

Chapitre 10

Méthodes itératives 4

Méthodes de Jacobi, Gauss-Seidel gradient, gradient conjugué Optimisation

Méthodes numériques 2001/2002 - D.Pastre
licence de mathématiques et licence MASS

1

Méthodes de Jacobi et Gauss-Seidel

Méthodes de point fixe

Pour résoudre $AX = B$, on décompose A sous la forme $A = G - H$ avec G inversible.

Alors

$$AX = B \Leftrightarrow GX = HX + B \Leftrightarrow X = MX + N$$

avec $M = G^{-1}H$ et $N = G^{-1}B$

On est alors ramené à un problème de recherche de point fixe :

On définit la suite récurrente

$$\begin{cases} X^{(0)} \text{ vecteur quelconque de } \mathbb{R}^n \\ X^{(k+1)} = M X^{(k)} + N \end{cases}$$

Si cette suite est convergente alors sa limite X est solution du système.

Dans la pratique, il faudra décomposer $A = G - H$ de façon que la suite soit convergente et que G^{-1} se déduise facilement de G ou plus simplement que le système $GX^{(k+1)} = HX^{(k)} + B$ se résolve simplement. On choisira pour G une matrice diagonale (**Jacobi**) ou triangulaire (**Gauss-Seidel**).

2

Définitions

Le **rayon spectral** $\rho(M)$ d'une matrice M est le plus grand des modules des valeurs propres de M .

Remarque : on considère toutes les valeurs propres, y compris les valeurs complexes.

Condition nécessaire et suffisante de convergence

Théorème

La suite $X^{(k)}$ ainsi définie converge pour tout $X^{(0)}$ si et seulement si $\rho(M) < 1$.

La convergence est d'autant plus rapide que $\rho(M)$ est petit.

3

Méthode de Jacobi

On écrit $A = D - E - F$ où D est la partie diagonale de A , E (resp. F) la partie triangulaire inférieure (resp. supérieure) changée de signe de A .

On suppose que tous les a_{ii} sont différents de 0, donc D est inversible. On choisit alors $G = D$ et $H = E + F$.

On a alors $M = D^{-1}(E + F) = D^{-1}(D - A) = I - D^{-1}A$

$$= \begin{bmatrix} 0 & -\frac{a_{12}}{a_{11}} & -\frac{a_{13}}{a_{11}} & \dots & -\frac{a_{1n}}{a_{11}} \\ -\frac{a_{21}}{a_{22}} & 0 & -\frac{a_{23}}{a_{22}} & \dots & -\frac{a_{2n}}{a_{22}} \\ \dots & & & & \\ -\frac{a_{n1}}{a_{nn}} & -\frac{a_{n2}}{a_{nn}} & -\frac{a_{n3}}{a_{nn}} & \dots & 0 \end{bmatrix}$$

Le calcul effectif se fait en résolvant directement le système

$$D X^{(k+1)} = (E + F) X^{(k)} + B$$

soit

$$x_i^{(k+1)} = \frac{b_i - \sum_{j=1, j \neq i}^n a_{ij} x_j^{(k)}}{a_{ii}}$$

4

Définition

On dit que A est à **diagonale strictement**

dominante si $\forall i \quad |a_{ii}| > \sum_{j=1, j \neq i}^n |a_{ij}|$

ou si $\forall i \quad |a_{ii}| > \sum_{j=1, j \neq i}^n |a_{ji}|$

Condition suffisante de convergence

Théorème

Si A est une matrice à diagonale strictement dominante alors la méthode de Jacobi appliquée au système $A X = B$ est convergente pour tout $X^{(0)}$

Méthode de Gauss-Seidel

A partir de $A = D - E - F$ on choisit

$$G=D-E \text{ et } H=F.$$

On a alors $M = (D-E)^{-1} F$ et $N = (D-E)^{-1} B$.

Le calcul effectif se fait en résolvant directement le système

$$(D-E) X^{(k+1)} = F X^{(k)} + B$$

soit

$$\sum_{j=1}^i a_{ij} x_j^{(k+1)} = - \sum_{j=i+1}^n a_{ij} x_j^{(k)} + b_i$$
$$x_i^{(k+1)} = \frac{b_i - \sum_{j=1}^{i-1} a_{ij} x_j^{(k+1)} - \sum_{j=i+1}^n a_{ij} x_j^{(k)}}{a_{ii}}$$

Remarque : le calcul de $x_i^{(k+1)}$ fait intervenir les valeurs des $x_j^{(k)}$ pour $j > i$ et des $x_j^{(k+1)}$ pour $j < i$. On fera donc le calcul pour i allant de 1 à n .

Conditions suffisantes de convergence

Théorème 1

Si A est une matrice à diagonale strictement dominante alors la méthode de Gauss-Seidel appliquée au système $A X = B$ est convergente pour tout $X^{(0)}$

Théorème 2

Si A est définie positive alors la méthode de Gauss-Seidel appliquée au système $A X = B$ est convergente pour tout $X^{(0)}$

Les méthodes de Jacobi et de Gauss-Seidel sont appelées techniques de relaxation. La méthode de Gauss-Seidel converge plus vite que la méthode de Jacobi.

Point de vue géométrique

Supposons A symétrique définie positive.

Au système $AX=B$ on associe la fonctionnelle

$$\text{quadratique } J(X) = \frac{1}{2} {}^t X A X - {}^t X B$$

Propriété

X_0 est solution du système $A X = B$ si et seulement si $J(X)$ peut se mettre sous la forme

$$J(X) = \frac{1}{2} ({}^t X - {}^t X_0) A (X - X_0) - C_0$$

où C_0 est une constante.

$J(X_0)$ est égal au minimum de la fonction $J(X)$

On a, pour tout X , $J(X) \geq C_0$ et X_0 est appelé le centre de la famille d'ellipsoïdes homothétiques $(J^{-1}(c))_{c \geq C_0}$

Définitions

On appelle **résidu** (ou *résidus*) relatif à Y le vecteur $r(X) = A X - B$

Le **gradient** de J en x est le vecteur $\frac{\partial J}{\partial x_i}(X)$

Notation $\nabla J(X)$

Proposition

Si $\nabla J(X)$ est le gradient de J en X , on a

$$\nabla J(X) = r(X)$$

Interprétation géométrique de la méthode de Jacobi

(pour une matrice A définie positive)

$X^{(k)}$ est sur l'ellipsoïde $J^{-1}J(X^{(k)})$

On construit les n points $X_i^{(k+1)}$ à partir de $X^{(k)}$ en annulant le $i^{\text{ème}}$ résidu. $X_i^{(k+1)}$ est sur l'ellipsoïde tangent à la droite parallèle à l'axe Ox_i passant par $X^{(k)}$. Le point $X^{(k+1)}$ est le point dont les coordonnées sont les coordonnées à l'ordre $k+1$ des $X_i^{(k+1)}$.

Interprétation géométrique de la méthode de Gauss-Seidel

On construit le point $X_{i+1}^{(k+1)}$ à partir de $X_i^{(k+1)}$ en annulant le $(i+1)^{\text{ème}}$ résidu.

Méthode du gradient (ou de la plus grande pente)

A partir du point $X^{(k)}$, qui est sur l'ellipsoïde $J^{-1}J(X^{(k)})$, on trace la normale à cet ellipsoïde. $X^{(k)}$ est le point de contact de cette normale avec l'ellipsoïde de la famille qui lui est tangent, soit

$$X^{(k+1)} = X^{(k)} + t_k r^{(k)}$$

avec

$r^{(k)} = AX^{(k)} - B$ et $r^{(k+1)}$ perpendiculaire à $r^{(k)}$

Le calcul donne (à vérifier en exercice)

$$r^{(k+1)} = r^{(k)} - \frac{t_{r^{(k)}} r^{(k)}}{t_{r^{(k)}} A r^{(k)}} A r^{(k)}$$

$$X^{(k+1)} = X^{(k)} - \frac{t_{r^{(k)}} r^{(k)}}{t_{r^{(k)}} A r^{(k)}} r^{(k)}$$

Théorème

Soit λ_M (resp. λ_m) la plus grande (resp. la plus petite) valeur propre de A . La méthode du gradient converge, quel que soit le vecteur initial $X^{(0)}$, vers la solution X_0 si $\frac{\lambda_M}{\lambda_m} < \sqrt{2}$

Méthode du gradient conjugué

On détermine $X^{(k+1)}$ à partir du point $X^{(k)}$ de la façon suivante :

$U^{(k)}$ et $U^{(k-1)}$ sont *conjugués* par rapport à l'ellipse section de l'ellipsoïde $J^{-1}(J(X^{(k)}))$ par le plan passant par $X^{(k)}$ et de vecteurs directeurs $U^{(k-1)}$ et $\nabla J(X^{(k)}) = r^{(k)}$, soit

$$U^{(k)} = -r^{(k)} + s_k U^{(k-1)} \text{ et } U^{(k-1)} A U^{(k)} = 0$$

(avec $U^{(0)} = -r^{(0)}$)

la droite passant par $X^{(k)}$ et de vecteur directeur $U^{(k)}$ est tangente en $X^{(k+1)}$ à l'ellipsoïde $J^{-1}(J(X^{(k+1)}))$, soit

$$X^{(k+1)} = X^{(k)} + t_k U^{(k)} \text{ et } t_{U^{(k)}} r^{(k+1)} = 0$$

Le calcul donne (à vérifier en exercice)

$$s_k = \frac{t_{U^{(k-1)}} A r^{(k)}}{t_{U^{(k-1)}} A U^{(k-1)}}$$

$$t_k = \frac{t_{U^{(k)}} r^{(k)}}{t_{U^{(k)}} A U^{(k)}} = \frac{t_{r^{(k)}} r^{(k)}}{t_{U^{(k)}} A U^{(k)}}$$

Théorème

La méthode du gradient conjugué converge en n itérations au plus. (C'est donc une méthode directe).

Optimisation

Les méthodes qui viennent d'être étudiées pour résoudre les systèmes linéaires consistent à chercher le minimum (*l'optimum*) d'une fonctionnelle $J(X)$. Ce sont des *problèmes d'optimisation*.

De nombreux autres problèmes sont aussi équivalents à des problèmes d'optimisation, c'est-à-dire à la recherche d'un optimum d'une fonctionnelle $J(X)$ (qui peut être beaucoup plus compliquée que dans les exemples qui viennent d'être vus). Les méthodes de relaxation et de gradient se généralisent.

Approximation au sens des moindres carrés

Exemple : si le système $AX=B$ n'a pas de solution, il peut être intéressant de trouver un X qui s'en approche le plus possible. On appelle approximation au sens des moindres carrés un X qui minimise la norme

$$\|AX - B\| = ({}^tX {}^tA - {}^tB)(AX - B)$$

Une application de cette approximation aux moindres carrés consiste à chercher une fonction polynomiale de degré donné qui passe aussi près que possible d'un grand nombre de points. Des méthodes d'interpolation permettent de trouver un polynôme de degré n dont la courbe passe par $n+1$ points donnés, mais il peut être préférable (rapidité de calcul ou inexactitude sur les données) de chercher un polynôme de moindre degré qui ne fait qu'approximer la courbe.

Pour trouver un tel polynôme $\sum_{i=1}^n a_i x^i$ passant au plus près des points (x_i, y_i) , on cherchera une approximation au sens des moindres carrés du système

$$\begin{bmatrix} 1 & x_0 & x_0^2 & \dots & x_0^n \\ 1 & x_1 & x_1^2 & & x_1^n \\ 1 & x_2 & x_2^2 & & x_2^n \\ \dots & & & & \\ 1 & x_n & x_n^2 & & x_n^n \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \dots \\ a_n \end{bmatrix} = \begin{bmatrix} y_0 \\ y_1 \\ y_2 \\ \dots \\ y_n \end{bmatrix}$$

où les inconnues sont les a_i .

Pour chercher une droite $y=ax+b$ par exemple passant au plus près des n points, on résoudra le système

$$\begin{bmatrix} 1 & x_0 \\ 1 & x_1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} y_0 \\ y_1 \end{bmatrix} \text{ d'inconnues } a \text{ et } b.$$

Attention, on ne minimise pas la somme des distances des points à la droite mais la somme des différences entre les ordonnées des points et de la droite pour les x_i .