INSTITUT FÜR MATHEMATIK UND
WISSENSCHAFTLICHES RECHNEN

Optimierung I

Problem Sheet 8

Due date: 3. June 2014

Problem 8.1: [Convexity]

i) Let K_1 and K_2 be two convex subsets of $\mathbb{R}^n$, let $\lambda \in \mathbb{R}$. Prove that the following subsets are convex:

   a) $K_1 \cap K_2$
   b) $K_1 + K_2$, defined by: $\{x + y \,|\, x \in K_1,\, y \in K_2\}$.
   c) λK_1, defined by: $\{\lambda x \,|\, x \in K_1\}$.

ii) Let $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}$ be two convex functions, prove that $\max(f, g)$ is also convex.

iii) Let $f : \mathbb{R}^n \to \mathbb{R}$. We define its *epigraph* as the following subset of $\mathbb{R}^{n+1}$:

$$\mathrm{Epi}(f) = \big\{(x, y) \in \mathbb{R}^n \times \mathbb{R} \,|\, y \geq f(x)\big\}. \tag{1}$$

Prove that f is convex if and only if $\mathrm{Epi}(f)$ is a convex subset of $\mathbb{R}^{n+1}$.

Problem 8.2: [Positive matrices of dimension 2]

i) Let S be a symmetric matrix of dimension 2. Prove that S is positive semi-definite if and only if:

$$\mathrm{tr}(S) \geq 0 \quad \text{and} \quad \det(S) \geq 0. \tag{2}$$

To this purpose, we recall the spectral theorem: for all symmetric matrix, there exist an invertible matrix B and a diagonal matrix D such that $S = BDB^{-1}$.

ii) Using the above characterization of positive semi-definite matrices, check if the following functions are convex on the open subset $(0, +\infty) \times (0, +\infty)$ of $\mathbb{R}^2$:

   a) $f(x_1, x_2) = x_1 x_2$
   b) $f(x_1, x_2) = \frac{1}{x_1 x_2}$
   c) $f(x_1, x_2) = \frac{x_1}{x_2}$
   d) $f(x_1, x_2) = \frac{x_1^2}{x_2}$.

iii) Let $A \in \mathbb{R}^{n \times n}$ be a positive definite symmetric matrix, $b \in \mathbb{R}^n$ and

$$f(x) = x^\top A x + b^\top x, \tag{3}$$

for $x \in \mathbb{R}^n$. With x^* denoting the global minimum of f, show that, if $x_0 - x^*$ is colinear to an eigenvector of A, then the steepest descent method starting at x_0 and using exact line search converges in one step.

Problem 8.3: [Stationary points]

i) Let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuous function satisfying the following property: for all sequence $(x_k)_k$,

$$\text{if } \|x_k\| \to +\infty, \text{ then } f(x_k) \to +\infty. \tag{4}$$

Prove the existence of (at least) one global minimizer.

ii) We consider the following function:

$$f(x) = x_1^4 + 2x_1^3 + 2x_1^2 - 2x_1x_2 + x_2^2. \tag{5}$$

Compute the stationary points. Compute the Hessian of f at these points, deduce which of them are local minimizers or local maximizers.

iii) Using the identity $f(x) = x_1^2(x_1+1)^2 + (x_1 - x_2)^2$, prove that property (4) is satisfied. Compute the global minimizer(s).

Problem 8.4: [Linear regression]

For given data points (x_1, y_1),...,(x_n, y_n) in $\mathbb{R}^2$, the linear regression problem problem consists in fitting the line $y = d + kx$ such that to minimize the residuals:

$$\min_{(d,k)\in\mathbb{R}^2} \sum_{i=1}^{n} (d + kx_i - y_i)^2. \tag{6}$$

We assume that there exist i and j such that $x_i \neq x_j$.

i) Rewrite the problem in the form of the following unconstrained problem:

$$\min_{z\in\mathbb{R}^2} \ (Az - b)^\top (Az - b), \tag{7}$$

where $A \in \mathbb{R}^{n\times 2}$ and $b \in \mathbb{R}^n$.

ii) Compute $A^\top A$, its determinant, and check that this last matrix is invertible.

iii) Compute the value of the unique stationary point (in function of (x_1, y_1),...,(x_n, y_n)).

Problem 8.5: [Programming exercise: gradient descent with Armijo and Wolfe rules]

We consider a gradient descent method in order to minimize a differentiable function $f : \mathbb{R}^n \to \mathbb{R}$. The parameters are the following: $x_0 \in \mathbb{R}^n$, σ_1 and $\sigma_2 \in (0, 1)$ such that $\sigma_1 < \sigma_2$, $\gamma > 1$, $\varepsilon > 0$, $R > 0$, $N \in \mathbb{N}$. The method generates the sequence $(x_k)_k$ starting at x_0 and defined by:

$$x_{k+1} = x_k + \alpha_k d_k, \tag{8}$$

where $d_k = -\nabla f(x_k)$. For a given $\alpha > 0$, we say that Armijo rule (resp. Wolfe rule) holds if

$$f(x_k + \alpha_k d_k) \leq f(x_k) - \sigma_1 \alpha_k \|d_k\|^2 \quad (\text{resp. } \nabla f(x_k + \alpha_k d_k) d_k \geq -\sigma_2 \|d_k\|^2). \tag{9}$$

We use a bisection method to find a step α_k satisfying these two rules. If we know that α_k belongs to an interval $[\beta_0, \beta_1]$, then, setting $\beta = (\beta_0 + \beta_1)/2$,

i) if β satisfies both rules, stop and set $\alpha_k = \beta$

ii) if β satisfies Armijo's rule (but not Wolfe's rule), investigate $[\beta, \beta_1]$

iii) if β does not satisfy Armijo's rule, investivate $[\beta_0, \beta]$.

Start with the interval $[0, \alpha'_k]$, where α'_k is defined as follows:

$$\alpha'_k = \min_{i \in \mathbb{N}} \{\gamma^i \mid f(x_k + \gamma^i d_k) > f(x_k) - \sigma_1 \gamma^i \|d_k\|^2\}. \tag{10}$$

As a result, you should have the following function:

```
[x e]= hollerm(xinit, sigma1, sigma2, gamma, epsilon, R, N)          (11)
```

Your method should stop according to the following criterion:

i) if $\|\nabla f(x_k)\| \leq \varepsilon$, then $x = x_k$ and $e = 0$

ii) if the total number of iterations is greater than N, then $e = 1$ and x is the last computed value of the sequence $(x_k)_k$ (too many iterations)

iii) if $\|x_k\| \geq R$, then $x = x_k$ and $e = 2$ (the problem may be unbounded).

Use the Euclidean norm. Your program should use the following functions:

`y = func(x)` and `y= gradient(x)` (12)

which give the value of $f(x)$ and $\nabla f(x)$, respectively. We will specify them when we will test your program. Your program should also use a column vector for `xinit`; the result of the function `gradient` will also be a column vector.
Test your program with the different functions:

- the Rosenbruck function (try different values of ε)
- a quadratic function of the form: $x^\top Ax + b^\top x$, with $A \in \mathbb{R}^{n\times n}$ symmetric, $b \in \mathbb{R}^n$, for example:

$$A = \begin{pmatrix} 14 & 9 & -1 \\ 9 & 18 & 6 \\ -1 & 6 & 5 \end{pmatrix}, \quad A = \begin{pmatrix} 2 & 27 & -7 \\ 27 & -5 & -46 \\ -7 & -46 & 9 \end{pmatrix}. \tag{13}$$

- $f(x) = -\exp(-\|x\|^2)$, for different starting points.

Follow the rules of the previous exercises, in particular send a mail with subject Optimierung 1, Abgabe 8.5. If your are using Octave, please indicate this by additionally writing Octave in the subject.