Comparaison de différentes méthodes d'optimisation sans contrainte

1 Banane de Rosenbrock

L'objectif de ce TP est de comparer différentes méthodes d'optimisation, afin de bien comprendre les différences fondamentales entre les méthodes de relaxation, de descente de gradient à pas constant, et de descente de gradient à pas optimal.

A cette fin, nous allons considérer une fonction classique en optimisation, appelée souvent "banane de Rosenbrock" :

$$J(x, y) = (x - 1)^2 + 10(x^2 - y)^2 \tag{1}$$

Cette fonction est-elle convexe ? En quel point J est elle minimum ?

Calculer $\nabla J(x, y)$ puis $\nabla^2 J(x, y)$.

Visualiser cette fonction en 3D avec matlab. On utilisera pour cela les fonctions *meshgrid* et *surf*.

Visualiser ensuite en 2D cette fonction, i.e. comme une image (le niveau de gris d'un pixel représente la valeur de la fonction). On pourra utiliser la fonction *imagesc* (et *colormap(gray)*).

Commentaires ? Expliquer pourquoi la fonction J n'est pas facile à minimiser.

Remarque : Dans la suite du TP vous pourrez toujours étudier la vitesse de convergence des algorithmes utilisés car, dans le cas de cette fonction, on sait où se trouve le minimum. Si x_* est le point optimum et que x_n est la suite des approximations trouvées par la méthode, étudier la vitesse de convergence revient à étudier la vitesse de décroissance de $\|x_n - x_*\|$.

2 Préliminaires : optimisation en dimension 1

Comme rappelé en cours, la dimension 1 est un cas spécial pour l'optimisation (essentiellement dû au fait que $\mathbb{R}$ est un ensemble ordonné). De nombreux algorithmes d'optimisation (comme le gradient à pas optimal) utilisent une méthode d'optimisation en dimension 1.

Ecrire une fonction qui prends en entrée un point x et un vecteur d, et qui calcule le minimum de la fonction $t \mapsto J(x + dt)$ en utilisant la méthode de Newton. Essayer votre méthode en prenant $(-1, 1)$ comme position de départ et $(1, 0)$ comme direction. Commentaires ?

3 Minimisation de J

3.1 Méthode de relaxation

Implémenter la méthode de relaxation. L'utiliser pour la fonction J, en partant du point $(-1, 1)$. Afficher la trajectoire des points calculés successivement par la méthode.

3.2 Méthode du gradient à pas constant

Implémenter la méthode du gradient à pas constant. L'utiliser pour la fonction J, en partant du point $(-1, 1)$. Afficher la trajectoire des points calculés successivement par la méthode. On étudiera en particulier la plage de pas qui font converger l'algorithme. On justifiera cette plage par l'étude de la différentielle seconde de la fonction J à son point optimum.

3.3 Méthode du gradient à pas optimal

Implémenter la méthode du gradient à pas optimal. L'utiliser pour la fonction J, en partant du point $(-1, 1)$. Afficher la trajectoire des points calculés successivement par la méthode. Expliquer précisément pourquoi cette trajectoire ressemble à la méthode par relaxation.

3.4 Méthode de Newton

Implémenter la méthodes de Newton. L'utiliser pour la fonction J, en partant du point $(-1, 1)$. Afficher la trajectoire des points calculés successivement par la méthode.

3.5 Méthode du gradient conjugué

Implémenter la méthodes du gradient conjugué (version Polak-Ribière). L'utiliser pour la fonction J, en partant du point $(-1, 1)$. Afficher la trajectoire des points calculés successivement par la méthode.

Rappel de la méthode de gradient conjugué, variante Polak-Ribière :

Initialisation : Poser $d_{-1} = [0, 0]^t$ et $\nabla F(u_{-1}) = [1, 0]^t$, et fixer $u_0 \in \mathbb{R}^n$ et commencer pour $k = 0$ ce qui suit

1) Vérifier le critère d'arrêt (par exemple module du gradient, ou nombre de pas)
2) Poser $\xi_k = \frac{<\nabla F(u_k) - \nabla F(u_{k-1}), \nabla F(u_k)>}{\|\nabla F(u_{k-1})\|^2}$
3) Poser $d_k = \nabla F(u_k) + \xi_k d_{k-1}$
4) Si $-d_k$ n'est pas une direction de descente (i.e. $< \nabla F(u_k), d_k >< 0$) remplacer d_k par $d_k = \nabla F(u_k)$.
5) Trouver l'optimum dans la direction $-d_k$ pour trouver un coefficient $\rho_k > 0$
6) $u_{k+1} = u_k - \rho_k d_k$ et retourner en 1.

3.6 Comparaisons

Comparer les différentes trajectoires obtenues. Commentaires (nombres d'itérations, temps de calcul, . . .).

4 (Pré)Conditionnement

Dans cette partie on s'intéresse au problème du conditionnement. On montre comment un mauvais conditionnement réduit l'efficacité d'un algorithme d'optimisation. On montre aussi, comment supprimer le problème par le préconditionnement. Cette section est aussi l'occasion de manipuler la différenciation d'opérateurs sur les images.

On se donne comme problème la reconstruction d'une image par la connaissance de son champ de gradient. Si A est un opérateur linéaire de $\mathbb{R}^N$ dans $\mathbb{R}^M$, calculer le gradient de la fonctionnelle

$$F(x) = \|Bx - b\|^2$$

où $b \in \mathbb{R}^M$ est l'observation. Vous ferez ce calcul en utilisant l'opérateur conjugué de A, que l'on note B^T.

Considérer l'opérateur noté D qui à une image $(x_{i,j}), i = 1 \dots N, j = 1 \dots N$ associe un champ de gradient $(Dx_{i,j}), i = 1 \dots N, j = 1 \dots N$ de manière suivante :

$$Dx_{i,j} = (x_{i,j} - x_{i-1,j} \quad , \quad x_{i,j} - x_{i,j-1})$$

avec la convention que $x_{0,j} = x_{1,j}$ et $x_{i,0} = x_{i,1}$ (c'est-à-dire que l'on imagine une colonne et une ligne supplémentaires de l'image qui sont des copies de la première ligne et colonne de l'image).

Expliciter l'opérateur conjugué de D (qui transforme un champ de vecteurs) en une image ? L'opérateur que vous avez défini est appelé (-) divergence (noté *div*). (pour vous aider imaginez cet opérateur dans le cas à une dimension).

Le but du préconditionnement est de changer la variable x en $x = Py$ de sorte que la matrice définissant la fonctionnelle quadratique (dans notre cas) passe de $B^T B$ à $P^T B^T BP$ et que cette dernière soit proche de l'identité. Dans l'exemple que l'on va voir on approxime la matrice B par sa version correspondant à la périodisation du signal (cela revient à prendre la convention $x_{0,j} = x_{N,j}$). L'inverse de $B^T B$ devient facile à calculer dans le domaine de Fourier.

Télécharger les programmes dans `http://perso.telecom-paristech.fr/~ladjal/TPMVA/TP0` et en analyser les résultats en termes d'accélération de la convergence. (un fichier contient la définition des opérateurs différentiels sur les images, et l'autre contient une comparaison entre une optimisation avec et sans preconditionnement).