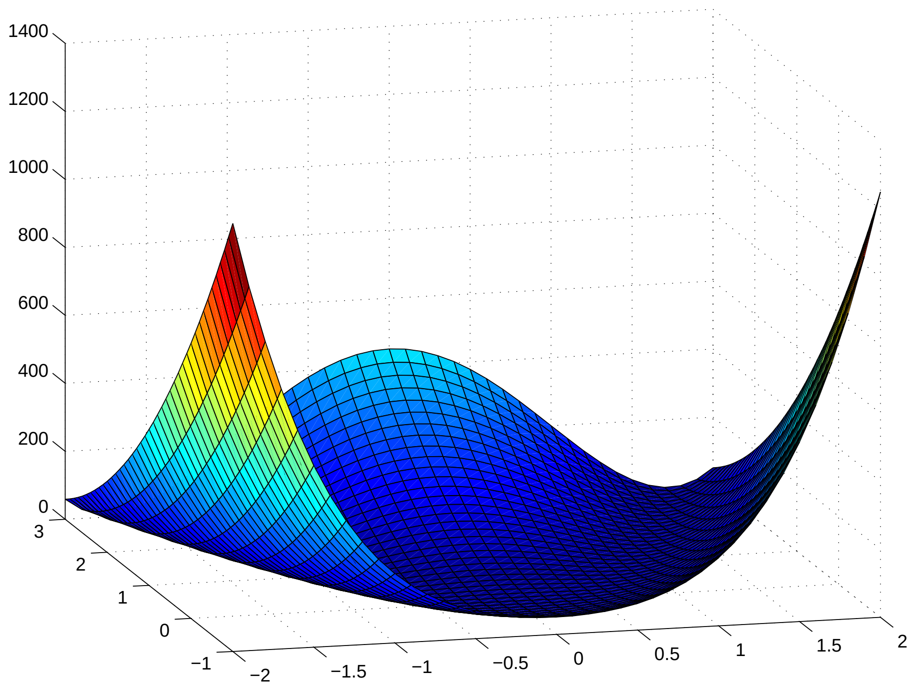
Optimisation non linéaire



Céline CATTOËN
Ranveig WALLACE

décembre 2002

Introduction

Pour tester différentes méthodes d'optimisation sans contrainte, on cherche le minimum de la fonction de Rosenbrock :

$$f(x) = 50(x_1^2 - x_2)^2 + (x_1 - 1)^2$$

Cette fonction est particulière du fait de sa courbure en coude autour de l'origine. Elle est très utilisée dans les exemples d'optimisation car la plupart des méthodes donnent une convergence assez lente pour résoudre ce problème. Le graphe de la fonction de Rosenbrock tracé avec l'option contour de Matlab montre, grâce aux lignes de niveaux, que le minimum de cette fonction se situe dans un voisinage du point (1,1).

Le point (1,1) est le minimum de la fonction de Rosenbrock, on peut vérifier que le gradient de f est nul en ce point avec :

$$\nabla f(x) = \begin{pmatrix} 200(x_1^2 - x_2)x_1 + 2(x_1 - 1) \\ -100(x_1^2 - x_2) \end{pmatrix}, \quad \nabla^2 f(x) = \begin{pmatrix} 600x_1^2 - 200x_2 + 2 & -200x_1 \\ -200x_1 & 100 \end{pmatrix}$$

et que la matrice hessienne de f definie ci-dessus est symétrique définie positive en (1,1).

On peut visualiser la courbe de Rosenbrock sur la figure de la page de garde et ainsi constater la particularité de ses vallées.

Comparaison des différentes méthodes :

On a utilisé plusieurs méthodes pour trouver le minimum de la fonction de Rosenbrock. Différents choix du point de départ ont permis de comparer la convergence et le temps de calcul de ces méthodes.

$x0 = [-1.2, 1]$	convergence	itérations	temps par itération	temps total
Plus forte pente	oui	2755	0.0017	4.7300
Fletcher-Reeves	oui	99	0.0094	0.9300
Polak-Ribière	oui	744	0.0189	14.0600
Newton sans r.l.	oui	6	0.0100	0.0600
Newton avec r.l.	oui	17	0.0029	0.0500
DFP	oui	52/1061	0.0021	0.1100
BFGS	oui	29	0.0017	0.0500

$x0 = [1.5, 0]$	convergence	itérations	temps par itération	temps total
Plus forte pente	oui	2913	0.0027	7.8500
Fletcher-Reeves	oui	153	0.0086	1.3200
Polak-Ribière	oui	1101	0.0203	22.3200
Newton sans r.l.	oui	4	0.0150	0.0600
Newton avec r.l.	oui	9	0.0067	0.0600
DFP	oui	6384/423	0.0028	1.2000
BFGS	oui	33	0.0033	0.1100

Remarques générales :

Pour les méthodes quasi-Newton (BFGS et DFP), il est préférable d'effectuer la recherche linéaire pour minimiser $f(x + td)$ dans la direction d en partant de $t = 1$. Ceci est dû au fait que ces méthodes approximent une descente 'normalisée' en quelque sorte. Les autres méthodes fournissent des directions de descente, mais la longueur n'est pas la bonne, donc $t = 1$ n'est pas forcément la meilleure initialisation.

Pour la méthode de Newton, faire une recherche linéaire (qui s'applique à la fonction elle-même) est incohérent avec l'esprit de la méthode qui travaille uniquement sur le gradient. *A posteriori* la recherche linéaire peut être utile pour éviter d'être en train de converger vers un maximum.

Plus forte pente :

Cette méthode est convergente et ne nécessite pas le stockage de beaucoup de vecteurs ou matrices (donc chaque itération est rapide), mais la convergence peut être très lente, surtout pour une fonction comme Rosenbrock qui a une vallée très allongée. On voit sur le graphe comment la méthode 'zigzague' dans la vallée en s'approchant de la solution.

Fletcher-Reeves et Polak-Ribière :

Ces méthodes sont des méthodes non linéaires du gradient conjugué qui ne nécessitent pas de mémoriser des matrices et sont en général plus rapides que la méthode de plus forte pente.

Les directions de descente de la méthode de Fletcher-Reeves (FR) et de Polak-Ribiere (PR) sont de la forme $d_{k+1} = -\nabla f(k+1) + \beta_{k+1} d_k$ avec $\beta_{k+1} = \nabla f_{k+1}^T \nabla f_{k+1} / \nabla f_k^T \nabla f_k$ pour FR et $\beta_{k+1} = \nabla f_{k+1}^T (\nabla f_{k+1} - \nabla f_k) / \nabla f_k^T \nabla f_k$ pour PR. Cette dernière est censée être une amélioration de FR, mais on a constaté ici qu'au moins pour la fonction de Rosenbrock, PR converge beaucoup moins vite que FR. D'après [?], il n'y a pas de garantie de convergence pour PR si on n'impose pas de contraintes plus strictes sur la recherche linéaire et sur β. En effet, on a pu observer que PR ne fournit pas toujours une direction de descente.

On a vu que FR a un peu de mal à changer de direction d'une itération à la suivante. Puisqu'à chaque itération on prend en compte le gradient du point précédent, la méthode peut d'abord s'approcher de la solution pour passer à côté et mettre plusieurs itérations à retourner vers la solution. En partant du point $(-1.2, 1)$, on voit clairement que FR fait deux ou trois passages dans la vallée à côté de la solution avant de converger.

PR n'a pas ce problème, et on voit qu'elle s'approche beaucoup plus vite de la solution. Par contre, à proximité de la solution elle fait beaucoup de petits allers-retours autour de la solution avant de converger. Le début montre donc que RB s'approche plus vite que FR, mais la fin peut être une indication de la non-convergence de la méthode.

BFGS et DFP :

En général, ces 2 méthodes convergent plus rapidement vers le point optimal que les méthodes non-linéaires du gradient conjugué car elles consistent en une approximation non seulement du gradient mais aussi de la Hessienne.

DFP travaille sur la norme de la différence entre la Hessienne et l'approximation de la Hessienne, alors que BFGS travaille sur la norme de la différence entre l'inverse de la Hessienne et l'approximation de l'inverse. Puisqu'on s'intéresse à l'inverse de Hessienne, BFGS sera la meilleure méthode.

La méthode BFGS a une propriété d'autocorrection : en effet lorsque l'on effectue une recherche linéaire adéquate et que la matrice H_k approxime mal la hessienne, cette approximation tend à s'autocorriger en quelques itérations.

Cependant, la méthode DFP ne possède pas cette propriété, ce qui explique en partie sa moins bonne performance. La différence entre les deux méthodes est flagrante pour le point de départ $(1.5, 0)$, pour lequel BFGS met 33 itérations à converger contre 423 itérations pour DFP. Sur le graphe on peut observer que BFGS s'approche assez vite et directement vers la solution, alors que DFP fait souvent des pas très petits et s'approche à la solution par bonds irréguliers.

Dans le tableau, deux nombres d'itérations sont indiqués pour DFP. Le premier est en posant $t = min(1, 1.01t)$ pour initialiser la recherche linéaire, le deuxième avec $t = 1$. Pour choisir entre les deux, il faut clairement faire une étude plus approfondie.

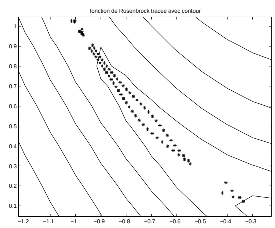
Newton :

Ces méthodes convergent en peu d'itérations mais nécessitent d'inverser des matrices, ce qui peut être coûteux en calculs. Cependant, ici on est en dimension 2, donc la matrice n'est pas très difficile à inverser. En effet, on voit que les méthodes de Newton sont les plus rapides au total, mais chaque itération prend plus long temps à cause de l'inversion de matrices.

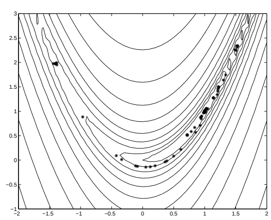
La méthode sans recherche linéaire fait toujours moins d'itérations que la méthode avec, mais elle a un inconvénient : on est sûr de trouver un extremum, mais pas forcément un minimum. La recherche linéaire assure une direction de descente donc on trouve forcément un minimum. Dans ce cas, Newton sans recherche linéaire trouve un minimum quel que soit le point de départ, mais seulement parce que la fonction de Rosenbrock n'a pas de maximum. La méthode sans recherche linéaire est donc la plus rapide, mais il faut s'assurer que l'on a bien trouvé un minimum.

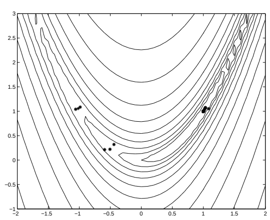
Newton sans recherche linéaire est plus rapide que les autres méthodes parce que le nouveau point trouvé est le meilleur point de départ pour l'itération suivante.

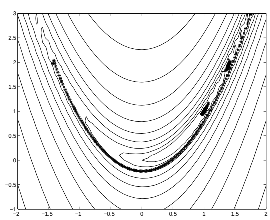
Sur les figures suivantes on peut observer le comportement de quelques méthodes :

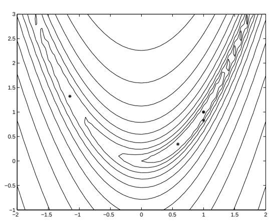


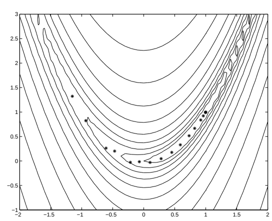
p.f. pente, $x_0 = [-1.2, 1]$

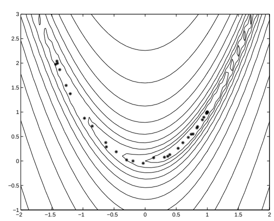

FR, $x_0 = [1.5, 1]$


PR, $x_0 = [-1.2, 1]$


DFP, $x_0 = [1.5, 1]$


Newton s.rl, $x_0 = [-1.2, 1]$


Newton a.rl, $x_0 = [-1.2, 1]$


BFGS, $x_0 = [1.5, 1]$

Références

[1] J. Nocedal, S.J. Wright, *Numerical Optimization*, Springer, 1999