The Chinese University of Hong Kong, Shenzhen
SDS · School of Data Science

Andre Milzarek · Fall Semester 2020/21

MDS 6106 – Introduction to Optimization

Solutions 2

Exercise E2.1 (Optimization Problem I):
Consider the function $f_\alpha : \mathbb{R}^2 \to \mathbb{R}$,

$$f_\alpha(x) := \alpha x_1^2 + x_2^2 - 2x_1x_2 - 2x_2,$$

where $\alpha \in \mathbb{R}$ is a scalar.

a) Find the stationary points (in case they exist) of f_α for each value of α.

b) For each stationary point x^* in part a), determine whether x^* is a local maximizer or a local minimizer or a saddle point of f_α.

c) For which values of α can f_α have a global minimizer?

Solution :

a) The gradient and Hessian of f_α are given by

$$\nabla f_\alpha(x) = \begin{pmatrix} 2\alpha x_1 - 2x_2 \\ 2x_2 - 2x_1 - 2 \end{pmatrix}, \quad \nabla^2 f_\alpha(x) = \begin{pmatrix} 2\alpha & -2 \\ -2 & 2 \end{pmatrix}$$

and it holds that

$$\nabla f_\alpha(x) = 0 \quad \iff \quad x_2 = x_1 + 1 \quad \text{and} \quad 2\alpha x_1 - 2x_1 = 2$$

which implies $(\alpha - 1)x_1 = 1$. This equation only has a solution if $\alpha \neq 1$. In this case, we obtain

$$x_1^* = \frac{1}{\alpha - 1}, \quad x_2^* = 1 + \frac{1}{\alpha - 1} = \frac{\alpha}{\alpha - 1}.$$

This is also the unique stationary point of f_α.

b) We have

$$\operatorname{tr}(\nabla^2 f_\alpha(x)) = 2\alpha + 2 = 2(\alpha + 1), \quad \det(\nabla^2 f_\alpha(x)) = 4\alpha - 4 = 4(\alpha - 1) \quad (\forall\ x).$$

Consequently, if $\alpha < 1$, the Hessian is indefinite and x^* is a saddle point. If $\alpha > 1$, then $\nabla^2 f_\alpha(x)$ is positive definite and x^* is a local minimizer of f_α.

c) The function f_α can only have a global minimizer if $\alpha > 1$. In the case $\alpha \leq 1$, f_α is unbounded and it does not possess a global minimizer (all stationary points are saddle points). Moreover, since the mapping f_α is quadratic (i.e., the Hessian $\nabla^2 f_\alpha(x)$ does not depend on x), the positive definiteness of the Hessian implies that f_α is strongly convex for all $\alpha > 1$. Hence, x^* is the unique global solution of $\min_x f_\alpha(x)$ in this situation.

Exercise E2.2 (Optimization Problem II):
We consider the optimization problem

$$\min_{x\in\mathbb{R}^2} \; f(x) = \frac{1}{2}x_1^2x_2^2 + \frac{1}{2}x_1^2 + \frac{1}{2}x_2^2 - 2x_1.$$

a) Is the function f coercive?

b) Calculate the gradient and Hessian of f and determine all stationary points of f.

c) Show that f has a unique global minimizer.

d) Is the mapping f convex?

Solution :

a) Yes, the function f is coercive. To see this, we use the estimate

$$f(x) \geq \frac{1}{2}(x_1^2 - 4x_1 + 4) + \frac{1}{2}x_2^2 - 2 = \frac{1}{2}(x_1 - 2)^2 + \frac{1}{2}x_2^2 - 2.$$

Obviously the latter lower bound converges to $+\infty$ if either $|x_1| \to \infty$ or $|x_2| \to \infty$. This establishes coercivity of f.

b) It holds that

$$\nabla f(x) = \begin{pmatrix} x_1x_2^2 + x_1 - 2 \\ x_1^2x_2 + x_2 \end{pmatrix}, \quad \nabla^2 f(x) = \begin{pmatrix} x_2^2 + 1 & 2x_1x_2 \\ 2x_1x_2 & x_2^2 + 1 \end{pmatrix}.$$

Furthermore, we have $\nabla f(x) = 0$ if and only if $x_2(1 + x_1^2) = 0$ which implies $x_2 = 0$ and $x_1 = 2$. Thus, $x^* = (2, 0)^\top$ is the single stationary point of f.

c) Since f is coercive, it possesses a global solution. Every global solution of $\min_x f(x)$ is a stationary point of f. Since f has only one stationary point, this implies that x^* is the unique global solution of the problem $\min_x f(x)$.

d) We have $\mathrm{tr}(\nabla^2 f(x)) = x_1^2 + x_2^2 + 2 \geq 0$ and

$$\det(\nabla^2 f(x)) = (x_1^2 + 1)(x_2^2 + 1) - 4x_1^2x_2^2 = -3x_1^2x_2^2 + x_1^2 + x_2^2 + 1.$$

Consequently, setting $x_1 = x_2 = \sqrt{2}$, we obtain $\det(\nabla^2 f(x)) = -3 \cdot 4 + 2 + 2 + 1 = -7$ which shows that $\nabla^2 f(x)$ is indefinite at $x = (\sqrt{2}, \sqrt{2})^\top$. Hence, f can not be convex on $\mathbb{R}^2$.

Exercise E2.3 (Convex Sets and Convex Functions):
In this exercise, we investigate convexity of sets and functions.

a) Let $A \in \mathbb{R}^{n\times n}$ be a given symmetric and positive semidefinite matrix and consider the set $X := \{x \in \mathbb{R}^n : x^\top Ax \leq 0\}$. Show that the set X is convex.

b) Show that the function $f : \mathbb{R}^2_{++} \to \mathbb{R}$, $f(x) = x_1^2 - 2x_1x_2 + x_2^2 - \ln(x_1x_2)$ is strictly convex on $\mathbb{R}^2_{++} := \{x \in \mathbb{R}^2 : x > 0\}$.

c) Determine whether the function $f(x) = -x_1^2 - x_2^2 - 2x_3^2 + x_1x_2$ is convex or concave.

d) Show that the hyperbolic set $\{x \in \mathbb{R}^2_+ : x_1x_2 \geq 1\}$ is convex, where $\mathbb{R}^2_+ := \{x \in \mathbb{R}^2 : x \geq 0\}$.

Hint: Rewrite the condition "$x_1x_2 \geq 1$" in a suitable way.

Solution :

a) Since the matrix A is symmetric and positive semidefinite, the mapping $f(x) = x^\top Ax$ is convex. As a consequence the level set $L_{\leq 0} := \{x \in \mathbb{R}^n : f(x) \leq 0\} = X$ is convex. (See lecture L-05, slide 15).

b) We calculate the gradient and Hessian of f. It holds that

$$\nabla f(x) = \begin{pmatrix} 2x_1 - 2x_2 - \frac{1}{x_1} \\ 2x_2 - 2x_1 - \frac{1}{x_2} \end{pmatrix}, \quad \nabla^2 f(x) = \begin{pmatrix} 2 + \frac{1}{x_1^2} & -2 \\ -2 & 2 + \frac{1}{x_2^2} \end{pmatrix}$$

Furthermore, we have $\mathrm{tr}(\nabla^2 f(x)) = 4 + \frac{1}{x_1^2} + \frac{1}{x_2^2} > 4$ for all $x \in \mathbb{R}^2_{++}$ and

$$\det(\nabla^2 f(x)) = \left(2 + \frac{1}{x_1^2}\right)\left(2 + \frac{1}{x_2^2}\right) - 4 = \frac{2}{x_1^2} + \frac{2}{x_2^2} + \frac{1}{x_1^2 x_2^2} > 0$$

for all $x \in \mathbb{R}^2_{++}$. This implies that $\nabla^2 f(x)$ is positive definite for all $x \in \mathbb{R}^2_{++}$ and hence, f is (strictly) convex on $\mathbb{R}^2_{++}$.

c) As in part b), we utilize definiteness of the Hessian to determine convexity. We have

$$\nabla f(x) = \begin{pmatrix} -2x_1 + x_2 \\ -2x_2 + x_1 \\ -4x_3 \end{pmatrix}, \quad \nabla^2 f(x) = \begin{pmatrix} -2 & 1 & 0 \\ 1 & -2 & 0 \\ 0 & 0 & -4 \end{pmatrix}.$$

We calculate the eigenvalues of $\nabla^2 f(x)$:

$$\det\left(\begin{bmatrix} -2-\lambda & 1 & 0 \\ 1 & -2-\lambda & 0 \\ 0 & 0 & -4-\lambda \end{bmatrix}\right) = -(4+\lambda)[(2+\lambda)^2 - 1] = -(4+\lambda)[3 + 4\lambda + \lambda^2].$$

This yields $\lambda_1 = -4$, $\lambda_{2/3} = \frac{1}{2}(-4 \pm \sqrt{16-12}) = \frac{1}{2}(-4 \pm 2) = -3\,/\,-1$. Consequently, $\nabla^2 f(x)$ is negative definite for all x which shows that f is a concave mapping.

d) We first notice $X := \{x \in \mathbb{R}^2_+ : x_1x_2 \geq 1\} = \{x \in \mathbb{R}^2_{++} : x_1x_2 \geq 1\}$, where $\mathbb{R}^2_{++} = \{x \in \mathbb{R}^2 : x > 0\}$. Since the (natural) logarithm $\ln(\cdot)$ is monotonically increasing, the condition $x_1x_2 \geq 1$ is equivalent to

$$f(x_1, x_2) := -\ln(x_1) - \ln(x_2) \leq 0.$$

The gradient and Hessian of f are given by:

$$\nabla f(x_1, x_2) = \begin{pmatrix} -x_1^{-1} \\ -x_2^{-1} \end{pmatrix}, \quad \nabla^2 f(x_1, x_2) = \begin{pmatrix} x_1^{-2} & 0 \\ 0 & x_2^{-2} \end{pmatrix}$$

Since $\nabla^2 f$ is positive definite on the open and convex set $\mathbb{R}^2_{++}$, it follows that f is convex on $\mathbb{R}^2_{++}$. This shows that the set X is convex.

Exercise E2.4 (Multiple Choice):
Answer the following multiple choice questions and decide whether the statements are true or false. Try to give short explanations of your answer.

a) *A point is a stationary point of f if and only if it is a local minimum or maximum of f.*

☐ True. ☐ False.

b) *The point x^* is a stationary point of f if and only if $\nabla f(x^*)^\top h = 0$ for all $h \in \mathbb{R}^n \backslash \{0\}$.*

☐ True. ☐ False.

c) *Let $X \subset \mathbb{R}^n$ be nonempty and let $x^* \in X$ be a global solution of $\min_{x \in X} f(x)$. Then, it holds that $\nabla f(x^*) = 0$.*

☐ True. ☐ False.

d) *The function $f(x) = x^4$ is strongly convex.*

☐ True. ☐ False.

e) *Let $I := [a, b]$ be given with $a, b \in \mathbb{R}$, $a < b$. Suppose that $g : I \to \mathbb{R}$ is a convex function. Then, g^2 is strictly convex.*

☐ True. ☐ False.

f) *Suppose that $g : \mathbb{R} \to \mathbb{R}$ is convex but not strictly convex. Then, g has either no global minimum or infinitely many.*

☐ True. ☐ False.

g) *Suppose we have $\lambda_{\min}(\nabla^2 f(x)) > 0$ for all $x \in \mathbb{R}^n$, then f is strongly convex.*

☐ True. ☐ False.

Solution :

a) *False.* Consider the function $f(x) = x^3$. Then, the point $x^* = 0$ is a stationary point, i.e., it satisfies $f'(x^*) = 3(x^*)^2 = 0$. However, x^* is a saddle-point, i.e., it's neither a minimum nor a maximum.

b) *True.* If x^* is a stationary point then we have $\nabla f(x^*) = 0$ and the condition is obviously satisfied. On the other hand, let us assume $\nabla f(x^*)^\top h = 0$ for all $h \in \mathbb{R}^n \backslash \{0\}$. In the case $\nabla f(x^*) = 0$, there is nothing to show. Hence, let us assume $\nabla f(x*) \neq 0$. But then setting $h = \nabla f(x^*) \neq 0$ implies $\|\nabla f(x^*)\|^2 = 0$ which is a contradiction to $\nabla f(x^*) \neq 0$ and thus, this case cannot occur. Together, we can infer $\nabla f(x^*) = 0$ which finishes the proof.

c) *False.* Consider the problem $\min_{x \in X} f(x) = x$ with $X = [1, \infty)$. Obviously, $x^* = 1$ is the global solution of the this problem, but we have $f'(x^*) = 1 \neq 0$.

d) *False.* It holds that $f''(x) = 12x^2$ and $f''(0) = 0$. Hence, by Theorem 4.15, $f(x) = x^4$ cannot be strongly convex.

e) *False.* Take $I = [0, 1]$ and $g(x) = 1$. This function is convex, but $g(x)^2 = 1 = g(x)$ is not strictly convex.

f) *False.* The function $g(x) = |x|$ is convex but not strictly convex. Moreover, it has a unique global minimum at $x = 0$.

g) *False.* For strong convexity a uniform lower bound for the minimal eigenvalue of the Hessian is required. A possible counterexample is the function $f(x) = e^x$.

Assignment A2.1 (Optimization Problem): (approx. *20* points)
Consider the unconstrained optimization problem

$$\min_{x\in\mathbb{R}^2} \ f(x) = \frac{1}{3}x_1^3 - x_1\left(\frac{3}{2} + x_2^2\right) + x_2^4.$$

a) Is the function f coercive?

b) Compute the gradient and Hessian of f and calculate all stationary points.

c) For each stationary point x^* found in part c) investigate whether x^* is a local maximizer, local minimizer, or saddle point and explain your answer.

d) Does the mapping f possess any strict local or global minimizer?

Solution :

a) No, the function f is not coercive. We have

$$f(x_1, 0) = \frac{1}{6}x_1(2x_1^2 - 9) \to \pm\infty \quad \text{if} \quad x_1 \to \pm\infty.$$

b) The gradient and Hessian of f are given by

$$\nabla f(x) = \begin{pmatrix} x_1^2 - x_2^2 - \frac{3}{2} \\ -2x_1x_2 + 4x_2^3 \end{pmatrix} \quad \text{and} \quad \nabla^2 f(x) = \begin{pmatrix} 2x_1 & -2x_2 \\ -2x_2 & -2x_1 + 12x_2^2 \end{pmatrix}.$$

Moreover, it holds that $\nabla f(x) = 0$ if and only if $2x_2(2x_2^2 - x_1) = 0$. Let us first consider the case $x_2 = 0$. Then, it follows $x_1 = \pm\sqrt{3/2}$. Otherwise, we have $x_1 = 2x_2^2$ which implies $4x_2^4 - x_2^2 - \frac{3}{2} = 0$, i.e.,

$$x_2^2 = \frac{1 \pm \sqrt{1 + 6 \cdot 4}}{2 \cdot 4} = \frac{1 \pm 5}{8} = \frac{3}{4} \text{ or } -\frac{1}{2}$$

This yields $x_2 = \pm\sqrt{3}/2$ and $x_1 = 3/2$. In total, f has the following for stationary points:

$$\bar{x}_1 = \left(\frac{\sqrt{6}}{2}, 0\right), \quad \bar{x}_2 = \left(-\frac{\sqrt{6}}{2}, 0\right), \quad \bar{x}_3 = \left(\frac{3}{2}, \frac{\sqrt{3}}{2}\right), \quad \bar{x}_3 = \left(\frac{3}{2}, -\frac{\sqrt{3}}{2}\right).$$

c) We have

$$\nabla^2 f(\bar{x}_1) = \begin{pmatrix} \sqrt{6} & 0 \\ 0 & -\sqrt{6} \end{pmatrix} \quad \text{and} \quad \nabla^2 f(\bar{x}_2) = \begin{pmatrix} -\sqrt{6} & 0 \\ 0 & \sqrt{6} \end{pmatrix}.$$

Both Hessians are diagonal matrices with eigenvalues $-\sqrt{6}$ and $\sqrt{6}$ and, hence $\nabla^2 f(\bar{x}_1)$ and $\nabla^2 f(\bar{x}_2)$ are indefinite and the stationary points $\bar{x}_1$ and $\bar{x}_2$ are saddle points. Furthermore, it holds that

$$\nabla^2 f(\bar{x}_3) = \begin{pmatrix} 3 & -\sqrt{3} \\ -\sqrt{3} & 6 \end{pmatrix} \quad \text{and} \quad \nabla^2 f(\bar{x}_2) = \begin{pmatrix} 3 & -\sqrt{3} \\ -\sqrt{3} & 6 \end{pmatrix}$$

and $\mathrm{tr}(\nabla^2 f(\bar{x}_3)) = \mathrm{tr}(\nabla^2 f(\bar{x}_4)) = 9 > 0$ and $\det(\nabla^2 f(\bar{x}_3)) = \det(\nabla^2 f(\bar{x}_4)) = 18 - 3 > 0$. This shows that $\nabla^2 f(\bar{x}_3)$ and $\nabla^2 f(\bar{x}_4)$ are positive definite. Thus, by the second order sufficient conditions, $\bar{x}_3$ and $\bar{x}_4$ are strict local minimizer.

d) Our results in a) and c) show that $\bar{x}_3$ and $\bar{x}_4$ are strict local solutions and that f does not possess a global minimum.

Assignment A2.2 (Another Optimization Problem): (approx. *15* points)

Consider the unconstrained optimization problem

$$\min_{x \in \mathbb{R}^3} \ f(x) = x_1^4 - 2x_1^2 + x_2^2 + 2x_2x_3 + 2x_3^2.$$

a) Is the mapping f coercive?

b) Verify whether the function f is convex on $\mathbb{R}^3$.

c) Find all stationary points of f and classify them according to whether they are saddle points, strict / non-strict, local / global, minimum / maximum points.

Solution :

a) Yes, the mapping f coercive. In particular, we have

$$\begin{aligned} f(x) &= x_1^2(x_1^2 - 2) + \frac{1}{3}x_2^2 + \frac{1}{2}x_3^2 + \left(\frac{2}{3}x_2^2 + 2x_2x_3 + \frac{3}{2}x_3^2\right) \\ &= x_1^2(x_1^2 - 2) + \frac{1}{3}x_2^2 + \frac{1}{2}x_3^2 + \left(\sqrt{\frac{2}{3}}x_2 + \sqrt{\frac{3}{2}}x_3\right)^2 \geq x_1^2(x_1^2 - 2) + \frac{1}{3}x_2^2 + \frac{1}{2}x_3^2. \end{aligned}$$

Since x_1^4 dominates x_1^2 (i.e., it grows faster than x_1^2 as $|x_1| \to \infty$, this establishes coercivity of f.

b) We calculate the gradient and Hessian of f. It holds that

$$\nabla f(x) = \begin{pmatrix} 4x_1^3 - 4x_1 \\ 2x_2 + 2x_3 \\ 2x_2 + 4x_3 \end{pmatrix}, \quad \nabla^2 f(x) = \begin{pmatrix} 12x_1^2 - 4 & 0 & 0 \\ 0 & 2 & 2 \\ 0 & 2 & 4 \end{pmatrix},$$

and

$$e_1^\top \nabla^2 f(x) e_1 = 12x_1^2 - 4.$$

For all $x_1 \in (-1/\sqrt{3}, 1/\sqrt{3})$, this term is negative which implies that f can not be convex on $\mathbb{R}^3$.

c) We have $4x_1(x_1^2 - 1) = 0$ and $x_2 = x_3 = 0$ which yields the stationary points

$$\bar{x}_1 = (0,0,0), \quad \bar{x}_2 = (1,0,0), \quad \bar{x}_3 = (-1,0,0).$$

Furthermore, it holds that $\det(\nabla^2 f(\bar{x}_1) - \lambda I) = (-4-\lambda)[(2-\lambda)(4-\lambda) - 4] = -(4+\lambda)(4 - 6\lambda + \lambda^2)$. Hence, the eigenvalues of $\nabla^2 f(\bar{x}_1)$ are given by -4, $3 \pm \sqrt{5}$. This shows that $\nabla^2 f(\bar{x}_1)$ is indefinite and thus, $\bar{x}_1$ is a saddle point. Similarly, we obtain $\det(\nabla^2 f(\bar{x}_{2/3}) - \lambda I) = (8-\lambda)(4 - 6\lambda + \lambda^2)$ and consequently, $\nabla^2 f(\bar{x}_{2/3})$ is positive definite and the stationary points $\bar{x}_2$ and $\bar{x}_3$ are strict local minimizer. Since the function f is coercive, it possesses at least one global minimizer that needs to be a stationary point. Due to $f(\bar{x}_2) = f(\bar{x}_3) = -1$, all local minimizer have the same function value and we can infer that $\bar{x}_2$ and $\bar{x}_3$ are global solutions of the problem $\min_x f(x)$.

Assignment A2.3 (Convex Sets): (approx. *20* points)

In this exercise, we study convexity of various sets.

a) Verify whether the following sets are convex or not and explain your answer!

$$X_1 = \{x \in \mathbb{R}^n : \alpha \le (a^\top x)^2 \le \beta\}, \quad \alpha, \beta \in \mathbb{R},\ \alpha \le \beta,\ a \in \mathbb{R}^n,$$
$$X_2 = \{x \in \mathbb{R}^n : \|x - a\|_2 \le \|x - b\|_2\}, \quad a, b \in \mathbb{R}^n,\ a \ne b.$$

b) Let the set of all positive semidefinite and symmetric $n \times n$ matrices be denoted by $\mathbb{S}^n_+$. Show that the set $X := \{A \in \mathbb{R}^{n\times n} : A \in \mathbb{S}^n_+ \text{ and } \mathrm{tr}(A) = 1\}$ is a convex subset of $\mathbb{R}^{n\times n}$.

c) Decide whether the following statements are true or false. Explain your answer and either present a proof / verification or a counter-example.

- The union of two convex sets $X_1, X_2 \subset \mathbb{R}^n$, $X_1 \ne X_2$ is never a convex set.
- Let $f : \mathbb{R}^n \to \mathbb{R}$ be a concave. Then, the set $X := \{x \in \mathbb{R}^n : f(x) \ge 0\}$ is convex.

Solution :

a) The set X_1 is not convex. We can choose $n = 1$ and $\alpha = \beta = a = 1$. Then, it follows $X_1 = \{x \in \mathbb{R} : x^2 = 1\} = \{\pm 1\}$. This set is obviously not convex.

The condition $\|x - a\| \le \|x - b\|$ in the definition of X_2 is equivalent to

$$\|x - a\|^2 \le \|x - b\|^2 \iff -2a^\top x + \|a\|^2 \le -2b^\top x + \|b\|^2 \iff 2(b - a)^\top x \le \|b\|^2 - \|a\|^2.$$

Hence, X_2 is a convex half space.

b) Let $A, B \in X$ be two symmetric, positive semidefinite matrices and let $\lambda \in [0, 1]$ be arbitrary. Then, the matrix $C = \lambda A + (1 - \lambda)B$ is symmetric and it follows

$$x^\top C x = \lambda[x^\top A x] + (1 - \lambda)[x^\top B x] \ge 0, \quad \forall\ x \in \mathbb{R}^n,$$

where we used the positive semidefiniteness of A and B. Hence, we have $C \in \mathbb{S}^n_+$ and by the linearity of the trace operator, it follows $\mathrm{tr}(C) = \lambda \mathrm{tr}(A) + (1 - \lambda)\mathrm{tr}(B) = \lambda + 1 - \lambda = 1$. This shows that X is a convex subset of $\mathbb{R}^{n\times n}$.

c) The first statement is wrong: consider $X_1 = [0, 1]$ and $X_2 = [-1, 0]$. Then $X_1 \ne X_2$ and $X_1 \cup X_2 = [-1, 1]$ is a convex set.

We verify the second statement briefly. Let $x, y \in X$ and $\lambda \in [0, 1]$ be arbitrary. To establish convexity of X, we need to show that $\lambda x + (1 - \lambda)y \in X$. Utilizing the concavity of f and $x, y \in X$, we have

$$g(\lambda x + (1 - \lambda)y) \ge \lambda g(x) + (1 - \lambda)g(y) \ge 0.$$

Hence, X is a convex set.

Assignment A2.4 (Convex Functions): (approx. *25* points)
In this exercise, convexity properties of different functions are investigated.

a) Verify that the following functions are convex over the specified domain:

- $f : \mathbb{R}_{++} \to \mathbb{R}$, $f(x) := \sqrt{1+x^{-2}}$, where $\mathbb{R}_{++} := \{x \in \mathbb{R} : x > 0\}$.
- $f : \mathbb{R}^n \to \mathbb{R}$, $f(x) := \frac{1}{2}\|Ax-b\|^2 + \mu\|Lx\|^2$, where $A \in \mathbb{R}^{m\times n}$, $L \in \mathbb{R}^{p\times n}$, $b \in \mathbb{R}^m$, and $\mu > 0$ are given.
- $f : \mathbb{R}^{n+1} \to \mathbb{R}$, $f(x,y) := \frac{\lambda}{2}\|x\|^2 + \sum_{i=1}^m \max\{0, 1 - b_i(a_i^\top x + y)\}$, where $a_i \in \mathbb{R}^n$ and $b_i \in \{-1,1\}$ are given data points for all $i = 1, ..., m$ and $\lambda > 0$ is a parameter.

b) Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex mapping and set $g(x) := (f(x))^2$. Is the function g convex? Explain your answer and either present a brief verification or a counter-example.

Is the mapping $x \mapsto \frac{1}{2}(\|x\|^2 - 1)^2$ convex?

Solution :

a) The mapping $f(x) = \sqrt{1+x^{-2}}$ is twice continuously differentiable on the convex set $\mathbb{R}_{++}$ and it holds that

$$f'(x) = \frac{1}{2\sqrt{1+x^{-2}}}\frac{-2}{x^3} = \frac{-1}{x^2\sqrt{1+x^2}}, \quad f''(x) = \frac{2x\sqrt{1+x^2}+\frac{x^3}{\sqrt{1+x^2}}}{x^4(1+x^2)} = \frac{3x^2+2}{x^3(1+x^2)\sqrt{1+x^2}}.$$

Notice that we have $f''(x) > 0$ for all $x > 0$ and thus, f is (strictly) convex on $\mathbb{R}_{++}$.

In the second example, we obtain

$$\nabla f(x) = A^\top(Ax-b) + \mu L^\top Lx \quad \text{and} \quad \nabla^2 f(x) = A^\top A + \mu L^\top L$$

Thus, it follows $h^\top \nabla^2 f(x) h = h^\top A^\top A h + \mu h^\top L^\top L h = \|Ah\|^2 + \mu\|Lh\|^2 \geq 0$ for all $h \in \mathbb{R}^n$. This establishes convexity of f on $\mathbb{R}^n$.

Finally, let us define $g(x,y) = \frac{\lambda}{2}\|x\|^2$ and $g_i(x,y) = \max\{0, 1 - b_i(a_i^\top x + y)\}$. Then, f can be interpreted as the sum of the functions g and g_i, $i = 1, ..., m$ and convexity follows if each of the functions g, g_i, $i = 1, ..., m$ is convex. The mapping g_i is the composition of the max-function $z \mapsto \max\{0, z\}$ and the affine-linear function $(x,y) \mapsto h_i(x,y) := 1 - b_i(a_i^\top x + y)$. Since h_i is convex (as a linear mapping), the max-function g_i also is convex! Finally, the Hessian of g is given by

$$\mathbb{R}^{(n+1)\times(n+1)} \ni \nabla^2 g(x,y) = \begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix} \succeq 0.$$

Combining these different results, we see that f is a convex function.

b) No, the mapping g generally does not need to be convex. We can consider the example $g(x) = \frac{1}{2}(x^2-1)^2 = \frac{1}{2}x^4 - x^2 + \frac{1}{2}$. The derivatives of g are given by $g'(x) = 2x^3 - 2x$ and $g''(x) = 6x^2 - 2$. Since g'' takes negative values around $x = 0$, the mapping g can not be convex. This also verifies that $\frac{1}{2}(\|x\|^2-1)^2$ is not a convex function.

Remark: Since the function $x \mapsto x^2$ is monotonically increasing on $[0, \infty)$, we can infer convexity in the case $f(x) \geq 0$ for all $x \in \mathbb{R}^n$.

Assignment A2.5 (A Penalty Problem): (approx. *20* points)

We consider the parametrized optimization problem

$$\min_x \ f_\beta(x) := \frac{1}{2}\|x-b\|^2 + \frac{\beta}{2}\left(\sum_{i=1}^n x_i\right)^2, \quad x \in \mathbb{R}^n, \tag{1}$$

where $b \in \mathbb{R}^n$ is given and $\beta \geq 0$ is a parameter.

a) Calculate the gradient and Hessian of f_β.

b) Show that the mapping f_β is strongly convex for all $\beta \geq 0$.

c) Show that f_β has a unique stationary point x^*_β and compute it explicitly. Determine whether x^*_β is a local minimizer, a local maximizer, or a saddle point of problem (1).

d) For $\beta \to \infty$, the solutions x^*_β converge to a point x^*. Calculate the limit $x^* = \lim_{\beta\to\infty} x^*_\beta$ explicitly and show that x^* satisfies the constraint $\mathbb{1}^\top x^* = \sum_{i=1}^n x^*_i = 0$.

Solution :

a) The objective function is given by $f_\beta(x) = \frac{1}{2}\sum_{i=1}^n (x_i - b_i)^2 + \frac{\beta}{2}\left(\sum_{i=1}^n x_i\right)^2$. Hence, we obtain

$$\frac{\partial f_\beta}{\partial x_i}(x) = (x_i - b_i) + \beta\left(\sum_{i=1}^n x_i\right)\cdot 1 \quad \text{and} \quad \frac{\partial^2 f_\beta}{\partial x_i \partial x_j}(x) = \begin{cases} 1+\beta & \text{if } i = j, \\ \beta & \text{if } i \neq j. \end{cases}$$

Due to $\mathbb{1}^\top x = \sum_{i=1}^n x_i$ and $(\mathbb{1}\mathbb{1}^\top)_{ij} = 1$ for all i, j, this shows that ∇f_β and $\nabla^2 f_\beta$ are given by:

$$\nabla f_\beta(x) = x - b + \beta \cdot (\mathbb{1}^\top x)\mathbb{1}, \quad \nabla^2 f_\beta(x) = I + \beta \cdot \mathbb{1}\mathbb{1}^\top.$$

b) We have

$$d^\top \nabla^2 f_\beta(x) d = d^\top d + \beta(\mathbb{1}^\top d)^2 \geq \|d\|^2 \quad \forall\ d \in \mathbb{R}^n$$

and hence, the Hessian $\nabla^2 f_\beta$ is uniformly positive definite for all x. This implies that f_β is 1-strongly convex for all $\beta \geq 0$.

c) Multiplying the condition $\nabla f_\beta(x) = 0$ with $\mathbb{1}^\top\cdot$, we obtain

$$\mathbb{1}^\top x - \mathbb{1}^\top b + \beta n \cdot (\mathbb{1}^\top x) = 0 \quad \Longrightarrow \quad \mathbb{1}^\top x = \frac{\mathbb{1}^\top b}{1+\beta n}.$$

Using this formula, it now follows $\nabla f_\beta(x) = 0$ if and only if $x = b - \frac{\beta}{1+\beta n}\cdot(\mathbb{1}^\top b)\mathbb{1}$. Since all stationary points are uniquely characterized in this way, this implies that $x^*_\beta = b - \frac{\beta}{1+\beta n}\cdot(\mathbb{1}^\top b)\mathbb{1}$ is the unique stationary point of problem (1). Moreover, due to part b), we know that x^* is a strict global minimizer of problem (1).

d) We have $x^*_\beta = b - \frac{1}{1/\beta + n}\cdot(\mathbb{1}^\top b)\mathbb{1} \to b - \frac{1}{n}\cdot(\mathbb{1}^\top b)\mathbb{1} =: x^*$ as $\beta \to \infty$. Moreover, we obtain

$$\mathbb{1}^\top x^* = \mathbb{1}^\top b - \frac{\mathbb{1}^\top\mathbb{1}}{n}\mathbb{1}^\top b = 0.$$