Norwegian University of Science and Technology
Department of Mathematical Sciences

TMA4180 Optimisation I Spring 2018

Exercise set 2

1 Consider the function

$$f(x, y, z) = 2x^2 + xy + y^2 + yz + z^2 - 6x - 7y - 8z + 9,$$

and the associated unconstrained minimization problem $\min_{(x,y,z)\in\mathbb{R}^3} f(x, y, z)$.

a) Find all points $(x, y, z) \in \mathbb{R}^3$ satisfying the first order necessary conditions for this problem (critical points).

b) Assess whether critical points satisfy second order necessary and/or sufficient optimality conditions.

c) Verify that the function f is convex. Conclude that all the critical points are points of global minimum.

d) Let $(\hat{x}, \hat{y}, \hat{z}) = (0, 0, 0)$, and let $(p_x, p_y, p_z) = (1, 2, 0)$. Verify that this is a descent direction for f at $(\hat{x}, \hat{y}, \hat{z})$. Find the range of steplengths $\alpha > 0$ satisfying the sufficient decrease condition for steps from $(\hat{x}, \hat{y}, \hat{z})$ along (p_x, p_y, p_z) with $c_1 = 4/5$.

e) In the assumptions/notation of **d)**, determine the steplength satisfying the sufficient decrease condition by utilizing the backtracking linesearch with the initial steplength $\alpha_0 = 1.0$ and the step reduction parameter $\rho = 1/4$.

f) In the assumptions/notation of **d)**, find the range of steplengths $\alpha > 0$ satisfying the (weak) curvature condition with $c_2 = 0.9$.

g) As in the previous case, use the bisection linesearch algorithm to find a steplength satisfying the weak Wolfe conditions (see the note on the wiki) with $c_1 = 4/5$ and $c_2 = 0.9$.

2 *(See N&W, Exercise 2.8)* Assume that $f : \mathbb{R}^n \to \mathbb{R}$ is a convex function. Show that the set of minimisers of f is convex (empty or non-empty).

3 Show that a strictly convex function $f : \mathbb{R}^n \to \mathbb{R}$ has at most one global minimiser. In addition, construct a strictly convex function that has no global minimisers.

4 Show that the function $f : \mathbb{R}^2 \to \mathbb{R}$,

$$f(x, y) = \log(e^x + e^y)$$

is convex.

5 Assume that f is a continuously differentiable function satisfying

$$\lim_{\|x\|\to\infty} \frac{f(x)}{\|x\|} = +\infty.$$

Show that the equation

$$\nabla f(x) = u$$

has a solution for every $u \in \mathbb{R}^n$.

Hint: Consider global minima of the function $f_u(x) := f(x) - u^T x$.

6 Consider a twice differentiable function $f : \mathbb{R}^n \to \mathbb{R}$. Suppose that $\inf_{x\in\mathbb{R}^n} f(x) = f^* > -\infty$. Assume further that the gradient of f is a Lipschitz continuous on $\mathbb{R}^n$ with constant $L > 0$, that is

$$\|\nabla f(x) - \nabla f(y)\| \le L\|x - y\|, \qquad \forall x, y \in \mathbb{R}^n.$$

a) Show that $\forall p, x \in \mathbb{R}^n : p^{\mathrm{T}}\nabla^2 f(x)p \le L\|p\|^2$.

b) Consider a steepest descent step $x_{k+1} = x_k + \alpha_k p_k$, where $p_k = -\nabla f(x_k)$. Use **a)** and a 2nd order Taylor series expansion of f to show that

$$f(x_{k+1}) \le f(x_k) - (1 - L\alpha_k/2)\alpha_k\|\nabla f(x_k)\|^2.$$

Note: this inequality implies that steepest descent method with any $0 < \alpha_k < 2/L$ is a *descent method*, i.e. $f(x_{k+1}) < f(x_k)$ if $\nabla f(x_k) \ne 0$.

c) Let us consider a steepest descent method with *fixed* steplength $\alpha_k \equiv \alpha \in (0, 2/L)$. Use **b)** to conclude that for steepest descent method with such a choice of steplength we have $\lim_{k\to\infty} \|\nabla f(x_k)\| = 0$.

Note: **c)** implies that any limit point of the sequence (x_k), if exists, must be a critical point.

The problem with algorithm above is that we may not have any information about the Lipschitz constant L, which is necessary for step size selection. One possible workaround could be to use the following algorithm. (An even better idea is to use a linesearch procedure.)

d) Consider now a steepest descent method with *variable* (positive) steplengths, which satisfy two conditions: $\lim_{k\to\infty} \alpha_k = 0$ while $\sum_{k=0}^{\infty} \alpha_k = +\infty$. Show that under these conditions we have $\liminf_{k\to\infty} \|\nabla f(x_k)\| = 0$.

Hint: algorithm becomes a descent method from some $k \in \mathbb{N}$. *Try to sum up the terms* $f(x_{k+1}) - f(x_k)$ *for such large* k *in order to arrive at the desired conclusion.*

Note, that **d)** implies that, for some subsequence k', we have that $\lim_{k'\to\infty} \|\nabla f(x_{k'})\| = 0$. Therefore, any limit point of this subsequence, if exists, must be stationary, similarly to **c)**.

As a side note we can go even further for convex twice differentiable functions f in the last exercise.

Namely, assume that f is, in addition to the previously assumed properties, a convex function, which attains (global) minimum at a point $x^* \in \mathbb{R}^n$. Also, assume that the steplength $\alpha_k \equiv \alpha \in (0, 1/L]$ so that

$$f(x_{k+1}) \leq f(x_k) - \frac{\alpha}{2}\|\nabla f(x_k)\|^2. \tag{1}$$

Since f is convex and differentiable, we have the lower bound

$$f(x^*) \geq f(x_k) + \nabla f(x_k)^{\mathrm{T}}(x^* - x_k). \tag{2}$$

Substituting this into (1) we get

$$\begin{aligned} f(x_{k+1}) &\leq f(x_k) - \frac{\alpha}{2}\|\nabla f(x_k)\|^2 \\ &\leq f(x^*) + \nabla f(x_k)^{\mathrm{T}}(x_k - x^*) - \frac{\alpha}{2}\|\nabla f(x_k)\|^2. \end{aligned}$$

The last two terms in this inequality can be further written as

$$\begin{aligned} \nabla f(x_k)^{\mathrm{T}}(x_k - x^*) - \frac{\alpha}{2}\|\nabla f(x_k)\|^2 &= \frac{1}{2\alpha}\left[\|x_k - x^*\|^2 - \|x_k - x^* - \alpha\nabla f(x_k)\|^2\right] \\ &= \frac{1}{2\alpha}\left[\|x_k - x^*\|^2 - \|x_{k+1} - x^*\|^2\right] \end{aligned}$$

Summing up the terms $f(x_k) - f(x^*)$ we obtain:

$$\begin{aligned} \sum_{k=0}^{n-1}[f(x_{k+1}) - f(x^*)] &\leq \frac{1}{2\alpha}\sum_{k=0}^{n-1}[\|x_k - x^*\|^2 - \|x_{k+1} - x^*\|^2] \\ &= \frac{1}{2\alpha}[\|x_0 - x^*\|^2 - \|x_n - x^*\|^2] \leq \frac{1}{2\alpha}\|x_0 - x^*\|^2. \end{aligned}$$

Finally, since $f(x^k)$ are non-increasing, we get that

$$f(x^n) - f(x^*) \leq \frac{1}{n}\sum_{k=0}^{n-1}[f(x_{k+1}) - f(x^*)] \leq \frac{\|x_0 - x^*\|^2}{2\alpha n}.$$

Thus (i) the function values converge linearly towards the optimal value, and (ii) it pays off to take longer steps α.

Assuming further conditions on f, e.g. *strong convexity* (i.e., the fact that the smallest eigenvalue of $\nabla^2 f(x)$ is uniformly bounded away from 0 for all $x \in \mathbb{R}$) we can further strengthen the lower bound (2) as

$$\begin{aligned} f(x^*) &= f(x_k) + \nabla f(x_k)^{\mathrm{T}}(x^* - x_k) + (x^* - x_k)^{\mathrm{T}}\nabla^2 f(x_k + \xi(x^* - x_k))(x^* - x_k) \\ &\geq f(x_k) + \nabla f(x_k)^{\mathrm{T}}(x^* - x_k) + \lambda_{\min}\|x^* - x_k\|^2, \end{aligned}$$

where $\lambda_{\min} > 0$ is the lower bound on the eigenvalues of $\nabla^2 f(x)$. With this additional assumption the convergence rate can be further improved to

$$f(x^n) - f(x^*) \leq \frac{c^n L\|x_0 - x^*\|^2}{2},$$

for steps $\alpha \in (0, 2/(L + \lambda_{\min}))$, where $0 < c < 1$.