MVA, Projet optimisation : Inégalité linéaires

Valentin DE BORTOLI, Louis THIRY

12 janvier 2017

Table des matières

1 Présentation du problème

On s'intéresse aux problèmes d'égalité et d'inégalité aux moindres carrés.

Définition 1. *Soit $A \in \mathcal{M}_{n,m}(\mathbb{R})$ avec $(n,m) \in \mathbb{N}^2$ et $b \in \mathbb{R}^n$. On appelle problème d'égalité aux moindres carrés et on note $\mathcal{P}_=(A,b)$ le probleme de minimisation :*

$$\mathcal{P}_=(A,b): \quad \hat{x} = \underset{x}{argmin}\, ||Ax-b||^2$$

Résoudre le problème d'égalité aux moindres carrés permet de trouver la meilleure solution possible au systeme linéaire $Ax = b$:

— lorsqu'il existe une solution à ce système linéaire, les deux problèmes sont équivalents
— lorsque l'on a plus d'équations que d'inconnues ($n > m$), le système peut être inconsistant et n'avoir aucune solution. Dans ce cas, une solution x_0 du problème d'égalité aux moindres carrés est la meilleure approximations telle que $Ax_0 = b$ au sens des moindres carrés. C'est une solution approximative au système linéaire $Ax = b$

Définition 2. *On appelle problème d'inégalité aux moindres carrés et on note $\mathscr{P}_{\leq}(A,b)$ le probleme de minimisation :*

$$\mathscr{P}_{\leq}(A,b): \quad \hat{x} = \underset{x}{argmin}\, ||(Ax-b)_+||^2$$

De même que pour le cas d'égalité, résoudre le problème d'inégalité aux moindres carrés permet de trouver la meilleure solution possible au systeme d'inégalité linéaires $Ax \leq b$:

— lorsqu'il existe une solution à ce système d'inégalité, les deux problèmes sont équivalents

— Dans le cas où ils n'y a pas de solutions aux système d'inégalité, c'est une solution approximative.

Le problème d'égalité est bien compris : on connait la forme des solutions. Ce problème est traité dans la première partie du rapport.

Pour le problème d'inégalité, on a simplement un résultat d'existence.

Le but de ce rapport est de proposer un algorithme pour trouver une solution au problème d'intégalité aux moindres carrés $\mathscr{P}_{\leq}(A,b)$.

2 Egalité aux moindres carrés

2.1 Rappels sur la pseudo-inverse

On introduit la notion de matrice pseudo inverse de Moore-Penrose [3] qui joue un rôle particulier dans l'étude des problèmes aux moindres carrés.

Théorème 1. *Soit $A \in \mathscr{M}_{n,m}(\mathbb{R})$ avec $(n,m) \in \mathbb{N}^2$. Il existe (M,N) deux matrices orthogonales de $\mathscr{M}_m(\mathbb{R})$, respectivement $\mathscr{M}_n(\mathbb{R})$ et $\Sigma = \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix}$ avec D diagonale strictement positive telles que $A = N\Sigma M$.*

Démonstration : On considère la matrice $A^TA \in \mathscr{M}_m(\mathbb{R})$. Celle-ci est symétrique, semi-définie positive et on peut appliquer le théorème spectral. On a donc une matrice orthogonales P qui vérifient $P^TA^TAP = \begin{matrix} D^2 & 0 \\ 0 & 0 \end{matrix}$ avec D^2 diagonale strictement positive. Il est à noter que D est une matrice de $\mathscr{M}_r(\mathbb{R})$ où r est le rang de la matrice A (et donc plus petit que m et n). On note $P = [P_1\ P_2]$ où $P_1 \in \mathscr{M}_{m,r}(\mathbb{R})$ et $P_2 \in \mathscr{M}_{m,m-r}(\mathbb{R})$. On a alors :

$$\begin{cases} P_1^TA^TAP_1 = D^2 \\ P_1^TA^TAP_2 = 0 \\ P_2^TA^TAP_1 = 0 \\ P_2^TA^TAP_2 = 0 \end{cases} \tag{1}$$

Posons $Q_1 = AP_1D^{-1}$. On a bien $Q_1^TQ_1 = D^{-1}P_1^TA^TAP_1D^{-1} = Id_r$. On complète Q_1 en une matrice orthogonale de $\mathscr{M}_n(\mathbb{R})$ et on a $Q^TAP = \begin{pmatrix} DP_1^TA^TAP_1 & DP_1^TA^TAP_2 \\ Q_2^TAP_1 & Q_2^TA^TAP_2 \end{pmatrix}$. Or $P_1^TA^TAP_2 = 0$ et $Q_2^TQ_1 = DQ_2^TAP_1 = 0$ et puisque $P_2^TA^TAP_2 = 0$, $AP_2 = 0$. Donc on trouve bien $Q^TAP = \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix}$.
On peut alors définir la pseudo-inverse.

Définition 3 (Pseudo-inverse de Moore-Penrose, 1920)**.** *On appelle pseudo-inverse de Moore-Penrose de A et on note A^+ la matrice $A^+ = M^T\begin{pmatrix} D^{-1} & 0 \\ 0 & 0 \end{pmatrix}N^T \in \mathscr{M}_{m,n}(\mathbb{R})$ (où M,N sont les matrices orthogonales qui apparaissent dans la décomposition en valeurs singulières).*

Proposition 1. *A^+ vérifie les propriétés suivantes :*

— $A^+AA^+ = A^+$

— $AA^+A = A$

— AA^+ *est symétrique*

— A^+A *est symétrique*

On peut alors montrer la propriété suivante :

Proposition 2. *Il existe une unique matrice qui vérifie ces quatre propriétés, la pseudo-inverse de Moore-Penrose.*

Démonstration : Soient (B,C) deux matrices satisfaisant les propriétés énoncées. On a alors :

$$\begin{aligned} AB &= ACAB \\ &= (AC)^T (AB)^T \\ &= (ABAC)^T \\ &= ABAC \\ &= AC \end{aligned} \tag{2}$$

De la même manière, $BA = CA$. Ensuite on a :

$$\begin{aligned} C &= CAC \\ &= CAB \\ &= BAB \\ &= B \end{aligned} \tag{3}$$

On établit maintenant une proposition concernant des décompositions orthogonales de l'espace.

Proposition 3. *On a les égalité suivantes :*

- $ker A^T \overset{\perp}{\oplus} Im A = \mathbb{R}^m$
- $ker A \overset{\perp}{\oplus} Im A^T = \mathbb{R}^n$
- $ker A^+ \overset{\perp}{\oplus} Im A = \mathbb{R}^m$ *et* AA^+ *projecteur orthogonal sur* $Im A$ *de noyau* $ker A^+$
- $ker A \overset{\perp}{\oplus} Im A^+ = \mathbb{R}^n$ *et* A^+A *projecteur orthogonal sur* $Im A^+$ *de noyau* $ker A$

Démonstration : concernant les deux premiers points, on vérifie rapidement l'orthogonalité de ces deux espaces. Puis on peut conclure sur le fait que leur somme directe correspond bien à tout l'espace en utilisant un argument de dimension (notamment $\text{rg}A = \text{rg}A^T$).
Pour les deux autres points on ne démontre que le premier (le second s'effectue exactement de la même manière). $(AA^+)^2 = AA^+$ donc c'est un projecteur et la condition de symétrie assure son orthogonalité. $\ker A^+$ est trivialement inclus dans son noyau. Soit un élément x tel que $AA^+x = 0$, alors $A^+AA^+x = A^+x = 0$ donc c'est un élément du noyau de A^+. On en déduit l'égalité des noyaux. De plus $\text{Im}A \subset \text{Im}AA^+$. Soit $y = Ax$. Alors $y = AA^+Ax$ donc $y \in \text{Im}A$. On a donc l'égalité des images. On peut donc conclure.

Proposition 4. *Une conséquence de cette proposition est* $ker A = ker A^+$ *et* $Im A = Im A^+$.

Remarque : cette identification des supplémentaires est vraie seulement parce que les sommes directes sont orthogonales.

2.2 Solutions au problème d'égalité aux moindres carrés

Revenons au problème des moindres carrés.

Proposition 5. *Un élément* x *est solution du problème* $\mathscr{P}_=(A,b)$ *si et seulement si* $A^T(Ax-b) = 0$.

Démonstration : Le problème d'égalité aux moindre carrés consiste en la minimisation de la fonction $F : x \mapsto ||Ax-b||^2$. F étant convexe, ses minimas sont ses points stationnaires, c'est à dire qui vérifient $\nabla F = 0$, i.e. $A^T(Ax-b) = 0$.

Proposition 6. *L'ensemble des solutions du problème* $\mathscr{P}_=(A,b)$ *est l'ensemble* $A^+b + ker A$.

Démonstration : Une solution x du problème vérifie $A^T(Ax-b) = 0$, i.e. le système linéraire $A^TAx = A^Tb$. L'ensemble des solutions de de système linéaire est $x_0 + \ker A^TA$ ou x_0 est une solution particulière ce ce système.

— $A^+ b$ est une solution de ce système :

$$A = N\begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix} M$$

$$A^+ = M^T\begin{pmatrix} D^{-1} & 0 \\ 0 & 0 \end{pmatrix} N^T.$$

$$A^T A A^+ b = M^T\begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix} N^T b$$

$$A^T b = M^T\begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix} N^T b$$

$$\Longrightarrow A^T(AA^+ b - b) = 0$$

— $\ker A = \ker A^T A$.

$$\ker A \subset \ker A^T A$$

$$\ker A^T A \subset \ker A : \quad x \in \quad \ker A^T A \Longrightarrow A^T Ax = 0 \Longrightarrow x^T A^T Ax = 0 = ||Ax||^2 \Longrightarrow Ax = 0 \Longrightarrow x \in \ker A$$

Donc l'ensemble des solutions du problème $\mathscr{P}_=(A,b)$ est $A^+ b + \ker A$.

Proposition 7. *$A^+ b$ est la solution de $\mathscr{P}_=(A,b)$ de norme minimale.*

Démonstration : Toute solutions x_0 de $\mathscr{P}_=(A,b)$ s'écrit $x_0 = A^+ b + y$, $y \in \ker A$. Comme $\ker A \perp \mathrm{Im} A^+$, par le théoreme de pythagore $\|x_0\|^2 = \|A^+ b\|^2 + \|y\|^2$ donc $|x_0| \leq |A^+ b|$.

Maintenant que nous avons traité le problème d'égalité aux moindres carrés, intéressons nous au problème d'inégalité aux moindres carrés.

3 Inégalité aux moindres carrés

3.1 Existence de solutions

Théorème 2. *Quelque soit la matrice $A \in \mathscr{M}_{n\times m}(\mathbb{R})$ et le vecteur $b \in \mathbb{R}^n$, le problème d'inégalité aux moindres carrés $\mathscr{P}_\leq(A,b)$ admet une solution x_0. Quelque soit la solution x_0, le vecteur résidu $z = (Ax_0 - b)_+$ est unique est x est une solution du problème ssi $(Ax-b)_+ = z$.*

Démonstration : Pour l'existence, trois preuves différentes sont proposées dans l'article [1]. La deuxième partie du théorème est évidente : s'il existe deux solutions x_0 et x_1 telles que $(Ax_1 - b)_+ \neq (Ax_0 - b)_+$ alors soit $(Ax_1 - b)_+ < (Ax_0 - b)_+$ ce qui contredit le fait que x_0 soit une solution, soit $(Ax_1 - b)_+ > (Ax_0 - b)_+$ ce qui contredit le fait que x_1 soit une solution. De même tout vecteur x de résidu minimal est solution.

Cependant, l'ensemble des solutions de $\mathscr{P}_\leq(A,b)$ n'est pas connu. C'est pourquoi nous commençons par nous restreindre à des cas plus simples.

3.2 Consistance de systèmes d'inégalité

Définition 4 (Consistance)**.** *Soient $A \in \mathscr{M}_{n\times m}(\mathbb{R})$ et $b \in \mathbb{R}^n$ On dit que le système d'inéquations linéaires (A,b) est consistant si $\exists x \in \mathbb{R}^m$ tel que $Ax \leq b$.*

Remarques :
— Dans le cas consitant, la résolutions de $Ax \leq b$ est équivalent à la résolution de $\mathscr{P}_\leq(A,b)$
— Trouver une solution au système $Ax \leq b$ n'est pas un problème facile bien que son expression soit très simple.

Définition 5 (Forte consistance)**.** *On dit qu'un système d'inéquations linéaires (A,b) est fortement consistant si $\exists x \in \mathbb{R}^m$ tel que $Ax < b$.*

Proposition 8. *Soit (A,b) un système fortement consistant alors (A,b) admet une infinité de solutions.*

Démonstration : On considère $(1+\epsilon)x$ avec ϵ assez petit pour que les inégalité linéaires soient toujours vérifiées (vrai par continuité). Puisque l'ensemble des solutions est convexe et qu'on a deux solutions distinctes, celui-ci est infini.

3.3 Consitance des systèmes d'inégalité homogènes

Les systemes d'inégalité homogènes sont tels que $b = 0$.

Nou allons ennoncer des théormes d'alternatives qui nous permettrons d'établir une condition nécessaire est suffisante à la consistance d'un système d'inégalité tel que $b = 0$.

Un théormème d'alternative consiste en un ensemble de deux assertions P et Q telles que l'une est fausse quand l'autre est vraie :

$$P \implies \text{not } Q \text{ and not } P \implies Q$$

3.3.1 Théorème de Gordan

Théorème 3 (Gordan, 1873)**.** *On note* $(a_i)_{i\in[\![1,m]\!]} \in (\mathbb{R}^m)^n$ *les lignes de* A $(\mathscr{P})$ $\exists x \in \mathbb{R}^m,\ \forall i \in [\![1,n]\!],\ \langle a_i, x\rangle < 0$ *est équivalent à la négation de la proposition suivante :* $(\mathscr{Q})$ $\exists (\lambda_i)_{i\in[\![1,n]\!]} \in {\mathbb{R}_+}^n \setminus \{0\},\ \sum_{i=1}^{n} \lambda_i a_i = 0$

Remarques :
— On peut donner une interprétation géométrique de ce théorème : aucun des a_i ne doit être dans le cône engendré par les vecteurs $(-a_j)_{j\in[\![1,n]\!]\setminus\{i\}}$. Cette remarque géométrique peut être comprise en termes de condition pour que les demi-hyperplans portés par les a_i s'intersectent (faire dessin).
— Ce théorème se rapproche d'une condition nécessaire et suffisante pour que le système d'inégalité $Ax \leq 0$ admette une solution consistante. Malheureusement les inégalité sont strictes.

Deux démonstrations du théorème de Gordan On va citer ici deux démonstrations du théorème de Gordan. La première est totalement géométrique, la seconde utilise à profit l'existence de solutions au sens des moindres carrés. Il est à noter que l'on ne considère que l'implication $(\neg\mathscr{Q} \Rightarrow \mathscr{P})$ que l'on démontre par contraposée. L'autre implication $(\mathscr{Q} \Rightarrow \neg\mathscr{P})$ est triviale.

Démonstration 1 : On considère Π_0 la projection de 0 sur le polyèdre convexe fermé formé par la famille $(-a_i)_{i\in[\![1,n]\!]}$, noté Δ. On écrit alors la caractérisation de cette projection :

$$\forall x \in \Delta \langle x - \Pi_0, -\Pi_0\rangle \leq 0 \tag{4}$$

Appliqué en $(-a_i)_{i\in[\![1,n]\!]}$ on trouve que $\forall i \in [\![1,n]\!],\ \langle a_i, \Pi_0\rangle + \|\Pi_0\|^2 \leq 0$. Or on a $\neg\mathscr{P}$ donc Π_0 ne peut être de norme différente de 0 (sinon on a une solution stricte aux inégalité). Ainsi $\Pi_0 = 0$. Donc 0 est combinaison convexe des $(-a_i)_{i\in[\![1,n]\!]}$ donc des $(a_i)_{i\in[\![1,n]\!]}$ également et on en déduit le théorème de Gordan.

Démonstration 2 : on considère le problème $\langle a_i, x\rangle \leq -\epsilon$ avec $\epsilon \in \mathbb{R}_+^*$ et ce $\forall i \in [\![1,n]\!]$. Le problème n'admet pas de solution (usage de $\neg\mathscr{P}$) mais en admet une au sens des moindres carrés. Elle vérifie alors $A^T(Ax-\epsilon)_+ = 0$. Mais on peut réécrire cela $\sum_{i=1}^{n} a_i((Ax-\epsilon)_+)_i = 0$. Enfin, on remarque que $\exists i \in [\![1,n]\!],\ ((Ax-\epsilon)_+)_i > 0$ sinon on a une solution consistante. En posant $\lambda_i = ((Ax-\epsilon)_+)_i$ on a donc montré $\mathscr{Q}$.

3.3.2 Lemme de Farkas, théorème de Motzkin et de Stiemke.

Le lemme de Farkas permet de passer aux inégalité larges.

Lemme 1 (Farkas, 1902)**.** $(\mathscr{P})$ $\exists x \in \mathbb{R}^m,\ \forall i \in [\![1,n]\!],\ \langle a_i, x\rangle \leq 0$ *et* $\langle a_1, x\rangle < 0$ *est équivalent à la négation de la proposition suivante :* $(\mathscr{Q})$ $\exists (\lambda_i)_{i\in[\![1,n]\!]} \in {\mathbb{R}_+}^n \setminus \{\lambda_1 = 0\},\ \sum_{i=1}^{n} \lambda_i a_i = 0$

Démonstration : Encore une fois $\mathscr{Q} \Rightarrow \neg\mathscr{P}$ est évidente. L'implication réciproque $\neg\mathscr{P} \Rightarrow \mathscr{Q}$ est plus délicate. On va raisonner par récurrence sur la dimension de l'espace. L'initialisation est triviale. On constate que si $\neg\mathscr{P}$ alors on est dans les conditions du théorème de Gordan et $\exists(\lambda_i)_{i\in[\![1,n]\!]} \in \mathbb{R}_+^n \setminus \{0\}$ telle que $\sum_{i=1}^{n} \lambda_i a_i = 0$. Si $\lambda_1 > 0$ alors on a terminé. Plaçons-nous désormais dans le cadre où $\lambda_1 = 0$. On a donc $\sum_{i=1}^{n} \lambda_i a_i = 0$ et au moins un des λ_i (avec $i \in [\![2,n]\!]$) est non nul, on le note λ_j. Plaçons-nous sur $a_j^\perp$ qui est de dimension strictement inférieure à celle de l'espace ambiant (on a supposé tous les $a_i \neq 0$, l'extension au cas où l'un d'entre eux ou plusieurs sont nulles est immédiate). On peut appliquer le lemme de Farkas sur $a_j^\perp$ et on a $(\alpha_i)_{i\in[\![1,n]\!]\setminus\{j\}} \in \mathbb{R}_+^{n-1}$ avec $\alpha_1 > 0$ tel que $\sum_{i=1, i\neq j}^{n} \alpha_i a_i = \mu a_j$ avec $\mu \in \mathbb{R}$ (en effet il a fallu projeter les a_i dans $a_j^\perp$ pour pouvoir appliquer le lemme, c'est-à-dire, il a fallu appliquer le lemme aux $a_i - \langle a_i, a_j\rangle a_j$). Si μ est négatif alors on a terminé. Si μ est positif alors on remplace a_j par $\frac{-1}{\lambda_j} \sum_{i=1, i\neq j}^{n} \lambda_i a_i$ et on conclut.

On peut maintenant citer le théorème de Motzkin, qui bien que semblant plus fort que le lemme de Farkas en est une conséquence directe [1].

Théorème 4 (Motzkin, 1936)**.** *($\mathscr{P}$) $\exists x \in \mathbb{R}^m,\ \forall i \in [\![1,n]\!],\ \langle a_i, x\rangle \leq 0$ et $\forall i \in [\![1,p]\!],\ \langle a_i, x\rangle < 0$ est équivalent à la négation de la proposition suivante : ($\mathscr{Q}$) $\exists(\lambda_i)_{i\in[\![1,n]\!]} \in {\mathbb{R}_+}^n \setminus \{(\lambda_1,\dots,\lambda_p) = 0\},\ \sum_{i=1}^{n} \lambda_i a_i = 0$*

Remarque : Ce théormème nous fournit une CNS pour qu'un systeme $Ax \leq b$ soit consistant.

Démonstration : Encore une fois $\mathscr{Q} \Rightarrow \neg\mathscr{P}$ est triviale. On considère seulement le sens $\neq \mathscr{P} \Rightarrow \mathscr{Q}$. On considère $C_i = \{x,\ \langle a_i, x\rangle \leq 0,\ \langle a_j, x\rangle \leq 0\}$. Si tous les C_i sont non vides pour $i \in [\![1,p]\!]$ alors on peut considérer un élément $x = \sum_{i=1}^{p} x_i$ avec $x_i \in C_i$. Cet élément est alors une solution de notre système d'inéquations linéaires et cela rentre en contradiction avec $\neg\mathscr{P}$. Ainsi, au moins un des C_i est vide. On peut alors appliquer le lemme de Farkas pour ce problème et donc on obtient une famille $(\lambda_i)_{i\in[\![1,n]\!]}$ avec λ_i non nul et donc $(\lambda_1,\dots,\lambda_p)$ non nul.

Le cas des inégalité larges est donc bien compris. Il s'agit maintenant d'entériner le cas où $b \neq 0$.

3.4 Consitance des systèmes d'inégalité dans le cas général $b \neq 0$

Pour cela on va se servir des équivalences suivantes :

$$\exists x \in \mathbb{R}^m,\ Ax \leq b \ \Leftrightarrow\ \exists x \in \mathbb{R}^{m+1} (A|b)x \leq 0, \text{ et } x_{m+1} = -1 \ \Leftrightarrow \exists x \in \mathbb{R}^{m+1} (A|b)x \leq 0, \text{ et } x_{m+1} < 0 \tag{5}$$

où $(A|b)$ est la matrice complétée de A par le vecteur b, c'est-à-dire : $(A|b) = [A\ b]$. La dernière équivalence est obtenue en utilisant la linéarité des équations.

Proposition 9. *Le système $Ax \leq b$ n'admet pas de solutions si et seulement si e_{m+1} est un élément du cône engendré par $([-a_i - b_i])_{i\in[\![1,n]\!]}$. On remarque que $-(A|b)^T$ est la matrice dont les colonnes sont les $[-a_i - b_i]$.*

Démonstration : c'est une simple conséquence du lemme de Farkas.
Désormais on note $\tilde{A} = -(A|b)^T$.

Proposition 10. *Soit (A,b) un système d'inéquations linéaires consistant et si $n \leq m$. Alors il admet une infinité de solutions.*

Démonstration : Si le système est fortement consistant alors on se sert de la proposition démontrée plus haut. Sinon, puisque le système est consistant sans être fortement consistant on peut utiliser le lemme de Farkas puis le théorème de Gordan pour écrire :

$$\begin{cases} \neg\left(\exists(\lambda_i)_{i\in[\![1,n+1]\!]} \in \mathbb{R}_+^{n+1} \setminus \{\lambda_{n+1} \neq 0\},\ \sum_{i=1}^{n} \lambda_i [a_i b_i] - \lambda_{n+1} e_{n+1} = 0\right) \\ \exists(\lambda_i)_{i\in[\![1,n+1]\!]} \in \mathbb{R}_+^{n+1} \setminus \{0\},\ \sum_{i=1}^{n} \lambda_i [a_i b_i] - \lambda_{n+1} e_{n+1} = 0 \end{cases} \tag{6}$$

1. Tout comme le théorème de Brouwer paraît plus simple que le théorème de Schauder !

Donc on peut rassembler ces deux informations pour écrire :

$$\exists(\lambda_i)_{i\in[\![1,n]\!]}, \in \mathbb{R}_+^n \setminus \{0\}, \sum_{i=1}^{n} \lambda_i [a_i b_i] = 0 \tag{7}$$

Cela donne a fortiori $\text{rg}(A^T) < n$. Or $\text{rg}(A) = \text{rg}(A^T)$. Donc puisque $m = \dim(\text{Ker}A) + \text{rg}A$, $\ker A \neq \{0\}$. Donc on peut ajouter n'importe quel élément du noyau et on a toujours une solution de notre système d'inéquations linéaires. On a donc une infinité de solutions

Remarque : le cas $n < m$ est trivial. Seul le cas $n = m$ est intéressant.

Proposition 11. *L'intersection non vide de deux demi-espaces de $\mathbb{R}^n$ est un convexe de dimension au pire $n-1$. Si c'est le cas alors c'est l'hyperplan affine séparateur.*

Ainsi pour obtenir une réduction de dimension, si la première équation était $\langle a, x\rangle \leq b$ on doit avoir $\langle -a, x\rangle \leq -b$ pour la seconde. Malheureusement, ce résultat ne persiste pas pour un polyèdre convexe (c'est-à-dire l'intersection de plusieurs demi-espaces).

Proposition 12. *Soit $\mathscr{P}$ un polyèdre convexe de $\mathbb{R}^n$. Alors, l'intersection non vide d'un demi-espace avec ce polyèdre convexe est un convexe de dimension au plus n. De plus celui-ci est de dimension k si et seulement si l'intersection du polyèdre convexe avec le demi-espace est une face de dimension k du polyèdre convexe.*

Pour déterminer les faces du polyèdre convexe on doit déterminer les sommets. Pour déterminer les faces il convient ensuite de considérer les combinaisons convexes des sommets qui préservent des égalité. Ce procédé peut être extrêmement couteux. En effet, le nombre de sommets est possiblement bien supérieur à la dimension. Par exemple, le cube $[0,1]^n$ a pour sommets 2^n points. Néanmoins pour des petits problèmes on peut toujours raison via des considérations géométriques.

Ayant obtenu des résultats sur le nombre de solutions lorsque la consistance est acquise on se tourne désormais vers le problème de déterminer si oui ou non il existe des solutions consistantes. On rappelle qu'on s'intéresse toujours au problème $Ax \leq b$ avec $A \in \mathscr{M}_{n,m}(\mathbb{R})$ et $b \in \mathbb{R}^n$.

Proposition 13. *Résoudre le problème $(\mathscr{P}_1)$: $Ax \leq b$ est consistant, revient à résoudre les problème $(\mathscr{P}_2)$: $-Bx \leq \tilde{A}^+ e_{m+1}$ est consistant et $(\mathscr{P}_2)'$: $\tilde{A}x = e_{m+1}$ où B est une matrice dont les colonnes forment une base de $\ker\tilde{A}$.*

Démonstration : On sait que $\mathscr{P}_1$ n'admet pas de solutions si et seulement si il existe un élément $c \geq 0$ tel que $\tilde{A}c = e_{m+1}$. L'ensemble des solutions de l'équation $\tilde{A}x = e_{m+1}$ (résolution du problème $\mathscr{P}_2'$) peut facilement être trouvé par élimination de Gauss. Si ce problème n'admet pas de solution alors $\mathscr{P}_1$ admet des solutions. Supposons maintenant que ce problème admette des solutions. L'ensemble de celle-ci est donné par $\tilde{A}^+ e_{m+1} + \ker\tilde{A}$. Il est à signaler qu'une base de $\ker\tilde{A}$ peut être identifiée lors de la recherche de solutions de $\mathscr{P}_2'$. Il s'agit alors de trouver un élément de cet ensemble dont toutes les coordonnées sont positives. Il faut donc résoudre le problème $\mathscr{P}_2$.

Remarque 1 : l'ajout du problème $\mathscr{P}_2'$ n'est pas un problème car celui-ci est facile à résoudre.

Remarque 2 : $B \in \mathscr{M}_{n,n-\text{rang}(A|b)}(\mathbb{R})$.

Remarque 3 : le problème des inégalité linéaires est très simple à résoudre si $\text{rang}(A) = 1$ ou $\text{rang}(A) = 0$. On dira alors que le problème est facile à résoudre.

Ainsi, via la proposition précédente on peut construire un nouveau problème de taille $(n, n - \text{rang}(A|b))$. Il peut être utile de considérer ce problème si $\text{rang}(A) = m$. Dans le cas général, il est compliqué de déterminer si le système est consistant ou non. On ne dispose que depuis récemment d'algorithmes qui convergent en temps polynomiaux et beaucoup de questions sont encore ouvertes à ce sujet.

3.5 Trouver une solution au problème d'inégalités aux moindres carrés.

3.5.1 Convexité

La fonction $F(x) = \|(Ax-b)_+\|^2$ que l'on cherche à minimiser est différentiable et convexe. Malheureusement elle est strictement convexe si et seulement si $\ker A = \{0\}$. Ainsi, si $\ker A \neq \{0\}$, on n'a aucune assurance que les algorithmes "classiques" de minimisation de fonctions convexes (gradient, gradient conjugué...) convergent vers la solution.

3.5.2 Elimination de Fourier-Motzkin

On va ici décrire une méthode découverte par Fourier pour déterminer si un système est consistant ou non. Malheureusement, comme on le verra, cette méthode n'est pas applicable car sa complexité est doublement exponentielle dans le pire cas.

4 L'algorithme de Han

L'algorithme de Han [2] qui permet de résoudre le problème des inégalité linéaires au moindre carré en temps fini.

4.1 Description

0n se donne $x_0 \in \mathbb{R}^m$. On définit $I = \{i \in [\![1,n]\!],\ (Ax_k - b)_i > 0\}$, ensemble des indices qui font que x_k n'est pas une solution. On considère la formule d'itération suivante (en omettant les indices sur l'ensemble I pour les notations) :

$$\begin{aligned} d_k &= A_I^+(A_I x_k - b) \\ \widehat{\lambda} &= \underset{\lambda\in\mathbb{R}}{\operatorname{argmin}}\ f(x-\lambda d_k) \\ x_{k+1} &= x_k - \widehat{\lambda} d_k \end{aligned}$$

Remarques :

— si l'on suppose que $I = [1;N]$ et que $A_I^T A_I = A^T A$ inversible, notre fonction F a minimiser est deux fois différentible de Hessienne inversible. L'algorithme de Newton s'appplique et donne comme direction de déscente :

$$\begin{aligned} d_k &= (A^T A)^{-1} A^T (Ax_k - b) \\ \equiv A^T A d_k &= A^T (Ax_k - b) \end{aligned}$$

— on peut essayer d'appliquer cette stratégie dans le cas général en cherchant d_k qui résolve le système

$$\begin{aligned} A_I^T A_I d_k &= A_I^T (A_I x_k - b_I) \\ \equiv ???? A_I d_k &= (A_I x_k - b_I) \end{aligned}$$

au sens des moindres carrés, c'est à dire $d_k \in A_I^+(A_I x_k - b_I)_+ \ker A_I$. On prend d_k de norm minimale, c'est à dire $d_k = A_I^+(A_I x_k - b_I)_+$. C'est la direction de descente proposée par l'algorithme de Han.

4.2 Preuve de convergence

Commençons par montrer que d_k ainsi défini est bien une direction de descente. On note $d(x) = A_I^+(A_I x - b_I)$.

Proposition 14. *On a* $\nabla f(x) = A_I^T A_I d(x)$ *et donc* $d(x)^T \nabla f(x) = \|A_I d(x)\|^2$

Démonstration : On a :

$$\begin{aligned} A_I^T A_I d(x) &= A_I^T A_I A_I^+ A_I x - A_I^T A_I A_I^+ b_I \\ &= A_I^T A_I x - A_I^T (A_I A_I^+)^T b_I \\ &= A_I^T A_I x - (A_I A_I^+ A_I)^T b_I \\ &= A_I^T A_I x - A_I^T b_I \\ &= \nabla f(x) \end{aligned} \tag{8}$$

Proposition 15. *$(f(x_k))_{k\in\mathbb{N}}$ décroit et $\nabla f(x_k) \underset{k\to+\infty}{\to} 0$.*

Démonstration : On écrit

$$\begin{aligned} f(x-\lambda d) - f(x) &= -\int_0^1 d^T \nabla(x-\lambda t d)\lambda \mathrm{d}t \\ &= -\int_0^1 d^T \nabla f(x-\lambda t d)\lambda \mathrm{d}t + \int_0^1 d^T \nabla f(x)\lambda \mathrm{d}t - d^T \nabla f(x)\lambda \\ &\leq \|A_I\|^2 \|d\| \lambda^2 \frac{1}{2} - \|A_I d\|^2 \lambda \end{aligned} \tag{9}$$

On s'est servi du caractère Lipschitz du gradient. $\|\nabla f(x) - \nabla f(y)\| \leq \|A\|^2 \|x - \|y\|$. On s'est également servi du fait que λ est positif (puisqu'on a une direction de descente et la fonction est convexe on peut se restreindre à ce cas). En minimisant le terme de droite en λ on trouve :

$$f(x-\lambda d) - f(x) \leq -\frac{1}{2} \frac{\|A_I d\|^4}{\|A_I\|^2 \|d\|^2} \tag{10}$$

Ainsi puisque la série de terme générale $f(x_k) - f(x_{k+1})$ est convergente, $\frac{\|A_I d_k\|^2}{\|d_k\|} \to 0$. Soit $d_k \to 0$ et dans ce cas $\nabla f(x_k) \to 0$. Sinon $A_I d_k \to 0$ et la conclusion est la même.
Il s'agit maintenant de montrer que le résidu $(Ax_k - b)_+$ converge vers le résidu (unique) d'une solution du problème.

Lemme 2. *Soit $(b_k, c_k)_{k\in\mathbb{N}}$ deux vecteurs tels que le systèmes $Ax \leq b_k$ et $Bx = c_k$ soit consistant pour tout $k \in \mathbb{N}$. On suppose que nos deux suites convergent vers b^*, respectivement c^*. Le système $Ax \leq b^*$ et $Bx = c^*$ est consistant.*

DEMO

Proposition 16. *Soit $z_k = (Ax_k - b)_+$ alors $z_k \to z^*$ où z^* est le résidu optimal.*

Démonstration : Par décroissance de $f(x_k)$ la suite z_k est contenue dans un compact. Soit une suite extraite qui converge vers $\overline{z}$. On la note encore z_k. Dans ce cas l'ensemble $I = \{i \in [\![1,n]\!],\ \langle a_i, x\rangle > b_i\}$ est constant à partir d'un certain rang. On estime que la suite z_k commence à partir de ce rang. Le système $A_I x = b_I + z_k$ est consistant quel que soit k. Donc via le lemme : $A_I x = b_I + \overline{z}$ est consistant. Donc on peut trouver un $\overline{x}$ tel que $\overline{z} = (A\overline{x} - b)_+$. Mais $\nabla f(\overline{x}) = A^T \overline{z}$ et est limite de $\nabla f(x_k)$ (donc vaut 0). On a alors $\overline{x}$ qui est solution aux moindres carrés et donc $\overline{z} = z^*$. On conclut dans le cas général en supposant que z_k ne tend pas vers z^*. Dans ce cas on extrait une sous-suite qui est toujours à distance ϵ de z^* mais on peut extraire une sous-suite convergente de cette sous-suite. Celle-ci converge vers z^* en vertu de ce qui a été dit au dessus. D'où l'absurdité et on a la convergence attendue.
On n'a toujours aucune information sur la suite x_k et sa convergence est dure à évaluer. On va montrer que la suite converge en un nombre fini d'étapes en s'intéressant au nombre d'indices $I(x_k) = \{i \in [\![1,n]\!],\ \langle a_i, x_k\rangle \geq b_i\}$. Si celui-ci stagne alors on aura une solution de notre problème, sinon il ne peut être que décroissant. Puisque c'est une suite d'entiers naturels, on pourra conclure sur la convergence.

Lemme 3. *Il existe $\epsilon \in \mathbb{R}_+^*$ tel que $\|(Ax-b)_+ - z^*\| \leq \epsilon$ implique que $A_{I(x)} y = b_{I(x)} + z^*_{I(x)}$ est consistant.*

Démonstration : On suppose que la proposition est fausse on a donc une suite x_k telle que $(Ax_k - b)_+$ tend vers z^* mais le système $A_{I(x_k)} y = b_{I(x_k)} + z^*_{I(x_k)}$ n'est jamais consistant. Puisqu'on a une infinité d'éléments de la suite et un nombre fini d'indices possibles on fixe I qui contient une infinité de termes. On renomme notre suite x_k en conséquence. On a toujours $A_I y = b_I + (z_k)_I$ qui est consistant. Donc $A_I y = b_I + z_I^*$ est consistant (par passage de la consistance à la limite). C'est absurde donc la proposition est vraie.

Lemme 4. *Il existe $\epsilon \in \mathbb{R}_+^*$ tel que si $\|(Ax-b)_+ - z^*\| \leq \epsilon$ alors $I(x-d) \subset I(x)$.*

Démonstration : On choisit $\epsilon \in \mathbb{R}_+^*$ tel que l'on soit dans les hypothèses du lemme précédent et qu'en plus $P = \{i \in [\![1,n]\!],\ z_i^* > 0\} \subset I(x)$. Dans ce cas $A_I y = b_I + z_I^*$ est consistant. On a $A^T z^* = A_I^T z_I^* = 0$ (puisque z^* est associé à une solution du problème aux moindres carrés et puisque $P \subset I$ on peut se limiter à z_I^* au lieu de z^*). On pose $\overline{y}$ une solution du système $A_I y = b_I + z_I^*$. C'est alors également une solution de $A_I^T A_I y = A_I^T b_I$. Donc $\overline{y}$ est une solution du problème aux moindres carrés $A_I y = b_I$. On a la même propriété sur $x-d$. Par égalité des résidus : $A_I(x-d) = b_I + z_I^*$. Par positivité de z_I^* on conclut sur l'inclusion.

Lemme 5. *Si $I(x) = I(x-d)$ alors $x - \lambda d$ est une solution aux moindres carrés. Si on a inclusion stricte alors l'inclusion est stricte également de $I(x)$ dans $I(x-\lambda d)$.*

Démonstration : Il est aisé de constater que si $I(x) = I(x-d)$ alors en se servant de $A_I^T A_I d = \nabla f(x)$ on a une solution de l'inégalité au sens des moindres carrés. Donc, $x - \lambda d$ est également solution (puisque $f(x - \lambda d) \leq f(x-d)$). On montre que $\lambda \in]0,1]$ et donc que $x - \lambda d = \lambda(x-d) + (1-\lambda)x$. Ainsi $I(x) \subset I(x-\lambda d)$ (le fait que $\lambda \in]0,1]$, vient en étudiant $f(x-\lambda d)$ en 0 et 1. On ne peut pas avoir égalité car dans ce cas $\lambda = 1$ (toujours en étudiant $g'(t) = \nabla f(x - td)$ en λ ici). Dans ce cas $x - \lambda d = x - d$ et c'est absurde au vu de l'hypothèse d'inclusion stricte. Donc on a inclusion stricte.
Ces trois lemmes permettent d'énoncer (sans démonstration, il s'agit simplement d'appliquer les deux derniers lemmes...) le théorème suivant.

Théorème 5. *L'algorithme de Han converge en un nombre fini d'étapes.*

Remarque : néanmoins, on ne dispose pas de borne sur le temps de convergence de l'algorithme, ni même sur la distance à une solution.

5 Une tentative de régularisation de l'algorithme de Han

Dans cette sous section on considère le problème des inégalité linéaires au moindre carré avec régularisation L^2. Le problème devient alors de trouver l'argument minimum de la fonction suivante :

$$F(x) = \|(Ax-b)_+\|^2 + \alpha\|x\|^2 \tag{11}$$

Avec $\alpha \in \mathbb{R}_+^*$. Cette fonction définie sur $\mathbb{R}^m$ est différentiable, convexe. De plus, elle est coercive donc l'existence du minimum est assuré (cette existence représentait une véritable difficulté sans la régularisation L^2). Tout argument minimum $\hat{x}$, celui-ci doit vérifier :

$$A^T(A\hat{x}-b)_+ + \alpha\hat{x} = 0 \tag{12}$$

Il s'agit donc de résoudre un problème de point fixe :

$$x = -\frac{1}{\alpha}\left(A^T(Ax-b)_+\right) \tag{13}$$

On pose $f_\alpha(x) = -\frac{1}{\alpha}\left(A^T(Ax-b)_+\right)$.

5.1 Itérées et minimum

Si α est assez grand alors la fonction f_α est contractante. Dans ce cas, on peut définir la suite $x_n = f_\alpha(x_{n-1})$ avec initialisation quelconque. Le minimum est alors unique et limite de cette suite. Dans les figures suivantes on présente une courte étude numérique de ce phénomène. Dans la suite on a choisi une matrice dont les coefficients sont choisis selon une loi uniforme entre 0 et 1. On prend cette matrice de taille 100×100. De la même manière on choisit un vecteur b dont les coefficients sont choisis selon une loi uniforme entre $-\frac{1}{2}$ et $\frac{1}{2}$. N correspond au nombre d'itérations de l'algorithme. La matrice et les vecteurs utilisés ici pour nos expériences définissent un système consistant (pour vérifier cela on a utilisé 13).

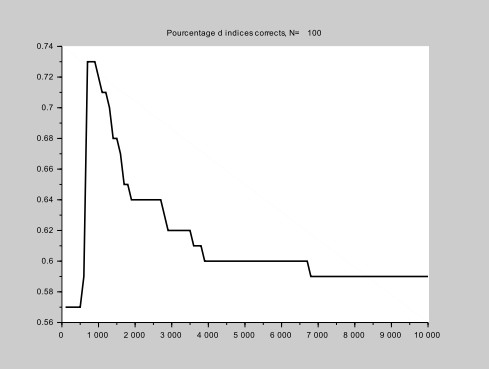

FIGURE 1 – Pourcentage d'indices valides pour $N = 100$

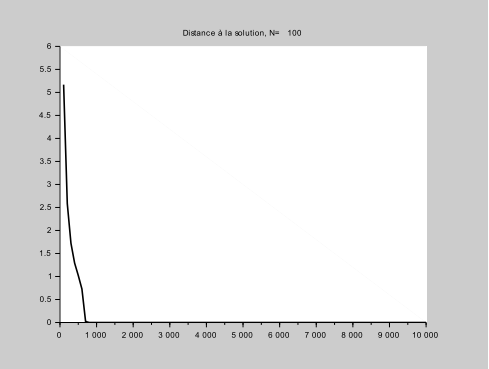

FIGURE 2 – Distance entre la dernière itération et l'avant-dernière pour $N = 100$

Ici on étudie l'évolution du nombres d'indices corrects, c'est-à-dire tels que $(Ax)_i \leq b_i$. On constate que les résultats se situent d'abord autour d'une moitié d'indices corrects ce qui n'est pas satisfaisant. On constate que pour ces α trop faibles l'algorithme n'a pas convergé. Ensuite, on observe une forte augmentation puis une décroissance stricte du nombre d'indices corrects. Cela est dû au fait qu'on cherche de plus en plus à minimiser la norme de x sans chercher à minimiser la norme de $(Ax-b)_+$. Conformément à ce qu'on attendait, il s'agit de prendre α le plus petit possible pour que le plus d'inégalité soient satisfaites.

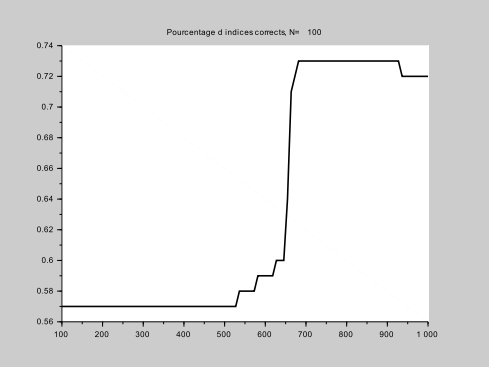

FIGURE 3 – Pourcentage d'indices valides pour $N = 100$

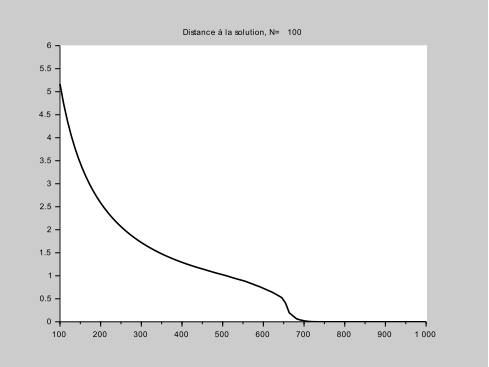

FIGURE 4 – Distance entre la dernière itération et l'avant-dernière pour $N = 100$

Ici, on a détaillé le comportement de l'algorithme lorsque α est petit. Il est clair que l'augmentation du nombre d'indices corrects est dû à la convergence de l'algorithme. Néanmoins, au maximum, seuls 74 indices sur 100 vérifient les contraintes. Pour faire mieux il serait souhaitable de diminuer la valeur de α. Malheureusement on perd alors la convergence.

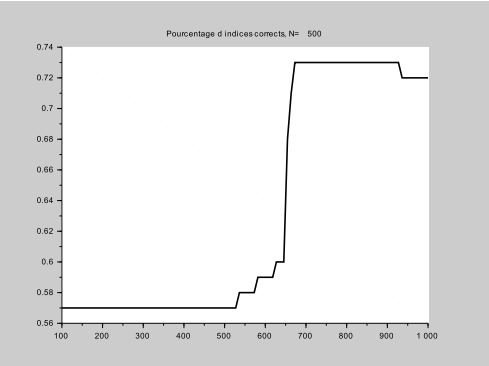

FIGURE 5 – Pourcentage d'indices valides pour $N = 500$

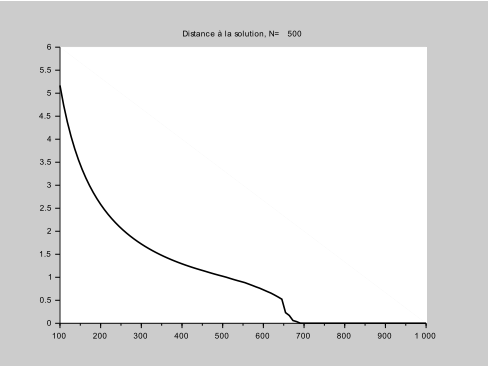

FIGURE 6 – Distance entre la dernière itération et l'avant-dernière pour $N = 500$

On a ici tenté d'augmenter le nombre d'itérations de l'algorithme afin de déterminer si le problème de la

non-convergence est dû à un α trop petit ou à un nombre d'itérations trop faible. Le fait que les résultats soient quasiment lorsque $N = 500$ nous incite à pencher vers la première hypothèse d'un α trop petit.

Références

[1] Penot Jean-Paul Contesse Luis, Hiriart-Urruty Jean-Baptiste. Least-squares solution of linear inequality systems : a pedestrian approach.

[2] Han. Least-squares solution of linear inequalities. 1980.

[3] Roger Penrose. A generalized inverse for matrices. In *Mathematical proceedings of the Cambridge philosophical society*, volume 51, pages 406–413. Cambridge Univ Press, 1955.

6 Une autre approche du problème

6.1 Du cas d'inégalité au cas d'égalité sous contrainte

TODO

6.2 Algorithme de gradient projeté

TODO