7 Exercises and Problems, and their Solutions

This chapter is dedicated to the concrete application of the methods and techniques from the previous chapters. All these applications are given as exercises, together with their solutions. At the same time, there are problems with complete solutions that highlight different theoretical aspects which were not included in the previous chapters. This means that in what follows, there appear several theoretical conclusions important on their own. The organization of the material in this chapter follows the main topics of the book.

7.1 Analysis of Real Functions of One Variable

Exercise 7.1. *Determine the extreme points of the functions below:*

(i) $f : \mathbb{R} \to \mathbb{R}$, $f(x) = (x+2)^2(x-1)^3$;

(ii) $f : \mathbb{R} \to \mathbb{R}$, $f(x) = \sin^3 x + \cos^3 x$;

(iii) $f : \mathbb{R} \to \mathbb{R}$, $f(x) = \sqrt[3]{x^2} - \sqrt[3]{x^2 - 1}$;

(iv) $f : \mathbb{R} \to \mathbb{R}$, $f(x) = \frac{x^2-5x+6}{x^2+1}$;

(v) $f : (0,\infty) \to \mathbb{R}$, $f(x) = \frac{|\ln x|}{\sqrt{x}}$;

(vi) $f : \mathbb{R} \setminus \{-\sqrt{2}, 0\} \to \mathbb{R}$, $f(x) = \frac{x^2 e^{\frac{1}{x}}}{x+\sqrt{2}}$;

(vii) $f : \mathbb{R} \to \mathbb{R}$, $f(x) = x \arcsin \frac{x-1}{\sqrt{2(x^2+1)}}$.

Solution (i) The derivative of the function is

$$f'(x) = (x+2)(x-1)^2(5x+4),$$

so the critical (stationary) points are $-2; 1; -\frac{4}{5}$. From the interval of monotonicity of the function (according to the sign of derivative) one immediately infers that -2 is a local maximum point, $-\frac{4}{5}$ is a local minimum point, while 1 is not an extremum point.

(ii) It is enough to study the function on the interval $[0, 2\pi)$ (taking into account its periodicity). The derivative is

$$f'(x) = 3\sin x \cos x(\sin x - \cos x),$$

with the roots

$$0, \frac{\pi}{4}, \frac{\pi}{2}, \pi, \frac{5\pi}{4}, \frac{3\pi}{2}.$$

From the analysis of the sign of the derivative on the corresponding intervals, and extending the argument to the entire real line, we obtain that $0+2k\pi$, $\frac{\pi}{2}+2k\pi$, $\frac{5\pi}{4}+2k\pi$ ($k \in \mathbb{Z}$) are local maxima, and $\frac{\pi}{4}+2k\pi$, $\pi+2k\pi$, $\frac{3\pi}{2}+2k\pi$ are local minima.

Another possible solution is to compute the second order derivative of f and to use the following result: for a critical point $\overline{x}$, if $f''(\overline{x}) > 0$, then $\overline{x}$ is a local minimum, and if $f''(\overline{x}) < 0$, then $\overline{x}$ is a local maximum.

(iii) The function is differentiable on $\mathbb{R} \setminus \{-1, 0, 1\}$, and the derivative is

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

The critical points are $-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}$, so the candidates for extrema are

$$-1, 0, 1, -\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}.$$

From the analysis of the derivative sign around those points, we infer that $-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}$ are local maxima and 0 is a local minimum.

(iv) The discussion is similar to that from item (ii). One gets that $1 - \sqrt{2}$ is a local maximum and $1 + \sqrt{2}$ is a local minimum.

(v) The function is differentiable on $(0, \infty) \setminus \{1\}$, and

$$f'(x) = \begin{cases} -\frac{2-\ln x}{2x\sqrt{x}}, & x \in (0,1) \\ \frac{2-\ln x}{2x\sqrt{x}}, & x \in (1, \infty). \end{cases}$$

The candidates for extrema are 1 and e^2. From the variation of f, we decide that 1 is a local minimum and e^2 is a local maximum.

(vi) Similar arguments show that $-1 - \sqrt{2}$ is a local maximum and $-\sqrt{2} + 2$ is a local minimum. We remark as well that for $|f|$ both points are local minima, but no one is global minimum, since $\inf_{x\in\mathbb{R}} \left|f(x)\right| = 0$ is obtained for $x \to 0+$.

(vii) In order to decide the behaviour of the first derivative, we should compute the second one as well. Then

$$f'(x) = \begin{cases} \arcsin \frac{x-1}{\sqrt{2(x^2+1)}} + \frac{x}{x^2+1}, & x > -1 \\ \arcsin \frac{x-1}{\sqrt{2(x^2+1)}} - \frac{x}{x^2+1}, & x < -1, \end{cases}$$

and

$$f''(x) = \begin{cases} \frac{2}{(x^2+1)^2}, & x > -1 \\ -\frac{2}{(x^2+1)^2}, & x < -1. \end{cases}$$

Finally, the function has neither critical points nor extrema (the only possible candidate is the point where f is not differentiable, i.e., -1). □

Exercise 7.2. *Decide if $x = 0$ is an extreme point for $f : \mathbb{R} \to \mathbb{R}, f(x) = e^x + e^{-x} + 2\cos x$.*

Solution We successively compute the derivative at this point, until we find a nonvanishing one. We obtain

$$f'(0) = 0;\ f''(0) = 0;\ f'''(0) = 0;\ f^{iv}(0) = 4.$$

So, $x = 0$ is a local minimum point, according to Theorem 1.3.14. □

Problem 7.3. *Let $a, b \in \mathbb{R}$, $a < b$, and $f, g : [a, b] \to \mathbb{R}$ be a continuous functions. One defines $h : \mathbb{R} \to \mathbb{R}$,*

$$h(t) = \sup\{f(x) + tg(x) \mid x \in [a, b]\}.$$

Show that h is a Lipschitz function.

Solution Let $s, t \in \mathbb{R}$. Since the applications $x \mapsto f(x) + tg(x)$ and $x \mapsto f(x) + sg(x)$ are continuous, by virtue of Weierstrass Theorem, on the compact interval $[a, b]$, there exist $x_t, x_s \in [a, b]$ such that

$$\begin{aligned} h(t) &= f(x_t) + tg(x_t) \\ h(s) &= f(x_s) + sg(x_s). \end{aligned}$$

Then

$$\begin{aligned} h(t) - h(s) &= f(x_t) + tg(x_t) - (f(x_s) + sg(x_s)) \\ &\leq f(x_t) + tg(x_t) - f(x_t) - sg(x_t) \\ &= g(x_t)(t - s). \end{aligned}$$

Similarly,

$$\begin{aligned} h(t) - h(s) &= f(x_t) + tg(x_t) - (f(x_s) + sg(x_s)) \\ &\geq f(x_s) + tg(x_s) - f(x_s) - sg(x_s) \\ &= g(x_s)(t - s). \end{aligned}$$

We denote $M := \max_{x\in[a,b]} \left|g(x)\right|$ and observe that $M \in \mathbb{R}$ (using again Weierstrass Theorem). Then we get

$$\left|h(t) - h(s)\right| \leq M \left|t - s\right|,$$

that is the conclusion. □

Exercise 7.4. *Let $f : \mathbb{R} \to \mathbb{R}$, $f(x) = \min\{x + 1, 0, 1 - x\}$. Show that $\overline{x} = 0$ is a local minimum point of f, but it is not global maximum for f.*

Solution The easy-to-draw graph of f proves the assertions. Clearly, a more rigorous proof could be given as well, also starting from the investigation of the graph. □

Exercise 7.5. *Let $x_1 < x_3 < x_2$ be real numbers, and $f : \mathbb{R} \to \mathbb{R}$ be a C^1 function on $[x_1, x_2]$. Suppose that*

$$\max\{f(x_1), f(x_2)\} < f(x_3).$$

Then there exists $\overline{x} \in (x_1, x_2)$ a critical point of f with

$$f(\overline{x}) = \max_{x\in[x_1,x_2]} f(x).$$

Solution On $[x_1, x_2]$, the function f admits a maximum point (Weierstrass' Theorem). From the assumption, this point cannot be x_1 or x_2. So, this maximum point (denoted by $\overline{x}$) lies in the interior of the interval, so it is a local maximum of f. Fermat Theorem gives the conclusion. □

Exercise 7.6. *Let $a_1, \dots, a_n$ be strictly positive real numbers such that*

$$a_1^x + a_2^x + \dots + a_n^x \geq n, \ \forall x \in \mathbb{R}.$$

Show that $a_1 a_2 \dots a_n = 1$.

Solution Let $f : \mathbb{R} \to \mathbb{R}$, $f(x) = a_1^x + a_2^x + \dots + a_n^x$. Since $f(0) = n$, from the hypothesis we infer that $x = 0$ is a global minimum point for f so, from Fermat Theorem, $f'(0) = 0$, which leads to the conclusion.

Another solution could be given using the well-known (fundamental) limit: $\lim_{x\to 0} \frac{a^x - 1}{x} = \ln a$ for $a > 0$. □

Problem 7.7. *Let $a, b \in \mathbb{R}$, $a < b$, and $f : [a, b] \to \mathbb{R}$ be a continuous function, derivable at a and b, with $f'(a)f'(b) < 0$. Show that f admits a local extremum in (a, b).*

Solution From Weierstrass Theorem, f admits a minimum and a maximum on $[a, b]$. If the conclusion would not hold, then these points should be a and b. Without loss of generality, suppose that $f'(a) < 0$ (whence, $f'(b) > 0$). Since

$$f'(a) = \lim_{x\to a+} \frac{f(x) - f(a)}{x - a} < 0,$$

for x sufficiently close to a, $f(x) < f(a)$, so a should be the maximum point. Whence b is the minimum point, so $f(x) \geq f(b)$ for every $x \in [a, b]$. Then

$$\frac{f(x) - f(b)}{x - b} \leq 0, \ \forall x \in [a, b].$$

We infer that $f'(b) \leq 0$, which is absurd. This contradiction can be resolved only if the conclusion holds. □

Problem 7.8. *Let $f : \mathbb{R} \to \mathbb{R}$ be a derivable function such that* $\lim_{x\to\infty} \frac{f(x)}{x} = \infty$ *and* $\lim_{x\to-\infty} \frac{f(x)}{x} = -\infty$. *Show that* $\operatorname{Im} f' = \mathbb{R}$ *(i.e., f' is surjective).*

Solution Let $r \in \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$, $g(x) := f(x) - rx$. It is clear that $\lim_{|x|\to\infty} g(x) = \infty$ and, therefore, g attains its (global) minimum at a point $\overline{x} \in \mathbb{R}$. Then, from Fermat Theorem, $0 = g'(\overline{x}) = f'(\overline{x}) - r$. □

Exercise 7.9. *Show that $f : (1, \infty) \to \mathbb{R}$, $f(x) = -\ln(\ln x)$ is convex. Deduce that for every $a, b > 1$, one has*

$$\sqrt{\ln a \ln b} \leq \ln\left(\frac{a+b}{2}\right).$$

Solution The second-order derivative on the definition interval is

$$f''(x) = \frac{1}{x^2 \ln x} + \frac{1}{x^2 \ln^2 x}.$$

Since $x > 1$, this function has only positive values, so f is convex.

Now, using this property, one gets

$$-\ln\left(\ln\frac{a+b}{2}\right) \le -\frac{1}{2}\left(\ln(\ln a) + \ln(\ln b)\right) = -\ln(\sqrt{\ln a \ln b}),$$

and the desired inequality follows. □

Problem 7.10. *Let $g : \mathbb{R} \to \mathbb{R}$ be a continuous function. Show that g is convex if and only if for every integrable (in the Riemann sense) function $f : [0, 1] \to \mathbb{R}$, the following inequality holds*

$$g\left(\int_0^1 f(u)du\right) \le \int_0^1 g(f(u))du.$$

Solution If g is convex then for any partition of $[0, 1]$ and for every system of intermediate points (with the notation from Definition 1.4.1) it holds that

$$g\left(\sum_{i=1}^{n} f(\xi_i)(x_i - x_{i-1})\right) \le \sum_{i=1}^{n} g(f(\xi_i))(x_i - x_{i-1}).$$

Passing to the limit for the norm of the partition going to 0, we get the inequality from the conclusion.

For the converse implication, we fix $x, y \in \mathbb{R}$, $\alpha \in (0, 1)$, and we consider $f : [0, 1] \to \mathbb{R}$,

$$f(u) = \begin{cases} x, & \text{if } u \in [0, \alpha] \\ y, & \text{if } u \in (\alpha, 1]. \end{cases}$$

Then f is Riemann integrable on $[0, 1]$ (see Theorems 1.4.9, 1.4.10). Applying the assumption of this stage, we find $g(\alpha x + (1-\alpha)y) \le \alpha g(x) + (1-\alpha)g(y)$, whence the convexity of g. □

Problem 7.11. *(i) Let $f : [1, \infty) \to \mathbb{R}$ given by $f(x) = x\sqrt[x]{x}$. Show that f is increasing and concave.*

(ii) Show that the sequence $a_n := (n+1)\sqrt[n+1]{n+1} - n\sqrt[n]{n}$, $n \in \mathbb{N}\setminus\{0, 1\}$ is monotone and bounded. Find $\lim a_n$.

Solution (i) The function f could be written as

$$f(x) = e^{\left(1+\frac{1}{x}\right)\ln x},$$

and

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

Obviously, f' has positive values, so f is increasing. In order to establish the sign of f'' we compare, by use of some auxiliary functions, the expressions $|1 - \ln x|$ and $\sqrt{x}$ considering, separately, the cases $x \le e$ and $x \ge e$. In both cases, we deduce that $f''(x) \le 0$, whence f is concave.

(ii) We can write, for every $n \in \mathbb{N}\setminus\{0,1\}$, $a_n = f(n+1) - f(n)$, and by the concavity of f we get

$$f(n+1) = f\left(\frac{n+2}{2} + \frac{n}{2}\right) \ge \frac{f(n+2)}{2} + \frac{f(n)}{2},$$

which leads to $a_n \ge a_{n+1}$. Therefore, (a_n) is a decreasing sequence. The monotonicity of f proves that the terms of (a_n) are positive, whence (a_n) is convergent. In order to find its limit, we can apply the Stolz-Cesàro Criterion (Proposition 1.1.24) to the sequence

$$b_n = \sqrt[n]{n} = \frac{n\sqrt[n]{n}}{n}, \; n \in \mathbb{N} \setminus \{0,1\}.$$

We have that

$$\lim b_n = 1 = \lim \frac{(n+1)\sqrt[n+1]{n+1} - n\sqrt[n]{n}}{n+1-n} = \lim a_n. \qquad \square$$

Problem 7.12. *Let $a, b \in \mathbb{R}$, $a < b$, and $f : [a,b] \to \mathbb{R}$ be a C^2 function with $f(a) = f(b) = 0$. Let $M := \sup_{x\in[a,b]} \left|f''(x)\right|$ and*

$$g(x) := f(x) - M\frac{(x-a)(b-x)}{2}; \; h(x) := -f(x) - M\frac{(x-a)(b-x)}{2}.$$

Show that g, h are convex and deduce the inequality

$$\left|f(x)\right| \le M\frac{(x-a)(b-x)}{2}, \; \forall x \in [a,b].$$

Solution We have

$$g''(x) = f''(x) + M \ge 0, \; \forall x \in [a,b]$$
$$h''(x) = -f''(x) + M \ge 0, \; \forall x \in [a,b],$$

which shows that both functions are convex.

The convexity of g and the fact that $g(a) = g(b) = 0$ lead to the conclusion that $g(x) \le 0$ for every $x \in [a,b]$. Hence,

$$f(x) \le M\frac{(x-a)(b-x)}{2}.$$

Analogously, arguing for h,

$$-f(x) \le M\frac{(x-a)(b-x)}{2}.$$

These two relations lead to the conclusion. $\square$

Problem 7.13. *Let $f : \mathbb{R} \to \mathbb{R}$ be a convex function.*

(i) Let $a, b \in \mathbb{R}$, $a < b$. Study the position of the graph of f with respect to the line joining the points $(a, f(a))$ and $(b, f(b))$.

(ii) Deduce that if f is bounded, then it is constant.

Solution (i) It is clear (from the geometrical interpretation of convexity) that for $x \in [a, b]$, the graph of f is under the line joining $(a, f(a))$ and $(b, f(b))$. Notice that this line has the equation

$$y = \frac{f(b) - f(a)}{b - a}(x - a) + f(a).$$

For $x > b$, from convexity and from $a < b < x$ we deduce

$$\frac{f(b) - f(a)}{b - a} \le \frac{f(x) - f(a)}{x - a},$$

that is

$$f(x) \ge \frac{f(b) - f(a)}{b - a}(x - a) + f(a),$$

whence the graph of f is above the mentioned line. The same conclusion holds analogously for $x < a$.

(ii) Suppose that f would not be constant. Then it would exist $a, b \in \mathbb{R}$ with $a < b$ and $f(a) \neq f(b)$. We can consider, without loss of generality, that $f(b) > f(a)$. Since for $x > b$,

$$f(x) \ge \frac{f(b) - f(a)}{b - a}(x - a) + f(a),$$

we get that $\lim_{x\to\infty} f(x) = +\infty$, that is f is not bounded. This is a contradiction, so f is constant. □

Problem 7.14. *Let $a, b \in \mathbb{R}, a < b$, and $f : (a, b) \to \mathbb{R}$ be convex. Show that f is bounded from below. It is true that f is bounded?*

Solution Let $x_0 < x_1 < x_2$ be three points in (a, b). For $x < x_1$, we have

$$\frac{f(x_1) - f(x)}{x_1 - x} \le \frac{f(x_2) - f(x_1)}{x_2 - x_1},$$

which gives

$$(x_1 - x)\frac{f(x_2) - f(x_1)}{x_2 - x_1} + f(x_1) \le f(x),$$

so f is bounded from below on $(a, x_1]$. Similarly, for $x > x_1$,

$$\frac{f(x) - f(x_1)}{x - x_1} \ge \frac{f(x_1) - f(x_0)}{x_1 - x_0}$$

that is

$$(x - x_1)\frac{f(x_1) - f(x_0)}{x_1 - x_0} + f(x_0) \le f(x).$$

We obtain that f is bounded from below on $[x_1, b)$, so it is also bounded from below on (a, b).

Generally, the upper boundedness is not ensured by the convexity. As an example, consider the function $f : (-\frac{\pi}{2}, \frac{\pi}{2}) \to \mathbb{R}$, $f(x) = |\operatorname{tg} x|$. □

Exercise 7.15. *Let $f : \mathbb{R} \to \mathbb{R}$ be a convex, increasing function. Show that f is constant or* $\lim_{x\to\infty} f(x) = \infty$.

Solution If f is not constant, then there exist $a, b \in \mathbb{R}$, $a < b$ and $f(a) < f(b)$. Let $y = mx + n$ be the line joining $(a, f(a))$ and $(b, f(b))$. Clearly, $m > 0$. For $x > b$, we have, as seen before, $f(x) \geq mx + n$, whence $\lim_{x\to\infty} f(x) = \infty$. □

Problem 7.16. *Let $f : \mathbb{R} \to \mathbb{R}$ be a convex function.*

(i) Show that if $\lim_{x\to\infty} f(x) = 0$, *then $f(x) \geq 0$, for every $x \in \mathbb{R}$.*

(ii) Show that if f admits asymptote to $+\infty$, then its graph is above the asymptote.

Solution (i) Suppose, by way of contradiction, that exists $x_0 \in \mathbb{R}$ with $f(x_0) < 0$. By hypothesis, there exists $x_1 > x_0$ with $f(x_1) > f(x_0)$. For $x > x_1$, by means of convexity

$$\frac{f(x_1) - f(x_0)}{x_1 - x_0} \leq \frac{f(x) - f(x_1)}{x - x_1},$$

whence

$$f(x_1) + (x - x_1)\frac{f(x_1) - f(x_0)}{x_1 - x_0} \leq f(x).$$

We arrive at the contradiction $\lim_{x\to\infty} f(x) = \infty$.

(ii) Let $y = ax + b$ be the equation of the asymptote. The function $g : \mathbb{R} \to \mathbb{R}$, $g(x) = f(x) - ax - b$ is convex (as a sum between a convex function and an affine one) and, moreover, $\lim_{x\to\infty} g(x) = 0$. The application of the above step gives us the conclusion. □

Problem 7.17. *(i) Let $a, b \in \mathbb{R}$, $a < b$, $M > 0$, and $(f_n) : [a, b] \to \mathbb{R}$ be a sequence of M–Lipschitz functions on $[a, b]$. Show that if (f_n) is pointwise convergent on $[a, b]$, then (f_n) is uniformly convergent on $[a, b]$.*

(ii) Let $a, b \in \mathbb{R}$, $a < b$, and $(f_n) : (a, b) \to \mathbb{R}$ be a sequence of convex functions which is pointwise convergent on (a, b). Show that (f_n) is uniformly convergent on every closed subinterval of (a, b).

Solution (i) It is clear that the pointwise limit of (f_n) is itself an M–Lipschitz function, which we denote by f. Take $\varepsilon > 0$. We fix as well a partition of the form

$$a = \alpha_0 < \alpha_1 < \ldots < \alpha_p = b$$

of the interval (a, b), with the norm smaller than $\frac{\varepsilon}{M}$. The pointwise convergence gives a natural number n_0 sufficiently large such that for every $i \in \overline{0, p}$,

$$\left|f_n(\alpha_i) - f(\alpha_i)\right| < \varepsilon.$$

Let $x \in [a, b]$. There exists $i \in \overline{0, p-1}$ with $x \in [\alpha_i, \alpha_{i+1}]$. We have

$$\begin{aligned}\left|f_n(x) - f(x)\right| &\le \left|f_n(x) - f_n(\alpha_i)\right| + \left|f_n(\alpha_i) - f(\alpha_i)\right| + \left|f(\alpha_i) - f(x)\right| \\ &< M\,|x - \alpha_i| + \varepsilon + M\,|x - \alpha_i| \le 3\varepsilon.\end{aligned}$$

So (f_n) is uniformly convergent towards f.

(ii) We reduce this situation to the preceding one. Let $[\alpha, \beta] \subset (a, b)$ and $\alpha' \in (a, \alpha)$, $\beta' \in (\beta, b)$. Since f_n is convex, for every $x, y \in [\alpha, \beta]$ with $x \neq y$,

$$\frac{f_n(\alpha) - f_n(\alpha')}{\alpha - \alpha'} \le \frac{f_n(x) - f_n(y)}{x - y} \le \frac{f_n(\beta) - f_n(\beta')}{\beta - \beta'}.$$

The outer members of this inequality are bounded (by the pointwise convergence), so there exists $M > 0$ with

$$\left|\frac{f_n(x) - f_n(y)}{x - y}\right| \le M, \ \forall x, y \in [\alpha, \beta], \ x \neq y, \ \forall n \in \mathbb{N}.$$

Therefore we can apply (i) to the functions (f_n) on $[\alpha, \beta]$. □

Problem 7.18. *(i) Let $f, g : \mathbb{R} \to \mathbb{R}$. Show that if f is convex and increasing and g is convex, then $f \circ g$ is convex.*

(ii) Let $f : \mathbb{R} \to (0, \infty)$. Show that $\ln f$ *is convex if and only if for every $\alpha > 0$, the function f^α is convex.*

Solution (i) Let $x, y \in \mathbb{R}$ and $\lambda \in (0, 1)$. Then using the properties of f and g, we have

$$\begin{aligned}(f \circ g)(\lambda x + (1 - \lambda)y) &\le f(\lambda g(x) + (1 - \lambda)g(y)) \\ &\le \lambda f(g(x)) + (1 - \lambda)f(g(y)).\end{aligned}$$

(ii) Suppose first that $\ln f$ is convex. For every $\alpha > 0$,

$$f^\alpha = e^{\alpha \ln f}.$$

Since the mapping $x \mapsto e^{\alpha x}$ is convex and increasing, the preceding item applies.

Conversely, suppose that for every $\alpha > 0$, the function f^α is convex. Hence, for every $x, y \in \mathbb{R}$ and $\lambda \in [0, 1]$, $u(\alpha) \le v(\alpha)$ where

$$\begin{aligned}u(\alpha) &= e^{\alpha \ln f(\lambda x + (1-\lambda)y)} \\ v(\alpha) &= \lambda e^{\alpha \ln f(x)} + (1 - \lambda)e^{\alpha \ln f(y)}.\end{aligned}$$

But $u(0) = v(0)$. From $u(\alpha) \le v(\alpha)$ and $u(0) = v(0)$ we obtain $u'(0) \le v'(0)$, which leads to

$$\ln f(\lambda x + (1 - \lambda)y) \le \lambda \ln f(x) + (1 - \lambda) \ln f(y),$$

that is the desired relation. □

Definition 7.1.1. *Let $I \subset \mathbb{R}$ be an interval, and $f : I \to \mathbb{R}$ be a function. We say that f admits a support functional at $x \in I$ if there exists an affine function of the form $s(u) = f(x) + m(u - x)$ (where $u \in \mathbb{R}$) such that $s(u) \le f(u)$ for every $u \in I$.*

Problem 7.19. *Let $a, b \in \mathbb{R}$, $a < b$. Show that $f : (a, b) \to \mathbb{R}$ is convex if and only if it admits a support functional at every $x \in (a, b)$.*

Solution Suppose first that f is convex. We know already that f has lateral derivatives at any point of the interval. Let $\overline{x} \in (a, b)$ be fixed and $m \in [f'_-(\overline{x}), f'_+(\overline{x})]$. Then, for every $x \in (a, b)$, $x > \overline{x}$ we have

$$\frac{f(x) - f(\overline{x})}{x - \overline{x}} \ge m,$$

and for every $x \in (a, b)$, $x < \overline{x}$ the inverse inequality holds. In both cases,

$$m(x - \overline{x}) \le f(x) - f(\overline{x}),$$

hence f admits a support functional at $\overline{x}$.

Conversely, suppose that f admits a support functional at every $x \in (a, b)$. Let $x, y \in (a, b)$ and $\lambda \in (0, 1)$. Then $\overline{x} = \lambda x + (1 - \lambda)y \in (a, b)$, so there exists $m \in \mathbb{R}$ such that

$$s(u) := f(\overline{x}) + m(u - \overline{x}) \le f(u), \ \forall u \in (a, b).$$

Consequently,

$$f(\overline{x}) = s(\overline{x}) = \lambda s(x) + (1 - \lambda)s(y) \le \lambda f(x) + (1 - \lambda)f(y).$$

Therefore, f is convex. □

From the above problem and its solution one infers the next result.

Theorem 7.1.2. *Let $a, b \in \mathbb{R}$, $a < b$. A function $f : (a, b) \to \mathbb{R}$ is convex if and only if for every $c \in (a, b)$ there exists $y \in \mathbb{R}$ such that*

$$f(x) \ge f(c) + y(x - c), \ \forall x \in (a, b).$$

Moreover, y can be arbitrarily chosen in the interval $[f'_-(c), f'_+(c)] = \partial f(c)$.

Problem 7.20. *Let $a, b \in \mathbb{R}$, $a < b$ and $f : [a, b] \to \mathbb{R}$ be a convex, continuous function. Show that for every $c \in (a, b)$, there exists $y \in \mathbb{R}$ such that*

$$f(c) + y\frac{a + b - 2c}{2} \le \frac{1}{b - a}\int_a^b f(x)dx.$$

Show that the equality holds if and only if f is affine, that is has the form

$$f(x) = f(c) + y(x - c), \ \forall x \in (a, b).$$

Infer that the first inequality in the Hermite-Hadamard Inequality holds as equality if and only if f is affine.

Solution Using the preceding theorem and applying the Riemann integral, we get the desired inequality. For the equality, if there exists $x \in (a, b)$ with

$$f(c) + y(x - c) < f(x),$$

then from the continuity of the involved functions we arrive at a contradiction. So, in order to have equality one must have an affine function.

The obtained inequality is a refinement of the first inequality in the Hermite-Hadamard Inequality: it is enough to take $c = \frac{a+b}{2}$. □

From the above facts and the already known theory, we get that, under the continuity assumption, in both parts of the Hermite-Hadamard Inequality, equality holds if and only if f is affine.

Exercise 7.21. *Show that the following functions are convex and write Hermite-Hadamard Inequality in every case on mentioned intervals:*
(i) $f : [0, \infty) \to \mathbb{R}$, $f(x) = (x + 1)^{-1}$ *on* $[0, x]$ *and* $[n - 1, n]$;
(ii) $f : \mathbb{R} \to \mathbb{R}$, $f(x) = e^x$ *on* $[a, b]$;
(iii) $f : [0, \pi] \to \mathbb{R}$, $f(x) = -\sin x$ *on* $[a, b]$.

Solution In all three cases, the convexity is easy to verify using the second-derivative criterion. No one of these function is affine, so both inequalities in Hermite-Hadamard Inequality are strict.

For the first function, Hermite-Hadamard Inequality on an interval of the form $[0, x]$, $x \in \mathbb{R}$ (for instance) leads to

$$x - \frac{x^2}{x+2} < \ln(x + 1) < x - \frac{x^2}{2(x+1)},$$

and on an interval of the form $[n - 1, n]$, $n \in \mathbb{N}^*$ to

$$\frac{2}{2n+1} < \ln(n + 1) - \ln n < \frac{1}{2}\left(\frac{1}{n} + \frac{1}{n+1}\right).$$

For the second function we get

$$e^{\frac{a+b}{2}} < \frac{e^b - e^a}{b - a} < \frac{e^a + e^b}{2}, \ \forall a, b \in \mathbb{R},\ a \neq b.$$

In particular,

$$\sqrt{xy} < \frac{x - y}{\ln x - \ln y} < \frac{x + y}{2}, \ \forall x, y \in (0, \infty),\ x \neq y.$$

For the third function we find

$$\frac{\sin a + \sin b}{2} < \frac{\cos a - \cos b}{b - a} < \sin\left(\frac{a+b}{2}\right), \ \forall a, b \in \mathbb{R},\ a \neq b.$$

From here one can deduce the well-known inequalities

$$\sin x < x < \operatorname{tg} x, \ \forall x \in \left[0, \frac{\pi}{2}\right]. \qquad \square$$

Problem 7.22. *Let $a, b \in \mathbb{R}$, $a < b$, and $f : [a, b] \to \mathbb{R}$ be a convex, continuous function. Show that for every $c \in [0, \frac{b-a}{4}]$ one has*

$$\frac{1}{2}\left(f\left(\frac{a+b}{2} - c\right) + f\left(\frac{a+b}{2} + c\right)\right) \le \frac{1}{b-a}\int_a^b f(x)dx.$$

Solution We apply the inequality from Problem 7.20 for f on $\left[a, \frac{a+b}{2}\right]$ and $\left[\frac{a+b}{2}, b\right]$, and take into account that one can choose y like in Theorem 7.1.2. Therefore, for $c \in \left[0, \frac{b-a}{4}\right]$ we find

$$\frac{2}{b-a}\int_a^{\frac{a+b}{2}} f(x)dx \ge f\left(\frac{b+a}{2} - c\right) + f'_-\left(\frac{b+a}{2} - c\right)\frac{\frac{a+b}{2} + a - 2\left(\frac{a+b}{2} - c\right)}{2}$$

and

$$\frac{2}{b-a}\int_{\frac{a+b}{2}}^{b} f(x)dx \ge f\left(\frac{b+a}{2} + c\right) + f'_-\left(\frac{b+a}{2} + c\right)\frac{\frac{a+b}{2} + b - 2\left(\frac{a+b}{2} + c\right)}{2}.$$

After computation and considering the fact that

$$f'_-\left(\frac{b+a}{2} + c\right) \ge f'_-\left(\frac{b+a}{2} - c\right)$$

we get the conclusion. $\square$

Problem 7.23. *Let $a, b \in \mathbb{R}$, $a < b$ and $f : [a, b] \to \mathbb{R}$ be a C^2 function. Take $m, M \in \mathbb{R}$ such that $m \le f''(x) \le M$ for every $x \in [a, b]$ (note that such constants do exist according to Weierstrass' Theorem). Show that*

$$m\frac{(b-a)^2}{24} \le \frac{1}{b-a}\int_a^b f(x)dx - f\left(\frac{a+b}{2}\right) \le M\frac{(b-a)^2}{24},$$

and

$$m\frac{(b-a)^2}{12} \le \frac{f(a)+f(b)}{2} - \frac{1}{b-a}\int_a^b f(x)dx \le M\frac{(b-a)^2}{12}.$$

Solution It is easy to observe that the mappings

$$x \mapsto f(x) - \frac{mx^2}{2}$$

$$x \mapsto \frac{Mx^2}{2} - f(x)$$

are convex and continuous. Both inequalities follow from the application of Hermite-Hadamard Inequality to these two functions. □

Problem 7.24. *Let $a, b \in \mathbb{R}$, $a < b$, and $f : [a, b] \to \mathbb{R}$ be a L-Lipschitz ($L > 0$) function. Show that*

$$\left| f(x) - \frac{1}{b-a} \int_a^b f(t)dt \right| \leq \left[\frac{1}{4} + \left(\frac{x - \frac{a+b}{2}}{b-a} \right)^2 \right] L(b-a).$$

Solution We can write successively,

$$\begin{aligned}
\left| f(x) - \frac{1}{b-a} \int_a^b f(t)dt \right| &= \left| \frac{1}{b-a} \int_a^b (f(x) - f(t))dt \right| \\
&\leq \frac{L}{b-a} \int_a^b |x-t| \, dt \\
&= \left[\frac{1}{4} + \left(\frac{x - \frac{a+b}{2}}{b-a} \right)^2 \right] L(b-a),
\end{aligned}$$

and the inequality is proved. □

7.2 Nonlinear Analysis

Problem 7.25. *Show that a nonempty set $D \subset \mathbb{R}^p$ is convex if and only if it contains all the convex combinations of its elements.*

Solution Clearly, if D contains all the convex combinations of its elements it contains as well all the convex combinations with two of its elements and this is the definition of convexity.

Conversely, if D is convex then, by definition, contains all the convex combinations of any two of its elements. Consider three elements $x_1, x_2, x_3 \in D$, $\alpha_1, \alpha_2, \alpha_3 \in (0, 1)$ with $\alpha_1 + \alpha_2 + \alpha_3 = 1$. Then

$$\alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 = (\alpha_1 + \alpha_2) \left(\frac{\alpha_1}{\alpha_1 + \alpha_2} x_1 + \frac{\alpha_2}{\alpha_1 + \alpha_2} x_2 \right) + \alpha_3 x_3.$$

Now $z := \frac{\alpha_1}{\alpha_1+\alpha_2} x_1 + \frac{\alpha_2}{\alpha_1+\alpha_2} x_2$ is a convex combination of two elements of D, so $z \in D$. Furthermore, $(\alpha_1 + \alpha_2)z + \alpha_3 x_3$ is again a convex combination of two elements of D, so, finally, $\alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 \in D$. Of course, the complete proof needs a mathematical induction argument which uses the idea from above: suppose that for a $n \in \mathbb{N}\setminus\{1,2\}$, D contains the convex combinations of n of its elements. Take $x_1, x_2, \ldots, x_n, x_{n+1} \in D$, $\alpha_1, \alpha_2, \ldots, \alpha_n, \alpha_{n+1} \in (0,1)$ with $\alpha_1 + \ldots + \alpha_{n+1} = 1$. Then

$$\sum_{i=1}^{n+1} \alpha_i x_i = \sum_{i=1}^{n} \alpha_i x_i + \alpha_{n+1}x_{n+1} = \sum_{i=1}^{n} \alpha_i \sum_{i=1}^{n} \frac{\alpha_i}{\sum_{i=1}^{n}\alpha_i} x_i + \alpha_{n+1}x_{n+1}.$$

Now $z := \sum_{i=1}^{n} \frac{\alpha_i}{\sum_{i=1}^{n}\alpha_i} x_i$ is a convex combinations of n elements of D, whence, from the assumption, $z \in D$. Next, $\sum_{i=1}^{n} \alpha_i z + \alpha_{n+1}x_{n+1}$ is a convex combination of two elements of D, so $\sum_{i=1}^{n+1} \alpha_i x_i \in D$ and the proof is complete. □

Problem 7.26. *Let $D \subset \mathbb{R}^p$ be a nonempty set. Show that* conv D *is the smallest convex set (with respect to the set inclusion order relation) which contains D.*

Solution The inclusion $D \subset \operatorname{conv} D$ is obvious (take the elements with $n = 1$ in the definition of $\operatorname{conv} D$). On the other hand, if C is a convex set which includes D, from the above problem, is contains, in particular, all the convex combinations of the elements of D, that is it contains $\operatorname{conv} D$. It remains to show that $\operatorname{conv} D$ is a convex set. Take $x, y \in \operatorname{conv} D$ and $\mu_1, \mu_2 \in (0,1)$ with $\mu_1+\mu_2 = 1$. Since $x \in \operatorname{conv} D$, there exist $n \in \mathbb{N}^\star$, $(\alpha_i)_{i\in\overline{1,n}} \subset (0,\infty)$, $\sum_{i=1}^{n} \alpha_i = 1$, $(x_i)_{i\in\overline{1,n}} \subset D$ with $x = \sum_{i=1}^{n} \alpha_i x_i$. Similarly, for y, there exist $m \in \mathbb{N}^\star$, $(\beta_i)_{i\in\overline{1,m}} \subset (0,\infty)$, $\sum_{i=1}^{m} \beta_i = 1$, $(y_i)_{i\in\overline{1,m}} \subset D$ with $y = \sum_{i=1}^{m} \beta_i y_i$. Then

$$\mu_1 x + \mu_2 y = \mu_1 \sum_{i=1}^{n} \alpha_i x_i + \mu_2 \sum_{i=1}^{m} \beta_i y_i = \sum_{i=1}^{n} \mu_1\alpha_i x_i + \sum_{i=1}^{m} \mu_2\beta_i y_i.$$

It is enough to observe that

$$\sum_{i=1}^{n} \mu_1\alpha_i + \sum_{i=1}^{m} \mu_2\beta_i = \mu_1 \sum_{i=1}^{n} \alpha_i + \mu_2 \sum_{i=1}^{m} \beta_i = \mu_1 + \mu_2 = 1$$

in order to deduce that $\mu_1 x + \mu_2 y$ is a convex combination of $n + m$ elements of D, so $\mu_1 x + \mu_2 y \in \operatorname{conv} D$. □

Problem 7.27. *Show that the convex hull of a closed set is not necessarily closed. Show that the convex hull of a compact set is compact.*

Solution For the first part, let us consider the following closed subset of $\mathbb{R}^2$:

$$D = \big([0,\infty) \times \{0\}\big) \cup \big(\{0\} \times [0,1]\big).$$

It is easy to observe that $\operatorname{conv} D = \big([0,\infty) \times [0,1)\big) \cup \{(0,1)\}$. This set is not closed: every point $(a, 1)$ with $a > 1$ is in $\operatorname{cl}\operatorname{conv} D \setminus \operatorname{conv} D$.

For the second part, consider $D \subset \mathbb{R}^p$ a compact convex set. We use the Carathéodory Theorem (Theorem 2.1.17). Since

$$\operatorname{conv} D = \left\{ \sum_{i=1}^{p+1} \alpha_i x_i \mid (\alpha_i)_{i\in\overline{1,p+1}} \subset [0,\infty),\ \sum_{i=1}^{p+1} \alpha_i = 1,\ (x_i)_{i\in\overline{1,p+1}} \subset D \right\},$$

we define the function $f : [0,1]^{p+1} \times D^{p+1} \to \mathbb{R}^p$ given as

$$f(\alpha_1, \ldots, \alpha_{p+1}, x_1, \ldots, x_{p+1}) = \sum_{i=1}^{p+1} \alpha_i x_i$$

and observe that $\operatorname{conv} D = f(M \times D^{p+1})$ where M is the unit simplex of $\mathbb{R}^{p+1}$. Since M and D are compact, we deduce that $M \times D^{p+1}$ is also compact. Moreover, f is continuous, whence $\operatorname{conv} D$ is the image of a compact set through a continuous function. Therefore, $\operatorname{conv} D$ is compact. □

Problem 7.28. *Let $A \subset \mathbb{R}^p$ be a nonempty set. Recall that the conic hull of A is* $\operatorname{cone} A := [0, \infty)A$. *Show that* $\operatorname{cone} A$ *is the smallest cone (with respect to the set inclusion order relation) which contains A. Show that* $\operatorname{cone}(\operatorname{conv} A) = \operatorname{conv}(\operatorname{cone} A)$ *and* $A^- = (\operatorname{cl} \operatorname{cone} A)^-$.

Solution All the affirmations are easy to prove by the use of the definitions of the involved objects. □

Problem 7.29. *Let $f, g : \mathbb{R}^p \to \mathbb{R}$ be convex functions. Show that the function* $\max(f, g)$ *is convex. Is it true for* $\min(f, g)$?

Solution Denote $h : \mathbb{R}^p \to \mathbb{R}$, $h(x) = \max(f, g)$. For every $x, y \in \mathbb{R}^p$ and $\alpha \in (0, 1)$,

$$\begin{aligned} f(\alpha x + (1-\alpha)y) &\le \alpha f(x) + (1-\alpha) f(y) \le \alpha h(x) + (1-\alpha) h(y) \\ g(\alpha x + (1-\alpha)y) &\le \alpha g(x) + (1-\alpha) g(y) \le \alpha h(x) + (1-\alpha) h(y), \end{aligned}$$

so $h(\alpha x + (1-\alpha)y) \le \alpha h(x) + (1-\alpha)h(y)$, whence h is convex. For the $\min(f, g)$ it is not longer true: it is enough to look at the convex functions $f, g : \mathbb{R} \to \mathbb{R}$, $f(x) = x^2$ and $g(x) = (x-1)^2$. □

Problem 7.30. *Let $K \subset \mathbb{R}^p$ be a closed convex cone with nonempty interior which does not coincide with $\mathbb{R}^p$, and let $e \in \operatorname{int} K$. Show that:*

(i) $K + [0, \infty)e \subset K$;

(ii) $K + (0, \infty)e = \operatorname{int} K$;

(iii) $\mathbb{R}e - K = \mathbb{R}^p$;

(iv) for any $x \in \mathbb{R}^p$, $x + \mathbb{R}e \not\subset K$;

Solution Observe first that $e \neq 0$ since otherwise $0 \in \operatorname{int} K$ implies $K = \mathbb{R}^p$ which is impossible.

(i) Since K is a convex cone, $\alpha K \subset K$ for any $\alpha \ge 0$ and $K + K \subset K$. Since $e \in K$, we have successively that $[0, \infty)e \subset K$ and then $K + [0, \infty)e \subset K$.

(ii) We have that $e \in \operatorname{int} K$, so there exists $\varepsilon > 0$ such that $e + B(0, \varepsilon) \subset K$. Then for any $\alpha > 0$, $\alpha e + B(0, \alpha\varepsilon) \subset K$, whence $\alpha e \in \operatorname{int} K$ for any $\alpha > 0$. Fix now $\alpha > 0$ and take $k \in K$. Since $B(0, \alpha\varepsilon)$ is absorbing, there exists $t > 0$ such that $tk \in B(0, \alpha\varepsilon)$.

But since $B(0, \alpha\varepsilon)$ is open, it is a neighborhood of tk, so there exists $\delta > 0$ such that $tk + B(0, \delta) \subset B(0, \alpha\varepsilon)$. Henceforth, $\alpha e + tk + B(0, \delta) \subset \alpha e + B(0, \alpha\varepsilon) \subset K$, so $\alpha e + tk \in \operatorname{int} K$. The inclusion $K + (0, \infty)e \subset \operatorname{int} K$ is proved. Let us prove the converse. Take $v \in \operatorname{int} K$, that is there exists $\varepsilon > 0$ such that $v - D(0, \varepsilon) \subset K$. But $\varepsilon \|e\|^{-1} e \in D(0, \varepsilon)$, whence $\varepsilon \|e\|^{-1} e - v \in -K$, so there exists $k \in K$ with $\varepsilon \|e\|^{-1} e - v = -k$, i.e., $v = k + \varepsilon \|e\|^{-1} e \subset K + (0, \infty)e$. The equality is proved.

(iii) Take $x \in \mathbb{R}^p$. As above, there exists $\varepsilon > 0$ such that $e + B(0, \varepsilon) \subset K$ and $t > 0$ such that $tx \in B(0, \varepsilon)$. Of course, we also have that $-tx \in B(0, \varepsilon)$, and we get that $e - tx \in K$, which means that $x \in t^{-1}e - K \subset \mathbb{R}e - K$.

(iv) Assume that there exists $x \in \mathbb{R}^p$ with $x + \mathbb{R}e \subset K$. Take $k \in K$ and $t \in \mathbb{R}$. The convexity of K allows us to deduce that

$$\frac{n-1}{n}k + \frac{1}{n}(x + tne) \in K$$

for every $n \in \mathbb{N} \setminus \{0\}$. Passing to the limit as $n \to \infty$ and taking into account the closedness of K, we get $k + te \subset K$, so $K + \mathbb{R}e \subset K$. Consequently, $\mathbb{R}e - K \subset -K$. From (iii) we infer that $\mathbb{R}^p \subset -K$, whence $\mathbb{R}^p = K$, which is a contradiction. Therefore, the conclusion holds. □

Problem 7.31. *Show that f is sublinear if and only if its epigraph is a convex cone.*

Solution Clearly the property $f(\alpha x) = \alpha f(x)$ for all $\alpha \geq 0$ and $x \in \mathbb{R}^p$ is equivalent to the fact that the epigraph of f is a cone, while the property $f(x + y) \leq f(x) + f(y)$ for all $x, y \in \mathbb{R}^p$ is equivalent to the fact that this cone is convex. □

Problem 7.32. *Let $K \subset \mathbb{R}^p$ be a closed convex cone with nonempty interior, and let $e \in \operatorname{int} K$. Show that for every $v \in K^- \setminus \{0\}$, $\langle v, e \rangle < 0$.*

Solution The fact that $\langle v, e \rangle \leq 0$ follows from the definition of K^-. Suppose that $\langle v, e \rangle = 0$. Since $e \in \operatorname{int} K$, there exists $\varepsilon > 0$ such that $e + B(0, \varepsilon) \subset K$. Then one obtains that $\langle v, u \rangle \leq 0$ for all $u \in B(0, \varepsilon)$. If follows that $v = 0$, a contradiction. □

Problem 7.33. *Let $K \subset \mathbb{R}^p$ be a closed convex cone with nonempty interior, and let $\emptyset \neq A \subset \mathbb{R}^p$. Show that the following assertions are equivalent:*

(i) there exists $v \in K^- \setminus \{0\}$, $\langle v, a \rangle \leq 0$ for all $a \in A$;

(ii) $\operatorname{conv} A \cap -\operatorname{int} K = \emptyset$.

Solution Firstly, we prove that (i) implies (ii). If $\operatorname{conv} A \cap -\operatorname{int} K \neq \emptyset$, then there exists $u \in \operatorname{conv} A$ with $u \in -\operatorname{int} K$. Then, from the preceding proposition, $\langle v, u \rangle > 0$, which contradicts (i). Whence (ii) holds.

Suppose that (ii) holds. Then, from the convex sets separation theorem, there exists $v \in \mathbb{R}^p \setminus \{0\}$ such that

$$\langle v, a \rangle \leq \langle v, u \rangle, \ \forall a \in \operatorname{conv} A, \ \forall u \in -\operatorname{int} K.$$

Easy arguments show that $v \in K^- \setminus \{0\}$ and $\langle v, a\rangle \le 0$, for all $a \in A$. □

Problem 7.34. *Let $K \subset \mathbb{R}^p$ be a closed convex cone with nonempty interior, and let $\emptyset \neq A \subset \mathbb{R}^p$. Show that the following assertions are equivalent:*

(i) $A \cap -\operatorname{int} K = \emptyset$;

(ii) $\operatorname{cl} A \cap -\operatorname{int} K = \emptyset$;

(iii) $(A + K) \cap -\operatorname{int} K = \emptyset$;

(iv) $\operatorname{cl}(\operatorname{cone}(A + K)) \cap -\operatorname{int} K = \emptyset$.

Solution These assertions are simple applications of the definitions. □

The next result is named after the American mathematician James Caristi who published it in 1976. A multifunction from $\mathbb{R}^p$ to $\mathbb{R}^q$ is an applications which maps every point from $\mathbb{R}^p$ into a subset of $\mathbb{R}^q$.

Theorem 7.2.1 (Caristi Fixed Point Theorem)**.** *Let $\varphi : \mathbb{R}^p \to \mathbb{R}$ be a lower semicontinuous and lower bounded function. Let $T : \mathbb{R}^p \rightrightarrows \mathbb{R}^p$ be a multifunction (with nonempty values) with the property that*

$$\varphi(y) \le \varphi(x) - \|x - y\|, \ \forall x \in \mathbb{R}^p, \ \forall y \in T(x).$$

Then there exists $\overline{x} \in \mathbb{R}^p$ with $\overline{x} \in T(\overline{x})$.

Let us shed light on the links between this result and Ekeland Variational Principle.

Problem 7.35. *Show that Caristi Fixed Point Theorem and the third conclusion from Ekeland Variational Principle are equivalent.*

Solution The proof of Caristi Fixed Point Theorem using Ekeland Variational Principle. For function φ and $\varepsilon > 0$, $\delta := \varepsilon + 1$, we apply Ekeland Variational Principle, and from its third conclusion we deduce that there exists $\overline{x} \in \mathbb{R}^p$ with

$$\varphi(\overline{x}) \le \varphi(x) + \frac{\varepsilon}{\varepsilon + 1} \|x - \overline{x}\|, \ \forall x \in \mathbb{R}^p,$$

from where

$$\varphi(\overline{x}) < \varphi(x) + \|x - \overline{x}\|, \ \forall x \in \mathbb{R}^p \setminus \{\overline{x}\}.$$

If, ab absurdum, $\overline{x} \notin T(\overline{x})$, then for every $y \in T(\overline{x})$ we have $y \neq \overline{x}$, whence

$$\varphi(\overline{x}) < \varphi(y) + \|y - \overline{x}\|,$$

in contradiction to the hypothesis of Caristi Theorem.

The proof of the third conclusion of Ekeland Variational Principle by the use of Caristi Fixed Point Theorem. Suppose that the conclusion does not hold. Then for every $x \in \mathbb{R}^p$, we consider the nonempty set

$$T(x) = \left\{ y \in \mathbb{R}^p \mid y \neq x, \ \frac{\delta}{\varepsilon} f(x) \ge \frac{\delta}{\varepsilon} f(y) + \|y - x\| \right\}.$$

For $\varphi(\cdot) = \frac{\delta}{\varepsilon} f(\cdot)$ we can apply Caristi Theorem, so there exists $\overline{x} \in \mathbb{R}^p$ with $\overline{x} \in T(\overline{x})$, which is, obviously, impossible. □

Problem 7.36. *Show that Caristi Fixed Point Theorem implies the existence part from Banach Fixed Point Principle.*

Solution Let $f : \mathbb{R}^p \to \mathbb{R}^p$ be a contraction of constant $\lambda \in (0, 1)$, which we identify to the mapping T from Caristi Fixed Point Theorem. Let $\varphi : \mathbb{R}^p \to \mathbb{R}$,

$$\varphi(x) = \frac{1}{1-\lambda} \left\|x - f(x)\right\| .$$

Clearly, φ is continuous and lower bounded (by 0). Moreover, the condition from Caristi Fixed Point Theorem is automatically fulfilled since $y \in T(x)$ (which here means $y = f(x)$). So T has a fixed point. □

Definition 7.2.2. *One says that $f : \mathbb{R}^p \to \mathbb{R}^p$ is a directional contraction if it is continuous and there exists $\lambda \in (0, 1)$ such that for every $x \in \mathbb{R}^p$ with $x \neq f(x)$ there exists*

$$y \in (x, f(x)) = \{u \in \mathbb{R}^p \mid \exists t \in (0, 1),\ u = tx + (1-t)f(x)\}$$

with the property

$$\left\|f(x) - f(y)\right\| \leq \lambda \left\|x - y\right\| .$$

Exercise 7.37. *Using the function $f : \mathbb{R}^2 \to \mathbb{R}^2$,*

$$f(x, y) = \left(\frac{3x}{2} - \frac{y}{3}, x + \frac{y}{3}\right),$$

deduce that every contraction is a directional contraction, while the converse is false.

Solution From $\left\|f(x, y) - f(z, y)\right\| = \frac{\sqrt{13}}{2} |x - z|$ for every $x, y, z \in \mathbb{R}$, we deduce that f is not a contraction. Let $(x, y) \in \mathbb{R}^2$ with $f(x, y) \neq (x, y)$. Denote $f(x, y) = (a, b)$. Notice that $f(x, y) = (x, y) \iff a = x \iff b = y$. Consider $x \neq a$ and observe, after calculations, that the point $(u, v) = 2^{-1}(a + x, b + y)$ satisfies the required property with $\lambda = 5/24$, whence f is a directional contraction. □

Problem 7.38. *Let $f : \mathbb{R}^p \to \mathbb{R}^p$ be a directional contraction of constant λ. Then f has a fixed point.*

Solution Let $g : \mathbb{R}^p \to \mathbb{R}$,

$$g(x) = \left\|x - f(x)\right\| .$$

Clearly, g is continuous and lower bounded. We apply the last conclusion of Ekeland Variational Principle for g and for $\varepsilon := \frac{1-\lambda}{2}$, $\delta := 1$. We infer the existence of an element $x \in \mathbb{R}^p$ such that for every $y \in \mathbb{R}^p$

$$\left\|x - f(x)\right\| \leq \left\|y - f(y)\right\| + \frac{1-\lambda}{2} \left\|x - y\right\| .$$

If $x = f(x)$ the proof is over. We want to show that this is the only possible situation. Suppose, by way of contradiction, that $f(x) \neq x$. From the directional contraction condition there exists $y \neq x$ with

$$\|x - y\| + \left\|y - f(x)\right\| = \left\|x - f(x)\right\|$$

and

$$\left\|f(x) - f(y)\right\| \le \lambda \|x - y\| .$$

Therefore, putting together all the relations, we get

$$\begin{aligned} 0 &\le \lambda \|x - y\| - \left\|f(x) - f(y)\right\| \\ &\le \lambda \|x - y\| - \left\|f(y) - y\right\| + \left\|y - f(x)\right\| \\ &= (\lambda - 1) \|x - y\| - \left\|f(y) - y\right\| + \left\|x - f(x)\right\| \\ &\le \frac{\lambda - 1}{2} \|x - y\| . \end{aligned}$$

Since $\lambda < 1$, we obtain that $x = y$, which is a contradiction. Therefore, $f(x) = x$ is the sole possibility, and this ends the proof. □

Exercise 7.39. *Decide if the directional contractions imply the uniqueness of the fixed point.*

Solution It is sufficient to analyze the case of the function from Exercise 7.37, where all the elements of the form $(x, \frac{3x}{2})$, $x \in \mathbb{R}$ are fixed points. So, the uniqueness property of the fixed point do not hold. □

Exercise 7.40. *Let $f, g : \mathbb{R}_+ \to \mathbb{R}$ be defined by*

$$f(x) = \begin{cases} \dfrac{x}{e^x - 1}, & x \neq 0 \\ 1, & x = 0. \end{cases} \quad , \quad g(x) = (x - 2)e^{2x} + (x + 2)e^x .$$

(i) Show that $g(x) \ge 0$, for every $x \in \mathbb{R}_+$.
(ii) Show that f is of class C^1 on $\mathbb{R}_+$.
(iii) Show that

$$f''(x) = \frac{g(x)}{(e^x - 1)^3}, \ \forall x \in (0, \infty)$$

and $\left|f'(x)\right| \le 2^{-1}$ for every $x \in \mathbb{R}_+$.

(iv) One defines the sequence (x_n) by $x_0 = 0$, and $x_{n+1} = f(x_n)$ for every $n \in \mathbb{N}$. Show that

$$|x_n - \ln 2| \le 2^{-n} \ln 2, \ \forall n \in \mathbb{N}.$$

Solution (i) It is enough to prove that $(x - 2)e^x + x + 2 \ge 0$, for every $x \in \mathbb{R}_+$. This can be done easily by studying the variation of this expression through its derivatives (up to the order two).

(ii) Clearly, f is derivable on $(0, \infty)$ and, on this interval,

$$f'(x) = \frac{e^x - 1 - xe^x}{(e^x - 1)^2}.$$

We compute the limit of this derivative at 0 (with a combination between a fundamental limit and the L'Hôpital Rule)

$$\begin{aligned}\lim_{x\to 0+} f'(x) &= \lim_{x\to 0+} \frac{e^x - 1 - xe^x}{(e^x - 1)^2} = \lim_{x\to 0+} \frac{e^x - 1 - xe^x}{x^2} \frac{x^2}{(e^x - 1)^2} \\ &= \lim_{x\to 0+} \frac{e^x - 1 - xe^x}{x^2} = \lim_{x\to 0+} \frac{-e^x}{2} = -\frac{1}{2}.\end{aligned}$$

By the use of some of Lagrange Theorem consequences, we deduce that f is differentiable at 0, and its derivative is continuous at 0. Moreover, $f'(0) = -2^{-1}$.

(iii) The function f is twice derivable on $(0, \infty)$, and the announced relation can be shown by direct calculation. According to (i), f' is increasing on $(0, \infty)$, and the continuity of f', relation $f'(0) = -2^{-1}$ and the remark

$$\lim_{x\to\infty} f'(x) = 0$$

show that $f'(x) \in [-2^{-1}, 0)$ for every $x \in \mathbb{R}_+$, whence the conclusion.

(iv) The preceding step shows that f is a 2^{-1}–contraction on $\mathbb{R}_+$ which takes values in $\mathbb{R}_+$. Since (x_n) is a Picard iteration with the initial data 0, we infer, by the Banach Principle, that (x_n) converges to the unique fixed point of f from $\mathbb{R}_+$, which proves to be, by direct calculus, $\overline{x} = \ln 2$. The estimation follows by induction as

$$|x_n - \overline{x}| = \left|f(x_{n-1}) - f(\overline{x})\right| \le 2^{-1} |x_{n-1} - \overline{x}| \le \ldots \le 2^{-n} |x_0 - \overline{x}| .$$

Consequently, the inequality holds. □

Exercise 7.41. *Let consider the function* $f : \mathbb{R}\setminus\{0\} \to \mathbb{R}$,

$$f(x) = 1 + \frac{1}{4} \sin \frac{1}{x}.$$

For initial data $x_0 \in \mathbb{R}\setminus\{0\}$ *consider the Picard iteration associated to* (x_n). *Study the convergence of this sequence.*

Solution The image of f is the interval $I := \left[\frac{3}{4}, \frac{5}{4}\right]$. We consider the restriction of f to this interval and we show that this is a contraction from I to I. To this end, we compute (for $x \in I$)

$$f'(x) = -\frac{1}{4x^2} \cos \frac{1}{x},$$

from where

$$\left|f'(x)\right| \le \frac{4}{9} < 1, \ \forall x \in I.$$

Since $x_1 = f(x_0) \in I$, we can apply Banach Principle in order to obtain that (x_n) is convergent towards the unique fixed point of f from I.

Notice that the approximate value of the fixed point of f can be found by using the Matlab code given in the previous chapter. □

Exercise 7.42. *Let consider the equation $x^3 - x - 1 = 0$ for $x \in I := [1, 2]$.*

(i) Transform this equation into a problem of finding a fixed point for a suitable contraction.

(ii) Deduce the existence and the uniqueness of the solution of the initial equation and indicate a sequence (x_n) convergent towards this solution. Determine a sufficient number of terms to be computed in order to approximate the solution with a less that 10^{-5} error.

Solution (i) The equation is equivalent to

$$x = x^3 - 1,$$

but in this formulation we should take $g(x) := x^3 - 1$, but $g([1, 2]) \not\subset [1, 2]$, and g is not a contraction. So, we write the initial equation equivalently as

$$x^3 = x + 1 \Leftrightarrow x = \sqrt[3]{x + 1}.$$

Consider $f : I \to \mathbb{R}$, $f(x) = \sqrt[3]{x + 1}$. It is easy to observe that $f(I) \subset I$ and

$$f'(x) = \frac{1}{3\sqrt[3]{(x+1)^2}} \leq \frac{1}{3\sqrt[3]{4}} < 1,$$

so f is a contraction from I to I.

(ii) From the Banach Principle and the above formulation we infer the existence and the uniqueness of the solution (denoted by $\overline{x}$) of the initial equation. Every Picard iteration associated to f is convergent to the solution. We take $x_0 = 1$ and $x_{n+1} = f(x_n)$ for every $n \in \mathbb{N}$. Furthermore, for every n,

$$|x_n - \overline{x}| \leq \left(\frac{1}{3\sqrt[3]{4}}\right)^n |x_0 - \overline{x}| \leq \left(\frac{1}{3\sqrt[3]{4}}\right)^n.$$

It is then sufficient to estimate n for which

$$\left(\frac{1}{3\sqrt[3]{4}}\right)^n \leq 10^{-5}.$$

Moreover, it is sufficient to have

$$\frac{1}{4^n} \leq 10^{-5},$$

that is $n \geq 5 \log_4 10$. Therefore, $n = 9$ satisfies the requirement. □

Problem 7.43. *Show that every weak contraction defined on a compact set K and taking values into K has a unique fixed point.*

Solution The function f is continuous (being Lipschitz). Further, the function $x \in K \mapsto \left\|f(x) - x\right\|$ is continuous on K, hence admits a global minimum point. Consequently, there exists $\overline{x} \in K$ such that

$$\left\|f(\overline{x}) - \overline{x}\right\| \le \left\|f(x) - x\right\|, \ \forall x \in K.$$

If $f(\overline{x}) \neq \overline{x}$, then, from weak contraction property,

$$\left\|f(\overline{x}) - f(f(\overline{x}))\right\| < \left\|f(\overline{x}) - \overline{x}\right\|.$$

Since $f(\overline{x}) \in K$, these two relations are in contradiction. So $f(\overline{x}) = \overline{x}$. The uniqueness is obvious. □

Problem 7.44. *Let $a, b \in \mathbb{R}$, $a < b$, and $f : [a, b] \to \mathbb{R}$ be a continuous function. Show that the next assertions hold:*

(i) If $[a, b] \subset f([a, b])$, then f has at least a fixed point.

(ii) Every closed interval from $f([a, b])$ is the image of a closed interval from $[a, b]$.

(iii) Let $n \in \mathbb{N}^\star$. If there exist n closed intervals $I_0, I_1, \ldots, I_{n-1}$ contained in $[a, b]$, such that for every $k \in \overline{0, n-2}$, $I_{k+1} \subset f(I_k)$ and $I_0 \subset f(I_{n-1})$, then f^n has at least one fixed point.

Solution (i) By continuity, $f([a, b])$ is a compact interval which we denote by $[m, M]$, where $m, M \in \mathbb{R}$. By hypothesis $[a, b] \subset f([a, b])$, we deduce that

$$m \le a < b \le M.$$

Since $m, M \in f([a, b])$, there exist $x_m, x_M \in [a, b]$ such that $m = f(x_m)$ and $M = f(x_M)$. But

$$f(x_m) - x_m = m - x_m \le a - x_m \le 0$$

and

$$f(x_M) - x_M = M - x_M \ge b - x_M \ge 0,$$

and the function $g(\cdot) = f(\cdot) - \cdot$ vanishes in $[a, b]$, that is f has a fixed point in this interval.

(ii) Let $I = [c, d] \subset f([a, b])$. Obviously, there exist $u, v \in [a, b]$ such that $f(u) = c$ and $f(v) = d$. We can suppose that $u \le v$. We consider the set $A := \{x \in [u, v] \mid f(x) = c\}$. This set is bounded (as a subset of $[a, b]$), nonempty ($u \in A$) and closed (since f is continuous and $A = [u, v] \cap f^{-1}(c)$) so, there exist $\alpha = \max A \in A$. Similarly, the set $B := \{x \in [\alpha, v] \mid f(x) = d\}$ has a minimum point, denoted by β. Then, $f(\alpha) = c, f(\beta) = d$ and for every $x \in (\alpha, \beta)$ one has $f(x) \neq c$ and $f(x) \neq d$. From the Darboux property, $(c, d) \subset f((\alpha, \beta))$, and the interval $f((\alpha, \beta))$ does not contain the points c and d. Consequently, $[c, d] = f([\alpha, \beta])$.

(iii) From the inclusion $I_0 \subset f(I_{n-1})$ and from (ii), there exists a closed interval $J_{n-1} \subset I_{n-1}$ such that $I_0 = f(J_{n-1})$. But $J_{n-1} \subset I_{n-1} \subset f(I_{n-2})$. Again, using (ii), there exists a closed interval $J_{n-2} \subset I_{n-2}$ such that $J_{n-1} = f(J_{n-2})$. We repeat this argument, and we infer the existence of n closed interval $J_0, J_1, \dots, J_{n-1}$ such that

$$J_k \subset I_k, \ \forall k \in \overline{0, n-1},$$

and

$$J_{k+1} = f(J_k), \forall k \in \overline{0, n-2} \text{ and } I_0 = f(J_{n-1}).$$

So,

$$J_0 \subset I_0 = f(J_{n-1}) = f(f(J_{n-2})) = \dots = f^n(J_0).$$

We now apply (i) for the continuous function f^n and for interval J_0, and we get the conclusion. □

Problem 7.45. *Let $f : [0, 1] \to [0, 1]$ be a continuous function with $f(0) = 0$ and $f(1) = 1$. Show that there exist $m \in \mathbb{N}^\star$ such that $f^m(x) = x$ for every $x \in [0, 1]$, then $f(x) = x$ for every $x \in [0, 1]$.*

Solution Since f^m is a bijection, it follows that f itself is a bijection, from a well-known result concerning the injectivity and surjectivity of the compositions. The continuity ensures through Theorem 1.2.27 that f is strictly monotone, and since $f(0) = 0$, $f(1) = 1$, we deduce that f is strictly increasing. Suppose, by way of contradiction, that there exists $x \in (0, 1)$ such that $f(x) > x$ (the case $f(x) < x$ is similar). Then from the monotonicity, for every $n \in \mathbb{N}$,

$$f^n(x) > f^{n-1}(x) > \dots > f(x) > x.$$

In particular, for $n = m$, we get a contradiction, hence the assumption made was false. □

Problem 7.46. *Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous function such that $f \circ f$ has a fixed point. Show that f has a fixed point.*

Solution If we suppose that f has no fixed point, from continuity, it follows that either $f(x) > x$, for every $x \in \mathbb{R}$, or $f(x) < x$, for every $x \in \mathbb{R}$. In the first situation, passing x into $f(x)$ we get: $f(f(x)) > f(x) > x$, for every $x \in \mathbb{R}$, so $(f \circ f)(x) > x$, for every $x \in \mathbb{R}$. Therefore, $f \circ f$ has no fixed point, which contradicts the assumptions. The second case is similar. □

Problem 7.47. *Let $a, b \in \mathbb{R}$, $a < b$, and $f : [a, b] \to [a, b]$ be a Lipschitz function of constant $L > 0$. Define the sequence $(x_n)_{n \in \mathbb{N}^\star}$ by $x_0 \in [a, b]$, and for every $n \geq 0$,*

$$x_{n+1} = (1 - \lambda)x_n + \lambda f(x_n),$$

where $\lambda := \frac{1}{L+1}$. Show that (x_n) is monotone and convergent towards a fixed point of f.

Solution We observe that if one of the terms x_n is a fixed point for f, then starting from n, the sequence is stationary, and the conclusion follows. Suppose that $f(x_n) \neq x_n$ for every $n \in \mathbb{N}^*$. Without loss of generality, suppose that $f(x_0) > x_0$, since the opposite case is similar. Since $f(b) \leq b$, the continuity of f tells us that there exists a fixed point in the interval $(x_0, b]$. Also from continuity, there exists the least fixed point, denoted by p, in this interval (otherwise, x_0 would be itself a fixed point since the set of fixed points is closed). Let us observe that

$$x_1 = (1 - \lambda)x_0 + \lambda f(x_0) > x_0.$$

We want to show, by induction, that the sequence is increasing and for every $n \in \mathbb{N}^*$, we have $x_n < p$ and $x_n < f(x_n)$. Suppose that these relations hold up to rank n, and we show it for rank $n + 1$. Suppose, by contradiction, that $p < x_{n+1}$. Then $x_n < p < x_{n+1}$, whence

$$0 < p - x_n < x_{n+1} - x_n = \lambda(f(x_n) - x_n),$$

from where

$$\begin{aligned} 0 &< \frac{1}{\lambda} |x_n - p| = (L + 1) |x_n - p| \\ &< \left|f(x_n) - x_n\right| \leq \left|f(x_n) - f(p)\right| + |p - x_n|, \end{aligned}$$

whence

$$L |x_n - p| < \left|f(x_n) - f(p)\right|$$

which contradicts the Lipschitz property of f. So $x_{n+1} < p$. Moreover,

$$x_{n+1} = (1 - \lambda)x_n + \lambda f(x_n) > x_n.$$

If we would have $f(x_{n+1}) < x_{n+1}$, then between x_n and x_{n+1}, a fixed point would exist, but since $x_n > x_0$ and $x_{n+1} < p$, this would contradict the choice of p. Therefore, the claims are proved. Now, since (x_n) is monotone and bounded, it converges towards a point $\overline{x} \in [a, b]$. We have

$$\begin{aligned} \left|\overline{x} - f(\overline{x})\right| &\leq |\overline{x} - x_n| + \left|x_n - f(x_n)\right| + \left|f(x_n) - f(\overline{x})\right| \\ &= |\overline{x} - x_n| + \frac{1}{\lambda} |x_n - x_{n+1}| + \left|f(x_n) - f(\overline{x})\right|. \end{aligned}$$

At this moment, it is obvious that the right-hand side goes to 0 for $n \to \infty$ and we get $\overline{x} = f(\overline{x})$. □

7.3 Smooth Optimization

Exercise 7.48. *Find the local extrema of the following functions:*

(i) $f : \mathbb{R}^2 \to \mathbb{R}$, $f(x_1, x_2) = 6x_1^2 x_2 + 2x_2^3 - 45x_1 - 51x_2 + 7$;

(ii) $f : \mathbb{R}^3 \setminus \{(0,0,0)\} \to \mathbb{R}$, $f(x_1, x_2, x_3) = \frac{x_1}{x_2} + \frac{x_2}{4} + \frac{x_3}{x_1} + \frac{1}{x_3}$;
(iii) $f : \mathbb{R}^2 \to \mathbb{R}$, $f(x_1, x_2) = x_1 x_2 (x_1^2 + x_2^2 - 4)$;
(iv) $f : \mathbb{R}^3 \to \mathbb{R}$, $f(x_1, x_2, x_3) = x_1^4 + x_2^3 + x_3^2 + 4x_1x_3 - 3x_2 + 2$;
(v) $f : \mathbb{R}^2 \to \mathbb{R}$, $f(x_1, x_2) = x_1^4 + x_2^4$;
(vi) $f : \mathbb{R}^2 \to \mathbb{R}$, $f(x_1, x_2) = x_1^2 + x_2^3$;
(vii) $f : \mathbb{R}^2 \to \mathbb{R}$, $f(x_1, x_2) = x_1 x_2^2 e^{x_1 - x_2}$.

Solution We have to deal with nonlinear optimization problems without restrictions. The general method for finding the extrema is as follows. One finds the stationary (critical) points by solving the equation $\nabla f(x) = 0$. In everyone of these points one computes $\nabla^2 f(\overline{x})$, which in fact identifies to Hessian matrix.

- If $\nabla^2 f(\overline{x})$ is positive definite, then $\overline{x}$ is a local minimum;
- if $\nabla^2 f(\overline{x})$ is negative definite, then $\overline{x}$ is a local maximum;
- if $\nabla^2 f(\overline{x})$ is indefinite, then $\overline{x}$ is not a local extremum point.

In order to verify these aspects, in some cases, one can use the method described after Corollary 3.1.29:

- if the determinants of the matrices $\left(\frac{\partial^2 f}{\partial x^i \partial x^j}(\overline{x})\right)_{i,j \in \overline{1,k}}$, $k \in \overline{1,p}$ are strictly positive, then $\overline{x}$ is a local minimum;
- if the determinants of the matrices $\left(\frac{\partial^2 f}{\partial x^i \partial x^j}(\overline{x})\right)_{i,j \in \overline{1,k}}$, $k \in \overline{1,p}$ are nonzero and alternate their signs starting as negative, then $\overline{x}$ is a local maximum;
- if the determinants of the matrices $\left(\frac{\partial^2 f}{\partial x^i \partial x^j}(\overline{x})\right)_{i,j \in \overline{1,k}}$, $k \in \overline{1,p}$ are nonzero, then every other configuration of signs apart from those described above implies that the point is not an extremum.

If no one of the above conclusions apply, then one should consider every case in its particularities in order to decide the nature of the critical point.

(i) We solve the system coming from relation $\nabla f(x) = 0$ in order to find the critical points. We obtain the system

$$\begin{aligned} 12x_1y_1 &= 45 \\ 6x_1^2 + 6x_2^2 &= 51 \end{aligned}$$

which have the solutions $\left(\frac{3}{2}, \frac{5}{2}\right), \left(\frac{5}{2}, \frac{3}{2}\right), \left(-\frac{3}{2}, -\frac{5}{2}\right), \left(-\frac{5}{2}, -\frac{3}{2}\right)$.

Therefore, by the application of the above method, we obtain the conclusions: $\left(\frac{3}{2}, \frac{5}{2}\right)$ is a local minimum, $\left(\frac{5}{2}, \frac{3}{2}\right), \left(-\frac{5}{2}, -\frac{3}{2}\right)$ are not local extrema and $\left(-\frac{3}{2}, -\frac{5}{2}\right)$ is a local maximum.

The item (*ii*) is similar.

(v) The only critical point is $(0,0)$, but the determinants given by the Hessian matrix are zero, so we cannot decide on this basis. Nevertheless, it is easy to observe that

$$f(x_1, x_2) \geq 0 = f(0,0), \ \forall (x_1, x_2) \in \mathbb{R}^2,$$

so $(0,0)$ is a global minimum.

(vi) Again, $(0,0)$ is the unique critical point, but we cannot decide its nature on the above theory. Observe that $f(0,0) = 0$, but for the sequence $x_n = \left(\frac{1}{n}, 0\right) \to (0,0)$, $f(x_n) > 0$, while for $y_n = (0, -\frac{1}{n}) \to (0,0)$, $f(y_n) < 0$. So, in every neighborhood of $(0,0)$ there exist points where the objective function takes greater or smaller values. Therefore, the point is not a local extremum.

For the other items one proceeds similarly: there exist critical points where the above method works and critical points where we should use the structure of the problem. There are as well situations when one cannot decide if $\nabla^2 f(\overline{x})$ is positive (negative) definite or not, by using the direct calculus of it, the Sylvester Criterion being not applicable. $\square$

Problem 7.49. *(i) Let us consider a differentiable function $f : \mathbb{R} \to \mathbb{R}$. Show that if it has only one critical (stationary) point $\overline{x}$, which turns to be local extremum, then it is necessarily a global extremum.*

(ii) By the study of the function $f : \mathbb{R}^p \to \mathbb{R}$ $(p \geq 2)$ defined by

$$f(x) = (1 + x_p)^3 \sum_{k=1}^{p-1} x_k^2 + x_p^2,$$

show that the assertion of (i) is no longer true in the case of several variables.

Solution (i) Suppose that $\overline{x}$ is a local minimum and it is not a global minimum point. Then, it would exist $u \in \mathbb{R}$ with $f(u) < f(\overline{x})$. One can suppose that $u < \overline{x}$. But, from the local minimality of $\overline{x}$, there exists $v \in \mathbb{R}$ with $u < v < \overline{x}$ and $f(\overline{x}) < f(v)$ (otherwise, f would be constant on an interval $(\overline{x} - \varepsilon, \overline{x})$ and therefore $\overline{x}$ would fail to be the only critical point). Therefore, $f(u) < f(\overline{x}) < f(v)$, whence the value $f(\overline{x})$ is attained inside the interval (u, v), at a point which we denote by w. By virtue of Rolle Theorem applied to f on $[w, \overline{x}]$, there exists $t \in (w, \overline{x})$ with $f'(t) = 0$, which is a contradiction.

(ii) We show that $0 \in \mathbb{R}^p$ is the only critical point of f, and it is a local strict minimum (of order $\alpha = 2$), but it is not a global minimum point (an illustration of the case $p = 2$ is given in the figure below).

We have an optimization problem without restrictions. After computation,

$$\frac{\partial f}{\partial x_k}(x) = 2x_k(1 + x_p)^3, \ \forall k \in \overline{1, p-1},$$

$$\frac{\partial f}{\partial x_p}(x) = 3(1 + x_p)^2 \sum_{k=1}^{p-1} x_k^2 + 2x_p,$$

and the only critical point $\overline{x}$ (i.e. $\nabla f(\overline{x}) = 0$, that is $\frac{\partial f}{\partial x_k}(\overline{x}) = 0$ for $k \in \overline{1, p}$) is $\overline{x} = 0$. An easy calculus shows that the Hessian matrix of f at $\overline{x}$ is the square matrix of dimensions $p \times p$ having the number 2 on the main diagonal and 0 in all the other positions, whence it is positive definite. According to Corollary 3.1.29, we deduce that

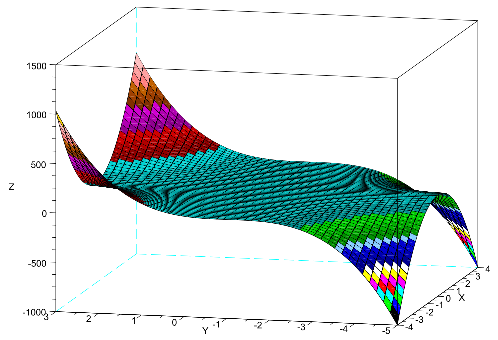

Figure 7.1: $f(x) = (1 + x_2)^3 x_1^2 + x_2^2$.

$\overline{x}$ is a local strict solution of order two. Let us observe that $f(1, 1, \ldots, x_p) = (p-1)(1 + x_p)^3 + x_p^2$ is a third degree polynomial expression which attains, when $x_p \in \mathbb{R}$, all the real values. It follows that f cannot have a global minimum. □

Exercise 7.50. *Let $f : \mathbb{R}^2 \to \mathbb{R}$, $f(x_1, x_2) = 3x_1^4 - 4x_1^2 x_2 + x_2^2$. Find the minimum points of the function f.*

Solution Let us compute the critical points. The system

$$\begin{cases} \frac{\partial f}{\partial x_1}(x_1, x_2) = 0 \\ \frac{\partial f}{\partial x_2}(x_1, x_2) = 0 \end{cases}$$

has as a unique solution the element $\overline{x} = (0, 0)$. However, the sufficient optimality condition of order two is not satisfied, since the Hessian matrix of f at $\overline{x}$ is the matrix $\begin{pmatrix} 0 & 0 \\ 0 & 2 \end{pmatrix}$. Therefore, we cannot decide, on the basis of Corollary 3.1.29, if $\overline{x}$ is a minimum point. In such cases, we use the structure of the problem to get the conclusion. In our specific case, we observe that $f(x_1, x_2) = (x_1^2 - x_2)(3x_1^2 - x_2)$, and for the sequence $x_k = (\sqrt{k^{-1}}, -k^{-1}) \to \overline{x}$

$$f(x_k) = \frac{8}{k^2} > f(\overline{x})$$

while, for the sequence $x_k' = (\sqrt{(2k)^{-1}}, k^{-1}) \to \overline{x}$,

$$f(x_k') = -\frac{1}{4k^2} < f(\overline{x}).$$

Then, $\overline{x}$ is not an extremum point of f. The picture of the localization of the graph of f around $\overline{x}$ shows what happens there. □

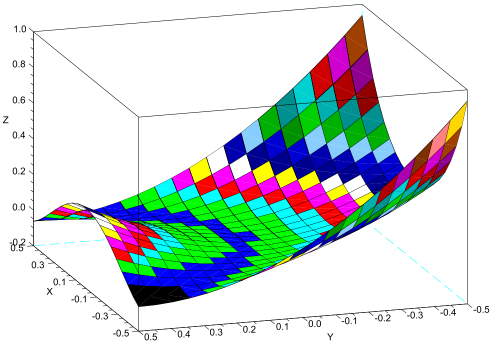

Figure 7.2: $f(x) = 3x_1^4 - 4x_1^2x_2 + x_2^2$.

Exercise 7.51. *Consider the Rosenbrock function* $f : \mathbb{R}^2 \to \mathbb{R}$, $f(x_1, x_2) = (1 - x_1)^2 + 100(x_2 - x_1^2)^2$. *Find the minima of this function.*

Solution The graph of f is depicted below.

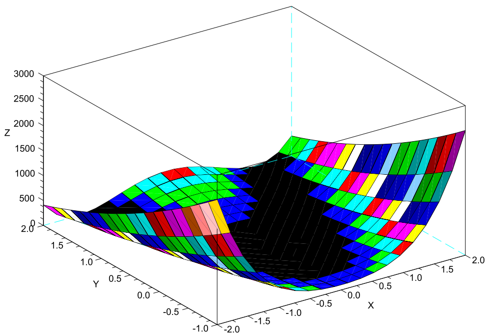

Figure 7.3: The Rosenbrock function.

The system

$$\begin{cases} \frac{\partial f}{\partial x_1}(x_1, x_2) = 0 \\ \frac{\partial f}{\partial x_2}(x_1, x_2) = 0 \end{cases}$$

is equivalent to

$$\begin{cases} -2(1-x_1) - 400x_1(x_2 - x_1^2) = 0 \\ 200(x_2 - x_1^2) = 0, \end{cases}$$

and its unique solution is $(x_1, x_2) = (1, 1)$. One could observe that $f(1, 1) = 0$, and $f(x_1, x_2) \geq 0$ for every $(x_1, x_2) \in \mathbb{R}^2$, whence $(1, 1)$ is a global minimum. In fact, the Hessian matrix at this point

$$\begin{pmatrix} 802 & -400 \\ -400 & 200 \end{pmatrix}$$

is positive definite, whence $(1, 1)$ is a local strict solution of order two.

This function is called the Rosenbrock function and it is used in order to test numerical algorithms for the approximation of solutions. Generally speaking, because of the fact that the minimum point is situated in a relatively planar region (see the graph), this is not easy to approximate numerically. □

Exercise 7.52. *Find the global extrema of $f : \mathbb{R}^3 \to \mathbb{R}$, $f(x_1, x_2, x_3) = x_1^3 + x_2^3 + x_3^3$ on the sphere $x_1^2 + x_2^2 + x_3^2 = 4$.*

Solution We have a problem with a restriction given as an equality $h(x) = 0$, where

$$h : \mathbb{R}^3 \to \mathbb{R},\ h(x_1, x_2, x_3) = x_1^2 + x_2^2 + x_3^2 - 4.$$

Let us observe that for every $x \neq 0$, $\nabla h(x) \neq 0$. Since $x = 0$ is not feasible, the linear independence condition holds for every feasible point.

On the other hand, since the sphere is compact and f is continuous, from Weierstrass Theorem, the global extrema of the problem do exist. We observe as well that since we have only equalities (in fact a single one) as restrictions the necessary optimality conditions from Karush-Kuhn-Tucker Theorem look the same for both minima and maxima. Applying these conditions, we get the system

$$\begin{aligned} x_1(3x_1 + 2\mu) &= 0 \\ x_2(3x_2 + 2\mu) &= 0 \\ x_3(3x_3 + 2\mu) &= 0 \\ x_1^2 + x_2^2 + x_3^2 &= 4, \end{aligned}$$

where μ is a real number. We distinguish between several situations.

- If $x_1 = x_2 = x_3 = 0$, then the point is not feasible.
- If $x_1 = x_2 = 0$ and $x_3 \neq 0$, we get $x_3 = \pm 2$ (μ is not important at this stage).
- If $x_1 = 0$, $x_2, x_3 \neq 0$, we get $x_2 = x_3 = \pm\sqrt{2}$.
- If $x_1, x_2, x_3 \neq 0$, we get $x_1 = x_2 = x_3 = \pm\frac{2}{\sqrt{3}}$.

Taking into account the fact that the other cases are symmetric to these ones, we finally get the points

$$(\pm 2, 0, 0); (0, \pm 2, 0); (0, 0, \pm 2);$$

$$(0, \sqrt{2}, \sqrt{2}); (0, -\sqrt{2}, -\sqrt{2}); (\sqrt{2}, 0, \sqrt{2});$$
$$(-\sqrt{2}, 0, -\sqrt{2}); (\sqrt{2}, \sqrt{2}, 0); (\sqrt{2}, \sqrt{2}, 0);$$
$$\left(\frac{2}{\sqrt{3}}, \frac{2}{\sqrt{3}}, \frac{2}{\sqrt{3}}\right); \left(-\frac{2}{\sqrt{3}}, -\frac{2}{\sqrt{3}}, -\frac{2}{\sqrt{3}}\right).$$

By direct computation of the function values, we infer that the maximum value is 8 and it is attained at $(2, 0, 0)$, $(0, 2, 0)$, $(0, 0, 2)$, and the minimum value is -8, and it is attained at $(-2, 0, 0)$, $(0, -2, 0)$, $(0, 0, -2)$. □

Exercise 7.53. *Find the global extrema of $f : \mathbb{R}^3 \to \mathbb{R}$, $f(x_1, x_2, x_3) = x_1^3 + x_2^3 + x_3^3$ on the set of points which satisfy $x_1^2 + x_2^2 + x_3^2 = 4$ and $x_1 + x_2 + x_3 = 1$.*

Solution The points we are looking for do exist by the same reasons as before. We have a restriction of the form $h(x) = 0$, where

$$h : \mathbb{R}^3 \to \mathbb{R}^2,\ h(x_1, x_2, x_3) = (x_1^2 + x_2^2 + x_3^2 - 4, x_1 + x_2 + x_3 - 1).$$

The linear independence condition is satisfied for all triples (x_1, x_2, x_3) for which at least two components are different. Since there are no feasible points with $x_1 = x_2 = x_3$, we deduce that the qualification condition do hold, so the extreme points are among the critical points of the Lagrangian. We obtain the system

$$\begin{aligned}
3x_1^2 + 2\mu_1 x_1 + \mu_2 &= 0\\
3x_2^2 + 2\mu_1 x_2 + \mu_2 &= 0\\
3x_3^2 + 2\mu_1 x_3 + \mu_2 &= 0\\
x_1^2 + x_2^2 + x_3^2 &= 4\\
x_1 + x_2 + x_3 &= 1,
\end{aligned}$$

where $\mu \in \mathbb{R}$. In order to have a compatible system in the variables μ_1, μ_2 from the first three equations, it is necessary and sufficient that

$$\begin{vmatrix} 3x_1^2 & 2x_1 & 1\\ 3x_2^2 & 2x_2 & 1\\ 3x_3^2 & 2x_3 & 1 \end{vmatrix} = 0,$$

whence

$$\begin{aligned}
(x_1 - x_2)(x_1 - x_3)(x_2 - x_3) &= 0\\
x_1^2 + x_2^2 + x_3^2 &= 4\\
x_1 + x_2 + x_3 &= 1.
\end{aligned}$$

We obtain

$$\left(\frac{1}{3} + \frac{\sqrt{22}}{6}, \frac{1}{3} + \frac{\sqrt{22}}{6}, \frac{1}{3} - \frac{\sqrt{22}}{3}\right)$$

$$\left(\frac{1}{3} - \frac{\sqrt{22}}{6}, \frac{1}{3} - \frac{\sqrt{22}}{6}, \frac{1}{3} + \frac{\sqrt{22}}{3}\right)$$

and their permutations. It is not difficult to verify that the first point (and its permutations) corresponds to the maximum, and the second one to the minimum. □

Exercise 7.54. *Let $n \geq 3$ and $a_i > 0$, for every $i \in \overline{1, n}$. Determine the minimum of*

$$\sum_{i=1}^{n} a_i x_i^2$$

under the constraint

$$\sum_{i=1}^{n} x_i = c,$$

where c is a given constant. What is the maximum of the above expression under the same constraint?

Solution Let us observe that if we denote by $M := \{x \in \mathbb{R}^n \mid \sum_{i=1}^{n} x_i = c\}$ the set of feasible points, and by $f : \mathbb{R}^n \to \mathbb{R}$, $f(x) = \sum_{i=1}^{n} a_i x_i^2$ the objective function, then for $v \geq \sum_{i=1}^{n} \frac{a_i c^2}{n^2}$, the set $M \cap N_v f$ is nonempty (it contains, for instance, the element $(cn^{-1}, \ldots, cn^1) \in \mathbb{R}^n$) and bounded. Hence, according to Theorem 3.1.7, there exists the minimum of the proposed problem. Notice that the restriction function is affine, so the quasiregularity qualification condition is fulfilled. The Karush-Kuhn-Tucker conditions say that there exists $\mu \in \mathbb{R}$ such that

$$2a_i x_i + \mu = 0, \ \forall i \in \overline{1, n},$$

whence

$$x_i = -\frac{\mu}{2a_i}.$$

We replace this into restriction and we get

$$c = -\sum_{i=1}^{n} \frac{\mu}{2a_i},$$

so

$$\mu = -\frac{2c}{\sum_{i=1}^{n} \frac{1}{a_i}}.$$

Therefore,

$$x_i = \frac{c}{a_i \sum_{i=1}^{n} \frac{1}{a_i}}, \ \forall i \in \overline{1, n}.$$

The given expression does not admit a maximum: we observe that for the sequence of feasible points $(p, -p, c, 0 \ldots, 0)_{p \in \mathbb{N}}$, the value of f goes to $+\infty$. □

Exercise 7.55. *Minimize* $f : \mathbb{R}^3 \to \mathbb{R}$,

$$f(x_1, x_2, x_3) = x_3 + \frac{1}{2}\left(x_1^2 + x_2^2 + \frac{x_3^2}{10}\right)$$

under constraints $x_1 + x_2 + x_3 = r$ $(r > 0)$, $x_1 \geq 0$, $x_2 \geq 0$, $x_3 \geq 0$.

Solution The existence of minimum is ensured by Weierstrass Theorem. The constraints are linear, so we do not have to verify the qualification conditions. In order to bring the problem to the standard form, the inequalities restrictions are written as $-x_1 \leq 0, -x_2 \leq 0, -x_3 \leq 0$. Karush-Kuhn-Tucker conditions ensure that at a minimum point x there are $\lambda_1, \lambda_2, \lambda_3 \geq 0$ and $\mu \in \mathbb{R}$ such that

$$\begin{aligned} x_1 - \lambda_1 + \mu &= 0 \\ x_2 - \lambda_2 + \mu &= 0 \\ 1 + \frac{x_3}{10} - \lambda_3 + \mu &= 0 \\ \lambda_1 x_1 &= 0 \\ \lambda_2 x_2 &= 0 \\ \lambda_3 x_3 &= 0 \\ x_1 + x_2 + x_3 &= r. \end{aligned}$$

After the study of all the possibilities, we find the solution

– for $r \leq 2$,

$$(x_1, x_2, x_3) = \left(\frac{r}{2}, \frac{r}{2}, 0\right) \text{ and } (\lambda_1, \lambda_2, \lambda_3, \mu) = \left(0, 0, 1 - \frac{r}{2}\right);$$

– for $r > 2$,

$$\begin{aligned} &(x_1, x_2, x_3) = \left(\tfrac{r+10}{12}, \tfrac{r+10}{12}, \tfrac{5(r-2)}{6}\right) \text{ and} \\ &(\lambda_1, \lambda_2, \lambda_3, \mu) = \left(0, 0, 0, -\tfrac{10+r}{12}\right). \end{aligned}$$

The problem is solved. □

Exercise 7.56. *Maximize* $f : \mathbb{R}^3 \to \mathbb{R}$, $f(x) = x_1 x_2 x_3$ *under the restrictions* $x_1, x_2, x_3 \geq 0$, $2x_1 + 2x_2 + 4x_3 \leq a$, *where* $a > 0$.

Solution It is clear that the maximum is attained (Weierstrass Theorem) and it is strictly positive. For standardization, we deal with the problem

$$\min \ -x_1 x_2 x_3$$

$-x_1, -x_2, -x_3 \leq 0$, $2x_1 + 2x_2 + 4x_3 \leq a$. Following the theory for a minimum point x, there exist $\lambda_1, \lambda_2, \lambda_3, \lambda_4 \geq 0$ such that

$$-x_2 x_3 - \lambda_1 + 2\lambda_4 = 0$$

$$\begin{aligned}
-x_1x_3 - \lambda_2 + 2\lambda_4 &= 0\\
-x_1x_2 - \lambda_3 + 4\lambda_4 &= 0\\
\lambda_1 x_1 &= 0\\
\lambda_2 x_2 &= 0\\
\lambda_3 x_3 &= 0\\
(2x_1 + 2x_2 + 4x_3 - a)\lambda_4 &= 0.
\end{aligned}$$

We infer that

$$\begin{aligned}
x_1(-x_2x_3 + 2\lambda_4) &= 0\\
x_2(-x_1x_3 + 2\lambda_4) &= 0\\
x_3(-x_1x_2 + 4\lambda_4) &= 0\\
(2x_1 + 2x_2 + 4x_3 - a)\lambda_4 &= 0
\end{aligned}$$

We add the first three equations and we get $-3x_1x_2x_3+\lambda_4(2x_1+2x_2+4x_3) = 0$, therefore

$$-3x_1x_2x_3 + a\lambda_4 = 0,$$

so $\lambda_4 = \frac{3x_1x_2x_3}{a}$. Replacing in $x_1(-x_2x_3 + 2\lambda_4) = 0$, since $x_1x_2x_3 \neq 0$, we find $x_1 = \frac{a}{6}$. Analogously, $x_2 = \frac{a}{6}$, $x_3 = \frac{a}{12}$. The multipliers are $(0, 0, 0, \frac{a^2}{144})$. So, the solution of the problem is $(\frac{a}{6}, \frac{a}{6}, \frac{a}{12})$. □

Exercise 7.57. *Find the extrema of* $f : \mathbb{R}^3 \to \mathbb{R}$, $f(x) = x_1x_2x_3$ *under the restriction* $h(x) = 0$, *where* $h : \mathbb{R}^3 \to \mathbb{R}^2$, $h(x) = (x_1x_2 + x_1x_3 + x_2x_3 - 8, x_1 + x_2 + x_3 - 5)$.

Solution We show that the set of feasible points is bounded (hence compact).

Using the equality $x_3 = 5 - x_1 - x_2$, we can eliminate x_3 from the first restriction, that is

$$x_1^2 + x_2^2 + x_1x_2 - 5x_1 - 5x_2 + 8 = 0,$$

whence

$$\left(\frac{x_1}{\sqrt{2}} + \frac{x_2}{\sqrt{2}}\right)^2 + \left(\frac{x_1}{\sqrt{2}} - \frac{5}{\sqrt{2}}\right)^2 + \left(\frac{x_2}{\sqrt{2}} - \frac{5}{\sqrt{2}}\right)^2 = 17.$$

We deduce that x_1 and x_2 lie in a bounded set, as the x_3 as well. Therefore, the set of feasible points is compact and both minimization and maximization problems have solutions.

We verify the linear independence qualification condition: we ask if there exist two real numbers α_1, α_2 with $(\alpha_1, \alpha_2) \neq (0, 0)$ such that

$$\begin{aligned}
\alpha_1(x_2 + x_3) + \alpha_2 &= 0\\
\alpha_1(x_1 + x_3) + \alpha_2 &= 0\\
\alpha_1(x_1 + x_2) + \alpha_2 &= 0.
\end{aligned}$$

Since we are interested only on the set of feasible points, we deduce

$$\begin{aligned}\alpha_1(5-x_1)+\alpha_2&=0\\ \alpha_1(5-x_2)+\alpha_2&=0\\ \alpha_1(5-x_3)+\alpha_2&=0.\end{aligned}$$

Then, we find that $x_1 = x_2 = x_3$, which cannot be true on the set of feasible points. Therefore, the linear independence qualification condition holds on the set of feasible points. The application of Theorem 3.2.6 (and the remarks afterwards) leads to the conclusion that if $\overline{x}$ is a minimum or maximum point of the problem, then there exist $\mu_1, \mu_2 \in \mathbb{R}$ such that

$$\begin{aligned}&\overline{x}_2\overline{x}_3+\mu_1(\overline{x}_2+\overline{x}_3)+\mu_2=0\\ &\overline{x}_1\overline{x}_3+\mu_1(\overline{x}_1+\overline{x}_3)+\mu_2=0\\ &\overline{x}_1\overline{x}_2+\mu_1(\overline{x}_1+\overline{x}_2)+\mu_2=0\\ &\overline{x}_1\overline{x}_2+\overline{x}_1\overline{x}_3+\overline{x}_2\overline{x}_3=8\\ &\overline{x}_1+\overline{x}_2+\overline{x}_3=5.\end{aligned}$$

Obviously, μ_1, μ_2 cannot be simultaneously 0. After easy manipulations we get

$$\begin{aligned}&(\mu_1\overline{x}_3+\mu_2)(\overline{x}_1-\overline{x}_2)=0\\ &(\mu_1\overline{x}_2+\mu_2)(\overline{x}_1-\overline{x}_3)=0\\ &(\mu_1\overline{x}_1+\mu_2)(\overline{x}_2-\overline{x}_3)=0\\ &\overline{x}_1\overline{x}_2+\overline{x}_1\overline{x}_3+\overline{x}_2\overline{x}_3=8\\ &\overline{x}_1+\overline{x}_2+\overline{x}_3=5.\end{aligned}$$

If $\mu_1 = 0$, then $\overline{x}_1 = \overline{x}_2 = \overline{x}_3$, which is not possible. Then $\mu_1 \neq 0$ and since $\overline{x}_1, \overline{x}_2, \overline{x}_3$ cannot be equal, we find $\overline{x} = (2, 2, 1)$, $\overline{x} = (1, 2, 2)$, $\overline{x} = (2, 1, 2)$ and $\overline{x} = (\frac{4}{3}, \frac{4}{3}, \frac{7}{3})$, $\overline{x} = (\frac{7}{3}, \frac{4}{3}, \frac{4}{3})$, $\overline{x} = (\frac{4}{3}, \frac{7}{3}, \frac{4}{3})$. By easy comparison of function values we find that the first ones are maxima, while the last ones, minima. □

Exercise 7.58. *Find the global minima of* $f : \mathbb{R}^3 \to \mathbb{R}$, $f(x_1, x_2, x_3) = x_1^3 + x_2^3 + x_3^3$ *on the set of points satisfying* $x_1^2 + x_2^2 + x_3^2 \leq 4$ *and* $x_1 + x_2 + x_3 \leq 1$.

Solution Once again, the existence of minimum is ensured by Weierstrass Theorem. It is the same objective function as in Exercise 7.53, but this time the constraints are written as inequalities. Observe that the restrictions are convex and, moreover, the Slater condition holds. So, the minima are to be found among the critical points of the Lagrangian. We have

$$\begin{aligned}3x_1^2+2\lambda_1x_1+\lambda_2&=0\\ 3x_2^2+2\lambda_1x_2+\lambda_2&=0\\ 3x_3^2+2\lambda_1x_3+\lambda_2&=0\\ \lambda_1(x_1^2+x_2^2+x_3^2-4)&=0\\ \lambda_2(x_1+x_2+x_3-1)&=0\end{aligned}$$

$$x_1^2 + x_2^2 + x_3^2 \leq 4$$
$$x_1 + x_2 + x_3 \leq 1$$
$$\lambda_1, \lambda_2 \geq 0.$$

Again, we distinguish several situations.

- If $\lambda_1 = \lambda_2 = 0$, then $x_1 = x_2 = x_3 = 0$.
- If $\lambda_1 = 0$, $x_1 + x_2 + x_3 = 1$, we deduce

  $$\lambda_2 = -3x_1^2 = -3x_2^2 = -3x_3^2,$$

  so the unique solution $\overline{x}$ is $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ which gives $\lambda_2 = -\frac{1}{3} < 0$, and this is not convenient.
- If $\lambda_2 = 0$, $x_1^2 + x_2^2 + x_3^2 = 4$, we have

  $$12 + 2(x_1 + x_2 + x_3)\lambda_1 = 0,$$

  that is
  $$\lambda_1 = -\frac{6}{x_1 + x_2 + x_3}.$$
  Replacing in the first three equations, we obtain the solutions

  $$(-2, 0, 0), (0, -2, 0),\ (0, 0, -2)$$
  $$(-\sqrt{2}, -\sqrt{2}, 0),\ (-\sqrt{2}, 0, -\sqrt{2}),\ (0, -\sqrt{2}, -\sqrt{2})$$
  $$\left(-\frac{2}{\sqrt{3}}, -\frac{2}{\sqrt{3}}, -\frac{2}{\sqrt{3}}\right).$$
- If $x_1 + x_2 + x_3 = 1$, $x_1^2 + x_2^2 + x_3^2 = 4$, we are again in the situation from Exercise 7.53.

After the computation of the function values we get that the minimum is attained in $(-2, 0, 0)$, $(0, -2, 0)$, $(0, 0, -2)$, so, finally, the situation is different from that in Exercise 7.53. □

Problem 7.59. *Obtain the mean inequality from Karush-Kuhn-Tucker conditions applied for an appropriate optimization problem.*

Solution Let $f, g : \mathbb{R}^n \to \mathbb{R}$,

$$f(x) = \frac{x_1 + x_2 + \ldots + x_n}{n}$$
$$h(x) = 1 - x_1 x_2 \ldots x_n.$$

We study the problem of minimization of f under the restrictions

$$h(x) = 0,\ x_i > 0,\ \forall i \in \overline{1, n}.$$

Firstly, we observe that f has a minimum on the compact set $\{x \in \mathbb{R}^n \mid h(x) = 0, x_i \geq 0, \forall i \in \overline{1, n}\} \cap N_\nu f$ (where $\nu > 0$). Since the point with at least one zero component is not in this set, we deduce that the minimum belongs to the set of feasible points of the problem under consideration. On the other hand, the linear independence condition is satisfied at every feasible point. So, the Karush-Kuhn-Tucker condition means that for a solution x of the problem, there exists $\mu \in \mathbb{R}$ such that

$$0 = \frac{1}{n} - \frac{\mu}{x_i}, \ \forall i \in \overline{1, n}.$$

Therefore, $x_1 = x_2 = \ldots = x_n = 1$. So, $f(x) \geq f(1, \ldots, 1) = 1$ for every feasible point x.

Let $a_1, a_2, \ldots, a_n > 0$. Denote $G = \sqrt[n]{a_1 \ldots a_n}$ and $x_i = \frac{a_i}{G} > 0$, for every n. Then $h(x) = 0$, therefore

$$\frac{x_1 + x_2 + \ldots + x_n}{n} \geq 1,$$

that is

$$\frac{1}{n}\left(\frac{a_1}{G} + \ldots + \frac{a_n}{G}\right) \geq 1,$$

and the conclusion follows. □

Exercise 7.60. *Find the global extrema for* $f : \mathbb{R}^2 \to \mathbb{R}$, $f(x_1, x_2) = -2x_1^2 + 4x_1x_2 + x_2^2$ *on the unit circle.*

Solution The existence of solutions is ensured by the Weierstrass Theorem. We take $h : \mathbb{R}^2 \to \mathbb{R}$, $h(x_1, x_2) = 1 - x_1^2 - x_2^2$ and we have to study the problem of extrema of f under the constraint $h(x) = 0$. The linear independence qualification condition holds and for both minima and maxima we have to find the critical points of the Lagrangian. We have

$$\begin{aligned} (-2-\mu)x_1 + 2x_2 &= 0 \\ 2x_1 + (1-\mu)x_2 &= 0 \\ x_1^2 + x_2^2 &= 1. \end{aligned}$$

We infer that

$$\begin{vmatrix} -2-\mu & 2 \\ 2 & 1-\mu \end{vmatrix} = 0,$$

so $\mu = 2$ or $\mu = -3$ and, correspondingly, $(x_1, x_2) = \left(-\frac{2}{\sqrt{5}}, \frac{1}{\sqrt{5}}\right)$, $(x_1, x_2) = \left(\frac{2}{\sqrt{5}}, -\frac{1}{\sqrt{5}}\right)$ or $(x_1, x_2) = \left(\frac{1}{\sqrt{5}}, -\frac{2}{\sqrt{5}}\right), \left(-\frac{1}{\sqrt{5}}, \frac{2}{\sqrt{5}}\right)$. By direct computation of the function f values at these points, we find that the first two points are maxima, while the last two are minima. □

Exercise 7.61. *Show that* $f : \mathbb{R}^3 \to \mathbb{R}$, $f(x_1, x_2, x_3) = x_1^2 + x_2^2 + x_3^2 - x_1 - x_2 - x_3$ *is convex. Find its minimum value under the restrictions*

$$x_1^2 + x_2^2 = 4, \ -1 \leq x_3 \leq 1.$$

Solution The convexity of f follows from the fact that $\nabla^2 f(x)$ is positive definite for every $x \in \mathbb{R}^3$. Obviously, the set M of feasible points is not convex. So, a point $x \in M$ is a minimum if

$$-\nabla f(x) \in N(M, x) = \{u \in \mathbb{R}^3 \mid \langle u, c - x\rangle \le 0,\ \forall c \in M\},$$

i.e.,

$$\nabla f(x)(c - x) \ge 0,\ \forall c \in M.$$

The expression of the normal cone leads to the result, but this can be complicated. For instance, for a feasible point x with $x_3 \in (-1, 1)$,

$$N(M, x) = \mathbb{R}\{(x_1, x_2, 0)\},$$

and the condition becomes

$$-(2x_1 - 1, 2x_1 - 1, 2x_1 - 1) \in \mathbb{R}\{(x_1, x_2, 0)\},$$

and we get $x_3 = \frac{1}{2}$ and $x_1 = x_2 = \pm\sqrt{2}$. Therefore $\left(-\sqrt{2}, -\sqrt{2}, \dfrac{1}{2}\right)$ and $\left(\sqrt{2}, \sqrt{2}, \dfrac{1}{2}\right)$ are minima. The global minimum is attained at $\left(\sqrt{2}, \sqrt{2}, \dfrac{1}{2}\right)$. The remaining reasoning is similar.

A useful remark which can lead to the solution in a simple manner is as follows: the cost function on the feasible set is $f(x_1, x_2, x_3) = 4 - x_1 - x_2 + x_3^2 - x_3$, and the variables are now separated since it is sufficient to find minima of $x_3^2 - x_3$ on $[-1, 1]$ and the minima of $4 - x_1 - x_2$ under the restriction $x_1^2 + x_2^2 = 4$. These two problems are much more easy to handle than the initial one (the former being elementary). We obtain the minimum of the initial problem at $\left(\sqrt{2}, \sqrt{2}, \dfrac{1}{2}\right)$. □

Exercise 7.62. *Consider the region in* $\mathbb{R}^2$ *defined by*

$$x_1^2 - x_2^2 \le 1,\ x_1^2 + x_2^2 \le 4.$$

Find the extreme values on this region for

$$f : \mathbb{R}^2 \to \mathbb{R},\ f(x_1, x_2) = x_1^2 + 2x_2^2 + x_1x_2.$$

Solution The feasible set is compact. The linear independence condition means that $x_1x_2 \ne 0$. If $x_1x_2 = 0$, then we have two situations: for $x_1 = 0$, we get $x_2 \in [-2, 2]$, and for those points the maximum of f is 8, while the minimum is 0; for $x_2 = 0$, we obtain $x_1 \in [-1, 1]$, and the maximum of f is 1, while the minimum is 0.

After these cases, we can suppose that $x_1x_2 \ne 0$ and we treat separately the minimization and the maximization problems. For minimization we have the conditions

$$2x_1 + x_2 + 2\lambda_1 x_1 + 2\lambda_2 x_1 = 0$$

$$\begin{aligned} x_1 + 4x_2 - 2\lambda_1 x_2 + 2\lambda_2 x_2 &= 0 \\ \lambda_1(x_1^2 - x_2^2 - 1) &= 0 \\ \lambda_2(x_1^2 + x_2^2 - 4) &= 0 \\ \lambda_1 &\geq 0 \\ \lambda_2 &\geq 0. \end{aligned}$$

The rest of the calculation is not very involved, and at the end we compare the objective function values. For maximization, the reasoning is similar. □

Exercise 7.63. *Find the closest point to the origin on the surfaces:*
(i) $x_1x_2 + x_1x_3 + x_2x_3 = 1$;
(ii) $x_1^2 + x_2^2 - x_3^2 = 1$.

Solution In both cases, the objective function is $f : \mathbb{R}^3 \to \mathbb{R}$,

$$f(x_1, x_2, x_3) = x_1^2 + x_2^2 + x_3^2,$$

and on the set of feasible points the linear independence qualification condition holds.

For instance, at (i), for $v > 1$, the set $M \cap N_v f$ is nonempty (it contains, for example, the point $(1, 0, 0)$) and bounded. According to Theorem 3.1.7 there exists the global minimum of the proposed problem. For for a local minimum point, there exists $\mu \in \mathbb{R}$ such that

$$\begin{aligned} 2x_1 + \mu(x_2 + x_3) &= 0 \\ 2x_2 + \mu(x_1 + x_3) &= 0 \\ 2x_3 + \mu(x_1 + x_2) &= 0 \\ x_1x_2 + x_1x_3 + x_2x_3 &= 1. \end{aligned}$$

For the first three equations we cannot have $x_1 = x_2 = x_3 = 0$ (because of the fourth equation), so we infer that

$$\begin{vmatrix} 2 & \mu & \mu \\ \mu & 2 & \mu \\ \mu & \mu & 2 \end{vmatrix} = 0,$$

and calculations lead to

$$x = \left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right) \text{ and } x = \left(-\frac{1}{\sqrt{3}}, -\frac{1}{\sqrt{3}}, -\frac{1}{\sqrt{3}}\right),$$

which both are minima.

Let us observe that there is no maximum point, since for every $n \in \mathbb{N}^*$ the point $\left(n, n, \frac{1-n^2}{2n}\right)$ is feasible, but $f\left(n, n, \frac{1-n^2}{2n}\right) \overset{n\to\infty}{\to} \infty$. □

Exercise 7.64. *Let* $f : \mathbb{R}^2 \to \mathbb{R}$,

$$f(x) = -x_1 - 2x_2 - 2x_1x_2 + \frac{x_1^2}{2} + \frac{x_2^2}{2}$$

and the set of feasible points

$$M := \left\{x \in \mathbb{R}^2 \mid x_1 + x_2 \le 1,\ x_1 \ge 0,\ x_2 \ge 0\right\}.$$

Solve the problem (P) *of minimizing* f *on* M.

Solution Let us remark that for every $x \in \mathbb{R}^2$,

$$\nabla^2 f(x) = \begin{pmatrix} 1 & -2 \\ -2 & 1 \end{pmatrix}.$$

Since this matrix is not positive definite (the determinant is negative), the function is not convex. If it would exist a minimum point $\overline{x}$ lying in the interior of M, then that point would be a minimum point without restrictions (according to Remark 3.1.2), whence, from Fermat Theorem, $\nabla f(\overline{x}) = 0$. But

$$\nabla f(\overline{x}) = (-1 + \overline{x}_1 - 2\overline{x}_2, -2 - 2\overline{x}_1 + \overline{x}_2),$$

and the resulting system has the solution $\overline{x} = (-\frac{5}{3}, -\frac{4}{3})$ which actually does not belong to M. Hence, the problem has no solutions in $\operatorname{int} M$. However f is continuous, M is compact, so the problem (P) has at least one solution.

We can approach the problem in two ways.

The first one takes advantage of the fact that the geometrical image of the set M is a simple one (a triangle with the vertices at $(0,0)$, $(0,1)$, and $(1,0)$). It is not difficult to compute the Bouligand tangent and normal cones to M at its boundary points and then to verify the necessary optimality condition: $-\nabla f(\overline{x}) \in N(M, \overline{x})$ (see Theorem 3.1.18).

Therefore, if the point $\overline{x}$ is:

– on the open segment joining the vertices $(0,1)$, $(1,0)$:

$$T(M,\overline{x}) = \{u \in \mathbb{R}^2 \mid u_1 + u_2 \le 0\};\ N(M,\overline{x}) = \mathbb{R}_+\{(1,1)\};$$

– on the open segment joining the vertices $(0,0)$, $(0,1)$:

$$T(M,\overline{x}) = \{u \in \mathbb{R}^2 \mid u_1 \ge 0\};\ N(M,\overline{x}) = \mathbb{R}_+\{(-1,0)\};$$

– on the open segment joining the vertices $(0,0)$, $(1,0)$:

$$T(M,\overline{x}) = \{u \in \mathbb{R}^2 \mid u_2 \ge 0\};\ N(M,\overline{x}) = \mathbb{R}_+\{(0,-1)\};$$

– exactly $(0,1)$:

$$T(M,\overline{x}) = \{u \in \mathbb{R}^2 \mid u_1 + u_2 \le 0,\ u_1 \ge 0\}$$

$$N(M,\overline{x}) = \left\{a(1,1) + b(-1,0) \mid a,b \geq 0\right\};$$

– exactly $(1,0)$:

$$T(M,\overline{x}) = \{u \in \mathbb{R}^2 \mid u_1 + u_2 \leq 0,\ u_2 \geq 0\}$$
$$N(M,\overline{x}) = \left\{a(1,1) + b(0,-1) \mid a,b \geq 0\right\};$$

– exactly $(0,0)$:

$$T(M,\overline{x}) = \{u \in \mathbb{R}^2 \mid u_1 \geq 0,\ u_2 \geq 0\};$$
$$N(M,\overline{x}) = \left\{a(-1,0) + b(0,-1) \mid a,b \geq 0\right\}.$$

A direct computation shows that there is only one point which satisfies the necessary optimality condition, and this is $\overline{x} = \left(\frac{1}{3}, \frac{2}{3}\right)$. Therefore, according to the preceding remark, this is the only minimum point of the problem.

One can also consider the problem of finding the global maximum of f on M (the existence of such a maximum is ensured by the Weierstrass' Theorem), which is equivalent to the finding the global minimum of $-f$ on M. With the same arguments as above, we find two points which verify the necessary optimality condition (i.e., $\nabla f(\overline{x}) \in N(M,\overline{x})$): $\overline{x} = (0,0)$ and $\overline{x} = (1,0)$. But $f(0,0) = 0$, while $f(1,0) = -2^{-1}$, so $(0,0)$ is the global maximum point.

The second possible solution is to consider the function

$$g : \mathbb{R}^2 \to \mathbb{R}^3,\ g(x) := (x_1 + x_2 - 1, -x_1, -x_2)$$

and to reinterpret the problem as a problem with three functional inequalities given by $g(x) \leq 0$. Since all the functions g_i are linear, it is not necessary to verify qualification conditions (according to Theorem 3.2.24). Then, if $\overline{x} \notin \operatorname{int} M$ is a solution of the problem, there exists $(\lambda_1, \lambda_2, \lambda_3) \in \mathbb{R}^3_+$ such that

$$\begin{cases} \nabla f(\overline{x}) + \lambda_1 \nabla g_1(\overline{x}) + \lambda_2 \nabla g_2(\overline{x}) + \lambda_3 \nabla g_3(\overline{x}) = 0 \\ \lambda_i g_i(\overline{x}) = 0,\ i \in \overline{1,3}. \end{cases}$$

Now, again, the discussion should be divided into six cases which mirror those from before. For instance, if $\overline{x}$ is on the open segment joining $(0,1)$ and $(1,0)$, then $g_2(\overline{x}) < 0$, $g_3(\overline{x}) < 0$, whence $\lambda_2 = \lambda_3 = 0$, and the above system reduces to

$$\begin{cases} -1 - 2\overline{x}_2 + \overline{x}_1 + \lambda_1 = 0 \\ -2 - 2\overline{x}_1 + \overline{x}_2 + \lambda_1 = 0 \\ x_1 + x_2 - 1 = 0, \end{cases}$$

which gives the solution $\lambda_1 = 2$, $\overline{x} = \left(\frac{1}{3}, \frac{2}{3}\right)$. In the same way, in the other situations, the Karush-Kuhn-Tucker system has no solution and the conclusion is the same as in the first approach.. □

Problem 7.65. *Let $f : \mathbb{R}^p \to \mathbb{R}$ be a convex and differentiable function. Write the optimality conditions for the minimization of f on the unit simplex.*

Solution According to Example 2.1.16 and Proposition 3.1.25, $\overline{x} \in M$ is a minimum point of f on M if and only if $-\nabla f(\overline{x}) \in N(M, \overline{x})$. From the particular form of $N(M, \overline{x})$, one infers that this condition become

$$\frac{\partial f}{\partial x_i}(\overline{x}) = c, \text{ (constant)}, \forall i \notin I(\overline{x})$$
$$\frac{\partial f}{\partial x_i}(\overline{x}) \geq c, \ \forall i \in I(\overline{x}). \qquad \square$$

Exercise 7.66. *Let us consider the objective function $f : \mathbb{R}^2 \to \mathbb{R}$, $f(x_1, x_2) = x_1 + x_2^2$ and a function which defines a equality constraint as $h : \mathbb{R}^2 \to \mathbb{R}$, $h(x_1, x_2) = x_1^3 - x_2^2$. Find the solution of the minimization of f under $h(x) = 0$.*

Solution We have $M = \{x \in \mathbb{R}^2 \mid h(x) = 0\}$. In order to verify the linear independence qualification condition at a point $\overline{x} \in M$, it is necessary and sufficient to have $\nabla h(\overline{x}) \neq 0$, and this happens for all points in M, but $\overline{x} = (0, 0)$. For the moment, we avoid this point. If $\overline{x} \in M \setminus \{(0, 0)\}$ is a minimum point of (P), then, according to Theorem 3.2.6, there exists $\mu \in \mathbb{R}$ such that

$$\nabla f(\overline{x}) + \mu \nabla h(\overline{x}) = 0.$$

A simple calculation shows that the resulting system has no solution, whence (P) has no solution different from $(0, 0)$. Let us remark that $\overline{x} = (0, 0)$ is a solution (even a global one) since $f(\overline{x}) = 0$, and for every $x \in M$, $x_1^3 = x_2^2 \geq 0$, so $f(x) = x_1 + x_2^2 \geq 0$. $\square$

Exercise 7.67. *Let $n_1, \ldots, n_p \in \mathbb{N}^*$, and let $f : \mathbb{R}^p \to \mathbb{R}$, $f(x) = -x_1^{n_1} x_2^{n_2} \ldots x_p^{n_p}$. Minimize this function on the unit simplex of $\mathbb{R}^p$.*

Solution Clearly, the problem has a solution, since f is continuos and M is compact. Since f vanishes if at least one of the components of the argument is zero, it is clear that the solutions will actually be from

$$\left\{ x \in \mathbb{R}^p \mid x_i > 0, \ \forall i \in \overline{1, p}, \ \sum_{i=1}^{p} x_i = 1 \right\}.$$

With the notation from Example 2.1.16, this means that $I(\overline{x}) = \emptyset$. Firstly, the necessary optimality condition, $-\nabla f(\overline{x}) \in N(M, \overline{x})$, could be written, with the expression of the normal cone to the unit simplex in mind (Example 2.1.16), as

$$\frac{n_i}{\overline{x}_i} f(\overline{x}) = c, \text{ (constant) } \forall i \in \overline{1, p},$$

that is

$$\frac{n_i}{\overline{x}_i} = c', \text{ (constant) } \forall i \in \overline{1, p}.$$

Because $\sum_{i=1}^{p} \overline{x}_i = 1$, by denoting $N := \sum_{i=1}^{p} n_i$, we find

$$\overline{x}_i = \frac{n_i}{N}, \ \forall i \in \overline{1,p}.$$

Since the problem admits a solution and only one point satisfies the necessary condition, we deduce that this point is the solution we are looking for.

Another approach consists of transforming the geometrical restriction into a functional one. Let $h : \mathbb{R}^p \to \mathbb{R}$, $h(x) = \sum_{i=1}^{p} x_i - 1$. Clearly, $M = \{x \in \mathbb{R}^p \mid h(x) = 0\}$. Let $\overline{x}$ be a solution of the problem. As $\nabla h(\overline{x}) \neq 0$, we can apply Theorem 3.2.6 in order to deduce the existence of a number $\mu \in \mathbb{R}$ such that

$$\nabla f(\overline{x}) + \mu \nabla h(\overline{x}) = 0,$$

that is

$$-\frac{n_i}{\overline{x}_i} f(\overline{x}) = \mu, \ \text{(constant)} \ \forall i \in \overline{1,p}.$$

So, as expected, the same conclusion follows. □

Exercise 7.68. *Let $a \in \mathbb{R}$, $f, h : \mathbb{R}^2 \to \mathbb{R}$, $f(x) = (x_1 - 1)^2 + x_2^2$, $h(x) = -x_1 + ax_2^2$. Decide if $\overline{x} = (0,0)$ is a minimum point for f under the restriction $h(x) = 0$.*

Solution We have $\nabla h(\overline{x}) = (-1, 2a\overline{x}_2) \neq (0,0)$, so the linear independence qualification condition is fulfilled. We study the necessary optimality condition from Theorem 3.2.6 for $\overline{x} = (0,0)$. There exists $\mu \in \mathbb{R}$ with

$$\nabla f(\overline{x}) + \mu \nabla h(\overline{x}) = 0,$$

and we obtain $\mu = -2$. We discuss now the second-order necessary optimality condition (Theorem 3.3.2). It is easy to see that $T_B(M, \overline{x}) = \{0\} \times \mathbb{R}$, and for $u = (0, u_2) \in T_B(M, \overline{x})$,

$$\nabla^2_{xx} L(\overline{x}, \mu)(u,u) = 2(1 - 2a)u_2^2.$$

If $a > \frac{1}{2}$, the condition from Theorem 3.3.2 is not satisfied, so the point is not a minimum.

If $a < \frac{1}{2}$, the sufficient condition from Theorem 3.3.3 is satisfied, so the point is even a strict solution of second order.

For $a = \frac{1}{2}$, only necessary optimality condition is satisfied. In this case we observe that for $x \in M$, $x_1 = 2^{-1}x_2^2$, whence

$$f(x_1, x_2) = (2^{-1}x_2^2 - 1)^2 + x_2^2 = 4^{-1}x_2^4 + 1 \geq 1 = f(0,0),$$

so $(0,0)$ is a minimum point. □

Exercise 7.69. *Solve the problem of minimization of $f : \mathbb{R}^3 \to \mathbb{R}$, $f(x) = -x_1 - x_2 - x_2x_3 - x_1x_3$ under the affine constraint $h(x) = 0$, where $h : \mathbb{R}^3 \to \mathbb{R}$, $h(x) = x_1 + x_2 + x_3 - 3$.*

Solution Let $\overline{x} \in \mathbb{R}^3$ be a feasible point, i.e., $h(\overline{x}) = 0$. We verify if the necessary optimality condition in Theorem 3.2.6 holds. There should exist some $\mu \in \mathbb{R}$ with

$$\nabla f(\overline{x}) + \mu \nabla h(\overline{x}) = 0.$$

We get the linear system

$$\begin{cases} \overline{x}_1 + \overline{x}_2 + \overline{x}_3 = 3 \\ -1 - \overline{x}_3 = -\mu \\ -\overline{x}_1 - \overline{x}_2 = -\mu \end{cases}$$

which gives $\mu = 2$, $\overline{x} = (\overline{x}_1, 2 - \overline{x}_1, 1)$, $\overline{x}_1 \in \mathbb{R}$.

Let us observe that these points also verify the second-order necessary optimality conditions (see Theorem 3.3.2 and Remark 3.3.1). It is clear (as in the case of unit simplex) that $T(M, \overline{x}) = \{u \in \mathbb{R}^3 \mid u_1 + u_2 + u_3 = 0\}$. Then for every $u \in T(M, \overline{x})$,

$$\begin{aligned} \nabla^2_{xx} L(\overline{x}, \mu)(u, u) &= (u_1, u_2, u_3) \begin{pmatrix} 0 & 0 & -1 \\ 0 & 0 & -1 \\ -1 & -1 & 0 \end{pmatrix} (u_1, u_2, u_3)^t \\ &= -2(u_1 + u_2)u_3 = 2(u_1 + u_2)^2 \geq 0. \end{aligned}$$

One can observe that for $u = (u_1, -u_1, 0)$ with $u_1 \neq 0$, the sufficient second-order optimality condition in Theorem 3.3.3 is not satisfied, so we should decide if the above points are solutions using different tools. More precisely, will try to exploit the particular form of the problem. We observe that for every $\overline{x}_1 \in \mathbb{R}$, $f(\overline{x}_1, 2 - \overline{x}_1, 1) = -4$, and for an arbitrary $x \in M$, $f(x) + 4 = (x_1 + x_2 - 2)^2 \geq 0$, so all the points which satisfy the necessary optimality conditions are solutions for our problem. □

Exercise 7.70. *Let* $a > 4^{-1}$, $f, g : \mathbb{R}^2 \to \mathbb{R}$, $f(x) = x_1^2 + ax_2^2 + x_1x_2 + x_1$ *and* $g(x) = x_1 + x_2 - 1$. *Solve the problem* (P), *with the usual notations.*

Solution Let us observe that

$$\nabla^2 f(x) = \begin{pmatrix} 2 & 1 \\ 1 & 2a \end{pmatrix}$$

is positive definite, whence f is convex (Theorem 2.2.10). Since the remaining problem data is convex (affine, in fact), according to Theorems 3.2.6 and 3.2.8, a point $\overline{x}$ is solution of (P) if and only is there exists $\lambda \geq 0$ such that

$$\begin{cases} \nabla f(\overline{x}) + \lambda \nabla g(\overline{x}) = 0 \\ \lambda g(\overline{x}) = 0. \end{cases}$$

This system can be written as $\overline{x} \in \operatorname{int} M$ (i.e., $g(\overline{x}) < 0$) or $\overline{x} \in \operatorname{bd} M$ (i.e., $g(\overline{x}) = 0$) in one of the following forms:

$$\begin{cases} \overline{x}_1 + \overline{x}_2 - 1 < 0 \\ 2\overline{x}_1 + \overline{x}_2 + 1 = 0 \\ \overline{x}_1 + 2a\overline{x}_2 = 0 \end{cases}$$

and

$$\begin{cases} \overline{x}_1 + \overline{x}_2 - 1 = 0 \\ 2\overline{x}_1 + \overline{x}_2 + 1 + \lambda = 0 \\ \overline{x}_1 + 2a\overline{x}_2 + \lambda = 0 \\ \lambda \geq 0, \end{cases}$$

respectively. The former one admits solution for $a > 3^{-1}$, and this is $\overline{x} = \left(-\frac{2a}{4a-1}, \frac{1}{4a-1}\right)$. The latter one has solution for $a \in (4^{-1}, 3^{-1}]$, and this is $\overline{x} = \left(1 - \frac{1}{a}, \frac{1}{a}\right)$, $\lambda = \frac{1-3a}{a}$. □

Exercise 7.71. *Let $f, g : \mathbb{R}^2 \to \mathbb{R}$, $f(x) = x_1^3 + x_2^2$ and $g(x) = x_1^2 + x_2^2 - 9$. Study the minima of f under the restriction $g(x) \leq 0$.*

Solution We remark from the beginning that the set of feasible points is compact. Since f is continuous, the problem admits a solution. Let us remark that g is convex, since its Hessian matrix is positive definite at every point. But $g(0, 0) < 0$ and the Slater condition holds, so a solution $\overline{x}$ of (P) verifies Karush-Kuhn-Tucker conditions that is, there exists $\lambda \geq 0$ such that

$$\begin{cases} \nabla f(\overline{x}) + \lambda \nabla g(\overline{x}) = 0 \\ \lambda g(\overline{x}) = 0. \end{cases}$$

As before, we have two distinct situations and we get the solutions

$$\overline{x} = (0, 0), \lambda = 0 \text{ and } \overline{x} = (-3, 0), \lambda = \frac{9}{2}.$$

But $(0, 0)$ is not a solution, because it is enough to consider the sequences $(x_n) = (\frac{1}{n}, 0) \to (0, 0)$ and $(y_n) = (-\frac{1}{n}, 0) \to (0, 0)$ for which

$$f(y_n) < f(0, 0) < f(x_n), \ \forall n \in \mathbb{N}^*.$$

The conclusion is that $\overline{x} = (-3, 0)$ is the only solution of the problem. □

Exercise 7.72. *Let A be a symmetric square matrix of dimension p. We consider the application $f : \mathbb{R}^p \to \mathbb{R}$, $f(x) = \left\langle (Ax^t)^t, x \right\rangle$. Solve the problems of minimization and maximization of this function on the unit sphere of $\mathbb{R}^p$.*

Solution Clearly, in both cases there is a solution. We observe that, following Remark 3.2.21, at a point x of the sphere, the Bouligand tangent cone to the sphere is $\{u \in \mathbb{R}^p \mid \langle x, u \rangle = 0\}$, whence the corresponding normal cone is $\{\alpha x \mid x \in \mathbb{R}\}$. Then, from Theorem 3.1.18, a point $\overline{x}$ on the unit sphere is a solution for minimization or maximization problems if there exists $\mu \in \mathbb{R}$ with $(A\overline{x}^t)^t = \mu \overline{x}$, that is $\mu = \left\langle (A\overline{x}^t)^t, \overline{x} \right\rangle$. We deduce that the points which satisfy the optimality conditions correspond to some eigenvectors of A, while the values of f in those points are eigenvalues for f. We conclude that the greater eigenvalue of A is $\lambda_1 := \max_{\|x\|=1} \left\langle (Ax^t)^t, x \right\rangle$, while the smallest is $\lambda_p := \min_{\|x\|=1} \left\langle (Ax^t)^t, x \right\rangle$. Therefore

$$\lambda_1 = \max_{x \in \mathbb{R}^p \setminus \{0\}} \frac{\left\langle (Ax^t)^t, x \right\rangle}{\|x\|^2} \text{ and } \lambda_p = \min_{x \in \mathbb{R}^p \setminus \{0\}} \frac{\left\langle (Ax^t)^t, x \right\rangle}{\|x\|^2}.$$ □

Exercise 7.73. *Let us consider the function $f : \mathbb{R}^p \to \mathbb{R}$ given by $f(x) = \frac{1}{2}\langle (Qx^t)^t, x\rangle + \langle c, x\rangle$, where Q is a symmetric positive definite square matrix of dimension p. We consider as well the restriction $Ax^t = b^t$ where A is a matrix of dimensions $q \times p$ of rank q, where $1 \le q \le p$ and $b \in \mathbb{R}^q$. Solve the problem of the minimization of f subject to the given restriction.*

Solution First of all, we see that f is strictly convex ($\nabla^2 f(x) = Q$ for every $x \in \mathbb{R}^p$), $\lim_{\|x\|\to\infty} f(x) = +\infty$ and the restriction is affine. Therefore, there is a unique minimum point $\overline{x}$ fully characterized by the relations

$$A\overline{x}^t = b^t$$
$$\exists \mu \in \mathbb{R}^q \text{ such that } \nabla f(\overline{x}) + \mu A = 0.$$

So

$$\begin{cases} A\overline{x}^t = b^t \\ (Q\overline{x}^t)^t + c + \mu A = 0, \end{cases}$$

and

$$\begin{cases} A\overline{x}^t = b^t \\ Q\overline{x}^t + c^t + A^t\mu^t = 0. \end{cases}$$

Since Q is invertible, we get

$$\begin{cases} A\overline{x}^t = b^t \\ \overline{x}^t + Q^{-1}c^t + Q^{-1}A^t\mu^t = 0 \end{cases}$$

and, by multiplication by A in the second relation,

$$\begin{cases} A\overline{x}^t = b^t \\ b + AQ^{-1}c^t + AQ^{-1}A^t\mu^t = 0. \end{cases}$$

But $AQ^{-1}A^t$ is a symmetric square matrix of dimension q and for every $y \in \mathbb{R}^q$,

$$\left\langle (AQ^{-1}A^ty^t)^t, y\right\rangle = \left\langle (Q^{-1}A^ty^t)^t, (A^ty^t)^t\right\rangle \ge 0.$$

The equality in the above relation holds if and only if $A^ty^t = 0$. Since the linear function associated with A is surjective (from the rank condition), we infer that the linear function associated to A^t is injective, so $A^ty^t = 0$ if and only if $y = 0$. Therefore $AQ^{-1}A^t$ is positive definite, whence invertible. We obtain

$$\mu^t = -(AQ^{-1}A^t)^{-1}(b + AQ^{-1}c^t)$$

and

$$\overline{x}^t = -Q^{-1}c^t + Q^{-1}A^t(AQ^{-1}A^t)^{-1}(b + AQ^{-1}c^t). \qquad \square$$

Exercise 7.74. *Solve the problem to minimize and to maximize the objective function $f : \mathbb{R}^3 \to \mathbb{R}$, $f(x) = x_1x_2x_3$ under the constraint $h(x) = 0$, where $h : \mathbb{R}^3 \to \mathbb{R}$, $h(x) = x_1x_2 + x_1x_3 + x_2x_3 - 8$.*

Solution We remark that these problems do not admit global solutions because, for instance, the points

$$\left(n, n, \frac{8-n^2}{2n}\right), \left(-n, -n, -\frac{8-n^2}{2n}\right), \; n \in \mathbb{N}^*$$

are feasible, but

$$\lim_n f\left(n, n, \frac{8-n^2}{2n}\right) = \lim_n n\frac{8-n^2}{2} \to -\infty$$
$$\lim_n f\left(-n, -n, -\frac{8-n^2}{2n}\right) = \lim_n n\frac{-8+n^2}{2} \to +\infty,$$

whence f is neither lower, nor upper bounded on the feasible set. So, we are looking for local solutions.

It is easy to see that the linear independence qualification condition holds in every feasible point. The application of Theorem 3.2.6 leads to the conclusion that if $\overline{x}$ is solution of one of the problems, then there exists $\mu \in \mathbb{R}$ such that

$$\begin{cases} \overline{x}_2\overline{x}_3 + \mu(\overline{x}_2 + \overline{x}_3) = 0 \\ \overline{x}_1\overline{x}_3 + \mu(\overline{x}_1 + \overline{x}_3) = 0 \\ \overline{x}_1\overline{x}_2 + \mu(\overline{x}_1 + \overline{x}_2) = 0 \\ \overline{x}_1\overline{x}_2 + \overline{x}_1\overline{x}_3 + \overline{x}_2\overline{x}_3 = 8. \end{cases}$$

Solving this system (by appropriate multiplication and subtraction of the equations), we get $\overline{x} = \left(-\frac{2\sqrt{6}}{3}, -\frac{2\sqrt{6}}{3}, -\frac{2\sqrt{6}}{3}\right)$, $\mu = \frac{\sqrt{6}}{2}$ and $\overline{x} = \left(\frac{2\sqrt{6}}{3}, \frac{2\sqrt{6}}{3}, \frac{2\sqrt{6}}{3}\right)$, $\mu = -\frac{\sqrt{6}}{2}$. We need to verify second-order optimality conditions. The set of critical directions in both cases is $\{u \in \mathbb{R}^3 \mid u_1 + u_2 + u_3 = 0\}$. For the first point, $\nabla^2_{xx}L(\overline{x}, \mu)(u, u) = \frac{-\sqrt{6}}{3}(u_1u_2 + u_1u_3 + u_2u_3)$. Since

$$0 = (u_1 + u_2 + u_3)^2 = u_1^2 + u_2^2 + u_3^2 + 2(u_1u_2 + u_1u_3 + u_2u_3)$$

we deduce that $u_1u_2 + u_1u_3 + u_2u_3 < 0$ for any nonzero critical direction, i.e., $\nabla^2_{xx}L(\overline{x}, \mu)(u, u) > 0$ for such a direction, whence the sufficient optimality condition in Theorem 3.3.3 do hold, and therefore the reference point is a local (strict) minimum. The same holds for the second point and the corresponding sufficient condition for maximization. Hence the second point is the solution of the maximization problem. □

Exercise 7.75. *The Karush-Kuhn-Tucker conditions tell us that a solution of (P) is a critical point with respect to x for the Lagrangian function.*

(i) Show that a solution of (P) is not necessarily a minimum point of $x \mapsto L(x, (\lambda, \mu))$. For this, consider $f : \mathbb{R}^2 \to \mathbb{R}$, $f(x_1, x_2) = x_1^2 - x_2^2 - 3x_2$ and $h : \mathbb{R}^2 \to \mathbb{R}$, $h(x_1, x_2) = x_2$.

(ii) Show that there exist as well situations when a solution of (P) is a minimum point of $x \mapsto L(x, (\lambda, \mu))$. For this, consider $f : \mathbb{R}^2 \to \mathbb{R}$, $f(x_1, x_2) = 5x_1^2 + 4x_1x_2 + x_2^2$ and $h : \mathbb{R}^2 \to \mathbb{R}$, $h(x_1, x_2) = 3x_1 + 2x_2 + 5$.

Solution (i) It is clear that $(0, 0)$ is a minimum point of f under linear restriction $h(x_1, x_2) = 0$. Implementing Karush-Kuhn-Tucker conditions, we get the multiplier $\mu = 3$, but $(0, 0)$ is not a minimum point for $f(x_1, x_2) + 3h(x_1, x_2) = x_1^2 - x_2^2$.

(ii) Clearly, $f(x_1, x_2) = (2x_1 + x_2)^2 + x_1^2$ satisfies the conditions of Proposition 3.1.8, whence the problem has solutions. Since the restriction $h(x) = 0$ is affine, we can apply Theorem 3.2.6 and we get the linear system

$$\begin{cases} 10x_1 + 4x_2 + 3\mu = 0 \\ 4x_1 + 2x_2 + 2\mu = 0 \\ 3x_1 + 2x_2 + 5 = 0, \end{cases}$$

which has the solution $(x_1, x_2, \mu) = (1, -4, 2)$. Then Lagrangian function in (x_1, x_2) is $l : \mathbb{R}^2 \to \mathbb{R}$,

$$\begin{aligned} l(x_1, x_2) &= 5x_1^2 + 4x_1x_2 + x_2^2 + 6x_1 + 4x_2 + 10 \\ &= (2x_1 + x_2 + 2)^2 + (x_1 - 1)^2 + 5, \end{aligned}$$

and $(1, -4)$ is a global minimum without restrictions. □

Exercise 7.76. *Find the local minima of the function* $f(x_1, x_2) = x_1^2 + (x_2 - 1)^2$ *under the constraints* $x_2 \le x_1^2$ *and* $x_2 \le x_1$.

Solution Consider the functions $g_1(x_1, x_2) = x_2 - x_1^2$, $g_2(x_1, x_2) = x_2 - x_1$. The the constraints set is

$$M = \left\{(x_1, x_2) \in \mathbb{R}^2 \mid g_i(x_1, x_2) \le 0,\ i = 1, 2\right\}.$$

By considering the gradients of the functions g_i, equal to $(-2x, 1)$, $(-1, 1)$, respectively, observe that the linear independence condition is not satisfied (for $x = \frac{1}{2}$).
To find the minima, we write then the Fritz John necessary conditions (i.e., Theorem 3.2.1): we must find the scalars $\lambda_0, \lambda_1, \lambda_2 \ge 0$, not all 0, such that

$$\begin{cases} \lambda_0 \nabla f(x_1, x_2) + \lambda_1 \nabla g_1(x_1, x_2) + \lambda_2 \nabla g_2(x_1, x_2) = 0 \\ \lambda_1 \cdot g_1(x_1, x_2) = 0,\ \lambda_2 \cdot g_2(x_1, x_2) = 0, \\ g_1(x_1, x_2) \le 0, g_2(x_1, x_2) \le 0 \end{cases}$$

hence we arrive at

$$\begin{cases} 2\lambda_0 x_1 - 2\lambda_1 x_1 + \lambda_2 = 0 \\ 2\lambda_0(x_2 - 1) + \lambda_1 + \lambda_2 = 0 \\ \lambda_1(x_2 - x_1^2) = 0 \\ \lambda_2(x_2 - x_1) = 0 \\ x_2 - x_1^2 \le 0, \lambda_1 \ge 0 \\ x_2 - x_1 \le 0, \lambda_2 \ge 0. \end{cases} \tag{7.3.1}$$

Suppose $\lambda_0 = 0$. Then we get from the second relation $\lambda_1 + \lambda_2 = 0$, and since $\lambda_1 \ge 0, \lambda_2 \ge 0$, it follows $\lambda_1 = \lambda_2 = 0$. We obtain the contradiction that not all scalars are 0, hence $\lambda_0 > 0$, and we may suppose without loss of generality $\lambda_0 = 1$.

Consider then the Lagrangian

$$\begin{aligned} L(x, x_2; \mu_1, \mu_2) &= f(x_1, x_2) + \lambda_1 g_1(x_1, x_2) + \lambda_2 g_2(x_1, x_2) \\ &= x_1^2 + (x_2 - 1)^2 + \lambda_1(x_2 - x_1^2) + \lambda_2(x_2 - x_1). \end{aligned}$$

The Karush-Kuhn-Tucker necessary conditions (see Theorem 3.2.6) will be then

$$\begin{cases} 2x_1 - 2\lambda_1 x_1 + \lambda_2 = 0 \\ 2(x_2 - 1) + \lambda_1 + \lambda_2 = 0 \\ \lambda_1(x_2 - x_1^2) = 0 \\ \lambda_2(x_2 - x_1) = 0 \\ x_2 - x_1^2 \le 0, \lambda_1 \ge 0 \\ x_2 - x_1 \le 0, \lambda_2 \ge 0. \end{cases} \tag{7.3.2}$$

For $\lambda_1 = \lambda_2 = 0$, we get from the first two relations the solution $(0, 1)$, which does not satisfy $x_2 - x_1 \le 0$.

For $\lambda_1, \lambda_2 > 0$, we have $x_2 - x_1^2 = 0$ and $x_2 - x_1 = 0$, with solutions $(0, 0)$ and $(1, 1)$. For $(0, 0)$, we obtain from the first relation $\lambda_2 = 0$, and then $\lambda_1 = 2$. For $(1, 1)$, we get from the second relation $\lambda_1 + \lambda_2 = 0$, which is impossible because $\lambda_1, \lambda_2 > 0$.

For $\lambda_1 = 0, \lambda_2 > 0$, we must solve the system

$$\begin{aligned} 2x_1 + \lambda_2 &= 0 \\ 2(x_2 - 1) + \lambda_2 &= 0 \\ x_2 - x_1 &= 0. \end{aligned}$$

Subtracting the first two relations and adding the third, one gets the contradiction $-1 = 0$.

For $\lambda_1 > 0, \lambda_2 = 0$, we must solve the system

$$\begin{aligned} x - \lambda_1 x_1 &= 0 \\ 2(x_2 - 1) + \lambda_1 &= 0 \\ x_2 - x_1^2 &= 0. \end{aligned}$$

From the first relation, if $x_1 = 0$, one gets $x_2 = 0$ and $\lambda_1 = 2$, that is, one obtains the above solution. Suppose $x_1 \neq 0$. Then $\lambda_1 = 1$, and $x_2 = \frac{1}{2}$, which gives $x_1 = \pm\frac{1}{\sqrt{2}}$. From the obtained solutions, the only feasible one is $\left(\frac{1}{\sqrt{2}}, \frac{1}{2}\right)$.

For concluding, we have got the following Karush-Kuhn-Tucker points and associated multipliers:

$$\begin{aligned} a &= (0, 0), \quad \lambda_a = (2, 0) \\ b &= \left(\frac{1}{\sqrt{2}}, \frac{1}{2}\right), \quad \lambda_b = (1, 0). \end{aligned}$$

We calculate $\nabla^2_{xx}L(x_1, x_2, \lambda_1, \lambda_2)$. Because

$$\frac{\partial^2 L}{\partial x_1^2}(x_1, x_2, \lambda_1, \lambda_2) = 2 - 2\lambda_1, \quad \frac{\partial^2 L}{\partial x_2^2}(x_1, x_2, \lambda_1, \lambda_2) = 2, \quad \frac{\partial^2 L}{\partial x_1 \partial x_2}(x_1, x_2, \lambda_1, \lambda_2) = 0,$$

one gets

$$\nabla^2_{xx}L(x_1, x_2, \lambda_1, \lambda_2)(h_1, h_2) = (2 - 2\lambda_1)h_1^2 + 2h_2^2.$$

Using Theorem 3.3.3, we must verify that

$$\nabla^2_{xx}L(a, \lambda_a)(h_1, h_2) = -2h_1^2 + 2h_2^2$$

is positive definite on the linear subspace which is orthogonal to the gradients of those restrictions which are active at a, for which the associated multiplier is strictly positive.

At a, both restrictions are active, but just $\lambda_1 > 0$ (i.e., $A(a) = \{1\}$). We have $\nabla g_1(a) = (0, 1)$, so we will consider the orthogonal subspace to this vector, that is

$$\begin{aligned} \left\{(u, v) \in \mathbb{R}^2 \mid \left\langle \nabla g_1(a), (u, v)\right\rangle = 0\right\} &= \left\{(u, v) \mid v = 0\right\} \\ &= \left\{(u, 0) \mid u \in \mathbb{R}\right\}. \end{aligned}$$

Since

$$\nabla^2_{xx}L(a, \lambda_a)(u, 0) = -2u^2$$

is not positive definite, we cannot apply Theorem 3.3.3 to decide if a is a local minimum.

Observe, though, that in a both restrictions are active, and the gradients $(0, 1)$, $(-1, 1)$ are linearly independent. We can apply then Theorem 3.3.2 for $A(a) = \{1\}$, and observe that the necessary optimality condition is not satisfied, hence a is not a local minimum.

For b, we have

$$\nabla^2_{xx}L(b, \lambda_b)(h_1, h_2) = 2h_2^2,$$

and again the first restriction is active. We consider the orthogonal subspace on $\nabla g_1(b) = \left(-\sqrt{2}, 1\right)$, that is

$$\left\{(u, v) \in \mathbb{R}^2 \mid -\sqrt{2}u + v = 0\right\} = \left\{(u, \sqrt{2}u) \mid u \in \mathbb{R}\right\}.$$

$\nabla^2_{xx}L(b, \lambda_b)$ is positive definite on this subspace, hence b is a local minimum.

We end by mentioning that in $(0, 0)$ the restriction g_2 is active, and the associated multiplier is 0. □

Problem 7.77. *Let $u \in \mathbb{R}^p \setminus \{0_{\mathbb{R}^p}\}$ and $a \in \mathbb{R}$. We consider the set (called hyperplane)*

$$M := \{x \in \mathbb{R}^p \mid \langle u, x\rangle = a\}$$

and a point $v \in \mathbb{R}^p \setminus M$. Show that M is a convex, closed set, and determine the explicit expression of the projection of v on M, as well as the value of the distance from v to M.

Solution The convexity and the closedness of M are obvious. So, from Theorem 2.1.5, there exists a unique projection of v on M, which we denote here by v_a.

Then v_a is the unique solution (see Theorem 2.1.5) of the problem of minimizing the function

$$f(x) = \frac{1}{2}\|x - v\|^2,$$

for $x \in M$. The choice of the objective function above has the same motivation as in the case of the least squares method. Since the constraint (in functional interpretation) is affine and f is convex, the element v_a is characterized by

$$\begin{cases} \langle u, v_a \rangle = a \\ \exists \mu \in \mathbb{R}, \nabla f(v_a) + \mu u = 0. \end{cases}$$

Then, by the differentiation of f,

$$\begin{cases} \langle u, v_a \rangle = a \\ v_a - v + \mu u = 0. \end{cases}$$

The second relation gives

$$\langle v_a - v, u \rangle + \mu \|u\|^2 = 0,$$

so

$$\mu = \frac{\langle u, v \rangle - a}{\|u\|^2}.$$

Therefore, we finally get

$$v_a = v - \frac{\langle u, v \rangle - a}{\|u\|^2} u.$$

To end, the distance from v to M has the value.

$$\|v - v_a\| = \frac{|\langle u, v \rangle - a|}{\|u\|}. \qquad \square$$

Problem 7.78. *(i) Consider a function $f : \mathbb{R} \to \mathbb{R}$ of class C^2, and let $\overline{x}$ be a simple root of f (i.e. $f(\overline{x}) = 0, f'(\overline{x}) \neq 0$). Deduce the algorithm of Newton's method from the Banach Principle, applied to an appropriate contraction.*

(ii) Show that, if the root is not simple, then the order of the convergence is not quadratic. In the case that the order of the root is known, appropriately modify the iterations such that a quadratic convergence holds.

Solution (i) Recall that the Newton iterations are given by:

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}.$$

Let put now this into a rigorous perspective. Let $L \in (0, 1)$. Let us denote by V a closed interval centered at $\overline{x}$ for which $f'(x) \neq 0$ for any $x \in V$ and

$$\left| \frac{f(x) f''(x)}{f'(x)^2} \right| < L, \ \forall x \in V. \tag{7.3.3}$$

The above relation is possible exactly because $f(\overline{x}) = 0$. Take now the function $g : V \to \mathbb{R}$

$$g(x) = x - \frac{f(x)}{f'(x)}.$$

Clearly, this function is well defined on V and since

$$g'(x) = 1 - \frac{f'(x)^2 - f(x)f''(x)}{f'(x)^2} = \frac{f(x)f''(x)}{f'(x)^2},$$

from the choice of V, we deduce that g is a contraction. On the other hand, $g(\overline{x}) = \overline{x}$, so, in particular,

$$\left|g(x) - \overline{x}\right| \leq L\,|x - \overline{x}|\,,\ \forall x \in V$$

which means that g applies V in V. Now, if we start with $x_0 \in V$, let us observe that the Newton iteration is in fact a Picard iteration associated to g. On the basis of the theory previously developed, the Newton iteration converges (for any initial date $x_0 \in V$) to the fixed point of which is exactly the root of f in V, that is $\overline{x}$. As shown before,

$$g'(\overline{x}) = 0$$

and we can apply Remark 6.1.1 in order to deduce that one has a quadratic convergence, that is

$$\begin{aligned}\lim_k \frac{x_{k+1} - \overline{x}}{(x_k - \overline{x})^2} &= \lim_{x \to \overline{x}} \frac{g(x) - \overline{x}}{(x - \overline{x})^2} \\ &= \lim_{x \to \overline{x}} \frac{1}{f'(x)} \frac{xf'(x) - f(x) - \overline{x}f'(x)}{(x - \overline{x})^2} = \frac{f''(\overline{x})}{2f'(\overline{x})} \in \mathbb{R}.\end{aligned}$$

(ii) The above discussion emphasizes that in order to be sure that the Newton iteration converges quadratically to the underlying root, it is necessary that x_0 to be in a neighborhood V of the solution where the derivative do not vanish and condition (7.3.3) takes place. Therefore, for functions with several roots, depending on initial data, we can find different solutions.

Let us suppose that f is as $f(x) = (x - \overline{x})^q u(x)$, where $q > 1$ is a natural number, u is of class C^2 and $u(\overline{x}) \neq 0$. Then g is

$$g(x) = x - \frac{(x - \overline{x})u(x)}{qu(x) + (x - \overline{x})u'(x)}$$

and, after calculation,

$$g'(x) = \frac{\left(1 - \frac{1}{q}\right) + (x - \overline{x})\frac{2u'(x)}{qu(x)} + (x - \overline{x})^2\frac{u''(x)}{q^2u(x)}}{\left[1 + (x - \overline{x})\frac{u'(x)}{qu(x)}\right]^2}.$$

For values close enough to $\overline{x}$, $\left|g'(x)\right| < 1$, so the Picard iterations converge to the fixed point of g which is exactly $\overline{x}$ (in the neighborhood one considers). On the other hand,

$g'(\overline{x}) = 1 - \frac{1}{q} \neq 0$, so the convergence is only linear. Therefore, the Newton procedure converges quadratically only for simple roots. If for a given root $\overline{x}$ the order of multiplicity q is known, then one can consider the function

$$g_q(x) = x - q\frac{f(x)}{f'(x)} = x - q\frac{(x-\overline{x})u(x)}{qu(x) + (x-\overline{x})u'(x)},$$

and a similar calculation shows that $g_q'(\overline{x}) = 0$, so, by means of the above reasonings, one gets again the quadratic convergence. □

7.4 Nonsmooth Optimization

Problem 7.79. *Let $g, h : \mathbb{R}^p \to \mathbb{R}$ be convex functions, and g of class C^1. We denote $f : \mathbb{R}^p \to \mathbb{R}$, $f = g + h$. The next assertions are equivalent:*

(i) $\overline{x}$ is a minimum point for f;

(ii) for every $x \in \mathbb{R}^p$,

$$\nabla g(\overline{x})(x - \overline{x}) + h(x) - h(\overline{x}) \geq 0;$$

(iii) for every $x \in \mathbb{R}^p$,

$$\nabla g(x)(x - \overline{x}) + h(x) - h(\overline{x}) \geq 0.$$

Solution $(i) \Rightarrow (ii)$ Let $\overline{x}$ be a minimum point of f on $\mathbb{R}^p$, let $x \in \mathbb{R}^p$ and $\lambda \in (0,1)$. Then

$$f(\overline{x}) \leq f(\lambda x + (1-\lambda)\overline{x}),$$

that is,

$$\begin{aligned} g(\overline{x}) + h(\overline{x}) &\leq g(\lambda x + (1-\lambda)\overline{x}) + h(\lambda x + (1-\lambda)\overline{x}) \\ &\leq g(\overline{x} + \lambda(x - \overline{x})) + \lambda h(x) + (1-\lambda)h(\overline{x}). \end{aligned}$$

We obtain

$$0 \leq g(\overline{x} + \lambda(x - \overline{x})) - g(\overline{x}) + \lambda(h(x) - h(\overline{x})),$$

from where, if we divide by λ,

$$0 \leq \frac{g(\overline{x} + \lambda(x - \overline{x})) - g(\overline{x})}{\lambda} + h(x) - h(\overline{x}).$$

Making $\lambda \to 0$, we get $0 \leq \nabla g(\overline{x})(x - \overline{x}) + h(x) - h(\overline{x})$.

$(ii) \Rightarrow (i)$ The convexity of g gives

$$g(x) \geq g(\overline{x}) + \nabla g(\overline{x})(x - \overline{x}),$$

which together with the hypothesis leads us to

$$g(x) + h(x) \geq g(\overline{x}) + h(\overline{x}).$$

$(ii) \Rightarrow (iii)$ Again, the convexity of g allows us to write

$$\left(\nabla g(x) - \nabla g(\overline{x})\right)(x - \overline{x}) \geq 0, \ \forall x \in \mathbb{R}^p,$$

which together with the hypothesis leads to the conclusion.

$(iii) \Rightarrow (ii)$ Let $\overline{x} \in \mathbb{R}^p$ with

$$\nabla g(x)(x - \overline{x}) + h(x) - h(\overline{x}) \geq 0, \ \forall x \in \mathbb{R}^p.$$

Let $x \in \mathbb{R}^p$ and $\lambda \in (0, 1)$. Then

$$\lambda \nabla g((1-\lambda)\overline{x} + \lambda x)(x - \overline{x}) + h((1-\lambda)\overline{x} + \lambda x) - h(\overline{x}) \geq 0,$$

and using the convexity of h,

$$\lambda \nabla g((1-\lambda)\overline{x} + \lambda x)(x - \overline{x}) + \lambda \left[h(x) - h(\overline{x})\right] \geq 0,$$

that is

$$\nabla g((1-\lambda)\overline{x} + \lambda x)(x - \overline{x}) + h(x) - h(\overline{x}) \geq 0.$$

But ∇g is a continuous application, so, for $\lambda \to 0$, we get the conclusion. □

Problem 7.80. *Show that every sublinear function is convex. Show that for any $x \in \mathbb{R}^p$,*

$$\partial f(x) = \{u \in \partial f(0) \mid \langle x, u \rangle = f(x)\}.$$

Solution The first part is obvious by direct calculation: for any $\alpha \in (0, 1)$ and any $x, x_2 \in \mathbb{R}^p$, the sublinearity of f yields

$$f(\alpha x + (1-\alpha)y) \leq f(\alpha x) + f((1-\alpha)y) = \alpha f(x) + (1-\alpha)f(y),$$

so f is convex.

Let us prove the second affirmation. Before that, notice that $f(0) = 0$. Take $u \in \partial f(0)$ and $x \in \mathbb{R}^p$ with $\langle x, u \rangle = f(x)$. Then for every $y \in \mathbb{R}^p$, $\langle y, u \rangle \leq f(y)$, whence

$$\langle y - x, u \rangle = \langle y, u \rangle - \langle x, u \rangle \leq f(y) - f(x).$$

This shows that $u \in \partial f(x)$. Conversely, consider $x \in \mathbb{R}^p$ and suppose that $u \in \partial f(x)$. So,

$$\langle y - x, u \rangle \leq f(y) - f(x), \ \forall y \in \mathbb{R}^p$$

which for $y = 0$ gives $\langle x, u \rangle \geq f(x)$ and for $y = x + tz$ with $t > 0$ and $z \in \mathbb{R}^p$ gives

$$t \langle z, u \rangle \leq f(x + tz) - f(x) \leq f(x) + tf(z) - f(x) = tf(z).$$

So, for every $z \in \mathbb{R}^p$, $\langle z, u - 0 \rangle \leq f(z) - f(0)$, i.e., $u \in \partial f(0)$. In particular, we also get $\langle x, u \rangle \leq f(x)$, so, in fact, coupled with the opposite inequality proved before, it follows that $\langle x, u \rangle = f(x)$. □

Problem 7.81. *Find the subdifferential of the norm.*

Solution The norm is a differentiable function away from 0. So, for any $x \in \mathbb{R}^p \setminus \{0\}$,

$$\partial \|\cdot\| (x) = \{\nabla \|\cdot\| (x)\} = \{(x_1 \cdot \|x\|^{-1}, x_2 \cdot \|x\|^{-1}, \ldots, x_p \cdot \|x\|^{-1})\}.$$

Let us find the subdifferential of the norm at 0. Following the definition,

$$\begin{aligned} u \in \partial \|\cdot\| (0) &\Leftrightarrow \langle u, x\rangle \le \|x\|, \ \forall x \in \mathbb{R}^p \\ &\Leftrightarrow \langle u, x\rangle \le 1, \ \forall x \in S(0,1) \Leftrightarrow |\langle u, x\rangle| \le 1, \ \forall x \in S(0,1) \Leftrightarrow \|u\| \le 1. \end{aligned}$$

Hence, $\partial \|\cdot\| (0) = D(0,1)$. Notice that the norm is a sublinear function, so $\partial \|\cdot\| (x)$ at $x \in \mathbb{R}^p \setminus \{0\}$ can be deduced as well from Problem 7.80.

Observe as well that, due to the convexity of the norm, one could also apply Example 5.1.5. □

Problem 7.82. *Let $K \subset \mathbb{R}^p$ be a closed convex cone with nonempty interior and let $e \in \operatorname{int} K$. Consider the function $s_e : \mathbb{R}^p \to \mathbb{R}$ given by*

$$s_e(x) := \inf\{t \in \mathbb{R} \mid x \in te - K\}. \tag{7.4.1}$$

Show that:

(i) s_e is well defined, i.e., the set in the right-hand side cannot be $\emptyset$ and the infimum cannot be $-\infty$.

(ii) for every $\lambda \in \mathbb{R}$ and $v \in \mathbb{R}^p$ one has

$$\{x \in \mathbb{R}^p \mid s_e(x) \le \lambda\} = \lambda e - K, \tag{7.4.2}$$

$$\{x \in \mathbb{R}^p \mid s_e(x) < \lambda\} = \lambda e - \operatorname{int} K, \tag{7.4.3}$$

$$\{x \in \mathbb{R}^p \mid s_e(x) = \lambda\} = \lambda e - \operatorname{bd} K, \tag{7.4.4}$$

and

$$s_e(v + \lambda e) = s_e(v) + \lambda; \tag{7.4.5}$$

(iii) s_e is sublinear;

(iv) s_e is strictly $\operatorname{int} K$–monotone, i.e., for all $x_1, x_2 \in \mathbb{R}^p$ with $x_2 - x_1 \in \operatorname{int} K$, one has $s_e(x_1) < s_e(x_2)$.

(v) the subdifferential of s_e at a point $x \in \mathbb{R}^p$ is

$$\partial s_e(x) = \{u \in -K^- \mid u(e) = 1, \langle u, x\rangle = s_e(x)\}.$$

Solution (i) The first item follows from Problem 7.30 (iii) and (iv).

(ii) Fix $\lambda \in \mathbb{R}$. The inclusion

$$\lambda e - K \subset \{x \in \mathbb{R}^p \mid s_e(x) \le \lambda\}$$

follows directly from the definition of s_e. Take now an element x from the second term of the above inclusion. Then for every $n \geq 1$, one has that $s_e(x) < \lambda + n^{-1}$, therefore there exists $t_n \in \mathbb{R}$ such that $t_n < \lambda + n^{-1}$ and $x \in t_n e - K$. Then, taking into account Problem 7.30 (i),

$$\begin{aligned} x &\in \left(\lambda + n^{-1}\right) e + \left(t_n - \lambda - n^{-1}\right) e - K \\ &\subset \left(\lambda + n^{-1}\right) e - K, \ \forall n \in \mathbb{N}^*. \end{aligned}$$

Making $n \to \infty$, the closedness of K yields the conclusion. So, we have proved that

$$\lambda e - K = \{x \in \mathbb{R}^p \mid s_e(x) \leq \lambda\}.$$

Notice that, in particular, a similar type of argument show that the infimum in definition of s_e is actually attained, i.e., for any $x \in \mathbb{R}^p$, $x \in s_e(x)e - K$. We prove now the relation concerning the strict level sets. To this end, take first $x \in \lambda e - \operatorname{int} K$. Then there exists $\varepsilon > 0$ such that $x \in \lambda e - \varepsilon e - \operatorname{int} K$, whence $s_e(x) \leq \lambda - \varepsilon < \lambda$. Take now $\lambda \in \mathbb{R}$ and $x \in \mathbb{R}^p$ with $s_e(x) < \lambda$. Then, by use of Problem 7.30 (ii),

$$\begin{aligned} x &\in s_e(x)e - K \subset \lambda e - (\lambda - s_e(y))e - K \\ &\subset \lambda e - \operatorname{int} K, \end{aligned}$$

and the converse inclusion follows. Then (7.4.3) holds. Now, (7.4.2) shows that s_e is lower semicontinuous, while (7.4.3) shows that it is upper semicontinuous. Then s_e is continuous. Again, using in conjunction (7.4.2) and (7.4.3), one obtains (7.4.4).

Let us prove the last equality of (ii). Fix $\lambda \in \mathbb{R}$ and $v \in \mathbb{R}^p$. There exists a sequence (t_n) such that $t_n \to s_e(v + \lambda e)$ and for every $n \in \mathbb{N}^*$,

$$v + \lambda e \in t_n e - K.$$

This means that

$$v \in (t_n - \lambda)e - K$$

and from (7.4.2), it follows that

$$s_e(v) \leq t_n - \lambda.$$

Passing to the limit as $n \to \infty$, one gets $s_e(v) + \lambda \leq s_e(v + \lambda e)$. Conversely, one has from (7.4.2) that

$$v \in s_e(v)e - K,$$

i.e.,

$$v + \lambda e \in (s_e(v) + \lambda)e - K,$$

whence

$$s_e(v + \lambda e) \leq s_e(v) + \lambda.$$

This final inequality completes the proof of (ii).

(iii) The relation (7.4.2) equally shows that

$$\operatorname{epi} s_e = \{(x,t) \in \mathbb{R}^p \times \mathbb{R} \mid x \in te - K\}, \tag{7.4.6}$$

and this is clearly a closed convex cone. So, s_e is sublinear (see Problem 7.31).

(iv) Take $x_1, x_2 \in \mathbb{R}^p$ with $x_2 - x_1 \in \operatorname{int} K$. Then

$$y_2 \in s_e(y_2)e - K,$$

whence

$$\begin{aligned} y_1 \in y_2 - \operatorname{int} K &\subset s_e(y_2)e - K - \operatorname{int} K \\ &\subset s_e(y_2)e - \operatorname{int} K. \end{aligned}$$

Therefore, $s_e(y_1) < s_e(y_2)$ and the conclusion follows.

(v) Taking into account Problem 7.80, it is enough to prove that the subdifferential of s_e at 0 is

$$\partial s_e(0) = \{u \in -K^- \mid u(e) = 1\}. \tag{7.4.7}$$

Notice that $s_e(0) = 0$, so an element $u \in \mathbb{R}^p$ is in $\partial s_e(0)$ if and only if

$$s_e(y) \geq \langle u, y \rangle \,,\ \forall x \in \mathbb{R}^p.$$

This means that for all $y \in \mathbb{R}^p$ and for all $\lambda \in \mathbb{R}$ with $\lambda \geq s_e(y)$, one has $\lambda \geq \langle u, y \rangle$. Consequently, taking into account (7.4.2), for all $y \in \lambda e - K$ one has $\lambda \geq \langle u, y \rangle$, i.e.,

$$\lambda \geq \lambda \langle u, e \rangle - \langle u, k \rangle \,,\ \forall k \in K.$$

Since the above inequality holds for all λ, making $k = 0$, one deduces that $\langle u, e \rangle = 1$. On the other hand, $\langle u, k \rangle \geq 0$ for all $k \in K$, whence $u \in -K^-$. The first inclusion in relation (7.4.7) is proved. For the converse, take $u \in -K^-$ such that $\langle u, e \rangle = 1$. Fix $y \in \mathbb{R}^p$ and take $\lambda \geq s_e(y)$. Then there exists $k \in K$ such that $y = \lambda e - k$. Accordingly,

$$\langle u, y \rangle = \lambda \langle u, e \rangle - \langle u, k \rangle \leq \lambda.$$

Since $\lambda \geq s_e(y)$ was arbitrarily chosen, one has

$$\langle u, y \rangle \leq s_e(y),\ \forall y \in \mathbb{R}^p,$$

i.e., $u \in \partial s_e(0)$. □

Remark 7.4.1. *The mapping studied in the previous problem is the Gerstewitz (Tammer) scalarization functional which successfully used in optimization problems with vector-valued functions. For more details, the reader is invited to consult (Göpfert et al., 2003).*

Up until now, we have avoided speaking about functions which can take $+\infty$ as a value. For the end of the discussion about convex functions we shall say something about this subject. We consider a function with real-extended values $f : \mathbb{R}^p \to \mathbb{R} \cup \{+\infty\}$. For such a function the domain is

$$\operatorname{dom} f := \{x \in \mathbb{R}^p \mid f(x) < +\infty\}.$$

The definition of convexity for f is formally the usual one (for any $x, y \in \mathbb{R}^p$ and $\alpha \in [0, 1]$, $f(\alpha x + (1 - \alpha)y) \le \alpha f(x) + (1 - \alpha)f(y)$) with the convention $+\infty + r = +\infty$ and $s(+\infty) = +\infty$ for any $r \in \mathbb{R} \cup \{+\infty\}$ and $s \in [0, \infty]$. It is clear that $x, y \in \operatorname{dom} f$ implies that $\alpha x + (1 - \alpha)y \in \operatorname{dom} f$, so $\operatorname{dom} f$ is a convex set. Moreover, it is enough to have the inequality in the definition of convexity fulfilled only for $x, y \in \operatorname{dom} f$. Many of the results given for convex functions with real values are still valid for functions with real-extended values. It is a useful exercise for the reader to verify how this adaptation can be made.

Problem 7.83. *Let $f, g : \mathbb{R}^p \to \mathbb{R} \cup \{+\infty\}$ be convex functions. Suppose that the function $f \square g : \mathbb{R}^p \to \mathbb{R} \cup \{+\infty\}$,*

$$(f \square g)(x) := \inf\{f(x - y) + g(y) \mid y \in \mathbb{R}^p\}$$

(called the convolution of f and g) is well defined (i.e., the infimum is not $-\infty$). Show that $\operatorname{dom} f \square g = \operatorname{dom} f + \operatorname{dom} g$, $f \square g = g \square f$, *and $f \square g$ is convex.*

Solution Take $x \in \operatorname{dom} f \square g$, that is $\inf\{f(x-y)+g(y) \mid y \in \mathbb{R}^p\} < +\infty$. Then there exists $y \in \mathbb{R}^p$ such that $f(x - y), g(y) \in \mathbb{R}$, so $x = (x - y) + y \in \operatorname{dom} f + \operatorname{dom} g$. Conversely, take $x \in \operatorname{dom} f + \operatorname{dom} g$. Then $x = x_1 + x_2$ with $x_1 \in \operatorname{dom} f$ and $x_2 \in \operatorname{dom} g$. We have

$$(f \square g)(x) = \inf\{f(x - y) + g(y) \mid y \in \mathbb{R}^p\} \le f(x_1) + f(x_2) \in \mathbb{R},$$

so $x \in \operatorname{dom} f \square g$.

The equality $f \square g = g \square f$ is obvious. We show that $f \square g$ is convex. Consider $x_1, x_2 \in \operatorname{dom} f \square g$ and $\alpha \in (0, 1)$. Take $t_1, t_2 \in \mathbb{R}$ with $t_1 > (f \square g)(x_1)$ and $t_2 > (f \square g)(x_2)$. Then there exist $y_1, y_2 \in \mathbb{R}^p$ with

$$\begin{aligned} f(x_1 - y_1) + g(y_1) &< t_1 \\ f(x_2 - y_2) + g(y_2) &< t_2. \end{aligned}$$

Then

$$\begin{aligned} (f \square g)(\alpha x_1 + (1 - \alpha)x_2) &\le f(\alpha x_1 + (1 - \alpha)x_2 - \alpha y_1 - (1 - \alpha)y_2) + g(\alpha y_1 + (1 - \alpha)y_2) \\ &= f(\alpha(x_1 - y_1) + (1 - \alpha)(x_2 - y_2)) + g(\alpha y_1 + (1 - \alpha)y_2) \\ &\le \alpha f(x_1 - y_1) + (1 - \alpha)f(x_2 - y_2) + \alpha g(y_1) + (1 - \alpha)g(y_2) \\ &= \alpha(f(x_1 - y_1) + g(y_1)) + (1 - \alpha)(f(x_2 - y_2) + g(y_2)) \end{aligned}$$

$$< \alpha t_1 + (1 - \alpha)t_2.$$

Since t_1, t_2 are chosen arbitrarily such that $t_1 > (f\Box g)(x_1)$ and $t_2 > (f\Box g)(x_2)$, we infer that

$$(f\Box g)(\alpha x_1 + (1 - \alpha)x_2) \le \alpha f\Box g)(x_1) + (1 - \alpha)(f\Box g)(x_2).$$

So, $f\Box g$ is convex. □

Problem 7.84. *If $f : \mathbb{R} \to \mathbb{R}$ is strictly convex and differentiable, compute $f\Box f$.*

Solution To compute a convolution means in fact to solve an optimization problem (in order to determine the infimum from the definition of the convolution). In our specific case, consider, for fixed x, $h : \mathbb{R} \to \mathbb{R}$, $h(y) = f(x - y) + f(y)$. The derivative of h is $h'(y) = -f'(x - y) + f'(y)$. Since f is strictly convex, its derivative is injective (being strictly increasing), so $\overline{y} = 2^{-1}x$ is the only critical point. A study of the monotonicity of h reveals that $\overline{y}$ is a global minimum point for h, so, $(f\Box f)(x) = f(x-2^{-1}x)+f(2^{-1}x) = 2f(2^{-1}x)$. □

Problem 7.85. *Let $A \subset \mathbb{R}^p$, $A \neq \mathbb{R}^p$ be a nonempty convex set. Show that the function $\mu : \mathbb{R}^p \to \mathbb{R} \cup \{+\infty\}$ given by*

$$\mu(y) := \begin{cases} -d_{\mathbb{R}^p \setminus A}(y), & y \in A \\ +\infty, & y \notin A. \end{cases}$$

is convex.

Solution Take $y_1, y_2 \in A = \operatorname{dom} \mu$ and observe that $D(y_1, d_{\mathbb{R}^p\setminus A}(y_1)) \subset \operatorname{cl} A$, and similarly for y_2. Then,

$$B := \operatorname{conv}\left[D(y_1, d_{\mathbb{R}^p\setminus A}(y_1)) \cup D(y_2, d_{\mathbb{R}^p\setminus A}(y_2))\right] \subset \operatorname{cl} A.$$

For every $\lambda \in [0, 1]$,

$$D(\lambda y_1 + (1 - \lambda)y_2, \lambda d_{\mathbb{R}^p\setminus A}(y_1) + (1 - \lambda)d_{\mathbb{R}^p\setminus A}(y_2)) \subset B,$$

whence

$$d_{\mathbb{R}^p\setminus A}(\lambda y_1 + (1 - \lambda)y_2) \ge \lambda d_{\mathbb{R}^p\setminus A}(y_1) + (1 - \lambda)d_{\mathbb{R}^p\setminus A}(y_2)$$

which is exactly the desired property. □

Problem 7.86. *Let $A \subset \mathbb{R}^p$, $A \neq \mathbb{R}^p$ be a nonempty set. The oriented distance function associated to A is $\Delta_A : \mathbb{R}^p \to \mathbb{R}$, given as*

$$\Delta_A(y) := d_A(y) - d_{\mathbb{R}^p\setminus A}(y).$$

Show that:

(i) Δ_A *is real-valued and* $1-$*Lipschitz;*

(ii) $\Delta_A(y) < 0$ *for every* $y \in \operatorname{int} A$, $\Delta_A(y) = 0$ *for every* $y \in \operatorname{bd} A$ *and* $\Delta_A(y) > 0$ *for every* $y \in \operatorname{int}(\mathbb{R}^p \setminus A)$;

(iii) if A *is closed, then* $A = \{y \in \mathbb{R}^p \mid \Delta_A(y) \le 0\}$;

(iv) if A *is convex, then* Δ_A *is convex;*

(v) if A *is a cone, then* Δ_A *is positively homogeneous;*

(vi) if A *is a closed convex cone, then* $-\Delta_A$ *is* $A-$*monotone*;

(vii) if A *is a closed convex cone with nonempty interior, then* $-\Delta_A$ *is strictly* $\operatorname{int} A-$*monotone;*

(viii) if A *is a closed convex cone, then for every* $y \in \mathbb{R}^p$, $\partial\Delta_{-A}(y) \subset -A^-$.

Solution (i) Let $y_1, y_2 \in \mathbb{R}^p$. If $y_1, y_2 \in A$ or $y_1, y_2 \in \mathbb{R}^p \setminus A$, the inequality

$$\left|\Delta_A(y_1) - \Delta_A(y_2)\right| \le \|y_1 - y_2\|$$

follows from the similar property of the distance to a set function. The same if at least one of the points is on the boundary of A. Suppose now that $y_1 \in \operatorname{int} A$ and $y_2 \in \operatorname{int}(\mathbb{R}^p \setminus A)$. Then it exists $\lambda \in (0,1)$ such that $y := \lambda y_1 + (1-\lambda)y_2 \in \operatorname{bd} A$. Then

$$\left|\Delta_A(y_1) - \Delta_A(y_2)\right| = d_{\mathbb{R}^p\setminus A}(y_1) + d_A(y_2) \le \|y_1 - y\| + \|y_2 - y\| = \|y_1 - y_2\|$$

and the proof of the property is complete.

(ii) Clearly, if $y \in \operatorname{int} A$, then $d_A(y) = 0$ and $d_{\mathbb{R}^p\setminus A}(y) < 0$, and similar for the last situation. If $y \in \operatorname{bd} A$, since $\operatorname{bd} A = \operatorname{bd}(\mathbb{R}^p \setminus A)$, then both $d_A(y)$ and $d_{\mathbb{R}^p\setminus A}(y)$ are zero.

(iii) Suppose that A is closed. If $y \in A$, then $d_A(y) = 0$ whence $\Delta_A(y) \le 0$. Conversely, if $\Delta_A(y) \le 0$, then supposing that $y \notin A$ we would have $y \in \operatorname{int}(\mathbb{R}^p \setminus A)$ whence $\Delta_A(y) > 0$. The conclusion follows.

(iv) Observe now that

$$\Delta_A(y) = \inf\{\|y - x\| + \mu(x) \mid x \in \mathbb{R}^p\}, \tag{7.4.8}$$

where μ is defined in the previous problem. Indeed, if $y \in A$, $\Delta_A(y) = -d_{\mathbb{R}^p\setminus A}(y)$, while

$$\inf\{\|y - x\| + \mu(x) \mid x \in \mathbb{R}^p\} = \inf\{\|y - x\| - d_{\mathbb{R}^p\setminus A}(x) \mid x \in A\} \le -d_{\mathbb{R}^p\setminus A}(y)$$

is obvious. Now observe that for every $x \in A$, $\|y - x\| + d_{\mathbb{R}^p\setminus A}(y) \ge d_{\mathbb{R}^p\setminus A}(x)$, which means that

$$\|y - x\| + \mu(x) \ge -d_{\mathbb{R}^p\setminus A}(y)$$

so $\inf\{\|y - x\| + \mu(x) \mid x \in \mathbb{R}^p\} \ge -d_{\mathbb{R}^p\setminus A}(y)$ and in this case (7.4.8) follows. If $y \in \mathbb{R}^p \setminus A$, $\Delta_A(y) = d_A(y)$. Clearly,

$$\inf\{\|y - x\| + \mu(x) \mid x \in \mathbb{R}^p\} \le \inf\{\|y - x\| \mid x \in A\} = d_A(y).$$

In order to prove the reverse inequality, observe first that there always exists a sequence $(y_n) \subset A$ with $\|y - y_n\| \to d_A(y)$. We claim that $d_{\mathbb{R}^p\setminus A}(y_n) \to 0$. Indeed, in the opposite case we can suppose, without loss of generality that there exists $\varepsilon > 0$ such that for every $n \in \mathbb{N}$, $B(y_n, \varepsilon) \subset A$. Take now $\delta \in (0, 1)$ such that $\delta \|y - y_n\| < \varepsilon$ for every n (note that such a δ does exist). Consider $z_n := y + (1 - \delta)(y - y_n)$. One has, on one hand, that

$$\|y_n - z_n\| = \delta \|y - y_n\| < \varepsilon,$$

hence $z_n \in A$ and, on the other hand,

$$d_A(y) \le \|y - z_n\| = (1 - \delta) \|y - y_n\| \to (1 - \delta) d_A(y) < d_A(y).$$

This contradiction can be eliminated only if we admit that $d_{\mathbb{R}^p \setminus A}(y_n) \to 0$. Therefore,

$$\|y - y_n\| - d_{\mathbb{R}^p \setminus A}(y_n) \ge \inf\{\|y - x\| + \mu(x) \mid x \in \mathbb{R}^p\},$$

whence, passing to the limit, we get

$$d_A(y) \ge \inf\{\|y - x\| + \mu(x) \mid x \in \mathbb{R}^p\},$$

so (7.4.8) is finally completely proven. So, according to relation (7.4.8), the function Δ_A appears as a convolution of two convex function, so it is itself a convex function.

(v) If A is a cone, then $\mathbb{R}^p \setminus A$ shares the property to be closed at multiplication with positive scalars, whence the property of Δ_A we are looking for comes from the similar property of the distance to a set function.

(vi) If A is a convex cone then, from (iv) and (v), we deduce that Δ_A is subadditive since

$$\begin{aligned} \Delta_A(y_1 + y_2) &= 2\Delta_A(2^{-1}y_1 + 2^{-1}y_2) \le 2\Delta_A(2^{-1}y_1) + 2\Delta_A(2^{-1}y_2) \\ &= \Delta_A(y_1) + \Delta_A(y_2). \end{aligned}$$

Now, if $y_2 - y_1 \in -A$, we can write successively

$$0 \ge \Delta_A(y_1 - y_2) \ge \Delta_A(y_1) - \Delta_A(y_2).$$

(vii) The last assertion is similar, taking into account (ii).

($viii$) Let $u \in \partial\Delta_{-A}(y)$. One has

$$\langle u, z - y\rangle \le \Delta_{-A}(z) - \Delta_{-A}(y), \quad \forall z \in \mathbb{R}^p. \tag{7.4.9}$$

From (vi), it follows $\Delta_{-A}(a + y) \le \Delta_{-A}(y)$ for every $a \in -A$ and whence, from (7.4.9), $\langle u, a\rangle \le 0$. This implies that for every $y \in \mathbb{R}^p$

$$\partial\Delta_{-A}(y) \subset -K^-.$$

The solution is complete. □

Problem 7.87. *Let $f : \mathbb{R}^p \to \mathbb{R} \cup \{+\infty\}$ be a function. One defines the conjugate of f as $f^* : \mathbb{R}^p \to \mathbb{R} \cup \{+\infty\}$,*

$$f^*(u) = \sup\{\langle x, u\rangle - f(x) \mid x \in \mathbb{R}^p\}.$$

Show that f^* is convex.

Solution Take $u_1, u_2 \in \operatorname{dom} f^\star$ and $\alpha \in (0,1)$. Consider $t \in \mathbb{R}$ with $t < f^\star(\alpha u_1 + (1-\alpha)u_2)$. Then there exists $x \in \mathbb{R}^p$ such that $t < \langle x, \alpha u_1 + (1-\alpha)u_2 \rangle - f(x)$. Thus,

$$\begin{aligned} t &< \alpha \langle x, u_1 \rangle - \alpha f(x) + (1-\alpha) \langle x, u_2 \rangle - (1-\alpha) f(x) \\ &\leq \alpha f^\star(u_1) + (1-\alpha) f^\star(u_2). \end{aligned}$$

Since $t < f^\star(\alpha u_1 + (1-\alpha)u_2)$ was chosen arbitrarily, we deduce that

$$f^\star(\alpha u_1 + (1-\alpha)u_2) \leq \alpha f^\star(u_1) + (1-\alpha) f^\star(u_2),$$

so $f^\star$ is convex. Note that this is a direct proof. □

Exercise 7.88. *Compute the conjugate for the following functions:*
(i) $f : \mathbb{R} \to \mathbb{R}$, $f(x) = \frac{1}{p} |x|^p$ *where* $p > 1$;
(ii) $f : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$,

$$f(x) = \begin{cases} -\ln x, \text{ if } x > 0 \\ +\infty, \text{ if } x \leq 0; \end{cases}$$

(iii) $f : \mathbb{R} \to \mathbb{R}$, $f(x) = e^x$;
(iv) $f : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$,

$$f(x) = \begin{cases} 0, \text{ if } |x| \leq 1 \\ +\infty, \text{ if } x > 1. \end{cases}$$

Solution As in the case of convolution, to compute a conjugate is equivalent to solving an optimization problem (in order to determine the supremum in the definition of the conjugate). For our functions, it is a routine to compute the derivatives for $h(x) = xu - f(u)$ and to get the variation of this function and, from that, the conclusions. We obtain:
(i) $f^\star : \mathbb{R} \to \mathbb{R}$, $f^\star(u) = \frac{1}{q} |u|^q$, where $\frac{1}{p} + \frac{1}{q} = 1$;
(ii) $f^\star : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$,

$$f^\star(u) = \begin{cases} -\ln(-u), \text{ if } u < 0 \\ +\infty, \text{ otherwise}; \end{cases}$$

(iii) $f^\star : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$,

$$f^\star(u) = \begin{cases} u \ln(u) - u, \text{ if } u > 0 \\ 0, \text{ if } u = 0 \\ +\infty, \text{ if } u < 0; \end{cases}$$

(iv) $f^\star : \mathbb{R} \to \mathbb{R}$, $f^\star(u) = |u|$. □

Problem 7.89. *Let* $f : \mathbb{R}^k \to \mathbb{R}$ *be a convex function. Show that for every* $x, u \in \mathbb{R}^k$,

$$\langle u, x \rangle \leq f(x) + f^\star(u)$$

with equality if and only if $u \in \partial f(x)$.
Write the above inequality for the convex function $f : \mathbb{R} \to \mathbb{R}, f(x) = \frac{1}{p}|x|^p$ *, where* $p > 1$. *Find* $\partial f(x)$.

Solution From the definition of the conjugate, for every $u \in \mathbb{R}^k$,

$$f^*(u) = \sup\{\langle x, u\rangle - f(x) \mid x \in \mathbb{R}^k\} \geq \langle u, x\rangle - f(x), \ \forall x \in \mathbb{R}^k,$$

so the inequality holds. The equality means that the supremum is attained at x, so for any $y \in \mathbb{R}^k$,

$$\langle x, u\rangle - f(x) \geq \langle y, u\rangle - f(y),$$

which means that $u \in \partial f(x)$.

Taking into account (i) in the preceding problem, the inequality becomes the inequality of Young already studied before (see Subsection 2.2.3), namely

$$xu \leq \frac{|x|^p}{p} + \frac{|u|^q}{q}, \ \forall x, u \in \mathbb{R}, \text{ where } p, q > 1, \ \frac{1}{p} + \frac{1}{q} = 1.$$

We have seen that

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q}, \ \forall p, q > 1, \ \frac{1}{p} + \frac{1}{q} = 1, \ a, b \geq 0,$$

with the equality if and only if $a^p = b^q$. Thus, for every $x, u \in \mathbb{R}$,

$$xu = \langle x, u\rangle \leq |x|\,|u| \leq \frac{|x|^p}{p} + \frac{|u|^q}{q},$$

with the equality if and only of $xu = |x|^p = |u|^q$, whence $\partial f(x) = \{u \in \mathbb{R} \mid xu = |x|^p = |u|^q\}$. □

Problem 7.90. *Let* $C \subset \mathbb{R}^p$ *be a nonempty closed convex set. Show that for every* $x \in \operatorname{bd} C$, $N(C, x) \neq \{0\}$. *Deduce that the distance function* d_C *is not differentiable at the points of* $\operatorname{bd} C$.

Solution Without loss of generality, we can suppose that $x = 0 \in \operatorname{bd} C$. There exists a sequence $(x_k) \subset \mathbb{R}^p \setminus C$ such that $x_k \to 0$. Take, for every k, y_k as the projection of x_k on C. From the continuity of the projection (Proposition 4.1.2), one deduces that $y_k \to 0$. Consider $z_k := x_k - y_k \neq 0$ (for every k) and the limit $\mu \neq 0$ of a convergent subsequence of the bounded sequence $\|z_k\|^{-1} z_k$. Take $x \in C$. Since for any $\alpha \in (0, 1)$, $\alpha x + (1-\alpha)y_k \in C$, one deduces that $\|y_k - x_k\| \leq \|\alpha x + (1-\alpha)y_k - x_k\|$, so

$$\|y_k - x_k\|^2 \leq \|\alpha x + (1-\alpha)y_k - x_k\|^2.$$

After some calculations, one gets that

$$2\langle z_k, x - y_k\rangle - \alpha\|x - y_k\|^2 \leq 0.$$

Making $\alpha \to 0$, one gets $\langle z_k, x - y_k \rangle \leq 0$ for any k. In particular,

$$\left\langle \|z_k\|^{-1} z_k, x - y_k \right\rangle \leq 0$$

for any k. Passing to the limit, we obtain that $\langle \mu, x \rangle \leq 0$ for all $x \in C$. Then $\mu \in N(C, 0)$, so the conclusion.

If d_C would be differentiable at a point $\overline{x}$ in its boundary, then $\partial d_C(\overline{x})$ reduces to a singleton (Proposition 4.2.3). But, as shown in Proposition 4.3.3, $N(C, \overline{x}) \cap D(0, 1) \subset \partial d_C(\overline{x})$. In view of the above proved fact, $N(C, \overline{x}) \cap D(0, 1)$ cannot be a singleton, so a contradiction arises. Therefore, d_C cannot be differentiable at $\overline{x}$.

Remark as well that a solution of the problem could be easily deduced from the more general result of Corollary 5.2.29. □

Problem 7.91. *Let $C \subset \mathbb{R}^p$ be a nonempty closed, convex set. Show that d_C^2 is differentiable and for every $x \in \mathbb{R}^p$, and $\nabla d_C^2(x) = 2(x - \operatorname{pr}_C x)$.*

Solution Let $x \in \mathbb{R}^p$ and $h \in \mathbb{R}^p \setminus \{0\}$. Therefore,

$$\begin{aligned} d_C^2(x+h) - d_C^2(x) &\geq d_C^2(x+h) - \left\| x - \operatorname{pr}_C(x+h) \right\|^2 \\ &= \left\| \operatorname{pr}_C(x+h) - (x+h) \right\|^2 - \left\| x - \operatorname{pr}_C(x+h) \right\|^2 \\ &= 2 \left\langle x - \operatorname{pr}_C(x+h), h \right\rangle + \|h\|^2 , \end{aligned}$$

so, using Proposition 4.1.2,

$$\begin{aligned} d_C^2(x+h) - d_C^2(x) - 2\langle x - \operatorname{pr}_C x, h\rangle &\geq 2\left\langle x - \operatorname{pr}_C(x+h), h \right\rangle - 2\langle x - \operatorname{pr}_C x, h\rangle + \|h\|^2 \\ &\geq -2 \left\| \operatorname{pr}_C x - \operatorname{pr}_C(x+h) \right\| \|h\| + \|h\|^2 \\ &\geq -\|h\|^2 . \end{aligned}$$

On the other hand,

$$\begin{aligned} d_C^2(x+h) - d_C^2(x) &\leq \left\| (x+h) - \operatorname{pr}_C x \right\|^2 - \|x - \operatorname{pr}_C x\|^2 \\ &= 2\langle x - \operatorname{pr}_C x, h\rangle + \|h\|^2 . \end{aligned}$$

Finally, we can write

$$-\|h\|^2 \leq d_C^2(x+h) - d_C^2(x) - 2\langle x - \operatorname{pr}_C x, h\rangle \leq \|h\|^2 ,$$

so,

$$-\|h\| \leq \|h\|^{-1} \left(d_C^2(x+h) - d_C^2(x) - 2\langle x - \operatorname{pr}_C x, h\rangle \right) \leq \|h\| ,$$

which confirms that d_C^2 is differentiable at every $x \in \mathbb{R}^p$ and $\nabla d_C^2(x) = 2(x - \operatorname{pr}_C x)$. □

Problem 7.92. *Let $a_1, a_2, a_3 \in \mathbb{R}^p$ non colinear points. Show that there exists a unique point $\overline{x} \in \mathbb{R}^p$ which minimize on $\mathbb{R}^p$ the function $f : \mathbb{R}^p \to \mathbb{R}$,*

$$f(x) = \|x - a_1\| + \|x - a_2\| + \|x - a_3\| .$$

Then prove that $\overline{x} \in \operatorname{conv}\{a_1, a_2, a_3\}$. Deduce that if $\overline{x} \notin \{a_1, a_2, a_3\}$, then the angle between $\overline{x} - a_i$ and $\overline{x} - a_j$ is $\frac{2\pi}{3}$ for all $i, j \in \overline{1,3}$, $i \neq j$. (The point $\overline{x}$ is called the Torricelli point of the triangle of vertices a_1, a_2, a_3.)

Solution It is easy to observe that f is convex and coercive, so it has a global minimum on $\mathbb{R}^p$. Moreover, f is strictly convex. Indeed, for $x, y \in \mathbb{R}^p$, the equality $\|x\| + \|y\| = \|x + y\|$ is possible if and only if x, y are on a half-line passing through the origin. If there exist $x, y \in \mathbb{R}^p$, $x \neq y$ and $\alpha \in (0, 1)$ with $f(\alpha x + (1 - \alpha)y) = \alpha f(x) + (1 - \alpha)f(y)$, then a_1, a_2, a_3 must be on the same straight-line determined by x and y, which is a contradiction. Therefore, there exists a unique element $\overline{x} \in \mathbb{R}^p$, global minimum of f. Thus, from Theorem 4.3.1, $0 \in \partial f(\overline{x})$. But, by use of Theorem 4.2.7 and Problem 7.81,

$$\partial f(x) = \begin{cases} \sum_{i=1}^{3} \|x - a_i\|^{-1} (x - a_i), & \text{if } x \notin \{a_1, a_2, a_3\} \\ \sum_{i=1, i\neq j}^{3} \|x - a_i\|^{-1} (x - a_i) + D(0, 1), & \text{if } x = a_j. \end{cases}$$

If $\overline{x}$ is one of the points a_1, a_2, a_3, then surely belongs to $\operatorname{conv}\{a_1, a_2, a_3\}$. Suppose that $\overline{x} \notin \{a_1, a_2, a_3\}$. Then

$$\sum_{i=1}^{3} \|\overline{x} - a_i\|^{-1} (\overline{x} - a_i) = 0,$$

whence

$$\overline{x} = \left(\sum_{i=1}^{3} \|\overline{x} - a_i\|^{-1} \right)^{-1} \sum_{i=1}^{3} \|\overline{x} - a_i\|^{-1} a_i \in \operatorname{conv}\{a_1, a_2, a_3\}.$$

For the last part, if $\overline{x} \notin \{a_1, a_2, a_3\}$, again from $\sum_{i=1}^{3} \|\overline{x} - a_i\|^{-1} (\overline{x} - a_i) = 0$ we deduce that $\left\langle \|\overline{x} - a_i\|^{-1} (\overline{x} - a_i), \|\overline{x} - a_j\|^{-1} (\overline{x} - a_j) \right\rangle = 2^{-1}$ for all $i, j \in \overline{1,3}$, $i \neq j$. Consequently, the angle between $\overline{x} - a_i$ and $\overline{x} - a_j$ is always $\frac{2\pi}{3}$. This ends the solution. □

Problem 7.93. *Consider the Fermat's principle concerning the propagation of light: "in an inhomogeneous medium a ray of light travels between two points along the path requiring the shortest time". Prove the law of refraction: when light passes from one medium to another, the directions of the light satisfy $\frac{\sin \alpha_1}{v_1} = \frac{\sin \alpha_2}{v_2}$, where α_1, α_2 are the angles between the directions of the ray and the normal to the surface which separates the two media, and v_1, v_2 are the speeds of the light in the two media, respectively.*

Solution Suppose that the light travels from the point $(0, a)$ in the first medium to the point (b, d), $b > 0$ in the second one (see the figure below).

For easy calculus, suppose that the surface which separates the two media is the Ox axis. Let v_1, v_2 be the speeds of the light in the two media, respectively. Suppose that the light passes from the first medium to the second one at the point $(x, 0)$, where x is to be determined following Fermat's principle. The amount of time spent by the

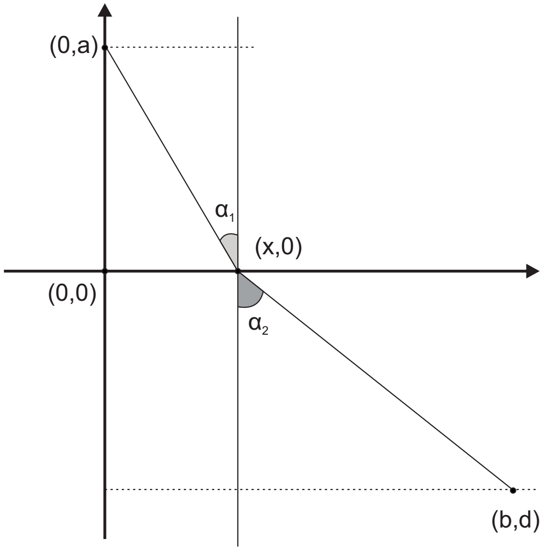

Figure 7.4: Propagation of light.

light in the first and the second media are, respectively,

$$t_1 = \frac{\sqrt{a^2 + x^2}}{v_1}$$

$$t_2 = \frac{\sqrt{(b - x)^2 + c^2}}{v_2},$$

so the total time to minimize is $\frac{\sqrt{a^2+x^2}}{v_1} + \frac{\sqrt{(b-x)^2+c^2}}{v_2}$. Take $f : \mathbb{R} \to \mathbb{R}$,

$$f(x) = \frac{\sqrt{a^2 + x^2}}{v_1} + \frac{\sqrt{(b - x)^2 + c^2}}{v_2}.$$

Now the problem is to minimize f on $\mathbb{R}$. The derivative of f is

$$f'(x) = \frac{x}{v_1\sqrt{a^2 + x^2}} + \frac{x - b}{v_2\sqrt{(x - b)^2 + c^2}}.$$

Since $f'(0) = \frac{-b}{v_2\sqrt{b^2+c^2}} < 0$ and $f'(b) = \frac{b}{v_1\sqrt{a^2+b^2}} > 0$ and

$$f''(x) = \frac{1}{v_1} \cdot \frac{a^2}{(a^2 + x^2)^{\frac{3}{2}}} + \frac{1}{v_2} \cdot \frac{c^2}{\left((x - b)^2 + c^2\right)^{\frac{3}{2}}} > 0, \ \forall x \in \mathbb{R},$$

there exists a unique critical point of f situated in the interval $[0, b]$. Denote by $\overline{x}$ this critical point. The variation of f shows that $\overline{x}$ is a minimum point. Then

$$\frac{\overline{x}}{v_1\sqrt{a^2 + \overline{x}^2}} = \frac{b - \overline{x}}{v_2\sqrt{(\overline{x} - b)^2 + c^2}}.$$

But,

$$\frac{\overline{x}}{\sqrt{a^2+\overline{x}^2}} = \sin\alpha_1$$

and

$$\frac{b-\overline{x}}{\sqrt{(\overline{x}-b)^2+c^2}} = \sin\alpha_2,$$

so,

$$\frac{\sin\alpha_1}{v_1} = \frac{\sin\alpha_2}{v_2}.$$

and the light refraction law in proved. □

Exercise 7.94. *Show that* $f:(0,\infty)\times(0,\infty)\to\mathbb{R}$,

$$f(x_1,x_2) = \frac{1}{x_1} + \frac{1}{x_2} - \frac{1}{x_1+x_2}$$

is convex.

Solution The calculus of second-order partial derivatives leads to the following form of the Hessian matrix at a point x in the domain of f :

$$\begin{pmatrix} \frac{2}{x_1^3} - \frac{2}{(x_1+x_2)^3} & -\frac{2}{(x_1+x_2)^3} \\ -\frac{2}{(x_1+x_2)^3} & \frac{2}{x_2^3} - \frac{2}{(x_1+x_2)^3} \end{pmatrix}.$$

On this basis, it is easy to verify that $\nabla^2 f(x)$ is positive definite for every $x\in(0,\infty)\times(0,\infty)$, so, according to Theorem 2.2.10, the function f is convex. □

Problem 7.95. *Let* $a,b,c,d\in\mathbb{R}$ *with* $a<b$, $c<d$, *and* $f:[a,b]\times[c,d]\to\mathbb{R}$. *Define* $\varphi:[a,b]\to\mathbb{R}$,

$$\varphi(x) = \inf\{f(x,y) \mid y\in[c,d]\}.$$

Show that φ *is well defined and continuous.*

Solution The mapping

$$[c,d]\ni y\longmapsto f(x,y)$$

is continuous and hence attains its minimum on $[c,d]$. Then there exist $y_x\in[c,d]$ such that $\varphi(x)=f(x,y_x)$.

From Cantor Theorem (Theorem 1.2.24), f is uniformly continuous on the compact set $[a,b]\times[c,d]$: for every $\varepsilon>0$, there exists $\delta_\varepsilon>0$, such that for every (x',y'), $(x'',y'')\in[a,b]\times[c,d]$ with $\left\|(x',y')-(x'',y'')\right\|<\delta_\varepsilon$, one has $\left|f(x',y')-f(x'',y'')\right|<\varepsilon$.

Let $x',x''\in[a,b]$ with $\left|x'-x''\right|\le\delta_\varepsilon$. Then

$$\begin{aligned}\varphi(x')-\varphi(x'') &= f(x',y_{x'})-f(x'',y_{x''})\\ &\le f(x',y_{x''})-f(x'',y_{x''})<\varepsilon.\end{aligned}$$

Changing the roles of x' and x'', we get

$$\varphi(x'') - \varphi(x') < \varepsilon,$$

from where the conclusion. □

Problem 7.96. *Let $f : \mathbb{R}^p \to \mathbb{R}$ be a function and $M \subset \mathbb{R}^p$ be a closed set. Consider the problem*

$$\min f(x), \ x \in M.$$

Suppose that $\overline{x}$ is a solution of this problem and that f is locally Lipschitz of constant $L \geq 0$ around $\overline{x}$. Let $\varphi : \mathbb{R}^p \to [0, \infty)$ with $\varphi(\overline{x}) = 0$ a lower semicontinuous function. Show that one and only one of the following holds:

(i) there exists $\lambda > 0$ such that $\overline{x}$ is a local minimum without constraints for $f + \lambda\varphi$;

(ii) there exists $(z_n)_{n\in\mathbb{N}^*} \subset \mathbb{R}^p \setminus M$, $z_n \to \overline{x}$ such that for every $n \in \mathbb{N}^*$, the mapping

$$x \mapsto \varphi(x) + n^{-1} \|x - z_n\|$$

attains its minimum at z_n.

Solution We can have one and only one of the following possibilities:

– there exists a neighborhood V of $\overline{x}$ and $a > 0$ such that for every $x \in V$,

$$a\varphi(x) \geq d(x, M);$$

– there exists $x_n \to \overline{x}$ such that

$$2n\varphi(x_n) < d(x_n, M).$$

In the first situation, let $\alpha > 0$ such that $B(\overline{x}, \alpha) \subset U \cap V$, where U is the neighborhood of $\overline{x}$ where the Lipschitz condition holds and where $\overline{x}$ is a minimum point.

For every $x \in B(\overline{x}, \alpha/3)$ there exists $u \in M$ with

$$\|x - u\| \leq 2d(x, M) \leq 2\|x - \overline{x}\|.$$

Hence, in particular,

$$\|u - \overline{x}\| \leq \|u - x\| + \|x - \overline{x}\| \leq 3\|x - \overline{x}\| < \alpha.$$

Then

$$f(x) \geq f(u) - L\|u - x\| \geq f(\overline{x}) - 2La\varphi(x)$$

so we are in the first alternative of the conclusion.

In the second situation, since φ has positive values, $(x_n) \subset \mathbb{R}^p \setminus M$ and

$$\varphi(x_n) \leq \inf_{x\in\mathbb{R}^p} \varphi(x) + 2^{-1}n^{-1}d(x_n, M).$$

We take $\varepsilon := 2^{-1}n^{-1}d(x_n, M) > 0$ and $\delta := 2^{-1}d(x_n, M)$, and we apply Ekeland Variational Principle to φ for the ε–minimum x_n. We infer that there exists $z_n \in \mathbb{R}^p$ with

$$\begin{aligned}
&\varphi(z_n) \leq \varphi(x_n) < n^{-1}d(x_n, M)\\
&\|z_n - x_n\| \leq 2^{-1}d(x_n, M)\\
&\varphi(z_n) \leq \varphi(x) + n^{-1}\|x - z_n\|, \ \forall x \in \mathbb{R}^p.
\end{aligned}$$

The last relation shows that for every $n \in \mathbb{N}^\star$ the mapping $x \mapsto \varphi(x) + n^{-1}\|x - z_n\|$ attains its minimum at z_n. Moreover,

$$\begin{aligned}
\|z_n - \overline{x}\| &\leq \|z_n - x_n\| + \|x_n - \overline{x}\|\\
&\leq d(x_n, M) + \|x_n - \overline{x}\| \leq 2\|x_n - \overline{x}\|,
\end{aligned}$$

whence $z_n \to \overline{x}$. If we would have $z_n \in M$, then

$$\|z_n - x_n\| \leq 2^{-1}d(x_n, M) < d(x_n, M) \leq \|z_n - x_n\|,$$

which is a contradiction. The solution is now complete. □

Let us remark that Problem 7.96 is a generalization of Theorem 4.3.2.

Exercise 7.97. *Consider the function $f : \mathbb{R}^2 \to \mathbb{R}$ given by $f(x, y) := |x_1| - |x_2|$. Prove that*

$$\partial_C f(0, 0) = \operatorname{conv}\left\{(-1, -1), (-1, 1), (1, -1), (1, 1)\right\}.$$

Solution The formula can be deduced in a similar manner to Example 5.1.9, taking into account formula (5.1.8). □

Other useful properties of the Clarke tangent and the normal cones are contained in the next exercise.

Problem 7.98. *Let $f, g : \mathbb{R}^p \to \mathbb{R}$ be locally Lipschitz functions around $\overline{x}$. The following hold:*

(i) $f \cdot g : \mathbb{R}^p \to \mathbb{R}$, given by $(f \cdot g)(x) := f(x) \cdot g(x)$ for every x is locally Lipschitz around $\overline{x}$, and

$$\partial_C(f \cdot g)(\overline{x}) \subset g(\overline{x})\partial_C f(\overline{x}) + f(\overline{x})\partial_C g(\overline{x}). \tag{7.4.10}$$

If, moreover, $f(x) \geq 0$ and $g(x) \geq 0$ for every x, and f and g are both regular, then fg is regular and equality holds in (7.4.10).

(ii) Suppose $g(x) \neq 0$ for every x. Then the function $\frac{f}{g} : \mathbb{R}^p \to \mathbb{R}$, given by $\left(\frac{f}{g}\right)(x) := \frac{f(x)}{g(x)}$ for every x is locally Lipschitz around $\overline{x}$ and

$$\partial_C\left(\frac{f}{g}\right)(\overline{x}) \subset \frac{g(\overline{x})\partial_C f(\overline{x}) - f(\overline{x})\partial_C g(\overline{x})}{g^2(\overline{x})}. \tag{7.4.11}$$

If, moreover, $f(x) \geq 0$ and $g(x) > 0$ for every x, and f and $-g$ are both regular, then $\frac{f}{g}$ is regular and equality holds in (7.4.11).

Solution For (i), take $h : \mathbb{R}^2 \to \mathbb{R}$ given by $h(x_1, x_2) := x_1 \cdot x_2$. Then Theorem 5.1.18 applies, and one has the conclusion. The proof of (ii) is similar. □

Problem 7.99. *Let $f_1, \ldots, f_k : \mathbb{R}^p \to \mathbb{R}$ be locally Lipschitz functions around x. Then the function $h : \mathbb{R}^p \to \mathbb{R}$ given as*

$$h(x) := \max \{f_1(x), \ldots, f_k(x)\}$$

is locally Lipschitz around x and

$$\partial_C h(x) \subset \operatorname{conv} \bigcup_{i \in A(x)} \partial_C f_i(x),$$

where, $A(x) = \left\{ i \in \overline{1, k} \mid h(x) = f_i(x) \right\}$ denotes the set of active indices at x.

Solution Observe that $h = g \circ f$, where $f : \mathbb{R}^p \to \mathbb{R}^k$ is $f(x) = (f_1(x), \ldots, f_k(x))$ and $g : \mathbb{R}^k \to \mathbb{R}$ is $g(y) = \max \{y_1, \ldots, y_k\}$. Since f and g are locally Lipschitz, one may apply Theorem 5.1.18.

Observe moreover that, since g is convex, its Clarke subdifferential coincides with the convex subdifferential, i.e.,

$$\begin{aligned}
\partial_C g(f(x)) &= \partial g(f(x)) \\
&= \left\{ \eta \in \mathbb{R}^k_+ \mid \eta_1 + \ldots + \eta_k = 1,\ \eta_1 f_1(x) + \ldots + \eta_k f_k(x) \geq h(x) \right\} \\
&= \left\{ \eta \in \mathbb{R}^k_+ \mid \eta_1 + \ldots + \eta_k = 1,\ \eta_1 f_1(x) + \ldots + \eta_k f_k(x) = h(x) \right\} \\
&= \left\{ \eta \in \mathbb{R}^k_+ \mid \eta_i = 0 \text{ if } i \notin A(x),\ \eta_1 + \ldots + \eta_k = 1 \right\}.
\end{aligned}$$

From (5.1.17), we get that

$$\begin{aligned}
\partial_C h(x) &\subset \operatorname{cl} \operatorname{conv} \left\{ \partial_C \langle \eta, f \rangle (x) \mid \eta \in \mathbb{R}^k_+,\ \eta_i = 0 \text{ if } i \notin A(x),\ \sum_{i=1}^{k} \eta_i = 1 \right\} \\
&= \operatorname{cl} \operatorname{conv} \left\{ \partial_C \left(\sum_{i \in A(x)} \eta_i f_i(x) \right) \mid \eta_i \geq 0,\ \sum_{i \in A(x)} \eta_i = 1 \right\} \\
&\subset \operatorname{cl} \operatorname{conv} \left\{ \sum_{i \in A(x)} \eta_i \partial_C f_i(x) \mid \eta_i \geq 0,\ \sum_{i \in A(x)} \eta_i = 1 \right\}.
\end{aligned}$$

Since $\sum_{i \in A(x)} \eta_i \partial_C f_i(x)$ from above is a convex combination of the sets $\partial_C f_i(x)$ for $i \in A(x)$, each of one being convex and compact, one gets the desired conclusion. □

Problem 7.100. *Suppose $A_1 \subset \mathbb{R}^p$ and $A_2 \subset \mathbb{R}^q$ are two sets, and take $x_1 \in A_1$, $x_2 \in A_2$. Prove that*

$$T_C(A_1 \times A_2, (x_1, x_2)) = T_C(A_1, x_1) \times T_C(A_2, x_2)$$

and

$$N_C(A_1 \times A_2, (x_1, x_2)) = N_C(A_1, x_1) \times N_C(A_2, x_2)$$

Solution The first equality follows easily from the characterization given by Theorem 5.1.25 (ii). Then the second equality follows by polarity. □

Problem 7.101. *Consider a nonempty set* $A \subset \mathbb{R}^p$ *and* $\overline{x} \in A$. *Define the indicator function of* A *as* $\iota_A : \mathbb{R}^p \to \mathbb{R} \cup \{+\infty\}$,

$$\iota_A(x) := \begin{cases} 0, & \text{if } x \in A \\ \infty, & \text{if } x \notin A. \end{cases}$$

Prove that

$$\partial_F \iota_A(\overline{x}) = N_F(A, \overline{x})$$

and

$$\partial_M \iota_A(\overline{x}) = \partial^\infty \iota_A(\overline{x}) = N_M(A, \overline{x}).$$

Solution It follows from the fact that $\operatorname{epi} \iota_A = A \times [0, \infty)$, hence by Proposition 5.2.3 (vi)

$$N_F(\operatorname{epi} \iota_A, (\overline{x}, 0)) = N_F(A, \overline{x}) \times (-\infty, 0]$$

and

$$N_M(\operatorname{epi} \iota_A, (\overline{x}, 0)) = N_M(A, \overline{x}) \times (-\infty, 0].$$ □

Exercise 7.102. *Consider the sets in* $\mathbb{R}^3$ *given by*

$$A = \left\{(t, p, q) \mid (p, q) \in \left[(0, 0), (\cos t, \sin t)\right]\right\},$$
$$Q = \left\{(t, p, q) \mid (p, q) \in \operatorname{cone} D\left((\cos t, \sin t), \frac{\sqrt{2}}{2}\right)\right\}.$$

Compute the Fréchet and the Mordukhovich normal cones to these sets at $(0, 0, 0)$.

Solution Observe first that the set A can be equivalently written as

$$\begin{cases} x(u, v) = u \\ y(u, v) = v \cos u, \qquad u \in \mathbb{R},\ v \in [0, 1]. \\ z(u, v) = v \sin u \end{cases}$$

Our intention is to compute the Fréchet normal cone to the set A using its equality to $T_B(A, (x, y, z))^-$. Consider hence points (x_0, y_0, z_0) from A close to $(0, 0, 0)$. If (x_0, y_0, z_0) is such that $v \in (0, 1)$, then the Fréchet normal cone to this point is the cone generated by the normal vector to the surface, whose expression is $(-v, -\sin u, \cos u)$. In fact, it is the line

$$\left\{(x, y, z) \in \mathbb{R}^3 \mid \frac{x - u}{-v} = \frac{y - v\cos u}{-\sin u} = \frac{z - v \sin u}{\cos u}\right\}. \tag{7.4.12}$$

When $(x_n, y_n, z_n) \to (0,0,0)$, it means that the corresponding $(u_n, v_n) \to (0,0)$, which gives, when passing to the limit when considering elements from the set (7.4.12), the line $\{0\} \times \{0\} \times \mathbb{R}$, i.e., the Oz axis. Remark also that, for the point $(x_0, y_0, z_0) \in A$, the Bouligand tangent cone is the tangent plane to the surface at (x_0, y_0, z_0) translated to $(0,0,0)$. This plane is

$$\left\{(x,y,z) \in \mathbb{R}^3 \mid -vx - y\sin u + z\cos u = 0\right\}. \tag{7.4.13}$$

Consider now $(x_0, y_0, z_0) \in A$ is such that $v = 0$, i.e., points of the type $(u, 0, 0)$. In this case, the Bouligand tangent cone is the half-plane obtained by taking $y \geq 0$ in the equation (7.4.13):

$$P := \left\{(x,y,z) \in \mathbb{R}^3 \mid -y\sin u + z\cos u = 0,\ y \geq 0\right\}.$$

In this case, the Fréchet normal cone is P^-, i.e.,

$$\left\{(x,y,z) \in \mathbb{R}^3 \mid x = 0,\ y\cos u + z\sin u = 0,\ y \leq 0\right\}.$$

When $(x_n, y_n, z_n) \to (0,0,0)$, one has again $(u_n, v_n) \to (0,0)$, hence passing to the limit when taking elements from the previous set one obtains the half-plane

$$\{(x,y,z) \in \mathbb{R}^3 \mid x = 0,\ y \leq 0\}.$$

In conclusion,

$$N_M(A, (0,0,0)) = \{(x,y,z) \in \mathbb{R}^3 \mid x = 0,\ y \leq 0\}.$$

Remark also that the structure of Q and the calculus of $N_M(Q, (0,0,0))$ are somehow similar. This is because Q is bounded by two helicoids, given parametrically as

$$(H_1): \begin{cases} x(u,v) = u \\ y(u,v) = v\cos\left(u + \dfrac{\pi}{4}\right) \\ z(u,v) = v\sin\left(u + \dfrac{\pi}{4}\right) \end{cases}, \quad \text{and } (H_2): \begin{cases} x(u,v) = u \\ y(u,v) = v\cos\left(u - \dfrac{\pi}{4}\right) \\ z(u,v) = v\sin\left(u - \dfrac{\pi}{4}\right) \end{cases},$$

both parametrized for $u \in \mathbb{R}$, $v \geq 0$.

Taking points from (H_1) or (H_2) such that $v > 0$, one gets the normal cones

$$\left\{(x,y,z) \in \mathbb{R}^3 \mid \frac{x-u}{-v} = \frac{y - v\cos\left(u \pm \frac{\pi}{4}\right)}{-\sin\left(u \pm \frac{\pi}{4}\right)} = \frac{z - v\sin\left(u \pm \frac{\pi}{4}\right)}{\cos\left(u \pm \frac{\pi}{4}\right)}\right\},$$

which tend when $(u, v) \to (0,0)$ to the lines

$$\left\{(0,y,y) \mid y \in \mathbb{R}\right\} \text{ and } \left\{(0,y,-y) \mid y \in \mathbb{R}\right\}.$$

When $(x_0, y_0, z_0) \in Q$ is such that $v = 0$, then

$$T_B(Q, (x_0, y_0, z_0)) = \left\{ (x, y, z) \in \mathbb{R}^3 \mid \begin{array}{c} -y \sin\left(u + \frac{\pi}{4}\right) + z \cos\left(u + \frac{\pi}{4}\right) \leq 0 \\ -y \sin\left(u - \frac{\pi}{4}\right) + z \cos\left(u - \frac{\pi}{4}\right) \geq 0 \\ y \geq 0 \end{array} \right\}$$

and

$$N_F(Q, (x_0, y_0, z_0)) = \left\{ (x, y, z) \in \mathbb{R}^3 \mid \begin{array}{c} x = 0 \\ y \cos\left(u + \frac{\pi}{4}\right) + z \sin\left(u + \frac{\pi}{4}\right) \leq 0 \\ y \cos\left(u - \frac{\pi}{4}\right) + z \sin\left(u - \frac{\pi}{4}\right) \geq 0 \\ y \leq 0 \end{array} \right\}.$$

Passing to the limit for $(u, v) \to (0, 0)$, one gets

$$N_M(Q, (0, 0, 0)) = \{(x, y, z) \in \mathbb{R}^3 \mid x = 0,\ y \leq 0,\ y \leq z \leq -y\}. \qquad \square$$

Problem 7.103. *Let $f, g : \mathbb{R}^p \to \mathbb{R}$ be Lipschitz around $\overline{x} \in \mathbb{R}^p$.*

(i) If $\partial_F(-f(\overline{x})g)(\overline{x}) \neq \emptyset$, then

$$\partial_F(f \cdot g)(\overline{x}) \subset \bigcap_{\xi \in \partial_F(-f(\overline{x})g)(\overline{x})} [\partial_F(g(\overline{x})f)(\overline{x}) - \xi],$$

which holds with equality if g is Fréchet differentiable at $\overline{x}$.

(ii) If $g(\overline{x}) \neq 0$, and if $\partial_F(f(\overline{x})g)(\overline{x}) \neq \emptyset$, then

$$\partial_F\left(\frac{f}{g}\right)(\overline{x}) \subset \bigcap_{\xi \in \partial_F(f(\overline{x})g)(\overline{x})} \frac{[\partial_F(g(\overline{x})f)(\overline{x}) - \xi]}{(g(\overline{x}))^2},$$

which holds with equality if g is Fréchet differentiable at $\overline{x}$.

Solution (i) Define $F : \mathbb{R}^p \to \mathbb{R}^2$ and $G : \mathbb{R}^2 \to \mathbb{R}$ by

$$F(x) := (f(x), g(x)) \quad \text{and} \quad G(y_1, y_2) := y_1 \cdot y_2.$$

Then $f \cdot g = G \circ F$, hence one can apply the chain rule from Theorem 5.2.36 to get that

$$\partial_F(f \cdot g)(\overline{x}) = \partial_F(G \circ F)(\overline{x}) = \partial_F \langle \xi, F \rangle (\overline{x}),$$

where $\xi \in \nabla G(\overline{y})$, since G is Fréchet differentiable at $\overline{y} := (f(\overline{x}), g(\overline{x}))$. Moreover, $\nabla G(\overline{y}) = (\overline{y}_2, \overline{y}_1)$, hence we have from above that

$$\partial_F(f \cdot g)(\overline{x}) = \partial_F(\overline{y}_2 f + \overline{y}_1 g)(\overline{x}) = \partial_F \left[g(\overline{x}) \cdot f - (-f(\overline{x})) \cdot g\right](\overline{x}),$$

which gives, using Theorem 5.2.33, the conclusion.

The proof of (ii) is similar. $\square$

Problem 7.104. *Let $f_1, \dots, f_k : \mathbb{R}^p \to \mathbb{R}$ be some functions, and $\overline{x} \in \mathbb{R}^p$. Then the function $g : \mathbb{R}^p \to \mathbb{R}$ given as*

$$g(x) := \min \left\{ f_1(x), \dots, f_k(x) \right\}, \ \forall x \in \mathbb{R}^p$$

satisfies

$$\partial_F g(\overline{x}) \subset \bigcap_{i \in A(\overline{x})} \partial_F f_i(\overline{x}),$$

where $A(\overline{x}) = \left\{ i \in \overline{1, k} \mid g(\overline{x}) = f_i(\overline{x}) \right\}$ denotes the set of active indices at $\overline{x}$.

Solution Take $\xi \in \partial_F g(\overline{x})$. Hence for any $\varepsilon > 0$, one can find $\delta > 0$ such that

$$\langle \xi, x - \overline{x} \rangle \leq g(x) - g(\overline{x}) + \varepsilon \left\| x - \overline{x} \right\|$$

whenever $\|x - \overline{x}\| < \delta$. For such x, for any $i \in A(\overline{x})$, one has

$$\begin{aligned} \langle \xi, x - \overline{x} \rangle &\leq g(x) - g(\overline{x}) + \varepsilon \left\| x - \overline{x} \right\| \\ &= g(x) - f_i(\overline{x}) + \varepsilon \left\| x - \overline{x} \right\| \\ &\leq f_i(x) - f_i(\overline{x}) + \varepsilon \left\| x - \overline{x} \right\|, \end{aligned}$$

which gives us that $\xi \in \partial_F f_i(\overline{x})$. □