TMA4180 Optimisation I — Sample Solutions

21 May 2026

Problem 1.

Consider $f(x_1, x_2) = (x_1 - 2)^2 + (x_1^2 - x_2 - 2)^2$.

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

(b) Determine whether $f(x_1, x_2)$ is globally convex. (5 points)

(c) Perform one step of gradient descent with backtracking Armijo line search starting at $(x_1, x_2) = (0, 0)$. Use $c = 0.1$, $\rho = 0.25$, and $\bar{\alpha} = 1$. (10 points)

Solution.

The gradient and Hessian of f are

$$\nabla f = \begin{pmatrix} 2(x_1 - 2) + 4x_1(x_1^2 - x_2 - 2) \\ -2(x_1^2 - x_2 - 2) \end{pmatrix}, \quad \nabla^2 f = \begin{pmatrix} 12x_1^2 - 4x_2 - 6 & -4x_1 \\ -4x_1 & 2 \end{pmatrix}.$$

a) The second component of $\nabla f = 0$ gives $x_2 = x_1^2 - 2$; substituting into the first yields $2x_1 - 4 = 0$, so $x_1 = 2$ and $x_2 = 2$. Thus $(2, 2)$ is the unique stationary point. At this point $\nabla^2 f(2, 2) = \begin{pmatrix} 34 & -8 \\ -8 & 2 \end{pmatrix}$ has determinant $4 > 0$ and positive diagonal, so it is positive definite. Since f is a sum of squares with $f \geq 0$ and $f(2, 2) = 0$, the point $(2, 2)$ is the unique global (and hence local) minimiser.

b) Convexity requires $\det \nabla^2 f \geq 0$ everywhere. We have $\det \nabla^2 f = 2(12x_1^2 - 4x_2 - 6) - 16x_1^2 = 8x_1^2 - 8x_2 - 12$. At $(x_1, x_2) = (0, 0)$ this equals $-12 < 0$, so f is not globally convex.

c) At $(0, 0)$ we have $\nabla f(0, 0) = (-4, 4)^T$ and $f(0, 0) = 4 + 4 = 8$. The search direction is $p = -\nabla f = (4, -4)^T$, with $\nabla f^T p = -32$.

Step length $\alpha = 1$*:* $x = (4, -4)$, $f(4, -4) = 4 + 18^2 = 328$. The Armijo condition requires $328 \leq 8 + 0.1 \cdot 1 \cdot (-32) = 4.8$, which fails.

Step length $\alpha = 0.25$*:* $x = (1, -1)$, $f(1, -1) = 1 + 0 = 1$. The Armijo condition requires $1 \leq 8 + 0.1 \cdot 0.25 \cdot (-32) = 7.2$, which holds. The accepted step is $(1, -1)$.

Problem 2.

Consider the constrained optimisation problem

$$\min_{x \in \Omega} \; f(x_1, x_2) = (x_1^2 - 1)^2 + x_2^2$$

posed on $\Omega = \{(x_1, x_2) \in \mathbb{R}^2 : x_1 + x_2 \geq 0\}$.

(a) Show that the LICQ holds everywhere in Ω and determine all KKT points. (6 points)

(b) For each KKT point, state the Lagrange Hessian $\nabla^2_{xx} \mathcal{L}(\mathbf{x}^*, \lambda^*)$ and determine the critical cone $\mathcal{C}(\mathbf{x}^*, \lambda^*)$ explicitly. Apply both necessary and sufficient second-order conditions to decide which points are local minimisers. (6 points)

(c) Identify the global minimiser of f over Ω. (3 points)

Solution.

Writing $c(x_1, x_2) = x_1 + x_2$, we have

$$\nabla f = \begin{pmatrix} 4x_1(x_1^2 - 1) \\ 2x_2 \end{pmatrix}, \quad \nabla c = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \nabla^2 f = \begin{pmatrix} 12x_1^2 - 4 & 0 \\ 0 & 2 \end{pmatrix}.$$

Since c is linear, $\nabla^2_{xx}\mathcal{L}(x;\lambda) = \nabla^2 f(x)$ for every λ.

a) The gradient $\nabla c = (1,1)^T$ is non-zero everywhere, so with a single constraint the LICQ holds at every feasible point. The KKT conditions are

$$4x_1(x_1^2 - 1) = \lambda, \quad 2x_2 = \lambda, \quad x_1 + x_2 \geq 0, \quad \lambda \geq 0, \quad \lambda(x_1 + x_2) = 0.$$

If $\lambda = 0$, then $x_2 = 0$ and $x_1 \in \{-1, 0, 1\}$; feasibility leaves $(1,0)$ and $(0,0)$. If $\lambda > 0$, complementarity gives $x_2 = -x_1$ and the second equation gives $\lambda = -2x_1 > 0$, so $x_1 < 0$. Substituting into the first equation, $4x_1(x_1^2 - 1) = -2x_1$, i.e. $2x_1(2x_1^2 - 1) = 0$, with negative root $x_1 = -1/\sqrt{2}$. The KKT points are

$$(1,0) \text{ with } \lambda = 0, \qquad (0,0) \text{ with } \lambda = 0, \qquad (-1/\sqrt{2}, 1/\sqrt{2}) \text{ with } \lambda = \sqrt{2}.$$

b) At $(1,0)$ with $\lambda = 0$ the constraint is inactive, so $\mathcal{C} = \mathbb{R}^2$. The Hessian $\nabla^2_{xx}\mathcal{L}(1,0;0) = \operatorname{diag}(8,2)$ is positive definite, so the sufficient conditions hold and $(1,0)$ is a strict local minimiser.

At $(-1/\sqrt{2}, 1/\sqrt{2})$ with $\lambda = \sqrt{2}$ the constraint is active with positive multiplier, so $\mathcal{C} = \{w : w_1 + w_2 = 0\} = \mathbb{R}(1,-1)^T$. The Hessian $\nabla^2_{xx}\mathcal{L} = 2I$ is positive definite on $\mathbb{R}^2$, hence on $\mathcal{C}$. This is a strict local minimiser.

At $(0,0)$ with $\lambda = 0$ the constraint is active but the multiplier vanishes, so the critical cone is the half-plane $\mathcal{C} = \{w : w_1 + w_2 \geq 0\}$. The Hessian $\nabla^2_{xx}\mathcal{L}(0,0;0) = \operatorname{diag}(-4,2)$ is indefinite, and the vector $w = (1,0)^T \in \mathcal{C}$ gives $w^T \nabla^2 \mathcal{L}\, w = -4 < 0$. The necessary condition fails, so $(0,0)$ is not a local minimiser.

c) The local minimisers are $(1,0)$ with $f = 0$ and $(-1/\sqrt{2}, 1/\sqrt{2})$ with $f = 3/4$. Since $f \geq 0$ on $\mathbb{R}^2$ with equality only at $(\pm 1, 0)$, of which only $(1,0)$ is feasible, the unique global minimiser is $(1,0)$.

Problem 3.

Consider $\min_{\mathbf{x}\in\mathbb{R}^2} f(x) = \frac{1}{2}x_1^2 + x_2^2$ subject to $c(x) = x_1 - x_2 - 1 = 0$.

(a) Formulate the quadratic penalty method $Q(\mathbf{x};\mu)$. (4 points)

(b) Find the unconstrained minimiser $\mathbf{x}(\mu)$. (6 points)

(c) Calculate the limit as $\mu \to \infty$ and verify it matches the true constrained minimiser. (5 points)

Solution.

a) The quadratic penalty function is

$$Q(\mathbf{x};\mu) = f(x) + \frac{\mu}{2}c(x)^2 = \frac{1}{2}x_1^2 + x_2^2 + \frac{\mu}{2}(x_1 - x_2 - 1)^2.$$

b) The stationarity conditions $\nabla_x Q = 0$ read

$$x_1 + \mu(x_1 - x_2 - 1) = 0, \qquad 2x_2 - \mu(x_1 - x_2 - 1) = 0.$$

Adding the two equations gives $x_1 + 2x_2 = 0$, i.e. $x_1 = -2x_2$. Substituting into the second equation yields $2x_2 + 3\mu x_2 + \mu = 0$, hence

$$x_2(\mu) = \frac{-\mu}{3\mu + 2}, \qquad x_1(\mu) = \frac{2\mu}{3\mu + 2}.$$

c) As $\mu \to \infty$,

$$x_1(\mu) \to \frac{2}{3}, \qquad x_2(\mu) \to -\frac{1}{3}.$$

The true constrained problem is found by substituting $x_1 = x_2 + 1$ into f, giving $\frac{3}{2}x_2^2 + x_2 + \frac{1}{2}$, with derivative $3x_2 + 1 = 0$, so $x_2 = -1/3$ and $x_1 = 2/3$. The QPM limit matches the true minimiser.

Problem 4.

Let $A \in \mathbb{R}^{n\times n}$ be SPD, and x^* the unique solution to $Ax = b$. Suppose $\{p_i\}_{i=0}^{n-1}$ is conjugate with respect to A. Consider the exact line search $x_{k+1} = x_k + \alpha_k p_k$.

(a) Show that for an exact line search minimising $\phi(x) = x^TAx - b^Tx$, $\alpha_k = -r_k^Tp_k/p_k^TAp_k$, where $r_k = Ax_k - b$. (3 points)

(b) Prove that $\{p_i\}_{i=0}^{n-1}$ are linearly independent. (3 points)

(c) Writing $x^* - x_0 = \sum_{i=0}^{n-1} a_ip_i$, isolate a_k and show $a_k = -p_k^Tr_0/p_k^TAp_k$. (3 points)

(d) Using $x_k = x_0 + \sum_{i=0}^{k-1}\alpha_ip_i$, prove that $p_k^Tr_k = p_k^Tr_0$. (3 points)

(e) Show that an exact line search in conjugate directions converges in at most n steps. (3 points)

Solution.

a) Exact line search requires $\frac{d}{d\alpha}\phi(x_k+\alpha p_k) = 0$, i.e. $\nabla\phi(x_{k+1})^Tp_k = 0$. Since $\nabla\phi(x) = Ax-b = r(x)$, this gives $r_{k+1}^Tp_k = 0$. Writing $r_{k+1} = r_k + \alpha_kAp_k$ and solving yields

$$\alpha_k = -\frac{r_k^Tp_k}{p_k^TAp_k}.$$

b) Suppose $\sum_{i=0}^{n-1} c_ip_i = 0$. Multiplying from the left by p_k^TA and using conjugacy $p_k^TAp_i = 0$ for $i \neq k$ leaves $c_k(p_k^TAp_k) = 0$. Since A is SPD and $p_k \neq 0$, we have $p_k^TAp_k > 0$, hence $c_k = 0$. This holds for every k, so $\{p_i\}$ are linearly independent.

c) Multiplying $x^* - x_0 = \sum_i a_ip_i$ from the left by p_k^TA and using conjugacy gives $p_k^TA(x^* - x_0) = a_k\,p_k^TAp_k$. Since $Ax^* = b$, the left-hand side equals $p_k^T(b - Ax_0) = -p_k^Tr_0$. Therefore

$$a_k = -\frac{p_k^Tr_0}{p_k^TAp_k}.$$

d) From $x_k = x_0 + \sum_{i=0}^{k-1}\alpha_ip_i$ we get $r_k = Ax_k - b = r_0 + \sum_{i=0}^{k-1}\alpha_iAp_i$. Multiplying from the left by p_k^T and using conjugacy $p_k^TAp_i = 0$ for $i < k$ kills the sum, so $p_k^Tr_k = p_k^Tr_0$.

e) From (a) and (d), $\alpha_k = -p_k^Tr_k/p_k^TAp_k = -p_k^Tr_0/p_k^TAp_k = a_k$ by (c). Hence after n steps $x_n = x_0 + \sum_{k=0}^{n-1}\alpha_kp_k = x_0 + \sum_{k=0}^{n-1}a_kp_k = x^*$.

Problem 5.

Quasi-Newton methods approximate the Hessian by B_k. Let $s_k = x_{k+1} - x_k$, $y_k = \nabla f_{k+1} - \nabla f_k$. The update must satisfy the secant equation $B_{k+1}s_k = y_k$. Consider symmetric rank-1 updates $B_{k+1} = B_k + \sigma vv^T$ with $\sigma \in \{-1, 1\}$.

(a) Suggest an initial B_0. Give one pro and one con. (3 points)

(b) Show that $v = \delta(y_k - B_ks_k)$ for some constant δ. (5 points)

(c) Find δ and the resulting SR1 update. (4 points)

(d) Decide true or false: (i) SR1 preserves SPD when B_k is SPD. (ii) BFGS preserves SPD when B_k is SPD and $s_k^Ty_k > 0$. (iii) SR1 is often paired with a trust-region framework rather than line search. (3 points)

Solution.

a) Take $B_0 = I$. *Pro:* cheap, requires no Hessian information, and is symmetric positive definite. *Con:* carries no curvature information, so the first few iterations are typically slow.

b) Inserting the rank-1 update into the secant equation,

$$(B_k + \sigma v v^T)s_k = y_k \quad \Longleftrightarrow \quad \sigma(v^T s_k)\, v = y_k - B_k s_k.$$

Both σ and $v^T s_k$ are scalars, so v must be a scalar multiple of $y_k - B_k s_k$, i.e. $v = \delta(y_k - B_k s_k)$ for some $\delta \in \mathbb{R}$.

c) Substituting $v = \delta(y_k - B_k s_k)$ into $\sigma(v^T s_k)v = y_k - B_k s_k$ and cancelling the vector factor yields

$$\sigma\delta^2 (y_k - B_k s_k)^T s_k = 1.$$

This determines $\sigma\delta^2 = \big((y_k - B_k s_k)^T s_k\big)^{-1}$, with σ matching the sign of $(y_k - B_k s_k)^T s_k$. Either way, the rank-1 update $\sigma v v^T = \sigma\delta^2 (y_k - B_k s_k)(y_k - B_k s_k)^T$ becomes

$$B_{k+1} = B_k + \frac{(y_k - B_k s_k)(y_k - B_k s_k)^T}{(y_k - B_k s_k)^T s_k}.$$

d) *(i) False.* The SR1 denominator $(y_k - B_k s_k)^T s_k$ can be negative, in which case the rank-1 correction is subtracted and may destroy positive definiteness even when B_k is SPD.

(ii) True. The BFGS update is designed so that $s_k^T y_k > 0$ together with SPD B_k guarantees SPD B_{k+1}.

(iii) True. Since SR1 may produce indefinite Hessian approximations, line search methods (which require descent) become problematic. Trust-region methods handle indefinite models naturally via the subproblem.

Problem 6.

Consider the linear program $\min_{x_1,x_2} -3x_1 - 2x_2$.

(a) Find a sequence of points in $\mathbb{R}^2$ showing that the problem with only the constraint $x_1 + x_2 \le 4$ is unbounded. (2 points)

(b) With the constraints $x_1 + x_2 \le 4$, $2x_1 + x_2 \le 6$, $x_1 \ge 0$, $x_2 \ge 0$, rewrite in standard form $\min c^T z$ s.t. $Az = b$, $z \ge 0$, introducing slack variables $s_1, s_2 \ge 0$. (3 points)

(c) Formulate the dual problem. Solve it and verify strong duality. (6 points)

(d) Recalling that a solvable LP has a basic feasible point as solution, sketch $\Omega \subset \mathbb{R}^2$ and identify these points. (4 points)

Solution.

a) Take $(x_1^k, x_2^k) = (k, 4-k)$ for $k \in \mathbb{N}$. Then $x_1^k + x_2^k = 4$, so feasibility holds, while $-3x_1^k - 2x_2^k = -k - 8 \to -\infty$ as $k \to \infty$.

b) With slack variables $x_1 + x_2 + s_1 = 4$ and $2x_1 + x_2 + s_2 = 6$, writing $z = (x_1, x_2, s_1, s_2)^T$, we obtain

$$c = \begin{pmatrix} -3 \\ -2 \\ 0 \\ 0 \end{pmatrix}, \quad A = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 2 & 1 & 0 & 1 \end{pmatrix}, \quad b = \begin{pmatrix} 4 \\ 6 \end{pmatrix}.$$

c) The slack-formulation dual of $\min c^T z$ s.t. $Az = b$, $z \ge 0$ is $\max b^T \lambda$ s.t. $A^T\lambda + s = c$, $s \ge 0$. Substituting,

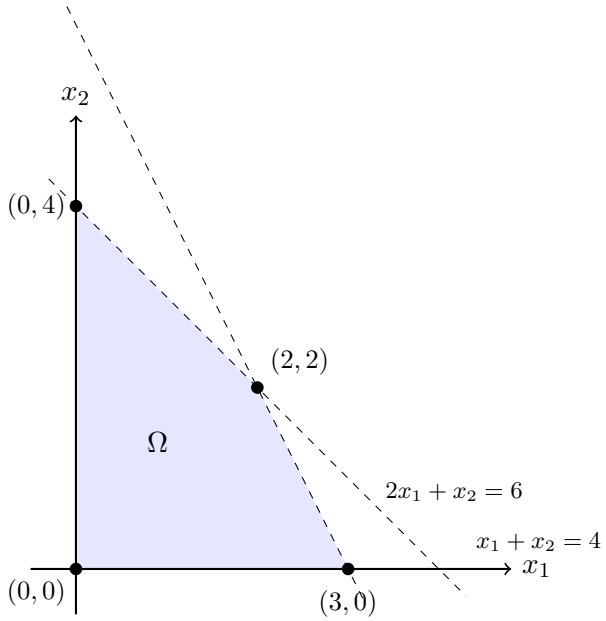
$$\max_{\lambda \in \mathbb{R}^2} 4\lambda_1 + 6\lambda_2 \quad \text{s.t.} \quad \lambda_1 + 2\lambda_2 \le -3,\ \lambda_1 + \lambda_2 \le -2,\ \lambda_1 \le 0,\ \lambda_2 \le 0.$$

Setting the first two constraints to equality, $\lambda_1 + 2\lambda_2 = -3$ and $\lambda_1 + \lambda_2 = -2$, gives $\lambda_2 = -1$ and $\lambda_1 = -1$. The remaining sign constraints hold, and the dual value is $4(-1) + 6(-1) = -10$. The

other vertices of the dual feasible region give strictly smaller values, so $\lambda^* = (-1, -1)$ is the dual optimum.

The primal point $z^* = (2, 2, 0, 0)$ is feasible with $c^T z^* = -10$. Primal and dual values agree, so strong duality holds.

d) The set $\Omega \subset \mathbb{R}^2$ is the quadrilateral with vertices $(0, 0)$, $(3, 0)$, $(2, 2)$, $(0, 4)$.



Evaluating the objective $-3x_1 - 2x_2$ at the vertices yields $0, -9, -10, -8$ respectively, confirming the optimal value -10 at $(2, 2)$, in agreement with the dual solution.