THESIS / THÈSE

MASTER EN SCIENCES MATHÉMATIQUES

Conditions de convergence pratiques pour les problèmes d'optimisation d'une fonction sans contrainte

Romain, Nadine

Award date:
1976

Awarding institution:
Universite de Namur

Conditions de convergence pratiques pour les problèmes d'optimisation d'une fonction sans contrainte.

ROMAIN

NADINE

Table des matières.

Chapitre 1.

– Partie théorique –

I Introduction

Dans ce travail, nous nous intéressons au problème suivant : nous cherchons à résoudre le problème d'optimisation sans contrainte : $\text{Inf. } F(x) \, , \, x \in \mathbb{R}^n$ (P)

où F est une fonction réelle de n variables réelles. Généralement, on suppose F différentiable.

Les algorithmes constructifs, susceptibles d'applications numériques pour l'approximation du (ou d'un) $x \in \mathbb{R}^n$ réalisant l'infimum (appelé dès lors " solution optimale " du problème (P)), sont des méthodes itératives. Ces procédés de calcul démarrent avec un point x^0 donné.

En utilisant les valeurs calculées de la fonction et généralement des dérivées partielles premières, ils engendrent une suite de points x^k ($k = 1, 2, \ldots$)

Ils demandent évidemment que cette suite de points tendent vers le point où F atteint son minimum.

La plupart des algorithmes calculent la suite de points en utilisant des directions de recherche linéaires.

En un point x^k (point obtenu à la $k^{\text{ème}}$ itération) on choisit une direction de recherche d^k. A ce moment, la fonction

$$f(t) \equiv F(x^k + t\, d^k) \qquad (2)$$

est considérée comme une fonction de la variable réelle t.

Le point x^{k+1} (point à la $k+1^{\text{ème}}$ itération) est défini par

$$x^{k+1} = x^k + t\, d^k \qquad (3)$$

la valeur de t étant déterminée de manière à avoir l'inégalité

$$F(x^{k+1}) \leq F(x^k)$$

De nombreux algorithmes considéraient la minimisation EXACTE de la fonction $F(x)$

Or, en pratique, on a constaté qu'il était numériquement impossible d'atteindre cette minimisation exacte

C'est pourquoi, nous nous intéressons dans ce mémoire à des algorithmes ne considérant plus cette recherche exacte.

Le problème est toujours :

$$\min F(x) \ , \quad x \in \mathbb{R}^n$$

où F est une fonction de plusieurs variables.

MAIS au lieu de considérer le paramètre t comme un paramètre exact, ces algorithmes essayent de trouver une APPROXIMATION DE t et c'est là que réside l'innovation par rapport aux algoritmes précédents

Ils procèdent comme suit : le point à la $k+1^{\text{ème}}$ itération sera calculé à partir du point à la $k^{\text{ème}}$ itération par la relation

$$x^{k+1} = x^{k} + t^{k} d^{k}$$

où t^{k} est cette fois une approximation de t_m^{k} défini comme suit :

$$f^{k}(t_m^{k}) = \min \{ F(x^{k} + t d^{k}) \mid t > 0 \}$$

et où d^{k} est la direction de recherche.

A cause de cette approximation, une erreur sera introduite dans la solution et ce à chaque itération (puisqu'il faut approximer t à chaque itération)

Nous présentons des conditions suffisantes de convergence des méthodes gradient en tenant compte de l'erreur introduite par Wolfe [1] à savoir

$$\theta^k = \frac{\nabla F(x^{k+1})^t d^k}{\nabla F(x^k)^t d^k}$$

Nous aurons dans ce mémoire deux parties:

La 1ère partie sera la partie théorique

Nous y exposerons quelques propriétés des méthodes de descente ; un théorème de convergence et de vitesse de convergence. Nous les appliquerons à la méthode gradient conjugué et plus particulièrement à la méthode G.

La 2^de^ partie sera la partie pratique où nous présenterons un algorithme de résolution pour la méthode G.

II Préliminaires.

Notre but est donc de minimiser une fonction $F(x)$ sans contrainte et sans recherche linéaire exacte et d'établir des propriétés de convergence pour les méthodes permettant de réaliser ce fait.

Nous supposerons dans la suite que $F(x)$ possède des dérivées secondes continues et que la matrice des dérivées secondes est définie positive. Remarquons que la seconde hypothèse entraîne que la fonction $F(x)$ est convexe [2]. Elle implique aussi que tout minimum local est un minimum global [3].

Pour résoudre ce problème, nous procédons comme indiqué ci-après.

Plaçons-nous à la $k^{ème}$ itération.

Le point x^{k+1} est obtenu à partir de x^k par : $x^{k+1} = x^k + t^k d^k$

où t^k est une approximation de t_m^k défini par (3)

t^k est la longueur du pas

d^k est la direction de recherche.

A l'endroit du minimum, nous avons les relations : $\dfrac{df^k(0)}{dt} < 0$

et $\dfrac{df^k(t_m^k)}{dt} = 0$

$$\frac{df^k(t)}{dt} = \nabla F^t (x^k + t\,d^k)\, d^k$$

c'est à dire : en posant : $g^k = \nabla F(x^k)$

$(g^k)^t d^k < 0$ et $(g^{k+1})^t d^k = 0$ qui sont les conditions de conjugaison de la recherche exacte.

§ 1 : Notes théoriques concernant les erreurs de minimisation.

Les propriétés de convergence des méthodes gradient dépendent de la précision des directions de recherche à chaque itération. Cependant, atteindre un haut degré de précision dans les recherches linéaires demande beaucoup de temps et donc coûte très cher.

On a constaté que ces attentes de précision pouvaient être relachées sans affecter sérieusement la performance de l'algorithme. En fait, pour la méthode de la plus forte pente, on a trouvé plus intéressant dans certains cas de mettre la longueur du pas sous la forme :

$$t^{k} = \rho\, t_{m}^{k} \quad \text{avec} \quad 0 < \rho < 1$$

technique connue sous le nom de « sous-descente » [4]

Ljubic [5] a étudié les vitesses de convergence pour les méthodes de descente appliquées à la minimisation de fonctions quadratiques. Il a trouvé que les méthodes de descente convergeaient linéairement [6] (voir appendice 1) si et seulement si

$$\liminf_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} \rho(2-\rho) \cos^2 \psi^{k} > 0$$

où ψ^{k} est l'angle entre le gradient et la direction de recherche qui est défini par : $\cos \psi^{k} = (g^{k})^{t} d^{k} / |g^{k}||d^{k}|$

- ρ est le facteur de relaxation, $0 < \rho < 2$

Goldstein et Pria [7] ont trouvé que la condition :

$$\varepsilon < \frac{F(x^{k+1}) - F(x^k)}{\nabla F^t(x^k)(x^{k+1} - x^k)} < 1 - \varepsilon$$

où $0 < \varepsilon < 1/2$ serait intéressante puisqu'elle garantit la propriété :

$$\nabla F^t(x^k)\, d^k < 0$$

Wolfe [1] a montré que $1 - \varepsilon$ serait équivalent à $\theta^k < 1 - \varepsilon$ où

$$\theta^k = \frac{\nabla F^t(x^{k+1})\, d^k}{\nabla F^t(x^k)\, d^k}$$

De plus, il a noté que ε pouvait être remplacé par la condition que la fonction $f^k(t) = F(x^k + t d^k)$ soit monotonément décroissante : $0 < t < t_m^k$

Klessig et Polak [9] ont découvert un schéma pour contrôler les effets des erreurs dans la recherche linéaire en exigeant que $\theta^k < \delta^k$ où δ^k est choisi (dépendant de la valeur de ψ^k) tel que $\delta^k \leq \delta^{k-1} < 1$.

Ils ont montré que lorsque la méthode gradient conjugué suggérée par Polak et Ribière [10] est appliquée à la minimisation d'une fonction convexe, il existe des constantes positives δ et ρ telles que

$$\delta^k \geq \delta \qquad k = 1, 2, \ldots$$

$$\cos \psi^k \geq \rho$$

Alors l'algorithme révisé converge linéairement [6].

§ 2. Vitesse de convergence.

Les méthodes que nous considérons dans ce mémoire sont du type "gradient conjugué" c'est à dire qu'elles ont la propriété suivante :

Si elles sont appliquées directement à la minimisation d'une fonction quadratique $F(x) = x^t A x$, alors, les directions de recherche successives sont conjuguées par rapport à A c'est à dire :

$$(d^{k+1})^t A d^k = 0$$

Si $d^1 = -\nabla F(x^1)$, on peut montrer que de telles méthodes se terminent en n pas au plus pour un problème n-dimentionnel quadratique.

Powell [11] a montré que l'algorithme

de Davidon - Fletcher - Powell produisait la convergence super-linéaire [6] (voir appendice 1) quand on l'appliquait à une fonction convexe, en supposant une minimisation exacte.

Puisque les erreurs s'accumulent au fur et à mesure des itérations, nous aurons des problèmes en ce qui concerne les fonctions non quadratiques

C'est pourquoi on suggère de redémarrer l'algorithme toutes les n (ou $n+1$) itérations en posant $d^k = -\nabla F(x^k)$ si $\mathrm{mod}(k, n) = 0$. Une telle procédure converge vers le minimum, puisque le pas de descente est pris selon des intervalles réguliers.

§ 3 Quelques résultats classiques.

On considère deux classes de méthodes de gradient conjugué :

a) - les méthodes "vectorielles"

b) - les méthodes "matricielles"

a) Les méthodes "vectorielles" utilisent comme direction de recherche une combinaison linéaire du gradient et d'un autre vecteur, habituellement la direction de recherche précédente

$$d^k = a\, d^{k-1} + b\, g^k$$

où $g^k = \nabla F(x^k)$

Ce sont de telles méthodes que nous considérerons dans la suite

b) Les méthodes "matricielles" calculent la direction de recherche en multipliant le gradient par une certaine matrice

$$d^k = - M^k g^k$$

où M^k est calculée à partir de M^{k-1}, d^{k-1}, g^{k-1} et g^k

L'erreur de la solution d'un problème uni-dimentionnel sera mesurée à la $k^{\text{ème}}$ itération par

$$\theta^k = (g^{k+1})^t d^k / (g^k)^t d^k$$

qui représente le rapport de la dérivée de la fonction dans la direction d^k sur la dérivée dans la direction d^k en x^k

Remarquons que si x^{k+1} est le lieu exact du minimum alors $\theta^k = 0$

Pour la méthode Fletcher-Reeves, on a montré [12] que $0 \leq \theta^k \leq 1 - \eta$ où

$0 < \eta < 1$ est suffisante pour garantir la convergence vers le minimum. On n'a pas encore établi de vitesse de convergence.

Pour la méthode Davidon-Fletcher-Powell, les conditions $0 \leq \theta^k \leq 1 - m$

$$\theta^k \leq R |g^k|^2 / (g^k)^t M^k g^k$$

$$\theta^k \leq S (g^k)^t M^k g^k / |g^k|^2$$

$$\theta^k \leq T (g^{k+1})^t M^k g^{k+1} / |g^k|^2$$

$$RT \leq 1 - c$$

où $0 < m < 1$, $0 < c < 1$, R, S, T des constantes arbitraires positives, sont suffisantes [12] pour avoir une vitesse de convergence linéaire [6].

Une condition plus forte :

$$\theta^k \leq (1-c)(g^{k+1}) M^k g^{k+1} / (g^k)^t M^k g^k$$

est suffisante [12] pour que le rapport soit superlinéaire [6]

Nous démontrerons dans

que la condition

$$[(g^{k+1})^t g^k]\,\theta^k = (1-c)\,|g^{k+1}|^2$$

est suffisante pour avoir une vitesse de convergence linéaire pour la "méthode G".

La "méthode G" est une méthode gradient conjugué dans laquelle on fixe

$b = -1$

III Propriétés des méthodes de descente

§1 Nous allons démontrer dans les pages suivantes, des théorèmes se rapportant à la vitesse de convergence et à la convergence des méthodes de descente en général. Nous énoncerons en premier lieu les hypothèses dans lesquelles nous travaillons, ainsi que quelques définitions.

Nous nous plaçons sous les hypothèses :

a/ : $F(x)$ a des dérivées secondes continues.

b/ : les valeurs propres de la matrice hessian $\nabla^2 F(x)$ sont limitées supérieurement et inférieurement respectivement par E et $\mathcal{E}$.

c/ : les dérivées secondes de F satisfont la condition de Lipschitz :

$$|\nabla^2 F(x) - \nabla^2 F(\zeta)| \leq L\,|x - \zeta|$$

où L est une constante et ζ la position du minimum de $F(x)$

<u>Définition 1</u> : Une <u>méthode de descente</u> est un procédé itératif pour trouver le minimum sans contrainte d'une fonction objective $F(x)$, pour laquelle, étant donné x^0 on a : $F(x^{k+1}) < F(x^k)$ $k = 0, 1 ..$ où x^k est le point à l'itération k

<u>Définition 2</u> : Le vecteur d est une <u>direction de descente</u> à partir du point x, si il existe un $\delta > 0$ tel que pour tout $t \in [0, \delta]$ on ait :

$$F(x + t\,d) < F(x)$$

§2 Nous allons développer des résultats semblables à ceux donnés par Wolfe [1] mais en tenant compte des erreurs dans les recherches linéaires. A cet effet, nous établirons deux lemmes qui nous conduiront à un théorème de vitesse de convergence et un théorème de convergence.

Lemme 1 :

Soit $\bar{x} = x + \bar{t}\,d$ où d est tel qu'on ait :

$\nabla F^t(x)\,d < 0$ et $\nabla F^t(\bar{x})\,d = \theta\,\nabla F^t(x)\,d$

$0 \leq \theta < 1$ et

Soit $\cos\psi = \nabla F^t(x)\,d * |\nabla F(x)|^{-1} * |d|^{-1}$

Alors :

$$\frac{1}{2} E^{-1}(1-\theta)^2 |\nabla F(x)|^2 \cos^2\psi \leq F(x) - F(\bar{x})$$

$$\leq \frac{1}{2} \varepsilon^{-1} (1-\theta^2) |\nabla F(x)|^2 \cos^2\psi. \quad (1*)$$

Démonstration :

Considérons une fonction à une variable :

$h(t)$ définie pour tout $t > 0$

Supposons que $h(0) = 0$, $h'(0) < 0$ et

$$\varepsilon \leq h''(t) \leq E .$$

Par intégration de ces inégalités, nous obtenons : $\varepsilon t + h'(0) \leq h'(t) \leq Et + h'(0)$ (1)

Soit $\bar{t}_f$ une approximation du minimum de $h(t)$ telle que $h'(\bar{t}_f) = \theta h'(0)$ où $0 \leq \theta \leq 1$. En remplaçant t par $\bar{t}_f$ dans (1) nous avons :

$$-\frac{(1-\theta)\, h'(0)}{E} \leq \bar{t}_f \leq -\frac{(1-\theta)\, h'(0)}{\varepsilon} \qquad (2)$$

Intégrons à nouveau (1) :

$$q(t) \leq h(t) \leq Q(t) \qquad (3)$$

où $q(t) = \frac{1}{2}\, \varepsilon t^2 + [h'(0)]\, t$

$Q(t) = \frac{1}{2}\, E t^2 + [h'(0)]\, t$

Soient $\bar{t}_q$ et $\bar{t}_Q$ les deux points où :

$q'(\bar{t}_q) = \theta q'(0)$ et $Q'(\bar{t}_Q) = \theta Q'(0)$. Alors :

$$\bar{t}_q = -\frac{(1-\theta)[h'(0)]}{\varepsilon} \; ; \; \bar{t}_Q = -\frac{(1-\theta)[h'(0)]}{E}$$

A ce moment, on peut écrire (2) par :

$$\bar{t}_Q \leq \bar{t}_h \leq \bar{t}_q \qquad (4)$$

Puisque $q(t)$ est monotonément décroissante dans l'intervalle $(0, \bar{t}_q)$, nous pouvons utiliser (3) et (4) pour obtenir :

$$h(\bar{t}_h) \geq q(\bar{t}_h) \geq q(\bar{t}_q) \qquad (5)$$

De la même manière :

$$h(\bar{t}_h) \leq h(\bar{t}_Q) \leq \bar{Q}(\bar{t}_Q) \qquad (6)$$

En combinant (5) et (6), nous avons :

$$q(\bar{t}_q) \leq h(\bar{t}_h) \leq Q(\bar{t}_Q)$$

ou

$$-\frac{1}{2}\,\varepsilon^{-1}(1-\theta^2)[h'(0)]^2 \leq h(\bar{t}_h) \leq$$

$$-\frac{1}{2}\,E^{-1}(1-\theta^2)[h'(0)]^2 \qquad (7)$$

Supposons que $h(t)$ soit définie comme

$h(t) = F(x + td) - F(x)$. Alors :

$h'(t) = \nabla F^t(x + td)\, d$

$h''(t) = d^t \nabla^2 F(x + td)\, d$

Or, $h(0) = 0$. et si d est une direction de descente à partir de x, alors $h'(0) < 0$.

En utilisant les limites des valeurs propres du Hessien : $\varepsilon |d|^2 \leq h''(t) \leq E |d|^2$

Donc, à partir de (2), nous avons :

$$\varepsilon |d|^2 \bar{t}_h \leq -(1-\theta) \nabla F^t(x)\, d \leq E |d|^2 \bar{t}_h \quad (8)$$

Posons : $\bar{x} = x + \bar{t}_h d$. L'équation (7) nous donne :

$$-\frac{1}{2} \varepsilon^{-1} (1-\theta^2) [F^t(x)\, d]^2 |d|^{-2} \leq F(\bar{x}) - F(x)$$

$$\leq -\frac{1}{2} E^{-1} (1-\theta^2) [\nabla F^t(x)\, d]^2 |d|^{-2}$$

CQFD.

<u>Commentaire</u> : En imposant $\theta \geq 0$, nous

avons exigé que la dérivée au point minimum approché et au point de départ aient le même signe. Nous avons donc exigé que l'approximation du minimum soit entre le point de départ et le minimum exact. A ce moment, nous sommes sûrs que la valeur de la fonction au minimum approché sera inférieure à la valeur de la fonction au point de départ

Lemme 2

Si ξ est la position du minimum de $F(x)$ alors :

$$\frac{1}{2} E^{-1} |\nabla F(x)|^2 \leq F(x) - F(\xi) \leq \frac{1}{2} \varepsilon^{-1} |\nabla F(x)|^2 \qquad (2*)$$

Démonstration :

Soit x^* le point obtenu à partir de x

dans la direction d jusqu'au moment

où $\theta = 0$ c'est à dire $\nabla F^t(x^*)\, d = 0$

Si ξ est la position du minimum

de $F(x)$ alors, $F(x) - F(\xi) \geq F(x) - F(x^*)$

A partir de l'inégalité du lemme précé-

dent en prenant $d = -\nabla F(x)$ on a :

$$\frac{1}{2} E^{-1} |\nabla F(x)|^2 \leq F(x) - F(\xi)$$

et puisque d peut être choisi telle que

x^* soit arbitrairement près de ξ, on a :

$$F(x) - F(\xi) \leq \frac{1}{2} \varepsilon^{-1} |\nabla F(x)|^2 \qquad \text{CQFD.}$$

<u>Théorème 1</u> (de vitesse de convergence)

Si une méthode de descente est telle que

$1 - (\theta^k)^2 \geq c > 0$ et $\cos^2 \psi^k \geq r > 0$ où

r et c sont des constantes et $k = 1, 2, \ldots$

alors, la suite de points engendrée

par la méthode converge vers la solution avec une vitesse qui est au moins linéaire.

<u>Démonstration</u> :

Ce théorème est une conséquence immédiate de la relation :

$$1 - E^{-1}\varepsilon(1-\theta^2)\cos^2\psi \geq \frac{F(\bar{x}) - F(\xi)}{F(x) - F(\xi)} \geq 1 - E^{-1}\varepsilon(1-\theta^2)\cos^2\psi \qquad (3*)$$

obtenue en divisant la relation (1 *) du lemme 1 par la relation (2*) du lemme 2.

<u>Théorème 2</u> (de convergence)

Si une méthode de descente est telle que $1-(\theta^k)^2 \geq c > 0$ où c est une constante et $k = 1, 2, \ldots$

Alors, la suite de points engendrée

par la méthode converge vers la solution si et seulement si

$$\sum_{j=1}^{\infty} \cos^2 \psi^j = \infty$$

Démonstration :

Soit $\bar{M} = \lim_{j \to \infty} F(x^j)$ (à cause des conditions sur θ, $F(x^j)$ est une suite monotoniment décroissante limitée inférieurement par $F(z)$ et donc $\bar{M}$ existe.) Alors : $\bar{M} \geq F(z)$

Si $\sum_j \cos^2 \psi^j < \infty$, alors il existe un k tel que la partie droite de l'inégalité (3*) dans le théorème 1 satisfait :

$1 - \varepsilon^{-1} E(1-(\theta^j)^2) \cos^2 \psi^j \geq 1 - \varepsilon^{-1} E \cos^2 \psi^j > 0$

pour tout $j \geq k$. Nous pouvons donc prendre le produit de la partie droite de $(3*)$ pour tout $j > h$.

Supposons que $F(x^h) \neq F(\mathfrak{Z})$, nous avons

$$\frac{\bar{M} - F(\mathfrak{Z})}{F(x^h) - F(\mathfrak{Z})} \geq \prod_{j \geq h} \left(1 - \varepsilon^{-1} E \cos^2 \psi^j\right) \quad (4*)$$

A partir de $(4*)$, nous concluons que

$\bar{M} > F(\mathfrak{Z})$ et $F(x^h)$ ne converge pas vers $F(\mathfrak{Z})$.

Pour prouver l'inverse, nous prenons la partie gauche de $(3*)$ pour tout $j > h$, où nous supposons que $F(x^h) \neq F(\mathfrak{Z})$

$$\prod_{j \geq h} \left[1 - E^{-1} \varepsilon \left(1 - (\theta^j)^2\right) \cos^2 \psi^j\right] \geq \frac{\bar{M} - F(\mathfrak{Z})}{F(x^h) - F(\mathfrak{Z})} \quad (5*)$$

La condition $\Sigma_j \cos^2 \psi^j = \infty$ implique que la partie gauche de $(5*)$ est nulle ; donc $\bar{M} = F(\mathfrak{Z})$, et $F(x^h)$ converge vers $F(\mathfrak{Z})$ et la preuve est complète.

IV Application de la théorie précédente à la méthode gradient conjugué

Nous présentons dans les pages suivantes une application de la théorie donnée dans le paragraphe précédent.

Plus particulièrement, nous parlerons de la vitesse de convergence d'une méthode gradient particulière : la "méthode 6".

Nous donnerons à cet effet des conditions sur l'erreur dans la solution à chaque itération. Ces conditions sont suffisantes pour montrer que la méthode converge vers le minimum d'une fonction qui admet des dérivées secondes et dont les valeurs propres de la matrice hessian sont bornées inférieurement et supérieurement. Nous montrerons que sous ces conditions, la "méthode G" converge linéairement [6]

Nous supposerons aussi que la fonction à minimiser est quadratique.

Les méthodes gradient utilisent comme direction de recherche une combinaison linéaire du gradient et d'un autre vecteur à l'itération précédente

$$d^{k+1} = a^k d^k + b^k g^{k+1}$$

où a^k et b^k sont des scalaires et où g^{k+1} est le gradient de la fonction au point x^{k+1}

Pour simplifier les notations, nous

marquerons d'une astérisque les quantités se rapportant à la $k+1^{ème}$ itération et nous laisserons sans indice celles de la $k^{ème}$ itération.

Nous aurons donc :

$$d^* = a\,d + b\,g^*$$

Nous choisissons a et b tels que d^* soit conjugué à d par rapport à la matrice $G = \nabla^2 F$

Cette condition nous donne la relation suivante : $\dfrac{a}{b} = - \dfrac{g^{*t}\, G\, d}{d^t\, G\, d}$ (1)

(réalisable car G est définie positive donc $d^t G d \neq 0$ car $d \neq 0$)

Cette relation découle de la réalisation de la condition de conjugaison : $d^* G d = 0$

En effet : $d^{*t} G d = a\, d^t G d + b\, g^{*t} G d = 0$

$$a\, d^t G d = -\, b\, g^{*t} G d$$

$$a / b = -\, g^{*t} G d \,/\, d^t G d.$$

En posant $y = g^* - g$, la relation (1) peut s'écrire :

$$a / b = -\, g^{*t} y \,/\, d^t y.$$

Nous considérons, avons-nous dit, un cas particulier de la méthode gradient conjugué à savoir la "méthode G".

Cette méthode n'est rien d'autre que la méthode gradient conjugué où l'on pose :

$$b = -1$$

Avec cette valeur de b, nous obtenons donc a par :

$$a = g^{*t} y \,/\, d^t y$$

Avant d'examiner les propriétés de convergence de la "méthode 6" nous citerons quelques formules géométriques qui, développées, nous serviront dans la démonstration des théorèmes de vitesse de convergence.

Lemme technique.

i) Posons
$$\begin{cases} \cos\psi = g^t d\,|g|^{-1} \\ \cos\varphi = g^{*t} d\,|g^*|^{-1}|d|^{-1} \\ T = a\,|d|\,/\,b\,|g^*| \end{cases}$$

nous aurons : $b^{-1}\cos\psi^* = |d^*|^{-1}|g^*|(1+T\cos\varphi)$

$$|d^*|^2/|g^*|^2 \leq (1+|T|)^2$$

Démonstration : voir Annexe 2.

ii) Pour prouver la convergence linéaire de la "méthode 6" nous imposerons la condition suivante sur θ :

$$(g^{*t} g)\,\theta < (1-\eta)\,|g^*|^2$$

qui nous conduira à une condition nécessaire et suffisante pour que d^* soit une direction de descente, c'est à dire à la condition

$$(g^{*t} g)\,\theta \leq |g^*|^2$$

Démonstration : Voir Annexe 3.

En se servant de ces résultats, nous sommes maintenant en mesure de démontrer les théorèmes qui vont suivre.

Théorème 1.

La suite de points engendrée par la "méthode σ" converge linéairement vers le minimum de $F(x)$ si à chaque itération $k = 1, 2, \ldots$ l'erreur θ^k satisfait l'inégalité :

$$\theta^k (g^{k+1})^t g^k < (1-\eta) |g^{k+1}|^2$$

$$0 \leq \theta^k \leq 1-c$$

où c et η sont des constantes :

$0 < c < 1$ et $0 < \eta < 1$.

Démonstration :

Considérons la quantité

$$T = a |d| / b |g^*|$$

$$= - g^{*t} \eta \, |d| |\eta| / |g^*| |\eta| \, d^t \eta$$

(en remplaçant a / b par sa valeur)

$$|T| \leq \frac{|d| |\eta|}{d^t \eta} |\cos(g^*, \eta)|$$

Utilisons les limites des valeurs propres du Hessian, $\mathcal{E}$ et E

Nous aurons:

$$|T| \leq \mathcal{E}^{-1} E \qquad (1)$$

Considérons ensuite:

$$1 + T\cos\varphi = 1 + \frac{g^{*t} y}{d^t y} \frac{|d|}{|g^*|} \frac{g^{*t} d}{|g^*||d|}$$

$$= 1 + (1-\theta)^{-1} \theta\, g^{*t} y |g^*|^{-2}$$

$$= (1-\theta)^{-1} (1 - \theta g^{*t} g |g^*|^{-2})$$

Par hypothèse nous avons:

$$1 + T\cos\varphi > \eta / c \qquad (2)$$

Dès lors, en substituant:

$$\frac{|d^*|^2}{|g^*|^2} \leq (1 + |T|)^2 \; ; \quad (1) \text{ et } (2)$$

dans l'équation:

$$b^{-1} \cos\psi^* = |d^*|^{-1} |g^*| (1 + T\cos\varphi),$$

nous obtenons :

$$-\cos \psi^{k} > \eta \left[C \left(1 + \varepsilon^{-1} E \right) \right]^{-1}$$

Grâce au théorème 1 page 30, la suite engendrée par la méthode converge vers le minimum de $F(x)$ en une vitesse au moins linéaire.

Théorème 2

A chaque pas d'une méthode gradient conjugué, la valeur d'une fonction objective quadratique diminue d'une valeur au moins aussi grande qu'un pas de la plus forte pente pris à partir du même point, si la méthode gradient conjugué satisfait à l'inégalité :

$$|d^{k}|^{2} \, |g^{k}|^{-2} \cos^{2} \psi^{k} > 1 \quad , \quad k = 0, 1, 2 \ldots$$

Démonstration :

Par le même raisonnement que celui qui conduit à l'inégalité

$$-\frac{1}{2}\,\varepsilon^{-1}(1-\theta^2)[F^t(x)\,d]^2\,|d|^{-2} \leq F(\bar{x})-F(x)$$
$$\leq \frac{1}{2}\,E^{-1}(1-\theta^2)[\nabla F^t(x)\,d]^2\,|d|^{-2}$$

(voir lemme 1, p 25)

nous obtenons :

$$F(\bar{x})-F(x)=\frac{-(1-\theta^2)[\nabla F^t(x)\,d]^2}{2\,d^t[\nabla^2 F]\,d} \qquad (1)$$

En posant : $G=\nabla^2 F$, en élevant d^* au carré et en multipliant par G nous obtenons :

$$d^{*t}G\,d^*=g^{*t}G\,g^*-2a\,g^{*t}G\,d+a^2\,d^tG\,d$$

En utilisant : $a/b=-g^{*t}G\,d/d^tG\,d$,

nous avons :

$$d^{*t}G\,d^*=g^{*t}G\,g^*-2a\,g^{*t}G\,d+a^2\,d^tG\,d$$
$$=g^{*t}G\,g^*-2\,\frac{g^{*t}G\,d}{d^tG\,d}\,g^{*t}G\,d$$

$$+\frac{(g^{*t} G d)^2}{(d^t G d)^2} d^t G d$$

$$= g^{*t} G g^* - \frac{2(g^{*t} G d)^2}{d^t G d} + \frac{(g^{*t} G d)^2}{d^t G d}$$

$$= g^{*t} G g^* - \frac{(g^{*t} G d)^2}{d^t G d}$$

ce qui implique :

$$d^* G d^* > g^* G g^* \qquad (2)$$

Considérons maintenant

$$(g^{*t} d^*)^2 = |g^*|^4 \, |d^*|^2 \cos^2 \varphi^* \, |g^*|^{-2}$$

Alors, par hypothèse

$$(g^{*t} d^*)^2 > |g^*|^4 \quad (\text{car } |d^*|^2 \cos^2 \varphi^* \, |g^*|^{-2} > 1)$$

(3)

Grâce aux relations (1), (2), (3), on établi le résultat demandé.

Remarque.

Si nous choisissons dans le théorème 1 les paramètres tels que $\eta > c$, alors la "méthode 6" satisfait les hypothèses du théorème 2.

Cela peut se voir en combinant :

$$b^{-1}\cos\psi^* = |d^*|^{-1}|g^*|(1+\tau\cos\varphi)$$

et $|d^k|^2|g^k|^{-2}\cos^2\psi^k > 1$, $k = 0, 1, 2 \ldots$

nous obtenons :

$$|d^*|^2|g^*|^{-2}\cos^2\psi^* = (1+\tau\cos\varphi)^2 > (\eta/c)^2 > 1$$

Nous avons donc montré que chaque direction choisie par la "méthode 6" sera au moins aussi bonne que la direction de la plus forte pente quand la méthode est appliquée à la minimisation d'une fonction quadratique.

V Conclusions.

Nous avons considéré la minimisation d'une fonction à n variables réelles sans contrainte.

Comme la minimisation exacte n'est numériquement pas réalisable, nous avons introduit un terme d'erreur θ^k dans la solution du problème à chaque itération. La grandeur de cette erreur est mesurée par :

$$\theta^k = \frac{\nabla F^t(x^{k+1})\, d^k}{\nabla F^t(x^k)\, d^k}$$

Cette erreur est due au fait que nous avons considéré le paramètre t, dans l'expression donnant le point x à la $k+1^{\text{ème}}$ itération à savoir :

$x^{k+1} = x^k + t^k d^k$ (où d^k est la direction de recherche à la $k^{ième}$ itération) non plus comme une valeur exacte mais comme une approximation.

Nous avons démontré des théorèmes se rapportant à la vitesse de convergence et à la convergence d'une méthode quelconque. Nous avons ensuite présenté comme application des conditions sur θ^k et α pour la "méthode G". Ces conditions nous permettent d'assurer la convergence linéaire de la méthode. Elles ne sont que suffisantes.

VI Annexes.

Annexe 1.

Convergence linéaire.

La notion de vitesse de convergence est définie comme suit :

Soit une suite $\{z_i\}_{i=0}^{\infty}$ dans un espace de Banach B, qui converge vers z^*

On dira que $\{z_i\} \longrightarrow z^*$ au moins linéairement (ou que la vitesse de convergence vers z^* est au moins linéaire) si il existe un entier $k \geq 0$, une constante E et un $\theta \in [0, 1]$, tels que

$$\| z_i - z^* \|_B \leq E\theta^i \qquad \forall i \geq k$$

c'est-à-dire que la converge de $\{z_i\}$

est au moins linéaire si $\| z_i - z^* \|_B \to 0$

quand $i \to \infty$ au moins aussi vite

qu'une progression géométrique

Convergence superlinéaire.

Définissons l'ordre-Q d'une méthode

itérative I au point x^* comme :

$$O_Q = \infty \text{ si } Q_p = 0 \text{ pour tout } p.$$

$$= \inf\{p : Q_p = \infty\} \text{ autrepart.}$$

(Pour la définition de Q_p voir [4])

La méthode est dite Q-superlinéaire si

$Q_1 = 0$

La convergence superlinéaire ne spécifie

pas l'ordre, par exemple, la converge

quadratique est superlinéaire.

(Dans cette optique, la convergence est

dite linéaire si $O_Q = 1$ et $0 < Q_1 < 1$)

Annexe 2

Démonstration de la partie i) du lemme technique page 38 à savoir :

Posons $\cos\psi = g^t d\, |g|^{-1}$

$$\cos\varphi = g^{*t} d\, |g^*|^{-1} |d|^{-1}$$

$$T = -\frac{a\,|d|}{b\,|g^*|}$$

Nous aurons :

1) $b^{-1}\cos\psi^* = |d^*|^{-1}|g^*|(1 + T\cos\varphi)$

2) $|d^*|^2 / |g^*|^2 \leq (1 + |T|)^2$

1) Multiplions $d^* = ad + bg^*$ par g^*

$$g^{*t} d^* = \alpha g^{*t} + b\,|g^*|^2$$

$$= b\,|g^*|^2 (1 + T\cos\varphi)$$

Alors : $b^{-1}\cos\psi^* = |d^*|^{-1}|g^*|(1 + T\cos\varphi)$

En effet : $b\,|g^*|^2 + b\,|g^*|^2 + \cos\varphi =$

$$= b|g^*|^2 + b|g^*|^2 \frac{a|d|}{b|g^*|} g^{*t} d \, |g^*|^{-1} |d|^{-1}$$

$b|g^*|^2 + a g^{*t} d =$ 1er membre.

$$b^{-1} \cos \varphi^* = |d^*|^{-1} |g^*| \left(1 + \frac{a|d|}{b|g^*|} g^{*t} d \, |g^*|^{-1} |d|^{-1}\right)$$

$$= |d^*|^{-1}|g^*| + \frac{|d^*|^{-1}|g^*| \, a|d| \, g^{*t} d \, |g^*|^{-1} |d|^{-1}}{b|g^*|}$$

$$= |d^*|^{-1}|g^*| + \frac{|d^*|^{-1} a g^{*t} d}{b|g^*|}$$

$$b^{-1} g^{*t} d^* |g^*|^{-1} |d^*|^{-1} = \frac{(b|g^*||d^*|^{-1}|g^*|) + (|d^*|^{-1} a g^{*t} d)}{b|g^*|}$$

$$b|g^*| b^{-1} g^{*t} d^* |g^*|^{-1} |d^*|^{-1} = b|g^*||d^*|^{-1}|g^*| + a|d^*|^{-1} g^{*t} d$$

$$g^{*t} d^* |d^*|^{-1} = b|g^*||d^*|^{-1}|g^*| + a|d^*|^{-1} g^{*t} d$$

donc $g^{*t} d^* = b|g^*|^2 + a g^{*t} d$ CQFD.

2) Prenons le carré de $d^* = ad + bg^*$

$$|d^*|^2 = a^2|d|^2 + 2ab\, g^{*t} d + b^2 |g^*|^2$$

$$\frac{|d^*|^2}{|g^*|^2} = b^2 + \frac{2ab\, g^{*t} d}{|g^*|^2} + a^2 \frac{|d|^2}{|g^*|^2}$$

$$= b^2\left(1 + 2\frac{a}{b}\frac{g^{*t} d}{|g^*|^2} + \frac{a^2}{b^2}\frac{|d|^2}{|g^*|^2}\right)$$

Puisque $b = -1$.

on a : $$\frac{|d^*|^2}{|g^*|^2} = 1 - 2a\frac{g^{*t} d}{|g^*|^2} + a^2\frac{|d|^2}{|g^*|^2}$$

or $$T = -a\frac{|d|}{|g^*|}$$

$$T\cos\varphi = \frac{a}{b}\frac{|d|}{|g^*|}\, g^{*t} d\, |g^*|^{-1}\, |d|^{-1}$$

$$= \frac{a}{b}\frac{g^{*t} d}{|g^*|^2} = -a\frac{g^{*t} d}{|g^*|^2}$$

$$\Rightarrow \frac{|d^*|^2}{|g^*|^2} = 1 + 2T\cos\varphi + T^2 = (1 + |T|)^2$$

CQFD.

Annexe 3.

Démontrons maintenant la seconde partie du lemme c'est-à-dire :

Pour prouver la convergence linéaire de la "méthode G" nous imposerons la condition suivante sur θ à savoir : $(g^{*t}g)\theta < (1-\eta)|g^*|^2$ (où $0 < \eta < 1$), qui nous conduira à une C.N.S pour que d^* soit une direction de descente c'est à dire à la condition :

$$(g^{*t}g)\theta < |g^*|^2$$

La mesure de l'erreur dans la solution à l'itération k : est : $\theta = \dfrac{g^{*t}d}{g^t d}$

Nous imposons une condition sur θ : celle de l'énoncé.

Puisque $g^{*t} d^* < 0$ si et seulement si

$$-|g^*|^2 + g^{*t} y (d^t y)^{-1} (g^{*t} d) < 0$$

ou : $-[|g^*|^2 + \theta(1-\theta)^{-1}(|g^*|^2 - g^{*t} g)] < 0$

ou : $-(1-\theta)^{-1}(|g^*|^2 - \theta g^{*t} g) < 0$,

une condition nécessaire et suffisante pour que d^* soit une direction de descente est :

$$(g^{*t} g)\,\theta < |g^*|^2$$

Chapitre 2.

– Partie numérique –

Expérience numérique

Afin d'illustrer la théorie exposée dans le 1er chapitre, nous allons l'appliquer à un problème bien défini à savoir :

$$\min F(x_1, x_2) = 100\,(x_2 - x_1^2)^2 + (1 - x_1)^2$$

L'algorithme général de résolution sera le suivant :

Etape 0 : lire $x^0 = (-1, 2, 1.0)$

Etape 1 : plaçons-nous à la $k+1$ ème itération. Soient

$$d^{k-1},\ x^k,\ F(x^k),\ g(x^k) \text{ connus.}$$

a) - calculer la direction de recherche d^k

$$d^k = a^{k-1} d^{k-1} + b^{k-1} g^k$$

$$b^{k-1} = -1$$

$$a^{k-1} = \frac{(g^k)^t\, y^{k-1}}{d^{k-1}\, y^{k-1}} \quad \text{où } y^{k-1} = g^k - g^{k-1}$$

b) calculer t^k, approximation de t_m^k

(Ce calcul étant assez complexe, nous le détaillons dans un algorithme auxiliaire)

Etape 2 : calculer x^{k+1} par la relation :

$$x^{k+1} = x^k + t^k d^k$$

Calculer $g(x^{k+1})$

Si $g(x^{k+1}) < \varepsilon$ STOP, on est arrivé au minimum

Sinon passer à l'étape 3.

Etape 3 : Calculer l'erreur faite dans la solution : $\theta^k = \dfrac{g(x^{k+1})^t d^k}{g(x^k)^t d^k}$

Etape 4 : Tests de convergence sur θ^k

Si $\theta^k \leq 1 - \eta \qquad 0 < \eta < 1$

et si $[g(x^{k+1})^t g^k]\, \theta^k \leq (1-c)|g^{k+1}|^2$

$0 < c < 1$

aller à l'étape 2

sinon aller en 6

<u>Etape 6</u> : Il n'y a pas convergence. STOP.

La plus grande difficulté est de calculer le paramètre t^k

Nous devons pour cela résoudre un problème uni-dimentionnel à chaque itération.

Il faut trouver t^* tel que

$$f(t^*) = \min \{ f(t) = F(x + td) \mid t > 0 \}$$

Pour résoudre ce problème, nous utilisons un procédé recommandé par Coggins [13]

Nous allons rechercher une approximation du paramètre suivant l'algorithme auxiliaire suivant:

Algorithme auxiliaire.

1) on se donne un point initial x_0 et on calcule la valeur de la fonction en ce point

2) Si on se trouve à la première itération on donne à t une valeur $t_0 = 1$

Sinon : $t_0^{k+1} = 0.5 \, t_m^k \, |d^k| / |d^{k+1}|$

3) On a un nouveau point : $x_0 + t_0 d_0$

4) On calcule la valeur de la fonction en ce point

5) Si la valeur de la fonction est inférieure ou égale à la valeur de la fonction au point antérieur, on va en 6

Sinon on va en 7.

6) On double la longueur du pas et on calcule la valeur de la fonction en ce nouveau point

Aller ensuite en 5)

7) on réduit de moitié la longueur du pas et un pas est à nouveau franchi

à partir du dernier point ayant donné une amélioration du résultat. On obtient de cette manière 4 points également espacés sur l'axe de recherche

8) On rejette le point terminal le plus éloigné du point pour lequel on obtient la plus petite valeur ; on a donc localisé trois points équidistants : t_1, t_2, t_3 tels que : $t_1 < t_2 < t_3$

$$t_3 - t_2 = t_2 - t_1$$

et $f(t_1) > f(t_2)$

$$f(t_2) < f(t_3)$$

9) Partant avec ces 3 points, on utilise une interpolation quadratique qui nous permet de localiser un point t^* tel que $t_1 < t^* < t_3$ (ce qui est possible

puisque la fonction est convexe)

10) On effectue des approximations quadratiques successives pour localiser le minimum. Après chaque itération, un point est éliminé.

11) Les itérations se terminent quand

a) $| f_m^* - f_m | < \bar{\varepsilon} \, | t_m^* - t_m |$

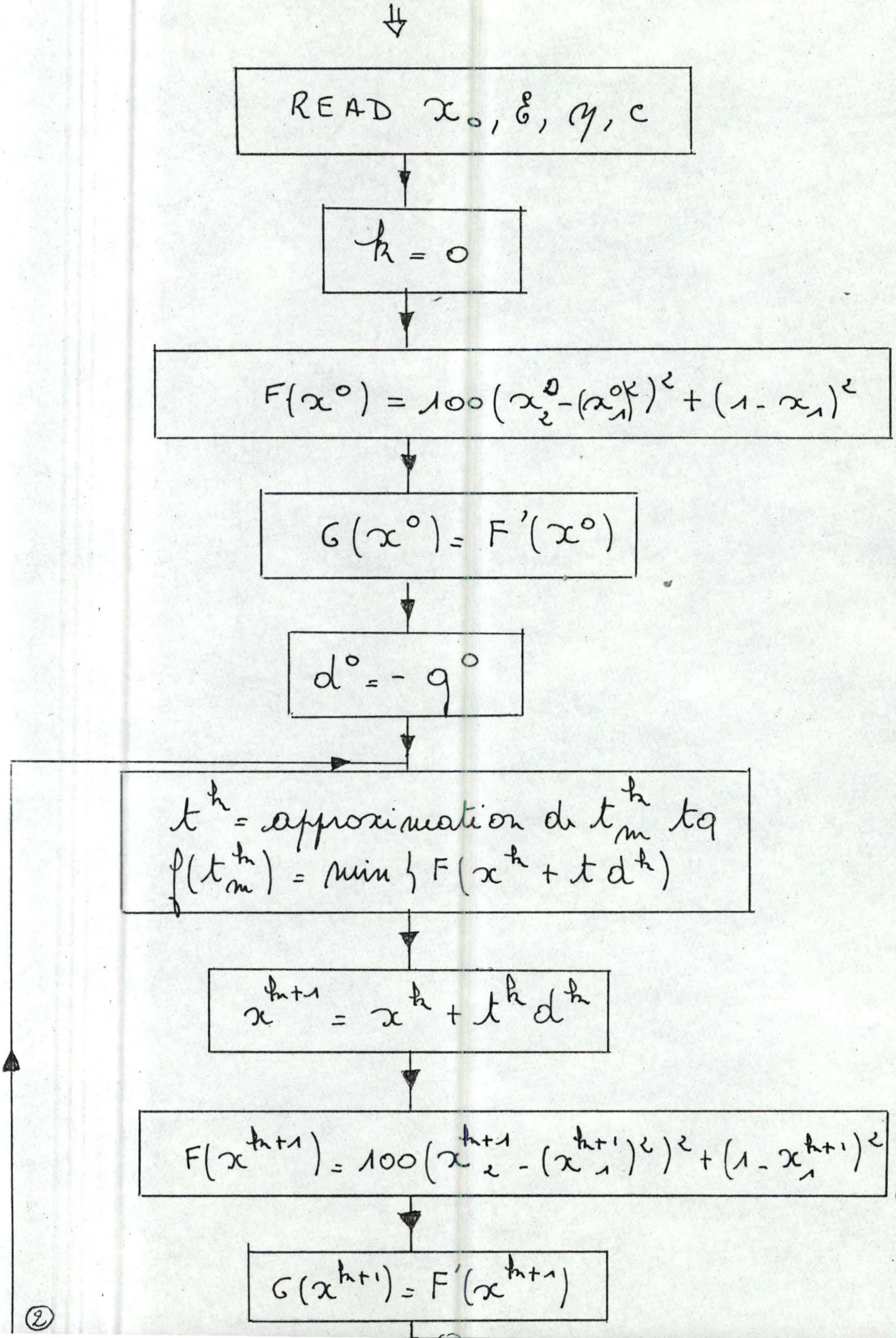
où $\bar{\varepsilon} = \varepsilon \, [g^t (x^k) \, d^k]$

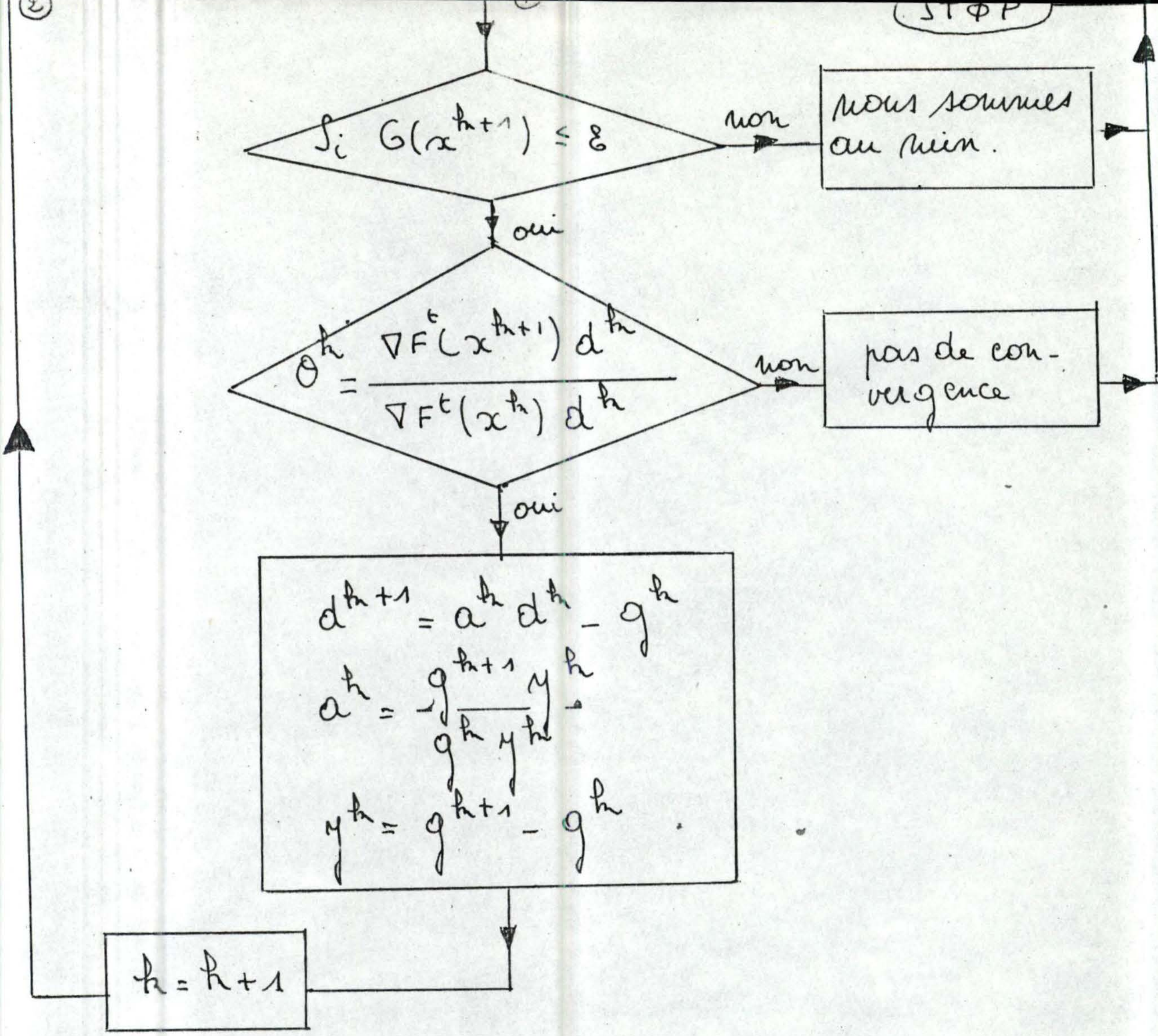
où f_m et f_m^* sont deux approximations successives du minimum de $f(t)$ se passant en $t = t_m$ et $t = t_m^*$

et b) $\bar{\varepsilon} \, | t_m^* - t_m | < 10^{-5} \left(| f_m^* + f_m | / 2 \right)$

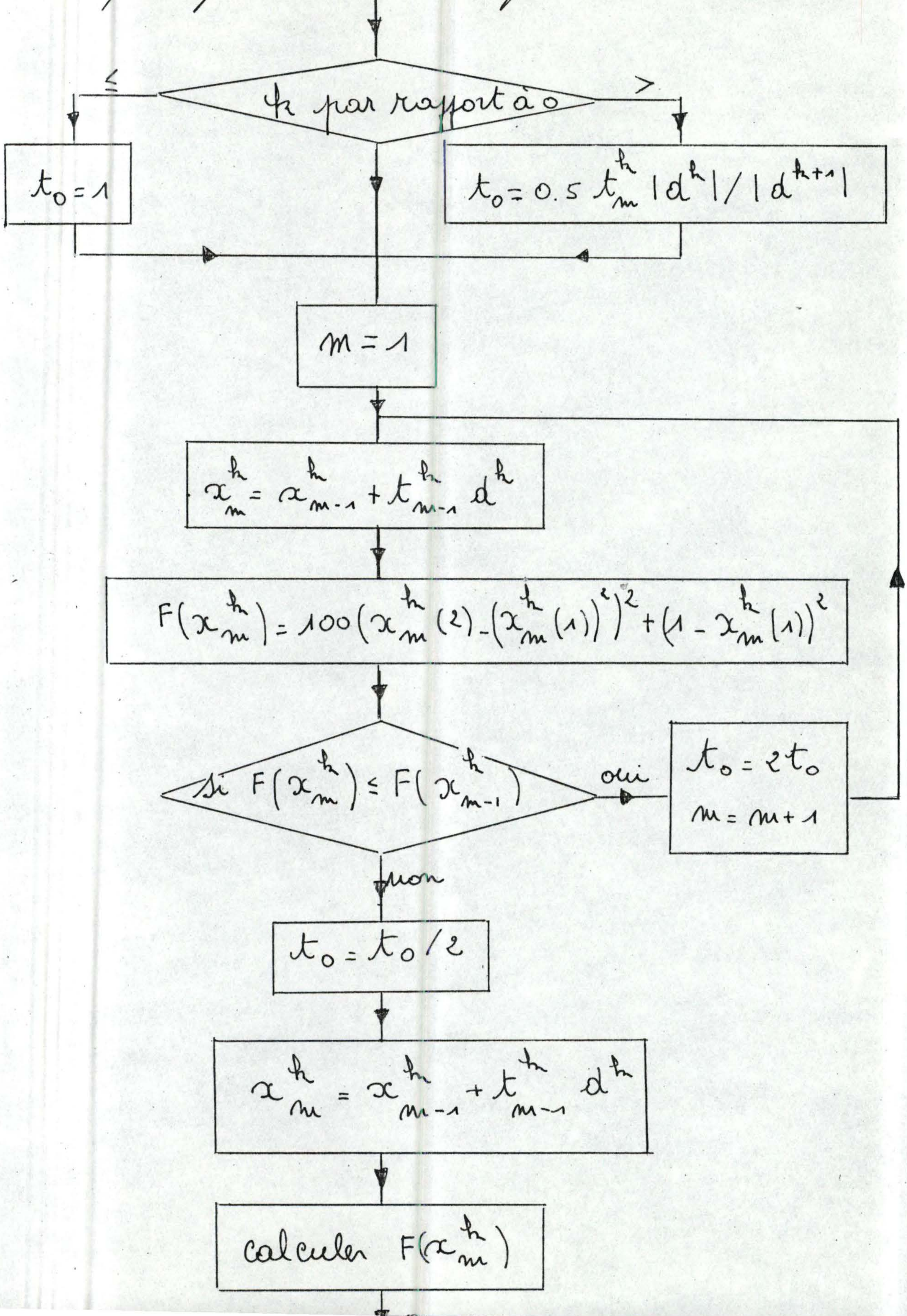
A ce moment, le t_m^* sera la valeur que l'on prendra dans le calcul de x^{k+1}.

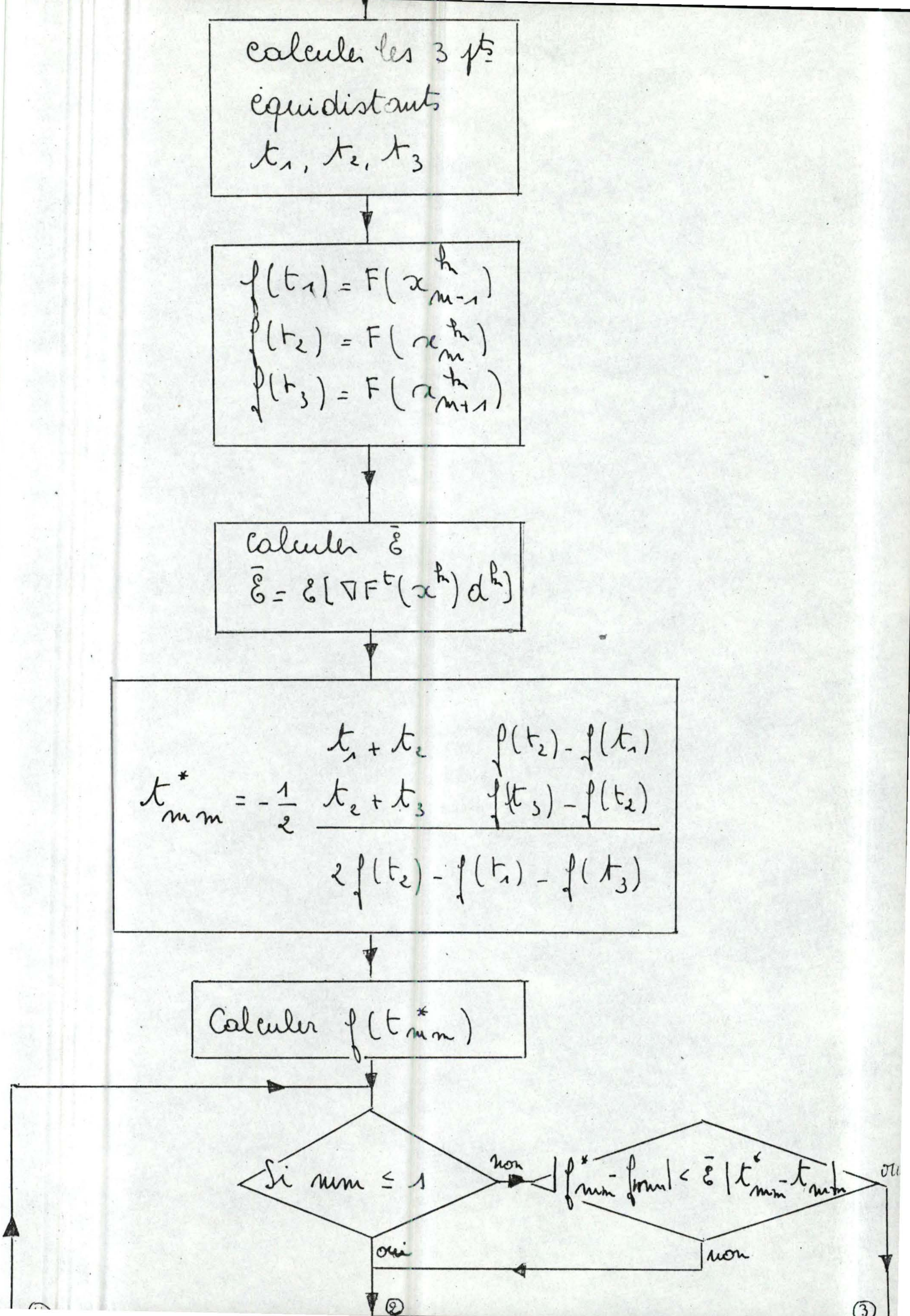
Organigramme de l'algorithme général



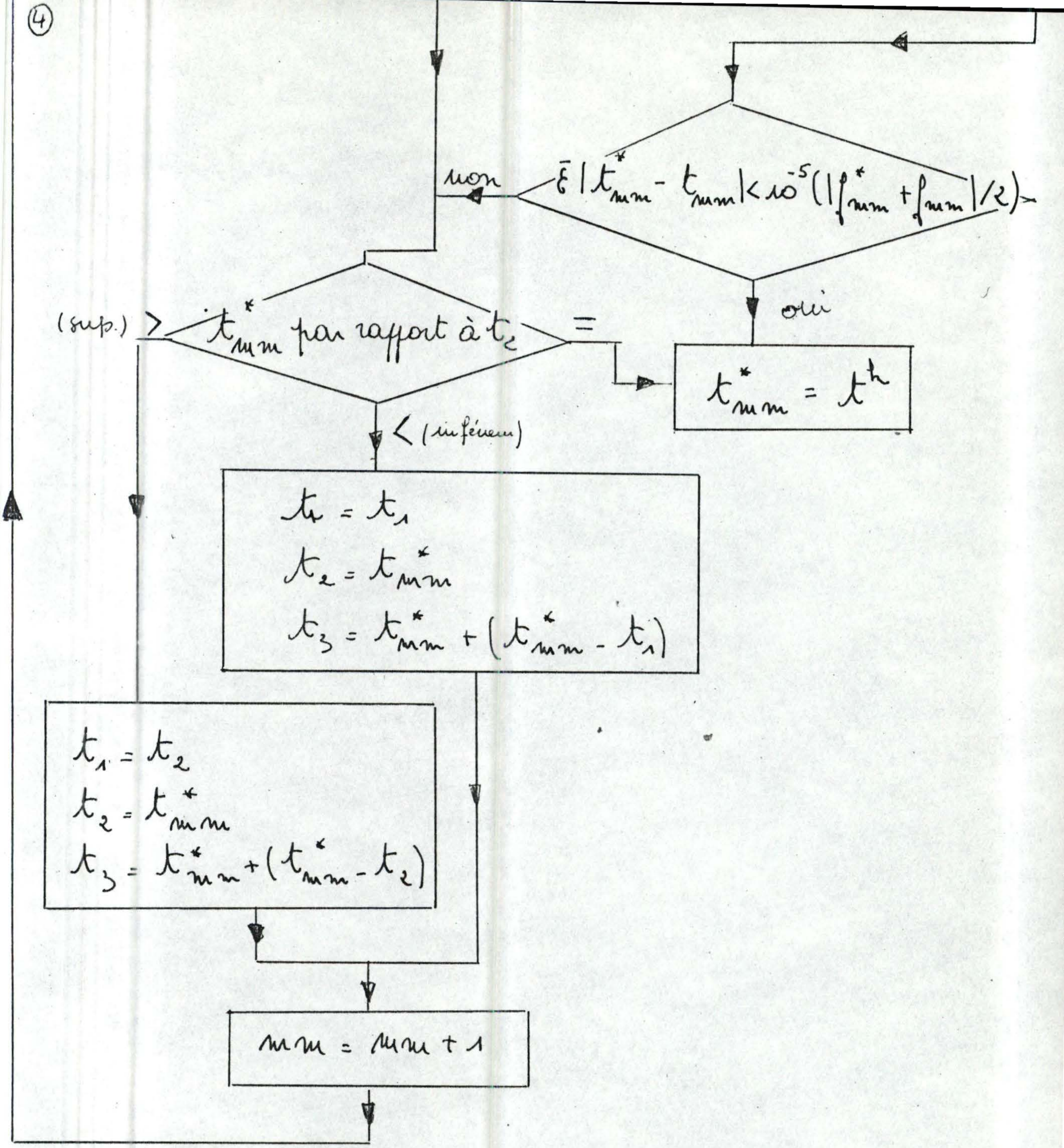

2
STOP
Si $G(x^{k+1}) \leq \varepsilon$
non
nous sommes au min.
oui
$\theta^k = \dfrac{\nabla F^t(x^{k+1})\, d^k}{\nabla F^t(x^k)\, d^k}$
non
pas de convergence
oui
$d^{k+1} = a^k d^k - g^k$
$a^k = -\dfrac{g^{k+1} y^k}{g^k y^k}$
$y^k = g^{k+1} - g^k$.
$k = k+1$

Organigramme de l'algorithme auxiliaire





calculer les 3 pts équidistants t_1, t_2, t_3
$f(t_1) = F(x^k_{m-1})$
$f(t_2) = F(x^k_m)$
$f(t_3) = F(x^k_{m+1})$
Calculer $\bar{\varepsilon}$
$\bar{\varepsilon} = \varepsilon[\nabla F^t(x^k) d^k]$
$t^*_{mm} = -\frac{1}{2} \frac{(t_1 + t_2)(f(t_2) - f(t_1)) + (t_2 + t_3)(f(t_3) - f(t_2))}{2f(t_2) - f(t_1) - f(t_3)}$
Calculer $f(t^*_{mm})$
Si $mm \leq 1$
non
oui
$|f^*_{mm} - f_{mm}| < \bar{\varepsilon} |t^*_{mm} - t_{mm}|$
oui
non
②
③



Remarque : k représente les itérations sur x

m représente les itérations sur t

Si $t^*_{nm} = t^k$, on met cette valeur dans le programme principal.

Résultats.

Suivant les schémas des organigrammes, nous avons rédigé le programme en utilisant pour la recherche de l'approximation de t une méthode de 3 points équidistants. Nous avons donc été amené à faire des itérations dans des itérations. Comme les itérations sur x sont déjà assez nombreuses, et que chacune exige des itérations sur t, il faut un temps machine considérable. ce qui à mon avis n'est pas un "bon point" pour la méthode.

De plus, quand on obtient t^* à la 1ère itération, il faut à nouveau chercher 3 points équidistants. Nous l'avons fait de la manière suivante :

La fonction étant convexe, nous étions sûrs que : $t^1 < t^* < t^3$ (où t^* a pris la place de t_2). Nous regardons de quel côté de t^2 se situe t^*. S'il se trouve à gauche, nous prenons le plus grand écart entre $t^2 - t^*$ ou $t^* - t^1$ et nous prenons les 3 points suivants :

$$t_1 = t_1 \quad \text{si } t^* - t^1 \text{ est } > \text{ à } t^2 - t^*$$
$$= t^* - (t_2 - t^*) \quad \text{sinon}$$

$$t_2 = t^*$$

$$t_3 = t^* + (t^* - t^1) \quad \text{si } t^* - t^1 \text{ est } > t^2 - t^*$$
$$= t_2 \quad \text{sinon.}$$

Là aussi la méthode serait sans aucun doute à améliorer.

Mélanie Lenard [12] a testé ce problème. Elle a comparé les résultats obtenus en employant 3 méthodes différentes : la méthode G, la méthode Davidon-Fletcher-Powell, la méthode Fletcher-Reeves.

Elle a obtenu :

méthode G : 170 évaluations de F, 25 évaluations de G ; (avec $\varepsilon = .5$, $T = .9$)

méthode DFP : 136 éval. de F, 21 évaluations de G ; (avec $\varepsilon = .5$, $T = 10^6$)

méthode F.R : 508 évaluations de F, 77 évaluations de G ; (avec $\varepsilon = .5$).

Elle n'a pas précisé sa méthode pour la recherche de l'approximation du paramètre t.

Conclusions.

Nous constatons donc que la non précision dans les recherches linéaires n'est pas un facteur critique pour atteindre une vitesse de convergence satisfaisante pour les méthodes gradient conjugué

Les résultats varient évidemment suivant le choix des constantes ε et T

La grande difficulté à mon avis réside dans la recherche d'une approximation du paramètre t

Je suppose qu'il y aurait moyen de trouver une méthode plus perfectionnée que celle proposée, et qui permettrait d'accélérer encore la vitesse de convergence

Références.

[1] Wolfe P.: " Convergence theory in non linear Programming", in Integer and non linear Programming, Ed. J. Abadie (North Holland, Amsterdam, 1930) 1-36.

[2] S. Vajda: " Theory of linear and non linear Programming" ch. 1, p. 7.

[3] J. Fichefet: " Programmation linéaire. cours de 1ère licence.

[4]: Blum: " Numerical Analysis and Computation: Theory and Practice" Addison - Wesley Publishing Company, ch. 5, p 215-229.

[5]: Ljubic: " General theorems on Quadratic Relaxation", Soviet mathematics

6, p. 588 - 591.

[6] E. Polak : " Computational Methods in Optimization. A Unified Approach ", Academic Press. New York and London, ch 6 [242 - 268]

[7] Goldstein and Price : " An Effective Algorithm for Minimization ", Numerische Mathematik 10, p 184 - 189.

[9] Klessig R. et Polak E. : " Efficient Implementations of the Polak - Ribière Conjugate Gradient Algorithm ", Memo No ERL M 479 Electronics Research Laboratory, University of California, Berkeley.

[10] Polak E and Ribière : " Notes sur la convergence de méthodes de directions conjuguées ", Revue française d'Informatique et de Recherche opérationnelle 3, p 35-43.

[11] Powell: "On the Convergence of a Variable Metric Algorithm", Atomic Energy Research Establishment, Harwell TP 332.

[12] Lenard Melanie: "Practical Convergence Conditions for Unconstrained Optimization", Columbia University.

[13] Coggins, G.F: "Univariate Search Methods", Research Note 64/11, Central Instrument Research Laboratory Imperial Chemical Industries, Ltd.