Norwegian University of Science and Technology
Department of Mathematical Sciences

TMA4180
Optimisation I
Spring 2021

Solutions to exercise set 3

1 Consider the quadratic function

$$f(x) = \tfrac{1}{2}x^{\mathrm{T}}Qx - b^{\mathrm{T}}x,$$

where $Q \in \mathbb{R}^{d\times d}$ is a symmetric and positive definite matrix and $b \in \mathbb{R}^d$.

a) Compute the gradient and the Hessian of the function f, verify that f is strictly convex, and find the unique global minimum of f.

Solution: $\nabla f = Qx - b$ and $\nabla^2 f = Q$ from calculus. Since Q is symmetric positive definite (SPD), it follows that f is strictly convex on $\mathbb{R}^d$, and as such, there is at most one global minimum of f. Furthermore, this global minimum x^* must be a stationary point satisfying $\nabla f(x^*) = 0$. We conclude that $x^* = Q^{-1}b$, since Q is invertible (all eigenvalues of Q are positive, and hence, different from zero).

b) Let $x \in \mathbb{R}^d$, and let $p \in \mathbb{R}^d$ be a direction satisfying the inequality $\nabla f(x)^{\mathrm{T}}p < 0$. Compute analytically the step length $\alpha_{x,p}$ that solves the (exact) linesearch problem $\min_{\alpha>0} f(x + \alpha p)$.

Solution: First of all, to avoid trivial cases let us note that $p \neq 0$ and $\nabla f(x) = Qx - b \neq 0$ owing to the inequality $\nabla f(x)^{\mathrm{T}}p < 0$.
Now, let us look at the first order necessary conditions for $\alpha_{x,p}$ to be a minimizer:

$$\frac{\mathrm{d}}{\mathrm{d}\alpha} f(x + \alpha_{x,p}p) = p^{\mathrm{T}}\nabla f(x + \alpha_{x,p}p) = p^{\mathrm{T}}[Q(x + \alpha_{x,p}p) - b] = 0,$$

or

$$\alpha_{x,p} = -\frac{p^{\mathrm{T}}[Qx - b]}{p^{\mathrm{T}}Qp} > 0,$$

since $p^{\mathrm{T}}Qp > 0$ owing to Q being positive definite, and $p^{\mathrm{T}}[Qx-b] = p^{\mathrm{T}}\nabla f(x) < 0$ by our assumption.
Since $\mathrm{d}^2/\mathrm{d}\alpha^2 f(x + \alpha p) = p^{\mathrm{T}}Qp > 0$ the linesearch problem is strictly convex, and therefore $\alpha_{x,p}$ is the unique global minimum.

c) Recall the strong Wolfe conditions:

$$f(x + \alpha p) \leq f(x) + c_1\alpha\nabla f(x)^T p,$$
$$|\nabla f(x + \alpha p)^T p| \leq c_2|\nabla f(x)^T p|.$$

Let x, $p \in \mathbb{R}^d$, and $\alpha_{x,p}$ be as in the previous question. Show that the step length $\alpha_{x,p}$ satisfies the strong Wolfe conditions if and only if $c_1 \leq 1/2$.

Solution: Clearly the strong curvature condition is satisfied because $\nabla f(x + \alpha_{x,p}p)^{\mathrm{T}}p = \mathrm{d}/\mathrm{d}\alpha f(x+\alpha_{x,p}p) = 0$, thus the "new" slope is 0 and must be smaller than or equal in magnitude than the slope we have started with.
We check the sufficient decrease condition now:

$$f(x + \alpha_{x,p}p) - f(x) = \frac{1}{2}\alpha_{x,p}^2 p^{\mathrm{T}}Qp + \alpha_{x,p}p^{\mathrm{T}}(Qx - b) = -\frac{1}{2}\frac{[p^{\mathrm{T}}(Qx-b)]^2}{p^{\mathrm{T}}Qp} < 0,$$

while

$$c_1\alpha_{x,p}\nabla f(x)^{\mathrm{T}}p = -c_1\frac{[p^{\mathrm{T}}(Qx-b)]^2}{p^{\mathrm{T}}Qp}.$$

Thus the sufficient decrease condition is equivalent to the inequality $c_1 \leq 1/2$.

2 Assume that $f\colon \mathbb{R}^d \to \mathbb{R}$ is a strongly convex function, meaning that there exists $c > 0$ such that the function $x \mapsto f(x) - \frac{c}{2}\|x\|^2$ is convex.

a) Consider the Newton method with line search, the so-called *damped Newton method*

$$x_{k+1} = x_k + \alpha_k p_k \qquad \text{with} \qquad p_k = -\nabla^2 f(x_k)^{-1}\nabla f(x_k),$$

where the step length parameter α_k is chosen according to backtracking Armijo line search with parameters $\bar{\alpha} > 0$, $0 < c_1 < 1$, and $0 < \rho < 1$.
Show that p_k is a descent direction in each step, and that the sequence x_k converges to the unique minimizer x^* of f.
Hint: Use Theorem 1 in the note "Convergence of descent methods with backtracking (Armijo) linesearch…" by Anton Evgrafov.

We will now show that this method is invariant under affine transformations:
Assume that $B \in \mathbb{R}^{d\times d}$ is a non-singular matrix (not necessarily orthogonal) and that $c \in \mathbb{R}^d$. Define the function

$$g(x) := f(Bx + c).$$

b) Find expressions for $\nabla g(x)$ and $\nabla^2 g(x)$ in terms of f, B, and c.

Solution: For clarity, we set $y := Bx + c$, so that $g(x) = f(y(x))$. We present two ways of solving this problem. The first method is the explicit method,

$$\frac{\partial g}{\partial x_i}(x) = \sum_{k=1}^{n}\frac{\partial f}{\partial y_k}(Bx+c)\frac{\partial y_k}{\partial x_i} = \sum_{k=1}^{n}\frac{\partial f}{\partial y_k}(Bx+c)B_{ki},$$
$$\frac{\partial^2 g}{\partial x_i \partial x_j}(x) = \frac{\partial}{\partial x_j}\sum_{k=1}^{n}\frac{\partial f}{\partial y_k}(Bx+c)B_{ki} = \sum_{k=1}^{n}\sum_{\ell=1}^{n}\frac{\partial^2 f}{\partial y_k \partial y_\ell}(Bx+c)B_{ki}B_{\ell j},$$

and therefore

$$\nabla g(x) = B^{\mathrm{T}}\nabla f(Bx+c),$$
$$\nabla^2 g(x) = B^{\mathrm{T}}\nabla^2 f(Bx+c)B.$$

These calculations at 'index' level can be messy and exhaustive; there are alternative methods of solving this problem, but perhaps the easiest is to simply memorize two basic chain-rules for ∇ and ∇^2. For any (smooth) maps $h\colon \mathbb{R}^d \to \mathbb{R}$ and $H\colon \mathbb{R}^d \to \mathbb{R}^d$, the composition $h \circ H$ satifies

1. $\nabla(h(H)) = (J_H)^{\mathrm{T}}\nabla h(H)$,
2. $\nabla^2(h(H)) = (J_H)^{\mathrm{T}}\nabla^2 h(H)J_H$,

where J_H is the Jacobian of H. With these in mind, the results follows immediately as $J_y = B$.

c) Let $x \in \mathbb{R}^d$ and denote by x_1 the result of one Newton step starting at x for the minimisation of g with (Armijo) backtracking line search with parameters $0 < c_1 < 1$ and $0 < \rho < 1$.

Moreover, let $y = Bx+c$ and denote by y_1 the result of one Newton step starting at y for the minimisation of f with (Armijo) backtracking line search with the same parameters $0 < c_1 < 1$ and $0 < \rho < 1$.

Show that

$$y_1 = Bx_1 + c.$$

Solution: By Newtons method we have $p_0 = -[\nabla^2 g(x_0)]^{-1}\nabla g(x_0)$, and so $x_1 = x_0 + \alpha_0 p_0$, where $\alpha_0 = \rho^{k_0}$ and $k_0 \in \{0, 1, 2, ...\}$ is the smallest non-negative integer where Armijo's condition is satisfied:

$$g(x_0 + \alpha_0 p_0) \leq g(x_0) + c_1\alpha_0 p_0^{\mathrm{T}}\nabla g(x_0).$$

Turning to the minimization of f with respect to y, we similarly obtain $\tilde{p}_0 = -[\nabla^2 f(y_0)]^{-1}\nabla g(y_0)$ and $y_1 = y_0 + \tilde{\alpha}_0\tilde{p}_0$, where $\tilde{\alpha}_0 = \rho^{\tilde{k}_0}$ and $\tilde{k}_0$ is the smallest non-negative integer where Armijo's condition is satisfied:

$$f(y_0 + \tilde{\alpha}_0\tilde{p}_0) \leq f(y_0) + c_1\tilde{\alpha}_0\tilde{p}_0^{\mathrm{T}}\nabla f(y_0).$$

Our task is to prove that $y_1 = Bx_1 + c$, given that $y_0 = Bx_0 + c$. We start by showing that $Bp_0 = \tilde{p}_0$. Indeed,

$$\begin{aligned}
Bp_0 &= -B[\nabla^2 g(x_0)]^{-1}\nabla g(x_0) \\
&= -B[B^{\mathrm{T}}\nabla^2 f(Bx_0 + c)B]^{-1}B^{\mathrm{T}}\nabla f(Bx_0 + c) \\
&= -[\nabla^2 f(y_0)]^{-1}\nabla f(y_0) \\
&= \tilde{p}_0.
\end{aligned}$$

This further implies that all $\alpha \in \mathbb{R}$ satisfies the two equations

$$\begin{aligned}
g(x_0 + \alpha p_0) &= f(y_0 + \alpha\tilde{p}_0), \\
g(x_0) + c_1\alpha_0 p_0^{\mathrm{T}}\nabla g(x_0) &= f(y_0) + c_1\tilde{\alpha}_0\tilde{p}_0^{\mathrm{T}}\nabla f(y_0).
\end{aligned}$$

As a result, Armijo's condition is satisfied in the (g, x)-regime exactly when it is satisfied in the (f, y)-regime, that is, $k_0 = \tilde{k}_0$ and consequently $\alpha_0 = \tilde{\alpha}_0$. We arrive at the desired conclusion

$$\begin{aligned}
y_1 &= y_0 + \tilde{\alpha}_0\tilde{p}_0 \\
&= (Bx_0 + c) + \alpha_0 Bp_0 \\
&= B(x_0 + \alpha_0 p_0) + c \\
&= Bx_1 + c.
\end{aligned}$$

3 Consider the function

$$f(x, y) = 2x^2 + y^2 - 2xy + 2x^3 + x^4.$$

a) Compute all stationary points of f and find all global or local minimisers of f.

Solution: We have

$$\nabla f = [4x - 2y + 6x^2 + 4x^3, 2y - 2x]^{\mathrm{T}}$$

and

$$\nabla^2 f = \begin{bmatrix} 4 + 12x + 12x^2 & -2 \\ -2 & 2 \end{bmatrix},$$

Hence, stationary points satisfy $y = x$ by the first component of ∇f, while the second component yields that $0 = 2x(1 + 3x + 2x^2) = x(x + 1)(2x + 1)$. Thus critical points of f are $(0, 0)$, $(-\frac{1}{2}, -\frac{1}{2})$, and $(-1, -1)$. Now,

$$\nabla^2 f(0,0) = \begin{bmatrix} 4 & -2 \\ -2 & 2 \end{bmatrix} = \nabla^2 f(-1,-1) \qquad \text{and} \qquad \nabla^2 f(-\tfrac{1}{2}, -\tfrac{1}{2}) = \begin{bmatrix} 1 & -2 \\ -2 & 2 \end{bmatrix}$$

has eigenvalues $3 \pm \sqrt{5} > 0$ and $(3 \pm \sqrt{17})/2$ (one positive, and one negative), respectively. We conclude that $(0, 0)$, and $(-1, -1)$ are strict local minima, while $(-\frac{1}{2}, -\frac{1}{2})$ is a saddle point. Moreover, since $\nabla^2 f$ remains SPD both for $x > 0$ and $x < -1$ (the value of y is irrelevant), it follows that $(0, 0)$ and $(-1, -1)$ are the only candidates for global minima. Evaluating $f(0, 0) = 0 = f(-1, -1)$, shows that both are global minimisers of f.

b) Consider the gradient descent method with backtracking for the minimisation of f. Use the parameters $\rho = 1/2$ and $c_1 = 1/4$. Perform one step with starting value $(x_0, y_0) = (-1, 0)$. Does the method converge to a minimiser of f?

Solution: Gradient descent method gives $(x_{k+1}, y_{k+1}) = (x_k, y_k) + p_k$, with $p_k = -\nabla f_k$. Starting with preliminary step length α, $\rho = 1/2$, and $c_1 = 1/4$, we accept a new step provided

$$f\big((x_0, y_0) + \alpha p_0\big) \le f(x_0, y_0) + c\alpha \nabla f(x_0, y_0)^{\mathrm{T}} p_0 = 1 - 2\alpha$$

using that $p_0 = -\nabla f(x_0, y_0) = (2, -2)$.
Beginning with $\alpha = 1$, we reject the first try since $f\big((x_0, y_0) + \alpha p_0\big) = 13 > -1$. Reducing to $\alpha \mapsto \rho\alpha = 1/2$, still gives rejection, but $\alpha = 1/4$ succeeds, because $f\big((x_0, y_0) + \alpha p_0\big) = 1/16 \le 1/2$. Hence, we put $(x_1, y_1) = (-\frac{1}{2}, -\frac{1}{2})$,and proceed with a new round. However, (x_1, y_1) is a critical (saddle) point for f, so the gradient method stops here, thereby failing to converge to a minimiser.

c) Consider Newton's method with backtracking for the minimisation of f. Use the parameters $\rho = 1/2$ and $c_1 = 1/4$. Perform one step with starting value $(x_0, y_0) = (-1, 0)$. Does the method converge to a minimiser of f?

Solution: Similarly as in the previous exercise, the backtracking acceptance criterion for Newton's method reads

$$f\big((x_0, y_0) + \alpha p_0\big) \le f(x_0, y_0) + c\alpha \nabla f(x_0, y_0)^{\mathrm{T}} p_0 = 1 - \tfrac{1}{2}\alpha,$$

since $p_0 = -\nabla^2 f(x_0, y_0)^{-1} \nabla f(x_0, y_0) = (0, -1)$ and $c_1 = 1/4$. Starting with $\alpha = 1$, we have $f\big((x_0, y_0) + \alpha p_0\big) = 0 \leq 1/2$, so the step is accepted. We then put $(x_1, y_1) = (x_0, y_0) + p_0 = (-1, -1)$. This point is a global minimiser, the conclusion being that Newton's method converged in one step.