Exercise #5

February 14, 2023

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) Consider now the case $d = 2$ and let $f(x) = x_1^2 + (x_2 + 2)^2$. Find the global solution of the problem $\min_{x \in \Omega_2} f(x)$. Also, perform one step of the gradient projection method with the step length $\alpha = \frac{1}{2}$ and initial point $x^0 = (1, 1)$.

Solution.

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

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

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 the following four smooth inequality constraints

$$c_1(x) = -x_1 + 1 \geq 0,$$

$$c_2(x) = x_1 + 1 \geq 0,$$

$$c_3(x) = -x_2 + 1 \geq 0,$$

$$c_4(x) = x_2 + 1 \geq 0.$$

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 linearized feasible directions at $\hat{x}$ is defined as

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

which gives $F(\hat{x}) = \{d = (d_1, d_2) \in \mathbb{R}^2 \colon d_1 \le 0\}$. Moreover, the set of active constraint gradient $\{\nabla c_1(\hat{x})\}$ is linearly independent. Therefore LICQ (linear independence constraint qualification) holds. Now, Lemma 12.2 (N&W Book) implies that $F(\hat{x}) = T_{\Omega_1}(\hat{x}) = \{d = (d_1, d_2) \in \mathbb{R}^2 \colon d_1 \le 0\}$ (or we can use the Lemma 12.7 (N&W Book)). Further, $N_{\Omega_1}(\hat{x}) = \{p = (p_1, p_2) \in \mathbb{R}^2 \colon p^T q \le 0,\ \forall\, q \in T_{\Omega_1}(\hat{x})\} = \{p = (p_1, p_2) \in \mathbb{R}^2 \colon p_1 \ge 0 \text{ and } p_2 = 0\}$.
For the set Ω_2, the inequality constraint $c(x) = 1 - x_1^2 - x_2^2 \ge 0$ 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 defined as

$$F(\hat{x}) = \{d = (d_1, d_2) \in \mathbb{R}^2 \colon d^T \nabla c(\hat{x}) \ge 0\},$$

which gives $F(\hat{x}) = \{d = (d_1, d_2) \in \mathbb{R}^2 \colon d_1 \le 0\}$. Moreover, the set of active constraint gradient $\{\nabla c(\hat{x})\}$ is linearly independent. Therefore LICQ (linear independence constraint qualification) holds. Now, Lemma 12.2 (N&W Book) implies that $F(\hat{x}) = T_{\Omega_2}(\hat{x}) = \{d = (d_1, d_2) \in \mathbb{R}^2 \colon d_1 \le 0\}$. Further, $N_{\Omega_2}(\hat{x}) = \{p = (p_1, p_2) \in \mathbb{R}^2 \colon p^T q \le 0,\ \forall\, q \in T_{\Omega_2}(\hat{x})\} = \{p = (p_1, p_2) \in \mathbb{R}^2 \colon p_1 \ge 0 \text{ and } p_2 = 0\}$. The conclusion is $T_{\Omega_1}(\hat{x}) = T_{\Omega_2}(\hat{x})$ and $N_{\Omega_1}(\hat{x}) = N_{\Omega_2}(\hat{x})$.

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 defined as

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

which gives $F(\hat{x}) = \{d = (d_1, d_2) \in \mathbb{R}^2 \colon d_1,\ d_2 \le 0\}$. Moreover, the set of active constraints gradient $\{\nabla c_1(\hat{x}), \nabla c_3(\hat{x})\}$ is linearly independent. Therefore LICQ (linear independence constraint qualification) holds. Now, Lemma 12.2 (N&W Book) implies that $F(\hat{x}) = T_{\Omega_1}(\hat{x}) = \{d = (d_1, d_2) \in \mathbb{R}^2 \colon d_1, d_2 \le 0\}$ (or we can use the Lemma 12.7 (N&W Book)). Further, $N_{\Omega_1}(\hat{x}) = \{p = (p_1, p_2) \in \mathbb{R}^2 \colon p^T q \le 0,\ \forall\, q \in T_{\Omega_1}(\hat{x})\} = \{p = (p_1, p_2) \in \mathbb{R}^2 \colon p_1, p_2 \ge 0\}$.

c) If $\|z\|_2 \le 1$, 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, $\|x\|_2 \le 1$. Then

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

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

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

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

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

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

d) 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 $x^* = (0, -1)$ (this is quite easy to understand if you sketch Ω_2).

We now perform the gradient projection method for the step length $\alpha = \frac{1}{2}$ and initial point $x^0 = (1, 1)$. We have $\nabla f(x^0) = (2x_1^0, 2(x_2^0 + 2))^T = (2, 6)^T$. By gradient projection method, we have

$$x^1 = \pi_{\Omega_2}(x^0 - \alpha \nabla f(x^0)) = \pi_{\Omega_2}(0, -2) = (0, -1).$$

Therefore, the gradient projection method converges in one step.

Problem 2.

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

$$\|\pi_\Omega(x) - \pi_\Omega(y)\|_2 \leq \|x - y\|_2 \;\forall\, x, y \in \mathbb{R}^d.$$

Solution.

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

$$\langle z_1 - \pi_\Omega(z_1), x - \pi_\Omega(z_1)\rangle \leq 0,\; \forall\, x \in \Omega. \tag{1}$$

Since $\pi_\Omega(z_2) \in \Omega$, the inequality (1) renders

$$\langle z_1 - \pi_\Omega(z_1), \pi_\Omega(z_2) - \pi_\Omega(z_1)\rangle \leq 0. \tag{2}$$

By interchanging z_1 and z_2 in inequality (2), 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

$$\langle \pi_\Omega(z_2) - z_2, \pi_\Omega(z_2) - \pi_\Omega(z_1)\rangle \leq 0. \tag{3}$$

By adding inequalities (2) and (3), we get

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

By rearranging the above inequality implies

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

By applying Cauchy-Schwartz inequality property in the above inequality, we get

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

which implies

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

Since $z_1, z_2 \in \mathbb{R}^d$ are arbitrary,

$$\|\pi_\Omega(z_2) - \pi_\Omega(z_1)\|_2 \leq \|z_2 - z_1\|_2,\; \forall\, z_1, z_2 \in \mathbb{R}^d.$$

□

Problem 3.

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

$$\min_{x\in\Omega} f(x) \qquad \text{with } f(x) = \frac{1}{2}\|Ax - b\|_2^2 \tag{4}$$

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 (4) 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

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

where $I \in \mathbb{R}^{d\times d}$ is the d-dimensional identity matrix. Next we note that the eigenvalues of the matrix $I - \alpha A^T A$ are the values $1 - \alpha\sigma_i^2$, where $0 < \sigma_1 \le \sigma_2 \le \ldots \le \sigma_d$ 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 $I - \alpha A^T A$ 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 $I - \alpha A^T A$ 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 $I - \alpha A^T A$ are strictly smaller than 1. Since the spectral norm of a matrix is its largest singular value, it follows that

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

Thus

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

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

Using the previous exercise we now have that

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

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 (4) (as discussed in the lecture/the note on optimisation with convex constraints).

Problem 4. (Exercise 12.4, N&W Book)

If $f\colon \mathbb{R}^d \mapsto \mathbb{R}$ is convex and the feasible region Ω is convex, show that local solutions of the problem $\min_{x\in\Omega} f(x)$ are also global solutions. Show that the set of global solutions is convex.

Solution.

Let $x_0 \in \Omega$ be a local solution. It follows that there exists a neighborhood of x_0, $N(x_0)$ such that

$$f(x_0) \le f(x),\ \forall\, x \in N(x_0) \cap \Omega. \tag{5}$$

Now, we have to show that x_0 is global solution too. We assume to the contrary that x_0 is not global solution, it follows that there exists $x \in N(x_0) \cap \Omega$ such that

$$f(x) < f(x_0). \tag{6}$$

Since, $N(x_0)$ and Ω are convex, $N(x_0) \cap \Omega$ is convex. Thus, for all $\lambda \in [0,1]$, $\lambda x_0 + (1-\lambda)x \in N(x_0) \cap \Omega$. Convexity of the objective function f implies

$$f(\lambda x_0 + (1-\lambda)x) \le \lambda f(x_0) + (1-\lambda) f(x). \tag{7}$$

By combining inequalities (6) and (7), we obtain

$$f(\lambda x_0 + (1-\lambda)x) < f(x_0),$$

which contradicts the inequality (5). Therefore, x_0 is the global solution.

For the next part, we consider S is 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

$$f(x_1) \leq f(x) \text{ and } f(x_2) \leq f(x), \; \forall\, x \in \Omega. \tag{8}$$

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

$$f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda) f(x_2). \tag{9}$$

Inequalities (8) and (9) imply

$$f(\lambda x_1 + (1-\lambda)x_2) \leq f(x), \; \forall\, x \in \Omega,$$

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

Problem 5.

Consider the set

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

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

Solution.

We denote the feasible set by Ω, and 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 linearized feasible directions at $(\hat{x},\hat{y})$ is defined as

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

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 constraints gradient $\{\nabla c_1(\hat{x},\hat{y}), \nabla c_2(\hat{x},\hat{y})\}$ is linearly independent. Therefore LICQ (linear independence constraint qualification) holds. Now, Lemma 12.2 (N&W Book) implies that $F(\hat{x},\hat{y}) = T_\Omega(\hat{x},\hat{y}) = \{d = (d_1,d_2) \in \mathbb{R}^2 \colon d_1 \geq d_2 \geq 0\}$.

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 linearized feasible directions at $(\hat{x},\hat{y})$ is defined as

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

which gives $F(\hat{x},\hat{y}) = \{d = (d_1,d_2) \in \mathbb{R}^2 \colon d_2 \geq 0\}$. Moreover, the set of active constraints gradient $\{\nabla c_1(\hat{x},\hat{y})\}$ is linearly independent. Therefore LICQ (linear independence constraint qualification) holds. Now, Lemma 12.2 (N&W Book) implies that $F(\hat{x},\hat{y}) = T_\Omega(\hat{x},\hat{y}) = \{d = (d_1,d_2) \in \mathbb{R}^2 \colon d_2 \geq 0\}$.

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 linearized feasible directions at $(\hat{x},\hat{y})$ is defined as

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

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 constraints gradient $\{\nabla c_1(\hat{x},\hat{y}), \nabla c_2(\hat{x},\hat{y})\}$ is linearly dependent. Thus LICQ (linear independence constraint qualification) does not hold. Now, we cannot be sure that $F(\hat{x}, \hat{y}) = T_\Omega(\hat{x}, \hat{y})$. Therefore, we have to find the tangent cone $T_\Omega(\hat{x}, \hat{y})$ by 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 $t_k = \frac{\tau}{k}$ for $\tau > 0$. Then,

$$\lim_{k\to\infty} \frac{z_k - (0,0)}{t_k} = \left(\pm\frac{1}{\tau}, 0\right) \in T_\Omega(0,0). \tag{10}$$

Clearly, the point $(0, 0) \in \Omega$, and the Lemma 12.2 (N&W Book) implies $T_\Omega(0, 0) \subset F(0, 0)$. Moreover, for the feasible sequence $z_k = (0, 0)$,

$$\lim_{k\to\infty} \frac{z_k - (0,0)}{t_k} = (0,0) \in T_\Omega(0,0). \tag{11}$$

Now, by (10) and (11), we can say that $F(0, 0) \subset T_\Omega(0, 0)$. Finally, we have $T_\Omega(0, 0) = F(0, 0) = \{d = (d_1, d_2) \in \mathbb{R}^2 : d_2 = 0\}$.