**Problem 1.**

Consider the unconstrained optimisation problem

$$\min_{x,y} f(x, y), \tag{1}$$

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

$$f(x, y) = e^{y^2} + 2x^2 + xy + 4x - \frac{1}{2}y^2 + y.$$

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

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

c) Assume that we want to use the gradient descent method with backtracking Armijo line search for the minimisation of (1). Does the resulting iteration converge for all initialisations to a local or global solution? Briefly explain your reasoning.
*(5 points)*

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

**Solution.**

a) We start by computing the gradient of $f$: We have

$$\nabla f(x, y) = \begin{pmatrix} 4x + y + 4 \\ 2ye^{y^2} + x - y + 1 \end{pmatrix}.$$

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

$$\begin{aligned} 4x + y + 4 &= 0, \\ 2ye^{y^2} + x - y + 1 &= 0. \end{aligned}$$

From the first equation we get

$$x = -\frac{1}{4}y - 1,$$

which we can insert in the second equation. This results in

$$0 = 2ye^{y^2} - \frac{1}{4}y - 1 - y + 1 = 2ye^{y^2} - \frac{5}{4}y = y\left(2e^{y^2} - \frac{5}{4}\right).$$

Since $e^{y^2} \geq 1$ for all $y \in \mathbb{R}$, the term $2e^{y^2} - \frac{5}{4}$ is always positive. Thus the only solution of this equation is $y = 0$. Inserting this in the expression for $x$, we obtain $x = -1$. Thus the only critical point for the function $f$ is $(x, y) = (-1, 0)$.

The Hessian of $f$ is

$$H_f(x, y) = \begin{pmatrix} 4 & 1 \\ 1 & 4y^2e^{y^2} + 2e^{y^2} - 1 \end{pmatrix}.$$

Since

$$4y^2e^{y^2} + 2e^{y^2} - 1 \geq 2e^{y^2} - 1 \geq 1$$

for all $y \in \mathbb{R}$, both diagonal entries of the Hessian are always strictly positive, and the determinant is always at least 3 (and in particular positive). Thus the Hessian is always positive definite, which implies that $f$ is strictly convex. In particular, the unique critical point is also the unique global and local minimiser.

b) In the previous part, we have already concluded that $f$ is convex.

c) The function $f$ is obviously coercive,[1] and thus its level sets are bounded. Moreover, $f$ is obviously smooth. As shown in the lecture, this implies that the iterates $x_k$ of the gradient descent method, irrespective of the initialisation, satisfy $\nabla f(x_k) \to 0$ and that every accumulation point of the iterates is a critical point of $f$. Since $f$ has only a single critical point, which is at the same time the unique global minimum, we can conclude that the only possible accumulation point of the iterates $x_k$ is this global minimum of $f$. Now the coercivity of $f$ implies that the iterates stay bounded (by line search, the function values $f(x_k)$ cannot increase). Thus the sequence $(x_k)_{k\in\mathbb{N}}$ is bounded with at most a single accumulation point—which implies that the whole seqence converges to this accumulation point, which is the global minimum of $f$.

d) At the point $(0, 0)$ we have

$$\nabla f(0, 0) = \begin{pmatrix} 4 \\ 1 \end{pmatrix}.$$

[1] We can for instance estimate $f(x, y) = e^{y^2} + (x + 2)^2 + (x + y/2)^2 - 3y^2/4 + y - 4 \geq (x + 2)^2 + e^{y^2} - 3y^2/4 + y - 4$, where we see that both constituent parts $x \mapsto (x + 2)^2$ and $y \mapsto e^{y^2} - 3y^2/4 + y - 4$ are coercive functions of their respective variable. However, it is also not too difficult to prove the general result that a strictly convex function that attains its minimum is necessarily coercive. Try to do so yourself!

For the gradient descent method, we thus obtain the search direction $p = -\nabla f(0,0) = (-4,-1)^T$. Moreover,

$$\langle p, \nabla f(0,0)\rangle = -17$$

and $f(0,0) = 1$. The Armijo condition with $c_1 = 0.1$ thus requires that

$$f(-4\alpha, -\alpha) \leq 1 - \frac{17}{10}\alpha.$$

For $\alpha = \hat{\alpha} = 1$ we obtain

$$1 - \frac{17}{10}\alpha = -\frac{7}{10} < 0$$

and

$$f(-4,-1) = e^1 + 32 + 4 - 16 - \frac{1}{2} - 1 > 0.$$

(Concretely, $f(-4,-1) \approx 21.218$.) Thus Armijo's condition fails, and we need to decrease the step length to $\alpha = \rho\hat{\alpha} = \frac{1}{4}$.

Now we obtain

$$1 - \frac{17}{10}\alpha = 1 - \frac{17}{10}\frac{1}{4} = \frac{23}{40} > 0$$

and (using the estimate $e^{1/16} < 4^{1/16} = 2^{1/8} < 2$)

$$f(-1,-1/4) = e^{1/16} + 2 + \frac{1}{4} - 4 - \frac{1}{32} - \frac{1}{4} < -\frac{1}{32} < 0.$$

(Concretely, $f(-1,-1/4) \approx -0.967$.) Thus we see that Armijo's condition is satisfied and the step length $\alpha = \frac{1}{4}$ is accepted.

**Problem 2.**

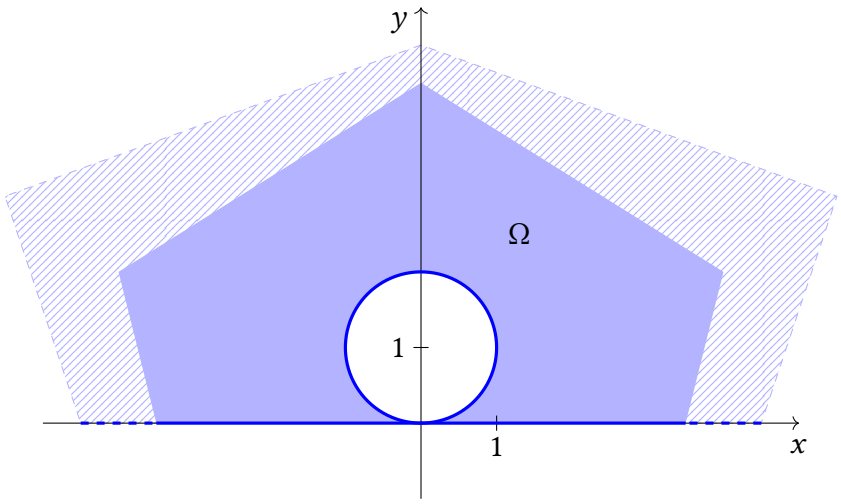
Consider the optimisation problem

$$\min_{x,y} 2x^2 - x^2 y - 3y^3 + 4y^2 + 2y$$

subject to the constraint $(x, y) \in \Omega$, where $\Omega \subset \mathbb{R}^2$ is given by the inequalities

$$x^2 + (y-1)^2 \geq 1 \quad \text{and} \quad y \geq 0.$$

Here is a sketch of the set $\Omega$ (the set continues infinitely in all directions of the upper half-plane):




a) Determine for each point in $\Omega$ whether the 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 $\Omega$ in the points $(x, y) = (0, 2)$ and $(x, y) = (0, 0)$.
*(10 points)*

c) Find all KKT points and all local minimisers for this optimisation problem. Use the second order necessary / sufficient conditions whenever applicable. Which of the local minimisers are global minimisers?
*(15 points)*

**Solution.**

As a preparation, we define

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

and

$$c_1(x, y) = x^2 + (y-1)^2 - 1 \qquad \text{and} \qquad c_2(x, y) = y,$$

and we compute

$$\nabla f(x, y) = \begin{pmatrix} 4x - 2xy \\ -x^2 - 9y^2 + 8y + 2 \end{pmatrix}, \qquad \nabla c_1(x, y) = \begin{pmatrix} 2x \\ 2(y-1) \end{pmatrix}, \qquad \nabla c_2(x, y) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

a) We see immediately that $\nabla c_2(x, y) \neq 0$ for all $x$, $y$. Moreover, the only point where $\nabla c_1(x, y) = 0$ is the point $(x, y) = (0, 1)$, which is not feasible. Thus the LICQ holds whenever none or only a single constraint is active.

It remains to check the points where both constraints are active. Here this is only the point $(x, y) = (0, 0)$, where $\nabla c_1(0, 0) = (0, -2)^T$ and $\nabla c_2(0, 0) = (0, 1)^T$, which are obviously linearly independent. Thus the LICQ holds everywhere apart from the point $(x, y) = (0, 0)$.

b) At $(0, 2)$, only the constraint $c_1$ is active and the LICQ holds. Moreover, $\nabla c_1(0, 2) = (0, 2)^T$. Thus

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

At $(0, 0)$, both constraints are active with $\nabla c_1(0, 0) = (0, -2)^T$ and $\nabla c_2(0, 0) = (0, 1)^T$. Thus

$$\mathcal{F}(0, 0) = \big\{(p, q) : -2q \geq 0 \text{ and } q \geq 0\big\} = \big\{(p, q) : q = 0\big\} = \mathbb{R}\begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

Since LICQ fails at $(0, 0)$, we can only infer that $T_\Omega(0, 0) \subset \mathcal{F}(0, 0) = \mathbb{R}(1, 0)^T$. However, as the whole $x$-axis is feasible, it follows readily that $\mathbb{R}(1, 0) \subset T_\Omega(0, 0)$, and thus

$$T_\Omega(0, 0) = \mathcal{F}(0, 0) = \mathbb{R}\begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

c) The KKT conditions for this problem are

$$\begin{aligned} 4x - 2xy &= 2\lambda x, \\ -x^2 - 9y^2 + 8y + 2 &= 2\lambda(y-1) + \mu, \\ x^2 + (y-1)^2 &\geq 1, \\ y &\geq 0, \\ \lambda &\geq 0, \\ \mu &\geq 0, \\ \lambda(x^2 + (y-1)^2 - 1) &= 0, \\ \mu y &= 0. \end{aligned}$$

For finding the KKT points of the problem, we consider all possible active sets:

- If no constraints are active, we have that $\lambda = \mu = 0$ and thus

$$4x - 2xy = 0 \qquad \text{and} \qquad -x^2 - 9y^2 + 8y + 2 = 0.$$

  For the first equation, we have two possibilities: Either $x = 0$, or $x \neq 0$, in which case we can divide by $x$ and obtain that $y = 2$.
  First first consider the case $x = 0$. Here the second equation reduces to

$$-9y^2 + 8y + 2 = 0,$$

  or

$$y^2 - \frac{8}{9}y - \frac{2}{9} = 0,$$

  which has the two solutions

$$y_\pm = \frac{4}{9} \pm \sqrt{\frac{16}{81} + \frac{2}{9}} = \frac{4 \pm \sqrt{34}}{9}.$$

  Since neither of the points $(0, y_\pm)$ is feasible, we have no KKT point in this case. (We see immediately that $y_- < 0$ and $0 < y_+ < 10/9 < 2$.)
  Now we consider the case $y = 2$. Here, the second equation reduces to

$$x^2 = -9y^2 + 8y + 2 = -36 + 16 + 2 = -18,$$

  which is clearly not solvable in $\mathbb{R}$. Again, we have no KKT point.

- Now we consider the case where the second constraint is active, that is, $y = 0$, and the first Lagrange parameter is 0, that is, $\lambda = 0$. Here we obtain the equations

$$4x - 2xy = 0 \qquad \text{and} \qquad -x^2 - 9y^2 + 8y^2 + 2 = \mu.$$

  Inserting $y = 0$ into the first equation, we obtain that $x = 0$. Then, the second equation reduces to $\mu = 2$. Since $(x, y) = (0, 0)$ is feasible and $\mu = 2 \geq 0$, we can conclude that $(0, 0)$ is a KKT point with Lagrange multiplier $(0, 2)$.
  It happens that the point $(0, 0)$ is at the same time the only point where both constraints are active. Since we have already concluded that this point is a KKT point, we need not consider it again in the future considerations.

- We are left with the case where the first constraint is active, that is, $x^2 + (y - 1)^2 = 1$, and $\mu = 0$. Here we obtain the equations

$$4x - 2xy = 2\lambda x \qquad \text{and} \qquad -x^2 - 9y^2 + 8y + 2 = 2\lambda(y - 1).$$

For the first equation, we have the two possibilities $x = 0$ or $x \neq 0$, in which case we can divide by $x$ and obtain that $2\lambda = 4 - 2y$.

We consider first the case $x = 0$. Since $x^2 + (y-1)^2 = 1$, it follows that either $y = 0$ or $y = 2$. The first possibility leads us to $(x, y) = (0, 0)$, which we have already discussed. For $y = 2$, the second equation simplifies to

$$2\lambda = -9y^2 + 8y + 2 = -18,$$

which implies $\lambda = -9 < 0$. Thus this yields no KKT point.

Now consider the case $x \neq 0$ and thus $2\lambda = 4 - 2y$. Inserting this value for $\lambda$ into the second equation, we obtain that

$$-x^2 - 9y^2 + 8y + 2 = (4 - 2y)(y - 1).$$

Since $c_1$ is active, we have that

$$-x^2 = (y-1)^2 - 1 = y^2 - 2y.$$

Thus we can simplify the equation for $y$ to

$$y^2 - 2y - 9y^2 + 8y + 2 = 6y - 2y^2 - 4.$$

This further simplifies to

$$-6y^2 = -6 \qquad \text{or} \qquad y = \pm 1.$$

Taking into account the inequality $y \geq 0$, we obtain that $y = 1$ and $\lambda = 2 - y = 1$. The equation $x^2 + (y-1)^2 = 1$ finally yields the corresponding values $x = \pm 1$ for $x$. We thus obtain the two KKT points $(x, y) = (1, 1)$ and $(x, y) = (-1, 1)$, both with Lagrange multiplier $(\lambda, \mu) = (1, 0)$.

Thus we have the following KKT points:

- $(0, 0)$ with Lagrange multiplier $(0, 2)$.
- $(1, 1)$ with Lagrange multiplier $(1, 0)$.
- $(-1, 1)$ with Lagrange multiplier $(1, 0)$.

Now we will try to use the second order conditions to determine whether these points are actually local minimisers. For this, we need to compute the Hessian of the Lagrangian, which is

$$H_{\mathcal{L}}(x, y; \lambda, \mu) = \begin{pmatrix} 4 - 2y & -2x \\ -2x & -18y + 8 \end{pmatrix} - \lambda \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}.$$

At the point $(0,0)$ with Lagrange multiplier $(0,2)$ we have that

$$H_{\mathcal{L}}(0,0;0,2) = \begin{pmatrix} 4 & 0 \\ 0 & 8 \end{pmatrix},$$

which is positive definite. Thus the second order sufficient optimality conditions are satisfied, and $(0,0)$ is a strict local solution. (Recall here that the second order sufficient conditions do not require LICQ to hold!)

At the point $(1,1)$ with Lagrange multiplier $(1,0)$ we have that

$$H_{\mathcal{L}}(1,1;1,0) = \begin{pmatrix} 0 & -2 \\ -2 & -12 \end{pmatrix},$$

which is an indefinite matrix. In order to decide whether $(1,1)$ is a local solution or not, we thus need to compute the critical cone $C(1,1;1,0)$. Since only the first constraint is active and the Lagrange multiplier for this constraint is strictly positive, we have that

$$\begin{aligned} C(1,1;1,0) &= \left\{(p,q) \in \mathbb{R}^2 : \langle (p,q)^T, \nabla c_1(1,1)\rangle = 0\right\} \\ &= \left\{(p,q) \in \mathbb{R}^2 : \langle (p,q)^T, (2,0)^T\rangle = 0\right\} = \left\{(p,q) \in \mathbb{R}^2 : p = 0\right\} = \mathbb{R}\begin{pmatrix} 0 \\ 1 \end{pmatrix}. \end{aligned}$$

Since

$$\langle (0,q)^T, H_{\mathcal{L}}(1,1;1,0)(0,q)^T\rangle = -12q^2 < 0$$

for all $q \neq 0$, the second order necessary condition at this point fails and thus $(1,1)$ is no local solution.

For the point $(-1,1)$ with Lagrange multiplier $(1,0)$, the situation is similar: We have

$$H_{\mathcal{L}}(-1,1;1,0) = \begin{pmatrix} 0 & 2 \\ 2 & -12 \end{pmatrix},$$

and again

$$C(-1,1;1,0) = \left\{(p,q) \in \mathbb{R}^2 : \langle (p,q)^T, \nabla c_1(-1,1)\rangle = 0\right\} = \mathbb{R}\begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Again we have that

$$\langle (0,q)^T, H_{\mathcal{L}}(1,1;1,0)(0,q)^T\rangle = -12q^2 < 0$$

for all $q \neq 0$, and thus the second order necessary conditions fail. Thus also $(-1,1)$ is no local solution.

To summarise, we have the following KKT points and local minimisers:

- The KKT point and local minimiser $(0,0)$.
- The KKT points $(1,1)$ and $(-1,1)$, which are not local minimisers.

Since $\lim_{y\to\infty} f(0,y) = -\infty$ and the points $(0,y)$ with $y \geq 2$ are all feasible, the function $f$ is unbounded below on the feasible set, and thus no global solutions exist.

**Problem 3.**

We consider the constrained optimisation problem

$$\min_x f(x) \qquad \text{s.t. } c(x) \geq 0, \tag{2}$$

where $f\colon \mathbb{R}^d \to \mathbb{R}$ is twice continuously differentiable, and $c\colon \mathbb{R}^d \to \mathbb{R}$ is concave and once continuously differentiable. Assume that $x^*$ is a KKT point of (2) and that the Hessian $H_f(x^*)$ is positive definite. Show that $x^*$ is a local solution of (2).
*(10 points)*

*Hint: Use a second order Taylor series expansion of $f$! Be aware, though, that the function $c$ is* not *assumed to be twice differentiable, so the Hessian $H_c$ cannot be assumed to exist.*

**Solution.**

Since $x^*$ is a KKT point, there exists $\lambda^* \geq 0$ such that $\nabla f(x^*) = \lambda^* \nabla c(x^*)$. Moreover, the complementarity condition $\lambda^* c(x^*) = 0$ holds. A second order Taylor series expansion of $f$ around $x^*$ now yields that

$$\begin{aligned} f(y) &= f(x^*) + \langle \nabla f(x^*), y - x^* \rangle + \frac{1}{2}\langle y - x^*, H_f(x^*)(y - x^*)\rangle + o(\|y - x^*\|^2) \\ &= f(x^*) + \lambda^* \langle \nabla c(x^*), y - x^* \rangle + \frac{1}{2}\langle y - x^*, H_f(x^*)(y - x^*)\rangle + o(\|y - x^*\|^2) \end{aligned}$$

as $y \to x^*$. Now the concavity of $c$ implies that

$$c(y) \leq c(x^*) + \langle \nabla c(x^*), y - x^* \rangle$$

for all $y$. Since $\lambda^* \geq 0$ and $\lambda^* c(x^*) = 0$, we obtain for all $y$ satisfying $c(y) \geq 0$ that

$$\lambda^* \langle \nabla c(x^*), y - x^* \rangle \geq \lambda^* (c(y) - c(x^*)) = \lambda^* c(y) \geq 0.$$

Thus it follows that

$$f(y) \geq f(x^*) + \frac{1}{2}\langle y - x^*, H_f(x^*)(y - x^*)\rangle + o(\|y - x^*\|^2)$$

for all $y$ satisfying $c(y) \geq 0$. The positive definiteness of $H_f(x^*)$ now implies that there exists $C > 0$ such that

$$\frac{1}{2}\langle y - x^*, H_f(x^*)(y - x^*)\rangle \geq C\|y - x^*\|^2$$

for all $y$. Thus we obtain that

$$f(y) \geq f(x^*) + C\|y - x^*\|^2 + o(\|y - x^*\|^2)$$

for all $y$ satisfying $c(y) \geq 0$. For $y$ sufficiently close to $x^*$, the term $C\|y - x^*\|^2$ dominates the term $o(\|y - x^*\|^2)$. As a consequence, $f(y) > f(x^*)$ for all $y \neq x^*$ that are sufficiently close to $x^*$ and satisfy $c(y) \geq 0$, which proves that $x^*$ is a local solution of (2).

**Problem 4.**

Assume that we want to solve an optimisation problem $\min_x f(x)$, where $f\colon \mathbb{R}^d \to \mathbb{R}$ with a (nonlinear) CG method with *exact line search.* Show that the Polak–Ribière method and the Hestenes–Stiefel method yield identical steps, if they are both initialised with $p_0 = -\nabla f(x_0)$.
*(10 points)*

**Solution.**

We show by induction over $k$ that both methods yield identical steps $p_k$ and thus also identical iterates $x_{k+1}$. For $k = 0$ this is trivial, as $p_0 = -\nabla f(x_0)$ in both cases.

Now assume that both methods yield identical steps up to iteration $k \geq 0$. Denote moreover

$$\beta_{k+1}^{\mathrm{PR}} := \frac{\langle \nabla f(x_{k+1}), \nabla f(x_{k+1}) - \nabla f(x_k) \rangle}{\|\nabla f(x_k)\|^2}$$

and

$$\beta_{k+1}^{\mathrm{HS}} := \frac{\langle \nabla f(x_{k+1}), \nabla f(x_{k+1}) - \nabla f(x_k) \rangle}{\langle \nabla f(x_{k+1}) - \nabla f(x_k), p_k \rangle}.$$

In order to show that the search direction $p_{k+1}$ is the same for both methods, we have to show that $\beta_{k+1}^{\mathrm{PR}} = \beta_{k+1}^{\mathrm{HS}}$, which is the same as showing that

$$\|\nabla f(x_k)\|^2 = \langle \nabla f(x_{k+1}) - \nabla f(x_k), p_k \rangle.$$

We now recall that we are using an exact line search for the computation of the step length $\alpha_k$ in each step. That is, $\alpha_k$ is a solution of the one-dimensional optimisation problem

$$\min_{\alpha > 0} f(x_k + \alpha p_k),$$

and thus the first order optimality condition

$$\langle \nabla f(x_k + \alpha_k p_k), p_k \rangle = 0$$

is satisfied. Since $x_{k+1} = x_k + \alpha_k p_k$, we can rewrite this as

$$\langle \nabla f(x_{k+1}), p_k \rangle = 0$$

for all $k$. As a consequence,

$$\langle \nabla f(x_{k+1}) - \nabla f(x_k), p_k \rangle = \langle \nabla f(x_{k+1}), p_k \rangle - \langle \nabla f(x_k), p_k \rangle = -\langle \nabla f(x_k), p_k \rangle.$$

Now, for $k = 0$, it follows that $p_0 = -\nabla f(x_0)$, and thus

$$\langle \nabla f(x_1) - \nabla f(x_0), p_0 \rangle = -\langle \nabla f(x_0), -\nabla f(x_0) \rangle = \|\nabla f(x_0)\|^2.$$

Moreover, for $k \geq 1$ we have that

$$p_k = -\nabla f(x_k) + \beta_k p_{k-1},$$

where $\beta_k$ is either equal to $\beta_k^{\mathrm{PR}}$ or $\beta_k^{\mathrm{HS}}$ (both of which are anyway equal according to the induction hypothesis). In any case, we obtain by using the equalities $\langle \nabla f(x_{k+1}), p_k \rangle = 0$ and $\langle \nabla f(x_k), p_{k-1} \rangle = 0$ that

$$\begin{aligned}\langle \nabla f(x_{k+1}) - \nabla f(x_k), p_k \rangle = -\langle \nabla f(x_k), p_k \rangle &= -\langle \nabla f(x_k), -\nabla f(x_k) + \beta_k p_{k-1} \rangle \\ &= \|\nabla f(x_k)\|^2 - \beta_k \langle \nabla f(x_k), p_{k-1} \rangle = \|\nabla f(x_k)\|^2.\end{aligned}$$

This shows that $\beta_{k+1}^{\mathrm{PR}} = \beta_{k+1}^{\mathrm{HS}}$, which proves the claim.

**Problem 5.**

Define the function $f\colon \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}$,

$$f(x, y) = \sum_{i=1}^{d} \exp(x_i),$$

let $A \in \mathbb{R}^{m \times d}$ and $b \in \mathbb{R}^m$. We consider the constrained optimisation problem

$$\min_{(x,y) \in \mathbb{R}^d \times \mathbb{R}} f(x, y) \qquad \text{s.t. } Ax \leq yb. \tag{3}$$

*(Here the variable $y \in \mathbb{R}$ only appears in the constraint; the function $f$ only depends on $x$.)*

a) Compute the Lagrangian dual of the optimisation problem (3).
*(10 points)*

b) Show that the dual problem has a solution for all $A \in \mathbb{R}^{m \times d}$ and $b \in \mathbb{R}^m$.
*(5 points)*

c) Find an example of a matrix $A \in \mathbb{R}^{m \times d}$ and a vector $b \in \mathbb{R}^m$ such that the primal problem has no solution. Find a solution of the dual problem in your example.
*(5 points)*

**Solution.**

a) The Lagrangian of (3) is the function $\mathcal{L}\colon \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^m \to \mathbb{R}$,

$$\mathcal{L}(x, y; \lambda) = \sum_{i=1}^{d} \exp(x_i) - \langle yb - Ax, \lambda \rangle.$$

Thus the dual objective function is

$$\begin{aligned} q(\lambda) &:= \min_{(x,y)\in\mathbb{R}^d\times\mathbb{R}} \Big( \sum_{i=1}^{d} \exp(x_i) - \langle yb - Ax, \lambda \rangle \Big) \\ &= \min_{x\in\mathbb{R}^d} \Big( \sum_{i=1}^{d} \exp(x_i) + \langle A^T\lambda, x \rangle \Big) + \min_{y\in\mathbb{R}} \big( -y\langle b, \lambda \rangle \big) \\ &= \sum_{i=1}^{d} \min_{x_i\in\mathbb{R}} \big( \exp(x_i) + (A^T\lambda)_i \big) - \max_{y\in\mathbb{R}} y\langle b, \lambda \rangle. \end{aligned}$$

Defining $g\colon \mathbb{R} \to \mathbb{R}$, $g(t) = \exp(t)$, we note that (using the attached formula sheet)

$$\begin{aligned} \min_{x_i\in\mathbb{R}} \big( \exp(x_i) + (A^T\lambda)_i \big) &= -\max_{x_i\in\mathbb{R}} \big( -(A^T\lambda)_i - g(x_i) \big) = -g^*(-(A^T\lambda)_i) \\ &= \begin{cases} (A^T\lambda)_i - (A^T\lambda)_i \log(-(A^T\lambda)_i) & \text{if } (A^T\lambda)_i \le 0, \\ -\infty & \text{if } (A^T\lambda)_i > 0. \end{cases} \end{aligned}$$

Here we set for simplicity $t - t\log t := 0$ for $t = 0$. Moreover, we have that

$$-\max_{y} y\langle b, \lambda \rangle = \begin{cases} 0 & \text{if } \langle b, \lambda \rangle = 0, \\ -\infty & \text{if } \langle b, \lambda \rangle \neq 0. \end{cases}$$

Thus the dual problem reads as

$$\max_{\lambda\in\mathbb{R}^m} \sum_{i=1}^{d} \big( (A^T\lambda)_i - (A^T\lambda)_i \log(-(A^T\lambda)_i) \big) \qquad \text{s.t.} \begin{cases} A^T\lambda \le 0, \\ \langle \lambda, b \rangle = 0, \\ \lambda \ge 0. \end{cases}$$

Here the constraint $\lambda \ge 0$ stems from the fact that we have started with the inequality constraint $Ax \le yb$.

b) The primal objective function is non-negative, and thus bounded below. Moreover, we only have linear inequality constraints, and the feasible set is non-empty (e.g. the point $(x, y) = (0, 0)$ is feasible irrespective of the definitions of $A$ and $b$). Thus Slater's constraint qualification holds, which further implies that strong duality is applicable. In particular, this implies that the dual problem admits a solution.

c) They maybe simplest examples here is choosing $A = 0$ and $b = 0$, in which case we have no effective constraints and the problem simply reads

$$\min_{(x,y)\in\mathbb{R}^d\times\mathbb{R}} \sum_{i=1}^{d} \exp(x_i).$$

Since the exponential function is always positive but $\lim_{t\to-\infty} \exp(t) = 0$, this problem has no solution. For this example, the dual problem reduces to

$$\max_{\lambda\in\mathbb{R}^m} 0 \qquad \text{s.t. } \lambda \geq 0.$$

Every vector $\lambda \geq 0$ solves this trivial problem.