Norwegian University of Science and Technology
Department of Mathematical Sciences

TMA4180 Optimisation I Spring 2018

Solutions to exercise set 2

$\boxed{1}$ 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).

Solution: $\nabla f(x,y,z) = (4x + y - 6, x + 2y + z - 7, y + 2z - 8)$, so critial/stationary point(s) of f, that is, points for which $\nabla f = 0$, must satisfy the system

$$\begin{bmatrix} 4 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 6 \\ 7 \\ 8 \end{bmatrix}.$$

The unique solution is $(x,y,z) = (6,6,17)/5$.

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

Solution: We evaluate the Hessian of f:

$$\nabla^2 f = \begin{bmatrix} 4 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 2 \end{bmatrix}.$$

All three eigenvalues of $\nabla^2 f$ are positive, and thus it is positive definite. Therefore, both necessary and sufficient conditions for a local minimum are satisfied at $(x,y,z) = (6,6,17)/5$.

Alternatively, one can look at the leading diagonal minors of $\nabla^2 f$. Indeed $\det(4) = 4 > 0$;

$$\det \begin{bmatrix} 4 & 1 \\ 1 & 2 \end{bmatrix} = 4 \cdot 2 - 1 \cdot 1 = 7 > 0,$$

and finally

$$\det \begin{bmatrix} 4 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 2 \end{bmatrix} = 10 > 0.$$

Since all leading minors are positive, the matrix is positive definite.

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

Solution: The function is twice differentiable with positife semi-definite Hessian (in fact, positive definite). Therefore, it is convex (in fact, strictly and even strongly convex) and all critical points (and owing to strict convexity there can be at most one) 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 $c_1 = 4/5$.

Solution: $p = (p_x, p_y, p_z)$ is a descent direction if $f(x, y, z)^{\mathrm{T}}(p_x, p_y, p_z) < 0$. Substituting the numbers we get $f(x, y, z)^{\mathrm{T}}(p_x, p_y, p_z) = -6 - 7 \cdot 2 = -20 < 0$.
Let us define $\phi(\alpha) = f(x + \alpha p_x, y + \alpha p_y, z + \alpha p_z)$. In our case, $\phi(\alpha) = 2\alpha^2 + 2\alpha^2 + 4\alpha^2 - 6\alpha - 14\alpha + 9 = 8\alpha^2 - 20\alpha + 9$. Similarly, consider $\ell(\alpha) = f(x, y, z) + c_1\alpha\nabla f(x, y, z)^{\mathrm{T}}(p_x, p_y, p_z) = 9 + 0.8\alpha(-6 - 7 \cdot 2) = 9 - 16\alpha$.
Finally, $\alpha > 0$ satisfies the sufficient decrease condition if $\phi(\alpha) \leq \ell(\alpha)$, or $8\alpha^2 - 20\alpha + 9 + 16\alpha - 9 = 4\alpha(2\alpha - 1) \leq 0$. Thus $\alpha \in (0, 1/2]$ would be acceptable.

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 $\rho = 1/4$.

Solution: We use the same notation as in the previous exercise.
We begin by setting $\alpha = 1.0$. Since $\phi(1.0) = 8 - 20 + 9 = -3 > \ell(1.0) = 9 - 16 = -7$, whis steplength is unacceptable and is therefore reduced to $\alpha := \alpha \cdot \rho = 1/4$.
At this point $\phi(0.25) = 8/16 - 20/4 + 9 = 4.5 \leq \ell(0.25) = 9 - 16/4 = 5$, the step is accepted by the backtracking linesearch.

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

Solution: We need to solve the inequality $\phi'(\alpha) \geq c_2\phi'(0)$, or $16\alpha - 20 \geq 0.9(16 \cdot 0 - 20) = -18$. Thus $16\alpha \geq 2$ or $\alpha \geq 1/8$.

g) As in the previous case, use the bisection linesearch algorithm to find a steplength satisfying the weak Wolfe conditions.

Solution: We simply follow the bisection algorithm.
Initialization: $\alpha = 1$, $\alpha_{\min} = 0$, $\alpha_{\max} = +\infty$.
Sufficient decrease condition is not satisfied at $\alpha = 1$, therefore we put $\alpha_{\max} = 1$, $\alpha = 0.5$.
Sufficient decrease condition is satisfied at $\alpha = 0.5$. We now check the curvature condition: $\phi'(0.5) = 16 \cdot 0.5 - 20 = -12$ $\phi'(0) = 16 \cdot 0 - 20 = -20$. Since $-12 \not< 0.9 \cdot (-20)$ we stop and return $\alpha = 0.5$.

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

Solution: Let $C \subseteq \mathbb{R}^n$ be the set of minimisers of f, and let $x, y \in C$ and $\lambda \in [0,1]$ be arbitrary. Then $f(x) = \min f = f(y)$ by definition of C. Owing to the convexity of f we have that

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y) = \min f.$$

Hence, by definition of $\min f$, we must have $f(\lambda x + (1-\lambda)y) = \min f$. In other words, $\lambda x + (1-\lambda)y \in C$, so C is convex.

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

Solution: Suppose that f has two distinct global minimisers x and y. Let us take a convex combination of x and y, namely, consider the point $z = \lambda x + (1-\lambda)y$ for some $\lambda \in (0,1)$. Owing to the strict convexity of f we have the strict inequality

$$f(z) < \lambda f(x) + (1-\lambda) f(y) = \lambda \min f + (1-\lambda) \min f = \min f,$$

which contradicts the definition of $\min f$. Therefore f has at most one global minimiser.

Every exponential map $f\colon x \mapsto a^x$ with $a > 0$ and $a \neq 1$ is strictly convex, but admits no local (and thus no global) minimiser on $\mathbb{R}$. Indeed, since $f'(x) = a^x \ln a$, it follows that f is strictly increasing for $a > 1$ and strictly decreasing for $0 < a < 1$. Moreover, strict convexity is a consequence of the inequality $f''(x) = a^x (\ln a)^2 > 0$.

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

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

is convex.

Solution: We utilise the fact that twice continuously differentiable functions are convex if and only if their Hessian matrix is (symmetric) positive semi-definite. Routine calculations yield that

$$\nabla^2 f(x,y) = \frac{\mathrm{e}^{x+y}}{(\mathrm{e}^x + \mathrm{e}^y)^2} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}.$$

The exponential terms are positive, and the constant matrix has eigenvalues 0 and 2. As such, $\nabla^2 f$ is positive semi-definite, and f is convex.

An alternative criterion for positive semi-definiteness of the matrix

$$\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$$

is that all diagonal (not only leading) minors are non-negative. Thus we can check $\det 1 \geq 0$, $\det 1 \geq 0$, and finally

$$\det \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} = 0 \geq 0.$$

$\boxed{5}$ 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 $f_u(x) := f(x) - u^T x$.

Solution: Let $u \in \mathbb{R}^n$ be arbitrary. Using the hint, note first that critical points of f_u are solutions to the equation $\nabla f(x) = u$, because $\nabla f_u(x) = \nabla f(x) - u$. Thus it suffices to show that f_u has a critical point, and in particular, we look for a global minimum, which is guaranteed to exist provided f_u is coercive (it is continuously differentiable and therefore lower semi-continuous). Now, f is certainly more than coercive: it grows superlinearly—faster than a linear function in the sense of the norm—to $+\infty$ as $\|x\| \to \infty$. (Indeed,

$$f(x) = \frac{f(x)}{\|x\|} \cdot \|x\| \to +\infty \cdot (+\infty) = +\infty \qquad \text{as} \qquad \|x\| \to \infty.)$$

Moreover, owing to Cauchy–Schwarz' inequality $u^\top x \le \|u\|\|x\|$ we can derive the estimate

$$f_u(x) = \frac{f(x)}{\|x\|}\|x\| - u^\top x \ge \left(\frac{f(x)}{\|x\|} - \|u\|\right)\|x\|.$$

Since $f(x)/\|x\| \to +\infty$ as $\|x\| \to \infty$, the first part in the parenthesis will eventually be larger than $\|u\|$, no matter which u we consider. Therefore $f_u(x) \to +\infty$ as $\|x\| \to \infty$, and f_u is coercive.

$\boxed{6}$ 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 $\forall p, x \in \mathbb{R}^n : p^{\mathrm{T}}\nabla^2 f(x)p \le L\|p\|^2$.

Solution: Per definition of (directional) derivatives, we have

$$p^{\mathrm{T}}\nabla^2 f(x)p = \lim_{\delta\to 0} \frac{[\nabla f(x+\delta p) - \nabla f(x)]^{\mathrm{T}} p}{\delta}.$$

It remains to use the Lipschitz continuity of $\nabla f(\cdot)$ and Cauchy–Schwarz inequality on the right hand side of this inequality.

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 $\nabla f(x_k) \neq 0$.

Solution: The desired inequality follows immediately from Taylor expansion:

$$\begin{aligned} f(x_{k+1}) &= f(x_k) + \nabla f(x_k)^{\mathrm{T}}(x_{k+1} - x_k) + \frac{1}{2}(x_{k+1} - x_k)^{\mathrm{T}}\nabla^2 f(x_k + \xi\alpha_k p_k)(x_{k+1} - x_k) \\ &= f(x_k) + \alpha_k \nabla f(x_k)^{\mathrm{T}} p_k + \frac{\alpha_k^2}{2} p_k^{\mathrm{T}} \nabla^2 f(x_k + \xi\alpha_k p_k) p_k \\ &\leq f(x_k) - \alpha_k \|\nabla f(x_k)\|^2 + \frac{\alpha_k^2 L}{2}\|\nabla f(x_k)\|^2, \end{aligned}$$

for some $\xi \in [0, 1]$.

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 $\lim_{k\to\infty} \|\nabla f(x_k)\| = 0$.

Solution: From **b)** it follows that the sequence of function values $f(x_{k+1})$ must be monotonically non-increasing. If $\lim_{k\to\infty} \|\nabla f(x_k)\| \neq 0$ then this sequence is strictly decreasing and unbounded from below. (Indeed, this would mean that $\exists \epsilon > 0 : \forall N \exists N(\epsilon) \geq N : \|\nabla f(x_{N(\varepsilon)})\| \geq \epsilon$). However, we also know that $f(x_{k+1}) \geq f^*$, and therefore $\lim_{k\to\infty} \|\nabla f(x_k)\|$ must be zero.

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.

Solution: Since $\alpha_k \to 0$ from some $k_1 \in \mathbb{N}$ we have that $\alpha_k \in (0, 2/L)$, $k \geq k_1$, and as a result the method becomes a descent method. In fact, from some $k_2 \in \mathbb{N}$ we have that $\alpha_k \in (0, 1/L]$, $k \geq k_2$, which means that $-(1 - L\alpha_k/2) \leq -1/2$ and as a result

$$f(x_{k+1}) \leq f(x_k) - \alpha_k/2\|\nabla f(x_k)\|^2, \qquad k \geq k_2.$$

Therefore, as in **c)**, we conclude that the sequence of function values f_k will be non-increasing starting from $k = k_1$, and since it is bounded from below it must also converge to some limit $\hat{f} \geq f^*$.

Let us sum up the terms $f(x_{k+1}) - f(x_k)$ starting from some $k = m \geq k_2$ up to $k = n \geq m$:

$$f(x_{n+1}) - f(x_m) = \sum_{k=m}^{n} [f(x_{k+1}) - f(x_k)] \leq -\sum_{k=m}^{n} \alpha_k/2 \|\nabla f(x_k)\|^2$$
$$\leq -\frac{1}{2}\big[\inf_{m \leq k \leq n} \|\nabla f(x_k)\|^2\big] \sum_{k=m}^{n} \alpha_k$$

If we take a limit on both sides of the inequality above with respect to $n \to \infty$ while recalling that $f_{n+1} \to \hat{f} > -\infty$ and $\lim_{n\to\infty} \sum_{k=m}^{n} \alpha_k = +\infty$, we can see that it can only hold if $\inf_{m \leq k} \|\nabla f(x_k)\|^2 = \lim_{n\to\infty} \inf_{m \leq k \leq n} \|\nabla f(x_k)\|^2 = 0$, for every $m \geq k/2$. Therefore, per definition $\liminf_{m\to\infty} \|\nabla f(x_m)\|^2 = 0$.

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)**.