Norwegian University of Science and Technology
Department of Mathematical Sciences

Solutions to exercise set 2

1 Consider the function

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

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

Solution: A calculation shows that

$$\nabla f(x,y,z) = [4x + y - 6, 2y + x + z - 7, 2z + y - 8]^\top$$

The only critical point, i.e., where $\nabla f = 0$ is

$$\mathbf{x}^* = [x^*, y^*, z^*] = \left[\frac{6}{5}, \frac{6}{5}, \frac{17}{5}\right]^\top.$$

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

Solution: The Hessian of f is constant and equal to

$$\nabla^2 f(x,y,z) = \begin{bmatrix} 4 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 2 \end{bmatrix}$$

A numerical computation of the eigenvalues of $\nabla^2 f$ shows that they are all positive. So the second order sufficient optimality condition is satisfied.

c) Verify that the function f is convex, and conclude that all critical points of f are global minima.

Solution: The function is twice differentiable and its Hessian is positive definite, hence it is convex.

d) Let now $(\hat{x}, \hat{y}, \hat{z}) = (0,0,0)$ and let $p = (1,2,0)$. Verify that p is a descent direction for f at $(0,0,0)$. Find the range of step lengths $\alpha > 0$ that satisfy the Armijo condition for steps from $(0,0,0)$ in direction p with $c_1 = 4/5$.

Solution: We see that $\nabla f(0,0,0) = [-6,-7,-8]^\top$, and $\nabla f(0,0,0)^\top p = -20$ is negative. p is a descent direction. We have that

$$\phi(\alpha) = f((0,0,0) + \alpha p) = f(\alpha, 2\alpha, 0) = 8\alpha^2 - 20\alpha + 9$$

and

$$l(\alpha) = f(0,0,0) + c_1 \alpha \nabla f(0,0,0)^\top p = 9 - \frac{4}{5} \cdot \alpha \cdot 20 = 9 - 16\alpha.$$

We see that the Armijo condition, $\phi(\alpha) \leq l(\alpha)$ is equivalent to

$$8\alpha^2 - 20\alpha \leq -16\alpha,$$

or

$$2\alpha^2 - \alpha \leq 0,$$

which is satisfied for $0 < \alpha \leq \frac{1}{2}$.

e) With the notation/assumptions of **d)**, determine the step length you would obtain by using a backtracking line search with the Armijo condition (with parameter $c_1 = 4/5$) with an initial step length $\hat{\alpha} = 1$, and a step length reduction parameter of $\rho = 1/4$.

Solution: From the previous point, we know that $\hat{\alpha} = 1$ is not acceptable, but $\rho \cdot \hat{\alpha} = \frac{1}{4}$ is. Therefore $\alpha = \frac{1}{4}$.

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

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

is convex.

Solution: We will utilise that a smooth function is convex if and only if its Hessian is positive semidefinite everywhere on $\mathbb{R}^2$

Straightforward calculations show

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

The eigenvalues of the matrix $\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$ are 0 and 2, so the eigenvalues of $\nabla^2 f(x, y)$ are 0 and $2\frac{e^x + e^y}{(e^x + e^y)^2}$. Specifically, they are both nonnegative, so f is convex. However, it is not strictly convex.

3 *(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 (possibly empty, though).

Solution: An empty set is (vacuously) convex, so assume f has at least one minimiser.

Let x and y be two minimisers of f (possibly $x = y$). We need to show that, for $0 \leq \lambda \leq 1$, $\lambda x + (1 - \lambda)y$ is also a minimiser.

Since x and y are minimisers, $f(x) = f(y) = m$ and for all $z \in \mathbb{R}^n$, we have $f(z) \geq m$. Specifically, for $0 \leq \lambda \leq 1$, we have

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

But f is convex, so we also have

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

Consequently, $f(\lambda x + (1-\lambda)) = m$ and $\lambda x + (1-\lambda)$ is a minimiser.

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

Solution: Assume that f has two distinct minimisers, x and y. By the previous task, the line segment connecting x and y also consists of minimizers, and we have

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

for all $0 \leq \lambda \leq 1$. This violates the assumption that f was strictly convex.

An example of a strictly convex funcion with no minimiser is given by

$$f\colon \mathbb{R}^1 \to \mathbb{R}$$
$$f(x) = a^x$$

with $a > 0$.

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 abitrary. Using the hint, we notice that critical points of f_u are solutions to $\nabla f(x) = u$, because $\nabla f_u(x) = \nabla f_u(\hat{x}) = \nabla f(\hat{x}) - u$. Thus, it suffices to show that f_u has a critical point. In particular, we try to show that f_u has a global minimum. f_u is continuously differentiable and therefore lower semi-continuous. We want to show that it is also coercive.

Now, f is certainly more than coercive: it grows superlinearly – faster than any linear function – in the sense of the norm.

By Cauchy–Schwarz, we have $u^\top x \leq \|u\|\|x\|$, and we get (for $x \neq 0$)

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

Since $f(x)/\|x\| \to +\infty$ as $\|x\| \to \infty$, the term in the parenthesis will eventually be strictly positive, regardless of what u is. Therefore $f_u(x) \to +\infty$ as $\|x\| \to \infty$, and f_u is coercive.

6 Implement both the gradient descent method and Newton's method with a line search satisfying the Wolfe conditions (you may want to use a bisection algorithm for the implementation of these conditions)

Apply your method to the minimisation of the Rosenbrock function

$$f(x, y) := 100(y - x^2)^2 + (1 - x)^2.$$

The Newton direction is not necessarily a descent direction for this function, as f is not convex, and thus it might be necessary to modify the search directions in the Newton method. Do this by switching to the negative gradient direction, whenever the inequality

$$-\nabla f(x_k)^T p_k^{\text{Newton}} \leq \varepsilon \|\nabla f(x_k)\| \|p_k^{\text{Newton}}\|$$

holds (here, $\varepsilon > 0$ is some fixed, small parameter).

Solution: See the wiki.