TMA4180
Optimization I
Spring 2020

Norwegian University of Science and Technology
Department of Mathematical Sciences

Solutions to exercise set 5

Assume that $\Omega \subset \mathbb{R}^n$ is a closed and convex set. Recall that we can define in this case for $x \in \Omega$ the normal cone $N_\Omega(x)$ as

$$N_\Omega(x) = \{q \in \mathbb{R}^n : q^T(\hat{x} - x) \le 0 \text{ for all } \hat{x} \in \Omega\},$$

and the tangent cone $T_\Omega(x)$ as

$$T_\Omega(x) = \{p \in \mathbb{R}^n : p^T q \le 0 \text{ for all } q \in N_\Omega(x)\}.$$

1 Assume that $\Omega \subset \mathbb{R}^n$ is non-empty, convex, and closed, and that $x \in \Omega$. Show that $N_\Omega(x)$ and $T_\Omega(x)$ are non-empty, closed, and convex cones.[1]

Solution: We shall prove said properties for $N_\Omega(x)$ only; the proofs for $T_\Omega(x)$ are similar. As we have $0 \in N_\Omega(x)$, it is not empty. To prove that $N_\Omega(x)$ is closed, we note that any sequence $(q_i) \subset N_\Omega(x)$ satisfies (by definition) ${q_i}^T(\hat{x} - x) \le 0$ for all $\hat{x} \in \Omega$, and so if $q_i \to q \in \mathbb{R}^n$, we must have by continuity $q^T(\hat{x} - x) \le 0$ for all $\hat{x} \in \Omega$; thus $q \in N_\Omega(x)$ and so $N_\Omega(x)$ is closed. To show that $N_\Omega(x)$ is a cone, note that for $t > 0$ we have

$$tq^T(\hat{x} - x) \le 0 \quad \Leftrightarrow \quad q^T(\hat{x} - x) \le 0,$$

and so $tq \in N_\Omega(x) \Leftrightarrow q \in N_\Omega(x)$. For convexity, we pick $q_0, q_1 \in N_\Omega(x)$ and $0 < t < 1$. By the cone property we have $tq_0 \in N_\Omega(x)$ and $(1-t)q_1 \in N_\Omega(x)$, and so for any $\hat{x} \in \Omega$

$$tq_0^T(\hat{x} - x) \le 0, \quad (1-t)q_1^T(\hat{x} - x) \le 0.$$

Adding this two inequalities together, we obtain (for any $\hat{x} \in \Omega$)

$$\begin{aligned} 0 &\ge tq_0^T(\hat{x} - x) + (1-t)q_1^T(\hat{x} - x), \\ &= (tq_0 + (1-t)q_1)^T(\hat{x} - x), \end{aligned}$$

and so $(tq_0 + (1-t)q_1) \in N_\Omega(x)$; in particular, $N_\Omega(x)$ is convex.

2 Consider the sets

$$\Omega_1 := \{x \in \mathbb{R}^n : \|x\|_\infty \le 1\}$$

and

$$\Omega_2 := \{x \in \mathbb{R}^n : \|x\|_2 \le 1\}.$$

[1] A subset $C \subset \mathbb{R}^n$ is called a cone, if, whenever $p \in C$, then also $\lambda p \in C$ for all $\lambda > 0$.

a) Show that Ω_1 and Ω_2 are non-empty, convex, and closed.

Solution: $0 \in \Omega_i$, $i = 1, 2$ so the sets are non-empty. Since both sets are unit balls with respect to to some norm, we have for arbitrary $x, y \in \Omega_i$, $i = 1, 2$, $\lambda \in [0, 1]$: $\|\lambda x + (1-\lambda)y\| \leq \|\lambda x\| + \|(1-\lambda)y\| = \lambda\|x\| + (1-\lambda)\|y\|$; thereby we have shown convexity.

Closedness follows immediately from the continuity of the norm.

b) In dimension $d = 2$, determine the normal and tangent cones to the sets Ω_1 and Ω_2 at the point $x = (1, 0)$. In addition, determine the normal and tangent cones to Ω_1 at the point $x = (1, 1)$.

Hint: This exercise will probably become easier if you start with drawing a sketch.

Solution: Ω_1 can be defined by the four (smooth) inequality constraints

$$\begin{aligned} 0 &\geq c_1(x) = -x_1 - 1, \\ 0 &\geq c_2(x) = x_1 - 1, \\ 0 &\geq c_3(x) = -x_2 - 1, \\ 0 &\geq c_4(x) = x_2 - 1. \end{aligned}$$

At the point $\hat{x} = (\hat{x}_1, \hat{x}_2) = (1, 0)$ we see that only $c_2 = 0$, that is c_2 is the only 'active' constraint here. The normal cone is then generated by $-\nabla c_2(\hat{x})$ (meaning that all elements of $N_{\Omega_1}(\hat{x})$ are given by this vector times a positive constant). Since $-\nabla c_2(\hat{x}) = [1, 0]^T$, we obtain

$$N_{\Omega_1}(\hat{x}) = \left\{ \begin{bmatrix} t \\ 0 \end{bmatrix} : t \geq 0 \right\}.$$

We easily get the tangent cone by noting that $p^T(-\nabla c_2(\hat{x})) = p_1 * 1 + p_2 * 0 = p_1$, and so for all (nonzero) $q \in N_{\Omega_1}(\hat{x})$, we have $p^T q \leq 0$ if, and only if, $p_1 \leq 0$. Thus

$$T_{\Omega_1}(\hat{x}) = \{p \in \mathbb{R}^2 : p_1 \leq 0\}.$$

For Ω_2, we see that it is given by the inequality constraint

$$0 \leq c(x) = 1 - x_1^2 - x_2^2.$$

As $c(\hat{x}) = 0$ the normal cone is generated by $-\nabla c(\hat{x}) = [2, 0]^T$, or equivalently generated by $[1, 0]^T$, thus $N_{\Omega_2}(\hat{x}) = N_{\Omega_1}(\hat{x})$, and consequently we also have $T_{\Omega_2}(\hat{x}) = T_{\Omega_2}(\hat{x})$.

For the last part of the exercise, we set $\hat{x} = (1, 1)$. At this point we see that both $c_2(\hat{x}) = 0$ and $c_4(\hat{x}) = 0$, while $c_1(\hat{x}) \neq 0 \neq c_3(\hat{x})$. Consequently the normal cone is generated by $-\nabla c_2(\hat{x}) = [1, 0]^T$ and $-\nabla c_4(\hat{x}) = [0, 1]^T$ and so

$$N_{\Omega_1}(\hat{x}) = \left\{ \begin{bmatrix} t \\ s \end{bmatrix} : t, s \geq 0 \right\}.$$

To calculate the tangent cone we see that a vector p satisfies

$$p_1 \cdot t + p_2 \cdot s \leq 0,$$

for all $t, s \geq 0$ if, and only if, $p_1, p_2 \leq 0$. Thus

$$T_{\Omega_1}(\hat{x}) = \left\{ \begin{bmatrix} t \\ s \end{bmatrix} : t, s \leq 0 \right\}.$$

c) Consider the projection problems

$$\pi_{\Omega_i}(z) = \arg\min_{x \in \Omega_i} \frac{1}{2}\|x - z\|_2^2.$$

Use the variational characterisation of the projection ($(\pi_{\Omega_i}(z)-z)^T(x-\pi_{\Omega_i}(z)) \geq 0$ for all $x \in \Omega_i$) to verify the formulas

$$\pi_{\Omega_1}(z) = \big(\max\{-1, \min\{1, z_i\}\}\big)_{i=1,\dots,n}$$

and

$$\pi_{\Omega_2}(z) = \frac{z}{\max\{1, \|z\|_2\}}.$$

Solution: Starting with Ω_1, we denote for simplicity $x^* = \pi_{\Omega_1}(z)$. By the variational characterisation, we get

$$0 \leq (x^* - z)^T(x - x^*) = \sum_{i=1}^{n}(x_i^* - z_i)(x_i - x_i^*), \tag{1}$$

for all $x \in \Omega_1$. We can analyse this sum term by term, by simply selecting $x = (x_1^*, \dots, x_{i-1}^*, x_i, x_{i+1}^*, \dots, x_n^*)$. Then all the terms in the sum above vanish, except for the ith, and we obtain

$$0 \leq (x_i^* - z_i)(x_i - x_i^*). \tag{2}$$

We break the analysis down to three cases.
Case 1: $-1 \leq z_i \leq 1$. We set $x_i = z_i$ and (2) becomes $0 \leq -(x_i^* - z_i)^2$. For this to be satisfied, we deduce that $x_i^* = z_i$.
Case 2: $z_i < -1$. We set $x_i = -1$, and (2) becomes $0 \leq -(x_i^* - z_i)(1 + x_i^*)$. Note that, since $\|x^*\|_\infty \leq 1$, we have $x_i^* \geq -1$ and so consequently $x_i^* - z_i > 0$ while $(1 + x_i^*) \geq 0$. Thus $0 \leq -(x_i^* - z_i)(1 + x_i^*)$ if, and only if, $x_i^* = -1$.
Case 3: $z_i \geq 1$. By similar analysis as in the previous case (now with $x_i = 1$), we deduce that $x_i^* = 1$.
We have componentwise computed x_i^* given z_i. The three cases can be summarized as $x_i^* = \min\{1, \max\{-1, z_i\}\}$.

Let us now deal with projections onto Ω_2. Again, we denote $x^* = \pi_{\Omega_2}(z)$. In the previous part, we used different values of x to compute x^*. Although this could similarly be done here, it is easier to insert the given expression for x^* and see that the optimality condition (variational characterisation) is satisfied. If $\|z\|_2 \leq 1$ then we can put $x^* = z$, and the optimality conditions will be satisfied (left hand side of the inequality will be identically equal to zero). If, on the other hand $\|z\|_2 > 1$, we plug $x = z/\|z\|_2$ into (1) and get

$$\begin{aligned} 0 &\leq (x^* - z)^T(x - x^*) \\ &= (1 - 1/\|z\|_2)(\|z\|_2 - z^T x). \end{aligned}$$

As $1 - 1/\|z\|_2 > 0$ and $\|z\|_2 - z^T x \geq \|z\|_2(1 - \|x\|_2) \geq 0$ (as $\|x\|_2 \leq 1$), we see that the optimality condition is satisfied for all $\|x\|_2 \leq 1$.
In summary, $x^* = z/\max\{1, \|z\|\}$.

d) Consider now the case $n = 2$ and let

$$f(x) = x_1^2 + (x_2 + 2)^2.$$

Find the global solution of the problem

$$\min_{x \in \Omega_1} f(x)$$

(you can do this graphically, if you want) and perform one step of the gradient projection method

$$x^{(k+1)} \leftarrow \pi_{\Omega_1}(x^{(k)} - \alpha \nabla f(x^{(k)}))$$

with step length $\alpha = 1/2$ starting at $x^{(0)} = (1, 1)$.

Solution: We are looking for the point in Ω_1 which is closest to $z = (0, -2)$. This is evidently $x^* = (0, -1)$, which is a projection of z onto Ω_1.
We now compute the projected gradient step. First note that $\nabla f(x^{(0)}) = [2x_1^{(0)}, 2(x_2^{(0)} + 2)]^\top = [2, 6]^\top$. And so $x^{(0)} - \frac{1}{2}\nabla f(x^{(0)}) = (0, -2)$. Therefore $x^{(1)} = \pi_{\Omega_1}((0, -2)) = (0, -1)$, that is, the method converges in one step.

3 Assume that $\Omega \subset \mathbb{R}^n$ is non-empty, convex and closed. Show that the projection mapping $\pi_\Omega \colon \mathbb{R}^n \to \Omega$ is a non-expansive map in the sense that

$$\|\pi_\Omega(x) - \pi_\Omega(y)\|_2 \le \|x - y\|_2$$

for all x, $y \in \mathbb{R}^n$.

Hint: Show first that

$$(\pi_\Omega(x) - x)^T(\pi_\Omega(y) - \pi_\Omega(x)) \ge 0 \qquad \text{and} \qquad (\pi_\Omega(y) - y)^T(\pi_\Omega(x) - \pi_\Omega(y)) \ge 0.$$

Then consider the function $g \colon \mathbb{R} \to \mathbb{R}$,

$$g(\lambda) := \|\lambda(x - y) + (1 - \lambda)(\pi_\Omega(x) - \pi_\Omega(y))\|_2^2.$$

Show that $g'(0) \ge 0$ *and deduce from the fact that* g *is quadratic that* $g(1) \ge g(0)$.

Solution: Fix $x \in \mathbb{R}^n$. By the variational conditions given in the previous exercise, we know that

$$(\pi_\Omega(x) - x)^T(v - \pi_\Omega(x)) \ge 0,$$

for all $v \in \Omega$. In particular, $\pi_\Omega(y) \in \Omega$ for all $y \in \mathbb{R}^n$, and so we get

$$(\pi_\Omega(x) - x)^T(\pi_\Omega(y) - \pi_\Omega(x)) \ge 0.$$

By swapping x and y we also get the second inequality from the hint. Let $g(\lambda)$ be defined as in the hint. Then

$$g'(\lambda) = (x - y - \pi_\Omega(x) + \pi_\Omega(y))^T(\lambda(x - y) + (1 - \lambda)(\pi_\Omega(x) - \pi_\Omega(y))).$$

And so for $\lambda = 0$ we get

$$\begin{aligned} g'(0) &= (x - y - \pi_\Omega(x) + \pi_\Omega(y))^T(\pi_\Omega(x) - \pi_\Omega(y)) \\ &= (\pi_\Omega(x) - x)^T(\pi_\Omega(y) - \pi_\Omega(x)) + (\pi_\Omega(y) - y)^T(\pi_\Omega(x) - \pi_\Omega(y)) \\ &\geq 0. \end{aligned}$$

As g is a positive quadratic function g' is non-decreasing function, and so $g'(\lambda) \geq 0$ for all $0 \leq \lambda \leq 1$. In particular, this implies $g(1) - g(0) = \int_0^1 g'(\lambda) d\lambda \geq 0$. Writing this inequality out in full, we get the result

$$\|\pi_\Omega(x) - \pi_\Omega(y)\|_2^2 = g(0) \leq g(1) = \|x - y\|_2^2.$$

4 Let $A \in \mathbb{R}^{m\times n}$ with $m \geq n$ have full rank, let $b \in \mathbb{R}^m$, and let $\Omega \subset \mathbb{R}^n$ be non-empty, convex, and closed. Consider the restricted least squares problem

$$\min_{x\in\Omega} f(x) \qquad \text{with } f(x) = \frac{1}{2}\|Ax - b\|_2^2 \tag{3}$$

and the gradient projection algorithm

$$x^{(k+1)} \leftarrow \pi_\Omega\big(x^{(k)} - \alpha\nabla f(x^{(k)})\big).$$

Show that this algorithm converges to the unique solution of (3) provided that $0 < \alpha < 2/\sigma_{\max}^2$, where $\sigma_{\max}$ denotes the largest singular value of A.

Hint: Show that the gradient descent step $x \mapsto x - \alpha\nabla f(x)$ is a contraction on $\mathbb{R}^n$, and then use the result of the previous exercise.

Solution: We want to show that the map $T : x \mapsto x - \alpha\nabla f(x)$ is a contraction mapping. First note that

$$\nabla f(x) = A^T Ax - A^T b,$$

and so a quick calculations gives

$$\begin{aligned} \|T(x) - T(y)\|_2 &= \|x - y - \alpha(\nabla f(x) - \nabla f(y))\|_2 \\ &= \|(I - \alpha A^T A)(x - y)\|_2 \\ &=: \|B(x - y)\|_2, \end{aligned}$$

where we define $B = (I - \alpha A^T A)$. Denote the singular values of A by $\sigma_1, \sigma_2, ..., \sigma_n$. The singular values of $A^T A$ are then given by $\sigma_1^2, \sigma_2^2, ..., \sigma_n^2$, and furthermore, the singular values of B are exactly $\tilde{\sigma}_i = |1 - \alpha\sigma_i^2|$ for $i = 1, 2, ..., n$. For simplicity, denote

$$\rho := \max_{i=1,2,\dots,n} |1 - \alpha\sigma_i^2|.$$

As ρ is the largest singular value of B, we have $\|B(x - y)\|_2 \leq \rho\|x - y\|_2$; from this, and the calculation above, we see that T is a contraction on $\mathbb{R}^n$ if $\rho < 1$. To see that this is indeed true, define

$$\sigma_{\min} := \min_{i=1,2,\dots,n} \sigma_i, \qquad \sigma_{\max} := \max_{i=1,2,\dots,n} \sigma_i.$$

We have $1 - \alpha\sigma_{\max}^2 \le 1 - \alpha\sigma_i^2 \le 1 - \alpha\sigma_{\min}^2$ for all i, since α is positive. Exploiting $\alpha < 2/\sigma_{\max}^2$, we get

$$-1 < 1 - \alpha\sigma_{\max}^2,$$

and since A has full rank we also have $0 < \sigma_{\min}$, which implies

$$1 - \alpha\sigma_{\min}^2 < 1.$$

Thus $|1 - \alpha\sigma_i^2| < 1$ for all i, and so $\rho < 1$, which implies that T is a contraction mapping.

Finally, the gradient projection algorithm can be summarized as

$$x^{(k+1)} \leftarrow \pi_\Omega \circ T(x^{(k)}).$$

By the previous exercise we saw that π_Ω is a non-expansive map, and so $\pi_\Omega \circ T$ is also a contraction mapping. The algorithm then converges by the Banach fixed point theorem.