INSTITUT FÜR MATHEMATIK UND
WISSENSCHAFTLICHES RECHNEN

Optimierung I

Problem Sheet 10

Due date: 24. June 2014

Problem 10.1: [Finding a descent direction with Newton's method]

We consider a twice-differentiable function $f : \mathbb{R}^n \to \mathbb{R}$, with a continous second-order derivative. At a given point $x \in \mathbb{R}^n$, if $\nabla^2 f(x)$ is invertible, Newton's method provides a direction $d \in \mathbb{R}^n$ which is such that:

$$\nabla f(x) + \nabla^2 f(x) d = 0. \tag{1}$$

We assume that $\nabla f(x) \neq 0$.

i) Give an example where (1) has a solution which is not a descent direction.

ii) Let us assume that $\nabla^2 f(x)$ is positive definite. Prove that d is a descent direction.

iii) Now, we do not make any assumption on the positivity of $\nabla^2 f(x)$. For all C, we denote by d_C, if it exists, the solution to:

$$\nabla f(x) + (\nabla^2 f(x) + C\,\mathrm{Id}_n) d = 0. \tag{2}$$

Let ρ_- be the smallest eigenvalue of $\nabla^2 f(x)$, prove that for all C with $C > -\rho_-$, (2) has a solution which is a descent direction.

iv) Prove that

$$d_C \underset{C\to\infty}{\longrightarrow} 0 \quad \text{and} \quad \frac{d_C}{\|d_C\|} \underset{C\to\infty}{\longrightarrow} -\frac{\nabla f(x)}{\|\nabla f(x)\|}. \tag{3}$$

v) Consider the function $f(x_1, x_2) = x_1^2 - \frac{1}{2}x_2^2 + x_1 + x_2$ and the point $x = (0,0)^\top$. Compute the solution to (1), prove that it is not a descent direction, compute the solution to (2) for any $C > -\rho_-$.

Problem 10.2: [Analysis of the order of convergence]

In this exercise, we analyse the convergence of Newton's method for a particular case (in dimension 1). Let f be defined by:

$$f(x) = -e^{-x^2}, \tag{4}$$

let us define:

$$d(x) = \frac{-f'(x)}{f''(x)}. \tag{5}$$

For a given $x_0 \in \mathbb{R}$, we define the sequence $(x_k)_k$ as follows: for all k,

$$x_{k+1} = x_k + d(x_k). \tag{6}$$

Note that $d(x)$ is not defined neither at $x = -\frac{1}{\sqrt{2}}$ nor at $x = -\frac{1}{\sqrt{2}}$. We say that the sequence $(x_k)_k$ is well-defined if for all k, $x_k \neq -\frac{1}{\sqrt{2}}$ and $x_k \neq \frac{1}{\sqrt{2}}$.

i) Draw a graph of f, draw the tangents at $x = \frac{1}{\sqrt{2}}$ and $x = -\frac{1}{\sqrt{2}}$.

ii) Compute $d(x)$, and motivate the definition of the sequence $(x_k)_k$.

iii) Let us define $g(x) = \left|\frac{x+d(x)}{x}\right|$, for $|x| \neq \frac{1}{\sqrt{2}}$. Study the variation of g, and prove that for all x, if $|x| \leq \frac{1}{2}$, then $g(x) < 1$, and if $|x| \geq \frac{1}{\sqrt{2}}$, $g(x) > 1$.

iv) Prove that if $x_0 > \frac{1}{\sqrt{2}}$, then $(x_k)_k$ is well-defined and $x_k \to +\infty$, and similarly, if $x_0 < -\frac{1}{\sqrt{2}}$, then $(x_k)_k$ is well-defined and $x_k \to -\infty$.

v) Prove that if $|x_0| < \frac{1}{2}$, the sequence converges to 0. Compute the order of convergence (german: *Q-Konvergenzordnung*) and the associated factor.

vi) Is the order of convergence the one that was expected from results of the lecture ?

Problem 10.3: [Cholesky decomposition]

For any positive definite symmetric matrix S in $\mathbb{R}^{n\times n}$, we call Cholesky decomposition the following decomposition:

$$S = LL^\top, \tag{7}$$

where L is a lower triangular matrix of $\mathbb{R}^{n\times n}$ with strictly positive diagonal coefficients. The goal of this exercise is to prove the existence and the uniqueness of such a matrix L. The proof is a proof by induction of the dimension n.

i) Consider the case $n = 1$.

ii) Let $n \in \mathbb{N}$, let us assume to have proved the result for this dimension. Let S be a positive definite symmetric matrix of size $n+1$, that we write as follows:

$$S = \begin{pmatrix} a & b^\top \\ b & \tilde{S} \end{pmatrix}, \tag{8}$$

where $a \in \mathbb{R}$, $b \in \mathbb{R}^n$, $\tilde{S} \in \mathbb{R}^{n\times n}$. Let L be a lower triangular matrix of size $n+1$, with

$$L = \begin{pmatrix} c & 0 \\ d & \tilde{L} \end{pmatrix}, \tag{9}$$

with $c \in \mathbb{R}$, $d \in \mathbb{R}^n$, $\tilde{L} \in \mathbb{R}^{n\times n}$ lower triangular. Prove that if $S = LL^\top$, then there is a unique possible choice for c and d, with $c > 0$.

iii) Let $\ell = \begin{pmatrix} c \\ d \end{pmatrix} \in \mathbb{R}^{n+1}$, compute the symmetric matrix $Q \in \mathbb{R}^{n\times n}$ which is such that:

$$S - \ell\ell^\top = \begin{pmatrix} 0 & 0 \\ 0 & Q \end{pmatrix}. \tag{10}$$

iv) Prove that Q is positive definite. To do so, you can prove that for all $x \in \mathbb{R}^n$, there exists $x_0 \in \mathbb{R}$ such that for $y = \begin{pmatrix} x_0 \\ x \end{pmatrix}$, $y^\top S y = x^\top Q x$. Conclude.

v) Using a Cholesky decomposition, solve the equation:

$$\begin{pmatrix} 10 & 5 & 4 \\ 5 & 25 & 20 \\ 4 & 20 & 26 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 8 \\ 30 \\ 64 \end{pmatrix}. \tag{11}$$

Problem 10.4: [An optimal triangle]

In this problem, we compute the maximal area of a triangle ABC having its three vertices A, B, and C on the same circle. Without loss of generality, we can assume that the three vertices belong to the circle of radius 1 and center $(0,0)$, and we can assume that $A = (0,1)$, $B = (\cos(\theta_1), \sin(\theta_1))$, $C = (\cos(\theta_2), \sin(\theta_2))$, with θ_1 and θ_2 in $[0, 2\pi]$.

Denoting by $\mathcal{A}$ the area of the triangle and by M the matrix of size 2 formed with the vectors $\vec{AB}$ and $\vec{AC}$, we recall the formula: $\mathcal{A} = \frac{1}{2}|\det(M)|$.
Find the optimal triangles.

Problem 10.5: [Regularization]
Suppose we want to recover a signal $\hat{u} \in \mathbb{R}^n$ from a given a noise signal $u_0 = \hat{u} + \delta$, where $\delta \in \mathbb{R}^n$ is some (Gaussian) noise (see Figure 1). If we know that our original signal was approximately piecewise constant, a one approach to recover $\hat{u}$ is to solve

$$\min_{u \in \mathbb{R}^n} F(u) \quad \text{with} \quad F(u) = \frac{1}{2}\|u - u_0\|_2^2 + \mu\|\nabla u\|_1 = \frac{1}{2}\sum_{i=1}^{n}(u_i - (u_0)_i)^2 + \mu \sum_{i=1}^{n} |(\nabla u)_i|$$

where

$$(\nabla u)_i = \begin{cases} u_{i+1} - u_i & \text{if } 1 \leq i < n \\ 0 & \text{else} \end{cases}$$

denotes a discret gradient. Now the difficulty is that $|\cdot|$ is not differentiable at 0. One approach is to solve an approximate, regularized problem:

i) Given a parameter $\epsilon > 0$, find an approximation $|\cdot|_\epsilon : \mathbb{R} \to \mathbb{R}_+$ of the absolute value $|\cdot|$ such that

- $|\cdot|_\epsilon$ is convex and continuously differentiable
- $\sup_{x \in \mathbb{R}} ||x|_\epsilon - |x|| \to 0$ as $\epsilon \to 0$
- There exists $R(\epsilon) > 0$ with $R(\epsilon) \to 0$ as $\epsilon \to 0$ and $|x|_\epsilon = |x|$ for all $|x| > R(\epsilon)$.

ii) Use $|\cdot|_\epsilon$ to define an approximation $F_\epsilon \in C^1(\mathbb{R}^n, \mathbb{R})$ of F such that $\|F_\epsilon - F\|_\infty \to 0$ as $\epsilon \to 0$ and calculate ∇F_ϵ

iii) Ensure existence of a unique solution to $\min_u F(u)$ and $\min_u F_\epsilon(u)$.

iv) Show that for any approximation as in 2), denoting u_ϵ the minimizer of F_ϵ, $u_\epsilon \to u^*$, u^* being the unique minimizer of F, in particular $F(u_\epsilon) \to F(u)$ as $\epsilon \to 0$.

v)* Use your program of exercise 8.5 to solve the approximate problem. An example of noisy and true data is provided on the webpage. Try different values of μ and ϵ and bring pictures of your results to class (on USB). You can also generate your own data.

NOTE: v)* is an additional point, it is not necessary to cross the exercise. After completing i)-iv), we will as if one students wants to show v)* and give 2 extra points for the presentation.

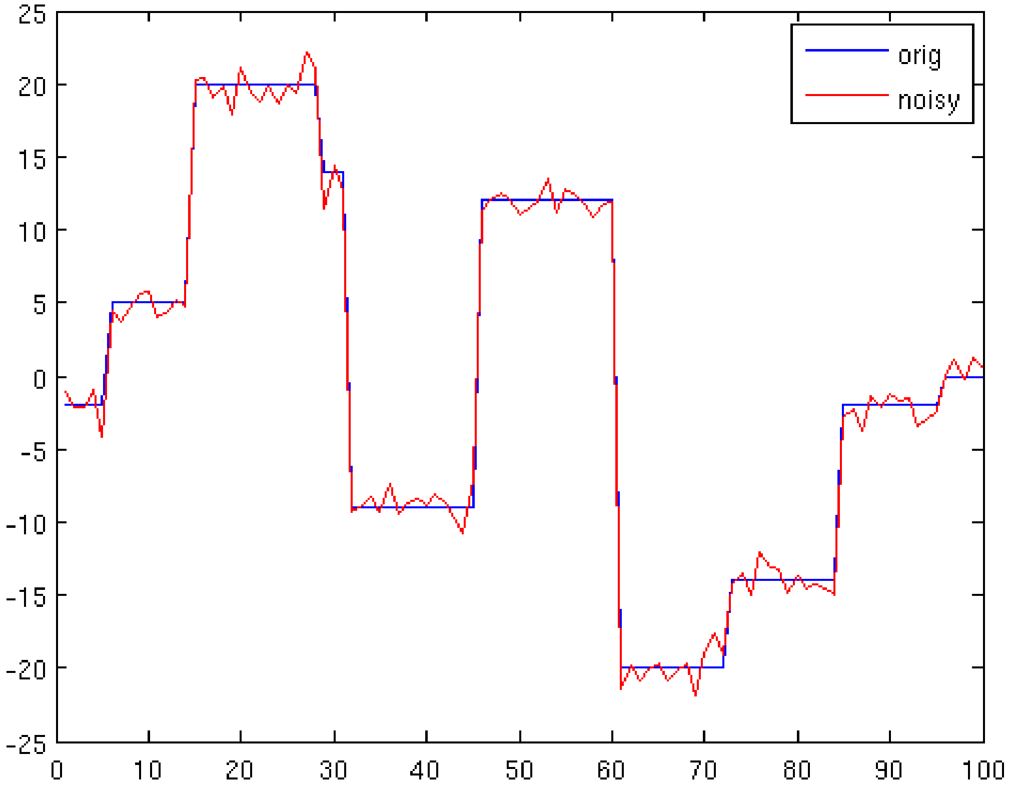

Abbildung 1: Example of original (blue) and noisy (red) data