OPTIMISATION

LICENCE MATHÉMATIQUES ET GESTION 2013–2014

Table des matières

3 Analyse matricielle **1**
3.1 Le problème 1
3.2 Généralités sur les matrices 2
3.3 Erreurs 2
3.4 Pivot de Gauss 3
3.5 Factorisation des matrices 4
3.6 Méthodes itératives 4
3.7 Exercices 6
3.8 Un peu de R 6

Plan général

1. **Modélisation**
2. **Calcul différentiel**
3. **Analyse matricielle**
4. **Programmation linéaire**
5. **Introduction à l'optimisation nonlinéaire**

3 Analyse matricielle

3.1 Le problème

Résoudre

$$(1) \quad Ax = b,\ A \in M_n(\mathbb{R}),\ b \in \mathbb{R}^n$$

ici, A est inversible

Une source possible
La régression linéaire multiple : trouver

$$\min_{x\in\mathbb{R}^n} |Bx - c|^2,\ B \in M_{m,n}(\mathbb{R}), c \in \mathbb{R}^m$$

Pour comprendre l'origine du problème (identification de paramètres en statistique) : `http://fr.wikipedia.org/wiki/Régression_linéaire_multiple`

Solution

On suppose B injective (hypohèse statistique raisonnable, cf supra). Alors $|Bx| \geq C|x|$ (vérifier !), d'où $|Bx-c|^2$ coercive (vérifier !). On trouve que le minimum est atteint par un x qui vérifie

$$(Bx) \cdot (Bh) - (Bh) \cdot c = 0, \ \ \forall\, h \in \mathbb{R}^n,$$

d'où $\underbrace{B^*B}_{A}\, x = \underbrace{B^*c}_{b}$ (vérifier !)

Par ailleurs, $A \in M_n(\mathbb{R})$ est inversible, car $Ax = 0 \Longrightarrow (Ax)\cdot x = 0 \Longrightarrow |Bx|^2 = 0 \Longrightarrow x = 0$

Remarque
A est définie positive : $(Ax)\cdot x > 0$ si $x \in \mathbb{R}^n \setminus \{0\}$

3.2 Généralités sur les matrices

Identité de Lagrange
Si $A \in M_{m,n}(\mathbb{R})$, $x \in \mathbb{R}^n$ et $y \in \mathbb{R}^m$, alors $(Ax)\cdot y = x \cdot (A^* y)$

Définitions
Une matrice $A = (a_i^j) \in M_n(\mathbb{R})$ est :
- diagonale, $A =$diag$(\lambda_1, \ldots, \lambda_n)$, si $a_i^i = \lambda_i$ et $a_i^j = 0$ si $i \neq j$
- triangulaire inférieure (resp. triangulaire inférieure stricte) si $a_i^j = 0$ pour $j > i$ (resp. pour $j \geq i$)
- triangulaire supérieure (resp. triangulaire supérieure stricte) si $a_i^j = 0$ pour $j < i$ (resp. pour $j \leq i$)

Théorème 3.1. Toute matrice carrée est semblable à une matrice triangulaire supérieure : $A = P^{-1}TP$, avec T triangulaire supérieure, ayant sur les diagonales les valeurs propres de A (multiplicités comprises)

Norme subordonnée
Si $\| \ \|$ est une norme sur $\mathbb{R}^n$ et $A \in M_n(\mathbb{R})$, alors la norme (matricielle)subordonnée à $\| \ \|$ est

$$\|A\| = \sup\{\|Ax\|/\|x\|; x \in \mathbb{R}^n \setminus \{0\}\} =$$

la plus petite constante $C \geq 0$ telle que $\|Ax\| \leq C\|x\|$, $\forall\, x \in \mathbb{R}^n$

Proposition 3.2.
- $A \mapsto \|A\|$ est une norme
- $\|AB\| \leq \|A\|\|B\|$

Exercice 1. Calculer, en fonction des éléments de A, la norme de A subordonnée à
- $\|x\|_1 := \displaystyle\sum_{j=1}^{n} |x_j|$
- $\|x\|_\infty := \displaystyle\max_{j=[\![1,n]\!]} |x_j|$

Définition 3.3. Le rayon spectral d'une matrice $A \in M_n(\mathbb{R})$ est

$$\rho(A) := \sup\{|\lambda|; \lambda \text{ valeur propre de } A\}$$

Définition 3.4. Le spectre d'une matrice $A \in M_n(\mathbb{R})$ est

$$\sigma(A) := \{\lambda \in \mathbb{C}; \lambda \text{ valeur propre de } A\}$$

Donc

$$\rho(A) = \sup\{|\lambda|; \lambda \in \sigma(A)\}$$

Théorème 3.5 (Formule du rayon spectral)**.** Si $A \in M_n(\mathbb{R})$ et $\| \ \|$ est une norme subordonnée, alors

$$\rho(A) = \lim_{k\to\infty} \|A^k\|^{1/k}$$

3.3 Erreurs

Erreurs
Si b est la valeur exacte d'une variable, et $b + \delta b$ la valeur calculée, alors :
- $\|\delta b\|$ est l'erreur absolue
- $\dfrac{\|\delta b\|}{\|b\|}$ est l'erreur relative (qui est une mesure plus pertinente de l'erreur)

Conditionnement
Si $A \in M_n(\mathbb{R})$ est inversible, alors le conditionnement de A est

$$\text{cond }(A) = \|A\|\|A^{-1}\|$$

Intuition
cond (A) mesure à quelle point A est proche/loin d'un multiple de l'identité (si le conditionnement est petit/grand), sachant que le conditionnement est toujours ≥ 1

Conditionnement et erreur
Si on a le système $Ax = b$, avec A inversible, et b est mesuré avec une erreur, $b \rightsquigarrow b + \delta b$, alors la solution devient $x + \delta x$

Estimation de l'erreur commise :

$$\frac{\|\delta x\|}{\|x\|} = \frac{\|A^{-1}\delta b\|}{\|x\|} \leq \text{cond }(A)\frac{\|\delta b\|}{\|b\|}$$

Intuition (bis)
cond (A) mesure la robustesse du système $Ax = b$ (=peu de sensibilité à des erreurs de mesure)

Aperçu des méthodes de résolution

- Echelonnement de la matrice (pivot de Gauss)
- Factorisation de la matrice en matrices simples
- Méthodes itératives

3.4 Pivot de Gauss

Etape $\#1$: réduction à un système triangulaire supérieur $Bx = c$
- Pour $i = [\![1, n]\!]$
- Sur la i-ème ligne : on cherche le premier indice $j \geq i$ tel que $a_i^j \neq 0$
- On permute les colonnes i et j, de sorte à avoir $a_i^i \neq 0$
- Pour $k = [\![i + 1, n]\!]$
- Remplacer la ligne l^k du système (1) par $l^k - \dfrac{a_i^k}{a_i^i} l^i$

Etape $\#2$: résolution du système triangulaire supérieur $Bx = c$
- Pour $i = [\![n, 1]\!]$
- Calculer x_i en fonction de x_k, $k = [\![i + 1, n]\!]$, à partir de l'égalité

$$b_i^i x_i = c_i - \sum_{k=i+1}^{n} b_i^j x_j$$

Analyse de l'algorithme

- L'algorithme fonctionne (au sens où on peut toujours trouver $j \geq i$ tel que $a_i^j \neq 0$)
- Le nombre de calculs est de l'ordre de n^3
- Algorithme sensible aux erreurs d'arrondi

Exemple

Pour le système

$$\begin{cases} \varepsilon x + y = \varepsilon \\ x + y = 1 \end{cases},$$

la solution exacte est $(1,0)^T$, mais une erreur d'arrondi d'ordre ε dans les calculs entraîne une erreur d'ordre 1 dans la solution

Pivot de Gauss modifié

On ne change que l'étape $\#1$

- Pour $i = [\![1, n]\!]$
- On cherche un couple (l, j), avec $l \geq i$ et $j \geq i$, tel que $|a_l^j|$ soit maximal
- On permute les lignes i et l, et les colonnes i et j, de sorte à avoir $|a_i^i| \geq |a_s^t|$ si $s \geq i$ et $t \geq i$
- Pour $k = [\![i+1, n]\!]$
- Remplacer la ligne l^k du système (1) par $l^k - \dfrac{a_i^k}{a_i^i} l^i$

3.5 Factorisation des matrices

Aperçu de la factorisation LU

Le but du jeu est d'écrire $A = LU$, avec L triangulaire inférieure et U triangulaire supérieure

On résout alors $Ax = b$ en deux temps : $Ly = b$, puis $Ux = y$

Aperçu de la factorisation QR

On cherche à mettre A sous la forme $A = QR$, avec Q orthogonale, R triangulaire supérieure

On ramène alors $Ax = b$ à : $Qy = b$ (ou encore $y = Q^* b$) et $Rx = y$

Factorisation de Cholesky

Donnée : $A > 0$. Sortie : R triangulaire supérieure telle que $A = R^* R$.

Algorithme :

- Pour $i = [\![1, n]\!]$
- Prendre $r_i^i = \left(a_i^i - \sum_{k=1}^{i-1} (r_k^i)^2 \right)^{1/2}$
- Pour $j = [\![i+1, n]\!]$
- Poser $r_i^j = \dfrac{1}{r_i^i} \left(a_i^j - \sum_{k=1}^{i-1} r_k^i r_k^j \right)$

3.6 Méthodes itératives

Aperçu

Pour résoudre (1) $Ax = b$

- On décompose $A = M - N$, avec M inversible
- (1) devient alors (2) $Mx = Nx + b$, ou encore (3) $x = M^{-1} N x + c$, avec $c = M^{-1} b$
- On essaie alors de résoudre (3) par la méthode des itérations de Picard

Algorithme

On se donne des critères d'arrêt : une norme $\| \ \|$ sur $\mathbb{R}^n$ et un nombre $\varepsilon > 0$

- On choisit $x^0 \in \mathbb{R}^n$ (dans la pratique, $x^0 = 0$)
- Pour $k \geq 0$
- On s'arrête si $\|Ax^k - b\| \leq \varepsilon$
- Sinon, on pose $x^{k+1} = \underbrace{M^{-1}N}_{B} x^k + c$

Théorème 3.6. L'algorithme de Picard converge (pour tout b et x^0) si et seulement si $\rho(B) < 1$

Démonstration de $\Longrightarrow$.
On prend $b = 0$ (d'où $x = 0$) et x^0 vecteur propre de B, de valeur propre λ
On a alors $x^j = \lambda^j x^0 \to 0$, d'où $|\lambda| < 1$ □

Démonstration de $\Longleftarrow$.
On écrit $B = P^{-1}TP$, avec T triangulaire supérieure stricte et $|t_i^i| < 1$
Soit x la solution de (1) et $y^k := P(x^k - x)$. Alors $y^{k+1} = Ty^k$
Par récurrence rétrograde sur $j \in [\![1, n]\!]$, on montre que $y_j^k \to 0$ pour $k \to \infty$, d'où $x^k \to x$ □

Proposition 3.7. On a $\rho(B) \leq \|B\|$, quelle que soit la norme subordonnée

Démonstration. Soient λ valeur propre de B et x vecteur propre associé. Alors $\|Bx\| = |\lambda|\|x\|$, d'où $\|B\| \geq |\lambda|$. On trouve

$$\|B\| \geq \sup\{|\lambda|; \lambda \text{ valeur propre de } B\} = \rho(B) \quad \square$$

Corollaire 3.8. S'il existe une norme subordonnée telle que $\|B\| < 1$, alors l'algorithme de Picard converge

Méthode de Jacobi
Prendre $M =$ la partie diagonale de A et $N = M - A$

Avantage principal
M^{-1} facile à calculer

Théorème 3.9. On suppose

$$|a_i^i| > \sum_{j \neq i} |a_i^j|$$

(on dit que A est à diagonale dominante)
Alors la méthode de Jacobi converge

Démonstration. Soient $x \neq 0$ et $\lambda \in \mathbb{C}$ tels que $M^{-1}Nx = \lambda x \Longrightarrow Nx = \lambda Mx \Longrightarrow -\sum_{j \neq i} a_i^j x_j = \lambda a_i^i x_i$, pour tout i
On prend i tel que $|x_i| \geq |x_j|$ pour tout $j \Longrightarrow$

$$|\lambda| \leq \frac{1}{|a_i^i|} \sum_{j \neq i} \frac{|x_j|}{|x_i|} |a_i^j| \leq \frac{1}{|a_i^i|} \sum_{j \neq i} |a_i^j| < 1 \quad \square$$

Méthode de Gauss-Seidel
Prendre M=la partie inférieure (ou supérieure) de A et $N = M - A$
Donc $m_i^j = \begin{cases} a_i^j, & \text{si } i \geq j \\ 0, & \text{sinon} \end{cases}$ et $n_i^j = \begin{cases} -a_i^j, & \text{si } i < j \\ 0, & \text{sinon} \end{cases}$

Théorème 3.10. On suppose $A > 0$. Alors la méthode de Gauss-Seidel converge

Démonstration. On écrit $A = D + L + U$, avec D diagonale, L triangulaire inférieure stricte et U triangulaire supérieure stricte, de sorte que $M = D + L$ et $N = -U$
On a A symétrique$\Longrightarrow U^* = L \Longrightarrow M + N^* = D$
Par ailleurs, on a $d_i^i = (Ae^i) \cdot e^i > 0$, d'où $M + N^* = D > 0$
En rappelant que $A = M - N > 0$, on conclût grâce au lemme suivant □

Lemme 3.11. Si $M - N > 0$ et $M + N^* > 0$, alors $\rho(M^{-1}N) < 1$

Démonstration. Si $x \neq 0$ et $M^{-1}Nx = \lambda x$, alors $Nx = \lambda Mx$

On trouve

- $\underbrace{(Nx) \cdot x}_{\beta} = \lambda \underbrace{(Mx) \cdot x}_{\alpha}$
- $\underbrace{(Mx) \cdot x}_{\alpha} - \underbrace{(Nx) \cdot x}_{\beta} > 0$
- $\underbrace{(Mx) \cdot x}_{\alpha} + \underbrace{(N^*x) \cdot x}_{\overline{\beta}} > 0$

Les deux dernières inégalités donnent $|\alpha| > |\beta|$, d'où $|\lambda| < 1$ □

3.7 Exercices

Exercice 1. Soit $A > 0$.

- Trouver les valeurs de α telles que la méthode de Richardson, consistant à prendre $M = \alpha I_n$, converge
- Y a-t-il un α qui semble meilleur que les autres ?

Exercice 2. Mettre en œuvre, sous R, la méthode de Gauss-Seidel

Exercice 3. Etudier la méthode suivante (dite "des relaxations successives") : on décompose $A > 0$ sous la forme $A = D + L + U$ (comme dans la méthode de Gauss-Seidel), puis on considère la méthode itérative

$$\begin{cases} x^{k+1} = \alpha y^{k+1} + (1-\alpha)x^k \\ Dy^{k+1} + Lx^{k+1} = -Ux^k + b \end{cases}$$

3.8 Un peu de R

Une façon de rentrer une matrice A :

- On définit le vecteur c des éléments de A **colonne par colonne**
  $> x < -c($ici, on met les éléments de $A)$
- Puis, on définit A
  $> A < -x; dim(A) < -c($nombre de lignes de A, nombres de colonnes de A $)$

La transposée d'une matrice A est appelée avec la commande
$> B < -t(A)$

Le produit des matrices A et B est appelé avec la commande
$> C < -A \,\%*\%\, B$

Inversion d'une matrice triangulaire inférieure

- $> n < -$taille de la matrice
- $> x < -c($éléments de A par colonnes$)$
- $> A < -x \; ; dim(A) < -c(n, n)$
- $> B < -matrix(nrow = n, ncol = n)$
- $> for\ (i\ in\ 1 : n)\ for\ (j\ in\ 1 : n)\ B[i, j] < -0$
- $> for\ (i\ in\ 1 : n)\ B[i, i] < -1/A[i, i]$
- $> for\ (i\ in\ 2 : n)\ for\ (j\ in\ 1 : (i-1))\ \ for\ (k\ in\ 1 : (i-1))\ B[i, j] < -B[i, j] - A[i, k] * B[k, j]/A[i, i]$

Méthode de Gauss-Seidel

- On calcule M^{-1}, $B = M^{-1}N$ et $c = M^{-1}b$
- $> test < -$test d'arrêt
- $> X < -numeric(n)$
- $> for\ (i\ in\ 1 : n)\ X[i] = 0$
- $> while\ (crossprod(A \,\%*\%\, X - b) > test^2)\ X < -B \,\%*\%\, X + c$
- $> X$