Exercise 5

February 13, 2024

Problem 1.

Consider the sets $\Omega_1 = \{x \in \mathbb{R}^d : \|x\|_\infty \leq 1\}$ and $\Omega_2 = \{x \in \mathbb{R}^d : \|x\|_2 \leq 1\}$.

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

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 $(1, 1)$.

c) Show that the projection π_{Ω_2} onto Ω_2 is explicitly given as

$$\pi_{\Omega_2}(z) = \begin{cases} \dfrac{z}{\|z\|_2} & \text{if } \|z\|_2 > 1, \\ z & \text{else.} \end{cases}$$

d) Show that the projection π_{Ω_1} onto Ω_1 is explicitly given as

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

e) Consider now the case $d = 2$ and let $f(x) = x_1^2 + (x_2 + 2)^2$. Find the global solution of the problem $\min_{x\in\Omega_2} f(x)$.

Solution.

a) Evidently, $0 \in \Omega_1,\ \Omega_2$. Therefore, Ω_1 and Ω_2 are non-empty. Moreover, closedness follows immediately from the continuity of the norm. Let $x_1, x_2 \in \Omega_i,\ i = 1, 2$ be arbitrary and $\lambda \in [0, 1]$. Then we get

$$\|\lambda x_1 + (1 - \lambda)x_2\| \leq \lambda\|x_1\| + (1 - \lambda)\|x_2\| \leq 1,$$

where $\|\cdot\|$ is any of the considered norms, which implies that $\lambda x_1 + (1 - \lambda)x_2 \in \Omega_i,\ i = 1, 2$. Therefore, Ω_1 and Ω_2 are convex sets. □

b) We can rewrite the set Ω_1 in terms of the following four linear inequality

constraints

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

It is evident that, at the point $\hat{x} = (\hat{x}_1, \hat{x}_2) = (1, 0)$, only the inequality constraint $c_1(\hat{x})$ is active, and $\nabla c_1(\hat{x}) = (-1, 0)^T$. Therefore, the cone of linearised feasible directions at $\hat{x}$ is given as

$$F(\hat{x}) = \big\{d = (d_1, d_2) \in \mathbb{R}^2 \colon \langle d, \nabla c_1(\hat{x})\rangle \geq 0\big\},$$

which gives $F(\hat{x}) = \{d = (d_1, d_2) \in \mathbb{R}^2 \colon d_1 \leq 0\}$. Since we only have linear constraints, the tangent cone coincides with the cone of linearised feasible directions, and thus $F(\hat{x}) = T_{\Omega_1}(\hat{x})$. Further,

$$\begin{aligned} N_{\Omega_1}(\hat{x}) &= \{p = (p_1, p_2) \in \mathbb{R}^2 \colon \langle p, q\rangle \leq 0,\ \forall\ q \in T_{\Omega_1}(\hat{x})\} \\ &= \{p = (p_1, p_2) \in \mathbb{R}^2 \colon p_1 \geq 0 \text{ and } p_2 = 0\}. \end{aligned}$$

For the set Ω_2, the inequality constraint $c(x) = 1 - x_1^2 - x_2^2 \geq 0$ is active at $\hat{x} = (1, 0)$, and $\nabla c(\hat{x}) = (-2, 0)^T$. Therefore, the cone of linearized feasible directions at $\hat{x}$ is given as

$$\begin{aligned} F(\hat{x}) &= \{d = (d_1, d_2) \in \mathbb{R}^2 \colon \langle d, \nabla c(\hat{x})\rangle \geq 0\} \\ &= \{d = (d_1, d_2) \in \mathbb{R}^2 \colon d_1 \leq 0\}. \end{aligned}$$

Since c is concave and $c(0, 0) = 1 > 0$, Slater's constraint qualification holds. Thus $F(\hat{x}) = T_{\Omega_2}(\hat{x})$. Furthermore,

$$\begin{aligned} N_{\Omega_2}(\hat{x}) &= \{p = (p_1, p_2) \in \mathbb{R}^2 \colon \langle p, q\rangle \leq 0,\ \forall\ q \in T_{\Omega_2}(\hat{x})\} \\ &= \{p = (p_1, p_2) \in \mathbb{R}^2 \colon p_1 \geq 0 \text{ and } p_2 = 0\}. \end{aligned}$$

For the last part, we see that the inequality constraints $c_1(x)$ and $c_3(x)$ are active at $\hat{x} = (1, 1)$, and $\nabla c_1(\hat{x}) = (-1, 0)^T,\ \nabla c_3(\hat{x}) = (0, -1)^T$. Therefore, the cone of linearized feasible directions at $\hat{x}$ is given as

$$F(\hat{x}) = \{d = (d_1, d_2) \in \mathbb{R}^2 \colon \langle d, \nabla c_1(\hat{x})\rangle \geq 0 \text{ and } \langle d, \nabla c_3(\hat{x})\rangle \geq 0\},$$

which gives $F(\hat{x}) = \{d = (d_1, d_2) \in \mathbb{R}^2 \colon d_1,\ d_2 \leq 0\}$. Since we only have linear constraints, this coincides with the tangent cone. Further,

$$\begin{aligned} N_{\Omega_1}(\hat{x}) &= \{p = (p_1, p_2) \in \mathbb{R}^2 \colon \langle p, q\rangle \leq 0,\ \forall\ q \in T_{\Omega_1}(\hat{x})\} \\ &= \{p = (p_1, p_2) \in \mathbb{R}^2 \colon p_1, p_2 \geq 0\}. \end{aligned}$$

c) If $\|z\|_2 \leq 1$, then we already have that $z \in \Omega_2$. Thus the projection of z is equal to z.

Now assume that $\|z\|_2 > 1$ and let $x \in \Omega_2$, that is, $\|x\|_2 \leq 1$. Then

$$\left\langle \frac{z}{\|z\|_2} - z, x - \frac{z}{\|z\|_2} \right\rangle = \frac{\langle z, x\rangle}{\|z\|_2} - \langle z, x\rangle - 1 + \|z\|_2 = \left(\|z\|_2 - 1\right)\left(1 + \frac{\langle z, x\rangle}{\|z\|_2^2}\right).$$

Since $\|z\|_2 > 1$, the first term in the last product is positive. Moreover, the Cauchy–Schwarz inequality implies that

$$\langle z, x\rangle \geq -\|z\|\|x\|_2.$$

Since $\|x\|_2 \leq 1 < \|z\|_2$, it follows that

$$1 + \frac{\langle z, x\rangle}{\|z\|_2^2} \geq 1 - \frac{\|x\|_2}{\|z\|_2} > 0,$$

that is, the second term is positive as well. Together, we have thus shown that

$$\left\langle \frac{z}{\|z\|_2} - z, x - \frac{z}{\|z\|_2} \right\rangle > 0$$

for all $x \in \Omega_2$. Since this is precisely the characterisation of the projection onto Ω_2, this shows that $\pi_{\Omega_2}(z) = z/\|z\|_2$.

d) We first look at the case where $z \in \Omega_1$. Then we look at the i'th component of z_1. Then we get first that $\min\{+1, z_i\} = z_i$ and that $\max\{-1, z_i\} = z_i$. Since this holds for an arbitrary component, we can conclude that $\pi_{\Omega_1}(z) = z$.

Next, consider the case $z \notin \Omega_1$. Firstly we can see from the definition of π_{Ω_1}, that all components of the suggested projection must be in the interval $[-1, 1]$, so the projection will clearly land within Ω_1. Let $x \in \Omega_2$. Then, we need to show that

$$\left\langle \pi_{\Omega_1}(z) - z, x - \pi_{\Omega_1}(z)\right\rangle = \sum_{i=1}^{d} (\pi_{\Omega_1}(z)_i - z_i)(x_i - \pi_{\Omega_1}(z)_i) \qquad (1)$$

is non-negative. We look at each individual term in the sum and note that there are three possibilities:

- $z_i < -1$: Then $\pi_{\Omega_1}(z)_i - z_i > 0$ and $x_i - \pi_{\Omega_1}(z)_i \geq 0$.
- $z_i > 1$: Then $\pi_{\Omega_1}(z)_i - z_i < 0$ and $x_i - \pi_{\Omega_1}(z)_i \leq 0$.
- $-1 \leq z_i \leq -1$: Then $\pi_{\Omega_1}(z)_i - z_i = 0$

Thus, every term in the sum on the right of (1) will be non-negative, which gives that the conditions from proposition 2.3 in the lecture note on convex optimization are satisfied. So we can conclude that π_{Ω_1} is indeed the projection onto Ω_1.

e) If we compare the objective function $f(x)$ with the optimization problem (5) of the lecture note on convex optimization, we can say that we have to find a point in Ω_2 (global solution) which is closest to $z = (0, -2)$. This is evidently the projection of $z = (0, -2)$ on to Ω_2, which is $x^* = (0, -1)$ by part c).

Problem 2. (cf. N&W, Exercise 12.4)

Assume that $f\colon \mathbb{R}^d \to \mathbb{R}$ is convex and that $\Omega \subset \mathbb{R}^d$ is convex and non-empty. Show that every local solution of the problem $\min_{x\in\Omega} f(x)$ is also a global solution. In addition, show that the set of global solutions of this problem is convex.

Solution.

Let $x_0 \in \Omega$ be a local solution. It follows that there exists some $\varepsilon > 0$ such that

$$f(x_0) \leq f(x) \text{ for all } x \in \Omega \text{ with } \|x - x_0\|_2 \leq \varepsilon. \tag{2}$$

Now, we have to show that x_0 is a global solution too. We assume to the contrary that x_0 is not a global solution. It follows that there exists $x \in \Omega$ such that

$$f(x) < f(x_0). \tag{3}$$

Since Ω is convex, it follows that for all $\lambda \in [0, 1]$ we have that $\lambda x_0 + (1-\lambda)x \in \Omega$. Convexity of the objective function f implies

$$f(\lambda x_0 + (1 - \lambda)x) \leq \lambda f(x_0) + (1 - \lambda)f(x). \tag{4}$$

By combining inequalities (3) and (4), we obtain

$$f(\lambda x_0 + (1 - \lambda)x) < f(x_0). \tag{5}$$

Notice that by selecting $1 > \lambda > 1 - \frac{\varepsilon}{\|x_0 - x\|_2}$, we get that $\|\lambda x_0 + (1 - \lambda)x - x_0\|_2 < \varepsilon$. When we combine this choice of λ with (5), we get a contradiction to the inequality (2). Therefore, x_0 is a global solution.

For the next part, we denote by S the set of global minimizers. Now, we have to show that S is convex. Let $x_1, x_2 \in S$ be arbitrary. We have

$$f(x_1) \leq f(x) \text{ and } f(x_2) \leq f(x) \qquad \text{for all } x \in \Omega. \tag{6}$$

Since, Ω is convex, $\lambda x_1 + (1 - \lambda)x_2 \in \Omega$ for all $\lambda \in [0, 1]$. The convexity of the function f yields

$$f(\lambda x_1 + (1 - \lambda)x_2) \leq \lambda f(x_1) + (1 - \lambda)f(x_2). \tag{7}$$

Inequalities (6) and (7) imply

$$f(\lambda x_1 + (1-\lambda)x_2) \leq f(x) \text{ for all } x \in \Omega,$$

which implies that $\lambda x_1 + (1-\lambda)x_2 \in S$. Therefore, S is a convex set. □

Problem 3.

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

$$\|\pi_\Omega(x) - \pi_\Omega(y)\|_2 \leq \|x - y\|_2 \quad \text{for all } x,\, y \in \mathbb{R}^d.$$

Solution.

Let $z_1, z_2 \in \mathbb{R}^d$ be arbitrary. The variational characterization of the projection operator on to Ω yields that

$$\langle z_1 - \pi_\Omega(z_1), x - \pi_\Omega(z_1)\rangle \leq 0, \ \forall\ x \in \Omega.$$

Since $\pi_\Omega(z_2) \in \Omega$, we can in particular use this with $x = \pi_\Omega(z_2)$, which yields

$$\langle z_1 - \pi_\Omega(z_1), \pi_\Omega(z_2) - \pi_\Omega(z_1)\rangle \leq 0. \tag{8}$$

By interchanging z_1 and z_2 in inequality (8), we obtain

$$\langle z_2 - \pi_\Omega(z_2), \pi_\Omega(z_1) - \pi_\Omega(z_2)\rangle \leq 0,$$

which can be rewritten as

$$\langle \pi_\Omega(z_2) - z_2, \pi_\Omega(z_2) - \pi_\Omega(z_1)\rangle \leq 0. \tag{9}$$

By adding inequalities (8) and (9), we get

$$\langle z_1 - z_2 + \pi_\Omega(z_2) - \pi_\Omega(z_1), \pi_\Omega(z_2) - \pi_\Omega(z_1)\rangle \leq 0.$$

We can rearrange this inequality as

$$\|\pi_\Omega(z_2) - \pi_\Omega(z_1)\|_2^2 \leq \langle z_2 - z_1, \pi_\Omega(z_2) - \pi_\Omega(z_1)\rangle.$$

By applying the Cauchy-Schwarz inequality, we further obtain that

$$\|\pi_\Omega(z_2) - \pi_\Omega(z_1)\|_2^2 \leq \|z_2 - z_1\|_2 \|\pi_\Omega(z_2) - \pi_\Omega(z_1)\|_2,$$

which implies

$$\|\pi_\Omega(z_2) - \pi_\Omega(z_1)\|_2 \leq \|z_2 - z_1\|_2.$$

Since $z_1, z_2 \in \mathbb{R}^d$ were arbitrary, we have that

$$\|\pi_\Omega(z_2) - \pi_\Omega(z_1)\|_2 \leq \|z_2 - z_1\|_2 \quad \text{for all } z_1, z_2 \in \mathbb{R}^d.$$

□

Problem 4.

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

$$\min_{x\in\Omega} f(x) \qquad \text{with } f(x) = \frac{1}{2}\|Ax - b\|_2^2 \tag{10}$$

and the gradient projection algorithm (without line search)

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

Show that this algorithm converges to the unique solution of (10) 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 $\mathbb{R}^d$, and then use the result of the previous exercise and Banach's fixed point theorem.

Solution.

We note first that the gradient of f is given as

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

Define now the mapping $G\colon \mathbb{R}^d \to \mathbb{R}^d$,

$$G(x) := x - \alpha A^T(Ax - b),$$

which is just the result of one gradient descent step for f with step length α. We first want to show that G is a contraction. We have

$$\|G(x)-G(y)\| = \|x-\alpha A^T(Ax-b)-y+\alpha A^T(Ay-b)\| = \|(I-\alpha A^TA)(x-y)\|,$$

where $I \in \mathbb{R}^{d\times d}$ is the d-dimensional identity matrix. Next we note that the eigenvalues of the matrix $I - \alpha A^TA$ are the values $1 - \alpha\sigma_i^2$, where $0 < \sigma_1 \leq \sigma_2 \leq \ldots \leq \sigma_d$ are the non-zero singular values of A. (Since A has rank d, it has d non-zero singular values.) In particular, the smallest eigenvalue of $I-\alpha A^TA$ is $1-\alpha\sigma_d^2$, which is larger than -1, since we had chosen $\alpha < 2/\sigma_d^2$. On the other hand, the largest eigenvalue is $1 - \alpha\sigma_1^2$, which is smaller than $+1$, since $\alpha > 0$ and $\sigma_1 > 0$. Since $I - \alpha A^TA$ is symmetric, its singular values are precisely the absolute values of its eigenvalues. Thus we obtain from the previous considerations that all singular values of $I - \alpha A^TA$ are strictly smaller than 1. Since the spectral norm of a matrix is its largest singular value, it follows that

$$\|I - \alpha A^TA\|_2 =: L < 1.$$

Thus

$$\|G(x) - G(y)\| = \|(I - \alpha A^T A)(x - y)\| \leq L\|x - y\|$$

with $0 < L < 1$, which shows that G is a contraction.

Using the previous exercise we now have that

$$\|\pi_\Omega(G(x)) - \pi_\Omega(G(y))\| \leq \|G(x) - G(y)\| \leq L\|x - y\|,$$

which shows that the projected gradient descent algorithm is generated by a contraction as well. From the Banach fixed point theorem we now obtain that the iteration $x^{(k)}$ converges to the unique point x^* satisfying

$$x^* = \pi_\Omega\big(x^* - \alpha A^T(Ax^* - b)\big),$$

which is precisely the solution of the optimisation problem (10) (as discussed in the lecture/the note on optimisation with convex constraints).

Problem 5.

Consider the set

$$\Omega := \big\{(x, y) \in \mathbb{R}^2 : y \geq 0 \text{ and } x^2(x + 1) - y \geq 0\big\}.$$

Determine the tangent cone and the cone of linearised feasible directions to Ω at the points $(x, y) = (-1, 0)$, $\big(-\frac{2}{3}, 0\big)$, and $(0, 0)$.

Solution.

We denote the feasible set by Ω, and the inequality constraints by $c_1(x, y) = y \geq 0$ and $c_2(x, y) = x^2(x + 1) - y \geq 0$. It is evident that, at $(\hat{x}, \hat{y}) = (-1, 0)$, both the inequality constraints $c_1(\hat{x}, \hat{y})$ and $c_2(\hat{x}, \hat{y})$ are active, and $\nabla c_1(\hat{x}, \hat{y}) = (0, 1)^T$, $\nabla c_2(\hat{x}, \hat{y}) = (1, -1)^T$. Therefore, the cone of linearised feasible directions at $(\hat{x}, \hat{y})$ is given as

$$F(\hat{x}, \hat{y}) = \{d = (d_1, d_2) \in \mathbb{R}^2 \colon \langle d, \nabla c_1(\hat{x}, \hat{y})\rangle \geq 0 \text{ and } \langle d, \nabla c_2(\hat{x}, \hat{y})\rangle \geq 0\},$$

which gives $F(\hat{x}, \hat{y}) = \{d = (d_1, d_2) \in \mathbb{R}^2 \colon d_1 \geq d_2 \geq 0\}$. Moreover, the set of active constraint gradients $\{\nabla c_1(\hat{x}, \hat{y}), \nabla c_2(\hat{x}, \hat{y})\}$ is linearly independent. Therefore the LICQ and thus $F(\hat{x}, \hat{y}) = T_\Omega(\hat{x}, \hat{y}) = \{d = (d_1, d_2) \in \mathbb{R}^2 \colon d_1 \geq d_2 \geq 0\}$.

It is evident that at $(\hat{x}, \hat{y}) = (-\frac{2}{3}, 0)$, only the inequality constraint $c_1(\hat{x}, \hat{y})$ is active, and $\nabla c_1(\hat{x}, \hat{y}) = (0, 1)^T$. Therefore, the cone of linearised feasible directions at $(\hat{x}, \hat{y})$ is given as

$$F(\hat{x}, \hat{y}) = \{d = (d_1, d_2) \in \mathbb{R}^2 \colon \langle d, \nabla c_1(\hat{x}, \hat{y})\rangle \geq 0\},$$

which gives

$$F(\hat{x},\hat{y}) = \{d = (d_1, d_2) \in \mathbb{R}^2 \colon d_2 \geq 0\}. \tag{11}$$

Moreover, the set of active constraint gradients $\{\nabla c_1(\hat{x},\hat{y})\}$ is linearly independent. Therefore the LICQ holds and $F(\hat{x},\hat{y}) = T_\Omega(\hat{x},\hat{y}) = \{d = (d_1, d_2) \in \mathbb{R}^2 \colon d_2 \geq 0\}$.

It is evident that at $(\hat{x},\hat{y}) = (0,0)$, both the inequality constraints $c_1(\hat{x},\hat{y})$ and $c_2(\hat{x},\hat{y})$ are active, and $\nabla c_1(\hat{x},\hat{y}) = (0,1)^T$, $\nabla c_2(\hat{x},\hat{y}) = (0,-1)^T$. Therefore, the cone of linearised feasible directions at $(\hat{x},\hat{y})$ is given as

$$F(\hat{x},\hat{y}) = \{d = (d_1, d_2) \in \mathbb{R}^2 \colon \langle d, \nabla c_1(\hat{x},\hat{y})\rangle \geq 0 \text{ and } \langle d, \nabla c_2(\hat{x},\hat{y})\rangle \geq 0\},$$

which gives $F(\hat{x},\hat{y}) = \{d = (d_1, d_2) \in \mathbb{R}^2 \colon d_2 = 0\}$. Moreover, the set of active constraint gradients $\{\nabla c_1(\hat{x},\hat{y}), \nabla c_2(\hat{x},\hat{y})\}$ is linearly dependent. Thus the LICQ does not hold and we cannot be sure whether $F(\hat{x},\hat{y}) = T_\Omega(\hat{x},\hat{y})$. Therefore, we have to find the tangent cone $T_\Omega(\hat{x},\hat{y})$ by other means, for instance by using its definition. For that, we consider that $z_k = (p_k, q_k) = \left(\pm\frac{1}{k}, 0\right)$ is a feasible sequence, which clearly converges to $(0,0)$, and choose $t_k = \frac{\tau}{k}$ for some $\tau > 0$. Then,

$$\lim_{k\to\infty} \frac{z_k - (0,0)}{t_k} = \left(\pm\frac{1}{\tau}, 0\right) \in T_\Omega(0,0). \tag{12}$$

Clearly, the direction $(0,0)$ is in $T_\Omega(0,0)$. Since (12) holds for all τ, it follows that the cone of linearised feasible directions as computed in (11) is contained in $T_\Omega(0,0)$. Since the opposite inclusion $T_\Omega(0,0) \subset F(0,0)$ always holds, it follows that

$$T_\Omega(0,0) = F(0,0) = \{d = (d_1, d_2) \in \mathbb{R}^2 \colon d_2 = 0\}.$$