Norwegian University of Science and Technology
Department of Mathematical Sciences

TMA4180
Optimization Theory
Spring 2016

Exercise set 5 solution

1 First off, we find:

$$g = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \qquad B = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix},$$

and from this we find the *full step*

$$p_0^B = \begin{bmatrix} -1 \\ -1 \end{bmatrix}.$$

a) We observe that $||p_0^B|| = \sqrt{2}$, so when $\Delta = 2$, the full step is accepted and the next step is simply $x_1 = x_0 + p_0^B = (0,0)$, the global minimizer of f.

However, when $\Delta = 5/6$, the full step is not accepted, and we need to calculate the minimizer lying on the circle centered in x_0 with radius $5/6$. To this end, we employ Theorem 4.1 in N&W, which states that p is a global solution of the trust-region problem

$$\min_{x\in\mathbb{R}^n} m(x) = f + g^T x + \frac{1}{2}x^T Bx \quad \text{s.t.} \quad ||x|| \leq \Delta$$

if and only if p is feasible and there exists a scalar $\lambda \geq 0$ such that the three conditions below are satisfied:

$$(B + \lambda I)p = -g, \tag{1}$$
$$\lambda(\Delta - ||p||) = 0, \tag{2}$$
$$(B + \lambda I) \text{ is positive semidefinite.} \tag{3}$$

In our case, we immediately see that condition (3) is satisfied for any $\lambda \geq 0$. Condition (1), when solved with respect to $p = (p_1, p_2)$, states:

$$p_1 = -\frac{1}{1+\lambda}$$
$$p_2 = -\frac{2}{2+\lambda}.$$

To satisfy condition (2), we must have either $\lambda = 0$ or $\Delta - ||p|| = 0$. The former option gives $p = (-1, -1)$, which is the unfeasible full step. Therefore, we must consider the latter option, and try to solve the equation

$$||p||^2 = \frac{1}{(1+\lambda)^2} + \frac{4}{(2+\lambda)^2} = \Delta^2 = \frac{25}{36}.$$

Multiplying by $(1+\lambda)^2(2+\lambda)^2$ and simplifying yields the fourth order polynomial equation for λ:

$$g(\lambda) = 25\lambda^4 + 150\lambda^3 + 145\lambda^2 - 132\lambda - 188 = 0. \tag{4}$$

By inspection, we find one solution as $\lambda = 1$. We may then (by e.g. polynomial division) factorize and find:

$$g(\lambda) = (\lambda - 1)(25\lambda^3 + 175\lambda^2 + 320\lambda + 188).$$

The second factor is clearly nonzero for all $\lambda \geq 0$, so the only non-negative λ satisfying (4) is $\lambda = 1$. Now, conditions (1)-(3) are satisfied. With $\lambda = 1$, we find $p = (-1/2, -2/3)$, and $x_1 = x_0 + p = (1/2, 1/3)$.

b) First off, we may observe that if $\Delta \geq \sqrt{2}$, the full step is accepted, and we have $x_1 = (0, 0)$.

Next, we find that the steepest descent step is

$$p_0^U = -\frac{g^T g}{g^T B g} g = \begin{bmatrix} -5/9 \\ -10/9 \end{bmatrix},$$

and observe that if $\Delta \leq ||p_0^U|| = \sqrt{125/81}$, the step will be taken in the steepest descent direction, to the edge of the trust region, i.e. we will have

$$p = \Delta \frac{p_0^U}{||p_0^U||} = \begin{bmatrix} -\Delta/\sqrt{5} \\ -2\Delta/\sqrt{5} \end{bmatrix},$$

i.e. we will have $x_1 = (1\text{-}\frac{\Delta}{\sqrt{5}},\ 1\text{-}\frac{2\Delta}{\sqrt{5}})$. Note that in the case $\Delta = 5/6$, this gives us $x_1 \simeq (0.63, 0.25)$, which is not too far from the optimal x_1 found in a).

This leaves the case $\sqrt{125/81} \leq \Delta \leq \sqrt{2}$. Here, we follow the dogleg path

$$p(\tau) = p_0^U + \tau(p_0^B - p_0^U), \qquad 0 \leq \tau \leq 1$$

all the way to the boundary of the trust region, that is, we solve the quadratic problem

$$||p_0^U + \tau(p_0^B - p_0^U)||^2 = ||p_0^U|| + 2\tau(p_0^B - p_0^U)^T p_0^U + \tau^2 ||p_0^B - p_0^U||^2 = \Delta^2$$

to find the value of τ. Inserting our values for p_0^U and p_0^B and solving the resulting quadratic equation, we find

$$\tau = -\frac{10}{17} + \sqrt{\frac{100 - 17(125 - 81\Delta^2)}{17^2}}, \tag{5}$$

where the negative solution is discarded since it results in $\tau < 0$. Note that at $\Delta = \sqrt{125/81}$ and $\Delta = \sqrt{2}$ we recover $\tau = 0$ and $\tau = 1$, respectively. Thus, we find $x_1 = x_0 + p_0^U + \tau(p_0^B - p_0^U)$, with τ as indicated by (5).

2 We denote, for $\Delta > 0$, by $p^*(\Delta)$, the solution of the minimization problem

$$\min_{||p|| \leq \Delta} m(p), \tag{6}$$

where

$$m(p) = f(x_0) + \nabla f(x_0)^T p + \frac{1}{2} p^T \nabla^2 f(x_0) p$$

for some function $f : \mathbb{R}^n \to \mathbb{R}$ and $x_0 \in \mathbb{R}^n$. We wish to show that in general, the curve $\Delta \mapsto p^*(\Delta)$ is not planar.

We will do this by constructing a counterexample where we find four points p_1, p_2, p_3 and p_4 which solve the problem (6) for different values of Δ, construct a plane containing p_1, p_2 and p_3, and show that p_4 does not lie in this plane. Following the hint, and taking $f(x) = x_1^2 + 2x_2^2 + 3x_3^2$, we find

$$\nabla f(x) = \begin{bmatrix} 2x_1 \\ 4x_2 \\ 6x_3 \end{bmatrix}, \qquad \nabla^2 f(x) = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 6 \end{bmatrix}.$$

Let us for simplicity assume that $x_0 = (1, 1, 1)$, such that $\nabla f(x_0) = [2, 4, 6]^T$. Furthermore, from Theorem 4.1 in N&W, as stated in the solution of exercise 1a), we know that for a given Δ, the solution of (6) must be of the form

$$p^*(\Delta) = -(\nabla^2 f(x_0) + \lambda I)^{-1} \nabla f(x_0) = - \begin{bmatrix} 2/(2+\lambda) \\ 4/(4+\lambda) \\ 6/(6+\lambda) \end{bmatrix},$$

with $\lambda \geq 0$ chosen such that $\lambda(||p^*(\Delta)|| - \Delta) = 0$. We may observe that $\lambda = 0$ corresponds to the global minimizer of $m(p)$, $p^B = -x_0$. This minimizer will only be accepted if $\Delta \geq ||p^B|| = \sqrt{3}$. Otherwise, we will have $\lambda > 0$ and $||p^*(\Delta)|| = \Delta$. We can therefore construct solutions to (6) by picking some value of λ and calculating the resulting $||p^*(\Delta)||$ to find which Δ the solution corresponds to. Choosing

$$\lambda_1 = 2, \qquad \lambda_2 = 4, \qquad \lambda_3 = 6 \qquad \lambda_4 = 8,$$

we find the minimizers

$$p_1 = - \begin{bmatrix} 1/2 \\ 2/3 \\ 3/4 \end{bmatrix}, \qquad p_2 = - \begin{bmatrix} 1/3 \\ 1/2 \\ 3/5 \end{bmatrix}, \qquad p_3 = - \begin{bmatrix} 1/4 \\ 2/5 \\ 1/2 \end{bmatrix}, \qquad p_4 = - \begin{bmatrix} 1/5 \\ 1/3 \\ 3/7 \end{bmatrix},$$

corresponding to $\Delta = \sqrt{181/144}$, $\Delta = \sqrt{649/900}$, $\Delta = \sqrt{189/400}$ and $\Delta = \sqrt{3691/11025}$, respectively. The first three points lie in the plane given by

$$-6x_1 + 15x_2 - 10x_3 - 1/2 = 0,$$

but since

$$6\frac{1}{5} - 15\frac{1}{3} + 10\frac{3}{7} - 1/2 = -\frac{1}{70} \neq 0,$$

the fourth point does not. Therefore, the curve $\Delta \mapsto p^*(\Delta)$ is not planar.

3 The four points defining the path of the double-dogleg method are

- $p_1 = 0$;
- $p_2 = p^C = -\frac{g^Tg}{g^TBg}g$;
- $p_3 = \bar{\gamma}p^B = -\bar{\gamma}B^{-1}g, \quad \bar{\gamma} \in (\gamma, 1)$; and
- $p_4 = p^B = -B^{-1}g$,

where $\gamma = \frac{||g||^4}{(g^TBg)(g^TB^{-1}g)}$. The path can be parametrized by τ as following:

$$p(\tau) = \begin{cases} \tau p^C, & 0 \leq \tau \leq 1 \\ p^C + (\tau - 1)(\hat{\gamma}p^B - p^C), & 1 \leq \tau \leq 2 \\ \hat{\gamma}p^B + (\tau - 2)(p^B - \hat{\gamma}p^C), & 2 \leq \tau \leq 3. \end{cases}$$

To prove that $||p||$ is monotonically increasing we need to show that $\frac{\mathrm{d}||p||}{\mathrm{d}\tau} > 0$ for all $\tau \in [0, 3]$. Or, equivalently, we can show that $\frac{\mathrm{d}||p||^2}{\mathrm{d}\tau} > 0$. For $\tau \in [0, 1]$, this is obviously satisfied since $\frac{\mathrm{d}||\tau p^C||}{\mathrm{d}\tau} = ||p^C|| > 0$

For $\tau \in [1, 2]$, we have

$$\begin{aligned} ||p||^2 &= ||p^C + (\tau - 1)(\hat{\gamma}p^B - p^C)||^2 \\ &= (p^C + (\tau - 1)(\hat{\gamma}p^B - p^C))^T(p^C + (\tau - 1)(\hat{\gamma}p^B - p^C)) \\ &= ||p_C^2|| + 2(\tau - 1)(p^C)^T(\hat{\gamma}p^B - p^C) + (\tau - 1)^2||\hat{\gamma}p^B - p^C||^2, \\ \Rightarrow \frac{\mathrm{d}||p||^2}{\mathrm{d}\tau} &= 2(p^C)^T(\hat{\gamma}p^B - p^C) + 2(\tau - 1)||\hat{\gamma}p^B - p^C||^2. \end{aligned}$$

The second term of $\frac{\mathrm{d}||p||^2}{\mathrm{d}\tau}$ is clearly positive. For the first term we have that

$$\begin{aligned} (p^C)^T(\hat{\gamma}p^B - p^C) &= \bar{\gamma}(p^C)^Tp^B - (p^C)^Tp^C \\ &= \bar{\gamma}\left(-\frac{g^Tg}{g^TBg}\right)g^T(-B^{-1}g) - \left(-\frac{g^Tg}{g^TBg}\right)^2 g^Tg \\ &> \gamma\frac{g^TB^{-1}g}{g^TBg}g^Tg - \frac{(g^Tg)^3}{(g^TBg)^2} \\ &= \frac{(g^Tg)^2}{(g^TBg)(g^TB^{-1}g)}\frac{g^TB^{-1}g}{g^TBg}g^Tg - \frac{(g^Tg)^3}{(g^TBg)^2} = 0. \end{aligned}$$

Hence, we see that $\frac{\mathrm{d}||p||^2}{\mathrm{d}\tau} > 0$ for $\tau \in [1, 2]$.

For $\tau \in [2, 3]$ we have

$$\begin{aligned} ||p||^2 &= ||\bar{\gamma}p^B + (\tau - 2)(p^B - \bar{\gamma}p^B)||^2 \\ &= ||p^B(\bar{\gamma} + (\tau - 2)(1 - \bar{\gamma}))||^2 \\ &= (\bar{\gamma} + (\tau - 2)(1 - \bar{\gamma}))^2||p^B||^2 \\ \Rightarrow \frac{\mathrm{d}||p||^2}{\mathrm{d}\tau} &= 2(\bar{\gamma} + (\tau - 2)(1 - \bar{\gamma}))(1 - \bar{\gamma})||p^B||^2 > 0. \end{aligned}$$

The last inequality follows since all terms in $\frac{\mathrm{d}||p||^2}{\mathrm{d}\tau}$ are positive.

4 **a)** We start by defining the constraint functions:

$$c_1(\mathbf{x}) = y - x$$
$$c_2(\mathbf{x}) = x^3 - y^4.$$

The feasible set is sketched in figure 1. To characterize the tangent cone $T(\mathbf{x})$ and the set of linearized feasible directions $\mathcal{F}(\mathbf{x})$, we employ lemma 12.2 in N&W, which states that if the LICQ conditions hold at a feasible point $\mathbf{x}$, then $T(\mathbf{x}) = \mathcal{F}(\mathbf{x})$. In the interior of Ω, there are no active constraints, and so the LICQ condition is vacuously true. Also, since no constraints are active, $T(\mathbf{x}) = \mathcal{F}(\mathbf{x}) = \mathbb{R}^2$.

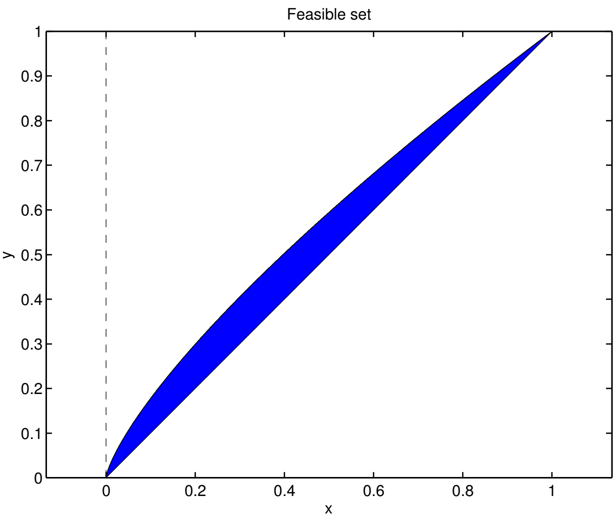

Figure 1: Feasible set, exercise 4a.

Next, we look at the points with one active constraint, starting with the line $c_1(\mathbf{x}) = 0$ (excluding the points where $c_2(\mathbf{x}) = 0$), and observing

$$\nabla c_1(\mathbf{x}) = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

This is nonzero, and so the LICQ condition holds. We find that

$$T(\mathbf{x}) = \mathcal{F}(\mathbf{x}) = \{d \in \mathbb{R}^2 : \nabla c_1(\mathbf{x})^T d \geq 0\} = \{d \in \mathbb{R}^2 : d_2 \geq d_1\}$$

Considering the line $c_2(\mathbf{x}) = 0$ (excluding the points where $c_1(\mathbf{x}) = 0$), we observe

$$\nabla c_2(\mathbf{x}) = \begin{bmatrix} 3x^2 \\ -4y^3 \end{bmatrix},$$

which is also nonzero as the point (0,0) is not under consideration yet since two constraints are active there. Thereby, the LICQ conditions hold, and we have

$$T(\mathbf{x}) = \mathcal{F}(\mathbf{x}) = \{d \in \mathbb{R}^2 : \nabla c_2(\mathbf{x})^T d \geq 0\} = \{d \in \mathbb{R}^2 : 3x^2 d_1 \geq 4y^3 d_2\}.$$

Lastly, we consider the corner points (1,1) and (0,0), where both constraints are active. In (1,1) we have

$$\nabla c_1(1,1) = \begin{bmatrix} -1 \\ 1 \end{bmatrix} \quad \text{and} \quad \nabla c_2(1,1) = \begin{bmatrix} 3 \\ -4 \end{bmatrix},$$

so the LICQ condition holds. Thereby,

$$\begin{aligned} T(1,1) = \mathcal{F}(1,1) &= \{d \in \mathbb{R}^2 : \nabla c_1(1,1)^T d \geq 0 \text{ and } \nabla c_2(1,1)^T d \geq 0\} \\ &= \{d \in \mathbb{R}^2 : d_2 \geq d_1 \text{ and } 3d_2 \geq 4d_1\}. \end{aligned}$$

In the last point, (0,0), the LICQ condition does not hold, since

$$\nabla c_1(0,0) = \begin{bmatrix} -1 \\ 1 \end{bmatrix} \quad \text{and} \quad \nabla c_2(0,0) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Therefore, we cannot expect that $T(0,0) = \mathcal{F}(0,0)$. Indeed, we have

$$\begin{aligned} \mathcal{F}(0,0) &= \{d \in \mathbb{R}^2 : \nabla c_1(0,0)^T d \geq 0 \text{ and } \nabla c_2(0,0)^T d \geq 0\} \\ &= \{d \in \mathbb{R}^2 : d_2 \geq d_1\}, \end{aligned}$$

defining a half-space. On the other hand, we can find the tangent cone by looking at the limiting vectors along the lines $c_1(\mathbf{x}) = 0$ and $c_2(\mathbf{x}) = 0$ as $\mathbf{x} \to 0$. Traveling toward (0,0) along $c_1(\mathbf{x}) = 0$, we take

$$z_k = \begin{bmatrix} 1/k \\ 1/k \end{bmatrix}, \quad t_k = ||z_k|| = \sqrt{2}/k,$$

and find the limiting direction

$$d = \lim_{k \to \infty} \frac{z_k}{t_k} = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}.$$

Along $c_2(\mathbf{x}) = 0$, we take

$$z_k = \begin{bmatrix} 1/k \\ 1/k^{3/4} \end{bmatrix}, \quad t_k = ||z_k|| = \frac{\sqrt{\sqrt{k}+1}}{k},$$

and find the limiting direction

$$d = \lim_{k \to \infty} \frac{z_k}{t_k} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

The tangent cone at (0,0) contains all vectors between these limiting cases, which can be shown to be:

$$T(0,0) = \{d \in \mathbb{R}^2 : d_1 \geq 0 \text{ and } d_2 \geq d_1\}.$$

b) We begin by once again defining the constraint functions

$$\begin{aligned} c_1(\mathbf{x}) &= y - x^4 \\ c_2(\mathbf{x}) &= x^3 - y \end{aligned}$$

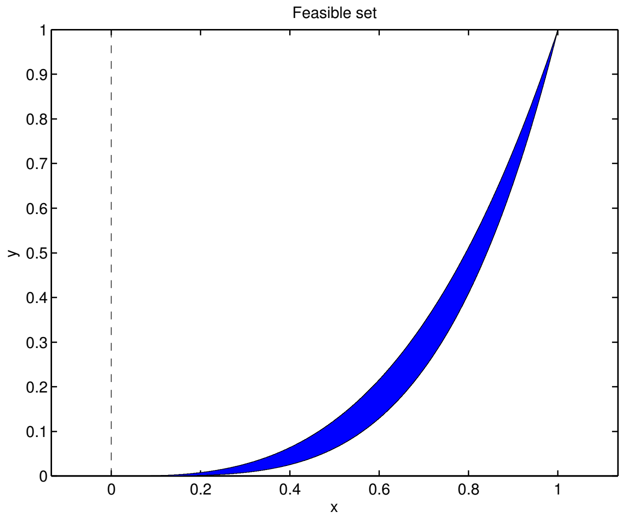

Figure 2: Feasible set, exercise 4b.

The feasible set is sketched in figure 2. To characterize the tangent cone $T(\mathbf{x})$ and the set of linearized feasible directions $\mathcal{F}(\mathbf{x})$, we employ lemma 12.2 in N&W as in the previous exercise. In the interior of Ω, there are no active constraints, and so the LICQ condition is vacuously true. Also, since no constraints are active, $T(\mathbf{x}) = \mathcal{F}(\mathbf{x}) = \mathbb{R}^2$.

Next, we look at the points with one active constraint, starting with the line $c_1(\mathbf{x}) = 0$ (excluding the points where $c_2(\mathbf{x}) = 0$), and observing

$$\nabla c_1(\mathbf{x}) = \begin{bmatrix} -4x^3 \\ 1 \end{bmatrix}$$

This is nonzero, and so the LICQ condition holds. We find that

$$T(\mathbf{x}) = \mathcal{F}(\mathbf{x}) = \{d \in \mathbb{R}^2 : \nabla c_1(\mathbf{x})^T d \geq 0\} = \{d \in \mathbb{R}^2 : d_2 \geq 4x^3 d_1\}$$

Considering the line $c_2(\mathbf{x}) = 0$ (excluding the points where $c_1(\mathbf{x}) = 0$), we observe

$$\nabla c_2(\mathbf{x}) = \begin{bmatrix} 3x^2 \\ -1 \end{bmatrix},$$

which is also nonzero. Thereby, the LICQ conditions hold, and we have

$$T(\mathbf{x}) = \mathcal{F}(\mathbf{x}) = \{d \in \mathbb{R}^2 : \nabla c_2(\mathbf{x})^T d \geq 0\} = \{d \in \mathbb{R}^2 : 3x^2 d_1 \geq d_2\}.$$

Lastly, we consider the corner points (1,1) and (0,0), where both constraints are active. In (1,1) we have

$$\nabla c_1(1,1) = \begin{bmatrix} -4 \\ 1 \end{bmatrix} \quad \text{and} \quad \nabla c_2(1,1) = \begin{bmatrix} 3 \\ -1 \end{bmatrix},$$

so the LICQ condition holds. Thereby,

$$\begin{aligned} T(1,1) = \mathcal{F}(1,1) &= \{d \in \mathbb{R}^2 : \nabla c_1(1,1)^T d \geq 0 \text{ and } \nabla c_2(1,1)^T d \geq 0\} \\ &= \{d \in \mathbb{R}^2 : d_2 \geq 4d_1 \text{ and } 3d_1 \geq d_2\}. \end{aligned}$$

In the last point, (0,0), the LICQ condition does not hold, since

$$\nabla c_1(0,0) = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad \text{and} \quad \nabla c_2(0,0) = \begin{bmatrix} 0 \\ -1 \end{bmatrix}.$$

Therefore, we cannot generally expect that $T(0,0) = \mathcal{F}(0,0)$. We have

$$\begin{aligned} \mathcal{F}(0,0) &= \{d \in \mathbb{R}^2 : \nabla c_1(0,0)^T d \geq 0 \text{ and } \nabla c_2(0,0)^T d \geq 0\} \\ &= \{d \in \mathbb{R}^2 : d_2 \geq 0 \text{ and } d_2 \leq 0\} \\ &= \{d \in \mathbb{R}^2 : d_2 = 0\} = \{d \in \mathbb{R}^2 : d = [d_1, 0]^T\}. \end{aligned}$$

We can find the tangent cone by looking at the limiting vectors along the lines $c_1(\mathbf{x}) = 0$ and $c_2(\mathbf{x}) = 0$ as $\mathbf{x} \to 0$. Traveling toward (0,0) along $c_1(\mathbf{x}) = 0$, we take

$$z_k = \begin{bmatrix} 1/k \\ 1/k^4 \end{bmatrix}, \quad t_k = ||z_k|| = \frac{k^4}{\sqrt{k^6+1}}$$

and find the limiting direction

$$d = \lim_{k \to \infty} \frac{z_k}{t_k} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

Along $c_2(\mathbf{x}) = 0$, we take

$$z_k = \begin{bmatrix} 1/k \\ 1/k^3 \end{bmatrix}, \quad t_k = ||z_k|| = \frac{k^3}{\sqrt{k^4+1}},$$

and find the limiting direction

$$d = \lim_{k \to \infty} \frac{z_k}{t_k} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

Thus, the tangent cone at (0,0) is

$$T(0,0) = \{d \in \mathbb{R}^2 : d = [d_1, 0]^T, d_1 > 0\},$$

and we see that it does not coincide with the set of linearized feasible directions.