MASTER EN SCIENCES MATHÉMATIQUES

Convergence des méthodes quasi-Newton pour les problèmes de contrôle optimal linéaires quadratiques

Gilson, Bernard

Award date:
1979

Awarding institution:
Universite de Namur

CONVERGENCE DES METHODES QUASI-NEWTON POUR LES PROBLEMES DE CONTROLE OPTIMAL LINEAIRES QUADRATIQUES

Promoteur : NGUYEN VAN HIEN

GILSON BERNARD

Je tiens à exprimer toute ma gratitude envers Monsieur Nguyen Van Hien : ses précieux conseils ont permis l'élaboration de ce travail.

Bernard Gilson

table des matières

introduction

Lorsqu'on résout des problèmes de contrôle optimal, on recherche généralement des contrôles qui satisfont le principe du maximum de Pontryagin [P.3]. De tels contrôles sont appelés contrôles extrémaux, et il est bien connu que le contrôle optimal, s'il existe, est un contrôle extrémal. En d'autres termes, le principe du maximum est une condition nécessaire d'optimalité. Dans certains problèmes de contrôle optimal, le principe du maximum n'apporte aucune information constructive. De tels problèmes sont apelés problèmes singuliers. Intuitivement, ils peuvent être considérés comme des cas où un test du 3ième ordre ou d'un ordre supérieur est nécessaire pour déterminer le comportement de la fonction de coût au voisinage d'un arc extrémal.

D'un point de vue algorithmique, de nombreux algorithmes ont été créés pour résoudre les problèmes de minimisation en dimension finie, et leurs généralisations aux espaces fonctionnels ont été appliquées aux problèmes de contrôle optimal. Un problème de contrôle optimal linéaire quadratique peut en effet être transformé en un problème de minimisation quadratique dans un espace fonctionnel. De ce fait, tous les résultats trouvés pour les problèmes de minimisation dans un espace fonctionnel sont aussi valables pour les problèmes de contrôle optimal correspondant.

C'est la raison pour laquelle la plupart des algorithmes ont été généralisés pour résoudre les problèmes de contrôle optimal, et les algorithmes les plus connus (les méthodes du gradient conjugué [F.1] et du D.F.P. [F.2]) ont entre autres été appliqués à des problèmes de contrôle optimal non singuliers, avec des résultats de convergence satisfaisants.

Pour les problèmes de minimisation quadratique non singuliers, les vitesses de convergence ont été étudiées par plusieurs auteurs, avec les résultats suivants :

(i) la méthode du gradient (à chaque itération, on descend dans une direction qui est l'opposé du gradient de la fonction au point d'itération) a une vitesse de convergence linéaire pour les problèmes dans des espaces de dimension finie et infinie ;

(ii) dans les espaces de dimension finie, les méthodes du gradient-conjugué, quasi-Newton et de Huang ([H-1], [H-2], [H-4]) ont la propriété de convergence finale quadratique.

(iii) dans les espaces fonctionnels, des résultats de convergence pour les méthodes du gradient-conjugué et du D.F.P. ont été présentés dans [H-4], [S-1] notamment.

Le principal objectif de ce travail était l'étude des propriétés de convergence et de vitesse de convergence des algorithmes quasi-Newton appliqués aux problèmes de minimisation quadratique singuliers dans les espaces fonctionnels. Le travail se base sur la thèse de B. Cheng [C-3]. Voici les résultats essentiels de cette étude :

cas singulier, espace de dimension finie.

Il est connu que les algorithmes quasi-Newton appliqués aux problèmes quadratiques non singuliers dans un espace de dimension n convergent en au plus n itérations. Dans le chapitre 3, on montrera que les mêmes algorithmes, appliqués cette fois aux problèmes singuliers quadratiques, convergent en au plus m ($\leq n$) itérations, où m est le rang de la matrice Hessienne ($n \times n$).
En outre, la vitesse de convergence pour le problème quadratique singulier est supérieure ou égale à la vitesse du même algorithme appliqué au problème quadratique non singulier associé.
Mais la méthode du gradient a exactement les propriétés opposées. Cette lente convergence pour les problèmes singuliers finidimensionnels est en fait une propriété de la méthode elle-même, et n'est pas due à la singularité du problème.

cas régulier, espace fonctionnel

Dans le chapitre 4, 2 résultats de vitesse de convergence seront obtenus.
Le premier est un résultat de vitesse de convergence linéaire; mais cette vitesse linéaire est trop lente comparée aux résultats obtenus.
Dès lors, une vitesse de convergence superlinéaire est obtenue si l'opérateur initial $H_0 = Id$ et si l'opérateur Hessien A a la forme $A = Id + K$, où K est un opérateur positif

autoadjoint complètement continu.
Cette propriété de convergence reste vraie pour tous les algorithmes quasi-Newton, puisque ceux-ci générent les mêmes directions de recherche à chaque pas.

cas singulier, espace fonctionnel

On considérera 2 cas mutuellement exclusifs : le cas où l'image $R(A)$ de l'opérateur Hessien A est fermé, et le cas où $R(A)$ est non fermé.
Cette différence affecte fortement les résultats de vitesse de convergence des algorithmes quasi-Newton.
On montrera que tous les résultats de vitesse de convergence obtenus pour le cas non singulier sont préservés pour ces algorithmes appliqués aux problèmes quadratiques singuliers, avec $R(A)$ fermé.
Pour le cas où $R(A)$ est non fermé, la propriété de convergence de l'algorithme est prouvée, mais aucun résultat de vitesse de convergence n'est obtenu.
Cependant, on verra que l'estimation initiale $x_0 \in R(A) + N(A)$, affecte fortement la vitesse de convergence de la suite des fonctionnelles.

Remarquons enfin que la notion de pseudo-inverse d'un opérateur est appelée à jouer un rôle important dans la recherche d'une solution d'un problème singulier.
C'est ce qui justifie une étude détaillée de cette notion dans le chapitre introductif.

chapitre 1 :

rappel de quelques notions - pseudo-inverse d'un opérateur.

1.1. adjoint d'un opérateur

1.2. propriétés dans la théorie des espaces de Hilbert

1.3. la notion de pseudo-inverse.

- 1.3.1. image fermée
- 1.3.2. image arbitraire
- 1.3.3. méthode de la plus grande descente.

1.4. inverse d'un opérateur

1.5. le spectre d'un opérateur

Chapitre 1 : rappels de quelques notions - pseudo-inverse d'un opérateur.

1.1. adjoint d'un opérateur [S.2, G.1]

1] définition :

un espace de Hilbert réel H est un espace de Banach (c-à-d. un espace vectoriel réel normé et complet) muni d'un produit intérieur $\langle \cdot, \cdot \rangle : H \times H \longrightarrow \mathbb{R}$, ce produit vérifiant les propriétés :

$\forall x, y, z \in H$, $\forall \alpha, \beta \in \mathbb{R}$:

a) $\langle x, y \rangle = \overline{\langle y, x \rangle}$

b) $\langle \alpha . x + \beta . y, z \rangle = \alpha . \langle x, z \rangle + \beta . \langle y, z \rangle$

c) $\langle x, x \rangle \geq 0$, $\langle x, x \rangle = 0$ ss : $x = 0$

2] théorème de représentation de Riesz :

> si ϕ est une fonctionnelle linéaire continue sur l'espace de Hilbert H, alors il existe un élément unique $y \in H$ tel que $\phi(x) = \langle x, y \rangle$ pour chaque $x \in H$.

3] définition :

si H_1 et H_2 sont deux espaces de Hilbert munis respectivement des produits intérieurs $\langle \cdot, \cdot \rangle_1$ et $\langle \cdot, \cdot \rangle_2$, et si A est un opérateur de $\mathcal{L}(H_1, H_2)$ ($\mathcal{L}$ l'espace des opérateurs linéaires et continus de H_1 dans H_2) alors l'adjoint de A, noté A^*, est l'unique opérateur linéaire dans $\mathcal{L}(H_2, H_1)$ qui satisfait la relation :

$\langle Ax, y \rangle_2 = \langle x, A^* y \rangle_1 \quad \forall x \in H_1, \forall y \in H_2$

En fait, cet opérateur A^* est bien défini grâce au théorème de représentation de Riesz. En effet, considérons $\langle Ax, y \rangle$, où y est un élément fixé de H_2 et x varie dans H_1. Il est clair que $\langle Ax, y \rangle$ est une fonctionnelle linéaire bornée en x (bornée, car $|\langle Ax, y \rangle| \leq \|Ax\| . \|y\|$ par l'inégalité de Cauchy-Swarz. Or, $\|y\| < \infty$ et $\|Ax\| < \infty$, puisque A est borné).
Par le théorème de représentation de Riesz, il existe un élément univoquement déterminé g tel que $\langle Ax, y \rangle = \langle x, g \rangle \quad \forall x \in H_1$.
Cet élément g dépend évidemment du choix de y. On le notera $g = A^* y$, où A^* est un opérateur linéaire défini sur H_2.

4] Rappelons les propriétés importantes de cet opérateur A^* ; soient A, B deux opérateurs de $\mathcal{L}(H_1, H_2)$,

1. $(A . B)^* = A^* . B^*$
2. $(A^*)^* = A$
3. $(A + B)^* = A^* + B^*$
4. $(\alpha . A)^* = \overline{\alpha} . A^*$
5. $\|A^*\| = \|A\|$
6. $\|A^* A\| = \|A A^*\| = \|A\|^2$
7. A^* est toujours linéaire et fermé.
8. le domaine D_{A^*} de A^* contient au moins le vecteur nul

Rappelons aussi qu'un opérateur A est dit fermé si la propriété suivante est vérifiée :

si $\{x_n\}$ est une suite dans D_A, le domaine de A,

- $x_n \longrightarrow x$
- $Ax_n \longrightarrow f$

alors $x \in D_A$ et $Ax = f$

5] définition :

un opérateur $A \in \mathcal{L}(H, H)$ est appelé autoadjoint si $A = A^*$.

6] Les théorèmes suivants établissent les relations fondamentales entre l'image de l'opérateur A (noté $R(A) = \{y \in H_2 : y = Ax$ pour un $x \in H_1\}$), le noyau de l'opérateur A (noté $N(A) = \{x \in H_1 : Ax = 0\}$), et l'image et le noyau de l'opérateur adjoint A^*.

théorème 1.1 $\quad \boxed{(R(A))^\perp = N(A^*)}$

démonstration :

Montrons d'abord que $(R(A))^\perp \subset N(A^*)$. Soit $z \in (R(A))^\perp$. Donc, $\langle Ax, z\rangle = 0 = \langle x, 0\rangle \ \forall x \in D_{A^*}$. Ainsi, $(z, 0)$ est un couple admissible, ce qui veut dire que $z \in D_{A^*}$ et $A^* z = 0$. Par définition de $N(A^*)$, $A^* z = 0 \iff z \in N(A^*)$

Il reste donc à montrer que $N(A^*) \subset (R(A))^\perp$.

Soit $z \in N(A^*)$. Cela veut dire que $A^* z = 0$.

Dès lors, $\langle x, A^* z\rangle = 0 \ \forall x$, et à fortiori $\forall x \in D_A$.

Par conséquent, $\langle Ax, z\rangle = 0 \ \forall x \in D_A$. Donc $z \in (R(A))^\perp$

■

corollaire 1.1 : $\overline{R(A)} = (N(A^*))^{\perp}$

corollaire 1.2 : si $R(A)$ est un ensemble fermé, alors $R(A) = (N(A^*))^{\perp}$

démonstration : [S.2, p.171]

théorème 1.2 : si H_1, H_2 sont 2 espaces de Hilbert, si $A \in \mathcal{L}(H_1, H_2)$ alors $R(A)$ fermée ssi $R(A^*)$ fermée

démonstration : [G.1, p.15]

1.2. propriétés dans la théorie des espaces de Hilbert

Nous présentons ici quelques théorèmes fondamentaux, dont nous nous servirons fréquemment dans les chapitres suivants. des démonstrations de ces théorèmes se trouvent dans la plupart des livres traitant de la théorie des espaces de Hilbert. (notamment [T.1])

théorème 1.3 :

un sous-ensemble C convexe fermé d'un espace de Hilbert H contient un unique vecteur de norme minimale.

théorème 1.4 :

> si C est un sous ensemble convexe fermé d'un espace de Hilbert H,
> alors $\forall u \in H$, $\exists ! \, x \in C$ tel que
> $\|u - x\| = \inf \{ \|u - y\| , y \in C \}$

théorème 1.5 :

> si S est un sous espace de H,
> alors $S^\perp$, le complément orthogonal de S,
> est toujours un sous espace fermé.

théorème 1.6 :

> si S est un sous espace fermé de H,
> alors H peut s'écrire comme la somme directe
> de S et $S^\perp$ (ce qui se note $H = S \oplus S^\perp$),
> ce qui veut dire que tout élément $x \in H$ peut
> s'écrire de manière unique comme
> $x = x_1 + x_2$, où $x_1 \in S$ et $x_2 \in S^\perp$.
> x_1 est appelé la projection orthogonale de x sur S

1.3. la notion de pseudo - inverse [G-1]

1.3.1. image fermée

1] Il est bien connu qu'une matrice A possède un inverse si et seulement si elle est carrée et non singulière, ou en d'autres termes, ses colonnes (ou ses lignes) sont linéairement indépendants. On notera cet inverse par A^{-1}, et on aura : $A \cdot A^{-1} = Id = A^{-1} \cdot A$

Supposons que H_1, H_2 soient deux espaces de Hilbert sur le même corps de scalaire. Considérons le problème suivant : résoudre une équation linéaire générale du type $\underline{Ax = b}$, où $b \in H_2$, $A \in \mathcal{L}(H_1, H_2)$. (1.3)

Si l'opérateur A a un inverse, alors l'équation (1.3) possède l'unique solution $x = A^{-1} \cdot b$.

Mais, en général, une telle équation linéaire peut avoir plus d'une solution (si $N(A) \neq \{0\}$, alors A est non injectif), ou peut ne pas avoir de solution du tout (si $b \notin R(A)$).

Même si l'équation n'a pas de solution au sens classique, il est quand même possible de trouver ce qui, en quelque sorte, constitue la "meilleure solution" possible au problème.

De manière générale, on voudrait donc parler d'un "inverse généralisé" d'une matrice A rectangulaire ou singulière. Ainsi, pour pseudo-inverse d'une matrice donnée A, on veut parler d'une matrice X associée d'une certaine façon à A,

i) qui existe pour une classe de matrices plus grande que la classe des matrices non singulières

ii) qui a quelques propriétés des inverses habituels

iii) qui coïncide avec l'inverse habituel quand A est non singulière.

Illustrons le point iii) ; supposons X définie par $A \times A = A$. Si A est régulière, alors

$$A \times A = A \quad \text{ssi} \quad A^{-1} A \times A = Id$$
$$\text{ssi} \quad \times A = Id$$
$$\text{ssi} \quad \times = A^{-1} \ .$$

On se ramène bien dans ce cas à la notion ordinaire de l'inverse.

2] définition :

Soit A un opérateur de $\mathcal{L}(H_1, H_2)$, où H_1, H_2 sont 2 espaces de Hilbert. On suppose dans ce paragraphe que R(A) est fermé. Soit P la projection de H_2 sur R(A) (ceci a bien un sens par le théorème 1.6, et par le fait que R(A) est fermé). Dans l'équation (1.3), b est un vecteur de H_2. Dès lors, Pb est le vecteur de R(A) qui est le plus proche de b, et il semble raisonnable de considérer comme une solution généralisée de (1.3) toute solution $u \in H_1$ de l'équation

$$Ax = Pb \qquad (1.4)$$

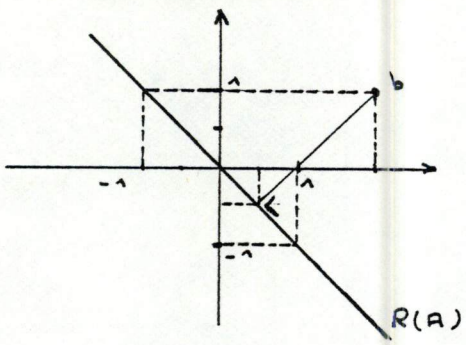
3] exemple :

Soit l'opérateur $A \in \mathcal{L}(\mathbb{R}^2, \mathbb{R}^2)$, représenté par la matrice $\begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}$. Soit $b = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$. On peut donc écrire :

$$Ax = b \iff \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \end{pmatrix} .$$

L'image de A, R(A), vaut : $R(A) = \{(y_1, y_2) : y_1 = -y_2\}$
$= \{\alpha . (1, -1) \text{ , où } \alpha \in \mathbb{R}\}$

Comme $b \notin R(A)$, le système n'admet pas de solution (au sens habituel). On voit bien que $P_b = \begin{pmatrix} 1/2 \\ -1/2 \end{pmatrix}$



Une solution généralisée sera donc un couple (x_1, x_2), solution de

$$\begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1/2 \\ -1/2 \end{pmatrix}$$

$$\Leftrightarrow \begin{cases} -x_1 + x_2 = 1/2 \\ x_1 - x_2 = -1/2 \end{cases}$$

l'ensemble des solutions généralisés est donc :

$$\left\{ (x_1, x_2) \in H_1 \times H_1 \ : \ x_2 = x_1 + \frac{1}{2} \right\}$$

4] Une autre approche naturelle pour trouver les solutions généralisées de l'équation (1.3) serait de trouver un élément $u \in H_1$ qui "vient le plus près" pour résoudre (1.3), dans le sens suivant : $\| Au - b \| \leq \| Ax - b \| \quad \forall x \in H_1$

Géométriquement, il est évident que cette notion est équivalente à la précédente (il suffit d'examiner l'exemple : le point de $R(A)$ qui "s'approche" le plus de b est la projection orthogonale de b sur $R(A)$).

On peut même montrer plus :

Théorème 1.7

Soit $A \in \mathcal{L}(H_1, H_2)$, $R(A)$ fermé, $b \in H_2$, alors les conditions suivantes sur $u \in H_1$ sont équivalentes :

(i) $Au = P_b$

(ii) $\| Au - b \| \leq \| Ax - b \| \quad \forall x \in H_1$

(iii) $A^* Au = A^* b$

démonstration :

(i) $\Longrightarrow$ (ii)

supposons que $Au = Pb$. En utilisant la propriété de Pythagore ($\|x+y\|^2 = \|x\|^2 + \|y\|^2$ si $x \perp y$), et le fait que $R(A)$ est fermé (donc $H = R(A) \oplus (R(A))^\perp$), on a :

$$\begin{aligned}\|Ax - b\|^2 &= \|Ax - Pb + Pb - b\|^2 \\ &= \|Ax - Pb\|^2 + \|Pb - b\|^2 \\ &= \|Ax - Pb\|^2 + \|Au - b\|^2 \\ &\geq \|Au - b\|^2\end{aligned}$$

(comme Pb est la projection orthogonale de b sur $R(A)$, $Pb - b \in (R(A))^\perp$)

(ii) $\Longrightarrow$ (iii)

si $\|Au - b\| \leq \|Ax - b\| \ \forall x \in H$, alors de nouveau par l'utilisation de la propriété de Pythagore, et le fait que $Ax = Pb$ pour un $x \in H$, on a :

$$\begin{aligned}\|Au - b\|^2 &= \|Au - Pb + Pb - b\|^2 \\ &= \|Au - Pb\|^2 + \|b - Pb\|^2 \\ &\geq \|Au - Pb\|^2 + \|b - Au\|^2 \text{ pour un}\end{aligned}$$

certain $x \in H$, t.q. $Ax = Pb$.

Donc $Au - b = Pb - b \in (R(A))^\perp = N(A^*)$ (par le théorème 1.1).

ainsi, $Au - b = Pb - b \in N(A^*)$

$\Longleftrightarrow A^*(Au - b) = 0$

$\Longleftrightarrow A^* Au = A^* b$, et (iii) est démontré

(iii) $\Longrightarrow$ (i)

si on a $A^*Au = A^*b$, alors $A^*(Au - b) = 0$, et donc $Au - b \in N(A^*) = (R(A))^\perp$. ainsi : $0 = P(Au - b) = Au - Pb$ ■

5] définitions :

> un vecteur $u \in H_1$ qui satisfait les conditions équivalentes (i), (ii), (iii) du théorème 1.7 est appelé une solution "au sens des moindres carrés" de l'équation $Ax = b$.

Remarquons que, comme $R(A)$ est fermé, une solution au sens des moindres carrés de l'équation $Ax = b$ existe, pour chaque $b \in H_2$ (puisque le théorème fondamental de décomposition d'un espace de Hilbert est applicable).
ainsi, si $N(A) \neq \{0\}$, alors il existe une infinité de solutions au sens des moindres carrés de $Ax = b$, étant donné que si u est une telle solution, alors $u+v$ en est une aussi pour chaque $v \in N(A)$.

Venons en maintenant au problème qui nous intéressait initialement. On voudrait trouver un "inverse" de l'opérateur A, associant à chaque $b \in H_2$ une solution $u \in H_1$ au sens des moindres carrés, solution univoquement déterminée.
Une manière naturelle de le faire est de remarquer que par le théorème 1.7, l'ensemble des solutions au sens des moindres carrés de (1.3) peut s'écrire comme :
$\{u \in H_1 : A^*Au = A^*b\}$.
Cet ensemble, par continuité et linéarité de A et de A^* est un ensemble convexe fermé.
Il contient donc un unique vecteur de norme minimale (par le théorème 1.3) ; on décide de choisir ce vecteur comme l'unique solution au sens des moindres carrés,

associée à b, par la voie du processus d'inversion généralisé.

> Soit $A \in \mathcal{L}(H_1, H_2)$, ayant une image fermée.
> Alors : $A^+ : H_2 \longrightarrow H_1$,
> définie par $A^+ b = u$, où u est l'unique solution au sens des moindres carrés de norme minimale de l'équation $Ax = b$,
> est appelé <u>l'inverse généralisé de A</u>.

Remarquons que nous restons cohérents avec ce que nous voulions. Si A est un opérateur inversible, on retrouve bien que $A^+ = A^{-1}$.
En effet, on a vu que u est une solution au sens des moindres carrés si $A^* A u = A^* b$. Or, A^+ est tel que $A^+ b = u$. Donc, $A^* A A^+ b = A^* b$,
d'où $A A^+ = Id$, et ainsi $A^+ = A^{-1}$.

6] <u>propriétés de l'opérateur A^+</u>

<u>Théorème 1.8</u>

> si $A \in \mathcal{L}(H_1, H_2)$ a une image fermée
> alors $R(A^+) = R(A^*) = R(A^+A)$

<u>démonstration</u> :

soit $b \in H_2$. Montrons d'abord que $A^+ b \in (N(A))^\perp = R(A^*)$ (a)
Pour le faire, supposons que $A^+ b = u_1 + u_2 \in (N(A))^\perp \oplus N(A)$

alors u_1 est une solution au sens des moindres carrés de $Ax = b$, étant donné que

$$\begin{aligned} Au_1 &= Au_1 + Au_2 && (\text{car } u_2 \in N(A)) \\ &= A(u_1 + u_2) && (\text{car } A \text{ linéaire}) \\ &= AA^+b \\ &= Au = Pb \end{aligned}$$

Supposons $u_2 \neq 0$. Par la propriété de Pythagore, $\|u_1 + u_2\|^2 = \|u_1\|^2 + \|u_2\|^2$, et donc $\|u_1\|^2 < \|u_1 + u_2\|^2 = \|A^+b\|^2$, en contradiction avec le fait que l'on a choisi pour définition de A^+ la solution de norme minimale.

Donc $u_2 = 0$, et on a bien que $\underline{A^+b = u_1 \in (N(A))^\perp}$

Supposons maintenant que $u \in (N(A))^\perp$.

Soit $b = Au$. Montrons que $A^+b = u$ (b)

On a certainement que $Au = PAu = Pb$, et donc u est une solution au sens des moindres carrés.

Montrons qu'elle est de norme minimale.

Si x est une autre solution au sens des moindres carrés, alors $Ax = Pb = Au$,

et, par conséquent, $x - u = \bar{u} \in N(A)$.

Comme $u \in (N(A))^\perp$, il suit que :

$$\begin{aligned} \|x - u + u\|^2 &= \|x\|^2 \\ &= \|x - u\|^2 + \|u\|^2 \\ &= \|\bar{u}\|^2 + \|u\|^2 \\ &\geq \|u\|^2 . \end{aligned}$$

Par conséquent, u est bien la solution au sens des moindres carrés de norme minimale, c-à-d. $\underline{u = A^+b}$

Les relations (a) et (b) que l'on vient de démontrer prouvent que la première relation ($R(A^+) = R(A^*)$)

En effet, montrer que $R(A^+) = R(A^*)$ revient à montrer que $R(A^+) = (N(A))^\perp$. Or,

en (a), on a pris $b \in H_2$, et on a montré que $A^+b \in (N(A))^\perp$
donc que $R(A^+) \subset (N(A))^\perp$

en (b), on a pris $u \in (N(A))^\perp$, et on a montré que $A^+b = u$, donc $u \in R(A^+)$,
et ainsi $(N(A))^\perp \subset R(A^+)$

Il reste à vérifier la seconde égalité annoncée.
Pour chaque $b \in H_2$,

$$A^+b = A^+Pb$$
$$= A^+Ax \quad \text{, pour un certain } x \text{ de } H_1.$$

Or, $A^+Ax \in R(A^+A)$,
et donc $R(A^+) = R(A^+A)$ ■

corollaire 1.3

> si $A \in \mathcal{L}(H_1, H_2)$, avec $R(A)$ fermé,
> alors $A^+ \in \mathcal{L}(H_2, H_1)$

démonstration :

Soit b, $\tilde{b} \in H_2$. alors $AA^+b = Au = Pb$
et $AA^+\tilde{b} = P\tilde{b}$

Donc, $AA^+b + AA^+\tilde{b} = Pb + P\tilde{b}$
$= P(b+\tilde{b})$
$= AA^+(b+\tilde{b})$,

et ainsi par le théorème 1.8,
$AA^+b + AA^+\tilde{b} - AA^+(b+\tilde{b}) = 0 = A^+b + A^+\tilde{b} - A^+(b+\tilde{b})$

Donc $A^+b + A^+\bar{b} - A^+(b+\bar{b}) \in (N(A))^\perp \cap N(A) = \{0\}$

On peut montrer de manière similaire que pour tout scalaire α, $A^+(\alpha b) = \alpha . A^+ b$.

Il reste à montrer que A^+ est borné. Pour cela, on se sert du lemme suivant :

Lemme 1.1

> si E_1, E_2 sont 2 espaces vectoriels normés complets, et si $A \in \mathcal{L}(E_1, E_2)$, avec $R(A)$ fermé, alors il existe un nombre $m > 0$ tel que $\|Ax\| \geq m . \|x\| \quad \forall x \in (N(A))^\perp$

Par le fait que $R(A^+) = R(A^*) = (N(A))^\perp$, et par le lemme cité, il existe un nombre positif m tel que $\|A\, A^+ b\| \geq m . \|A^+ b\| \quad \forall b \in H_2$.
Comme $AA^+ b = Pb$, il suit que
$\|b\| \geq \|Pb\| \geq m . \|A^+ b\|$, et A^+ est donc borné ■

Pour la démonstration du lemme 1.1, voir [6.1, p.14]

7] On sait que si u est une solution au sens des moindres carrés (donc $Au = Pb$), et si $N(A) \neq \{0\}$, alors $u + v$ est aussi une solution de ce type, avec $v \in N(A)$. Pour un tel u, $A^+ b = u$, donc $u \in R(A^+)$, qui est égal à $(N(A))^\perp$. Donc l'ensemble des solutions au sens des moindres carrés de $Ax = b$ a la forme :

$$\boxed{A^+ b + N(A)} \qquad (1.4)$$

Reprenons l'exemple développé en 1.3.1, 3].

On avait : $A = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}$, $b = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$.

On avait trouvé que $Pb = \begin{pmatrix} 1/2 \\ -1/2 \end{pmatrix}$, et que l'ensemble des solutions généralisées valait :

$$\left\{ (x_1, x_2) \text{ tel que } x_2 = x_1 + \frac{1}{2} \right\}.$$

Cherchons la solution de norme minimale.

$$\begin{aligned} \| (x_1, \tfrac{1}{2} + x_1) \|^2 &= \langle (x_1, \tfrac{1}{2} + x_1), (x_1, \tfrac{1}{2} + x_1) \rangle \\ &= 4 \cdot x_1^2 + 2 \cdot x_1 + \frac{1}{4} \end{aligned}$$

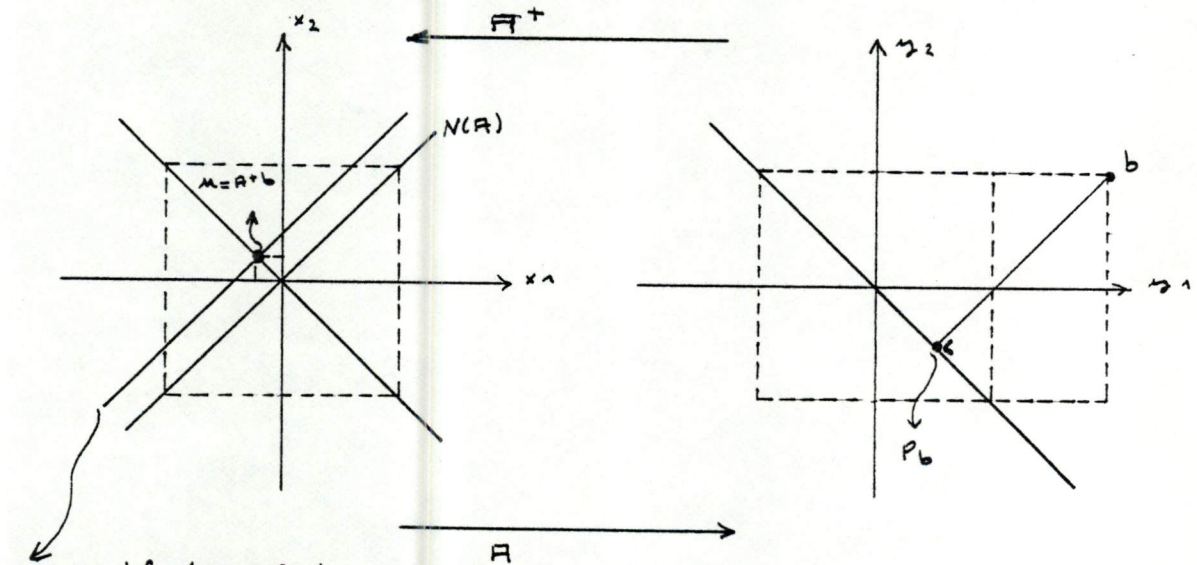
qui est nulle pour $x_1 = -\frac{1}{4}$.

Donc, $u = \left(-\frac{1}{4}, \frac{1}{4} \right)$

D'autre part,

$$\begin{aligned} N(A) &= \left\{ (x_1, x_2) \text{ t.q. } \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \right\} \\ &= \left\{ (x_1, x_2) \text{ t.q. } x_1 = x_2 \right\} \end{aligned}$$

Tenant compte de la remarque (1.4), on peut illustrer la situation par le schéma suivant :



ensemble des solutions au sens des moindres carrés.

On vérifie facilement que $A^+ = \begin{pmatrix} -\frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & -\frac{1}{4} \end{pmatrix}$

1.3.2. Image arbitraire

1] La distinction fondamentale entre le cas d'un opérateur avec image fermée et le cas d'un opérateur avec image non fermée est que l'inverse généralisé d'un opérateur avec image non fermée est un opérateur NON borné.

Si $A \in \mathcal{L}(H_1, H_2)$, avec $R(A)$ non nécessairement fermé, alors on ne peut plus toujours définir la projection P de H_2 sur $R(A)$, comme dans la section 1.3.1.

Bien sûr, en essayant de résoudre

$$Ax = b \qquad (1.5)$$

où $b \in H_2$ n'est pas nécessairement dans $R(A)$, on peut définir l'opérateur Q comme étant la projection de H_2 sur $\overline{R}(A)$, et alors désigner comme solutions généralisées de (1.5) toute solution u de l'équation :

$$Ax = Qb \qquad (1.6)$$

Mais cette équation peut très bien ne pas avoir de solution, puisque Qb peut ne pas appartenir à $R(A)$.

2] Essayons cependant de reprendre un raisonnement analogue au précédent (section 1.3.1);

ainsi : $\|Au - b\|^2$

$= \|Au - Qb + Qb - b\|^2$

$= \|Au - Qb\|^2 + \|Qb - b\|^2$

$\geq \|Au - Qb\|^2 + \|b - Au\|^2$ (pour (1.7)) ,

ce qui donne :

$$Au - b = Qb - b$$

et on sait que $Qb - b \in \overline{R(A)}^{\perp} = N(A^*)$,

ou encore $A^*Au = A^*b$ ■

définition :

> un vecteur $u \in H_1$ qui satisfait les conditions équivalentes (i), (ii), (iii) du théorème précédent est appelé une solution au sens des moindres carrés de l'équation $Ax = b$.

Il faut cette fois remarquer que, puisque l'hypothèse "$R(A)$ fermé" n'est pas satisfaite, il n'y a pas de solution au sens des moindres carrés de (1.5) si $Qb \notin R(A)$. Toutefois, le théorème suivant donne une condition d'existence pour les solutions :

Théorème 1.10

> une solution au sens des moindres carrés existe pour l'équation $Ax = b$
>
> ssi
>
> $b \in F_A$, où $F_A = \{ x : Qx \in R(A) \}$

démonstration : voir [T.3]

Bien sûr, pour tout vecteur b appartenant au sous espace $R(A) \oplus R(A)^{\perp} = \{ y + z \ , \ y \in R(A) \ , \ z \in R(A)^{\perp} \}$ de H_2, on peut résoudre l'équation (1.6). D'après le point (iii) du théorème 1.9, on peut voir que l'ensemble de toutes les solutions au sens des moindres carrés forme un ensemble convexe fermé, et par conséquent a un élément de norme minimale. On arrive naturellement à la définition suivante.

définition :

> soit $A \in \mathcal{L}(H_1, H_2)$, alors A^+, dont le domaine est $D(A^+) = R(A) \oplus R(A)^{\perp}$, défini pour $b \in D(A^+)$, par $A^+ b = u$, où u est une solution au sens des moindres carrés de norme minimale de l'équation $Ax = b$, est appelé l'inverse généralisé de A.

si $R(A)$ est fermé, alors $D(A^+) = H_2$; la nouvelle définition se ramène donc bien à celle du cas image fermée.
De même qu'à la section précédente, on peut montrer que

A^+ est linéaire

$N(A^+) = R(A)^{\perp}$

et $R(A^+) = N(A)^{\perp}$

Mais A^+ n'est pas borné. Le théorème suivant l'assure :

Théorème 1.9

> soit $A \in \mathcal{L}(H_1, H_2)$, et soit Q la projection de H_2 sur $\overline{R(A)}$;
> alors les conditions suivantes sur $u \in H_1$ sont équivalentes :
>
> (i) $Au = Qb$
>
> (ii) $\|Au - b\| \leq \|Ax - b\| \quad \forall x \in H_1$
>
> (iii) $A^*Au = A^*b$

Démonstration :

(i) $\Longrightarrow$ (ii)

(iii) $\Longrightarrow$ (i)

} même démonstration que celle du théorème 1.7 ; il suffit de remplacer P par Q.

(ii) $\Longrightarrow$ (iii)

(Remarquons qu'ici la démonstration du théorème 1.7 n'est pas applicable : on y supposait que $Ax = Pb$ pour un certain x de H_1. Ici, il n'existe pas de tel x si $Qb \notin R(A)$)

Comme $Qb \in \overline{R(A)}$, il existe une suite $\{x_n\}$ dans H_1 t.q.

$$Qb = \lim_{n} Ax_n .$$

Par conséquent,

$$\begin{aligned} \|b - Qb\|^2 &= \|b - \lim_{n} Ax_n\|^2 \\ &= \lim_{n} \|b - Ax_n\|^2 \\ &\geq \|b - Au\|^2 \end{aligned} \tag{1.7}$$

Comme $\overline{R(A)}$ est fermé,

$$H = \overline{R(A)} + \overline{R(A)}^{\perp}$$

et donc $Qb - b \in \overline{R(A)}^{\perp}$

Théorème 1.11

> si $A \in \mathcal{L}(H_1, H_2)$,
> alors A^+ est borné ssi $R(A)$ est fermé

démonstration :

$\Longleftarrow$ si $R(A)$ est fermé, alors A^+ est borné (corollaire 1.3)

$\Longrightarrow$ supposons $R(A)$ non fermé.
Comme $AA^+ y = Qy$ pour chaque $y \in D(A^+)$, et comme A^+ est borné, il y a un unique prolongement continu, noté $\hat{A}^+$, avec $\overline{D(A^+)} = H_2$.
Il suit que $A\hat{A}^+ y = Qy$, $y \in H_2$. Cela ne se peut que si $y \in \overline{R(A)}$. Mais $y \notin R(A)$ ■

3] Remarques :

* Il existe des méthodes itératives pour obtenir des solutions au sens des moindres carrés des équations linéaires algébriques.
D'autres dérivent des méthodes itératives pour calculer l'inverse généralisé d'une matrice.

* Le théorème, présenté ci-après sans démonstration, sera utilisé dans les chapitres ultérieurs. Le théorème est une condition nécessaire et suffisante d'existence d'une solution x de l'équation $Ax = b$.

Théorème 1.12

Considérons l'équation $Ax = b$, et la fonctionnelle

$$J(x) = \frac{1}{2} \cdot \langle Ax - b, Ax - b \rangle$$
$$= \frac{1}{2} \cdot \langle x, Qx \rangle + \langle x, \tilde{b} \rangle + \langle b, b \rangle$$

où $Q = A^T A$ et $\tilde{b} = -A^+ b$;

$\bar{x}$ est solution de $Ax = b$

ssi

$\bar{x}$ est une solution au sens des moindres carrés de $Ax = b$

et $J(\bar{x}) = 0$.

démonstration : [C.1]

1.3.3. méthode de la plus grande descente

1] L'idée de la méthode de la plus grande descente, proposée par Cauchy, pour la résolution de systèmes d'équations non linéaires, est la suivante :
on se propose d'interpréter la solution d'un système d'équations non linéaires comme un élément minimisant une certaine fonctionnelle non négative.
On construit alors une suite itérative de telle sorte qu'en passant d'une approximation à la suivante, on se dirige dans la direction de la décroissance la plus rapide de la fonctionnelle.
Dans cette section 1.3.3., on veut approximer $A^+ b$, où $A \in \mathcal{L}(H_1, H_2)$, $R(A)$ est fermé, $b \in H_2$, en utilisant une méthode de plus grande pente.

2] définition :

soient E_1, E_2 deux espaces vectoriels normés, et supposons que X soit un sous-ensemble ouvert de E_1. Une fonction $f : X \longrightarrow E_2$ est dite Fréchet différentiable en $x_0 \in X$ si il existe un opérateur $f'(x_0) \in \mathcal{L}(E_1, E_2)$ tel que :

$$\lim_{\|h\| \to 0} \frac{\| f(x_0 + h) - f(x_0) - f'(x_0).h \|}{\|h\|} = 0$$

On peut de plus montrer que l'opérateur $f'(x_0) \in \mathcal{L}(E_1, E_2)$ satisfaisant cette définition est unique.

On peut remarquer que pour une fonction réelle $f : \mathbb{R} \longrightarrow \mathbb{R}$, la définition précédente coïncide avec la notion usuelle de dérivée d'une fonction, si on considère $f'(x_0)$ non comme un nombre, mais plutôt comme l'opérateur correspondant dans $\mathcal{L}(\mathbb{R}, \mathbb{R})$, donné par : $x \longrightarrow f'(x_0)x$
La différentiation au sens de Fréchet est linéaire ; donc, si $f, g : E_1 \longrightarrow E_2$ sont différentiables en $x_0 \in E_1$, alors $(f+g)'(x_0) = f'(x_0) + g'(x_0)$,
et $(\alpha . f)'(x_0) = \alpha . f'(x_0)$

Si H est un espace de Hilbert, et $f : H \longrightarrow \mathbb{R}$ est une fonction à valeurs réelles, Fréchet différentiable sur H, alors pour tout $x_0 \in H$, $f'(x_0)$ est une fonctionnelle linéaire, continue sur H.
Par le théorème de représentation de Riesz, il existe un vecteur unique dans H, que l'on notera par $\nabla f(x_0)$,

et qui sera appelé le gradient de f en x_0, tel que

$$f'(x_0).h = \langle h, \nabla f(x_0) \rangle \quad \forall h \in H.$$

définition :

> La dérivée directionnelle d'une fonctionnelle $f : H \longrightarrow \mathbb{R}$ en x_0 dans la direction $h \neq 0$ est le nombre
> $$D f(x_0, h) = \lim_{t \longrightarrow 0^+} \frac{f(x_0 + t.h) - f(x_0)}{t}$$ si la limite existe.

En fait, si f est différentiable en x_0, on a :

$$\lim_{t \longrightarrow 0} \frac{f(x_0 + h) - f(x_0) - \langle \nabla f(x_0), h \rangle}{\| h \|} = 0$$

$$\Longrightarrow \lim_{\lambda \longrightarrow 0} \frac{f(x_0 + \lambda . d) - f(x_0) - \langle \nabla f(x_0), \lambda . d \rangle}{|\lambda| . \| d \|} = 0$$

$$\Longrightarrow \lim_{\lambda \longrightarrow 0} \frac{f(x_0 + \lambda . d) - f(x_0) - \langle \nabla f(x_0), \lambda . d \rangle}{|\lambda|} = 0$$

$$\Longrightarrow \lim_{\substack{\lambda \longrightarrow 0 \\ + }} \frac{f(x_0 + \lambda . d) - f(x_0) - \langle \nabla f(x_0), \lambda . d \rangle}{\lambda} = 0$$

$$\Longrightarrow \lim_{\substack{\lambda \longrightarrow 0 \\ + }} \frac{f(x_0 + \lambda . d) - f(x_0)}{\lambda} = \langle \nabla f(x_0), d \rangle$$

et donc

$$\boxed{D f(x_0, h) = \langle h, \nabla f(x_0) \rangle} \qquad (1.8)$$

3] méthode générale :

Supposons que H est un espace de Hilbert, et que $f : H \longrightarrow \mathbb{R}$ est une fonctionnelle non négative. On cherche un point $x^* \in H$ tel que $f(x^*) = \inf \{ f(x) : x \in H \}$.
On suppose que F est Fréchet différentiable en chaque point de H.
Donnant une approximation initiale x_0, on essaye de minimiser f en se déplaçant dans la direction de la plus grande décroissance de f. On doit donc choisir une direction $z \in H$ telle que la dérivée directionnelle est la plus petite possible.
Or, on a vu (1.8) que $D f(x_0, z) = \langle z, \nabla f(x_0) \rangle$.
Par l'inégalité de Cauchy-Schwarz,

$$- \| z \| . \| \nabla f(x_0) \| \leq \langle z, \nabla f(x_0) \rangle ,$$

avec l'égalité seulement si z est un multiple positif de $-\nabla f(x_0)$. Il suit donc que le point de départ x_0 de direction de décroissance la plus rapide de f se trouve dans la direction de $-\nabla f(x_0)$.
Par conséquent, on construit une nouvelle approximation x_1 en partant de x_0, et en prenant un pas $\alpha_0 > 0$, dans la direction $-\nabla f(x_0)$, c-à-d.

$$x_1 = x_0 - \alpha_0 . \nabla f(x_0)$$

Le paramètre α_0 est choisi de telle manière que x_1 donne le minimum de f sur la droite reliant x_0 à la direction $-\nabla f(x_0)$; en d'autres termes, α_0 doit satisfaire :

$$\frac{d}{d\alpha} f(x_0 - \alpha . \nabla f(x_0)) \Big|_{\alpha = \alpha_0} = 0$$

Dès lors, la suite itérative $\{x_n\}$ est générée par

$$x_{n+1} = x_n - \alpha_n . \nabla f(x_n)$$

où $\dfrac{d}{d\alpha} f(x_n - \alpha . \nabla f(x_n)) \Big|_{\alpha = \alpha_n} = 0$

Supposons maintenant que $A \in \mathcal{L}(H_1, H_2)$ ait une image fermée. Etant donné que les solutions au sens des moindres carrés de $Ax = b$ sont exactement le minimum de la fonctionnelle

$$f(x) = \frac{1}{2} \cdot \|Ax - b\|^2 \qquad \text{(théorème 1.7)},$$

on va approximer les solutions au sens des moindres carrés en appliquant la méthode de plus grande descente à f.
On,

$$\frac{1}{2} \cdot \|Ax-b\|^2 = \frac{1}{2} \cdot [\langle Ax-b, Ax-b \rangle]$$

$$= \frac{1}{2} \cdot [\langle Ax, Ax \rangle - \langle Ax, b \rangle - \langle b, Ax \rangle + \langle b, b \rangle]$$

$$= \frac{1}{2} \cdot [\langle x, A^*Ax \rangle - \langle x, A^*b \rangle - \langle A^*b, x \rangle$$

$$+ \langle b, b \rangle]$$

D'où $\nabla f(x) = A^*Ax - A^*b$
et la valeur optimale de α est donnée par

$$\alpha = \frac{\|r\|^2}{\|Ar\|^2} \quad , \text{ où } r = A^*Ax - A^*b$$

On considérera ainsi la suite $\{x_n\} \subset H_1$ générée par les relations :

$$x_{n+1} = x_n - \alpha_n \cdot r_n$$

$$\text{où} \begin{cases} \alpha_n = \dfrac{\|r_n\|^2}{\|Ar_n\|^2} \\ r_n = A^*Ax_n - A^*b \end{cases}$$

Remarquons que si $r_m = 0$, alors x_m est une solution au sens des moindres carrés, et la méthode se termine à x_m. On supposera donc que $r_m \neq 0$ pour tout m.
On peut alors démontrer les résultats suivants.

Lemme 1.2

> si $A \in \mathcal{L}(H_1, H_2)$
> alors $\lim_{m \to \infty} r_m = 0$

démonstration : [N-1]

* commentaire : ce lemme est vérifié pourtant A appartenant à $\mathcal{L}(H_1, H_2)$; aucune hypothèse n'est donc formulée à propos de $R(A)$.

théorème 1.13

> si $A \in \mathcal{L}(H_1, H_2)$ a une image fermée,
> alors la suite générée par la méthode de la plus grande descente converge vers une solution au sens des moindres carrés de $Ax = b$ pour chaque $x_0 \in H_1$.
> la suite converge vers $A^+ b$ ssi $x_0 \in R(A^*)$

démonstration : [N-1, G-1]

1.4. inverse d'un opérateur. [S.2]

1] Une transformation A est injective si: $Af = Ag \Rightarrow f = g$. Une transformation injective a un inverse A^{-1}, dont le domaine est l'image de A et dont l'image est le domaine de A. Limitons-nous à l'étude des transformations linéaires. Le théorème suivant est une caractérisation nécessaire et suffisante pour qu'un opérateur linéaire A soit injectif.

théorème 1.14.

> soit l'opérateur A linéaire,
> A est injectif ssi $\{ Ax = 0 \Longrightarrow x = 0 \}$

Le principal problème dans la théorie des opérateurs est la résolution de l'équation non homogène $Ax = f$, A étant une transformation donnée, et f un élément donné d'un espace de Hilbert H.

2] définitions

> La transformation A est régulière si les conditions suivantes sont satisfaites :
> a) $Ax = 0$ a seulement la solution nulle
> b) $R(A) = H$
> c) A^{-1} est borné.

* commentaire :

a) garantit que l'équation $Ax = f$ a au plus une solution, donc que A^{-1} existe

b) affirme que pour chaque $f \in H$, il y a au moins une solution de l'équation $Ax = f$

c) garantit que si f_1 et f_2 sont proches, disons $\|f_1 - f_2\| < \varepsilon$, les solutions respectives x_1 et x_2 sont proches, c-à-d. $\|x_1 - x_2\| < \|A^{-1}\| . \varepsilon$

définition :

> une transformation A est essentiellement régulière si les conditions a) et c) sont vérifiés.
> Cependant, $\overline{R(A)} = H$

Ainsi, comme A^{-1} est borné et défini sur un ensemble dense dans H, il peut être étendu à tout H par continuité.
Notons cette extension B. B est inversible, et son inverse B^{-1} est une extension de A. De cette manière, une transformation essentiellement régulière peut être étendue par continuité à une transformation régulière.

définition :

> A est singulière si A n'est ni régulière, ni essentiellement régulière.

3] classification des opérateurs singuliers

Si A est singulier, cela veut dire qu'une des 3 possibilités suivantes est vérifiée :

1. $Ax = 0$ a une autre solution que 0 (donc A^{-1} n'existe pas)
2. $Ax = 0$ a la seule solution triviale, mais A^{-1} n'est pas borné et $\overline{R(A)} = H$
3. $Ax = 0$ a la seule solution triviale, mais $\overline{R(A)} \neq H$ (A^{-1} peut être borné ou non)

Il faut remarquer que dans un espace de Hilbert H de dimension infinie, un opérateur peut être singulier de différentes manières. Dans un espace de dimension finie, ce n'est pas le cas.
En effet, soit A un opérateur linéaire tel que l'équation homogène $Ax = 0$ a la seule solution triviale nulle. Dans un espace de dimension finie,
$R(A) = H$ si $Ax = 0$ a seulement la solution nulle (1.9)
Mais dans un espace H de dimension infinie,
il est possible que $R(A) \neq H$ même si $Ax = 0$ a la seule solution nulle. (1.10)

démonstration de (1.9)

Supposons que $Ax = 0$ a seulement la solution nulle, et montrons que $R(A) = \mathbb{R}^n$ (on suppose que $H = \mathbb{R}^n$) tout entier.
Soit $\{e_1, \dots, e_n\}$ une base de $\mathbb{R}^n$. Montrons d'abord que les vecteurs $Ae_1, \dots, Ae_n$ forment aussi une base de $\mathbb{R}^n$.
Si cela n'était pas vrai, il existerait des constantes $d_1 \dots d_n$

non toutes nulles t.q. $\sum_{k=1}^{n} \alpha_k . A e_k = 0$, ou encore

$$A \left(\sum_{k=1}^{n} \alpha_k . e_k \right) = 0 .$$

Comme $Ax = 0$ a la seule solution nulle, on a :
$\sum_{k=1}^{n} \alpha_k . e_k = 0$. Comme $\{ e_1, \ldots, e_n \}$ forme une base, on doit avoir $\alpha_1 = \ldots = \alpha_k = \ldots = \alpha_n = 0$, ce qui est contraire à l'hypothèse. Donc les vecteurs $Ae_1, \ldots, Ae_n$ forment une base.

Maintenant, tout vecteur f dans $\mathbb{R}^n$ peut s'écrire $f = \sum_{i=1}^{n} \beta_i . Ae_i$, et une solution de $Ax = f$ est donc $x = \sum_{i=1}^{n} \beta_i . e_i$. Cette solution est unique. ■

Donnons maintenant un exemple de (1.10)

Soit $H = \ell_2 [0, 1]$. Considérons un ensemble complet orthonormal $\varphi_1(t), \ldots, \varphi_n(t), \ldots$ sur H. Si $x(t)$ est un élément arbitraire de H, on a :

$x = a_1 . \varphi_1 + a_2 . \varphi_2 + \ldots$

$= \sum_{k=1}^{\infty} a_k . \varphi_k$, où $a_k = \langle x , \varphi_k \rangle$.

On définit la transformation linéaire A par

$Ax = y = \sum_{k=1}^{\infty} a_k . \varphi_{k+1}$ (on a donc déplacé les coefficients de Fourier pour obtenir y).

On sait que la condition nécessaire et suffisante pour que $x \in H$ est que $\sum_{k=1}^{\infty} |a_k|^2$ converge.

Donc, si $x \in H$, alors $y \in H$. En outre, $\| x \| = \| y \|$.

Ainsi, A est borné et $\| A \| = 1$. L'image de A est l'ensemble fermé constitué par tous les vecteurs de H ayant 0 pour composante de φ_1. Donc $R(A) \neq H$.

Mais néanmoins, si $f \in R(A)$,

$f = \sum_{k=2}^{\infty} b_k \cdot \varphi_k$, où $\sum_{k=2}^{\infty} |b_k|^2$ converge.

Donc $Ax = f$ a l'unique solution $x = \sum_{k=1}^{\infty} b_k + \varphi_k$.

1.5. le spectre d'un opérateur [S-2, p. 180]

1] La principale difficulté dans l'étude du spectre d'un opérateur A, défini sur un espace de Hilbert H de dimension infinie, est le fait qu'un opérateur peut être singulier de plusieurs manières.
Soit A un opérateur linéaire fermé, défini sur un domaine D(A) d'un espace de Hilbert H. L'opérateur $A - \lambda . I$, où λ est arbitraire et I est l'opérateur identité, est aussi fermé. Il est défini sur le même domaine D(A). L'image de $A - \lambda I$ peut dépendre de λ, et c'est pour cela qu'on la note $R_\lambda(A)$.
On peut alors donner une classification de tous les points λ du plan complexe suivant que $A - \lambda I$ est régulier ou singulier.

2] définitions

D.1. | Une valeur de λ pour laquelle $A - \lambda I$ est régulière est appelée valeur régulière. L'ensemble des valeurs régulières est appelé ensemble résolvant de A.
Toutes les valeurs de λ qui ne sont pas dans cet ensemble forment le spectre de A.
Donc, λ est dans le spectre de A ssi $A - \lambda I$ est singulier.
Les points du spectre de A seront classés suivant la façon dont l'opérateur fermé $A - \lambda I$ est singulier.

D.2. | si $(A - \lambda I).x = 0$ a une solution non triviale, λ est appelée une valeur propre de A. Toute solution

non triviale x est un vecteur propre de A correspondant à λ. L'ensemble de toutes les valeurs propres forment le spectre ponctuel de A.
L'espace vectoriel des vecteurs propres correspondant à λ a une certaine dimension : on l'appelle la multiplicité de λ.

D.3. Si $(A-\lambda I)x = 0$ a l'unique solution triviale, mais que $(A-\lambda I)^{-1}$ est non borné, et que $R_\lambda(A) = H$, ou $\overline{R_\lambda(A)} = H$, alors λ appartient au spectre continu de A

D.4. Si $(A-\lambda I).x = 0$ a l'unique solution triviale, mais que $R_\lambda(A) \neq H$, λ appartient au spectre résiduel de A.

3] théorème 1.15

> Si A est symétrique, $\langle Ax, x \rangle$ est réel $\forall x \in D(A)$, et toutes les valeurs propres de A sont réelles

démonstration :

A est symmétrique, donc $\langle Ax, x \rangle = \langle x, Ax \rangle \ \forall x \in D(A)$.
Par une des propriétés du produit $\langle .,. \rangle$, on a aussi que $\langle Ax, x \rangle = \overline{\langle x, Ax \rangle}$. Donc $\langle x, Ax \rangle = \overline{\langle x, Ax \rangle}$; d'où $\langle Ax, x \rangle$ est réel $\forall x \in D(A)$.

Si λ est une valeur propre de A, alors il existe un vecteur x non nul tel que $Ax = \lambda . x$.
Par conséquent, $\langle Ax , x \rangle = \lambda . \langle x , x \rangle$. Mais $\langle Ax , x \rangle$ est réel, et $\langle x , x \rangle$ est > 0. Donc λ est réel. ■

théorème 1.16

> Le spectre continu d'un opérateur symmétrique A est situé sur l'axe réel.

démonstration :

La démonstration se base sur les deux lemmes suivants :

lemme 1.3

> si B^{-1} est non borné, il existe une suite $\{x_n\}$ dans $D(B)$ tel que $\|x_n\| = 1$ et $\|Bx_n\| < \frac{1}{n}$

lemme 1.4

> si A est symmétrique, et si $\lambda = \xi + i.\mu$, où ξ et μ sont réels, alors $\|(A - \lambda I).x\|^2 \geq \mu^2 . \|x\|^2$

Soit $\lambda = \xi + i.\mu$, avec $\mu \neq 0$. Soit $\{x_n\}$ une suite quelconque dans $D(A)$, avec $\|x_n\| = 1$. Par le lemme 1.4, $\|(A - \lambda I)x_n\| \geq |\mu|$. Etant donné que cette inégalité est vraie pour toute suite $\{x_n\}$ t. q. $\|x_n\| = 1$,

le lemme 1.3 affirme que $(A - \lambda I)^{-1}$ doit être borné, et donc que λ ne peut pas être dans le spectre continu. ■

* commentaire : on peut résumer les théorèmes 1.15 et 1.16 comme suit : si A est symmétrique, il peut y avoir des points non réels dans le spectre résiduel, mais les spectres ponctuels et continus sont réels.

Théorème 1.17

> le spectre d'un opérateur autoadjoint est entièrement situé sur l'axe réel, et le spectre résiduel est vide

démonstration :

Un opérateur autoadjoint est symmétrique. Par les deux théorèmes précédents, il suffit donc de montrer que le spectre résiduel est vide.
Considérons λ appartenant au spectre résiduel de A. On a par définition (D.4) que $\overline{R_\lambda(A)} \neq H$. L'espace linéaire fermé $(R_\lambda(A))^\perp$ a donc une dimension positive. On l'appelle la défficience de λ. On se sert du lemme 1.4 pour conclure.

lemme 1.4

> soit λ un point de défficience m dans le spectre résiduel de A, alors λ est une valeur propre de multiplicité m de A*

Ici, $A = A^*$. Comme A est symmétrique, toutes les valeurs propres de A sont réelles (théorème 1.15), et par le lemme 1.4, le spectre résiduel est réel. Mais si λ est une valeur réelle dans le spectre résiduel de A, il doit aussi être une valeur propre de A^* (par le lemme 1.4), donc de A. Or, par définition, un point dans le spectre résiduel ne peut pas être une valeur propre. ■

démonstration du lemme 1.4

Soit $x \in \mathcal{D}(A)$ et $y \in [R_\lambda(A)]^\perp$. On a dès lors : $\langle (A - \lambda I).x , y \rangle = 0 = \langle x, 0 \rangle$. Le couple $(y, 0)$ est un couple admissible pour $(A - \lambda I)$, d'où on peut conclure que $y \in \mathcal{D}((A-\lambda I)^*)$, et que $(A-\lambda I)^* y = 0$. Comme $(A-\lambda I)^* = A^* - \overline{\lambda}.I$, c'est démontré ■

4] Supposons maintenant que A soit un opérateur borné

théorème 1.18

> si A est un opérateur borné sur H,
> si λ est dans le spectre de A,
> alors $|\lambda| \leq \|A\|$

démonstration :

(i) soit λ dans le spectre ponctuel de A. Il existe donc un élément x, avec $\|x\| \neq 0$, tel que $Ax - \lambda x = 0$.

ainsi,

$$\lambda . \langle x, x \rangle = \langle Ax, x \rangle \quad \text{et} \quad \lambda = \frac{\langle Ax, x \rangle}{\langle x, x \rangle} = \frac{\langle Ax, x \rangle}{\|x\|^2}$$

Or, par l'inégalité de Cauchy-Swarz,

$|\langle Ax, x \rangle| \leq \|Ax\| . \|x\| \leq \|Ax\| . \|x\|^2$

d'où $|\lambda| \leq \|A\|$ (1.11)

(ii) soit λ dans le spectre résiduel de A, $\overline{\lambda}$ est une valeur propre de A^* (par le lemme 1.4). Donc $|\overline{\lambda}| \leq \|A^*\|$.
Comme $\|A^*\| = \|A\|$, et $|\overline{\lambda}| = |\lambda|$. On a aussi que $|\lambda| \leq \|A\|$.

(iii) soit λ dans le spectre continue de A, alors $(A - \lambda I)^{-1}$ existe et est non borné. Donc il existe une suite $\{x_n\}$, avec $\|x_n\| = 1$, tel que $Ax_n - \lambda . x_n \longrightarrow 0$ (par le lemme 1.3). Par conséquent,

$\lambda = \lim_{n \to \infty} \langle Ax_n, x_n \rangle$, et ainsi $|\lambda| \leq \|A\|$

(par (1.11) avec $\|x_n\| = 1$) ■

théorème 1.13

> Pour tout $x \neq 0$, si A est borné,
> $$\frac{|\langle Ax, x \rangle|}{\|x\|^2} \leq \|A\|$$

démonstration.

Par l'inégalité de Schwarz, $|\langle Ax, x \rangle| \leq \|Ax\| . \|x\|$
$\leq \|A\| . \|x\|^2$. ■

* commentaire : à partir de ce théorème, on peut évidemment dire que

$$M_A = \sup_{x \neq 0} \frac{|\langle Ax, x \rangle|}{\|x\|^2} \leq \|A\|$$

théorème 1.20

> si A est borné, symmétrique
> alors $M_A = \|A\|$

démonstration.

Par la remarque précédente, on sait que $M_A \leq \|A\|$.
Il reste donc à montrer que $M_A \geq \|A\|$.
Partons de l'identité suivante :

$$\langle A(x+y), x+y \rangle - \langle A(x-y), x-y \rangle = 2 \cdot \langle Ax, y \rangle + 2 \cdot \langle Ay, x \rangle$$

En utilisant la définition de M_A, on a :

$$\langle A(x+y), x+y \rangle \leq M_A \cdot \|x+y\|^2 \quad \text{et}$$

$$\langle A(x-y), x-y \rangle \geq -M_A \cdot \|x-y\|^2$$

En soustrayant les 2 inégalités, on a :

$$2 \cdot \langle Ax, y \rangle + 2 \cdot \langle Ay, x \rangle \leq M_A \cdot (\|x+y\|^2 + \|x-y\|^2),$$

ou encore en utilisant la loi du parallélogramme

(donc $\|x+y\|^2 + \|x-y\|^2 = 2 \cdot \|x\|^2 + 2 \cdot \|y\|^2$),

$$\langle Ax, y \rangle + \langle Ay, x \rangle \leq M_A \cdot (\|x\|^2 + \|y\|^2)$$

si $Ax = 0$, remplaçons y par $\left(\frac{Ax}{\|Ax\|}\right) \cdot \|x\|$
pour obtenir :

$$\langle Ax, Ax \rangle \cdot \frac{\|x\|}{\|Ax\|} + \langle AAx, x \rangle \cdot \frac{\|x\|}{\|Ax\|} \leq M_A \cdot (\|x\|^2 + \|x\|^2)$$

ou encore

$\langle Ax, Ax \rangle + \langle AAx, x \rangle \leq 2 . \|Ax\| . \|x\| . M_A$

Mais $\langle AAx, x \rangle = \langle Ax, Ax \rangle$,

d'où $2 . \|Ax\|^2 \leq 2 . \|Ax\| . \|x\| . M_A$,

ou encore $\|Ax\| \leq M_A . \|x\|$.

Donc $\|A\| \leq M_A$ (puisque l'inégalité précédente est valable, même si $Ax = 0$) ∎

Théorème 1.21

> si A est complètement continu et symmétrique, au moins un des nombres $\|A\|$, $-\|A\|$ est une valeur propre de A. En outre, aucune autre valeur propre de A n'a de valeur absolue plus grande.

démonstration :

Elle se base de nouveau sur un lemme.

lemme 1.5

> si A est symmétrique, il existe une suite $\{x_k\}$, avec $\|x_k\| = 1$, pour laquelle $\lim_{k \to \infty} (Ax_k - \lambda_1 . x_k) = 0$, avec λ_1 est l'un des nombres $\|A\|$ ou $-\|A\|$

Considérons cette suite définie par le lemme. Par la définition d'un opérateur complètement continu,

la suite $\{x_k\}$ contient une sous-suite $\{u_k\}$ telle que Au_k converge. Comme $Au_k - \lambda_1 . u_k \longrightarrow 0$, il suit que $\{u_k\}$ converge et qu'il existe un élément u de norme unité telle que $Au - \lambda_1 . u = 0$ ■

démonstration du lemme 1.5

Remarquons d'abord que ce lemme ne dit pas que λ_1 est une valeur propre de A. Une telle conclusion est vraie seulement si $\{x_k\}$ ou une sous-suite de $\{x_k\}$ converge vers un élément x. Cela impliquerait en effet que $Ax - \lambda_1 . x = 0$.

Par le théorème 1.20, on sait que $\sup_{x \neq 0} \frac{|\langle Ax, x \rangle|}{\|x\|^2} = \|A\|$, ce qui implique

$\sup_{\|x\| = 1} |\langle Ax, x \rangle| = \|A\|$.

Donc il existe une suite $\{z_k\}$, avec $\|z_k\| = 1$, telle que $\lim |\langle Az_k, z_k \rangle| = \|A\|$.

Comme $\langle Az_k, z_k \rangle$ est réel, la suite $\{z_k\}$ doit contenir une sous-suite $\{x_k\}$ telle que $\langle Ax_k, x_k \rangle \longrightarrow \|A\|$ ou $\langle Ax_k, x_k \rangle \longrightarrow -\|A\|$.

Donc la suite a la propriété $\langle Ax_k, x_k \rangle \longrightarrow \lambda_1$, où λ_1 vaut $\|A\|$ ou $-\|A\|$. On a aussi :

$$\begin{aligned} \|Ax_k - \lambda_1 . x_k\|^2 &= \|Ax_k\|^2 + \lambda_1^2 . \|x_k\|^2 - 2 . \lambda_1 . \langle Ax_k, x_k \rangle \\ &\leq \|A\|^2 + \lambda_1^2 - 2\lambda_1 . \langle Ax_k, x_k \rangle \\ &= 2 . \lambda_1^2 - 2\lambda_1 . \langle Ax_k, x_k \rangle \end{aligned}$$

Comme $\langle Ax_k, x_k \rangle \longrightarrow \lambda_1$,

$\|Ax_k - \lambda . x_k\|^2 \longrightarrow 0$, et $Ax_k - \lambda_1 . x_k \longrightarrow 0$ ■

Chapitre 2 :

problèmes singuliers et réguliers - cas régulier, espaces de dimension finie

2.1. introduction

2.2. algorithme du D.F.P.

2.3. algorithmes à directions conjuguées

Chapitre 2 : problèmes singuliers et réguliers - cas régulier, espaces de dimension finie.

Les théorèmes et les remarques de ce chapitre seront, pour la plupart, repris et explicités dans la suite de l'exposé. Leur présence dans ce chapitre ne se justifie que pour une meilleure compréhension des problèmes qui se posent et des différences fondamentales qu'il existe entre un problème singulier et un problème non singulier.

2.1. introduction [C_2]

1] Beaucoup de conditions suffisantes pour le problème de contrôle optimal et pour l'étude de la convergence des algorithmes se basent sur une extension de la solution optimale au voisinage de l'optimum. Ceci justifie l'étude du problème quadratique linéaire suivant (noté L.Q.P.) :

$$\min J(x) = \frac{1}{2} \cdot \int_{t_0}^{t_f} [x^T . P(t) . x + R(t) . u^2] \, dt \qquad (2.1)$$

$$\text{s.c.} \begin{cases} \dot{x} = F(t) . x + G(t) . u \\ x(t_0) = x_0 \end{cases} \qquad (2.2)$$

où t_0, t_f sont donnés, x est un n-vecteur

et u un scalaire

Tous les résultats que l'on verra pourront facilement être généralisés au cas où le contrôle $u(.)$ est un vecteur, et non plus un scalaire.

La procédure utilisée est la suivante :
transformer les équations (2.1) et (2.2) en un problème de minimisation quadratique fonctionnel sans contrainte.
Comme l'équation (2.2) est linéaire, on sait que la solution générale de (2.2) peut s'écrire :

$$x(t) = \Phi(t, t_0).x_0 + \int_{t_0}^{t_f} \Phi(t, \tau).G(\tau).u(\tau)\,d\tau \qquad (2.3)$$

où $\Phi(t,\tau)$ est la matrice de transition du système (2.2), c-à-d. $\Phi(t,\tau)$ est solution de l'équation différentielle matricielle :

$$\frac{\delta \Phi(t,\tau)}{\delta t} = F(t).\Phi(t,\tau) \text{ , soumis à } \Phi(\tau,\tau) = Id$$

(2.3) peut encore s'écrire :

$$\underline{x(t) = Sx_0 + Tu} \qquad (2.4)$$

où

- $S : \mathbb{R}^n \longrightarrow L_2^n[t_0, t_f]$
- $T : L_2[t_0, t_f] \longrightarrow L_2^n[t_0, t_f]$

(S et T sont des opérateurs linéaires)

Rappelons que $L_2^l[t_0, t_f] = \left\{ z(.) \in \mathbb{R}^l : \int_{t_0}^{t_f} z(t)^T.z(t)\,dt < \infty \right\}$

Considérons le produit interne usuel dans $L_2^l[t_0, t_f]$, c-à-d.

$$\underline{\langle a(t), b(t) \rangle = \int_{t_0}^{t_f} a(t)^T.b(t)\,dt} \qquad (2.5)$$

où $a(t), b(t) \in L_2^l[t_0, t_f]$

Remplaçons maintenant (2.4) dans (2.1), en tenant compte de (2.5) ; on obtient la fonction objective exprimée sous la forme :

$$J[u] = \frac{1}{2} . \langle u, Au \rangle + \langle u, w \rangle + J_0 \qquad (2.6)$$

où
$$\begin{cases} A = T^* P T + R & (2.7) \\ w = T^* . P . S x_0 & (2.8) \\ J_0 = \frac{1}{2} . \langle S x_0, P S x_0 \rangle & (2.9) \end{cases}$$

Notons le problème de minimiser (2.6) s.c. (2.7), (2.8) et (2.9) par (P).
L'opérateur linéaire T^* est l'opérateur adjoint défini par : $\langle a(t), T . b(t) \rangle = \langle T^* a(t), b(t) \rangle$.

Le problème (P) (minimisation d'une fonction quadratique du contrôle u) est donc équivalent au L.Q.P. avec les contraintes (2.2).
On peut facilement montrer que l'opérateur A (défini en (2.7)) est un opérateur autoadjoint.
Dans la suite, on appellera A l'opérateur de seconde variation, ou encore le Hessien du problème (P).

Remarquons enfin que l'avantage majeur de la formulation (2.6) au lieu de la formulation initiale (2.1) et (2.2) est que toutes les conditions d'existence et les conditions suffisantes de convergence des algorithmes se rapporteront uniquement à des caractérisations des opérateurs linéaires A et w.

2] Il est évident que l'existence et l'unicité d'un minimum pour le problème (P) dépendront fortement des propriétés de l'opérateur A. Trois types d'opérateurs joueront un rôle important dans les développements ultérieurs. Soit A un opérateur linéaire, autoadjoint défini sur $L_2[t_0, t_f]$;

définition 2.1

(1) A est semi défini positif (noté $A \geq 0$)
si $\langle z, Az \rangle \geq 0 \qquad \forall z \in L_2[t_0, t_f]$

(2) A est positif (noté $A > 0$)
si $\langle z, Az \rangle > 0 \qquad \forall z \neq 0 \ , z \in L_2[t_0, t_f]$

(3) A est fortement positif
si $\exists m > 0$ t.q. $m \cdot \langle z, z \rangle \leq \langle z, Az \rangle \quad \forall z \neq 0$, $z \in L_2[t_0, t_f]$

Remarque :

Il est évident que A fortement positif $\Longrightarrow$ A positif. Dans un espace fini dimensionnel, positivité et forte positivité sont deux notions équivalentes. En effet, soit la sphère unité de $\mathbb{R}^n$, $S = \{x \in \mathbb{R}^n : |x| \leq 1\}$. S est compacte (fermée et bornée). On sait que $Q(x) = \langle x, Ax \rangle$ est la forme quadratique associée à la matrice A. $Q(x)$ est fortement continue, puisque définie à partir du produit scalaire. Elle est donc en particulier continue sur S : $\exists y, z \in S$ t.q. $Q(y) \leq Q(x) \leq Q(z) \ \forall x \in S$. Soit $x \in \mathbb{R}^n$; $\dfrac{x}{|x|}$ est un vecteur unitaire $\in S$, et donc :

$$Q\left(x/|x|\right) = \sum_i \sum_j a_{ij} \cdot \frac{x_i}{|x|} \cdot \frac{x_j}{|x|} = \frac{1}{|x|^2} \cdot Q(x).$$

Donc, si Q est une forme quadratique, il existe des éléments $y, z \in \mathbb{R}^n$ tel que $|y| = |z| = 1$, avec
$Q(y) \cdot |x|^2 \leq Q(x) \leq Q(z) \cdot |x|^2 \quad \forall x \in \mathbb{R}^n$.
Si de plus la forme quadratique est définie positive, on choisit $\begin{cases} m = Q(y) > 0 \\ n = Q(z) > 0 \end{cases}$
D'où $\exists m > 0, \exists n > 0$ t.q. $m \cdot |x|^2 \leq Q(x) \leq n \cdot |x|^2$
$\forall x \in \mathbb{R}^n$.

On voit que cette démonstration fait intervenir le fait que la boule unité est compacte. Or, en dimension infinie, cette sphère unité n'est pas compacte. En effet, le théorème de Riesz assure que une sphère unité compacte dans un espace vectoriel topologique est équivalent à la finidimensionalité de cet espace. En fait, dans $L^2[t_0, t_f]$, la différence entre "positive" et "fortement positive" est la différence essentielle entre un problème de contrôle optimal singulier et non singulier.

On peut démontrer que si $P(t)$ et $R(t)$ sont continues sur $[t_0, t_f]$, alors $\underline{A \text{ est borné}}$, c-à-d.
$\exists M < \infty$ t.q. $< u, Au > \leq M \cdot < u, u > = M \cdot \|u\|^2$ (2.10)

3] Une des propriétés les plus importantes d'un opérateur fortement positif est que l'opérateur inverse existe et est borné. Cette propriété est exprimée par le théorème suivant :

Théorème 2.1

> soit A fortement positif et borné
> alors A^{-1} existe et est fortement positif et borné.

On peut trouver la démonstration de ce théorème dans [R-1].
Par contre, si A est un opérateur positif, mais non fortement positif, alors l'opérateur inverse A^{-1} est non borné. A cause de cela, l'équation du gradient pour l'équation (2.6), c-à-d. l'équation $A.u + w = 0$, ne possède pas toujours une solution. L'existence d'une solution minimale pour l'équation (2.6) est caractérisée complétement comme suit :

Théorème 2.2

> le problème (P) a une solution minimale $\bar{u}$ si et seulement si :
> (i) $A \geq 0$
> (ii) $w \in R(A)$

Démonstration :

$\Longrightarrow$ soit $\bar{x}$ un élément minimal de (P).
Comme J est deux fois différentiable,
$g(\bar{x}) = A\bar{x} + w = 0$ et $\langle x, \frac{d^2 J(\bar{x})}{dx^2} . x \rangle \geq 0$
$\forall x \in X$

Mais $A\bar{x} \in R(A)$, et donc $w = -A\bar{x} \in R(A)$:
(ii) est donc démontré.
Etant donné que $\frac{d^2 J(\bar{x})}{dx^2} = A$, il suit que
$\langle x, Ax \rangle \geq 0 \quad \forall x \in X$, et donc (i) est aussi démontré.

$\Longleftarrow$ soit $w \in R(A)$; il existe donc un élément $\bar{x}$ t.q. $A.\bar{x} + w = 0$, et

$$J(\bar{x}) = \frac{1}{2} . \langle \bar{x}, A\bar{x} \rangle + \langle \bar{x}, w \rangle + J_0$$

$$= \frac{1}{2} . \langle \bar{x}, -w \rangle + \langle \bar{x}, w \rangle + J_0$$

$$= \frac{1}{2} . \langle \bar{x}, w \rangle + J_0$$

Soit x un élément quelconque dans $L_2[t_0, t_f]$.
Dès lors, $x = \bar{x} + y$, où $y = x - \bar{x} \in L_2[t_0, t_f]$.
On obtient (après quelques calculs) que :
$J(x) = J(\bar{x}) + \frac{1}{2} . \langle y, Ay \rangle$, ou encore
$J(x) - J(\bar{x}) = \frac{1}{2} . \langle y, Ay \rangle$.
Etant donné que A est semi définie positive, il suit que $J(x) \geq J(\bar{x}) \quad \forall x \in L_2$, c-à-d. que $\bar{x}$ est une solution minimale de (P). ■

Le théorème nous conduit à deux définitions :

<u>définition 2.2</u>

> Le problème (P) assujetti aux conditions (i) et (ii) est appelé <u>problème quadratique singulier</u>, et sera noté (SQP)

définition 2.3

> le problème (P), dans lequel A est un opérateur fortement positif, est appelé problème quadratique non singulier ou régulier, et sera noté (NSQP)

Par définition, le (L.Q.P.) est singulier si $R(t) = 0$, $t \in [t_0, t_f]$. Pour les problèmes non singuliers, $R(t) > 0$, $t \in [t_0, t_f]$. Mais remarquons que pour les 2 types de problèmes, l'opérateur A a une composante commune : $T^* P T$ (relation (2.7)). Concernant cette composante, on a le

théorème 2.3

> l'opérateur linéaire $T^* P T$ défini dans l'équation (2.7) est un opérateur complétement continu.

démonstration : [C-3, p. 90-92]

Le théorème jouera un rôle particulièrement important dans le développement des résultats de vitesse de convergence pour les problèmes de contrôle optimal régulier.

4] le cas non singulier (régulier)

Considérons maintenant le L.Q.P. non singulier, c-à-d. où $R(t) > 0$, $t \in [t_0, t_f]$. Il existe un théorème donnant une condition suffisante pour que A soit fortement positif.

Théorème 2.4

> l'opérateur $A = T^* P T + R$ est borné et fortement positif si
> $R(t) > 0 \quad , \quad t \in [t_0, t_f]$
> et la matrice $K(t)$ est finie $\forall t \in [t_0, t_f]$, où
> K est la solution de l'équation de Riccati
> $\dot{K} = -K.F - F^T.K - P + K.G.R^{-1}.G^T.K$
> avec $K(t_f) = 0$.
> En outre, si A est fortement positif, il existe
> un contrôle optimal unique pour le L.Q.P.

[E-3, p.35-38]

* commentaire :

Le théorème est particulièrement important pour 2 raisons.

1. La seconde partie assure l'existence d'un contrôle optimal unique si : $A > 0$.
2. Comparons le théorème 2.4 avec le théorème 2.2.
On sait que $A > 0 \Longrightarrow A \geq 0$, et donc la condition (i) du théorème 2.2 est satisfaite. La condition (ii) est satisfaite également : cela vient du fait que $w \in R(A)$ est toujours satisfait pour un opérateur fortement positif (c.à.d. $\bar{u} = A^{-1} w$ est un élément tel que $A\bar{u} = w$, où A^{-1} existe et est borné par le théorème 2.1). On peut donc conclure que si A est un opérateur fortement positif, l'existence d'une solution optimale ne dépend pas de w dans l'équation (2.6), et donc ne dépend pas de x_0 (puisque A est indépendant de x_0 et que w dépend fortement de x_0).

C'est la raison fondamentale pour laquelle un problème singulier dépend fortement de la condition initiale x_0, tandis qu'un problème régulier pas.

D'un point de vue algorithmique maintenant, un théorème assure que les méthodes "traditionnelles" appliquées au problème non singulier convergent effectivement, et ce linéairement.

théorème 2.5. [E-3, p. 55-58]

> soit l'application dans un espace fonctionnel des méthodes du gradient, du gradient conjugué, et du D.F.P. au problème défini par les équations (2.1) et (2.2). Tous les algorithmes convergent uniformément ($\forall \varepsilon > 0 \; \exists k_0 \; \forall k \geq k_0 \; \| u_k - \bar{u} \| \leq \varepsilon$, ou $\lim_{k \to \infty} \| u_k - \bar{u} \| = 0$) si A est fortement positif.

Mais d'un point de vue pratique, on remarque pour le DFP et le gradient conjugué, la vitesse de convergence est plus rapide (plus exactement, elle est superlinéaire : $\lim_{k \to \infty} \frac{\| u_{k+1} - \bar{u} \|}{\| u_k - \bar{u} \|} = 0$)

Ceci s'explique simplement par le fait que l'opérateur A a une forme particulière : il est la somme d'un opérateur complètement continu (voir théorème 2.3) et d'une composante non singulière.

5] le cas singulier

On considère donc le cas où $R(t) = 0$. On peut montrer que, puisque $A = T^* P T$, l'opérateur A est complètement continu. On obtient aussi une condition suffisante de positivité de A (analogue au théorème 2.4.

théorème 2.6

> l'opérateur A de l'équation (2.6) est borné et positif si $R(t) = 0$ et si :
> 1) $G^T P G > 0 \quad \forall t \in [t_0, t_f]$
> 2) $K(t)$ est fini $\forall t \in [t_0, t_f]$, où $K(t)$ est solution de l'équation :
> $$\dot{K} = -K.F - F^T.K - P + [K.(FG - \dot{G}) + P.G].[G^T.P.G]^{-1} [K.(FG - \dot{G}) + P.G]^T, \quad \text{avec } K(t_f) = 0$$
> la condition 1) est appelée condition généralisée de Kelbsh - Legendre.

démonstration :

Elle est basée sur l'utilisation de la transformation de Goh, qui est simplement l'application d'un opérateur intégral au contrôle $u(.)$. On définit alors un nouveau produit intérieur dans $R(A)$, de telle sorte que A devienne fortement positif dans le nouvel espace. On se ramène alors au cas 4]. Pour la démonstration complète, voir [E-3, p. 39-48] ■

Etant donné que A est un opérateur ≥ 0, le w n'appartient pas nécessairement à R(A). En fait, une condition suffisante pour l'existence d'un arc singulier est :

Théorème 2.7 [A-1, p. 282-287]

> Il existe un unique contrôle singulier pour le problème (2.1), (2.2), avec R(t) = 0, si :
> 1) $G^T P G > 0 \quad \forall t \in [t_0, t_f]$
> 2) K(t) est finie $\forall t \in [t_0, t_f]$ (où K(t) est solution de l'équation définie au théorème 2.6)
> 3) v(t) est différentiable, et $v(t_0) = 0$, où
> $v(t) = -(G^T P G)^{-1} \cdot [G^T P + (FG - \dot{G})^T \cdot S] \cdot y$
> et y est solution de :
> $\dot{y} = F \cdot y + (F \cdot G - \dot{G}) \cdot v$, avec $y(t_0) = x_0$.
> Le contrôle optimal est alors :
> $$\overline{u}(t) = \frac{dv(t)}{dt}$$

Enfin, d'un point de vue algorithmique, on peut montrer que si $\{u_i\}$ est une suite d'éléments générée par le D.F.P. appliqué à (2.1), (2.2), avec R(t) = 0, $H_0 = Id$, et $u_0 \in R(A)$, alors la suite $\{u_i\}$ converge vers l'unique solution en norme minimale $\overline{u} = -A^+ w$, où A^+ est le pseudo-inverse associé à l'équation $Au = w$. En plus de montrer que la suite $\{u_i\}$ converge (uniformément), cette proposition permet de voir que la solution d'un problème de contrôle optimal

singulier est exprimée en termes de l'opérateur pseudo-inverse.

D'un point de vue vitesse de convergence, pour le (L.Q.P.) (singulier ou non singulier), les méthodes du D.F.P. et du gradient conjugué génèrent les mêmes directions de recherche si $H_0 = Id$. Les seuls résultats connus (si $A \geq 0$) sont des résultats de convergence linéaire pour ces deux méthodes.
On a aussi pu montrer un résultat de convergence superlinéaire pour la suite des coûts $\{J(u_i)\}$
(voir [E-3, p. 72-76])

Le cadre dans lequel on travaillera dorénavant est le suivant :
soit le problème de minimisation d'une fonction J dans $\mathbb{R}^n$, où J a des dérivées partielles au moins du second ordre. Le problème (P) s'écrira :

(P) ≡

déterminer un élément $\overline{x} \in X$, X espace séparable de Hilbert, qui minimise la fonction quadratique
$J(x) = \frac{1}{2} \cdot \langle x, Ax \rangle + \langle x, w \rangle + J_0$
où $x, w \in X$, $J_0 \in \mathbb{R}$, A est un opérateur linéaire autoadjoint, et $\langle .,. \rangle$ détermine le produit intérieur dans X.

2.2. algorithme du D.F.P. [H-1, T-2, C-4]

1] Une étape de l'algorithme du D.F.P. peut être représentée par le schéma suivant : on décrit la $(k+1)^{ième}$ itération, et on dispose à ce moment de x_k, H_k, $g_k = \nabla f(x_k)$;

pas 1 : poser $d_k = -H_k . g_k$ (c'est la direction de descente)

pas 2 : faire une recherche unidimensionnelle pour obtenir $0 < \lambda_k^*$, où $\lambda_k^* = \arg.\min. f(x_k + \lambda . d_k)$, $\lambda > 0$.

pas 3 : poser $\delta_k = \lambda_k^* . d_k$

pas 4 : poser $x_{k+1} = x_k + \delta_k$

pas 5 : calculer $f(x_{k+1})$ et $g(x_{k+1})$

pas 6 : poser $\gamma_k = g_{k+1} - g_k$

pas 7 : poser $H_{k+1} = H_k + \Pi_k + B_k$

où $\Pi_k = \dfrac{\delta_k . \delta_k^T}{\delta_k^T . \gamma_k}$ et $B_k = -\dfrac{H_k . \gamma_k . \gamma_k^T . H_k}{\gamma_k^T . H_k . \gamma_k}$

pas 8 : stop si $|d_k| \leq \varepsilon$, $|g_k| \leq \varepsilon$ (ε donné)
sinon, poser $k = k+1$, et aller au pas 1.

L'algorithme du D.F.P. génère donc une suite $\{x_i\}$ d'éléments dans X, à l'aide de la formule :

$$x_{i+1} = x_i + \alpha_i . s_i \qquad (2.11)$$

où
- x_0 est une estimation initiale
- α_i est la grandeur du pas, déterminée par $f(x_i + \alpha_i . s_i) \leq f(x_i + \lambda . s_i)$ $\forall \lambda$,
- et la direction de recherche s_i est donnée par $s_i = -H_i . g_i$,

avec $H_i = H_{i-1} + \frac{p_{i-1} \cdot p_{i-1}^T}{\langle p_{i-1}, y_{i-1} \rangle}$

$$- \frac{(H_{i-1} \cdot y_{i-1}) \cdot (H_{i-1} \cdot y_{i-1})^T}{\langle H_{i-1} \cdot y_{i-1}, y_{i-1} \rangle}$$

- $p_{i-1} = x_i - x_{i-1}$
- $y_{i-1} = g_i - g_{i-1}$
- $g_i = g(x_i)$

Rappelons deux propriétés importantes obtenues pour le D.F.P. dans $\mathbb{R}^n$, dans le cas où A est définie positive :

1) $H_i \geq 0 \quad \forall i \geq 1$. Ce qui a pour conséquence que le d_i est bien une direction de descente (puisque $\langle \nabla f(x_k), d \rangle = \langle g_k, -H_k \cdot g_k \rangle = -g_k^T \cdot H_k \cdot g_k < 0$)
2) on a la propriété de convergence finale quadratique, c-à-d que l'algorithme appliqué à un problème de minimisation quadratique sur $\mathbb{R}^n$ converge en un nombre fini d'itérations.

2] En fait, l'algorithme du D.F.P. s'insère dans une classe plus générale d'algorithmes, appelés quasi-newton.

définition 2.4

Les algorithmes quasi-Newton emploient

$$\begin{cases} x_{i+1} = x_i + \alpha_i \cdot s_i \\ f(x_i + \alpha_i \cdot s_i) \leq f(x_i + \lambda \cdot s_i) \\ s_i = -H_i \cdot g_i \end{cases}$$

pour déterminer l'élément suivant x_{i+1}, avec une grandeur de pas α_i, et une direction de recherche s_i, où l'opérateur H_i est construit de telle sorte que pour $i = 1, 2, \ldots$

$H_i . y_{i-1} = p_{i-1}$, c-à-d.

$H_i . (g_i - g_{i-1}) = x_i - x_{i-1}$, lorsque appliqué à un problème quadratique du type (P) avec une estimation initiale x_0 donnée, et un opérateur autoadjoint, linéaire, fortement positif H_0.

Dans le cas non singulier, Myers [M-1] a montré que pour les fonctions quadratiques, les directions de recherche générées par la méthode du gradient conjugué de Fletcher et Reeves [F-1] et par la méthode du D.F.P. dans $\mathbb{R}^n$ sont des multiples scalaires l'une de l'autre. si le pas initial est pris dans la direction de la plus forte descente.

3] propriétés de l'algorithme

On a vu que pour le cas singulier et non singulier, le critère de convergence est $g(x) = Ax + w = 0$.

Supposons que H_0 soit un opérateur autoadjoint, estimation initiale des opérateurs H_i intervenant dans le D.F.P., et x_0 l'estimation initiale de l'élément minimum recherché.

Les suites $\{x_i\}$, $\{g_i\}$, $\{H_i\}$ générées par la méthode du D.F.P. satisfont les équations suivantes :

$$x_i = x_0 + \sum_{k=0}^{i-1} \alpha_k . s_k \qquad (2.12)$$

$$g_i = g_0 - \sum_{k=0}^{i-1} \alpha_k . A . s_k \qquad (2.13)$$

avec $s_k = - H_k . g_k$.

Si $g_k = 0$ pour un k fini, il suit que $x_k = x_0 + \sum_{i=0}^{k-1} \alpha_i . s_i$ est une solution minimale du S.Q.P.

Donc, à moins qu'il ne le soit explicitement mentionné, on supposera que la méthode du D.F.P. ne finit pas après un nombre fini d'étapes, c-à-d. que $g_i = 0$ pour aucun indice i.

Avec cette supposition, on peut montrer que :

$$\langle p_i , A p_j \rangle = 0 \qquad i \neq j \qquad (2.14)$$

$$\langle g_k , p_i \rangle = 0 \qquad 0 \leq i < k , \; k = 1, 2, \ldots \qquad (2.15)$$

$$H_k . A . p_i = p_i \qquad 0 \leq i < k , \; k = 1, 2, \ldots \qquad (2.16)$$

$$H_k^* = H_k > 0 \qquad k = 1, 2, \ldots \qquad (2.17)$$

Dans l'algorithme du D.F.P. intervient l'expression de α_k, grandeur du pas. Cette valeur est déterminée par une recherche unidimensionnelle, obtenue en calculant le minimum de $J(x_k + \lambda . s_k)$. On annule donc le gradient, et on trouve :

$$J(x_k + \lambda . s_k) = \frac{1}{2} . \Big[\langle x_k , A . x_k \rangle + \langle x_k , A \lambda s_k \rangle + \langle \lambda . s_k , A . x_k \rangle + \langle \lambda . s_k , A \lambda s_k \rangle + \langle x_k , \omega \rangle + \langle \lambda . s_k , \omega \rangle + J_0 \Big]$$

D'où

$$\frac{d}{d\lambda} J(x_k + \lambda . s_k) = 0$$

$$= \frac{1}{2} . x_k . A . s_k + \frac{1}{2} . s_k . A . x_k + s_k . A . \lambda . s_k + s_k . w$$

$$\Longleftrightarrow \quad 0 = x_k . A . s_k + \lambda . s_k . A . s_k + s_k . w$$

$$\Longleftrightarrow \quad \lambda = - \frac{x_k . A . s_k + s_k . w}{s_k . A . s_k}$$

$$\Longleftrightarrow \quad \lambda = - \frac{s_k . [x_k A + w]}{s_k . A . s_k}$$

$$\Longleftrightarrow \quad \lambda = - \frac{\langle s_k , g_k \rangle}{\langle s_k , A s_k \rangle}$$

$$\Longleftrightarrow \quad \lambda = \frac{\langle H_k . g_k , g_k \rangle}{\langle H_k . g_k , A . H_k . g_k \rangle} \qquad (2.18)$$

Considérant cette expression, α_k existe si $\langle H_k . g_k , A . H_k . g_k \rangle \neq 0$. Supposons la proposition suivante démontrée :

proposition 2.1

> soit A un opérateur linéaire autoadjoint et semi défini positif, alors
> $\langle x , Ax \rangle = 0$ ssi $x \in N(A)$

Supposons (par l'absurde) que au pas i, $\langle H_i . g_i , A . H_i . g_i \rangle = 0$. La proposition précédente entraîne que $H_i . g_i \in N(A)$. De plus, $g_i \in R(A)$. Donc $\langle g_i , H_i . g_i \rangle = 0$, d'où $g_i = 0$ (2.17).

Le paramètre α_i est donc bien déterminé aussi longtemps que $g_i \neq 0$, $\forall i$.

démonstration de la proposition :

$\Longrightarrow$ si $A^* = A \geq 0$, $A^{1/2}$ existe, et $(A^{1/2})^* = A^{1/2} \geq 0$

On peut dès lors écrire :

$\langle x, Ax \rangle = 0$ si $\langle A^{1/2}x, A^{1/2}x \rangle = 0$

d'où $A^{1/2}x = 0$.

Donc $Ax = A^{1/2}.(A^{1/2}x) = 0$, ce qui veut dire que $x \in N(A)$.

$\Longleftarrow$ si $x \in N(A)$, alors $Ax = 0$,

et donc $\langle x, Ax \rangle = 0$ ■

2.3. algorithmes à directions conjuguées. [H-1]

On a cité la définition 2.4 des algorithmes quasi-Newton. Les algorithmes, qui possèdent la propriété d'unicité des directions de recherche sont le cas particulier d'une classe beaucoup plus vaste : les algorithmes à direction conjuguée.

Le paragraphe a pour but de décrire une méthode générale de construction d'algorithmes ayant la propriété de descente et de convergence quadratique. La méthode décrite est intéressante, puisqu'on montrera que les algorithmes D.F.P., de Pearson, du gradient conjugué (Fletcher-Reeves), par exemple, peuvent être obtenus comme cas particulier de cet algorithme général. Le résultat auquel on va aboutir est le suivant : tout vecteur satisfaisant la condition de monotonicité et la condition de conjugaison peut être utilisé comme

direction de recherche pour un algorithme de descente à convergence quadratique. Il existe une infinité de moyens de choisir une telle suite de directions de recherche, et donc on pourra construire une infinité d'algorithmes satisfaisant les 2 propriétés citées.

1] introduction

Les propriétés suivantes sont imposées lors de la construction des algorithmes pour minimiser une fonction de plusieurs variables :

1. l'algorithme utilise seulement une recherche unidimensionnelle

2. pour une fonction quadratique, on demande une convergence quadratique au point minimal (en fait, on veut obtenir le point minimal en un nombre d'itérations inférieur au nombre de variables).

3. l'algorithme emploie seulement la fonction et son gradient.

4. l'algorithme emploie l'information présente et celle donnée par l'itération précédente.

Les justifications de ces exigences sont les suivantes :

2. on peut considérer qu'une fonction se comporte plus ou moins comme une fonction quadratique au voisinage du point minimal. La condition 2. assure donc la convergence rapide au stade final.

3. on veut éviter les algorithmes du second ordre.

4. la condition 4. réduit la place mémoire nécessaire lors de l'exécution sur ordinateur.

2] <u>propriété de descente</u>

On doit donc imposer que $J(x_{i+1}) < J(x_i)$. Soit donc un algorithme passant de x_i à x_{i+1} par un pas Δx_i. La suite de points est générée par la formule de récurrence : $x_{i+1} = x_i + \Delta x_i$ (2.19)

De manière générale, on écrit : $\Delta x_i = -\alpha_i . s_i$ (2.20)

où s_i est un vecteur (direction de recherche) et α_i un scalaire déterminant la grandeur du pas.

En combinant (2.19) et (2.20), on obtient bien sûr

$$x_{i+1} = x_i - \alpha_i . s_i \quad (2.21)$$

D'où $$J(x_{i+1}) = J(x_i - \alpha_i . s_i) \quad (2.22)$$

Imposons la condition 1. Considérons s_i fixé, et ainsi $J(x_{i+1})$ devient seulement une fonction de α_i. Donc le minimum de $J(x_{i+1})$ le long de la direction s_i est donné par : $\frac{dJ}{d\alpha_i}(x_i - \alpha_i . s_i) = 0$. C'est une recherche unidimensionnelle.

Après développement, cette relation donne :

$$\boxed{g_{i+1}^T . s_i = 0} \quad (2.23)$$

Elle détermine la grandeur optimale du pas α_i pour une direction s_i.

Comme $g(x) = Ax + w$, on tire que g_{i+1} peut être relié à g_i par : $g_{i+1} = g_i + A.\Delta x_i$.
(2.20) et (2.23) font que cette relation devient :

$$g_{i+1} = g_i - \alpha_i . A . s_i ,$$

d'où $$\alpha_i = \frac{g_i^T . s_i}{s_i^T . A . s_i} \qquad (2.24)$$

Voyons maintenant quelle condition impose pour que $J(x_{i+1}) < J(x_i)$. On a :

$$J(x_{i+1}) = \frac{1}{2} . (x_{i+1})^T . A . (x_{i+1}) + w^T . (x_{i+1}) + J_0$$
$$= J(x_i) + g_i^T . \Delta x_i + \frac{1}{2} . \Delta x_i^T . A . \Delta x_i \qquad (2.25)$$

Par (2.20) et (2.24), cette relation (2.25) devient :

$$J(x_{i+1}) - J(x_i) = - g_i^T . \alpha_i . s_i - \frac{1}{2} . \alpha_i . s_i^T . A . \alpha_i . s_i$$
$$= \frac{-(g_i^T . s_i)^2}{2 . s_i^T . A . s_i} \qquad (2.26)$$

Mais A, par hypothèse, est définie positive ; donc $2 . s_i^T . A . s_i > 0$. D'où la relation (2.26) nous indique que $J(x_{i+1}) - J(x_i) < 0$ aussi longtemps que

$$\boxed{g_i^T . s_i \neq 0} \qquad (2.27)$$

Cette condition (2.27) est une condition de non orthogonalité : s_i ne peut être orthogonal à g_i. De plus, (2.24) montre que α_i et $g_i^T . s_i$ sont de même signe.

Conclusion : si un quelconque vecteur satisfaisant la condition de non orthogonalité ($g_i^T . s_i \neq 0$) est utilisé comme direction de recherche, les équations

$$x_{i+1} = x_i - \alpha_i . s_i \quad \text{et} \quad \frac{dJ}{d\alpha_i}(x_i - \alpha_i . s_i) = 0$$

constituent un algorithme complet ayant la propriété de descente pour la fonction $J(x)$. Un choix particulier de $s_i = g_i$ donne ce qu'on appelle un algorithme de plus grande descente.

3] Convergence quadratique.

Bien qu'on ait maintenant une propriété de descente, il n'est pas du tout garanti que le point minimum est atteint en un nombre fini d'itérations.
Remarquons que g_k et g_j sont liés par la relation :

$$g_k - g_j = -\sum_{i=j}^{k-1} \alpha_i . A . s_i \quad \text{pour } k-1 \geq j \geq 0 \qquad (2.28)$$

(puisque $g_{i+1} = g_i - A\alpha_i s_i = g_{i-1} - A\alpha_{i-1}s_{i-1} - A\alpha_i s_i = \dots$)
Transposons (2.28) :

$$g_k^T - g_j^T = -\sum_{i=j}^{k-1} \alpha_i . s_i^T . A^T \qquad (2.29)$$

Multiplions (2.29) par s_j :

$$s_j . (g_k - g_j^T) = (-\sum_{i=j}^{k-1} \alpha_i . s_i^T . A^T) . s_j \qquad (2.30)$$

Remplaçons α_i par sa valeur donnée en (2.24), c-à-d. par $\alpha_j = \dfrac{g_j^T . s_j}{s_j^T . A . s_j}$, ou encore $\alpha_j . s_j^T . A . s_j = s_j . g_j^T$

d'où (2.30) devient :

$$g_k^T . s_j = -\sum_{i=j+1}^{k-1} \alpha_i . s_i^T . A . s_j \quad , \quad k-2 \geq j \geq 0 \qquad (2.31)$$

Supposons maintenant que s_i est conjuguée à toutes les directions s_j par rapport à la matrice A ; on suppose donc que la condition de conjugaison suivante est satisfaite

$$\boxed{s_i^T . A . s_j = 0} \quad , \quad k-1 \geq i > j \geq 0 \qquad (2.32)$$

Grâce à cela, on va pouvoir montrer que le point minimum est atteint en n itérations, au plus. En effet, (2.31) devient :

$$g_k^T \cdot s_j = 0 \qquad \text{pour } k-2 \geq j \geq 0 .$$

Mais que se passe-t-il en x_{k-1} ? Par (2.23), $g_k^T \cdot s_j = 0$. D'où finalement :

$$g_k^T \cdot s_j = 0 \qquad \text{pour } k-1 \geq j \geq 0 \qquad (2.33)$$

Prenons $k = n$. La relation devient :

$$g_n^T \cdot s_j = 0 \qquad \text{pour } n-1 \geq j \geq 0 \qquad (2.34)$$

Par l'algèbre linéaire, on sait que une suite de n directions non nulles $s_0, s_1, \ldots, s_{n-1}$ peuvent être générés de cette façon ; les directions sont linéairement indépendantes et forment une base dans l'espace de dimension n. Ceci étant le cas, le seul vecteur g_n qui satisfait (2.34) est le vecteur nul.

conclusion : si un vecteur satisfaisant

- la condition de non orthogonalité (2.27)
- la condition de conjugaison (2.32)

est employé pour direction de recherche, on a un algorithme ayant les propriétés de descente et de convergence quadratique.

Mais remarquons que la condition de conjugaison pour $k = n$ conduit à une infinité de solutions. En effet, si $k = n$, la condition s'écrit :

$$s_i^T \cdot A \cdot s_j = 0 \qquad \text{pour } n-1 \geq i > j \geq 0 ,$$

et elle conduit à un système de $\frac{n.(n-1)}{2}$ équations scalaires dans lesquelles les vecteurs inconnus sont

$s_0, s_1, \ldots, s_{m-1}$. Etant donné que chaque vecteur s_i a m composantes scalaires, le nombre d'inconnues scalaires est m^2.

4] On a donc trouvé une infinité d'algorithmes satisfaisant les 2 premières conditions imposées. Employons maintenant les conditions 3. et 4. Exprimons s_i sous la forme :

$$s_i = H_i^T . g_i \qquad (2.35)$$

où H_i est une matrice $(m \times m)$, dont la forme sera spécifiée plus tard.
La condition de conjugaison devient :

$$g_i^T . H_i . A . s_j = 0 \quad \text{pour } i-1 \geq j \geq 0 \qquad (2.36)$$

Comparons alors (2.36) et (2.33) ; on voit que (2.36) peut être satisfait si la matrice H_i est choisie telle que

$$H_i . A . s_j = \beta . s_j \quad \text{pour } i-1 \geq j \geq 0 , \qquad (2.37)$$

où β est une constante arbitraire.
Donc, <u>la condition de conjugaison est satisfaite si la matrice H_i a la propriété (2.37)</u>.
Comme de toute manière on a une relation entre s_i et H_i, donnée par $s_i = H_i^T . g_i$, on va directement raisonner sur la matrice H_i.

Réécrivons (2.37) pour l'itération précédente :

$$H_{i-1} . A . s_j = \beta . s_j \quad \text{pour } i-2 \geq j \geq 0 \qquad (2.38)$$

On sépare alors cette condition en 2 groupes :

$$\begin{cases} H_i . A . s_j = \beta . s_j & \text{pour } i-2 \geq j \geq 0 \qquad (2.39) \\ H_i . A . s_{i-1} = \beta . s_{i-1} & \qquad (2.40) \end{cases}$$

On soustrait alors (2.38) de (2.39) pour obtenir :

$$(H_i - H_{i-1}) . A . s_j = 0 \quad \text{pour } i-2 \geq j \geq 0 \qquad (2.41)$$

De manière analogue à $x_{i+1} = x_i + \Delta x_i$, on notera que

$$H_i = H_{i-1} + \Delta H_{i-1} \quad , \tag{2.42}$$

où ΔH_{i-1} est une matrice $(n \times m)$ qui démote la différence.

On voit donc que la condition (2.39) est satisfaite si ΔH_{i-1} a la propriété :

$$\Delta H_{i-1} \cdot A \cdot s_j = 0 \qquad \text{pour } i-2 \geq j \geq 0 \tag{2.43}$$

Traduisons maintenant (2.40) en les mêmes termes : $H_i \cdot A \cdot s_{i-1} = g \cdot s_{i-1}$

$$\Longleftrightarrow (H_{i-1} + \Delta H_{i-1}) \cdot A \cdot s_{i-1} = g \cdot s_{i-1}$$

$$\Longleftrightarrow \Delta H_{i-1} \cdot A \cdot s_{i-1} = g \cdot s_{i-1} - H_{i-1} \cdot A \cdot s_{i-1} \tag{2.44}$$

On a donc (2.43) et (2.44) qui traduisent simplement (2.39) et (2.40).

Multiplions (2.43) par α_j et (2.44) par α_{i-1} pour éliminer A. Les relations deviennent :

$$\begin{cases} \Delta H_{i-1} \cdot A \cdot s_j \cdot \alpha_j = 0 \\ \Delta H_{i-1} \cdot A \cdot s_{i-1} \cdot \alpha_{i-1} = (g \cdot s_{i-1} \cdot \alpha_{i-1}) - \\ \qquad\qquad (H_{i-1} \cdot A \cdot s_{i-1} \cdot \alpha_{i-1}) \end{cases}$$

Mais n'oublions pas que

$$\Delta x_i = -\alpha_i s_i$$

et $\quad g_{i+1} = g_i + A \cdot \Delta x_i \quad ;$

d'où

$$\begin{cases} -\Delta H_{i-1} \cdot A \cdot \Delta x_j = 0 \\ -\Delta H_{i-1} \cdot A \cdot \Delta x_{i-1} = (g - H_{i-1} \cdot A) \cdot \Delta x_{i-1} \end{cases}$$

$$\begin{cases} \Delta H_{i-1} \cdot \Delta g_j = 0 \qquad \text{pour } i-2 \geq j \geq 0 & (2.45) \\ \Delta H_{i-1} \cdot \Delta g_{j-1} = g \cdot \Delta x_{i-1} - H_{i-1} \cdot \Delta g_{i-1} & (2.46) \end{cases}$$

où $\underline{\Delta g_{i-1} = g_i - g_{i-1}}$

Le sont de nouvelles conditions, exprimées en fonction

de H_i, ΔH_i, Δg_j.

Supposons ΔH_i de la forme suivante, pour satisfaire (2.46) :

$$\Delta H_{i-1} = \rho \cdot \frac{\Delta x_i - g_{i-1}^T}{y_{i-1}^T \cdot \Delta g_{i-1}} - \frac{H_{i-1} \cdot \Delta g_{i-1} \cdot z_{i-1}^T}{z_{i-1}^T \cdot \Delta g_{i-1}} \quad (2.47)$$

où y_{i-1}, z_{i-1} sont des $(n \times 1)$ vecteurs.

Pour satisfaire (2.45), y_{i-1} et z_{i-1} doivent vérifier les relations :

$$\begin{cases} y_{i-1}^T \cdot \Delta g_j = 0 & \text{pour } i-2 \geq j \geq 0 \\ z_{i-1}^T \cdot \Delta g_j = 0 & \text{pour } i-2 \geq j \geq 0 \end{cases} \quad (2.48)$$

conclusion : si on emploie y_{i-1}, z_{i-1} vérifiant (2.48), l'équation (2.47) donne la valeur de ΔH_{i-1}, et (2.42) donne la matrice H_i.

En fait, on essaye de trouver les vecteurs y_{i-1} et z_{i-1} en imposant la condition 4., c-à-d. en utilisant seulement l'information disponible à cette itération et au point le précédent immédiatement.

C'est ainsi qu'après de longs calculs, dont on pourra trouver les détails dans [H-1], on montre que :

$$\begin{cases} y_{i-1} = c_1 \cdot \Delta x_{i-1} + c_2 \cdot H_{i-1}^T \cdot \Delta g_{i-1} & (2.49) \\ z_{i-1} = \kappa_1 \cdot \Delta x_{i-1} + \kappa_2 \cdot H_{i-1}^T \cdot \Delta g_{i-1} & (2.50) \end{cases}$$

où $c_1, c_2, \kappa_1, \kappa_2$ sont des coefficients scalaires.

conclusion : la condition de conjugaison $s_i^T \cdot A \cdot s_j = 0$ pour $i-1 \geq j \geq 0$ est satisfaite si la matrice H est transformée suivant les équations :

$$
\begin{cases}
H_i = H_{i-1} + \Delta H_{i-1} \\
y_{i-1} = c_1 . \Delta x_{i-1} + c_2 . H_{i-1}^T . \Delta g_{i-1} \\
z_{i-1} = k_1 . \Delta x_{i-1} + k_2 . H_{i-1}^T . \Delta g_{i-1}
\end{cases}
$$

Ce sont les différents choix des constantes dans l'équation (2.47) qui engendreront des algorithmes différents. Par exemple, $\delta = 1$, $c_2 = k_1 = 0$, $c_1 = k_2 = 1$ donnent l'algorithme du D.F.P.

5] Pour terminer, il est intéressant de constater que malgré un choix différent des constantes δ, c_1, c_2, k_1, k_2, tous les algorithmes génèrent des directions de recherche parallèles les unes aux autres.

L'expression de la direction de recherche était :

$$s_i = H_i^T . g_i$$

Après quelques développements, on peut constater que s_i peut encore s'écrire sous la forme :

$$s_i = \beta_i . q_i .$$

où β_i est un scalaire défini par

$$\beta_i = 1 - \frac{k_2 . (\Delta g_{i-1}^T . H_{i-1}^T . g_i)}{(z_{i-1}^T . \Delta g_{i-1})}$$

et q_i est un $(n \times 1)$ vecteur défini par :

$$q_i = \left[Id - \frac{\Delta x_{i-1} . \Delta g_{i-1}^T}{\Delta x_{i-1}^T . \Delta g_{i-1}} \right] . H_{i-1}^T . g_i$$

On constate ainsi que la direction q_i est indépendante des constantes c_i et k_i. Cela veut dire que l'équation $s_i = \beta_i . q_i$ montre que les directions de recherche s_i générées par les différents choix de δ, c_i, k_i, $i = 1, 2$, sont parallèles les unes aux autres si la matrice utilisée H_{i-1} au point x_{i-1} est la même.

De même, on représente le déplacement Δx_i par $\Delta x_i = -\alpha_i \cdot \beta_i \cdot q_i$. On peut aussi montrer [H-1] que la grandeur optimale du pas le long de la direction q_i est donnée par :

$$\gamma_i = \alpha_i \cdot \beta_i = \frac{g_i^T \cdot q_i}{q_i^T \cdot A \cdot q_i} \quad .$$

γ_i est indépendant de S, ζ_1, ζ_2, π_1, π_2, et est donc le même pour tous les algorithmes.
(les détails sont présentés en annexe : voir [APPENDICE I])

---

chapitre 3 :

le cas singulier, espace de dimension finie

3.1. introduction

3.2. théorème de convergence

3.3. les pseudo-inverses

3.4. différences (d'un point de vue vitesse de convergence) entre les méthodes quasi-Newton et la méthode du gradient.

Chapitre 3 : le cas singulier, espace de dimension finie.

3.1. introduction

Il est bien connu que l'algorithme du D.F.P. converge en au plus n pas dans $\mathbb{R}^n$, lorsqu'appliqué à un N.S.Q.P. [F-2] Dans ce chapitre, on obtiendra un résultat de convergence semblable non seulement pour le D.F.P., mais pour tous les algorithmes à direction conjuguée, appliqués au S.Q.P.
En outre, on montrera que la vitesse de convergence dans le cas singulier est supérieure ou égale à la vitesse obtenue pour les mêmes méthodes quasi-Newton appliquées aux problèmes quadratiques non singuliers associés.

3.2. théorème de convergence [C-3]

On a introduit dans la section précédente [2.3] la classe des algorithmes à directions conjuguées.
Rappelons en la définition :

définition 3.1 :

> un algorithme à directions conjuguées génère une suite $\{x_i\}$ avec la formule de récurrence $x_{i+1} = x_i + \alpha_i . s_i$, avec une estimation x_0, où la grandeur du pas α_i est déterminée par $J(x_i + \alpha_i . s_i) \leq J(x_i + \lambda . s_i) \quad \forall \lambda$;

s_i est une direction de recherche non nulle satisfaisant :

1. $\langle g_i, s_i \rangle \neq 0$; c'est la condition de non orthogonalité assurant à l'algorithme la propriété de descente.
2. $\langle s_i, A s_j \rangle = 0$, $i \neq j$; c'est la condition de conjugaison des directions de recherche assurant à l'algorithme la propriété de convergence finale quadratique.

Le résultat de convergence finale quadratique peut être étendu au cas singulier comme suit :

Théorème 3.1

> considérons le problème (S.Q.P.) dans $\mathbb{R}^n$. Soit $\{x_i\}$ une suite de vecteurs dans $\mathbb{R}^n$ générée par un algorithme à directions conjuguées. Alors, la suite converge vers un vecteur minimum $\bar{x}$ en au plus m itérations, où m est le rang de A.

Démonstration :

Notons
- $f(x_i + \alpha_i . s_i) \leq f(x_i + \lambda . s_i) \quad \forall \lambda$ (3.1)
- $\langle g_i, s_i \rangle \neq 0$ (3.2)
- $\langle s_i, A s_j \rangle = 0$, $i \neq j$ (3.3)

La dérivation de f par rapport à λ donne

$$\langle g_{i+1}, s_i \rangle = 0 \tag{3.4}$$

Comme $g(x) = A.x + w$, on a que $g_{i+1} = g_i + A.\Delta x_i$,

d'où $g_{i+1} = g_i + A.\alpha_i.s_i$, et ainsi

$$\alpha_i = - \frac{\langle g_i, s_i \rangle}{\langle s_i, A.s_i \rangle} \qquad (3.5)$$

Remarquons que $\langle s_i, A.s_i \rangle \neq 0$.

Etant donné que $g(x) = Ax + w$, on peut relier g_{i+1} à g_i par la relation

$$g_{i+1} = g_i + \alpha_i . A.s_i \quad ,$$

et on peut répéter le procédé jusqu'à obtenir :

$$g_k = g_j + \sum_{i=j}^{k-1} \alpha_i . A.s_i \qquad (k > j) \qquad (3.6)$$

On transpose (3.6), puis on multiplie par s_j dans les deux membres :

$$\langle g_k , s_j \rangle = \langle g_j + \sum_{i=j}^{k-1} \alpha_i . A.s_i , s_j \rangle$$

$$= \langle g_j , s_j \rangle + \langle \sum_{i=j}^{k-1} \alpha_i . A.s_i , s_j \rangle$$

$$= \langle g_j , s_j \rangle + \alpha_j . \langle A s_j , s_j \rangle + \sum_{i=j+1}^{k-1} \alpha_i . \langle A.s_i , s_j \rangle$$

Or, (3.5) entraîne que $\alpha_j . \langle A.s_j , s_j \rangle = - \langle g_j , s_j \rangle$

Donc

$$\langle g_k , s_j \rangle = 0 + \sum_{i=j+1}^{k-1} \alpha_i . \langle A s_i , s_j \rangle \qquad (3.7)$$

pour $k-2 \geq j \geq 0$

Par (3.3), (3.7) peut s'écrire

$$\langle g_k , s_j \rangle = 0 \qquad \text{pour } k-2 \geq j \geq 0 \qquad (3.8)$$

D'autre part, (3.4) $\langle g_k , s_{k-1} \rangle = 0$; d'où finalement on réécrit (3.8) :

$$\underline{\langle g_k , s_j \rangle = 0} \qquad \text{pour } k-1 \geq j \geq 0 \qquad (3.9)$$

Considérons la décomposition de $\mathbb{R}^n$:

$$\mathbb{R}^n = R(A) \oplus N(A) \qquad (3.10)$$

Cela veut dire que chaque élément x de $\mathbb{R}^n$ peut se décomposer de manière unique en $x = x^r + x^n$, où $x^r \in R(A)$ et $x^n \in N(A)$. En particulier,

$$s_i = s_i^r + s_i^n .$$

Les relations (3.2), (3.3), (3.9) deviennent, en remplaçant s_i par $s_i^r + s_i^n$,

$$\langle g_i, s_i^r \rangle \neq 0 \qquad (3.11)$$

$$\langle s_i^r, A s_j^r \rangle = 0 \quad , i \neq j \qquad (3.12)$$

$$\langle g_k, s_j^r \rangle = 0 \quad , k-1 \geq j \geq 0 \qquad (3.13)$$

Il s'ensuit que $\langle s_i^r, A s_i^r \rangle > 0$ (3.14)

En effet, si $\langle s_i^r, A s_i^r \rangle = 0$, alors la propriété 2.1 entraîne que $s_i^r \in N(A)$, et donc que $\langle g_i, s_i^r \rangle = 0$, ce qui est contradictoire avec (3.11).

Les résultats (3.12) et (3.14) veulent dire que $\{s_i^r\}$ est un ensemble d'éléments non nuls, A-conjugués de $R(A)$.

Utilisons alors le lemme suivant :

Lemme 3.1

> m directions A-conjuguées de $\mathbb{R}^n$, $v_1, \ldots, v_m$, sont automatiquement linéairement indépendantes.

Considérons la relation (3.13), avec $k = m$, le rang de la matrice A. Cela donne donc :

$$\langle g_m, s_j^r \rangle = 0 \quad , m-1 \geq j \geq 0 \qquad (3.15)$$

Etant donné que $\{s_0^r, \ldots, s_{m-1}^r\}$ est un ensemble de

vecteurs linéairement indépendants dans $R(A)$, et que le rang de A est m, alors $\{s^0, \ldots, s^{m-1}\}$ forme une base dans $R(A)$. Cela implique que le vecteur gradient qui satisfait (3.15) doit appartenir à $N(A)$.

Mais $g_m = g(x_m)$

$= A \cdot x_m + w \quad \in R(A)$.

Donc $g_m \in N(A) \cap R(A) = \{0\}$,

c-à-d. $g_m = A \cdot x_m + w = 0$, et la suite $\{x_i\}$ converge en au plus m itérations. ∎

* commentaire : le théorème 3.1 montre que le résultat de convergence en un nombre fini de pas est valable pour tous les algorithmes qui génèrent des directions conjuguées. Par exemple, les méthodes quasi-Newton de Broyden [B-2] et de Huang [H-1] possèdent cette propriété. En particulier, on peut montrer que cette propriété de convergence en au plus m pas est vraie pour les méthodes du gradient conjugué, du D.F.P., de Powell [P-1].

3.3. Les pseudo-inverses.

Dans le premier chapitre, section 1.3, on a vu qu'une solution au sens des moindres carrés de l'équation $Ax = b$, où $A \in \mathcal{L}(H_1, H_2)$, $b \in H_2$, est donnée par un élément $u \in H_1$ tel que $\|Au - b\| \leq \|Ax - b\| \quad \forall x \in H_1$. On appellera meilleure solution approximative (notée BAS) la solution au sens des moindres carrés de norme minimale.

définition 3.2

> on appelle $\hat{x} \in X$ la meilleure solution approximative (notée BAS) de l'équation $Ax = b$ si
> 1. $\|A\hat{x} - b\| = \delta b$, où $\delta b = \inf_{x \in D(A)} \|Ax - b\|$
> 2. $\|\hat{x}\| < \|u\|$, où u est toute autre valeur qui atteint l'infimum.

Soient
- A un opérateur linéaire autoadjoint dense défini sur l'espace de Hilbert X
- Y l'espace de Hilbert image
- $D(A)$ le domaine de A
- $R(A)$ l'image de A
- $\overline{R}(A)$ la fermeture de $R(A)$
- $N(A)$ le noyau de A
- $(.)^{\perp}$ le complément orthogonal de $(.)$
- P_R la projection orthogonale sur $\overline{R}(A)$
- r un indice qui indique la restriction de n'importe quel opérateur A sur X au sous-espace $(N(A))^{\perp}$

A propos de l'opérateur pseudo-inverse, on pourrait encore montrer la proposition suivante :

théorème 3.2 [C-3]

> le pseudo-inverse existe comme l'unique opérateur défini linéaire, dense, fermé tel que $A^{+} = A_r^{-1} P_R$

On va maintenant voir que la notion même de pseudo-inverse entapelée à jouer un rôle important dans les problèmes de minimisation quadratique singulier.
Dans le cas présent, l'équation qui nous intéresse est

$$g(x) = A.x + w = 0 \qquad \text{où } w \in R(A) \tag{3.16}$$

Puisque dans le S.Q.P. A est semi définie positive ($A \geq 0$), la solution de (3.16) ne peut être donnée immédiatement. Il est alors intéressant de rechercher la BAS de (3.16).
Par le théorème 3.2, on sait que $A^+ = A^{-}_{\hat{R}} P_R$.
Comme 1. $P_R(w) = w \in R(A)$ (puisque $w \in R(A)$)

2. $R(A) \subset F_A = \{ b : P_R b \in R(A) \}$,

on utilise les 2 propositions suivantes

1. une BAS de $Ax = b$ existe ssi $b \in F_A$
   (c'est la proposition 1.10)
2. par la définition du pseudo-inverse,
   $D(A^+) = F_A$, et
   $\forall b \in F_A$, $A^+ b = \tilde{x}$ est la BAS de $Ax = b$

pour conclure :
la BAS de (3.16), notée x^*, existe et est donnée par

$$\boxed{x^* = -A^+ w} \tag{3.17}$$

De plus, par (1.4), on sait que l'ensemble des solutions en norme minimale de l'équation $Ax + w = 0$ est donné par

$$S = \{ \bar{x} \text{ t.q. } \bar{x} = x^* + x^{\sim}, \; x^{\sim} \in N(A) \}.$$

Chaque élément $\bar{x} \in S$ est une solution du S.Q.P., et $x^* = -A^+ w$ est la solution en norme minimale.

On a montré dans [C-4] que si $H_0 = Id$, la suite $\{x_0, x_1, \ldots\}$ générée par la méthode du D.F.P. et la méthode du gradient conjugué converge vers

$$\bar{x} = x^* + (I - P_R)x_0 \quad ,$$

où P_R est la projection orthogonale de x sur $R(A)$.

Par conséquent, si on veut obtenir la solution de norme minimale x^* par les méthodes D.F.P. ou du gradient conjugué, le meilleur moyen est de choisir $x_0 = 0$ comme estimation initiale.

Pour un S.Q.P., $A^T = A \geq 0$, et A^{-1} n'existe pas. Mais la matrice H_m a un comportement intéressant. On peut montrer à partir de la relation (2.16) que, si la suite $\{x_i\}$ converge en exactement m étapes pour le S.Q.P., on a :

$$(H_m . A).x = x \qquad \forall x \in R(A)$$

Cela veut dire que $H_m . A$ joue le rôle d'une matrice identité dans $R(A)$, ou encore $H_m A = P_R$.

Donc $\bar{x} = x^* + (I - P_R)x_0$

$= x^* + (I - H_m A)x_0$

si la méthode du D.F.P. converge en m pas, avec $H_0 = Id$ comme approximation initiale.

3.4. Différences (d'un point de vue vitesse de convergence) entre les méthodes quasi-Newton et la méthode du gradient [C-3, J-1]

Les résultats de cette section sont présentés sans démonstration. Ils sont donnés pour illustrer le fait suivant : les vitesses

de convergence obtenues pour la méthode du gradient sont largement insuffisants, tandis que les vitesses de convergence pour les algorithmes quasi-Newton sont déjà nettement plus accessibles.

Considérons le (S.Q.P.), c-à-d. le problème

$$\begin{cases} \min J(x) = \frac{1}{2} \cdot \langle x, Ax \rangle + \langle x, w \rangle + J_0 \\ \text{où } x \in \mathbb{R}^n,\ J_0 \in \mathbb{R},\ A^T = A \geq 0,\ w \in \mathcal{R}(A). \end{cases} \tag{3.18}$$

On définit alors le (A.N.S.Q.P.), c-à-d. le problème quadratique non singulier associé, comme suit :

$$\begin{cases} \min J_\eta(x) = \frac{1}{2} \cdot \langle x, A_\eta x \rangle + \langle x, w \rangle + J_0 \\ \text{où } A_\eta = A + \frac{1}{\eta} \cdot Id,\ \eta > 0. \end{cases} \tag{3.19}$$

On introduit ce nouveau problème pour pouvoir étudier le comportement des algorithmes comme un problème tendant vers une singularité. On peut montrer le

Théorème 3.3

(i) (A.N.S.Q.P.) approche (S.Q.P.) quand $\eta \to 0$

(ii) chaque vecteur propre z_i de A dans le (S.Q.P.) avec la valeur propre correspondante $\lambda_i \geq 0$, $i = 1, 2, \ldots, n$, est aussi un vecteur propre de A_η pour le (A.N.S.Q.P.), avec des valeurs propres correspondantes $\beta_i = \lambda_i + \frac{1}{\eta}$

En fait, c'est surtout le théorème suivant qui est intéressant du point de vue de la vitesse de convergence des algorithmes.

Théorème 3.4

> si la suite $\{\overline{x}_i\}$ générée par la méthode quasi-Newton appliquée au (A.N.S.Q.P.) converge en $k \leq n$ pas,
> alors la suite $\{x_i\}$ générée par la même méthode appliquée au (S.Q.P.), avec la même estimation initiale x_0, converge en au plus k pas.

Le théorème montre que la convergence d'une méthode quasi-Newton appliquée au (S.Q.P.) n'est jamais plus lente que la convergence de la même méthode, appliquée cette fois au (A.N.S.Q.P.).

Cependant, la méthode du gradient se comporte exactement d'une façon opposée. Plus précisément, on a montré [J-1] que la méthode du gradient a une convergence plus lente pour les problèmes singuliers que pour les approximations non singulières correspondantes, pour lesquelles les méthodes aux directions conjuguées ont une convergence plus rapide.

Les résultats peuvent être exprimés sous la forme suivante : pour la méthode du gradient, la convergence sera lente toutes les fois que le champ des valeurs propres de A est large, et lorsque la matrice est singulière, la vitesse de convergence est sous-linéaire.

Tout ceci implique que la lente convergence attribuée aux problèmes singuliers est en fait une propriété

de la méthode du gradient elle-même, et n'est pas
due à la singularité du problème.

Chapitre 4 :

le cas régulier, espace de dimension infinie.

4.1. introduction

4.2. convergence linéaire pour algorithmes généraux.

4.3. théorème de convergence superlinéaire.

4.4. conclusion

Chapitre 4 : le cas régulier, espace de dimension infinie

4.1. introduction

Il n'existe pas, à notre connaissance, de résultat concernant les vitesses de convergence des méthodes quasi-Newton dans un espace fonctionnel. Le problème est pourtant intéressant, puisque de tels algorithmes peuvent être appliqués notamment dans la résolution de problèmes de contrôle optimal. [voir [E1, E2]]
Après une approche générale du problème, une seconde approche sera présentée ; elle permettra de déterminer des résultats de vitesse de convergence superlinéaire dans certains cas particuliers.
Seul l'algorithme du D.F.P. sera étudié dans ce chapitre, étant donné que toutes les méthodes quasi-Newton possèdent la même direction de recherche à chaque itération pour les problèmes quadratiques.

4.2. convergence linéaire pour algorithmes généraux. [C-3]

1] Considérons le problème de minimisation quadratique régulier (dans le cadre d'un espace fonctionnel), c-à-d. le problème suivant :

$$\min J(x) = \frac{1}{2} \langle x, Ax \rangle + \langle x, w \rangle + J_0 \qquad (4.1)$$

où A est un opérateur linéaire autoadjoint fortement positif.

A est un opérateur fortement positif : cela veut donc dire (voir définition 2.1) qu'il existe une constante m, $0 < m < +\infty$, telle que :

$$0 < m.\langle x, x\rangle \leq \langle x, Ax\rangle \qquad \forall x \in X$$

Par l'inégalité de Cauchy-Schwarz, on a que

$$|\langle x, y\rangle| \leq \|x\|.\|y\| .$$

Or, $\langle x, Ax\rangle \geq 0$, d'où $|\langle x, Ax\rangle| = \langle x, Ax\rangle$,

et ainsi on peut écrire :

$$\underline{\exists\ 0 < m \leq M < \infty \text{ t.q.}}$$

$$\underline{m.\langle x, x\rangle \leq \langle x, Ax\rangle \leq M.\langle x, x\rangle \quad \forall x \in X}. \qquad (4.2)$$

La méthode du D.F.P. demande que H_0 soit un opérateur linéaire fortement positif, donc qu'il existe des constantes $\underline{0 < m_0 \leq M_0 < \infty}$ t.q.

$$\underline{m_0.\langle x, x\rangle \leq \langle x, H_0.x\rangle \leq M_0.\langle x, x\rangle \ \forall x \in X} \qquad (4.3)$$

On supposera dans la suite que H_0 est un opérateur <u>linéaire</u>, <u>fortement positif</u> et <u>autoadjoint</u>. D'où on peut écrire l'égalité :

$$\langle x, H_0 x\rangle = \frac{1}{2}.\langle x, H_0 x\rangle + \frac{1}{2}.\langle x, H_0^* x\rangle$$

$$= \frac{1}{2}.\langle x, (H_0 + H_0^*).x\rangle .$$

ainsi, la relation (4.3) peut s'écrire :

$$\underline{\exists\ 0 < m_0 \leq M_0 < \infty \text{ t.q.}}$$

$$\underline{m_0.\langle x, x\rangle \leq \langle x, \tfrac{1}{2}.(H_0 + H_0^*).x\rangle \leq M_0.\langle x, x\rangle} \qquad (4.4)$$

$$\forall x \in X$$

* <u>Remarque</u> : supposer que H_0 est un opérateur autoadjoint ne fait pas perdre la généralité de la suite de l'exposé.

En effet, pour tout opérateur H,

$$H = \frac{1}{2} \cdot (H + H^*) + \frac{1}{2} \cdot (H - H^*) \qquad (4.5)$$

et donc

$$\langle x, Hx \rangle = \frac{1}{2} \cdot \langle x, (H + H^*) \cdot x \rangle \qquad (4.6)$$

Dans le cas où H_0 est linéaire, fortement positif, mais pas nécessairement autoadjoint, les mêmes résultats pourront être obtenus en employant les équations (4.4), (4.5), (4.6) et en utilisant $(H_i + H_i^*)$, $(R_i + R_i^*)$, $(G_i + G_i^*)$ et $(S_i + S_i^*)$ au lieu de H_i, R_i, G_i et S_i respectivement quand ceux-ci apparaissent dans les produits intérieurs.

2] La propriété suivante, qui a été montrée dans [P-2, chapitre 6] pour résoudre le problème

$$\min \{ f(x) : x \in \mathbb{R}^n \}$$

peut être appliquée pas à pas à des problèmes de minimisation quadratique réguliers :

théorème 4.1

soit le problème (4.1) soumis à la condition (4.2), soit la suite $\{x_i\}$ construite par un algorithme général qui satisfait les conditions suivantes, pour $i = 0, 1, 2, \ldots$:

(i) $x_{i+1} = x_i + \alpha_i \cdot s_i$

(ii) $-\langle g_i, s_i \rangle \geq \beta \cdot \|g_i\| \cdot \|s_i\|$, $\beta \in [0, 1]$

(iii) $f(x_i + \alpha_i \cdot s_i) = \min \{ f(x_i + \alpha \cdot s_i), \alpha \geq 0 \}$

alors

$$\|x_i - x^*\| \leq \frac{2}{m} \cdot [f(x_0) - f(x^*)] \cdot \left[1 - \left(\frac{\beta \cdot m}{M}\right)^2\right]^i \qquad (4.7)$$

Le théorème donne donc une vitesse de convergence linéaire. A cet effet, rappelons la définition d'une convergence linéaire.

définition 4.1

> Soit $\{z_i\}$ une suite dans un espace de Hilbert H qui converge vers z^*. On dira que $\{z_i\}$ converge vers z^* au moins linéairement si il existe un entier $k \geq 0$, une constante E et un $\theta \in [0,1]$ t.q.
> $$\|z_i - z^*\|_H \leq E \cdot \theta^i \qquad \forall i \geq k$$

Cette définition exprime le fait que la convergence de $\{z_i\}$ est au moins linéaire si: $\|z_i - z^*\| \longrightarrow 0$ (quand $i \longrightarrow \infty$) au moins aussi rapidement qu'une progression géométrique.
La convergence exprimée par le théorème est-elle bien linéaire?
Le théorème nous donne le résultat (4.7).
or, $m \leq M$ et $\beta \in [0,1]$, d'où $\left(\frac{\beta \cdot m}{M}\right)^2 \in [0,1]$
et ainsi $\left(1 - \left(\frac{\beta \cdot m}{M}\right)^2\right) \in [0,1]$.

Posons $\theta = 1 - \left(\frac{\beta \cdot m}{M}\right)^2$, et $E = f(x_0) - f(x^*)$,
et on a bien que

$$\|x_i - x^*\| \leq \frac{2}{m} \cdot E \cdot \theta^i$$

Montrons maintenant que la méthode du D.F.P. vérifie bien les hypothèses du théorème, et donc qu'elle possède cette propriété de convergence.

Par construction même de l'algorithme du D.F.P., les conditions (i) et (iii) sont satisfaites. Il reste à vérifier (ii). Pour ce faire, on utilise les relations suivantes :

$$\bullet \ \frac{1}{M_0} \cdot \| H_k \cdot g_k \|^2 \leq \langle g_k , H_k \cdot g_k \rangle \qquad (4.8)$$

$$\bullet \ \frac{m \cdot m_0}{M} \cdot \| g_k \| \leq \| H_k \cdot g_k \| \ , \ k = 0, 1, 2, \ldots \qquad (4.9)$$

où $0 < m_0 \leq M_0$, $m_0 = \inf_{x \neq 0} \frac{\| H_0 \cdot x \|}{\| x \|}$

et $M_0 = \sup_{x \neq 0} \frac{\| H_0 \cdot x \|}{\| x \|}$

Les relations (4.8) et (4.9) sont présentées dans [H-4].
Dès lors :

$$\begin{aligned} -\langle g_i, s_i \rangle &= -\langle g_i, -H_i \cdot g_i \rangle \\ &= \langle g_i, H_i \cdot g_i \rangle \\ &\geq \frac{1}{M_0} \cdot \| H_i \cdot g_i \|^2 \quad \text{(par (4.8))} \\ &\geq \frac{m_0 \cdot m}{M \cdot M_0} \cdot \| g_i \| \cdot \| H_i \cdot g_i \| \\ &\qquad \text{(par (4.9))} \end{aligned}$$

Or, $s_i = -H_i \cdot g_i$, d'où $\| s_i \| = \| H_i \cdot g_i \|$.
Posons $\beta = \frac{m_0 \cdot m}{M \cdot M_0}$, et on obtient (ii).
Comme $m_0 \leq M_0$ et $m \leq M$, on a bien que $\beta \in [0, 1]$.

conclusion : un résultat de vitesse de convergence pour le D.F.P. est :

$$\| x_i - x^* \|^2 \leq \frac{2}{m} \cdot [f(x_0) - f(x^*)] \cdot \left[1 - \left(\frac{m^2 \cdot m_0}{M^2 \cdot M_0} \right)^2 \right]^i \qquad (4.10)$$

Dans [D-1] , un résultat similaire est obtenu pour la méthode du gradient conjugué, appliqué à (4.1) soumis à (4.2) :

$$\|x_i - x^*\|^2 \leq \frac{2}{m} \cdot [f(x_0) - f(x^*)] \cdot \left[1 - \frac{m}{M}\right]^i \quad , i = 0, 1, 2 \ldots \quad (4.11)$$

En employant les inégalités (4.8) et (4.9), et la procédure utilisée dans [D-1] , on peut montrer aussi que pour la méthode du D.F.P. :

$$\|x_i - x^*\|^2 \leq \frac{2}{m} \cdot [f(x_0) - f(x^*)] \cdot \left[1 - \left(\frac{m_0}{M_0}\right)^2 \cdot \left(\frac{m}{M}\right)^3\right]^i \quad (4.12)$$

$$i = 0, 1, 2, \ldots$$

qui est en fait une vitesse de convergence <u>strictement meilleure</u> que le résultat donné par l'équation (4.10). En effet,

$$1 - \left(\frac{m_0}{M_0} \cdot \frac{m^2}{M^2}\right)^2 > 1 - \left(\frac{m_0}{M_0}\right)^2 \cdot \left(\frac{m}{M}\right)^3 \quad ,$$

d'où

$$\left[1 - \left(\frac{m_0}{M_0}\right)^2 \cdot \left(\frac{m}{M}\right)^3\right]^i \text{ tendra plus vite vers 0}$$

que $\left[1 - \left(\frac{m_0}{M_0} \cdot \frac{m^2}{M^2}\right)^2\right]^i$ lorsque $i \longrightarrow \infty$

3] Considérons maintenant l'application de la méthode du gradient, c-à-d. lorsque $s_i = -g_i$, au même problème (4.1) avec (4.2).

Comme $s_i = -g_i$, on a

$$-\langle g_i, s_i \rangle = \langle g_i, g_i \rangle = \|g_i\|^2.$$

Donc $\beta = 1$ dans le théorème 4.1 , point (ii), et on peut dès lors trouver la vitesse de convergence suivante pour la méthode du gradient :

$$\|x_i - x^*\|^2 \leq \frac{2}{m} \cdot [f(x_0) - f(x^*)] \cdot \left[1 - \left(\frac{m}{M}\right)^2\right]^i \quad (4.13)$$

$$i = 0, 1, 2, \ldots$$

Les comparaisons de (4.10), (4.12) et (4.13) montrent que les vitesses de convergence obtenues pour l'algorithme du D.F.P., développés dans les 2 approches précédents, sont plus mauvaises que la vitesse de convergence de la méthode du gradient.
Mais les résultats sont contraires à l'expérience. C'est ce qui va conduire, dans la section suivante, à employer une autre approche, de manière à développer un résultat de convergence plus rapide pour la méthode du D.F.P.

4.3. Théorème de convergence superlinéaire

1] Dans cette section, on montrera que pour les problèmes de minimisation quadratique (ou régulier, espace fonctionnel) dont l'opérateur hessien peut s'écrire sous la forme $A = I + K$, où K est un opérateur complètement continu, autoadjoint, positif, le développement d'une vitesse de convergence superlinéaire est possible pour la méthode du D.F.P.
Avant d'étudier ce résultat, il est nécessaire d'introduire quelques définitions.

définition 4.2

1. un ensemble S d'éléments de X est borné si il existe une constante c telle que pour tout $x \in S$, $\|x\| \leq c$

2. un ensemble S est compact si chaque suite $\{x_n\}$ d'éléments dans S contient une sous-suite convergente.
3. un opérateur linéaire est complétement continu si il transforme des ensembles bornés en des ensembles compacts.

Remarques : [S-2, p. 184]

1. Soit $A \subset X$, on a les relations suivantes :
   A compact $\Longrightarrow$ A borné, et
   A compact $\Longleftarrow$ A borné seulement dans les espaces finidimensionnels (la démonstration est basée sur le théorème de Bolzano-Weierstrass).

2. Tout opérateur borné dont l'image est fini-dimensionnelle est complétement continu. En particulier, tout opérateur appartenant à H^* (le dual topologique de H) est complétement continu.
   Les opérateurs complétement continus les plus importants sont les opérateurs intégraux de Hilbert-Schmidt, dont on verra un exemple dans la section suivante 2.]

2] Les 2 propositions suivantes, dont les démonstrations seront présentées en annexe [APP. II], seront utile ultérieurement.

proposition 4.1

> si A est un opérateur complétement continu, inversible dans un espace de dimension infinie,
> alors son inverse est non borné'.

proposition 4.2

> si A est complétement continu, et $\{\varphi_n\}$ est une suite infinie orthonormale dans H, alors $\lim_{n \to \infty} A\varphi_n = 0$

* commentaire :

  En fait, un opérateur continu envoie des suites convergentes sur des suites convergentes. La proposition 4.2 montre simplement qu'un opérateur complétement continu envoie aussi certaines suites non convergentes sur des suites convergentes. L'opérateur inverse doit donc envoyer certaines suites convergentes sur des suites non convergentes. C'est la proposition 4.1. *

3] Voici le théorème fondamental de cette section :

Théorème 4.2 [c-3]

soit le problème (4.1), avec les conditions (4.2) et (4.3) ;
si $\{x_i\}$ et $\{H_i\}$ sont des suites infinies générées par l'algorithme du D.F.P,
alors
(i) H_k $(k = 0, 1, 2, \ldots)$ sont des opérateurs linéaires, autoadjoints, et il existe des constantes $0 < m \leq M < \infty$ t.q.
$m . \|x\|^2 \leq \langle x, H_k x \rangle \leq M . \|x\|^2$, $x \in X$,
$k = 0, 1, 2, \ldots$
(ii) si A a la forme $A = I + K$, où K est un opérateur complètement continu, autoadjoint, positif,
et si $H_0 = I$,
alors $\{x_i\}$ converge vers la solution minimale $x^* = -A^{-1}.w$ de manière superlinéaire, c-à-d.
$$\lim_{i \to \infty} \frac{\|x_{i+1} - x^*\|}{\|x_i - x^*\|} = 0$$

Le théorème montre donc que la méthode du D.F.P., avec $H_0 = Id$, a une vitesse de convergence superlinéaire si l'opérateur Hessien A du problème de minimisation quadratique a la forme $A = I + K$, où K est un opérateur autoadjoint, positif, complètement continu. Il serait donc utile de montrer qu'une large classe de problèmes de contrôle optimal quadratique

linéaire possède effectivement un opérateur Hessien A de la forme $A = R + K$, où K a les propriétés exigées.

Pour cela, considérons le problème de déterminer u^* qui minimise la fonctionnelle quadratique linéaire suivante :

$$J[u(.)] = \frac{1}{2} . x^T(t_f) . Q_f . x(t_f) + \int_{t_0}^{t_f} \left(\frac{1}{2} . x^T . Q . x + u^T . c . x + \frac{1}{2} . u^T . R . u \right) dt \quad (4.14)$$

soumis à

$$\begin{cases} \dot{x} = \bar{A} . x + B . u \\ x(t_0) = x_0 \end{cases} \quad (4.15)$$

où $Q, C, R, \bar{A}, B$ sont des matrices continues, fonction du temps. On suppose en outre que Q, R, Q_f sont symmétriques.

Soit $\langle a(t), b(t) \rangle = \int_{t_0}^{t_f} a^T(t) . b(t) \, dt$.

De cette manière, (4.14) s'écrit :

$$J(u) = \frac{1}{2} . \frac{1}{t_f - t_0} . \langle x(t_f), Q_f . x(t_f) \rangle + \frac{1}{2} . \langle x, Q x \rangle + \langle u, c . x \rangle + \frac{1}{2} . \langle x, R . u \rangle \quad (4.16)$$

On considère $\Phi(t, \tau)$, la matrice de transition de (4.15). Dès lors,

$$x(t) = \Phi(t, t_0) . x_0 + \int_{t_0}^{t_f} \Phi(t, \tau) . B(\tau) . u(\tau) \, d\tau$$

ou encore

$$\begin{cases} x(t) = E x_0 + F u & (4.17) \\ \text{et} \\ x(t_f) = E_f x_0 + F_f u & (4.18) \end{cases}$$

Ceci nous permet d'écrire $J(u)$ d'une manière plus concise :

$$J(u) = J_0 + \langle u, w \rangle + \frac{1}{2} \cdot \langle u, A u \rangle \qquad (4.19)$$

où
$$\cdot\ J_0 = \frac{1}{2} \cdot \frac{1}{t_f - t_0} \cdot \langle E_f x_0 , Q_f \cdot E_f \cdot x_0 \rangle + \frac{1}{2} \cdot \langle E x_0 , Q \cdot E \cdot x_0 \rangle \qquad (4.20)$$

$$\cdot\ w = \left(\frac{1}{t_f - t_0} \cdot F_f^* \cdot Q_f \cdot E_f + F^* \cdot Q \cdot E + C \cdot E \right) \cdot x_0 \qquad (4.21)$$

$$\cdot\ A = \frac{1}{t_f - t_0} \cdot F_f^* \cdot Q_f \cdot E_f + F^* \cdot Q \cdot F + 2 \cdot C \cdot F + R \qquad (4.22)$$

Sans perdre de généralité, les hypothèses suivantes sont faites pour simplifier le raisonnement :

(1) en regardant (4.22), on constate que l'opérateur Hessien A est indépendant de l'état initial $x(t_0) = x_0$. Comme on ne considère ici que les propriétés de A, on peut supposer que $x_0 = 0$.

(2) soit $R = 0$; montrons que $A = K$.

La proposition suivante exprime le résultat qui nous intéressait :

proposition 4.3

> considérons le L.Q.P. défini en (4.14) et (4.15) avec $R = 0$ et $x_0 = 0$. Si l'existence d'une unique solution minimale est garantie, alors l'opérateur Hessien A donné par (4.22) est autoadjoint, positif, complétement continu.

démonstration :

On sait que la condition nécessaire et suffisante pour que le problème possède une solution minimale est

$$\boxed{\begin{array}{ll}(i) & A^* = A \geq 0 \qquad (4.23)\\ (ii) & w \in R(A) \qquad (4.24)\end{array}}$$

Comme l'existence d'une unique solution minimale est garantie, (4.23) implique que A est un opérateur auto-adjoint positif.

Comme il a été montré dans [K-5, équa. 10], A peut aussi s'écrire :

$$(Au)(t) = \frac{1}{2} \cdot \int_{t_0}^{t_f} k(t,s) \cdot u(s)\, ds \qquad (4.25)$$

où
$$\begin{aligned} k(t,s) = {} & B^T(t) \cdot \Big[\Phi^T(t_f,t) \cdot Q_f \cdot \Phi(t_f,s) \\ & + \int_{t \vee s}^{t_f} \Phi^T(\tau,t) \cdot Q(\tau) \cdot \Phi(\tau,s)\, d\tau \Big] \cdot B(s) \\ & + C(t) \cdot \Phi(t,s) \cdot B(s) \cdot 1(t-s) + B^T(t) \cdot \Phi^T(s,t) \\ & \cdot C(s)^T \cdot 1(s-t) \end{aligned} \qquad (4.26)$$

- $t \vee s = \max(t,s)$
- $1(t-s) = \begin{cases} 1 & t \geq s \\ 0 & t < s \end{cases}$

c-à-d. A est un opérateur intégral avec le noyau $k(t,s)$.

Remarquons que $k(t,s)$ est une matrice dont chaque entrée $k_{ij}(t,s)$ est une fonction à valeur réelle de 2 variables définie sur le carré $t_0 \leq t,s \leq t_f$.

Pour montrer que A est un opérateur complètement continu, il est suffisant de montrer que l'opérateur intégral A_{ij} (pour chaque i,j) défini par

$$(A_{ij}u)(t) = \frac{1}{2} \cdot \int_{t_0}^{t_f} k_{ij}(t,s) \cdot u(s)\, ds \qquad (4.27)$$

est un opérateur complètement continu.

Pour le voir, il suffit de se rappeler la définition d'un

opérateur complètement continu : il transforme des ensembles bornés en des ensembles compacts.
Il suffit de montrer que A_{ij} est un opérateur Hilbert-Schmidt, c-à-d.

$$\int_{t_0}^{t_f}\int_{t_0}^{t_f} |k_{ij}^2(\tau,s)|\, d\tau\, ds < \infty \qquad (4.28)$$

(rappelons que un noyau $k(x,\xi)$ pour lequel $\int_a^b\int_a^b |k^2(x,\xi)|\,dx\,d\xi < \infty$ est appelé Hilbert-Schmidt, [S-2, p.191] et la transformation K générée par ce noyau est appelée opérateur intégral Hilbert-Schmidt. L'intérêt de ce type de noyau est qu'il génère toujours un opérateur borné et complètement continu, même si $k(x,\xi)$ est une fonction non bornée de x et de ξ).

Pour montrer (4.28), remarquons d'abord que comme $Q, C, \bar{A}, B$ sont des matrices continues fonction du temps, et comme Q_f est une matrice constante (par hypothèse), la matrice de transition $\Phi(\tau,s)$ est une fonction différentiable sur $t_0 \le \tau, s \le t_f$. Par conséquent, à partir de (4.26), on peut voir que $k_{ij}(\tau,s)$ est une fonction bornée sur $t_0 \le \tau, s \le t_f$ pour chaque i,j.
Soit $M_{ij} = \max_{t_0 \le \tau, s \le t_f} |k_{ij}(\tau,s)|$.

Il est clair que pour chaque i,j, on a $M_{ij} < \infty$, ce qui implique que

$$\int_{t_0}^{t_f}\int_{t_0}^{t_f} |k_{ij}^2(\tau,s)|\, d\tau\, ds \le (t_f - t_0)^2 \cdot M_{ij}^2 < \infty,$$

c-à-d. A_{ij} est complètement continu pour chaque i,j.
Ceci implique donc que A est un opérateur complètement continu. ∎

démonstration du théorème 4.2

La démonstration se base sur les six lemmes qui suivent ;

Lemme 4.1 [S_1]

> soit $A : X \longrightarrow X$ un opérateur sur X linéaire, autoadjoint, continu, fortement positif,
> alors $A(x) = x$

Lemme 4.2 [P_2, p. 272]

> soit B un opérateur sur X linéaire, auto-adjoint, continu, fortement positif,
> soit $\overline{J}(x) = J(B^{-1}(x))$ pour tout $x \in X$ (4.29)
> supposons que $\{x_i\}$ est une suite générée par la méthode du D.F.P. appliqué au problème (4.1), et $\{\overline{x}_i\}$ est une suite générée par la méthode du D.F.P. appliqué à $\overline{J}(x)$, avec $\overline{H}_0 = B.H_0.B$ dans le premier pas, et $\overline{x}_0 = B.x_0$,
> alors pour $i = 0, 1, 2, \ldots$, on a :
>
> 1. $\overline{J}(\overline{x}_i) = J(x_i)$ (4.30)
> 2. $\overline{x}_i = B.x_i$ (4.31)
> 3. $\overline{g}_i = B^{-1}.g_i$ (4.32)
> 4. $\overline{H}_i = B.H_i.B$ (4.33)
>
> où les « —— » indiquent les quantités construites par l'algorithme du D.F.P. appliqué à (4.29)

L'opérateur A défini dans le théorème 4.2 satisfait les conditions du lemme 4.1, tout comme $A^{-1/2}$; donc $A^{1/2}(x) = x$.

Posons $B = A^{1/2}$ dans le lemme 4.2. Le problème (4.29) devient dès lors :

$$\overline{J}(x) = J(B^{-1}x)$$
$$= \frac{1}{2} \cdot \langle A^{-1/2} \cdot x, A \cdot A^{-1/2} \cdot x \rangle + \langle A^{-1/2} \cdot x, w \rangle + J_0$$

ou encore

$$\overline{J}(x) = \frac{1}{2} \cdot \langle x, x \rangle + \langle x, \overline{w} \rangle + J_0 \qquad (4.34)$$

avec $\overline{w} = A^{-1/2} \cdot w$

Soient maintenant $\{x_i\}$ et $\{\overline{x}_i\}$ des suites générées par la méthode du D.F.P. appliquée au problème (4.1) et (4.34) respectivement, avec

$\overline{H}_0 = A^{1/2} \cdot H_0 \cdot A^{1/2}$ et $\overline{x}_0 = A^{1/2} \cdot x_0$.

Alors, par le lemme 4.2, on a pour $i = 0, 1, 2, \ldots$:

$$\begin{cases} \overline{x}_i = A^{1/2} \cdot x_i & (4.35) \\ \overline{g}_i = A^{-1/2} \cdot g_i & (4.36) \\ \overline{H}_i = A^{1/2} \cdot H_i \cdot A^{1/2} & (4.37) \end{cases}$$

Montrons maintenant que si le théorème 4.2 est vrai pour les suites $\{\overline{x}_i\}, \{\overline{H}_i\}$, alors il sera vrai aussi pour les suites $\{x_i\}$ et $\{H_i\}$. Cela nous permettra de travailler dans la suite avec le problème transformé $\overline{J}$.

Reprenons la relation (4.34) ; elle permet de dire que le minimum $\overline{x}^*$ est donné par :

$$\overline{x}^* = -\overline{w} = -A^{-1/2} \cdot w. \qquad (4.38)$$

Comme la relation (4.35) est vraie pour tout i, on peut dire :

$$\lim_{i\to\infty} \bar{x}_i = \bar{x}^* = \lim_{i\to\infty} A^{1/2}.x_i$$

$$= A^{1/2}.\lim_{i\to\infty} x_i$$

$$= A^{1/2}.x^*$$

ou encore

$$x^* = A^{-1/2}.x^*$$

$$= -A^{-1/2}.A^{-1/2}.w \qquad \text{(par (4.38))}$$

$$= -A^{-1}.w$$

Donc $x^* = -A^{-1}.w$; si $\bar{x}^*$ est obtenu, alors x^* est bien défini par la relation $x^* = A^{-1/2}.\bar{x}^*$ (4.39)

Si on peut prouver que la suite $\{\bar{x}_i\}$ converge vers $\bar{x}^*$ superlinéairement, c-à-d. $\dfrac{\|\bar{x}_{i+1} - \bar{x}^*\|}{\|\bar{x}_i - \bar{x}^*\|} \xrightarrow[i\to\infty]{} 0$,

alors $\{x_i\}$ converge superlinéairement aussi, c-à-d. $\dfrac{\|x_{i+1} - x^*\|}{\|x_i - x^*\|} \xrightarrow[i\to\infty]{} 0$.

En effet,

$$\frac{\|x_{i+1} - x^*\|}{\|x_i - x^*\|} = \frac{\|A^{1/2}.x_{i+1} - A^{1/2}.x^*\|}{\|A^{1/2}.x_i - A^{1/2}.x^*\|}$$

$$= \frac{\|A^{1/2}.(x_{i+1} - x^*)\|}{\|A^{1/2}.(x_i - x^*)\|}$$

$$\geq \frac{\overline{m}}{\overline{M}} . \frac{\|x_{i+1} - x^*\|}{\|x_i - x^*\|}$$

où
$$\begin{cases} \overline{m} = \inf\limits_{x\neq 0} \dfrac{\|A^{1/2}.x\|}{\|x\|} \\ \overline{M} = \sup\limits_{x\neq 0} \dfrac{\|A^{1/2}.x\|}{\|x\|} \end{cases}$$

De même pour le résultat (i) du théorème 4.2. Si il existe $0 < m' \leq M' < \infty$ t.q.

$$m' \leq \frac{\langle x, Hi.x \rangle}{\langle x, x \rangle} \leq M' \quad , \; x \neq 0 \; , \; x \in X \; , \; i = 0, 1, 2, \ldots$$

on peut alors déterminer un $\widetilde{m}$ et un $\widetilde{M}$ ($0 < \widetilde{m} \leq \widetilde{M} < \infty$) tels que :

$$\widetilde{m} \leq \frac{\langle x, Hi.x \rangle}{\langle x, x \rangle} \leq \widetilde{M} \quad , \; x \neq 0 \; , \; x \in X \; , \; i = 0, 1, 2, \ldots$$

et cela en se servant de la relation (4.37).

Toutes ces considérations nous permettent donc de travailler avec les suites $\{\overline{v}i\}$ et $\{\overline{H}i\}$, donc avec le problème défini dans le lemme 4.2, pour $B = A^{1/2}$. On peut évidemment se demander quel est l'avantage de travailler avec ce nouveau problème.
Il faut pour cela remarquer que, puisque $\overline{J}(x)$ s'écrit $\overline{J}(x) = \frac{1}{2}.\langle x, x \rangle + \langle x, \overline{w} \rangle + J_0$, l'opérateur Hessien $\overline{A}$ dans ce problème est en fait l'opérateur identité.
De ce fait, les relations suivantes sont vérifiées :

- $\overline{r}i = \overline{v}i$ (4.40)
- $\langle \overline{r}i, \overline{r}j \rangle = 0 \quad , \; i \neq j$ (4.41)
- $\overline{H}k \cdot \overline{r}i = \overline{r}i \quad , \; 0 \leq i < k$ (4.42)

(se référer aux relations (2.14) à (2.17)).
D'autre part, la formule pour $\overline{H}i+1$ se réduit à :

$$\overline{H}i+1 = \overline{H}i + \frac{\overline{r}i \rangle\langle \overline{r}i}{\langle \overline{r}i, \overline{r}i \rangle} - \frac{\overline{H}i.\overline{r}i \rangle\langle \overline{H}i.\overline{r}i}{\langle \overline{H}i.\overline{r}i, \overline{r}i \rangle} \qquad (4.43)$$

où la notation dyadique " $\rangle\langle$ " est définie :

$$(x \rangle\langle y).z = x.\langle y, z \rangle \quad \forall x, y, z \in X .$$

Remarquons enfin que dans la suite, on emploiera x_i, r_i, y_i, H_i au lieu de $\bar{x}_i, \bar{r}_i, \bar{y}_i, \bar{H}_i$ respectivement, puisqu'on ne travaille plus qu'avec les valeurs "barres", et que cela ne pourrait donc pas prêter à confusion.

Lemme 4.3 [T-2]

> supposons que $\{H_k\}$ est une suite infinie d'opérateurs sur x construits par la méthode du D.F.P., employée pour résoudre le problème (4.34) avec $H_0 = A^{1/2} \cdot H_0' \cdot A^{1/2}$, où H_0' est un opérateur arbitraire choisi tel qu'il satisfasse les hypothèses du lemme 4.1,
>
> alors
>
> (i) H_k est un opérateur linéaire, autoadjoint, positif, et
> $\langle f, H_k \cdot f \rangle = 0$ seulement si $f = 0$, $k = 0, 1, \ldots$
>
> (ii) il existe des constantes m', M',
> $0 < m' \leq M' < \infty$ t.q.
> $\forall k$, $\forall y \in X$:
> $m' \cdot \langle y, y \rangle \leq \langle y, H_k y \rangle \leq M' \cdot \langle y, y \rangle$ (4.44)

Démonstration :

(i) voir [T-2]

(ii)

* On remarque d'abord que le spectre d'un opérateur autoadjoint est entièrement situé sur l'axe réel et que son spectre résiduel est vide (voir

section 1.5 , théorème 1.17).

En outre,

$$m_i = \inf_{x \neq 0} \frac{|\langle x, H_i \cdot x \rangle|}{\|x\|^2} = \inf \{ |\alpha| , \alpha \in \text{spectre de } H_i \}$$

et

$$M_i = \sup_{x \neq 0} \frac{|\langle x, H_i \cdot x \rangle|}{\|x\|^2} = \sup \{ |\alpha| , \alpha \in \text{spectre de } H_i \}$$

existent pour chaque i.

Il est donc équivalent de montrer qu'il existe un m', M' , $0 < m' \leq M'$ t.q.

$$\begin{cases} M_i \leq M' & i = 0, 1, 2, \ldots \quad (4.45) \\ m' \leq m_i & i = 0, 1, 2, \ldots \quad (4.46) \end{cases}$$

* A partir de la définition de $H_0 = A^{1/2} \cdot H_0' \cdot A^{1/2}$, il existe (voir relation (4.3)) m_0', M_0' , $0 < m_0' \leq M_0'$ t.q.

$$m_0' \cdot \|x\|^2 \leq \langle x, H_0 \cdot x \rangle \leq M_0' \cdot \|x\|^2 , \forall x \in X$$

On va donc essayer de construire m', M' vérifiant (4.45) et (4.46)

a) <u>construction de M'</u>

* A partir de la relation (4.43) , on peut écrire :

$$H_{i+1} = R_i + P_i$$

où . $R_i = H_i - \dfrac{H_i \cdot p_i \otimes H_i \cdot p_i}{\langle H_i \cdot p_i, p_i \rangle}$

. $P_i = \dfrac{p_i \otimes p_i}{\langle p_i, p_i \rangle}$

On voit que R_i est un opérateur autoadjoint (par construction) , et que $R_i \cdot p_i = 0$. Cela veut dire que p_i est un vecteur propre de R_i , correspondant à une

valeur propre nulle.

* En outre, pour chaque $y \in X$, on a :

$$\langle y, R_i . y \rangle = \langle y, H_i . y \rangle - \frac{\langle H_i . p_i, y \rangle^2}{\langle H_i . p_i, p_i \rangle}$$

$$= \frac{\langle H_i^{1/2} . y, H_i^{1/2} . y \rangle . \langle H_i^{1/2} p_i, H_i^{1/2} p_i \rangle - \langle H_i^{1/2} p_i, H_i^{1/2} y \rangle^2}{\langle p_i, H_i . p_i \rangle}$$

$$\geq 0 \qquad \text{(par l'inégalité de Cauchy-Schwarz)}$$

et donc R_i est un opérateur autoadjoint semi défini positif.

Par définition, $\| R_i \| = \sup \langle x, R_i . x \rangle$. Donc il existe une suite $\{ z_k , \| z_k \| = 1 \}$ telle que $\lim_{k \to \infty} \langle z_k, R_i . z_k \rangle = \| R_i \|$. Comme $\langle z_k, R_i . z_k \rangle$ est réel (car un opérateur autoadjoint est symmétrique ; on utilise alors le théorème 1.15), la suite $\{ z_k \}$ doit contenir une sous-suite $\{ v_k \}$ t.q.

$$\langle v_k , R_i . v_k \rangle \xrightarrow[k \to \infty]{} \| R_i \| .$$

Mais pour chaque v_k, on a :

$$\langle v_k , R_i . v_k \rangle = \langle v_k , H_i . v_k \rangle - \frac{\langle H_i . p_i, v_k \rangle^2}{\langle H_i . p_i, p_i \rangle}$$

$$\leq \langle v_k , H_i . v_k \rangle$$

De ce fait,

$$\lim_{k \to \infty} \langle v_k, R_i . v_k \rangle \leq \lim_{\substack{k \to \infty \\ \| v_k \| = 1}} \langle v_k, H_i . v_k \rangle$$

$$\leq \| H_i \|$$

ce qui implique que $\underline{\| R_i \| \leq \| H_i \|}$ (4.47)

* On sait que $R_i . p_i = 0$, donc que p_i est un vecteur propre de R_i, correspondant à une valeur propre nulle. D'autre part, (4.42) indique que

$$H_{i+1} \cdot \psi_i = \psi_i$$

et donc ψ_i est aussi un vecteur propre de H_{i+1} avec une valeur propre de H_{i+1}, avec une valeur propre correspondante égale à un.

Soit λ un point fixe' (quelconque) du spectre ponctuel H_{i+1} t.q. $\lambda \neq 1$. Cela implique qu'il existe un élément $y \in X$ t.q.

$$H_{i+1} \cdot y = \lambda \cdot y$$

En outre, $\langle y, \psi_i \rangle = 0$ (car $\lambda \cdot \langle \psi_i, y \rangle = \langle \lambda \cdot \psi_i, y \rangle = \langle H_{i+1} \cdot \psi_i, y \rangle = \langle \psi_i, H_{i+1} \cdot y \rangle = \langle \psi_i, y \rangle$. Or, le spectre ponctuel est réel, d'où λ et 1 sont réels ; d'où $\langle \psi_i, y \rangle = 0$).

Mais,
$$\begin{aligned} \lambda y &= H_{i+1} \cdot y \\ &= (R_i + P_i) y \\ &= R_i \cdot y + P_i \cdot y \\ &= R_i \cdot y + \frac{\langle \psi_i, y \rangle}{\langle \psi_i, \psi_i \rangle} \\ &= R_i \cdot y \end{aligned}$$

Donc, λ choisi comme point du spectre ponctuel de H_{i+1} appartient aussi au spectre ponctuel de R_i.

On peut ainsi conclure :

$$\underline{\rho_s(H_{i+1}) \triangleq \max\left\{1, \rho_s(R_i)\right\}} \qquad (4.48)$$

où $\rho_s(.) = \sup\{|\alpha|, \alpha \in \text{spectre ponctuel de } (.)\}$

* Soit maintenant $\gamma \neq 1$ dans le spectre continue de H_{i+1}. Par définition du spectre continue, cela veut dire que $(H_{i+1} - \gamma . I)^{-1}$ existe et est non borné'. Donc il existe une suite $\{v_n : \|v_n\| = 1\}$ telle que

$$H_{i+1} \cdot v_n - \gamma \cdot v_n \xrightarrow[n \to \infty]{} 0 \qquad (4.49)$$

(par le lemme 1.3),

ce qui implique que

$$\gamma = \lim_{n \to \infty} \langle H_{i+1} . v_n , v_n \rangle \quad , \tag{4.50}$$

et aussi que

$$\langle H_{i+1} . v_n - \gamma . v_n , v_i \rangle \xrightarrow[n \to \infty]{} 0$$

Mais

$$\begin{aligned} \langle H_{i+1} . v_n - \gamma . v_n , v_i \rangle &= \langle H_{i+1} . v_n , v_i \rangle - \langle \gamma . v_n , v_i \rangle \\ &= \langle v_n , H_{i+1} . v_i \rangle - \gamma . \langle v_n , v_i \rangle \\ &= \langle v_n , v_i \rangle - \gamma . \langle v_n , v_i \rangle \quad \left((4.42) \text{ avec } k = i \right) \\ &= (1 - \gamma) . \langle v_n , v_i \rangle \end{aligned}$$

Comme $\langle H_{i+1} . v_n - \gamma . v_n , v_i \rangle$ tend vers 0 quand n tend vers l'∞, et comme $(1 - \gamma) \neq 0$ (on a choisi $\gamma \neq 1$), cela implique que :

$$\langle v_n , v_i \rangle \xrightarrow[n \to \infty]{} 0 \tag{4.51}$$

* Maintenant, par la définition de H_{i+1}, on a

$$\langle H_{i+1} . v_n , v_n \rangle = \langle R_i . v_n , v_n \rangle + \frac{\langle v_i , v_n \rangle^2}{\langle v_i , v_i \rangle}$$

En employant les relations (4.50) et (4.51), la dernière relation peut s'écrire

$$\gamma = \lim_{n \to \infty} \langle H_{i+1} . v_n , v_n \rangle$$

$$= \lim_{n \to \infty} \langle R_i . v_n , v_n \rangle + \lim_{n \to \infty} \frac{\langle v_i , v_n \rangle^2}{\langle v_i , v_i \rangle}$$

$$= \lim_{\substack{n \to \infty \\ \| v_n \| = 1}} \langle R_i . v_n , v_n \rangle$$

et donc

$$\underline{\gamma \leq \| R_i \|} \tag{4.52}$$

* On peut ainsi conclure :

$$\|H_{i+1}\| = M_{i+1} \leq \max\{1, \|R_i\|\}$$
$$\leq \max\{1, \|H_i\|\} \qquad (4.53)$$

(en effet :

$$\|H_{i+1}\| = \sup_{\|x\|=1} \frac{|\langle x, H_{i+1}.x\rangle|}{\|x\|^2}$$

$$= \sup\{|\alpha|, \alpha \in \text{spectre de } H_{i+1}\}$$

Mais le spectre est constitué du spectre ponctuel et continu ;

pour le spectre ponctuel, on a (4.48) :

$$\sup\{|\alpha|, \alpha \in \text{spectre ponctuel de } H_{i+1}\}$$
$$\leq \max\{1, \rho_s(R_i)\}$$

pour le spectre continu, on a (4.52), qui a pour conséquence :

$$\sup\{|\gamma|, \gamma \in \text{spectre continu de } H_{i+1}\}$$
$$\leq \max\{1, \rho_s(R_i)\}$$

D'où $\sup\{|\alpha|, \alpha \in \text{spectre de } H_{i+1}\}$

$$= \|H_{i+1}\|$$
$$\leq \max\{1, \rho_s(R_i)\}$$
$$\leq \max\{1, \|R_i\|\}$$
$$\leq \max\{1, \|H_i\|\} \quad (\text{par } (4.47)) \quad)$$

Comme (4.37) est vraie pour tout i, on a :

$$\|H_i\| \leq \max\{1, \|H_{i-1}\|\}$$
$$\leq \max\{1, \max\{1, \|H_{i-2}\|\}\}$$
$$= \max\{1, \|H_{i-2}\|\}$$
$$\leq \ldots$$
$$\leq \max\{1, \|H_0\|\},$$

ou encore

$$\|H_i\| = M_i \leq M' \ \forall\, i, \text{ où } M' = \max\{1, \|H_0\|\}.$$

le M' est fini (car H_0 satisfait aux hypothèses du lemme 4.1 ; il est donc entre autre linéaire et continue, donc borné), et indépendant de i. Ceci démontre (4.45).

b) construction de m'

* Remarquons que par la partie (i) du lemme 4.3, H_i^{-1} existe pour tout i, mais peut être ou non borné. Montrer (4.46) est équivalent de montrer que $G_i \triangleq H_i^{-1}$ ($i = 0, 1, ..$) sont uniformément bornés en norme, c-à-d. qu'il existe un $l' > 0$ pour $i = 0, 1, 2, \ldots$:

$$\| G_i \| = \sup_{x \neq 0} \frac{\langle x, G_i x \rangle}{\langle x, x \rangle} \leq l' \qquad (4.54)$$

H_{i+1} est donné par la relation (4.43).
G_{i+1} est donné par :

$$G_{i+1} = \left(I - \frac{r_i \rangle \langle r_i}{\langle r_i, r_i \rangle} \right) . G_i . \left(I - \frac{r_i \rangle \langle r_i}{\langle r_i, r_i \rangle} \right) + \frac{r_i \rangle \langle r_i}{\langle r_i, r_i \rangle} \qquad (4.55)$$

(on peut vérifier (4.55) en montrant que $H_{i+1} . G_{i+1} = G_{i+1} . H_{i+1} = I$ sous les hypothèses $H_i . G_i = I = G_i . H_i$).

* Comme $G_0 = H_0^{-1}$, $\| G_0 \| \leq \frac{1}{m_0'}$.
Réécrivons G_{i+1} comme :

$$G_{i+1} = S_i + P_i \qquad (4.56)$$

où $S_i = (I - P_i) . G_i . (I - P_i)$.
On a : $\| S_i \| \leq \| I - P_i \| . \| G_i \| . \| I - P_i \| = \| G_i \|$ (4.57)
(car $(I - P_i)$ est un opérateur de projection asymétrique avec $\| I - P_i \| = 1$)

En outre, G_{i-1} et S_i sont des opérateurs autoadjoints avec la propriété :

$$\begin{cases} G_{i+1} \cdot p_i = p_i \\ S_i \cdot p_i = 0 \end{cases}$$

* On emploie alors la même procédure que celle employée en a) pour montrer que $\{H_{i+1}\}$ est uniformément borné ; on peut dès lors montrer que pour chaque i :

$$\begin{aligned} \| G_{i+1} \| &\leq \max \{1, \| S_i \|\} \\ &\leq \max \{1, \| G_i \|\} \end{aligned} \qquad (4.58)$$

et donc

$$\begin{aligned} \| G_{i+1} \| &\leq \max \{1, \| G_i \|\} \\ &\leq \dots \\ &\leq \max \{1, \| G_0 \|\} \end{aligned}$$

or $\underline{\| G_k \| \leq l'}$, $k = 0, 1, 2, \dots$

avec $l' = \max \{1, \| G_0 \|\}$

$\leq \max \{1, \frac{1}{m_0}\}$; l' est fini et indépendant de k. ■

Lemme 4.4 [T-2]

> la suite des opérateurs $\{H_i\}$ est uniformément bornée et converge sur X vers un opérateur linéaire H^*

Soit L un sous-ensemble de X, défini par les combinaisons linéaires de η_i, $i = 0, 1, 2, \ldots$.
La fermeture de L, notée $\overline{L}$, est un sous-espace de X.

Lemme 4.5 [T-2]

> $H_i f \xrightarrow[i \to \infty]{} A^{-1} f$ pour $f \in L$, où A est l'opérateur Hermien du problème.

Les 5 lemmes précédents sont valables pour tout opérateur autoadjoint positif A et H_0. Cependant, il faut remarquer que le lemme 4.5 implique seulement la convergence ponctuelle de H_i vers A^{-1}, et pas une convergence uniforme.
Pour développer un résultat de convergence superlinéaire pour le théorème 4.2, il est nécessaire que l'opérateur A possède une structure suffisante pour permettre de montrer qu'une certaine suite converge uniformément.
C'est l'explication de l'hypothèse : $A = I + K$, où K est complètement continu, autoadjoint, positif.
En effet, comme on va le voir, cette structure de A est suffisante pour le développement du résultat voulu.

Lemme 4.6

si A est de la forme $A = I + K$, où K est un opérateur complètement continu, autoadjoint, positif,
et si : $H_0 = Id$ (pour le problème original, pas pour le problème transformé "barre")
alors $\| (H_i^{-1} - Id) . z_i \| \longrightarrow_{i \to \infty} 0$
où $z_i = \dfrac{r_i}{\| r_i \|}$

Lemme 4.7

la suite $\{x_i\}$ générée par la méthode du D.F.P appliquée au problème (4.29), avec H_0 et A définis comme dans le lemme 4.6, converge vers une solution minimale $x^* = -A^{-1/2} . w$ superlinéairement,
c-à-d. $\dfrac{\| x_{i+1} - x^* \|}{\| x_i - x^* \|} \longrightarrow_{i \to \infty} 0$

On peut faire la dernière remarque suivante : la suite des coûts $\{ J(x_i) \}$, pour le cas où $A = I + K$ et $H_0 = I$, converge aussi superlinéairement ; par calcul direct, on a :
$J(x_{i+1}) - J(x^*) = \frac{1}{2} . \langle x_{i+1} - x^* , A . (x_{i+1} - x^*) \rangle$,
et donc par (4.2),
$$\frac{J(x_{i+1}) - J(x^*)}{J(x_i) - J(x^*)} \leq \frac{M}{m} . \frac{\| x_{i+1} - x^* \|^2}{\| x_i - x^* \|^2} \longrightarrow_{i \to \infty} 0$$

■

4.4. Conclusion

On a donc montré que la suite $\{x_i\}$ générée par la méthode du D.F.P. avec $H_0 = Id$, appliquée au problème de minimisation quadratique non singulier pour le cas où $A = Id + K$, dans les espaces fonctionnels, converge vers la solution minimale x^* superlinéairement. Comme tous les algorithmes quasi-Newton générent les mêmes direction de recherche à chaque pas pour de tels problèmes, cette propriété de convergence superlinéaire est valable aussi pour tous les algorithmes quasi-Newton.

chapitre 5 :

le cas singulier, espace de dimension infinie

5.1. introduction

5.2. R(A) fermé - théorème de convergence superlinéaire.

5.3. R(A) non fermé - théorème de convergence -

Chapitre 5 : le cas singulier, espace de dimension infinie

5.1. introduction

Le but de ce chapitre est le suivant : étendre les résultats du chapitre 3 (on a montré que dans le S.Q.P., cas fini dimensionnel, les méthodes aux directions conjuguées convergent en au plus m pas, où m est le rang de la matrice Hessienne ; on y a aussi vu un résultat de vitesse de convergence : la vitesse de convergence du S.Q.P. est meilleure ou égale à la vitesse de convergence de la même méthode aux directions conjuguées, mais appliquée au problème non singulier quadratique associé) et du chapitre 4 (on a montré que les méthodes quasi-Newton convergent superlinéairement pour les N.S.Q.P. (espace fonctionnels) pour le cas $A = I + K$, où K est un opérateur autoadjoint positif, complètement continu, et avec $H_0 = I$) au S.Q.P., cas fonctionnel.

Cette généralisation se fera de la manière suivante :

1. cas où $R(A)$ est fermé : on obtiendra un résultat de vitesse de convergence.

2. cas où $R(A)$ est non fermé : on obtiendra un autre résultat de vitesse de convergence.

Ensuite, un certain nombre d'approches pour le développement d'une vitesse de convergence (cas où $R(A)$ non fermé) seront présentés.

5.2. R(A) fermé - théorème de convergence superlinéaire

1] On considère le problème S.Q.P., non fonctionnel,

où $$J(x) = \frac{1}{2}.\langle x, Ax\rangle + \langle x, w\rangle + J_0 \qquad (5.1)$$

avec
- $w \in R(A)$
- A un opérateur autoadjoint, linéaire, semi défini positif, borné
- R(A) fermé

Comme A est semi-défini positif,

$$0 \leq \langle x, Ax\rangle$$

et comme A est borné,

$$\exists M > 0 : \sup_{x \neq 0} \frac{\langle x, Ax\rangle}{\|x\|} \leq M$$

ou encore $\langle x, Ax\rangle \leq M.\langle x, x\rangle$

Donc, finalement,

$$\exists 0 < M < \infty \text{ t.q. } \underline{0 \leq \langle x, Ax\rangle \leq M.\langle x, x\rangle} \ \forall x \in X \qquad (5.2)$$

2] On va évidemment tirer parti du fait que R(A) est fermé.

On sait que $A^+ = A_n^{-1} P_n$,

où P_n est la projection orthogonale sur $\overline{R(A)}$,

et A_n^{-1} la restriction de A au sous-espace $(N(A))^\perp$

(voir théorème 3.2) ;

on sait aussi que A^+ est borné ssi R(A) est fermé

(voir théorème 1.11). Or, R(A) est fermé. on peut en conclure que A^+ est borné.

On voit aussi que A^+ est borné ssi A_n^{-1} est borné.

Mais A est un opérateur autoadjoint : $A = A^*$.

D'où $A_n^* = A_n$ et $A_n^{-1} = (A_n^{-1})^*$

Tout ceci montre que A_n est un opérateur linéaire autoadjoint fortement positif de $(N(A))^\perp = \overline{R}(A) = R(A)$ (puisque l'inverse existe, borné)

Donc, $\exists \; 0 < m_n \leq M_n < \infty$ t.q.

$$m_n \cdot \langle x, x \rangle \leq \langle x, A_n \cdot x \rangle \leq M_n \cdot \langle x, x \rangle \quad \forall x \in R(A) \qquad (5.3)$$

3] Voici maintenant le théorème annoncé :

Théorème 5.1

(i) soit $\{x_i\}$ une suite d'éléments de x générée par la méthode du D.F.P., appliquée au S.Q.P. avec $H_0 = Id$.

si $R(A)$ est fermé, alors $\{x_i\}$ converge vers une solution $\overline{x} = x^* + (I - P_n) x_0$,

où $x^* = -A^+ w$ (l'unique solution en norme minimale) et x_0 est l'estimation initiale,

(ii) et cela avec une vitesse de convergence :

$$\| x_i - \overline{x} \|^2 \leq \frac{2}{m_n} \cdot [f(x_0) - f(\overline{x})] \cdot \left[1 - \frac{m_n}{M_n}\right]^i \qquad i = 0, 1, 2, \ldots$$

(iii) en outre,

si $A_n = I_n + K_n$ où K_n est un opérateur linéaire, autoadjoint, complètement continu, positif de $\overline{R}(A)$ dans $R(A)$

alors $\{x_i\} \longrightarrow \overline{x}$ superlinéairement .

démonstration :

* remarquons que, étant donné que

$$\begin{cases} H_0 = Id \text{ est fortement positif} \\ R(A) \text{ est fermé} \end{cases},$$

on a $H_0(R(A)) = R(A)$

ssi $R(H_0 A) = R(A)$ (5.4)

* dans [H-4], il a été montré que

$$-s_k = H_0 \cdot \sum_{j=0}^{k} \frac{\langle H_k \cdot g_k, H_k \cdot g_k \rangle}{\langle g_j, H_0 \cdot g_j \rangle} \cdot g_j$$

Utilisons la relation (2.12), c-à-d. que la suite $\{x_i\}$ générée par le D.F.P. satisfait $x_i = x_0 + \sum_{k=0}^{i-1} \alpha_k \cdot s_k$, et le fait que $g_j \in R(A)$, $\forall j$.

De cette manière,

$$x_i = x_0 + \sum_{k=0}^{i-1} \alpha_k \cdot s_k$$

$$= x_0 + \sum_{k=0}^{i-1} \alpha_k \cdot \left(-H_0 \cdot \sum_{j=0}^{k} \frac{\langle H_k \cdot g_k, H_k \cdot g_k \rangle}{\langle g_j, H_0 \cdot g_j \rangle} \cdot g_j \right)$$

$$= x_0 + \sum_{k=0}^{i-1} \alpha_k \cdot H_0(R(A))$$

On peut ainsi conclure que <u>la méthode du D.F.P. génère une suite $\{x_i\}$ dans E, où</u>

<u>$E = \{ x : x = x_0 + x^r, x^r \in R(A) \}$</u>

* Considérons maintenant $T = P_R|_E$, c-à-d. la restriction de P_R sur E. Utilisons la méthode du D.F.P., mais en démarrant avec $x_0' = T \cdot x_0$, et en employant A_R au lieu de A.

La méthode génère une suite $\{x'_i\}$ reliée à x_i par

$$x'_i = T.x_i \qquad , i = 0, 1, 2, \ldots$$

En outre, étant donné que l'opérateur A_n employé est un opérateur linéaire, autoadjoint, fortement positif de $R(A)$ dans $R(A)$, et que $H_0 = Id$, le résultat de convergence établi en (4.7) est valable :

$$\| x_i - x^* \|^2 \leq \frac{2}{m} \cdot \left[J(x_0) - J(x^*) \right] \cdot \left[1 - \frac{m}{M} \right]^i$$

Donc ici

$$\| x'_i - x^* \|^2 \leq \frac{2}{m_n} \cdot \left[J(x_0') - J(x^*) \right] \cdot \left[1 - \frac{m_n}{M_n} \right]^i , i = 0, 1, \ldots \qquad (5.5)$$

L'idée est donc de se ramener à la situation du chapitre précédent : on peut ainsi en appliquer les théorèmes.

Puisqu'on est dans les conditions d'application du chapitre 4, on peut utiliser le théorème 4.2, point ii). Donc, si $A_n = I_n + K_n$, où K_n est linéaire, autoadjoint, complètement continu, positif, on a que la suite $\{x'_i\}$ converge superlinéairement vers

$$x^* = -A \tilde{\lambda}^{\wedge} w$$
$$= -A \tilde{\lambda}^{\wedge} P_n(w)$$
$$= -A^+ w \qquad ,$$

c-à-d.

$$\lim_{k \to \infty} \sup \frac{\| x'_{k+n} - x^* \|}{\| x'_k - x^* \|} = 0 \qquad (5.6)$$

* reprenons la relation suivante :

$$x_i = x_0 + \sum_{k=0}^{i-1} \alpha_k . s_k$$

On peut la réécrire sous la forme

$$x_i = P_n x_0 + \sum_{k=0}^{i-1} \alpha_k . s_k + (I - P_n) x_0 \qquad (5.7)$$

où $\left(P_n x_0 + \sum_{k=0}^{i-1} \alpha_k . s_k \right) \in R(A)$

et $(I - P_n) . x_0 \in N(A)$

Donc, $x'_i = Tx_i$

$$= T\Big[\underbrace{P_n x_0 + \sum_{k=0}^{i-1} \alpha_k . s_k}_{\in R(A)} + \underbrace{(I - P_n) x_0}_{\in N(A)}\Big]$$

$$= P_n x_0 + \sum_{k=0}^{i-1} \alpha_k . s_k \quad \text{(par définition de } T\text{)} \quad (5.8)$$

Comparons (5.8) et (5.7), c-à-d.

$$\begin{cases} x_i = P_n x_0 + \sum_{k=0}^{i-1} \alpha_k . s_k + (I - P_n) x_0 \\ x'_i = P_n x_0 + \sum_{k=0}^{i-1} \alpha_k . s_k \end{cases}$$

On peut en conclure que

$$x_i - (I - P_n x_0) = x'_i$$

ou encore

$$\underline{x_i = x'_i + (I - P_n x_0)}$$

La suite $\{x_i\}$ converge donc vers $x^* + (I - P_n) x_0$ (la convergence est superlinéaire si $A_n = I_n + K_n$, puisqu'alors $\{x'_i\}$ converge superlinéairement vers x^*).
C'est le résultat (i) du théorème.

* Comme $x_i = x'_i + (I - P_n) x_0$, on a

$$|x_i - \bar{x}| = |x'_i + (I - P_n) x_0 - \bar{x}|$$

$$= |x'_i + (I - P_n) x_0 - x^* - (I - P_n) x_0|$$

$$= |x'_i - x^*|$$

En employant (5.5) et (5.6), on obtient

$$\|x_i - \bar{x}\|^2 \leq \frac{2}{m_n} . [J(x_0) - J(\bar{x})] . \left[1 - \frac{m_n}{M_n}\right]^i , \; i = 0, 1, \ldots \quad (5.9)$$

pour un opérateur A fortement positif, autoadjoint :
c'est le résultat (ii) du théorème.

* Enfin, $\lim\sup_{k \to \infty} \frac{\|x_{k+1} - x\|}{\|x_k - x\|} = \lim\sup_{k \to \infty} \frac{\|x'_{k+1} - x^*\|}{\|x'_k - x^*\|} = 0 \quad (5.10)$

si $A_n = I_n + K_n$. C'est le résultat (iii) du théorème ■

Remarque :

ce théorème a été démontré pour $H_0 = Id$. En fait le résultat (ii) de convergence, c-à-d. la relation (5.9), est valable pour tout H_0 satisfaisant $R(H_0 A) = R(A)$, donc entre autre pour $A = H_0$ et $H_0 = P_n$. Cependant, (5.10) n'est vraie que si $H_0 = Id$ (et bien sûr $A_n = I_n + K_n$)

5.3. R(A) non fermé - théorème de convergence.

1] Dans la section précédente, on a montré que la convergence de la suite $\{x_i\}$ est superlinéaire pour le cas où $R(A)$ est fermé, avec $H_0 = Id$ et $A_n = I_n + K_n$.
Cependant, certains problèmes (notamment le problème de contrôle optimal singulier transformé en un problème de minimisation sur un espace fonctionnel n'ont pas une image fermée.
D'où l'intérêt de trouver un résultat de convergence pour la suite $\{x_i\}$ et d'essayer de discuter des vitesses de convergence pour cette suite.

2] Théorème 5.2

soit $\{x_i\}$ une suite d'éléments générée par la méthode du D.F.P. appliquée au S.Q.P., avec $H_0 = Id$, et $x_0 = y_0 - z_0 \in X$, où $z_0 \in N(A)$ et $y_0 \in \overline{R(A)}$ (x_0 étant l'approximation initiale du point minimal),
alors $\{x_i\}$ converge vers une solution $\bar{x} = z_0 + x^*$, où $x^* = -A^+ w$ est l'unique solution de norme minimale.

démonstration :

* On a $H_0 = Id$; utilisons (5.7) et (5.8), c-à-d.

$$\begin{cases} x_i = P_n x_0 + \sum_{k=0}^{i-1} \alpha_k . s_k + (I - P_k) . x_0 \\ x'_i = T . x_i = P_n x_0 + \sum_{k=0}^{i-1} \alpha_k . s_k \end{cases}$$

pour conclure :

$$\begin{aligned} x'_i &= x_i - (I - P_n) x_0 \\ &= x_i - I . x_0 - P_n x_0 \\ &= x_i - H_0 . x_0 - P_n x_0 \end{aligned}$$

d'où $x_i = x'_i + (H_0 - P_n) x_0$,

avec $(H_0 - P_n) x_0 \in N(A)$.

(car $\begin{aligned} A (H_0 - P_n) x_0 &= A x_0 - A P_n x_0 \\ &= A x_0 - A y_0 \\ &= A y_0 - A z_0 - A y_0 \\ &= 0 \quad \text{puisque } z_0 \in N(A) \end{aligned}$)

On peut donc écrire

$$\boxed{x_i = x'_i + z_0} \quad , \text{ où } x'_i \in R(A) \text{ et } z_0 \in N(A)$$

Or, on veut montrer que $\{x_i\}$ converge vers $\overline{x} = z_0 + x^*$.
Il suffira donc de montrer que $\{x'_i\}$ converge vers
$x^* = -A^+ . w$, l'unique solution de norme minimale.

* On travaille avec un opérateur autoadjoint, semi défini positif, borné ; on a donc que

$$\begin{cases} A = A^* \geq 0 \\ \exists M > 0 \text{ t.q. } 0 \leq \langle x, Ax \rangle \leq M . \langle x, x \rangle \quad \forall x \in X \end{cases}$$

d'où

$$\underline{0 \leq \langle x, A_n x \rangle \leq M . \langle x, x \rangle} \quad \underline{\forall x \in \overline{R}(A)}$$

En outre, $A^{1/2}$ existe, et $(A^{1/2})^* = A^{1/2} \geq 0$. On peut donc appliquer à $A^{1/2}$ la relation précédente :

$$0 < \langle x, A_n^{1/2} x \rangle \leq M^{1/2} \cdot \langle x, x \rangle \quad \forall x \in \overline{R}(A^{1/2})$$

où $$M^{1/2} = \sup_{\|x\| \neq 0} \left[\frac{\|A^{1/2} x\|^2}{\|x\|^2} \right]^{1/2}$$

Donc, $(A_n^{1/2})^{-1} = (A_n^{-1/2})$ existe sur $R(A^{1/2})$, et $(A_n^{-1/2})^* = (A_n^{-1/2})$.

Cela veut dire que pour chaque $y \in R(A^{1/2})$, il existe un x unique $\in \overline{R}(A^{1/2})$ t.q. $y = A_n^{1/2} \cdot x$

Cela implique immédiatement que

$$\forall y \in R(A^{1/2}),$$

$$\frac{\langle A_n^{1/2} \cdot y, A_n^{-1/2} \cdot y \rangle}{\langle y, y \rangle} = \frac{\|A_n^{-1/2} \cdot A_n^{1/2} \cdot x\|^2}{\|A_n^{1/2} \cdot x\|^2}$$

$$= \frac{\langle x, x \rangle}{\langle x, A_n x \rangle}$$

$$\geq \frac{1}{M}$$

ou encore

$$\boxed{\frac{1}{M} \cdot \langle y, y \rangle \leq \langle A_n^{1/2} y, A_n^{-1/2} y \rangle} \quad \forall y \in R(A^{1/2}) \quad (5.11)$$

* L'idée que l'on développe alors est la suivante : on définit une nouvelle norme dans $R(A^{1/2})$ t.q. $R(A^{1/2})$ devienne un espace de Hilbert. On considère alors une fonction G qui transformera le problème S.Q.P. dans $\overline{R}(A)$ en un problème de norme minimale dans $Y = R(A^{1/2})$.

Voyons cela plus en détail.

1. Soit $Y = R(A^{1/2})$. On définit un nouveau produit interne $\langle ., . \rangle'$ comme suit :

$$\langle y_1, y_2 \rangle' = \langle A_n^{-1/2} y_1, A_n^{-1/2} y_2 \rangle \quad \text{où } y_1, y_2 \in R(A^{1/2}) \quad (5.12)$$

2. avec x restreint à $\overline{R}(A)$, le S.Q.P. peut se réécrire comme

$$\underline{J(x) = \frac{1}{2} \cdot \langle x, A_r x \rangle + \langle x, w \rangle + J_0} \qquad (5.13)$$

où $x \in R(A)$, $w \in R(A_r)$

Comme $A_r : \overline{R}(A) \longrightarrow R(A_r)$, et comme $w \in R(A_r)$,

$\exists u \in \overline{R}(A)$ t.q. $w = A_r . u$.

On peut donc ainsi écrire que le gradient $g(x)$ vaut

$g(x) = A_r . x + w = A_r x + A_r u = A_r (x + u)$

3. On définit alors la fonction $G : \overline{R}(A) \longrightarrow Y$, continue, injective :

$$y = G(x) = A_r (x + u) \qquad (5.14)$$

Pour transformer le problème dans Y, la fonctionnelle $J(x)$ est remplacée par $J[G^{-1}(y)]$ (on suppose que celle-ci a exactement les mêmes valeurs sur Y que la forme sur $\overline{R}(A)$, et par conséquent a la même valeur minimale).

Etant donné que $x = G^{-1} y = A \bar{r}^1 . (y - w)$ (5.15)

alors

$$\begin{aligned} J(x) &= J(G^{-1} y) \\ &= \frac{1}{2} . \langle A \bar{r}^1 (y - w), (y - w) \rangle + \langle A \bar{r}^1 (y - w), w \rangle \\ &\quad + J_0 \\ &= \frac{1}{2} . \langle y, A \bar{r}^1 y \rangle - \frac{1}{2} . \langle A \bar{r}^1 w, w \rangle + J_0 \end{aligned}$$

ou encore

$$\underline{J(G^{-1} y) = \frac{1}{2} . \langle y, y \rangle' + J_1} \qquad (5.16)$$

où $y \in Y$ et $J_1 = J_0 - \frac{1}{2} . \langle w, w \rangle'$

On constate donc que G transforme le S.Q.P. dans $\overline{R}(A)$ en un problème de norme minimale dans Y, puisque on peut encore écrire (5.16) sous la forme

$$F(G^{-1}(y)) = J_1 + \frac{1}{2} . \| y \|^2$$

On peut en outre montrer les 2 propriétés suivantes (démontrées dans [5_1]), dont on se servira pour la suite de la démonstration du théorème 5.2.

Lemme 5.1

> Soit $y_0 = G(x_0) = A x_0 + w$
>
> $y_i = G(x_i) = A x_i + w$
>
> alors
>
> (i) $y_m = G(x_m)$ est la projection orthogonale de y_0 sur le complément orthogonal de $\{As_0, As_1, \dots, As_{m-1}\}$ dans Y.
>
> (ii) pour chaque $m = 0, 1, 2, \dots$, les vecteurs $As_0, As_1, \dots, As_m$ engendrent le même sous-espace que les vecteurs $Ay_0, A^2 y_0, \dots A^{m+1} y_0$ dans Y.

idée de la démonstration :

On sait que l'algorithme du D.F.P. génère une suite de points $x_0, x_1, x_2, \dots$. A la m ième étape de l'algorithme, on détermine une direction de recherche s_m. Un point de l'itération est donc numériquement déterminé le long de la demi-droite

$$L_m^+ = \{x_m + \alpha . s_m, \ \alpha \geq 0\}$$ qui minimise la fonction J le long de L_m^+.

On note ce point x_{m+1} et on continue la procédure. Par la transformation G, la demi-droite L^+ dans X est transformée en une demi-droite Λ^+ dans Y.
En fait, pour $x_m + \alpha . s_m \in L^+$, on a :

$$G(x_n + \alpha . s_n)$$
$$= A(x_n + \alpha . s_n) - w$$
$$= (Ax_n - w) + \alpha . A . s_n$$
$$= G(x_n) + \alpha . A . s_n$$

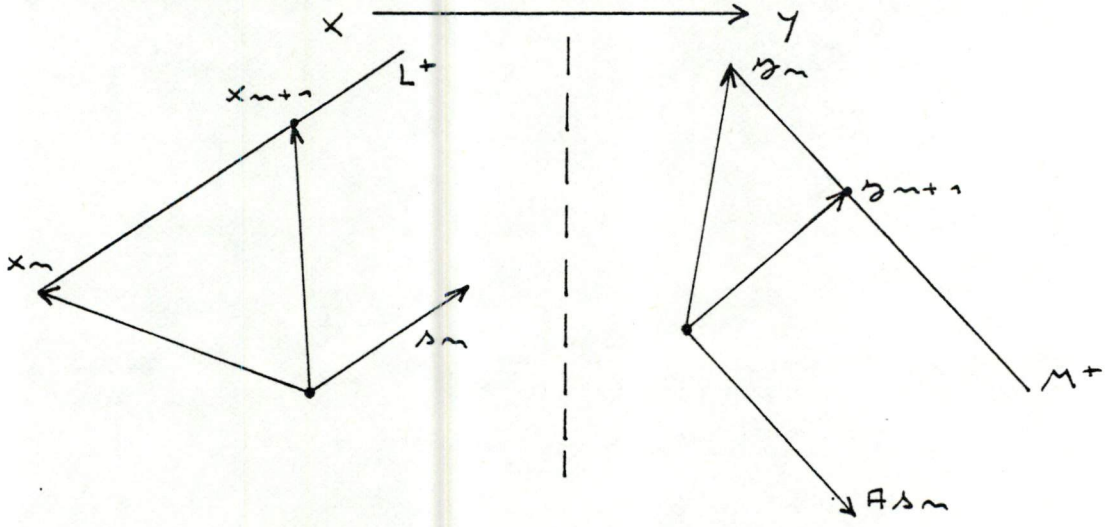
de telle sorte que

$$M^+ = \{ G(x_n) + \alpha . A . s_n \ , \alpha \geq 0 \}$$

Mais n'oublions pas qu'on a vu que, dans Y, la minimisation de J est en fait la minimisation de la norme. Par conséquent, la recherche linéaire dans X détermine un point x_{n+1} qui est envoyé sur le point $y_{n+1} = G(x_{n+1})$ dans Y, de norme minimale sur M^+.

Or, on sait qu'un tel point est orthogonal à la droite M^+, qui est dans la direction de As_n [voir APP. III]

La situation peut être décrite par le schéma suivant :



D'autre part, on peut montrer que dans l'algorithme du D.F.P., avec J fonction quadratique, les directions de recherche $s_0, s_1, \ldots$ sont A conjuguées dans X ($< s_i, A . s_j > = 0$ pour $i \neq j$). Que devient cette notion dans Y ?

On peut montrer que c'est exactement équivalent à dire que les directions sont orthogonales dans $\mathcal{Y}$.
A partir de là, on peut conclure ;
on a vu que $G(x_1)$ n'a pas de composante le long de $A s_0$, dans le sens où $G(x_1) \perp A s_0$. Etant donné que $A s_1$ est $\perp$ à $A s_0$, et que $G(x_2) = G(x_1)$ à un multiple scalaire de $A s_1$ près, on voit que $G(x_2)$ est $\perp$ à $A s_0$ et à $A s_1$. Par induction, on prouve ainsi facilement que si les directions de recherche sont A conjuguées dans x, alors pour $n = 0, 1, 2, \ldots$
$y_{n+1} = G(x_{n+1})$ est orthogonal à chacune $A s_0$, $A s_1, \ldots, A s_n$ dans $\mathcal{Y}$.
En outre, $G(x_{n+1})$ diffère de $G(x_0)$ par une combinaison linéaire de $A s_0, A s_1, \ldots, A s_n$, et une telle combinaison est elle même orthogonale à $G(x_{n+1})$.
D'où on peut conclure :
pour chaque $n = 0, 1, 2, \ldots$,
$G(x_{n+1})$ est la projection orthogonale de $G(x_0)$ sur le complément orthogonal de $A s_0, A s_1, \ldots A s_n$ (si $s_0, s_1 \ldots$ sont A conjugués dans x). ■

Lemme 5.2

> soit $y \in \mathcal{Y}$, et soit $\overline{S}$ le sous-espace fermé engendré par les points $Ay, A^2y, A^3y, \ldots$
> alors $y \in \overline{S}$

Démonstration :

Elle se base sur la proposition suivante : [voir [APP. IV]]

> Soit $A : X \longrightarrow X$ un opérateur linéaire continu sur l'espace de Hilbert X ;
> si N est un sous espace (fermé) invariant de A, alors $N^\perp$ est un sous espace invariant de A^*.

Le sous espace linéaire S généré par les points $Ax, A^2x, \dots$ est constitué de combinaisons linéaires finies de ces points. Il est évident que toute image par A d'une telle combinaison est une autre combinaison, et donc que S est un sous-espace invariant de A.
Donc, $\overline{S}$ est aussi un sous-espace invariant de A (par la proposition).
D'autre part, $S^\perp$ est un sous-espace invariant de $A^* = A$ (A autoadjoint), par la proposition.
Comme $X = \overline{S} \oplus \overline{S}^\perp$, un élément $x \in X$ se décompose en $x = y + y^\perp$, avec $y \in \overline{S}$ et $y^\perp \in \overline{S}^\perp$.
Donc $Ax = Ay + Ay^\perp$, avec $Ay \in \overline{S}$ et $Ay^\perp \in \overline{S}^\perp$.
Comme $Ax \in \overline{S}$ (A est un des générateurs de S), $Ay^\perp = Ax - Ay \in \overline{S}$, et donc $Ay^\perp = 0$.
Ceci donne $Ax = Ay$, donc $x = y$. Comme $y \in \overline{S}$, $x \in \overline{S}$. ∎

Suite de la démonstration du théorème 5.2

* par le lemme 5.1, on peut conclure que la suite $\{ y_n = G(x_n) \}$ converge vers la projection orthogonale

de y_0 dans le complément orthogonal de la fermeture de l'espace engendré par $\{A.y_0, A^2.y_0, \dots\}$.
Mais on a montré dans le lemme 5.2 que la fermeture de l'espace engendré par $\{A.y_0, A^2.y_0, \dots\}$ doit contenir y_0. Et donc la projection orthogonale de y_0 dans $\mathcal{I}^\perp$ est le vecteur nul, c-à-d.

$$\lim_{n \to \infty} y_n = 0 \quad \text{dans } \mathcal{Y}$$

ou encore

$$\lim_{n \to \infty} \langle y_n, y_n \rangle' = 0 \qquad (5.17)$$

Mais $\langle y_n, y_n \rangle' = \langle A^{-1/2}.y_n, A^{-1/2}.y_n \rangle$.
Donc, en tenant compte des relations (5.17), (5.11), on a :

$$\frac{1}{M} \cdot \lim_{n \to \infty} \langle y_n, y_n \rangle \leq \lim_{n \to \infty} \langle y_n, y_n \rangle' = 0$$

et finalement :

$$\lim_{n \to \infty} \| y_n \| = 0 \quad \text{dans } \overline{R}(A) \qquad (5.18)$$

---

* Etant donné que l'opérateur A n'est pas fortement positif, la relation (5.18) ne permet de tirer aucune conclusion quant à la convergence de la suite $\{x_i\}$.
Cependant, on va se servir du résultat suivant.
On a montré (lemme 3.2 de [K-3]) que la suite des éléments $[x_i]$ générée par la méthode du gradient conjugué appliquée au problème des moindres carrés

$$\frac{1}{2} \cdot \| A^{1/2}.x + \overline{\omega} \|$$

satisfait la relation :

$$\| x_{i+1} - x^* \|^2 \leq \| x_i - x^* \|^2 - \frac{1}{2} . \alpha_i . \| A^{1/2} . x_i + \bar{w} \|^2 \qquad (5.19)$$

où . $\bar{w} = (A^{1/2})^+ . w$

. $x^* = - (A^{1/2})^+ . \bar{w} = - A^+ . w$

Etant donné que les méthodes du C.G. et du D.F.P., avec $H_0 = Id$ (ce qui est une hypothèse du théorème ici), génèrent la même suite d'éléments $\{x_i\}$, et que

$$f(x) = \frac{1}{2} . \langle x, Ax \rangle + \langle x, w \rangle + f_0$$

$$= \frac{1}{2} . \| A^{1/2} . x + \bar{w} \| + f_0 - \frac{1}{2} . \langle \bar{w}, \bar{w} \rangle \qquad (5.20)$$

il suit que (5.19) est aussi valide pour la méthode du D.F.P.

Etant donné que $\alpha_i > 0$ et $\| A^{1/2} . x_i + \bar{w} \| > 0$, $\forall i$, la relation (5.19) entraîne

$$\| x_{i+1} - x^* \| < \| x_i - x^* \| \qquad , i = 0, 1, 2, \ldots \qquad (5.21)$$

ou encore que la suite $\{ \| x_i - x^* \| \}_i$ est une suite de nombres réels strictement positifs, décroissants.

De ce fait, <u>la limite de cette suite existe</u>.

En outre, étant donné que

$$\| x_i \| \leq \| x_i - x^* \| + \| x^* \|$$

$$\leq \| x_0 - x^* \| + \| x^* \| \qquad \text{(par (5.21))}$$

$$= c \qquad , i = 0, 1, 2, \ldots ,$$

<u>la suite $\{x_i\}$ est une suite bornée.</u>

* On a donc montré que

$$\lim_{i \to \infty} \| x_i - x^* \| \quad \text{existe}$$

et $\lim_{n \to \infty} \| g_n \| = 0$.

Cependant, pour prouver que $\lim_{i \to \infty} \| x_i - x^* \| = 0$,

la propriété géométrique suivante pour $\{x_i\}$ doit être employée :

une propriété connue du D.F.P. est que x_i minimise la fonctionnelle $f(x)$ à

$$x_0 + \overline{S}_i \qquad (S_i \text{ est l'espace engendré par } \{s_0, s_1, \ldots, s_{i-1}\})$$

étant donné que l'opérateur A est positif sur $\overline{R}(A)$, et que $\overline{S}_i \subset \overline{R}(A)$,

il suit que x_i est uniquement déterminé, et $x_i \in x_0 + \overline{S}_i$.

En outre, étant donné que $\|x_i\| \leq c$ et que l'élément $\overline{\overline{x}}$, qui minimise $f(x)$ sur $x_0 + \widehat{S}$, où $\widehat{S}$ est l'espace fermé engendré par $\{s_0, \ldots\}$, existe et est uniquement déterminé,

alors la suite $\{x_i\}$ doit converger vers $\overline{\overline{x}}$, c-à-d.

$$\lim_{i \to \infty} x_i = \overline{\overline{x}} \qquad (5.22)$$

Il suit alors que $\overline{\overline{x}} = x^*$.

Montrons le par contradiction, c-à-d. supposons que $\overline{\overline{x}} - x^* \neq 0$.

étant donné que $x^* \in \overline{R}(A)$, $\overline{\overline{x}} \in x_0 + \widehat{S} \subseteq \overline{R}(A)$,

alors $(\overline{\overline{x}} - x^*) \in \overline{R}(A)$

et

$\langle \overline{\overline{x}} - x^*, A.(\overline{\overline{x}} - x^*) \rangle > 0$.

En appliquant (5.22), (5.21), (5.18),

on obtient :

$$0 \leq \langle \bar{\bar{x}} - x^*, A.(\bar{\bar{x}} - x^*) \rangle$$

$$= \lim_{i \to \infty} \langle x_i - x^*, A.(x_i - x^*) \rangle$$

$$= \lim_{i \to \infty} \langle x_i - x^*, g_i \rangle$$

$$\leq \lim_{i \to \infty} \|x_i - x^*\| . \|g_i\|$$

$$\leq \|x_0 - x^*\| . \lim_{i \to \infty} \|g_i\|$$

$$= 0 \quad ,$$

ce qui donne la contradiction.

Donc : $\lim_{i \to \infty} x_i = \bar{\bar{x}} = x^*$ ■

3] Le théorème 5.2 montre que la suite $\{x_i\}$ générée par la méthode du D.F.P., avec $H_0 = Id$, converge vers $\bar{x} = x^* + z_0$, où $z_0 = (I - P_n) x_0 \in N(A)$, et lorsque $R(A)$ est non fermé.

Mais il ne nous apprend rien à propos de la vitesse de convergence de cette suite.

A ce propos, un théorème donne le résultat de vitesse de convergence suivant pour la suite $\{x_i\}$:

$\{x_i\}$ converge vers $\bar{x} = (I - P_n) x_0 + x^*$

où $x^* = -A^+ w \in R(A)$,

ET $\|x_i - \bar{x}\|^2 \leq \dfrac{M . \|x_0 - x^*\|^2 . \|(A^{1/2})^+ . (x_0 - x^*)\|^2}{M . \|(A^{1/2})^+ . (x_0 - x^*)\|^2 + i . \|x_0 - x^*\|^2}$ (5.23)

pour $i = 1, 2, \ldots$,

où $M = \|A\|$,

MAIS sous les restrictions suivantes :

$w \in R(A^2)$ ET $x_0 \in R(A) + N(A)$

4] On pourrait se poser la question suivante : on travaille avec le cas singulier, et on sait qu'alors l'opérateur pseudo-inverse A^+ joue le rôle de l'opérateur inverse dans le cas non singulier. Pourquoi alors ne pas développer un résultat de vitesse de convergence superlinéaire par l'approche du chapitre 4, avec A^+ jouant le rôle de A^{-1} ?

En fait, la démonstration du théorème 4.2 (donc du résultat de convergence superlinéaire) dépend fortement de la propriété de convergence uniforme de la suite $\{H_k\}$. On a en effet besoin du lemme 4.5, qui affirme que H_k converge vers A^{-1} uniformément sur $\overline{L}$.

Or, on a montré [dans [N-2], p. 38-43-38-6 3][illegible] que des séries d'approximation du pseudo-inverse A^+ convergent uniformément ssi l'opérateur A a une image fermée. Il est donc évident que le résultat de vitesse de convergence superlinéaire ne peut être développé par l'approche du chapitre 4.

5] Dans la section précédente, on a montré que tous les résultats de vitesse de convergence pour le cas non singulier s'appliquent au cas singulier si $R(A)$ est fermé.

Pour $R(A)$ non fermé, n'est-il pas possible de se ramener au cas précédent, et à partir de là déterminer une vitesse de convergence ?

Plus exactement, examinons les 2 questions suivantes :

1. est-il possible de changer la topologie de l'espace image, disons Y, t.q. l'opérateur original $A : X \longrightarrow Y$ a une image fermée dans Y (avec la nouvelle topologie) ?

réponse : oui.
En fait, si on considère $A^{1/2} : X \longrightarrow Y$, où $Y = R(A^{1/2}) + N(A^{1/2})$, avec le nouveau produit interne $\langle \cdot, \cdot \rangle'$ (déjà rencontré en (5.12)) :
$\langle y_1, y_2 \rangle' = \langle A^{-1/2} y_1, A^{-1/2} y_2 \rangle$,
on peut montrer que Y est lui même un espace de Hilbert, et que $A^{1/2}$ a une image fermée dans Y : [B-1]
En outre, la fonctionnelle $f(x)$ se réécrit :

$$\bar{f}(x) = \frac{1}{2} \cdot [\, \| A^{1/2} \cdot x + w \|' \,] \quad , x \in X \qquad (5.24)$$

2. est-il possible de dériver une vitesse de convergence pour la méthode du D.F.P. appliquée au S.O.P. où partir d'un résultat du D.F.P. appliqué au problème (5.24) ?

réponse : non.
Pour dériver une vitesse de convergence, il est nécessaire de développer une relation entre les 2 us. Par exemple, soient $\{x_i\}$ et $\{\bar{x}_i\}$ des suites générées par le D.F.P. appliqué respectivement au S.O.P. (5.1) et au problème (5.24) :
$\min \bar{f}(x) = \frac{1}{2} \cdot [\, \| A^{1/2} x - w \|' \,]^2$, $x \in X$
où $\bar{x}_0 = A^{1/2} \cdot x_0$, $\bar{H}_0 = A^{1/2} \cdot H_0 \cdot A^{1/2} = A$

Dès lors, les relations (4.35) et (4.36) restent valables.
Cependant, on voit que le problème (5.24) "contient" un opérateur singulier ($\bar{H}_0$ qui est égal à A), et par conséquent les résultats connus pour le cas non singulier ne peuvent pas être appliqués.
Par conséquent, aucun résultat ne peut être obtenu pour le cas singulier par un changement de topologie.

Remarquons pour terminer que l'approche par un changement de topologie est souvent utilisée pour les questions d'existence et d'unicité des solutions optimales pour les problèmes avec image non fermée.

ANNEXES

* On était donc parvenu au résultat suivant :
si on emploie y_{i-1}, z_{i-1} vérifiant (2.48), l'équation (2.47) donne la valeur de ΔH_{i-1}, et (2.42) donne la matrice H_i.
Maintenant, que peut-on dire d'autre à propos de y_{i-1} et de z_{i-1} ? On veut que la propriété 4. soit vérifiée, c-à-d. que l'algorithme emploie uniquement l'information présente et celle de l'itération précédente. On va donc essayer de trouver les vecteurs y_{i-1} et z_{i-1} en utilisant seulement l'information disponible à cette itération et au point le précédent immédiatement.

Tout d'abord, on note que la condition (2.32), c-à-d.

$$s_i^T . A . s_j = 0 \qquad , i-1 \geq j \geq 0 \ ,$$

à l'itération $i-1$ donne :

$$s_{i-1}^T . A . s_j = 0 \qquad , i-2 \geq j \geq 0 \qquad (A.1)$$

Reprenons les relations

$$\Delta x_i = -\alpha_i . s_i$$

et $\quad y_{i+1} = y_i + A . \Delta x_i$,

(A.1) devient $s_{i-1}^T . A . s_j = 0$

$$\iff \alpha_{i-1} . s_{i-1}^T . A . \alpha_j . s_j = 0$$

$$\iff \Delta x_{i-1}^T . A . \Delta x_j = 0$$

$$\iff \Delta x_{i-1}^T . (y_{j+1} - y_j) = 0$$

$$\iff \underline{\Delta x_{i-1}^T . \Delta y_j = 0} \qquad , i-2 \geq j \geq 0 \qquad (A.2)$$

Transformons de même la relation (2.23), c-à-d.

$$y_i^T . s_j = 0 \qquad , i-1 \geq j \geq 0 \ :$$

* On était donc parvenu au résultat suivant : si on emploie y_{i-1}, z_{i-1} vérifiant (2.48), l'équation (2.47) donne la valeur de ΔH_{i-1}, et (2.42) donne la matrice H_i.

Maintenant, que peut-on dire d'autre à propos de y_{i-1} et de z_{i-1} ? On veut que la propriété 4. soit vérifiée, c-à-d. que l'algorithme emploie uniquement l'information présente et celle de l'itération précédente. On va donc essayer de trouver les vecteurs y_{i-1} et z_{i-1} en utilisant seulement l'information disponible à cette itération et au point le précédent immédiatement.

Tout d'abord, on note que la condition (2.32), c-à-d.

$$s_i^T . A . s_j = 0 \qquad , i-1 \geq j \geq 0 \ ,$$

à l'itération $i-1$ donne :

$$s_{i-1}^T . A . s_j = 0 \qquad , i-2 \geq j \geq 0 \qquad (A.1)$$

Reprenons les relations

$$\Delta x_i = -\alpha_i . s_i$$

et $$y_{i+1} = y_i + A . \Delta x_i \ ,$$

(A.1) devient $s_{i-1}^T . A . s_j = 0$

$$\iff \alpha_{i-1} . s_{i-1}^T . A . \alpha_j . s_j = 0$$

$$\iff \Delta x_{i-1}^T . A . \Delta x_j = 0$$

$$\iff \Delta x_{i-1}^T . (y_{j+1} - y_j) = 0$$

$$\iff \underline{\Delta x_{i-1}^T . \Delta y_j = 0} \qquad , i-2 \geq j \geq 0 \qquad (A.2)$$

Transformons de même la relation (2.23), c-à-d.

$$y_i^T . s_j = 0 \qquad , i-1 \geq j \geq 0 \ :$$

$$g_i^T \cdot \Delta_j = 0 \iff (\Delta g_{i-1}^T + g_{i-1}^T) \cdot \Delta_j = 0$$
$$\iff \Delta g_{i-1}^T \cdot \Delta_j + g_{i-1}^T \cdot \Delta_j = 0$$
$$\iff \underline{\Delta g_{i-1}^T \cdot \Delta_j = 0} \quad , i-2 \geq j \geq 0 \qquad (A.3)$$

De plus, (A.3) $\Longrightarrow \Delta g_{i-1}^T \cdot H_{i-1} \cdot A \cdot \Delta_j = 0 \quad , i-2 \geq j \geq 0 \qquad (A.4)$
(car $H_{i-1} \cdot A \cdot \Delta_j = \delta \cdot \Delta_j \quad , i-2 \geq j \geq 0$)
Cette relation (A.4) peut encore être réécrite

$$\boxed{\Delta g_{i-1}^T \cdot H_{i-1} \cdot \Delta g_j = 0} \quad , i-2 \geq j \geq 0 \qquad (A.5)$$

Comparons maintenant (A.2) et (A.5) avec les expressions de y_{i-1} et de z_{i-1} :

$$\begin{cases} y_{i-1}^T \cdot \Delta g_j = 0 \\ z_{i-1}^T \cdot \Delta g_j = 0 \end{cases} \qquad \text{pour } i-2 \geq j \geq 0$$

On voit qu'on peut choisir y_{i-1} et z_{i-1} comme

$$\boxed{\begin{aligned} y_{i-1} &= \Delta x_{i-1} \\ z_{i-1} &= H_{i-1}^T \cdot \Delta g_{j-1} \end{aligned}}$$

ou une combinaison linéaire des deux.
En général, on écrit :

$$\begin{cases} y_{i-1} = C_1 \cdot \Delta x_{i-1} + C_2 \cdot H_{i-1}^T \cdot \Delta g_{i-1} & (A.6) \\ z_{i-1} = K_1 \cdot \Delta x_{i-1} + K_2 \cdot H_{i-1}^T \cdot \Delta g_{i-1} & (A.7) \end{cases}$$

où C_1, C_2, K_1, K_2 sont des coefficients scalaires.

<u>conclusion</u> : la condition de conjugaison $\Delta_i^T \cdot A \cdot \Delta_j = 0$, $i-1 \geq j \geq 0$, est satisfaite si la matrice H est transformée suivant les équations :

$$\left|\begin{array}{l} H_i = H_{i-1} + \Delta H_{i-1} \\ y_{i-1} = c_1 . \Delta x_{i-1} + c_2 . H_{i-1}^T . \Delta g_{i-1} \\ z_{i-1} = k_1 . \Delta x_{i-1} + k_2 . H_{i-1}^T . \Delta g_{i-1} \end{array}\right.$$

* <u>la matrice initiale H</u>

Il est possible de montrer que la direction de recherche $s_i = H_i^T . g_i$ peut s'écrire sous la forme

$$\boxed{s_i = \beta_i . q_i} \qquad (A.8)$$

où β_i est un scalaire défini par

$$\beta_i = 1 - \frac{k_2 . (\Delta g_{i-1}^T . H_{i-1}^T . g_i)}{(z_{i-1}^T . \Delta g_{i-1})} \qquad (A.9)$$

et q_i est un vecteur défini par

$$q_i = \left[I - \frac{\Delta x_{i-1} . \Delta g_{i-1}^T}{\Delta x_{i-1}^T . \Delta g_{i-1}} \right] . H_{i-1}^T . g_i \qquad (A.10)$$

Examinons l'équation (A.10) : on peut voir que la direction q_i est indépendante des constantes $c_i, k_i, i=1,2$; On peut donc conclure :

<u>l'équation $s_i = \beta_i . q_i$ montre que les directions de recherche s_i générées par les différents choix de $\beta, c_1, c_2, k_1, k_2$ sont parallèles les unes aux autres si la matrice utilisée H_{i-1} au point x_{i-1} est la même.</u>

* <u>le déplacement Δx_i</u> peut être représenté par

$$\boxed{\Delta x_i = -\gamma_i . q_i} \quad , \text{ où } \gamma_i = \alpha_i . \beta_i$$

Donc $\Delta x_i = -\alpha_i . \beta_i . q_i$

Par le fait que $\alpha_i = \dfrac{g_i^T . s_i}{s_i^T . A . s_i}$ et que $s_i = \beta_i . q_i$,

la grandeur optimale du pas le long de la direction q_i est donnée par

$$\boxed{\gamma_i = \frac{g_i^T . q_i}{q_i^T . A . q_i}} \qquad (A.11)$$

En effet, $s_i = \alpha_i . \beta_i = \dfrac{g_i^T . s_i}{s_i^T . A . s_i} . \beta_i$

$$= \frac{g_i^T . q_i . \beta_i}{q_i^T . \beta_i . A . q_i . \beta_i} . \beta_i$$

$$= \frac{g_i^T . q_i}{q_i^T . A . q_i}$$

Donc γ_i est indépendant de β, ϵ_1, ϵ_2, k_1, k_2, et ainsi sera le même pour tous les algorithmes .

* Par l'équation $s_i = \beta_i . q_i$, l'équation

$$f(x_{i+1}) - f(x_i) = - \frac{(g_i^T . s_i)^2}{2 . s_i^T . A . s_i}$$

devient

$$f(x_{i+1}) - f(x_i) = - \frac{(g_i^T . q_i)^2}{2 . q_i^T . A . q_i} \qquad (A.12)$$

Donc la condition de non orthogonalité $g_i^T . s_i \neq 0$

devient : $g_i^T . q_i \neq 0$, $n - 1 \geq i \geq 0$ (A.13)

Multiplions (A.10) par g_i^T, on obtient

$$g_i^T . q_i = g_i^T . \left[I - \frac{\Delta x_{i-1} . \Delta g_{i-1}^T}{\Delta x_{i-1}^T . \Delta g_{i-1}} \right] . H_{i-1}^T . g_i$$

En tenant compte de

$$g_i^T . s_j = 0 \qquad , i-1 \geq j \geq 0 ,$$

on obtient :

$$g_i^T \cdot q_i = g_i^T \cdot I \cdot H_{i-1}^T \cdot g_i - g_i^T \cdot \frac{\Delta x_{i-1} \cdot \Delta g_{i-1}^T}{\Delta x_{i-1}^T \cdot \Delta g_{i-1}} \cdot H_{i-1}^T \cdot g_i$$

$$\boxed{g_i^T \cdot q_i = g_i^T \cdot H_{i-1}^T \cdot g_i} \qquad (A.14)$$

la matrice H_{i-1} dans l'équation (A.14) peut être exprimée par des quantités se rapportant à la $(i-2)^{ième}$ itération, en tenant compte des formules itératives de H_i, ΔH_i, g_i et z_i.

Après d'assez longs calculs, on obtient :

$$g_i^T \cdot q_i - g_i^T \cdot H_{i-1}^T \cdot g_i = g_i^T \cdot H_{i-2}^T \cdot g_i \qquad (A.15)$$

On répète le processus, et finalement

$$\begin{aligned} g_i^T \cdot q_i &= g_i^T \cdot H_{i-1}^T \cdot g_i \\ &= g_i^T \cdot H_{i-2}^T \cdot g_i \\ &= \dots \\ &= g_i^T \cdot H_0^T \cdot g_i \end{aligned} \qquad (A.16)$$

<u>La condition de non orthogonalité peut donc être satisfaite si : $g_i^T \cdot H_0^T \cdot g_i \neq 0$, $n-1 \geq i \geq 0$</u>

* Un algorithme possédant les propriétés 1., 2., 3., 4., peut être généré de la manière suivante :

(i) le choix de la matrice initiale H_0 t.q. $A = \frac{1}{2} \cdot (H_0 + H_0^T)$ est définie positive ou négative.

(ii) l'itération de la matrice est

$$H_i = H_{i-1} + \rho \cdot \frac{\Delta x_{i-1} \cdot [c_1 \cdot \Delta x_{i-1} + c_2 \cdot H_{i-1}^T \cdot \Delta g_{i-1}]^T}{[c_1 \cdot \Delta x_{i-1} + c_2 \cdot H_{i-1}^T \cdot \Delta g_{i-1}]^T \cdot \Delta g_{i-1}}$$

$$- \frac{H_{i-1} \cdot \Delta g_{i-1} \cdot [k_1 \cdot \Delta x_{i-1} + k_2 \cdot H_{i-1}^T \cdot \Delta g_{i-1}]^T}{[k_1 \cdot \Delta x_{i-1} + k_2 \cdot H_{i-1}^T \cdot \Delta g_{i-1}]^T \cdot \Delta g_{i-1}} \qquad (A.17)$$

où β, c_1, c_2, κ_1, κ_2 sont des nombres réels arbitraires, avec la seule restriction que κ_1, κ_2 ne peuvent pas s'annuler simultanément.

(iii) pour le point, on utilise les relations :

$s_i = H_i^T \cdot g_i$

$\Delta x_i = -\alpha_i \cdot s_i$

$x_{i+1} = x_i + \Delta x_i$

$\boxed{\beta = 1}$

algorithme I : D.F.P.

on pose $c_1 = 1$, $c_2 = 0$, $\kappa_1 = 0$, $\kappa_2 = 1$, et on obtient :

$$H_i = H_{i-1} + \frac{\Delta x_{i-1} \cdot \Delta x_{i-1}^T}{\Delta x_{i-1}^T \cdot \Delta g_{i-1}} - \frac{H_{i-1} \cdot \Delta g_{i-1} \cdot \Delta g_{i-1}^T \cdot H_{i-1}}{\Delta g_{i-1}^T \cdot H_{i-1} \cdot \Delta g_{i-1}}$$

propriété : si H_0 est symétrique, toutes les matrices H_i sont symétriques

---

algorithme II : Mac Cormick :

on pose $c_1 = \kappa_1 = 1$, $c_2 = \kappa_2 = 0$, et on obtient :

$$H_i = H_{i-1} + \frac{(\Delta x_{i-1} - H_{i-1} \cdot \Delta g_{i-1}) \cdot \Delta x_{i-1}^T}{\Delta x_{i-1}^T \cdot \Delta g_{i-1}}$$

---

algorithme III : Pearson :

on pose $c_1 = \kappa_1 = 0$, $c_2 = \kappa_2 = 1$, et on obtient :

$$H_i = H_{i-1} + \frac{[\Delta x_{i-1} - H_{i-1} \cdot \Delta g_{i-1}] \cdot \Delta g_{i-1}^T \cdot H_{i-1}}{\Delta g_{i-1}^T \cdot H_{i-1} \cdot \Delta g_{i-1}}$$

---

algorithme IV :

on pose $K_1 = 0$, $K_2 = 1$, et on obtient

$$H_i = H_{i-1} - \frac{H_{i-1} \cdot \Delta g_{i-1} \cdot \Delta g_{i-1}^T \cdot H_{i-1}}{\Delta g_{i-1}^T \cdot H_{i-1} \cdot \Delta g_{i-1}}$$

propriété : si H_0 est symmétrique, toutes les matrices H_i sont symmétriques

---

$\boxed{\beta = 0}$

algorithme V :

on pose $K_1 = 0$, $K_2 = 1$, et on obtient :

$$H_i = H_{i-1} - \frac{H_{i-1} \cdot \Delta g_{i-1} \cdot \Delta g_{i-1}^T \cdot H_{i-1}}{\Delta g_{i-1}^T \cdot H_{i-1} \cdot \Delta g_{i-1}}$$

propriété : si H_0 est symmétrique, tous les H_i seront symmétriques

---

algorithme VI :

on pose $K_1 = 0$, $K_2 = 1$, et on obtient :

$$H_i = H_{i-1} - \frac{H_{i-1} \cdot \Delta g_{i-1} \cdot [\Delta x_{i-1} - H_{i-1}^T \cdot \Delta g_{i-1}]^T}{[\Delta x_{i-1} - H_{i-1}^T \cdot \Delta g_{i-1}]^T \cdot \Delta g_{i-1}}$$

$\boxed{\beta = -1}$

On remplace de la même manière β par -1 dans l'équation générale (A.17)

---

démonstration de la proposition 4.2

Par l'absurde. Si le contraire était vrai, il existerait une sous-suite $\{\psi_n\}$ de $\{\varphi_n\}$ t.q. $\|A.\psi_n\| > \varepsilon$ pour tout n suffisamment grand.
Etant donné que A est complétement continue, une sous suite $\{x_n\}$ peut être extraite de $\{\psi_n\}$, t.q.

$A.x_n$ converge vers f.

Cet élément f n'est pas l'élément nul, parce que $\|Ax_n\| > \varepsilon$ pour n suffisamment grand.
Par la continuité du produit intérieur, on trouve

$$\langle Ax_n, f\rangle \longrightarrow \|f\|^2 \neq 0 .$$

D'autre part,

$$\langle A.x_n, f\rangle = \langle x_n, A^* f\rangle$$

qui, par le lemme de Riemann Lebesgue, approche 0.
On arrive ainsi à une contradiction.
D'où $A.\varphi_n \longrightarrow 0$ ■

démonstration de la proposition 4.1

Soit $\{\varphi_n\}$ une suite orthonormale infinie dans H, et soit $x_n = A.\varphi_n$. Etant donné que A est inversible, $\varphi_n = A^{-1}.x_n$.
Maintenant, $\|\varphi_n\| = 1$ et $\|x_n\| \longrightarrow 0$ par la proposition 4.2, de telle manière que A^{-1} est évidemment non borné. ■

distance minimale d'un point à une droite

Soit X un espace réel de Hilbert. Une droite L dans X est un ensemble de la forme

$$L = \{ x_0 + \alpha . u \ , \ -\infty < \alpha < \infty \} \quad \text{où } x_0 \in X, \ u \in X, \ u \neq 0.$$

Soit $x \in X$, $x \notin L$.

Pour chaque α, la distance de x à $x_0 + \alpha . u$ est donnée par $\| x - x_0 - \alpha . u \|$. Il est donc évident que la distance minimale correspond à la valeur minimale de $\| x - x_0 - \alpha . u \|^2$.

Définissons $f(\alpha) = \| x - x_0 - \alpha . u \|^2$. La norme étant définie à partir du produit intérieur, on a

$$f(\alpha) = \langle x - x_0 - \alpha u , x - x_0 - \alpha u \rangle$$

$$= \alpha^2 . \langle u, u \rangle - 2 . \alpha \langle x - x_0 , u \rangle + \langle x - x_0 , x - x_0 \rangle$$

qui est une fonction quadratique.

En outre, $f'(\alpha) = 2 . \alpha . \langle u, u \rangle - 2 . \langle x - x_0 , u \rangle$,

et $f''(\alpha) = 2 . \langle u, u \rangle$, qui est toujours positif.

Il suit que $f(\alpha)$, et par conséquent aussi $\| x - x_0 - \alpha . u \|$ a une unique valeur minimale au point de L pour lequel $f'(\alpha) = 0$. En ce point,

$$\alpha . \langle u, u \rangle - \langle x - x_0 , u \rangle = 0$$

de telle sorte que

$$\langle x_0 + \alpha . u - x , u \rangle = 0 .$$

Soit $z = x_0 + \alpha . u$. Alors la distance minimale de x à L se produit au point $z \in L$ t.q. $z - x \perp L$.

définition :

> soit $A : X \longrightarrow X$ un opérateur linéaire continu sur l'espace de Hilbert X. Un sous-espace N de X est un sous-espace invariant de A si
> $x \in N \Longrightarrow Ax \in N$

Si N est un sous espace invariant de A, comme A est continu, il est facile de voir que la fermeture $\overline{N}$ de N est aussi un sous-espace invariant de A. En effet, si $x \in \overline{N}$, alors $x = \lim_{n \to \infty} x_n$ pour une suite $\{x_n\}$ dans N. Par continuité, $Ax = \lim_{n \to \infty} A.x_n$. Comme $A.x_n \in N$ pour chaque n, $Ax \in \overline{N}$.

Soit maintenant N un sous-espace fermé invariant de A ; considérons A^* l'opérateur adjoint de A :

$$\langle Ax, y \rangle = \langle x, A^* y \rangle$$

Soit $N^\perp$ le complément orthogonal de N ; donc $x = N \oplus N^\perp$

Considérons $y \in N^\perp$, $x \in N$.

Alors $\langle x, A^* y \rangle = \langle Ax, y \rangle$.

Mais $Ax \in N$ et $y \in N^\perp$, donc

$$\langle Ax, y \rangle = 0.$$

Cela veut dire que $A^* y \in N^\perp$, et donc que si N est un sous-espace (fermé) invariant de A, alors $N^\perp$ est un sous-espace invariant de A^*.

BIBLIOGRAPHIE

A.1 ANDERSON, B.D.O., and MOORE, J.B., "Linear optimal control", Prentice-Hall, inc., Englewood Cliffs, N.J., 1971.

B.1 BEUTLER, F.J., and ROOT, W.L., "the operator pseudo-inverse in control and system identification", generalized inverses and applications, (ed. by M.Z. Nashed), Academic press, New-York, 1976, pp. 397-494

B.2 BROYDEN, C.G., "quasi-Newton methods and their application to function minimization", mathematics of computation, vol. 21, 1967, pp. 368-381

B.3 BEN-ISRAEL, A., and COHEN, D., "on iterative computation of generalized inverses and associated projections", siam J. Numer. anal., vol. 3, no 3, 1966, pp. 410-419

C.1 CHENG, B.D., and POWERS, W.F., "convergence of gradient-type methods on singular parameter optimization problems", A.I.A.A. Journal, vol. 15, no 6 (U. of Michigan).

C.2 CHENG, B.D., and POWERS, W.F., "singular optimal control computation", Journal of guidance and control", vol. 1, no 1,

january - february 1978, pp. 83-89.

C.3 CHENG, B., "convergence of quasi-Newton algorithms with applications in singular optimal control", U. of Michigan, 1977.

C.4 CHENG, B.D., and POWERS, W.F., "rate of convergence of function space quasi-Newton algorithms: the non singular case", U. of Michigan, Dept. of Aerospace Eng., report no 013834-T-2, oct. 1976

C.5 CHENG, B.D., and POWERS, W.F., "convergence of finite-dimensional conjugate direction and quasi-Newton methods for singular problems", AIAA paper no. 76-791, presented at A.I.A.A. astrodynamics conference, august 1976.

D.1 DANIEL, J.W., "the conjugate gradient method for linear and non linear operator equations", SIAM J. Numer. anal., vol. 4, no 1, 1967, pp. 10-26

E.1 EDGE, E.R., and POWERS, W.F., "function space quasi-Newton algorithms for optimal control problems with bounded controls and singular arcs", to appear in JOTA, Dec. 1976

E.2 EDGE, E.R., and POWERS, W.F., "shuttle ascent trajectory optimization and function space quasi-Newton techniques", AIAA Journal, Vol. 14, NO. 10, 1976, pp. 1369-1376

E.3 EDGE, E.R., "function space quasi-Newton techniques with application to space shuttle trajectory optimization", doctoral thesis, the university of Michigan, aerospace Engineering department, 1977

F.1 FLETCHER, R., and REEVES, C.M., "function minimization by conjugate gradients", Computer J., vol. 7, 1964, pp. 149-154

F.2 FLETCHER, R., and POWELL, M.J.D., "a rapidly convergent descent method for minimization", Computer J., vol. 6, 1963, pp. 163-168

G.2 GOH, B.S., "the second variation for the singular Bolza problem", SIAM Journal on control, vol. 4, no 2, 1966, pp. 309-325

G.1 GROETSCH, C.W., "generalized inverses of linear operators", representation and approximation, Dekker NY, 1977

H.1 HUANG, H.Y., "unified approach to quadratically convergent algorithms for function minimization", JOTA, vol. 5, no 6, 1970, pp. 405-423

H.2 HESTENES, M.R., "the conjugate gradient method for solving linear systems", proc. symposia on applied mathematics, vol. 6,

Numerical analysis, Mc. Graw-Hill, New. York, 1956, pp. 83-102

H.3 HESTENES, M.R., "iterative method for solving linear equations", JOTA, vol. 11, no 4, 1973, pp. 323-334

H.4 HORWITZ, L.B., and SARACHIK, P.E., "Davidon's method in Hilbert space", SIAM J. appl. math., vol. 6, no 4, 1968, pp. 676-695

J.1 JONES, D.S., "the variable metric algorithm for non definite quadratic function", J. inst. maths applics, vol. 12, 1973, pp. 63-71.

J.2 JOHANSON, D.E., "convergence properties of the method of gradients", advances in control systems, vol. 4, (ed. by C.T. Leondes), academic press, New-York, 1966, pp. 279-316

K.1 KELLER, H.B., "on the solution of singular and semidefinite linear systems by iteration", SIAM J. numer. anal. vol. 2, no 2, 1965, pp. 281-290

K.2 KAMMERER, W.J., and NASHED, M.Z., "steepest descent for singular linear operators with non closed range", Applicable analysis, vol. 1, 1971, pp. 143-159

K.3 KAMMERER, W.J., and NASHED, M.Z., "on the convergence of the conjugate gradient method for singular linear operator equations", SIAM J. numer. anal., vol. 9, no 1, 1972, pp. 165-181

K.4 KAMMERER, W.J., and PLEMMONS, R.J., "direct iterative methods for least-squares solution to singular operator equation", journal of mathematical analysis and applications, vol. 49, 1975, pp. 512-526

K.5 KRASNER, N., and KAILATH, T., "a stochastic interpretation of singular quadratic minimization theory - part 1: general conditions for minimality".

M.1 MYERS, G.E., "properties of the conjugate-gradient and Davidon methods", JOTA, vol. 2, no 4, 1968, pp. 209-219

M.2 MC DANELL, J.P., and POWERS, W.F., "New-Jacobi type necessary and sufficient conditions for singular optimization problems", AIAA J., vol. 8, no 8, 1970, pp 1416-1420

M.3 MC DANELL, J.P., and POWERS, W.F., "necessary conditions for joining optimal singular and non singular subarcs", SIAM J. control, vol. 9, no 2, 1971, pp. 161-173

M.4 MC CORMICK, G.P., and RITTER, K., "methods of conjugate directions versus quasi-Newton methods", mathematical programming, vol. 3, 1972, pp. 101-116

N.1 NASHED, M.Z., "steepest descent for singular linear operator equations", SIAM J. numer. anal., vol. 7, no 3, 1970, pp. 358-368

N.2 NASHED, M.Z., generalized inverses and aplications, academic press, New-York, 1976.

N.3 NOBLE, B., "a method for computing the generalized inverse of a matrix", SIAM J. numer. anal., vol. 3, no 4, 1966, pp. 582-584

P.1 POWELL, M.J.D., "an efficient method for finding the minimum of a function of several variables without calculating derivatives", computer J., vol. 7, no 2, 1964, pp. 155-162

P.2 POLAK, E., computational methods in optimization, academic press, New York, 1971

P.3 PONTRYAGIN, L.S., BOLTYANSKII, V.G., GAMKRELIDZE, R.V., and MISHCHENKO, E.F., the mathematical theory of optimal processes, interscience, New-York, 1962

P.4 POWELL, M.J.D., "on the convergence of the variable metric algorithm", report no. T.P. 382, theoritical physics division, A.E.R.E. Harwell, 1969.

P.5 POWERS, W.F., and MCDANELL, J.P., "switching conditions and a synthesis technique for the singular saturn guidance problem", J. Spacecraft and Rockets, vol. 8, no 10, 1971, pp 1027-1032

R.1 RIESZ, F., and NAGY, B., "functional analysis", F. Ungar publishing CO., New-York, 1955

S.1 SILBER, R., "a characterization of Davidon's method", U. of Mich., Dept. aero. eng. report 071482-T-1, Feb. 1973

S.2 STAKGOLD, I., boundary value problems of mathematical physics, vol. 1, Macmillan CO., New-York, 1967, pp. 182-188.

T.1 TAYLOR, E., "introduction to functional analysis".

T.2 TOKUMARU, H., ADACHI, N., and GOTO, K., "Davidon's method for minimization problem in Hilbert space with an application to control problems", SIAM J. control, vol. 8, no 2, 1970, pp. 163-178

T.3 TURNER, P.R., and HUNTLEY, E., "The variable metric method in Hilbert space with applications to control problems", JOTA, vol. 19, no 3, July 1976, pp. 381 - 400