Problem 1.

Consider the unconstrained optimisation problem

$$\min_{x,y} f(x, y),$$

where $f : \mathbb{R}^2 \to \mathbb{R}$ is defined as

$$f(x, y) = x^4 + 2x^2y + y^2 - 4xy - 4x + 2y.$$

a) Find all local and global minimisers of f.
(10 points)

b) Determine whether the function f is convex.
(5 points)

c) Perform one step of the gradient descent method with backtracking Armijo line search starting from the point $(x, y) = (0, 0)$. Use the parameters $c_1 = 0.1$ (sufficient decrease parameter), $\rho = 0.25$ (contraction factor), and $\alpha_0 = 1$ (initial step length).
(10 points)

Solution.

a) We start by computing the gradient of f. We have

$$\nabla f(x, y) = \begin{pmatrix} 4x^3 + 4xy - 4y - 4 \\ 2x^2 - 4x + 2y + 2 \end{pmatrix}.$$

The first order optimality condition $\nabla f(x, y) = 0$ thus becomes

$$\begin{aligned} 4x^3 + 4xy - 4y - 4 &= 0, \\ 2x^2 - 4x + 2y + 2 &= 0. \end{aligned}$$

From the second equation, we obtain that

$$y = -x^2 + 2x - 1.$$

Inserting this expression for y into the first equation yields

$$4x^3 - 4x^3 + 8x^2 - 4x + 4x^2 - 8x + 4 - 4 = 0,$$

which simplifies to

$$12x^2 - 12x = 0$$

with the solutions $x = 1$ and $x = 0$. For $x = 1$, we obtain that $y = 0$; for $x = 0$, we obtain that $y = -1$. We thus have the critical points $(x, y) = (1, 0)$ and $(x, y) = (0, -1)$.

Next we compute the Hessian of f. We have

$$H_f(x, y) = \begin{pmatrix} 12x^2 + 4y & 4x - 4 \\ 4x - 4 & 2 \end{pmatrix}.$$

In particular,

$$H_f(1, 0) = \begin{pmatrix} 12 & 0 \\ 0 & 2 \end{pmatrix},$$

which is obviously positive definite, and

$$H_f(0, -1) = \begin{pmatrix} -4 & -4 \\ -4 & 2 \end{pmatrix},$$

which cannot be positive semi-definite, as the upper left entry is negative.

This shows that $(x, y) = (1, 0)$ is the unique local minimiser of f.

Concerning globality of the minimiser, we note that

$$f(x, -x^2) = x^4 - 2x^4 + 4x^3 - 4x + x^4 - 2x^2 = 4x^3 - 2x^2 - 4x,$$

which tends to $-\infty$ as $x \to -\infty$. Thus f is unbounded below and therefore has no global minimiser.

b) In the first part of the problem, we have already seen that the Hessian of f is not always positive semi-definite, which implies that f cannot be convex.

c) We start by computing the gradient

$$\nabla f(0, 0) = \begin{pmatrix} -4 \\ 2 \end{pmatrix}.$$

The next iterate is therefore

$$(x_1, y_1) = (0, 0) - \alpha \nabla f(0, 0) = (4\alpha, -2\alpha)$$

for some step length $\alpha > 0$ determined via backtracking.

In order to determine α, we have to test whether the Armijo condition

$$f(4\alpha, -2\alpha) \leq f(0, 0) - c_1 \alpha \|\nabla f(0, 0)\|_2^2$$

is satisfied. Inserting the expression for f and using that $\|\nabla f(0,0)\|_2^2 = 20$ and $c_1 = 0.1$, this reads as

$$256\alpha^4 - 64\alpha^3 + 32\alpha^2 - 16\alpha + 4\alpha^2 - 4\alpha \leq -2\alpha.$$

For $\alpha = \alpha_0 = 1$, this clearly fails (on the left hand side we have 208; on the right hand side -2). Thus we reduce α by the contraction factor $\rho = 1/4$ and test with $\alpha = 1/4$ again. Here we obtain $-2\frac{3}{4}$ on the left hand side and $-\frac{1}{2}$ on the right hand side. Thus the Armijo condition holds and the step is accepted. The next iterate is therefore $(x_1, y_1) = (1, -1/2)$.

Problem 2.

Consider the constrained optimisation problem

$$\min_{x,y} y^2 - (x-1)^2 \qquad \text{s.t. } 4y \geq x^2.$$

Use the 2nd order optimality conditions for constrained optimisation to determine whether the point $(x, y) = (2, 1)$ is a local solution of this problem.
(15 points)

Solution.

Define the Lagrangian

$$\mathcal{L}(x, y; \lambda) = y^2 - (x-1)^2 - \lambda(4y - x^2).$$

The KKT conditions for the problem read

$$\begin{aligned} -2(x-1) &= -2\lambda x, \\ 2y &= 4\lambda, \\ 4y &\geq x^2, \\ \lambda &\geq 0, \\ \lambda(4y - x^2) &= 0. \end{aligned}$$

For $x = 2$ and $y = 1$, these equations and inequalities are satisfied with the Lagrange multiplier $\lambda = \frac{1}{2}$.

The Hessian of the Lagrangian is

$$H_{\mathcal{L}}(x, y; \lambda) = \begin{pmatrix} -2 & 0 \\ 0 & 2 \end{pmatrix} + \lambda \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}.$$

Specifically, we have that

$$H_{\mathcal{L}}(2,1;1/2) = \begin{pmatrix} -1 & 0 \\ 0 & 2 \end{pmatrix}.$$

Now denote $c(x,y) = 4y - x^2$. Then $\nabla c(x,y) = (-2x, 4)$, and specifically $\nabla c(2,1) = (-4,4)$. Thus the critical cone at $(2,1)$ is the set

$$C(2,1;1/2) = \left\{(p,q)^T \in \mathbb{R}^2 : \langle (p,q)^T, (-4,4)^T \rangle = 0\right\} = \left\{(p,p)^T \in \mathbb{R}^2 : p \in \mathbb{R}\right\}.$$

Let now $(p,p) \in C(2,1;1/2) \setminus \{(0,0)^T\}$. Then

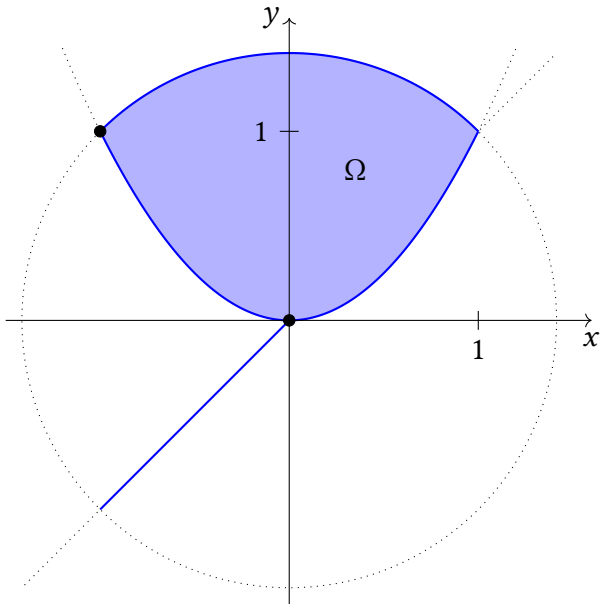
$$\langle (p,p)^T, H_{\mathcal{L}}(2,1;1/2)(p,p)^T \rangle = p^2 > 0.$$

Thus the second order sufficient optimality condition holds at $(2,1)$, which proves that this point is a (strict) local solution of the constrained problem.

Problem 3.

The set $\Omega \subset \mathbb{R}^2$ (see the sketch below) is given by the inequalities

$$(y - x^2)(x - y)^2 \geq 0 \qquad \text{and} \qquad 2 - x^2 - y^2 \geq 0.$$



a) Determine for each point in Ω whether the linear independence constraint qualification (LICQ) holds.
(5 points)

b) Determine the tangent cone $T_\Omega(x, y)$ and the cone of linearised feasible directions $\mathcal{F}(x, y)$ to the set Ω in the points $(x, y) = (-1, 1)$ and $(x, y) = (0, 0)$. (The points are marked in the sketch.)
(10 points)

Solution.

As a preparation, we define

$$c_1(x, y) = (y - x^2)(x - y)^2 \qquad \text{and} \qquad c_2(x, y) = 2 - x^2 - y^2$$

and compute

$$\nabla c_1(x, y) = \begin{pmatrix} -2x(x - y)^2 + 2(y - x^2)(x - y) \\ (x - y)^2 - 2(y - x^2)(x - y) \end{pmatrix} \qquad \text{and} \qquad \nabla c_2(x, y) = \begin{pmatrix} -2x \\ -2y \end{pmatrix}.$$

a) If no constraints are active, then LICQ holds trivially.

The constraint c_1 is active if either $x = y$ or $y = x^2$. We have that

$$\nabla c_1(x, y) = \begin{pmatrix} -2x(x-y)^2 + 2(y-x^2)(x-y) \\ (x-y)^2 - 2(y-x^2)(x-y) \end{pmatrix}.$$

If $x = y$, this is clearly equal to $(0, 0)$ and therefore linearly dependent. Thus LICQ fails in this case. If $y = x^2$, the second entry becomes $(x-y)^2$, which is different from zero unless we also have that $y = x$. Thus, in the cases where only c_1 is active, LICQ fails precisely when $x = y$.

The constraint c_2 is active whenever $x^2 + y^2 = 2$. In this case, at least one of the coordinates x or y is different from zero and thus also $\nabla c_2(x, y) \neq (0, 0)$. Thus LICQ holds whenever only c_2 is active.

Finally, we consider the case where both c_1 and c_2 are active. This happens for the points $(x, y) = (-1, 1)$, $(x, y) = (1, 1)$, and $(x, y) = (-1, -1)$. For $(x, y) = (-1, 1)$ we have

$$\nabla c_1(-1, 1) = \begin{pmatrix} 8 \\ 4 \end{pmatrix} \qquad \text{and} \qquad \nabla c_2(-1, 1) = \begin{pmatrix} 4 \\ -4 \end{pmatrix},$$

which are clearly linearly independent. Thus LICQ holds. For the points $(x, y) = (1, 1)$ and $(x, y) = (-1, -1)$ we have already seen that $\nabla c_1(1, 1) = (0, 0)$, and thus LICQ fails.

To summarise, LICQ fails at all points $(x, y) \in \Omega$ with $x = y$, and LICQ holds everywhere else in Ω.

b) At the point $(-1, 1)$, LICQ holds and thus

$$T_\Omega(-1, 1) = \mathcal{F}(-1, 1) = \left\{(p, q)^T \in \mathbb{R}^2 : 8p + 4q \geq 0 \text{ and } 4p - 4q \geq 0\right\}.$$

At the point $(0, 0)$ we have $\nabla c_1(0, 0) = (0, 0)$, and thus

$$\mathcal{F}(0, 0) = \mathbb{R}^2.$$

However, since LICQ fails, we cannot conclude that $T_\Omega(0, 0) = \mathcal{F}(0, 0)$. Instead, we see that, locally around $(0, 0)$, the set Ω is composed of two parts, namely the solid parabola

$$\Omega_1 = \left\{(x, y) \in \mathbb{R}^2 : y \geq x^2\right\}$$

and the line

$$\Omega_2 = \left\{(x, y) \in \mathbb{R}^2 : y = x\right\}.$$

Therefore,

$$T_\Omega(0, 0) = T_{\Omega_1}(0, 0) \cup T_{\Omega_2}(0, 0).$$

Since Ω_2 is a linear subspace, we have that

$$T_{\Omega_2}(0,0) = \Omega_2 = \{(p,q) \in \mathbb{R}^2 : p = q\}.$$

Moreover, the set Ω_1 is given by the inequality constraint

$$c_3(x,y) := y - x^2 \geq 0.$$

We compute $\nabla c_3(x,y) = (-2x, 1)$ for every $(x,y) \in \mathbb{R}^d$. Since $\nabla c_3(0,0) = (0,1)^T \neq 0$, LICQ holds for this constraint, and thus the tangent cone to Ω_1 at $(0,0)$ is equal to the cone of linearised feasible directions. That is, we have that

$$T_{\Omega_1}(0,0) = \{(p,q) \in \mathbb{R}^2 : q \geq 0\}.$$

Thus

$$T_{\Omega}(0,0) = T_{\Omega_1}(0,0) \cup T_{\Omega_2}(0,0) = \{(p,q) \in \mathbb{R}^2 : q \geq 0 \text{ or } p = q\}.$$

Problem 4.

Let $A \in \mathbb{R}^{d\times d}$ be symmetric and positive definite, let $c \in \mathbb{R}^d$, and let $b \in \mathbb{R}^d_{>0}$. We want to solve the constrained optimisation problem

$$\min_{x\in\mathbb{R}^d} \frac{1}{2}\langle x, Ax\rangle - \langle x, c\rangle \qquad \text{s.t. } 0 \leq x \leq b.$$

Here the inequalities $0 \leq x \leq b$ are to be interpreted componentwise. Propose an algorithm for the numerical solution of this problem under the assumption that the dimension d is large (e.g. $d \approx 10^4$). *(10 points)*

Solution.

In view of the large dimension of A, a reasonable choice could be a projected gradient method of the form

$$\tilde{x}_{k+1} = \min\{\max\{0, x_k - \alpha_k(Ax_k - c)\}, b\}$$

(that is, we compute a gradient descent step and project the result back to the feasible set) and

$$x_{k+1} = x_k + \theta_k(\tilde{x}_{k+1} - x_k).$$

The first step length $\alpha_k > 0$ may be determined using an Armijo condition and backtracking, although alternatives like Barzilai–Borwein step lengths (as in the project) may be preferable. The second step length $0 < \theta_k \leq 1$ can be determined by exact line search (and is not necessary if the Armijo condition was employed for α_k).

The main properties to take into consideration are the size of the problem (which makes active set methods unattractive) and the simplicity of the constraints (which favor gradient projection methods).

Problem 5.

Let $f\colon \mathbb{R}^d \to \mathbb{R}$ be continuously differentiable, let $x \in \mathbb{R}^d$, let $0 < c_1 < 1$, and let $p \in \mathbb{R}^d$ be a descent direction for f at the point x. Show that there exists some $\hat{\alpha} > 0$ such that the Armijo condition

$$f(x + \alpha p) \leq f(x) + c_1 \alpha \langle \nabla f(x), p \rangle$$

holds for all $0 < \alpha < \hat{\alpha}$.
(10 points)

Solution.

We recall that p is a descent direction for f at x, if $\langle \nabla f(x), p \rangle < 0$.

Assume now to the contrary that no such $\hat{\alpha} > 0$ exists. Then there exists a sequence $(\alpha_k)_{k\in\mathbb{N}}$ with $\alpha_k > 0$ for all k, $\lim_{k\to\infty} \alpha_k = 0$, and

$$f(x + \alpha_k p) > f(x) + c_1 \alpha_k \langle \nabla f(x), p \rangle$$

for all k. Dividing by $\alpha_k > 0$, we obtain that

$$c_1 \langle \nabla f(x), p \rangle < \frac{f(x + \alpha_k p) - f(x)}{\alpha_k}$$

for all $k \in \mathbb{N}$. Taking the limit $k \to \infty$, we conclude that

$$c_1 \langle \nabla f(x), p \rangle \leq \lim_{k\to\infty} \frac{f(x + \alpha_k p) - f(x)}{\alpha_k} = \langle \nabla f(x), p \rangle.$$

Thus

$$(1 - c_1)\langle \nabla f(x), p \rangle \geq 0$$

and thus, as $1 - c_1 > 0$, also

$$\langle \nabla f(x), p \rangle \geq 0,$$

which is a contradiction to the assumption that p is a descent direction for f at x.

Alternatively, if somewhat more error-prone because of the usage of the o-notation: A Taylor series expansion of f around x implies that

$$f(x + \alpha p) = f(x) + \alpha \langle \nabla f(x), p \rangle + o(\alpha)$$

as $\alpha \to 0$. Thus

$$f(x + \alpha p) - f(x) - c_1 \alpha \langle \nabla f(x), p \rangle = (1 - c_1)\alpha \langle \nabla f(x), p \rangle + o(\alpha)$$

as $\alpha \to 0$. Since p is by assumption a descent direction and $c_1 < 1$, the term

$$(1 - c_1)\langle \nabla f(x), p \rangle$$

is strictly negative. Dividing by α and taking the limit $\alpha \to 0$, we obtain that

$$\lim_{\alpha \to 0} \frac{1}{\alpha}\big(f(x + \alpha p) - f(x) - c_1 \alpha \langle \nabla f(x), p \rangle\big) = (1 - c_1)\langle \nabla f(x), p \rangle + \lim_{\alpha \to 0} \frac{o(\alpha)}{\alpha} = (1 - c_1)\langle \nabla f(x), p \rangle < 0.$$

As a consequence, the term on the left hand side needs to be strictly negative for all α sufficiently close to 0, which in turn implies that

$$f(x + \alpha p) - f(x) - c_1 \alpha \langle \nabla f(x), p \rangle < 0$$

for all sufficiently small $\alpha > 0$. This proves the claim.

Problem 6.

Let $f, g \colon \mathbb{R}^d \to \mathbb{R}$ be functions and let $\varphi \colon \mathbb{R} \to \mathbb{R}$ be a strictly increasing function. We consider the multicriteria optimisation problems

$$\min_{x \in \mathbb{R}^d} \big(f(x), g(x)\big) \tag{1}$$

and

$$\min_{x \in \mathbb{R}^d} \big(\varphi(f(x)), \varphi(g(x))\big). \tag{2}$$

Show that the problems (1) and (2) have precisely the same solutions.
(10 points)

Solution.

Assume that $x \in \mathbb{R}^d$ is *not* Pareto optimal for the problem (1). Then there exists $y \in \mathbb{R}^d$ such that $f(y) \leq f(x)$ and $g(y) \leq g(x)$ and at least one of those inequalities is strict. Without loss of generality, we may assume that $f(y) < f(x)$ (else we may re-name the functions). Because φ is strictly increasing, it follows that also $\varphi(f(y)) < \varphi(f(x))$ and $\varphi(g(y)) \leq \varphi(g(x))$, which shows that x is not Pareto optimal for the problem (2) either.

Conversely, assume that $x \in \mathbb{R}^d$ is *not* Pareto optimal for the problem (2). Then there exists $y \in \mathbb{R}^d$ such that $\varphi(f(y)) \leq \varphi(f(x))$ and $\varphi(g(y)) \leq \varphi(g(x))$ and at least one of those inequalities is strict. Again, we may assume without loss of generality that $\varphi(f(y)) < \varphi(f(x))$. Because φ is strictly increasing, it follows that also $f(y) < f(x)$ and $g(y) \leq g(x)$, which shows that x is not Pareto optimal for the problem (1) either.

Thus, the sets of points that are not Pareto optimal are the same for the problems (1) and (2), which implies that also the sets of Pareto optimal points are the same.

Problem 7.

Let $f, g\colon \mathbb{R}^d \to \mathbb{R}$ be convex functions. Define the function $h\colon \mathbb{R}^d \to \mathbb{R}$ by

$$h(x) := \inf_{z\in\mathbb{R}^d} \big(f(x-z) + g(z)\big).$$

a) Show that h is convex.
(5 points)

b) Show that the Legendre–Fenchel transform h^* of h satisfies

$$h^*(y) = f^*(y) + g^*(y)$$

for every $y \in \mathbb{R}^d$.
(5 points)

c) Assume now that $f = g$. Show that in this case we have that $h(x) = 2f(x/2)$ for every $x \in \mathbb{R}^d$. That is, for convex $f\colon \mathbb{R}^d \to \mathbb{R}$ we have for every $x \in \mathbb{R}^d$ that

$$\inf_{z\in\mathbb{R}^d} \big(f(x-z) + f(z)\big) = 2f\Big(\frac{x}{2}\Big).$$

(5 points)

Solution.

a) Let $x,\ y \in \mathbb{R}^d$, and let $0 < \lambda < 1$. Then

$$\begin{aligned}
\lambda h(x) + (1-\lambda)h(y) &= \lambda \inf_{z\in\mathbb{R}^d} \big(f(x-z) + g(z)\big) + (1-\lambda) \inf_{z'\in\mathbb{R}^d} \big(f(y-z') + g(z')\big) \\
&= \inf_{\substack{z\in\mathbb{R}^d \\ z'\in\mathbb{R}^d}} \big(\lambda f(x-z) + (1-\lambda) f(y-z') + \lambda g(z) + (1-\lambda) g(z')\big) \\
&\geq \inf_{\substack{z\in\mathbb{R}^d \\ z'\in\mathbb{R}^d}} \big(f(\lambda(x-z) + (1-\lambda)(y-z')) + g(\lambda z + (1-\lambda)z')\big) \\
&= \inf_{\substack{z\in\mathbb{R}^d \\ z'\in\mathbb{R}^d}} \big(f(\lambda x + (1-\lambda)y - (\lambda z + (1-\lambda)z')) + g(\lambda z + (1-\lambda)z')\big) \\
&= \inf_{\tilde{z}\in\mathbb{R}^d} \big(f(\lambda x + (1-\lambda)y - \tilde{z}) + g(\tilde{z})\big) \\
&= h(\lambda x + (1-\lambda)y).
\end{aligned}$$

Thus h is convex.

b) We have

$$\begin{aligned} h^*(y) &= \sup_{x\in\mathbb{R}^d} \big(\langle x, y\rangle - h(x)\big) \\ &= \sup_{x\in\mathbb{R}^d} \big(\langle x, y\rangle - \inf_{z\in\mathbb{R}^d} \big(f(x-z) + g(z)\big)\big) \\ &= \sup_{x\in\mathbb{R}^d} \sup_{z\in\mathbb{R}^d} \big(\langle x, y\rangle - f(x-z) - g(z)\big) \\ &= \sup_{x\in\mathbb{R}^d} \sup_{z\in\mathbb{R}^d} \big(\langle x-z, y\rangle - f(x-z) + \langle z, y\rangle - g(z)\big). \end{aligned}$$

Now we use the change of variables $(x-z, z) \mapsto (x', z)$, and obtain further that

$$h^*(y) = \sup_{x'\in\mathbb{R}^d} \sup_{z\in\mathbb{R}^d} \big(\langle x', y\rangle - f(x') + \langle z, y\rangle - g(z)\big).$$

The resulting expression contains two terms that only depend on x' but not on x, and two terms that only depend on z but not on x'. Thus we can split up the supremum and write

$$h^*(y) = \sup_{x'\in\mathbb{R}^d} \big(\langle x', y\rangle - f(x')\big) + \sup_{z\in\mathbb{R}^d} \big(\langle z, y\rangle - g(z)\big) = f^*(y) + g^*(y),$$

which was to show.

c) From part b) of the problem we have that

$$h^*(y) = 2f^*(y)$$

for every $y \in \mathbb{R}^d$. As a consequence (see the attached formula sheet),

$$h(x) = h^{**}(x) = (2f^*)^*(x) = 2f^{**}(x/2) = 2f(x/2).$$

Here we have used in the first and last equality that $f^{**} = f$ and $h^{**} = h$, since both f and h are convex functions.