INSA INSTITUT NATIONAL DES SCIENCES APPLIQUÉES ROUEN NORMANDIE

Projet MMSN
GM3 2022 - 2023

Gradient Conjugué

Étudiants :
LANGOLFF Clément
KESSLER Aymeric

Enseignant-responsable du projet :
El Bouchairi Imad

Table des matières

Première partie
Introduction

0.1 Fonctionnelle à minimiser

La méthode du gradient conjugué fait partie des méthodes de descente. À chaque itération, on determine un vecteur direction d_k et un scalaire α_k permettant de calculer une nouvelle approche de la solution $x_{k+1} = x_k + \alpha_k d_k$. L'objectif des méthodes de descente est de minimiser la fonctionnelle :

$$J(x) = \frac{1}{2}(Ax|x) - (b|x) = \frac{1}{2}x^T Ax - x^T b \tag{1}$$

où A est une matrice symétrique définit positive ($A^T = A$ et $(Ax|x) > 0 \vee x \neq 0$). Dans ce cas, J est aussi définit positive et est quadratique.

Trouver le minimum de la fonctionnelle J revient à trouver la solution du système $Ax = b$.
En effet, comme J est quadratique est positive, J est connexe et son minimum unique $\overline{x}$ est obtenu en annulant le gradient de J : $\nabla J(\overline{x}) = A\overline{x} - b = 0 \iff \overline{x}$ solution du système.

On note $r(x) = b - Ax = A(\overline{x} - x)$ le résidu du système et $e(x) = x - \overline{x}$ l'erreur ou la différence entre la solution calculée et la solution exacte. Minimiser J est équivalant à minimiser la fonctionnelle E définit par

$$E(x) = (A(x - \overline{x})|x - \overline{x}) = (Ae(x)|e(x)) \tag{2}$$

En effet, $E(x) = 2J(x) + \underbrace{(A\overline{x}|\overline{x})}_{cst>0}$.

Comme A est symétrique définit positive, $(Ax, y) = (x, Ay)$ définit un produit scalaire et $E(x) = (Ae(x)|e(x)) = \| e(x) \|_A{}^2$ où $\| e(x) \|_A$ est la norme associée à ce produit scalaire. Dans la suite, nous utiliserons la fonctionnelle E pour trouver la solution du système $Ax = b$.

0.2 Choix optimal de α_k pour une direction fixée d_k

On suppose que la direction d_k est fixée. L'objectif est de minimiser la fonctionnelle $E(x_k)$ à chaque itération. Or

$$\begin{aligned} E(x_{k+1}) &= E(x_k + \alpha_k d_k) = (A(x_k + \alpha_k d_k - \overline{x})|x_k + \alpha_k d_k - \overline{x}) \\ &= E(x_k) - 2\alpha_k(r_k|d_k) + \alpha_k{}^2(Ad_k|d_k) \end{aligned}$$

qui est un polynôme de degrés 2 en α_k et atteint son minimum en $\frac{-b}{2a} = \frac{(r_k|d_k)}{(Ad_k|d_k)}$.
Ainsi, peut importe la direction d_k choisie, le minimum de E sera atteint pour

$$\alpha_{opt} = \frac{(r_k|d_k)}{(Ad_k|d_k)} \tag{3}$$

.

De plus, on a

$$r_{k+1} = b - Ax_{k+1} = r_k - \alpha_k Ad_k \tag{4}$$

et le résidu obtenu à l'itération k est orthogonale à la direction d_k :

$$\begin{aligned} (r_{k+1}|d_k) &= (r_k - \alpha_k Ad_k|d_k) \\ &= (r_k|d_k) - \alpha_k(Ad_k|d_k) \\ &= (r_k|d_k) - \frac{(r_k|d_k)}{(Ad_k|d_k)}(Ad_k|d_k) \\ &= 0 \end{aligned} \tag{5}$$

.

0.3 Choix de la direction optimal pour α_{opt}

Pour ce α_{opt}, on a

$$\begin{aligned} E(x_{k+1}) &= E(x_k) - 2\alpha_k(r_k|d_k) + {\alpha_k}^2(Ad_k|d_k) \\ &= E(x_k) - 2\frac{(r_k|d_k)}{(Ad_k|d_k)}(r_k|d_k) + \frac{(r_k|d_k)^2}{(Ad_k|d_k)^2}(Ad_k|d_k) \\ &= E(x_k) - \frac{(r_k|d_k)^2}{(Ad_k|d_k)} \\ &= E(x_k)(1 - \frac{(r_k|d_k)^2}{E(x_k)(Ad_k|d_k)}) \end{aligned}$$

Or $E(x_k) = (Ae(x_k)|e(x_k))$ et $r_k = A(x_k - \overline{x}) = -Ae(x_k) \iff e(x_k) = -A^{-1}r_k$.
Ainsi $E(x_k) = (A^{-1}r_k|r_k)$.

Donc

$$E(x_{k+1}) = E(x_k)(1 - \frac{(r_k|d_k)^2}{(A^{-1}r_k|r_k)(Ad_k|d_k)}) \tag{6}$$

Mais $E(x)$ est une fonctionnelle minimisante qui s'approche le plus possible de la solution exacte $\overline{x}$, autrement dit on doit avoir $E(x_{k+1}) < E(x_k)$ et $\lim\limits_{k\to+\infty} E(x_k) = 0$.

On pose $(u_k)_{k\in\mathbb{N}}$ la suite définit par $u_k = 1 - \frac{(r_k|d_k)^2}{(A^{-1}r_k|r_k)(Ad_k|d_k)}$. On a

$$\begin{aligned} E(x_{k+1}) &< E(x_k)u_k \\ E(x_{k+1}) &< u_k^{k+1}E(x_0) \end{aligned}$$

Donc $E(x_k)$ définit une suite géométrique et converge si et seulement si à partir d'un certain rang k, $|u_k| = |1 - \frac{(r_k|d_k)^2}{(A^{-1}r_k|r_k)(Ad_k|d_k)}| < 1$. Autrement dit, s'il existe une constante $\mu \in]0,1[$ tel que $\frac{(r_k|d_k)^2}{(A^{-1}r_k|r_k)(Ad_k|d_k)} \geq \mu$.

Or $(Ad_k|d_k) \leq \lambda_1 {||d_k||_2}^2$ où λ_1 est la plus petite valeur propre de A. Et $(A^{-1}r_k|r_k) \leq \frac{1}{\lambda_n}{||r_k||_2}^2$ où λ_n est la plus grande valeur propre de A. Donc

$$\frac{(r_k|d_k)^2}{(A^{-1}r_k|r_k)(Ad_k|d_k)} \geq \frac{\lambda_1}{\lambda_n}\frac{(r_k|d_k)^2}{{||d_k||_2}^2{||r_k||_2}^2} = \underbrace{\frac{1}{cond_2(A)}}_{\in]0,1[}\underbrace{(\frac{r_k}{||r_k||_2}|\frac{d_k}{||d_k||_2})^2}_{\in[0,1]}$$

Remarquons que $E(x_k)_k$converge de plus en plus vite si u_k est proche de 0, c'est a dire si μ est le plus proche possible de 1. Ainsi, en choisissant une direction colinéaire au résidu r_k, on a

$$E(x_{k+1}) < \underbrace{(1 - \frac{1}{cond_2(A)})}_{\in]0,1[}^k E(x_0)$$

On a donc toujours une convergence de la suite $E(x_k)$ et le conditionnement de la matrice A a une influence sur cette vitesse de convergence. Plus le conditionnement est petit, plus la convergence est rapide.

On a vu que $\nabla J = -r$ donc le gradient est colinéaire au résidu (remarque $\nabla E = -2r$ aussi colinéaire au résidu).

Deuxième partie

Motivation du gradient conjugué

0.4 Analyse par méthode de projection

D'après le paragraphe d'avant, plus le conditionnement de la matrice est grand, plus la vitesse de convergence est faible. On souhaite donc maximiser $\frac{(r_k|d_k)^2}{(A^{-1}r_k|r_k)(Ad_k|d_k)}$ peu importe le conditionnement de la matrice A. L'idée est de faire une combinaison linéaire du résidu et de la direction précédente pour trouver la direction qui se rapproche le plus du centre (la solution du système). la nouvelle direction se trouvera dans le plan généré par d_{k-1}et r_k. On pose

$$d_k = r_k + \beta_k d_{k-1} \tag{7}$$

.

Comme $(r_{k+1}|d_k) = 0$, on a

$$\begin{aligned}\frac{(r_k|d_k)^2}{(A^{-1}r_k|r_k)(Ad_k|d_k)} &= \frac{(r_k|r_k + \beta_k d_{k-1})^2}{(A^{-1}r_k|r_k)(Ad_k|d_k)} \\ &= \frac{((r_k|r_k) + \beta_k(r_k|d_{k-1}))^2}{(A^{-1}r_k|r_k)(Ad_k|d_k)} \\ &= \frac{(r_k|r_k)^2}{(A^{-1}r_k|r_k)(Ad_k|d_k)}\end{aligned}$$

Ainsi maximiser $\frac{(r_k|d_k)^2}{(A^{-1}r_k|r_k)(Ad_k|d_k)}$ revient à minimiser $(Ad_k|d_k)$.

$$\begin{aligned}(Ad_k|d_k) &= (A(r_k + \beta_k d_{k-1})|r_k + \beta_k d_{k-1}) \\ &= {\beta_k}^2(Ad_{k-1}|d_{k-1}) + 2\beta_k(Ad_{k-1}|r_k) + (Ar_k|r_k)\end{aligned}$$

trinôme en β_k dont le coefficient sur le degrés principale est positif, le minimum est atteint en $\frac{-b}{2a} = -\frac{(Ad_{k-1}|r_k)}{(Ad_{k-1}|d_{k-1})}$.

Ainsi la direction maximisant la vitesse de convergence est obtenue pour

$$\beta_k = -\frac{(Ad_{k-1}|r_k)}{(Ad_{k-1}|d_{k-1})} \tag{8}$$

De plus, on a

$$\begin{aligned}(Ad_{k-1}|d_k) &= (Ad_{k-1}|r_k + \beta_k d_{k-1}) \\ &= (Ad_{k-1}|r_k) - \frac{(Ad_{k-1}|r_k)}{(Ad_{k-1}|d_{k-1})}(Ad_{k-1}|d_{k-1}) \\ &= 0\end{aligned} \tag{9}$$

On dit que les directions choisies à chaque itération sont A-conjugués.

Et

$$\begin{aligned}(r_k|d_k) &= (r_k|r_k + \beta_k d_{k-1}) \\ &= (r_k|r_k) + \beta_k \underbrace{(r_k|d_{k-1})}_{0} \\ &= (r_k|r_k)\end{aligned} \tag{10}$$

On peut récrire β_k en fonction du résidu :
on sait que $r_{k+1} = r_k - \alpha_k Ad_k$ donc $Ad_{k-1} = \frac{1}{\alpha_{k-1}}(r_{k-1} - r_k)$

et comme

$$\begin{aligned}(r_{k+1}|r_k) &= (r_k - \alpha_k Ad_k|r_k)\\ &= (r_k|r_k) - \alpha_k(Ad_k, r_k)\\ &= (r_k|r_k) - \alpha_k(Ad_k, d_k - \beta_k d_{k-1})\\ &= (r_k|r_k) - \alpha_k(Ad_k, d_k) + \alpha_k\beta_k \underbrace{(Ad_k|d_{k-1})}_{0}\\ &= (r_k|r_k) - \frac{(r_k|d_k)}{(Ad_k|d_k)}(Ad_k, d_k)\\ &= 0\end{aligned} \tag{11}$$

Donc

$$\begin{aligned}\beta_k &= -\frac{(\frac{1}{\alpha_{k-1}}(r_{k-1} - r_k)|r_k)}{(\frac{1}{\alpha_{k-1}}(r_{k-1} - r_k)|d_{k-1})}\\ &= -\frac{\overbrace{(r_{k-1}|r_k)}^{0} - (r_k|r_k)}{(r_{k-1}|d_{k-1}) - \underbrace{(r_k|d_{k-1})}_{0}}\\ &= \frac{(r_k|r_k)}{(r_{k-1}|d_{k-1})}\\ &= \frac{(r_k|r_k)}{(r_{k-1}|r_{k-1})}\end{aligned} \tag{12}$$

0.5 Algorithme pour la méthode du gradient conjugué

```
Fonction GradConj (Nmax, tol ,A,x,r0)
        d = r0
        normeR0 = (r0,r0)
        TANT QUE normeR0 < tol FAIRE
                Ad = A*d
                alpha = normeR0 / (Ad,d)
                x = x + alpha * d
                r1 = r0 - alpha * Ad
                normeR1 = (r1,r1)
                beta = normeR1 / normeR0
                d = r1 + beta * d
```

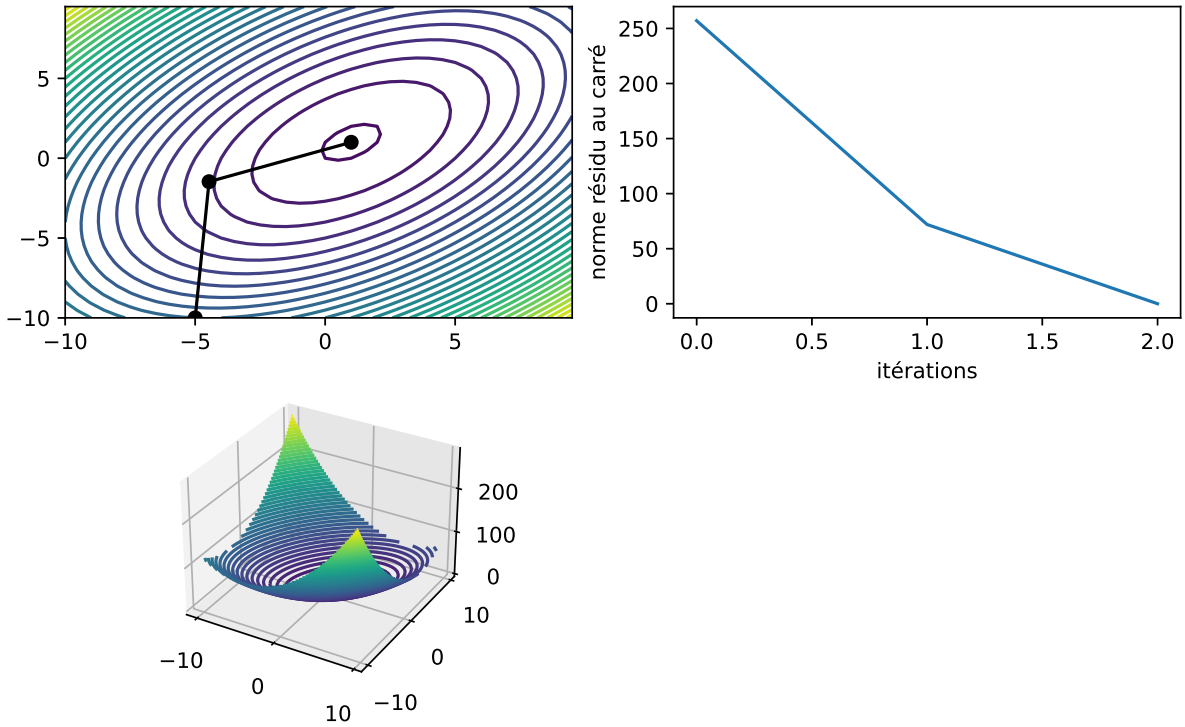
Troisième partie
Applications

0.6 Étude en dimension 2

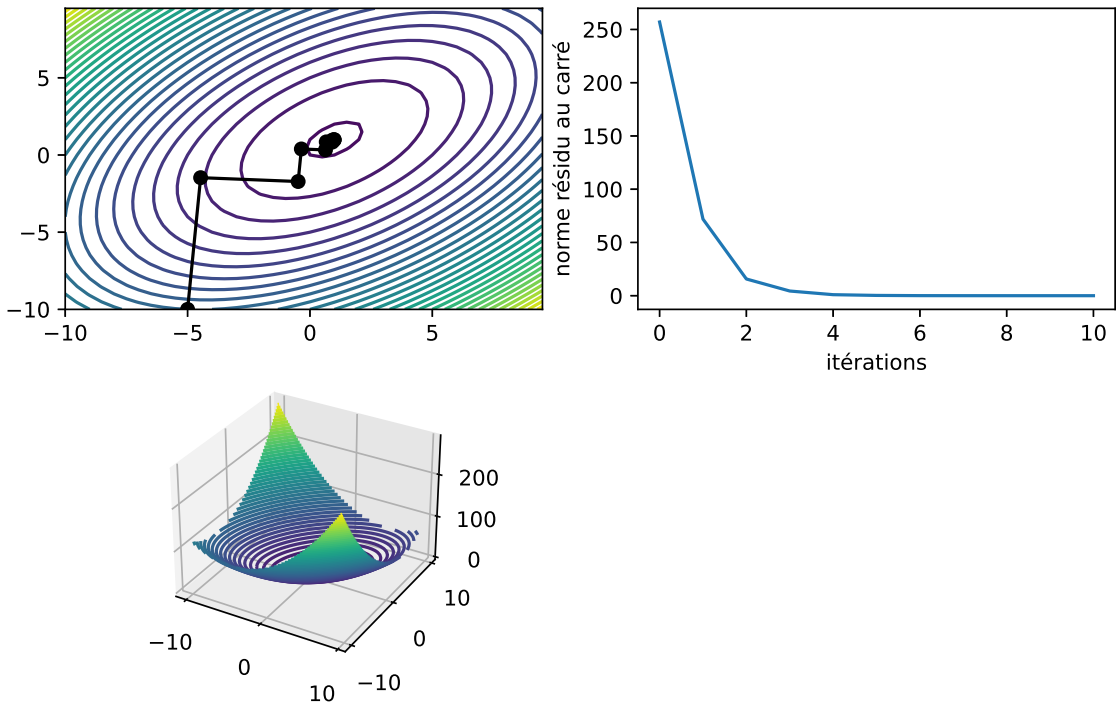
En dimension 2, la fonction bilinéaire symétrique définit par la matrice A est une paraboloïde et les lignes de niveau sont des ellipses. En dimension 2, l'algorithme du gradient conjugué doit converger en seulement 2 itérations peu importe le conditionnement de la matrice. A l'inverse, si le conditionnement de la matrice est mauvais, l'algorithme du gradient descente a une vitesse de convergence très mauvaise. On peut interpréter cette mauvaise convergence en observant des « sauts de puce » entre les courbes de niveau de la fonction. En effet, si la matrice est mal conditionnée, ces ellipses sont très allongées ce qui implique un déplacement très petit puisque la direction de descente est choisie orthogonale à la ligne de niveau.

Considérons la matrice $A = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}$ et $b = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$

Conditionnement de A = 2.99999



Algorithme du gradient conjugué appliqué à la matrice A pour $x_0 = (-5, -10)$

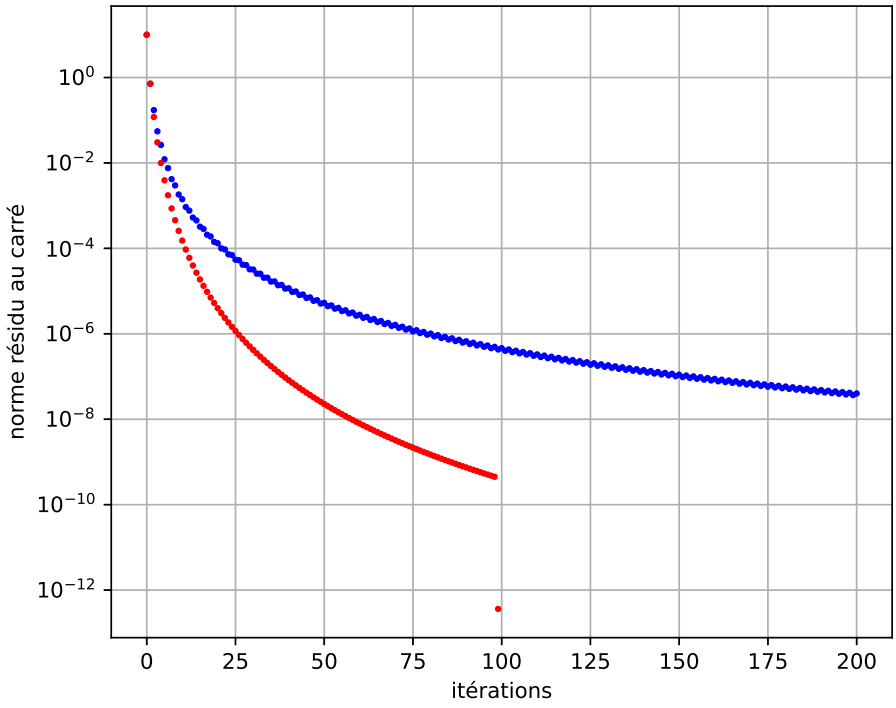


Algorithme du gradient descente appliqué à la matrice A pour $x_0 = (-5, -10)$

0.7 Application en dimension N

On se place maintenant dans le cas infini $A = \begin{pmatrix} 2 & -1 & 0 & \cdots & 0 \\ -1 & 2 & -1 & \ddots & \vdots \\ 0 & -1 & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & 2 & -1 \\ 0 & \cdots & 0 & -1 & 2 \end{pmatrix}$

Conditionnement de A = 16373



Comparaison des vitesses de convergence entre l'algorithme du gradient descente (en bleu) et du gradient conjugué (en rouge) pour N = 200

On remarque que l'algorithme du gradient conjugué converge beaucoup plus vite que l'algorithme de descente. De plus, on obtient une solution précise bien avant la dimension de la matrice. En moins de 100 itérations, on atteint déjà un résidu inférieur à 10^{-10} alors que le résidu du gradient descente stagne encore après 200 itérations.

Quatrième partie

Conclusion

L'algorithme du gradient conjugué fait partie des techniques les plus efficaces pour trouver la solution au système $Ax = b$. On peut encore améliorer cette convergence rapide avec un bon conditionnement de la matrice. Bien que cette méthode soit itérative, elle peut être considérée comme méthode directe puisqu'elle fournit la solution exacte en au plus n itérations (voir moins).

Références

[1] Luca Amodei and Jean-Pierre Dedieu. *Analyse numérique matricielle : cours et exercices corrigés*. 2008.

[2] Patrick Lascaux and Raymond Théodor. *Analyse numérique matricielle appliquée à l'art de l'ingénieur*. 2004.

[3] Frédéric Magoulès and François-Xavier Roux. *Calcul scientifique parallèle : Cours, exercices corrigés, exemples avec MPI et openMP*. October 2017.

[4] Marie-Hélène Meurisse. *Algorithmes numériques : fondements théoriques et analyse pratique : Cours, exercices et applications avec MATLAB*. January 2018.