Master M2 Optimization

by Grégoire Allaire, Antonin Chambolle, Thomas Wick

Basic course of the Master Program "Mathematical Modelling"
Ecole Polytechnique, ENPC and Pierre et Marie Curie University
Academic year 2016-2017

Exercise meeting 1, Sep 8, 2016

in Amphi Curie

Online version on:

`http://www.cmap.polytechnique.fr/~wick/m2_fall_2016_engl.html`

Exercise 1

1. Detect and classify (with a short justification) the stationary point of $f : \mathbb{R}^2 \to \mathbb{R}$ with $f(x_1, x_2) = x_1^2 - x_2^2$.

Answer of exercise 1

The stationary point $(0, 0)$ (check yourself why this is the only stationary point and how to obtain it) is a saddle point.

- Justification 1: Following x_1 we identify a minimum at $(0, 0)$ and following x_2 we identify a maximum at $(0, 0)$. Consequently, the curvatures are different and thus there is a saddle point.
- Justification 2: Compute the Hessian:

$$\nabla^2 f(0,0) = \begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix}$$

  The Hessian has two different eigenvalues and therefore, a saddle point can be classified. As alternative, one can compute the determinant: since $\det \nabla^2 f(0, 0) = -4 < 0$, a saddle point is obtained.

Exercise 2

Consider the gradient descent method.

1. Show with the help of an example that if the step length α is badly chosen, the gradient descent might diverge. [**Hint:** The iteration for gradient descent is given by $x_{k+1} = x_k - \alpha \nabla f(x_k)$. Consider the 1D case with a quadratic function for instance.]

Answer of exercise 2

Choose for example $f(x) = \frac{x^2}{2}$. Then $\nabla f(x) = f'(x) = x$. The general iteration reads:

$$x_{k+1} = x_k + \alpha_k p_k$$

where $p_k = -\nabla f(x_k)$. Here, $p_k = -x_k$. This yields:

$$x_{k+1} = x_k - \alpha_k x_k = x_k(1 - \alpha_k).$$

If $(1 - \alpha_k) \geq 1$, gradient descent does not converge. This is intuitively clear because the minimum is 0 and the sequence $\{x_k\}_k$ will produce increasing x_k with

$$x_0 < x_1 < x_2 < \dots$$

if $(1 - \alpha_k) > 1$ or just stagnates

$$x_0 = x_1 = x_2 = \dots$$

for $(1 - \alpha_k) = 1$.

Exercise 3

Consider the conjugate gradient method. Let $A \in R^{n\times n}$ be a symmetric, positive definite matrix.

1. If the set of vectors $\{p_0, \dots, p_k\}$ is A-conjugate, then these vectors are linearly independant.

Answer of exercise 3

Show that from $\sum_{i=0}^{k} \lambda_i p_i = 0$ it follows $\lambda_i = 0$ for all i. Proof:

$$\begin{aligned} 0 &= \sum_{i=0}^{k} \lambda_i p_i \\ &= p_j^T A(\sum_{i=0}^{k} \lambda_i p_i) \\ &= \sum_{i=0}^{k} p_j^T A p_i \\ &= \lambda_j p_j^T A p_j \end{aligned}$$

where $p_j^T A p_j > 0$ since $p_j \neq 0$ and $A \neq 0$. Therefore it follows: $\lambda_j = 0$ for all j.

Exercise 4

We consider now the Armijo step rule. Let U be an open set and $f : U \subset R^n \to R$ a C^1-function on U. Furthermore, let $\gamma \in (0, 1)$.

1. If $x \in U$ and $p \in R^n$ is a descent direction of f in x, then there exists a $\bar{\alpha} > 0$ such that

   $$f(x + \alpha p) - f(x) \leq \alpha\gamma\nabla f(x)^T p, \quad \forall\alpha \in [0, \bar{\alpha}].$$

   Thus, the Armijo rule is feasible.

Answer of exercise 4

The case $\alpha = 0$ is obviously satisfied. Let $\alpha > 0$ be sufficiently small. We first remark that $x + \alpha p \in U$. Now we calculate:

$$
\begin{aligned}
&\frac{f(x+\alpha p) - f(x)}{\alpha} - \gamma \nabla f(x)^T p \\
\rightarrow &\nabla f(x)^T p - \gamma \nabla f(x)^T p \\
= &(1-\gamma) \nabla f(x)^T p
\end{aligned}
$$

where $(1-\gamma) > 0$ and $\nabla f(x)^T p < 0$. Thus:

$$\frac{f(x+\alpha p) - f(x)}{\alpha} - \gamma \nabla f(x)^T p < 0.$$

Specifically we can now choose an $\bar{\alpha} > 0$ sufficiently small that for all $\alpha \in [0, \bar{\alpha}]$, it holds:

$$f(x+\alpha p) - f(x) \leq \alpha \gamma \nabla f(x)^T p.$$

Exercise 5

In this exercise we investigate that Newton is really a *local* method and does not converge from arbitrary starting points $x_0 \in R$. Let $f : R \rightarrow R$ with $f(x) = \sqrt{x^2+1}$.

1. Investigate for which starting points x_0 Newton does converge and for which starting points Newton will diverge.
2. What are possible procedures to globalize Newton's method and to increase the convergence radius?

Answer of exercise 5

- Answer to 1: We have:

  $$\nabla f(x) = \frac{x}{\sqrt{x^2+1}}, \quad \nabla^2 f(x) = \frac{1}{(\sqrt{x^2+1})^{3/2}}$$

  The Newton iteration reads:

  $$\nabla^2 f(x_k) p_k = -\nabla f(x_k)$$

  Inserting the specific expressions:

  $$\frac{1}{(\sqrt{x_k^2+1})^{3/2}} p_k = -\frac{x_k}{\sqrt{x_k^2+1}}$$

  Thus for the update we have:

  $$p_k = -x_k((x_k)^2+1).$$

The second step of Newton's method is to form the new iterate:

$$x_{k+1} = x_k + p_k = x_k - x_k((x_k)^2 + 1) = -(x_k)^3$$

For $|x_0| < 1$ we have (very) fast convergence of Newton's method. For your own studies, you might verify the rate of convergence for this example (normally in the literature Newton's method is considered to be quadratically convergent). In this example, however, Newton converges even with rate 3!! For $|x_0| = 1$ we have an oscillating behavior between $-x_0$ and x_0. For $|x_0| > 1$ we observe divergence. All these results are also intuitively clear since the minimum of $f(x)$ is $(0, 1)$. Thus, the sequence $\{x_k\}_k$ must converge to 0 and this is obviously not the case for $|x_0| \geq 1$.

- Answer to 2 (one possibility out of many - but extremely naive for this example). A classical procedure to increase Newton's convergence radius (i.e., to globalize Newton's method) is to add a line search parameter α_k. Thus the iteration reads:

$$x_{k+1} = x_k + \alpha_k p_k = x_k + \alpha_k(-x_k(x_k^2 +)) = (1 - \alpha_k)x_k - \alpha_k(x_k)^3.$$

Choose for example $\alpha_k = 0.5$. Then:

$$x_{k+1} = 0.5x_k - 0.5(x_k)^3.$$

Choose for instance now as $x_0 = 1.5$. Then we obtain for the first iterate:

$$x_1 = -0.93750.$$

Compute for some further steps for the second and so forth. Here we see that indeed $x_0 = 1.5$ works. Consequently, we see that in comparison to the standard Newton method with the 'natural' step length $\alpha_k = 1$, we could increase the convergence radius from $|x_0| < 1$ to (at least) $|x_0| = 1.5$.

S'il vous plait, tourner la page pour la version francaise.

Exercise 1

1. Détecter et classer (avec une justification) le point stationnaire de la fonction $f : \mathbb{R}^2 \to \mathbb{R}$ avec $f(x_1, x_2) = x_1^2 - x_2^2$.

Exercise 2

Considérer la méthode de la descente du gradient.

1. En utilisant un exemple, montrer si la longueur de pas n'est pas de bonne choix que la méthode s'écarte. [**Hint:** L'itération de la descente du gradient est donnée par $x_{k+1} = x_k - \alpha \nabla f(x_k)$. Considérer le cas 1D avec une fonction quadratique par exemple.]

Exercise 3

Considérer la méthode du gradient conjugué. Soit $A \in R^{n \times n}$ une matrice symétrique, définie positive.

1. Si l'ensemble des vecteurs $\{p_0, \ldots, p_k\}$ est A-conjugué, ils sont linéairement indépendants.

Exercise 4

Nous travaillons maintenant avec la règle d'Armijo. Soit U un ensemble ouvert et $f : U \subset R^n \to R$ une fonction C^1 sur U. En plus, soit $\gamma \in (0, 1)$.

1. Si $x \in U$ et $p \in R^n$ est une diréction descente de f dans x, il existe $\bar{\alpha} > 0$ tel que
$$f(x + \alpha p) - f(x) \leq \alpha \gamma \nabla f(x)^T p, \quad \forall \alpha \in [0, \bar{\alpha}].$$
En conclusion, la règle d'Armijo est faisable.

Exercise 5

Dans cet exercise nous enquêtons que la méthode de Newton est vraiment une méthode locale et Newton ne converge pas des points $x_0 \in R$ arbitraire. Soit $f : R \to R$ avec $f(x) = \sqrt{x^2 + 1}$.

1. Découvrir pour quels points x_0 la méthode de Newton ne converge pas.
2. Est-ce qu'il y a des procécures (lesquelles ?) pour globaliser la méthode de Newton pour augmenter le rayon de convergence ?