centre pour un polyèdre défini par un système d'inégalités linéaires $A_i x \geq a_i$, $i = 1, 2, \dots, m$, en résolvant le programme linéaire suivant :

Maximiser μ sous les conditions :

$$A_i x - \mu \geq a_i, \quad i = 1, 2, \dots, m$$

où μ est une variable supplémentaire, ce qui revient en fait à maximiser sur le polyèdre la F -distance $d(x) = \min \{ A_i x - a_i \mid i = 1, 2, \dots, m \}$.

Pour adapter une telle méthode à la présence de contraintes non linéaires, on pourrait s'inspirer du procédé utilisé dans la méthode des centres linéarisés.

Enfin, une méthode très générale de recherche d'un optimum par partitions successives du domaine des solutions réalisables est proposée vers 1960 aux USA sous le nom de « Branch and Bound », et en France par Roy et Berthier [6.9] sous le nom de « Séparation et Évaluation Progressive » (S.E.P.), avec une orientation particulière vers les problèmes en variables discrètes ou combinatoires. Nous donnons un schéma simplifié de cette dernière version.

Soit A l'ensemble des solutions réalisables, sur lequel on veut maximiser $f(x)$. On suppose que pour toute partie $A_k \subset A$, on est en mesure d'effectuer les deux opérations suivantes :

- 1) Partition de A_k en r sous-ensembles, selon une règle bien déterminée.
- 2) Approximation par excès (optimiste) de la valeur de $\sup \{ f(x) \mid x \in A_k \}$. Soit v_k cette valeur approchée, qui devient exacte si A_k est suffisamment petit.

L'algorithme procède alors comme suit : on partitionne A en r sous-ensembles $A_1, A_2, \dots, A_r$, pour lesquels on calcule les « valuations » $v_1, v_2, \dots, v_r$. On détermine la plus grande de ces valeurs, soit v_j , et on applique de nouveau l'opération de partition sur A_j . Et ainsi de suite, le sous-ensemble partitionné étant toujours celui de plus forte valuation. Sous des hypothèses générales, on a localisé la solution optimale $\hat{x}$ dès que l'on a obtenu un sous-ensemble A_k suffisamment petit (de valuation exacte).

Ajoutons que pour chaque sous-ensemble A_k apparu, on peut déterminer également une évaluation w_k par défaut de $\sup \{ f(x) \mid x \in A_k \}$, en déterminant par exemple une solution réalisable quelconque de A_k . Si pour une partie A_k , la valeur v_k est inférieure à la valeur w_k , d'une autre partie $A_{k'}$, on peut éliminer définitivement A_k pour la recherche ultérieure de l'optimum. Ce procédé d'élagage est important pour l'efficacité de la méthode, car, s'il n'était pas utilisé, cette efficacité serait *a priori* mauvaise. En effet, considérons un ensemble A de solutions réalisables, supposé fini pour simplifier. On peut

alors mesurer l'efficacité de la méthode par le rapport du nombre total de solutions de A au nombre de valuations effectuées (la valeur 1 correspondant à l'efficacité minimum, c'est-à-dire au dénombrement complet de tous les éléments de A). La méthode sera d'autant plus efficace que les partitions se feront systématiquement sur les plus petits des sous-ensembles déjà obtenus, afin d'atteindre le plus tôt possible un sous-ensemble de taille minimum, critère d'arrêt des calculs. Or, si à une étape donnée, on est en présence de k parties $A_1, A_2, \dots, A_k$, de valeurs $v_1 \geq v_2 \dots \geq v_k$, la partition sur A_1 fait disparaître A_1 , et le remplace par des sous-ensembles plus petits

$$A_{k+1}, A_{k+2}, \dots, A_{k+r},$$

de valeurs $v_{k+1}, v_{k+2}, \dots, v_{k+r}$. En faisant l'hypothèse plausible que la valeur approchée par excès de $\sup \{f(x) \mid x \in A_k\}$ est d'autant plus précise, donc plus petite, que A_k est petit (à la limite si A_k est de taille minimum, cette valuation est exacte), il faut s'attendre au nouveau classement probable suivant :

$$\max \{v_{k+1}, v_{k+2}, \dots, v_{k+r}\} < v_1$$

et par suite :

$$\max \{v_{k+1}, v_{k+2}, \dots, v_{k+r}\} \geq v_2 \geq \dots \geq v_k$$

La nouvelle partition a donc de fortes chances de ne pas s'appliquer à l'un des descendants de A_1 , et par suite d'homogénéiser la taille des sous-ensembles obtenus.

Cette méthode a cependant permis de résoudre des problèmes réels de natures diverses et de tailles importantes, principalement des problèmes en variables discrètes : grâce à une bonne connaissance de la nature concrète du problème traité, des critères d'élégage (calcul des w_k) suffisamment efficaces ont pu être trouvés.

On peut remarquer que la méthode des centres, exposée, en 2.2 entre dans ce schéma. Ici, la partition est une dichotomie, les valeurs des w_k sont exactes, les valeurs des v_k sont prises égales à $+\infty$ pour les ensembles nouvellement créés, et calculées exactement pour les autres, ce qui donne un choix du sous-ensemble à partitionner et un élégage efficaces.

TOURS D'HORIZON

- [1] DAVIES (D.), *Some practical methods of optimization*. Rapport Imperial Chemical. Industr. avril 1969.
- [2] DAVIES (D.) et SWANN (W. H.), *Review of constrained optimization*. Proceed. Confer. on optimization Keele, 1968 (Fletcher éd.), Academic Press, London.
- [3] DORN (W. S.), *Nonlinear programming, a survey*. IBM Research Paper n° RC 707, juin 1962.
- [4] FLETCHER (R.), *A class of methods for nonlinear programming with termination and convergence properties*. Rapport interne, Mathem. Branch, Theoretical Physics Division, mai 1969.

- [5] SPANG (H. A.), *A review of minimization techniques for nonlinear functions*. SIAM Review 4 (4), 1962, p. 343-365.
- [6] WOOD (C. E.), *Review of design optimization techniques*. IREE Transactions on systems Science and Cybernetics SSC-1, n° 1, 1965.
- [7] ZOUTENDIJK (G.), *Nonlinear programming : a numerical survey*. SIAM Control 4 (1), 1966, p. 194-210.

OUVRAGES, THESES

- [1] ARROW, HURWICZ et UZAWA, *Studies in linear and nonlinear programming*. Stanford University Press, 1958.
- [2] AUSLENDER (A.), *Méthodes numériques pour la résolution des problèmes d'optimisation avec contraintes*. Thèse de Doctorat Université de Grenoble juin 1969.
- [3] BERGE (C.) et GOUILHA-LOURI (A.), *Programmes, jeux et réseaux de transport*. Dunod, 1962.
- [4] BROISE (P.), HUARD (P.) et SENTENAC (J.), *Décomposition des programmes mathématiques*. Monographie de R.O. n° 6, A.F.I.R.O., Dunod, 1967.
- [5] CEA (J.), *Méthodes d'optimisation*. Cours de l'École d'Été d'Analyse Numérique CEA, EDF, 1969.
- [6] DANIEL (J. W.), *Theory and methods for the approximate minimization of functionals*. Cours École d'Été d'Analyse Numérique, CEA, IRIA, EDF, 1969.
- [7] DANTZIG (G. B.), *Linear programming and extensions*. Princeton University Press, 1963.
- [8] FIACCO (A. V.), *Sequential unconstrained minimization methods for nonlinear programming*. Thèse Northwestern University (Evanston, Illinois), juin 1967.
- [9] FIACCO (A. V.) et McCORMICK (G. P.), *Nonlinear programming : Sequential unconstrained minimization technique*. Wiley, 1968.
- [10] GOLDSTEIN (A. A.), *Constructive real analysis*. Harper, 1967.
- [11] GUIGNARD (M.), *Conditions d'optimalité et dualité en programmation mathématique*. Thèse Université de Lille, juin 1967.
- [12] KUNZI (H. P.) et EITTL (W.), *Nichtlineare Optimierung : Neuere Verfahren Bibliographie*. Springer Verlag, 1969.
- [13] LOOTSMA (F. A.), *Boundary properties of penalty functions for constrained minimization*. Thèse Université de Eindhoven, 1970.
- [14] MANGASARIAN (O. L.), *Nonlinear programming*. McGraw Hill, 1969.
- [15] POLAK (E.), *Computational methods in discrete optimal control and nonlinear programming : a unified approach*. Cours College of Engineering, University of California, Berkeley, 1969. Publié chez Academic Press (1970).
- [16] ROODE (J. D.), *Generalized Lagrangian functions in mathematical programming*. Thèse Université de Leiden, 1968.

1. METHODES DE PENALISATIONS

- [1] ABLOW (C. M.) et BRIGHAM (G.), « An analog solution of programming problems », *Oper. Research.*, 3 (4), 1955, p. 388-394.
- [2] BUTLER (T.) et MARTIN (A. V.), « On a method of Courant for minimizing functionals », *J. Math. and Physics*, 41, 1962, p. 291-299.
- [3] CAMP (G. D.), « Inequality. Constrained stationary value problems », *Oper. Res.*, 3 (4), 1955, p. 548-550.

- [4] COURANT (R.), « Variational methods for the solution of problems of equilibrium and vibrations », *Bull. Amer. Math. Soc.*, **49**, 1943, p. 1-23.
- [5] EREMIN (I. I.), « The penalty method in convex programming », *Soviet Math.*, **8** (2), 1967, p. 459-462.
- [6] FIACCO (A. V.), « Sequential unconstrained minimization methods for nonlinear Programs », *Thèse Northwestern University*, Evanston, Illinois, juin 1967.
- [7] MOTZKIN (T. S.), « New Techniques for linear inequalities and optimization ». *Symposium on linear inequalities and Programming*. U.S. Air Force Washington D.C., n° 10, 1952.
- [8] PIETRZYKOWSKI (T.), « Application of steepest descent method to concave programming », *Proceed. IFIP Congress*, Munich, 1962, p. 185-189.
- [9] POLAK (E.), « Computational methods in discrete optimal control and nonlinear programming : a unified approach ». *Electronics Research Laboratory, College of Engineering, University of California*. Memor. N° ERL-M261, février 1969. Édité en 1970 par Academic Press.
- [10] ROODE (J. D.), « Generalized Lagrangian functions in mathematical programming ». *Thèse Université de Leiden*, 1968.
- [11] RUBIN (H.) et UNGAR (P.), « Motion under a strong constraining force », *Comm. Pure Appl. Math.*, **10**, 1957, p. 65-87.
- [12] ZANGWILL (W. I.), « Nonlinear programming via penalty functions », *Management Sciences*, **13** (5), 1967, p. 344-358.

2. METHODES INTERIEURES

- [1] BOX (M. J.), « A new Method for constrained optimization and a comparison with other methods », *The Computer Journal*, **8**, 1965, p. 42-52.
- [2] BUI TRONG LIEU et HUARD (P.), « La méthode des Centres dans un espace topologique », *Numerische Math.*, (8), 1966, p. 56-67.
- [3] CAROLL (C. W.), « The created response surface technique for optimizing nonlinear restrained systems », *Op. Res.*, **9** (2), 1961, p. 169-184.
- [4] DAVIES (D.) et SWANN (W. H.), « Review of constrained optimization ». *Proceed. of the Conference on optimization*, Keele, 1968, Fletcher éd., Academic Press, London.
- [5] DEM'JANOV (V. F.) et RUBINOV (A. M.), « On the problem of minimization of a smothth functional with convex constraints », *Soviet Math.*, **6** (1), 1968, p. 9-11.
- [6] FIACCO (A. V.), « Sequential unconstrained minimization methods for nonlinear programming ». *Northwestern University* (Evanston, Illinois), *Thèse*, juin 1967.
- [7] FIACCO (A. V.) et McCORMICK (G. P.), « Computational algorithm for the sequential unconstrained minimization technique for nonlinear programming », *Manag. Sc.*, **10** (4), 1964, p. 601-617.
FIACCO (A. V.) et McCORMICK (G. P.), « The sequential unconstrained minimization technique for nonlinear programming. A primal-dual method », *Manag. Sc.*, **10** (2), 1964, p. 360-366.
- [8] FIACCO (A. V.) et McCORMICK (G. P.), « Extensions of SUMT for nonlinear programming Equality constraints and extrapolation », *Manag. Sc.*, **12** (11), 1966, p. 816-828.
- [9] FLETCHER (R.) et McCANN (A. P.), « Acceleration techniques for nonlinear programming using Carroll's method with Davidon's method ». *Proceed of the Conference on optimization*, Keele, 1968, Fletcher éd., Academic Press, London.
- [10] FREHEL (J.), « Une méthode de programmation non linéaire ». *Rapport interne IBM France*, n° FF2-0061-0, juillet 1968.

- [11] FRISCH (R.), « The double gradient method », *Université d'Oslo*, 1955.
- [12] GOLDSTEIN (A. A.) et KRIPKE (B. R.), « Mathematical programming by minimizing differentiable functions », *Numerische Math.*, **6**, 1964, p. 47-48.
- [13] GREENSTADT (J. L.), « A ricocheting gradient method for nonlinear optimization », *S.I.A.M. Appl. Math.*, **14** (3), 1966, p. 429-445.
- [14] HUARD (P.), « Resolution of mathematical programming with nonlinear constraints by the method of Centres », *Dans Nonlinear Programming* (Éd. Abadie), p. 206-219. North-Holland Publishing Co, Amsterdam, 1967.
- [15] HUARD (P.), « Programmation mathématique convexe », *R.I.R.O.* (7), 1968, p. 43-59.
- [16] HUARD (P.), « Méthode des Centres et méthode des centres par majorations », *Bull. EDF*, Direction des Études et Recherches, série C, n° 2, 1970.
- [17] LOOTSMA (F. A.), « Logarithmic programming : a method of solving nonlinear programming problems », *Philips Res. Reports*, **22**, 1967, p. 329-344.
- [18] LOOTSMA (F. A.), « Extrapolation in logarithmic programming », *Philips Res. Reports*, **23**, 1968, p. 108-116.
- [19] LOOTSMA (F. A.), « Constrained Optimization via penalty functions », *Philips Res. Reports*, **23**, 1968, p. 408-423.
- [20] LOOTSMA (F. A.), « Constrained optimization via parameter-free penalty functions », *Philips Res. Reports*, **23**, 1968, p. 424-437.
- [21] PARISOT (G. R.), « Résolution numérique approchée du problème de programmation linéaire par application de la programmation logarithmique », *Rev. Fr. de R.O.*, n° 20, 1961, p. 227-258.
- [22] PEUCHOT (M.), « Recherches concernant la résolution des problèmes de programmation linéaire », *C.R.A.S.*, Paris, **250** (20), 1960, p. 3271-3273.
- [23] POLAK (E.), « Computational methods in discrete optimal control and nonlinear programming : a unified approach », *Electronics Research Laboratory, College of Engineering, University of California*. Memo n° ERL, M 261, février 1969. Édité en 1970 par Academic Press.
- [24] PONENTALE (T.), « A new method for solving conditioned maxima problems », *J. Math. Analysis and Appl.*, **10**, 1965, p. 216-220.
- [25] ROSENBROCK (H. H.), « An automatic method for finding the greatest or least value of a function », *The Computer Journal*, **3** (1960), p. 175-184.
- [26] STONG (R. E.), « A note on the sequential unconstrained minimization technique for nonlinear programming », *J. of Management Sc.*, **12** (1), 1965, p. 142-144.
- [27] TOPKIS (M.) et VEINOTT (A.), « On the convergence of some feasible directions algorithms for nonlinear programming », *S.I.A.M. Control*, **5** (2), 1967, p. 268-279.
- [28] TREMOLIERES (R.), « Méthode des Centres à troncatures variables », *Bulletin Direction Études et recherches*, Électricité de France, Série C, n° 2, 1968, p. 57-64.
- [29] WOOD (C. E.), « Review of design optimization techniques », *IEEE Transactions on Systems Science and Cybernetics*, vol. SSC-1, n° 1, 1965.
- [30] ZOUTENDIJK (G.), « Methods of feasible directions », *Elsevier Publ. Co*, Amsterdam, 1960.
- [31] ZOUTENDIJK (G.), « Nonlinear programming : a numerical survey », *S.I.A.M. Control*, **4** (1), 1966, p. 194-210.

3. METHODES DE FRONTIERE OU DE PROJECTION

- [1] ABRHAM (J.), « An approximate method for convex programming », *Econometrica*, **29** (4), 1961, p. 700-703.

- [2] BIGG (M. D.), « The minimization of a general function subject to a set of non-linear constraints », *Proceed. Camb. Phil. Soc.*, **59**, 1963, p. 523-530.
- [3] CARPENTIER (J.) et ABADIE (J.), « Généralisation de la méthode du gradient réduit de Wolfe au cas de contraintes non linéaires ». *Acte du 4^e Congrès Ifors-Boston*, 1966, p. 1041-1052.
- [4] DAVIDSON (W. C.), « Variance algorithm for minimization », *Computer Journ.*, **10** (4), 1968, p. 406-410.
- [5] DAVIES (D.), « The use of Davidon's method in nonlinear programming ». *I.C.I.-Manag. Services Report*, N° MSDH-68/110, 1968.
- [6] DAVIES (D.) et SWANN (W. H.), « Review of constrained optimization ». *Proceed. Conference on optimization* (Fletcher Ed.), London Academic Press, Keele, 1968.
- [7] FAURE (P.) et HUARD (P.), « Résolution de programmes mathématiques à fonction non linéaire par la méthode du gradient réduit », *Revue Fr. R.O.*, n° 36, 1965, p. 167-206.
- [8] GOLDFARB (D.), « Extensions of Davidon's variable metric method to maximization under linear inequality and equality constraints ». *Rapport interne*. Courant Institute for Mathematical Sciences, 1966.
- [9] GUIGOU (J.), « Présentation et utilisation du code G.R.G. ». *Note E.D.F.* n° HI/102 du 9 juin 1969.
- [10] KALFON (P.), RIBIERE (G.) et SOGNO (J. C.), « A method of feasible directions using projection operators ». *IFIP Congress*, Edimburgh, 1968.
- [11] MURTAGH (B. A.) et SARGENT (R. W. H.), « A constrained minimization method with quadratic convergence ». *Proceeding Conference on optimization Keele*, 1968 (Fletcher éd.), Academic Press.
- [12] RIBIERE (G.), « Notice d'utilisation de la procédure Gradient Réduit (méthode révisée) ». *Publication n° MMC/11.9.7* (1968) de l'Institut Blaise Pascal, Paris.
- [13] ROBERTS (S. M.) et LYVERS (H. I.), « The gradient method on process control », *Ind. Eng. Chem.*, **53**, 1961, p. 877.
- [14] ROSEN (J. B.), « The gradient projection method for nonlinear programming ». Part I. Linear constraints. *S.I.A.M. Journal*, **8** (1), 1960, p. 181-217.
- [15] ROSEN (J. B.), « The gradient projection method for nonlinear programming ». Part II. Nonlinear constraints. *S.I.A.M. Journal*, **9**, 1961, p.
- [16] ROSENBRACK (H. H.), « An automatic method for finding the greatest or least value of a function », *The Computer Journal*, **3** (3), 1960, p. 175.
- [17] SWANN (W. H.), « Report on the development of a new direct search method of optimization ». *ICI Ltd. Central Instrument Labor Research*, Note 64/3 (1964).
- [18] TANABE (Kunio), « An algorithm for the constrained maximisation in nonlinear programming ». *Rapport Institut de Statistique Mathématique*, Tokio, 1969.
- [19] WOLFE (P.), « The reduced gradient method ». *Rand Document*, juin 1962.
- [20] WOLFE (P.), « On the convergence of gradient methods under constraints, *IBM Research, Zürich, Report N° RZ 204*, 1966.

4. METHODES DE LINEARISATION

- [1] AUSLENDER (A.) et BRODEAU (F.), « Convergence d'un algorithme de Frank-Wolfe appliqué à un problème de contrôle », *R.I.R.O.*, n° 7, 1965, p. 3-12.
- [2] BAUMOL (W. J.) et BUSHNELL (R. C.), « Error produced by linearization in mathematical programming », *Econometrica*, **35** (3), 1967, p. 444-471.
- [3] BROISE (P.), HUARD (P.) et SENTENAC (J.), « Décomposition des programmes mathématiques ». *Monographie de RO AFIRO*, n° 6, Dunod, 1967.

- [4] CANON (M. D.) et CULLUM (C. D.), « A tight upper bound on the rate of convergence of the Frank-Wolfe algorithm », *S.I.A.M. control*, **6** (4), 1968, p. 509-516.
- [5] CEA (J.), « Les méthodes de descentes en théorie de l'optimisation », *R.I.R.O.*, n° 13, 1968, p. 79-101.
- [6] CHENEY (E. W.) et GOLDSTEIN (A. A.), « Newton's method for convex programming and Tchebycheff approximation », *Numerische Math.*, **1** (5), 1959, p. 253-268.
- [7] DANTZIG (G. B.), « General Convex objective forms ». *The Rand Corpor.*, Rapport n° P. 1664, avril 1959.
- [8] DANTZIG (G. B.), « Linear programming and extensions ». *Princeton University Press*, Princeton, 1963.
- [9] DANTZIG (G. B.) et WOLFE (P.), « Principle of decomposition for linear programs », *Managt. Sc.*, **8** (1), 1960, p. 101-111.
- [10] DEM'YANOV (V. F.) et RUBINOV (A. M.), « The minimization of a smooth convex functional on a convex set », *S.I.A.M. control*, **5** (2), 1967, p. 280-294.
- [11] DESCLOUX (J.), « Note on convex programming », *S.I.A.M. Journ.*, **11** (3), 1963, p. 737-747.
- [12] FRANK (M.) et WOLFE (P.), « An algorithm for quadratic programming », *Naval Res. Logist. quarterly*, **3** (1-2), 1956, p. 95-120.
- [13] GILBERT (E. G.), « An iterative procedure for computing the minimum of a quadratic form on a convex set », *S.I.A.M. control*, **4** (1), 1966, p. 61-80.
- [14] GRIFFITH (R. E.) et STEWART (R. A.), « A nonlinear programming technique for the optimization of continuous processing systems », *Management Sc.*, **7** (4), 1961, p. 379-392.
- [15] HARTLEY (H. O.), « Nonlinear programming by the simplex method », *Econometrica*, **29** (2), 1961, p. 223-237.
- [16] HARTLEY (H. O.) et HOCKING (R. R.), « Convex programming by tangential approximation », *Managt. Sc.*, **9** (4), 1963, p. 600-612.
- [17] HOUGAZEAU (Y.), « Sur la minimisation des formes quadratiques avec contraintes ». C.R.A.S., Paris, 2 nov. 1966.
- [18] HUARD (P.), « Programmation mathématique convexe », *R.I.R.O.*, n° 7, 1968, p. 43-59.
- [19] HUARD (P.), « Méthode des Centres et méthode des centres par majorations ». *Bull. EDF, Direction des Ét. et Rech.*, série C, n° 2, 1970.
- [20] KAPLAN (A. A.), « Determination of the extremum of a linear function on a convex set », *Soviet Math.*, **9** (1), 1968, p. 269-271.
- [21] KELLEY (J. E.), « The cutting-plane method for solving convex programs », *S.I.A.M. Journal*, **8**, 1960, p. 703-712.
- [22] KUNZI (H. P.), « The duoplex method in nonlinear programming », *S.I.A.M. Control*, **4** (1), 1966, p. 130-138.
- [23] KUNZI (H. P.) et TZSCHACH (H.), « The duoplex algorithm », *Numerische Math.*, **7** (3), 1965, p. 222-225.
- [24] POLJAK (B. T.), « Existence theorems and convergence of minimizing sequences in extremum problems with restrictions », *Soviet Math.*, **7** (1), 1966, p. 72-75.
- [25] STIEFEL (E.), « Über diskrete und lineare Tschebyscheff-Approximationen », *Numerische Math.*, **1** (1), 1959, p. 1-28.
- [26] VALADIER (M.), « Extension d'un algorithme de Frank-Wolfe », *R.I.R.O.*, n° 36, 1965, p. 251-253.
- [27] WOLFE (P.), « Programming with nonlinear constraints. Preliminary report », *Notices Amer. Math. Soc.*, **5**, 1958, p. 508. Abstract 548-102.

5. METHODES « LAGRANGIENNES »

- [1] BUTZ (A. R.), « Iterative saddle-points techniques », *S.I.A.M. Appl. Math.*, **15** (3), 1967, p. 719-725.
- [2] FALK (J. E.), « Lagrange multiplier and nonlinear programming », *J. Math. Anal. Appl.*, **19**, 1967, p. 141-189.
- [3] FALK (J. E.) et THRALL (R. M.), « A constrained Lagrangian approach to nonlinear programming ». *Rapport University of Michigan*, mars 1965.
- [4] HUARD (P.), « Convex programming. Dual algorithm ». *Operations Research Center*, University of Berkeley. Rapport n° ORC 63-21 (RR), 1963.
- [5] ROODE (J. D.) (*), « Generalized Lagrangian functions in mathematical programming ». *Thèse Université de Leiden*, octobre 1968.
- [6] UZAWA (H.), « Iterative methods for concave programming » dans *Studies on linear and nonlinear programming*, Chap. 10, édit. par Arrow, Hurwicz et Uzawa, Stanford University Press, 1958.

6. RECHERCHES PONCTUELLES OU PAR DICHOTOMIE

- [1] BOX (M. J.), « A new method for constrained optimization and a comparison with other methods », *The Computer Journ.*, **8** (1), 1965, p. 42-52.
- [2] BROOKS (S. M.), « A discussion of random methods of seeking maxima », *J. Op. res. Soc.*, **6** (2), 1958, p. 244-251.
- [3] HOOKE (R.) et JEEVES (T. A.), « Direct search solution of numerical and statistical problems », *J. Association Computing Machinery*, **8** (2), 1961, p. 212.
- [4] LEVIN (A. J.), « On an algorithm for the minimization of convex functions », *Soviet Math.*, **6** (1), 1965, p. 286-290.
- [5] McARTHUR (D. S.), « Strategy in research. Alternative methods for design of experiments ». *I.R.E. Trans. On Engineering Management*, Vol. EM-8, 1961.
- [6] MITCHELL (R. A.) et KAPLAN (J. L.), « Nonlinear constrained optimization by a non random complex method ». *Journ. Research NBS (Engr. and Instr.)*, **72 C**, 1968, p. 249.
- [7] NELDER (J. A.) et MEAD (R.), « A simplex method for function minimization, *The Computer Journ.*, **7**, 1965, 308.
- [8] NEWMAN (D. J.), « Locating the maximum on a unimodale surface ». *Exposé Nat. Meeting of Oper. Res. Society*, octobre 1960.
- [9] ROY (B.), « Programmation mathématique et description segmentée », *METRA*, **2** (4), 1963, p. 523-535.
- [10] SPENDLEY (W.), « Nonlinear least squares fitting using a modified simplex minimization technique ». *Proceed. Conference on Optimization Keele*, 1968 (Fletcher éd.). Academic Press, London.
- [11] SPEEDLEY (W.), HEXT (G. R.) et HIMSWORTH (F. R.), « Sequential application of simplex designs in optimization and evolutionary operation », *Technometrics*, **4**, 1962, p. 441.
- [12] WEGNER (P.), « A nonlinear extension of the simplex method », *Managt. Sc.*, **7** (1), 1960, p. 43-55.
- [13] WOOD (C. E.), « Review of design optimization techniques » *IREE Transactions on Systems Science and Cybernetics*, SSC-1, n° 1, 1965.

(*) Cf. aussi même titre *Proceed. Conference on Optimization, Keele*, mars 1968 (Fletcher Ed.), Academic Press.