Problem 1

In this problem, you will evaluate several independent statements about minimisers of functions. For each of the statements, mark them as True or False.

a) If a function $f: \mathbb{R}^n \to \mathbb{R}$ is C^2 and convex, then f has a minimiser $x^* \in \mathbb{R}^n$.

Solution: False
Many counterexamples, for example $f(x) = a^T x$ for a nonzero vector a and $f(x) = e^{x_1}$.

b) If f is C^2 and a point x^* is such that $\nabla f(x^*) = 0$ and $\nabla^2 f(x^*)$ is symmetric positive definite, then x^* is a local minimiser of f.

Solution: True
This is (a weaker version of) Theorem 2.4 in N&W.

c) If a point x^* is a local minimiser of a constrained optimisation problem with continuously differentiable objective function f and constraint functions c_i (which could be equality constraints, inequality constraints, or both), then x^* is a KKT point.

Solution: False
Some extra condition on the constraints, such as LICQ is needed. A counterexample is given by the optimisation problem

$$\min_{(x,y)} x \quad \text{s.t.} \quad \begin{cases} y = 0, \\ y - x^2 = 0, \end{cases}$$

which has a global minimiser in $(0,0)$, but $(0,0)$ is not a KKT point.

d) If a constrained optimisation problem has c^2 objective function f and constraint functions c_i, a point x^* is a KKT point of the optimisation problem, with Lagrange multiplier λ^*, and

$$w^t \nabla^2_{xx} \updownarrow (x^*, \lambda^*) w > 0,$$

holds for all nonzero w in the critical cone at x^*, then x^* is a local minimiser.

Solution: True
This is Theorem 12.6 in N&W.

Problem 2

For each of the optimisation problems below, decide whether they are convex problems or not.

a)

$$\min_{(x,y)} (x-1)^4 + y^4 - x^2 + y$$

Solution: Not convex.
For example, at $(1,0)$, the Hessian is $\begin{bmatrix} -2 & 0 \\ 0 & 0 \end{bmatrix}$, which is not positive semidefinite.

b)

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} x^T A x - b^T x$$

where A is an $n \times n$ symmetric positive definite matrix and b is a vector in $\mathbb{R}^n$.
Solution: Convex.

c)

$$\min_{(x,y)} x^2 - y^2 \quad \text{s.t.} \quad x - 2y = 0.$$

Solution: Convex.
This is a tricky one. The function $x^2 - y^2$ is not convex on $\mathbb{R}^2$, however, on the line $x - 2y = 0$, convexity holds.

d)

$$\min_{(x,y)} x^2 + y^2 \quad \text{s.t.} \quad y \geq \sin x$$

Solution: Not convex.
The domain defined by $y \geq \sin x$ is not convex.

e)

$$\min_{(x,y)} x^2 - 2x + x^2 y^2 - 2xy$$

Solution: Not convex.
At the origin, the Hessian is $\begin{bmatrix} 2 & -2 \\ -2 & 0 \end{bmatrix}$, which is not positive semidefinite.

Problem 3 Consider the constrained optimisation problem:

$$\min_{(x,y)} y^3 - \frac{3}{2}y^2 + \frac{2}{3}y - x^3 \quad \text{s.t.} \quad \begin{cases} y \geq 0, \\ 1 - y - x^2 \geq 0. \end{cases}$$

Identify all KKT points and among the KKT points, the points that also satisfy the second order necessary conditions for optimality.

Solution: The gradient of the objective function $f(x, y) = y^3 - \frac{3}{2}y^2 + \frac{2}{3}y - x^3$ is:

$$\nabla f(x, y) = \begin{bmatrix} -3x^2 \\ 3y^2 - 3y + \frac{2}{3} \end{bmatrix},$$

and the gradients of the constraints $c_1(x, y) = y$, $c_2(x, y) = 1 - y - x^2$ are:

$$\nabla c_1(x, y) = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad \nabla c_2(x, y) = \begin{bmatrix} -2x \\ -1 \end{bmatrix}.$$

We can now consider the given points:

- $(0, 0)$: $c_1(0, 0) = 0$ $c_2(0, 0) = 1 > 0$, so it is a feasible point, c_1 is active, c_2 is not.

$$\nabla f(0, 0) = \begin{bmatrix} 0 \\ \frac{2}{3} \end{bmatrix}, \quad \nabla c_1(0, 0) = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

  We have $\nabla f(0, 0) = \frac{3}{2}\nabla c_1(0, 0)$, so $(0, 0)$ **is** a KKT point.

- $\left(0, \frac{1}{3}\right)$. Feasible point, none of the constraints are active. $\nabla f(0, \frac{1}{3}) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, so $\left(0, \frac{1}{3}\right)$ **is** a KKT point.

- $\left(0, \frac{2}{3}\right)$. Feasible point, none of the constraints are active. $\nabla f(0, \frac{2}{3}) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, so $\left(0, \frac{2}{3}\right)$ **is** a KKT point.

- $(0, 1)$. $c_1(0, 1) = 1 > 0, c_2(0, 1) = 0$. Feasible point, c_1 is inactive, c_2 is active.

$$\nabla f(0, 1) = \begin{bmatrix} 0 \\ \frac{2}{3} \end{bmatrix}, \qquad \nabla c_2(0, 1) = \begin{bmatrix} 0 \\ -1 \end{bmatrix}$$

  The solution of $\nabla f(0, 1) = \lambda_2 \nabla c_2(0, 1)$ is $\lambda_2 = -\frac{2}{3}$. Since $\lambda_2 < 0$, $(0, 1)$ **is not** a KKT point.

- $(-1,0)$. Here both c_1 and c_2 are active. We have

$$\nabla f(-1,0) = \begin{bmatrix} -3 \\ \frac{2}{3} \end{bmatrix} \quad \nabla c_1(-1,0) = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad \nabla c_2(-1,0) = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$$

  $\nabla f(-1,0) = \lambda_1 \nabla c_1(-1,0) + \lambda_2 \nabla c_2(-1,0)$ is a linear equation in λ_1, λ_2 with solution

$$\lambda_1 = -\frac{5}{6}, \lambda_2 = -\frac{3}{2}.$$

  The Lagrange multipliers are negative, so $(-1,0)$ **is not** a KKT point.

- $(1,0)$. Here both c_1 and c_2 are active. We have

$$\nabla f(1,0) = \begin{bmatrix} -3 \\ \frac{2}{3} \end{bmatrix} \quad \nabla c_1(1,0) = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad \nabla c_2(1,0) = \begin{bmatrix} -2 \\ -1 \end{bmatrix}$$

  $\nabla f(1,0) = \lambda_1 \nabla c_1(1,0) + \lambda_2 \nabla c_2(1,0)$
  is a linear equation in λ_1, λ_2 with solution

$$\lambda_1 = \frac{13}{6}, \lambda_2 = \frac{3}{2}.$$

  The Lagrange multipliers are positive, so $(-1,0)$ **is** a KKT point.

For the second part, we only consider the points that are KKT points and check if they also satisfy the second order *necessary* condtitions, that is $w^T \nabla^2 L(x^*, y^*, \lambda^*) w \geq 0$ for all w in the critical cone.

- $(0,0)$. The critical cone is $\mathcal{C}(0,0) = \left\{ \begin{bmatrix} p \\ 0 \end{bmatrix}, p \in \mathbb{R} \right\}$. The Hessian of the Lagrangian is

$$\nabla^2 f(0,0) - \frac{3}{2} \nabla^2 c(0,0) = \nabla^2 f(0,0) = \begin{bmatrix} 0 & 0 \\ 0 & -3 \end{bmatrix}$$

  We now see that

$$\begin{bmatrix} p & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & -3 \end{bmatrix} \begin{bmatrix} p \\ 0 \end{bmatrix} = 0$$

  So $(0,0)$ **satisfies** the necessary conditions for second order minima. (But not the sufficient conditions.)

- $\left(0, \frac{1}{3}\right)$ No active constraints, so the critical cone is all of $\mathbb{R}^2$. The Hessian of f is

$$\nabla^2 f(0, \frac{1}{3}) = \begin{bmatrix} 0 & 0 \\ 0 & -1 \end{bmatrix}$$

which is not positive semidefinite. $\left(0, \frac{1}{3}\right)$ **does not satisfy** the second order necessary conditions.

- $\left(0, \frac{2}{3}\right)$ No active constraints, so the critical cone is all of $\mathbb{R}^2$. The Hessian of f is

$$\nabla^2 f(0, \frac{2}{3}) = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

which is not positive semidefinite. $\left(0, \frac{2}{3}\right)$ **satisfies** the second order necessary conditions.

- $(1, 0)$ Both c_1 and c_2 are active and $\lambda_1 > 0, \lambda_2 > 0$. The critical cone is given by $\nabla c_1(1,0)^T w = 0, \nabla c_2(1,0)^T w = 0$ or

$$\begin{bmatrix} 0 & 1 \\ -2 & -1 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = 0.$$

The only solution is $w = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$. The critical cone is only the origin and

$$w^T \nabla^2 \mathcal{L}(1, 0, \lambda^*) w \geq 0$$

holds trivially when $w = 0$.

$(1, 0)$ **satisfies** the second order necessary conditions.

Problem 4 Imagine you wanted to solve the following optimisation problem numerically

$$\min_{x \in \mathbb{R}^n} \frac{\omega^2}{2} \sum_{i=1}^{n} x_i^2 + \sum_{i=1}^{n-1} \exp(x_{i+1} - x_i)$$

where ω is a non-zero real number and $n = 1\ 000\ 000$.

Explain how you would approach this problem.

That is: explain what method you would use, and any details of the implementation you consider critical.

Justify your choices.

Solution: The problem is an unconstrained optimisation problem. Furthermore, each of the functions

$$\frac{\omega^2}{2}x_i^2 \quad \text{and} \quad \exp(x_{i+1} - x_i)$$

are convex, so the problem is even convex.

The problem is that $n = 10^6$. This means that it is not feasible to store full $n \times n$ (nor to do matrix-vector products and much less to solve linear systems with full matrices.)

Line search methods are possible, but we have to be careful about how we find search directions. Due to the large dimension, BFGS is not possible, nor is a naive implementation of Newton's method.

It is possible to use gradient descent or conjugate gradient methods such as Fletcher–Reeves for search directions, but these methods are going to be slow (especially gradient descent.)[1]

A faster method for this optimisation problem is limited-memory BFGS, but the best algorithm is probably Newton's method where we take advantage of the structure of the problem: $\frac{\partial^2 f(x)}{\partial x_i \partial x_j}$ is only nonzero when $j = i-1, j = i$ or $j = i+1$. Therefore the Hessian is tridiagonal, and we can use an efficient, $\mathcal{O}(n)$, solver for tri-diagonal systems, such as the Thomas algorithm, to solve

$$\mathcal{H}f_k p = -\nabla f_k$$

So we can use a line search method with Newton's method to solve the optimisation problem. Safeguards to ensure we get descent directions are not needed since the problem is convex. For the line search we can use Wolfe condtions, strong Wolfe conditions or Armijo backtracking, (or other conditions) Important details that have to be observed to obtain quadratic convergence is that $\alpha = 1$ should be used if feasible, and that $c_1 \leq \frac{1}{2}$ in the Armijo condition.

Problem 5

Consider the unconstrained optimization problem:

$$\min_{(x,y)} x^2 - 2x + x^2y^2 - 2xy$$

[1] Nevertheless, suggesting Fletcher–Reeves with strong Wolfe conditions and $c_2 < \frac{1}{2}$ would score full points here.

a) Determine all stationary points of the objective function and determine if they are local minima.

Solution:

$$\nabla f(x,y) = \begin{bmatrix} 2x - 2 + 2xy^2 - 2y \\ 2x^2y - 2x \end{bmatrix}$$

Setting $\frac{\partial f}{\partial y}(x,y) = 0$, we get

$$x^2y - x = 0 \Rightarrow x = 0 \quad \text{or} \quad xy = 1$$

Inserting $x = 0$ into $\frac{\partial f}{\partial x}(x,y) = 0$, we get $y = -1$.

$xy = 1$ makes $\frac{\partial f}{\partial y}(x,y) = 0$ simplify to

$$2x - 2 = 0$$

with solution $x = 1$, $xy = 1$ then implies $y = 1$.

In conclusion, the stationary points are $(0,-1)$ and $(1,1)$.

The Hessian of f is

$$\mathcal{H}f(x,y) = \begin{bmatrix} 2 + 2y^2 & 4xy - 2 \\ 4xy - 2 & 2x^2 \end{bmatrix}$$

Now

$$\mathcal{H}f(0,-1) = \begin{bmatrix} 4 & -2 \\ -2 & 0 \end{bmatrix}$$

with determinant -4. $(0,-1)$ is therefore not a minimum.

$$\mathcal{H}f(1,1) = \begin{bmatrix} 4 & 2 \\ 2 & 2 \end{bmatrix}$$

This matrix has trace 6 and determinant 4, so it is positive definite. $(1,1)$ is therefore a local minimum.

b) Starting at $x_0 = [0,0]^\top$, compute one step of a line search method using steepest descent as the search direction. Ensure that the line search step satisfies the Wolfe conditions with $c_1 = \frac{1}{4}$ and $c_2 = \frac{3}{4}$.

Solution: The steepest descent direction is given by

$$p_0 = -\nabla f(0,0) = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$$

and

$$\Phi(\alpha) = f(x_0 + \alpha p_0) = f(2\alpha, 0) = 4\alpha^2 - 4\alpha$$

To satisfy the Armijo condition, we need

$$\Phi(\alpha) \leq \Phi(0) + c_1\Phi'(0)\alpha$$
$$4\alpha^2 - 4\alpha \leq \frac{1}{4} \cdot (-4) \cdot \alpha$$
$$\text{which implies}$$
$$0 \leq \alpha \leq \frac{3}{4},$$

and to satisfy the curvature condition, we need

$$\Phi'(\alpha) \geq c_2\Phi'(0)$$
$$8\alpha - 4 \geq \frac{3}{4} \cdot (-4) \quad \text{which implies}$$
$$\alpha \geq \frac{1}{8}.$$

To satisfy both the Armijo condition and the curvature condition, we need

$$\frac{1}{8} \leq \alpha \leq frac34$$

e.g. $\alpha = \frac{1}{2}$ which gives $x_1 = x_0 + \frac{1}{2}p_0 = (1, 0)$.

c) In part b), could you have used Newton as search direction instead?

Solution: Newton's method at $(0, 0)$ gives

$$\mathcal{H}f(0,0)p_0 = -\nabla f(0,0)$$
$$\begin{bmatrix} 2 & -2 \\ -2 & 0 \end{bmatrix} p^N = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$$

with solution $p^N = \begin{bmatrix} 0 \\ -1 \end{bmatrix}$. We then get that $\nabla f(0,0)^T p^N = 0$, so Newton's method does not give a descent direction in this case.

Problem 6 We study linear programs of the form

$$\min_{x \in \mathbb{R}^n} c^T x \quad \text{s.t.} \quad \begin{cases} \sum_{i=1}^n x_i = 1, & \\ x_i \geq 0, & \text{for } 1 \leq i \leq n, \end{cases}$$

where $c \in \mathbb{R}^n$.

a) Write the linear program in standard form.

Solution: The program is pretty much in standard form already, we only need to replace the sum $\sum_{i=1}^{n} x_i = 1$ with an inner product.

$$\min_x c^T x \quad \text{subject to} \begin{cases} j^T x = 1 \\ \quad x \geq 0 \end{cases}$$

with $j = \begin{bmatrix} 1 & 1 & \dots & 1 \end{bmatrix}^T$

b) Show that the optimal value of the linear program is

$$p = \min_{1 \leq i \leq n} c_i,$$

and describe the set of optimal solutions for all $c \in \mathbb{R}^n$.

Solution: There are several ways to proceed, the most elegant is probably to consider the dual problem. In this case there is only one equality constraint, $\lambda \in \mathbb{R}$, $A = j^T$ and $b = 1$.

The dual problem can be written

$$\max 1 \cdot \lambda \quad \text{subject to} \quad j\lambda \leq c$$

or

$$\max \lambda \quad \text{subject to} \quad \lambda \leq c_i \text{ for all } i$$

The optimal value of the dual problem is given by the most restrictive constraint, so $q = \lambda^* = \min_{1 \leq i \leq n} c_i$.

For linear programs, the optimal values of the primal and dual problem are equal, so $p = q = \min_{1 \leq i \leq n} c_i$.

For the second part, we look at the KKT conditions.

$$j\lambda + \mu = c \tag{1}$$
$$j^\top x = b \tag{2}$$
$$x \geq 0 \tag{3}$$
$$\mu \geq 0 \tag{4}$$
$$x_i \mu_i = 0 \;\; \text{for all } i \tag{5}$$

We know $\lambda = p = \min_i c_i$.

For the further discussion, we define $\mathcal{M} = \{i \text{ s.t. } c_i = p\}$, and $\mathcal{M}^c = \{i \text{ s.t. } c_i > p\}$.

For $i \in \mathcal{M}^c$, we have, by (1):

$$\mu_i = c_i - p > 0$$

and by (5),

$$x_i = 0 \quad \text{for } i \in \mathcal{M}^c$$

In other words, x_i can only be nonzero for $i \in \mathcal{M}$, and if x in addition satisfies the constraints, that is $x_i \geq 0$ and$\sum_i x_i = 1$, we have

$$\begin{aligned} c^T x &= \sum_{i=1}^{n} c_i x_i \quad [x_i = 0 \text{ for } i \in \mathcal{M}^c] \\ &= \sum_{i \in \mathcal{M}} c_i x_i \quad [c_i = p \text{ for } i \in \mathcal{M}] \\ &= p \sum_{i \in \mathcal{M}} x_i \qquad [\sum_i x_i = 1] \\ &= p \end{aligned}$$

So the set of optimal x is given by

$$\left\{ x \in \mathbb{R}^n \quad \text{such that} \quad \begin{aligned} x_i = 0 \text{ for } i \in \mathcal{M}^c, \\ x_i \geq 0 \text{ for } i \in \mathcal{M}, \\ \sum_{i \in \mathcal{M}} x_i = 0 \end{aligned} \right\}$$

where $\mathcal{M}$ is the set of indices i such that $c_i = \min_{1 \leq k \leq n} c_k$, and $\mathcal{M}^c$ its complement.