OPTIMISATION

1. a) The first order necessary condition of optimality is

$$0 = \nabla f = \begin{bmatrix} x_1 - x_2 \\ -x_1 + x_2^3 - x_2^2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

which gives the equations

$$x_1 = x_2,$$
$$x_2(x_2^2 - x_2 - 1) = 0.$$

The stationary points are therefore $P_1 = (0,0)$, $P_2 = \left(\frac{1+\sqrt{5}}{2}, \frac{1+\sqrt{5}}{2}\right)$ and $P_3 = \left(\frac{1-\sqrt{5}}{2}, \frac{1-\sqrt{5}}{2}\right)$. [2 marks]

b) The Hessian matrix of the function f is

$$\nabla^2 f(x) = \begin{bmatrix} 1 & -1 \\ -1 & 3x_2^2 - 2x_2 \end{bmatrix}.$$

Evaluating the matrix at the stationary points yields

$$\nabla^2 f(P_1) = \begin{bmatrix} 1 & -1 \\ -1 & 0 \end{bmatrix}$$

which is indefinite,

$$\nabla^2 f(P_2) = \begin{bmatrix} 1 & -1 \\ -1 & \frac{7+\sqrt{5}}{2} \end{bmatrix}$$

which is positive definite, and

$$\nabla^2 f(P_3) = \begin{bmatrix} 1 & -1 \\ -1 & \frac{7-\sqrt{5}}{2} \end{bmatrix}$$

which is also positive definite. Hence, P_1 is a saddle point, whereas P_2 and P_3 are two local minimizers. Computing the function at this points yields $f(P_2) = -1.0075$ and $f(P_3) = -0.0758$. Since f is radially unbounded, *i.e.*

$$\lim_{|x| \to +\infty} f(x) = +\infty,$$

the point P_2 is the global minimizer of f. [4 marks]

c) i) In the first step of the gradient method with extrapolation the term multiplied by β_0 is zero (because of the way the algorithm is initialized), hence the iteration coincides with the gradient iteration. [2 marks]

ii) Running one iteration of the algorithm from the indicated initial conditions yields the point

$$X_1 = \begin{bmatrix} 1 \\ 1+\alpha_0 \end{bmatrix}.$$

Note that, consistently with the answer to c.i), the parameter β_0 does not contribute to the point X_1. To check the descent condition note that

$$f(X_1) - f(X_0) = \frac{1}{12}\alpha_0(3\alpha_0^3 + 8\alpha_0^2 + 6\alpha_0 - 12),$$

hence α_0 should be selected such that

$$\alpha_0(3\alpha_0^3 + 8\alpha_0^2 + 6\alpha_0 - 12) < 0$$

which is the case for α_0 strictly positive and sufficiently small (approximately smaller than 0.82). [4 marks]

iii) Using $\alpha_k = 1/2$ and running one more iteration of the algorithm yields

$$X_2 = \begin{bmatrix} 5/4 \\ 23/16 + 1/2\beta_1 \end{bmatrix}.$$

The descent condition is now

$$f(X_2) - f(X_1) = -0.0788 - 0.1729\beta_1 + 0.415\beta_1^2 + 0.138\beta_1^3 + 0.015625\beta_1^4 < 0,$$

which shows that β_1 should be non-negative and smaller than approximately 0.6. The selection $\beta_1 = 0$ gives a descent condition because for such value of β_1 one has essentially the gradient iteration, for which the descent condition holds for the given selection of α_1. However, $\beta_1 = 0$ is not *optimal* since one could have a greater decrease selecting a strictly positive value of β_1. The optimal selection for this particular case is approximately $\beta_1 = 0.2$. [6 marks]

2. a) The optimization problem used in the proximal method is

$$\begin{aligned} x_{k+1} &= \min_x \left(\frac{1}{2} x' Q x + c' x + d + \frac{1}{2\gamma_k}(x - x_k)'(x - x_k) \right) = \\ &= \min_x \frac{1}{2} x' \left(Q + \frac{1}{\gamma_k} I \right) x + \left(c' - \frac{1}{\gamma_k} x_k' \right) x + d + \frac{1}{2\gamma_k} x_k' x_k. \end{aligned}$$

Note that the function to be minimized is again a quadratic function. Hence, it has a unique solution if and only if the matrix $\left(Q + \frac{1}{\gamma_k} I \right)$ is positive definite (which is always the case since $Q > 0$ and $\gamma_k > 0$).

[2 marks]

b) The solution of the optimization problem is

$$x^* = x_{k+1} = -\left(Q + \frac{1}{\gamma_k} I \right)^{-1} \left(c - \frac{1}{\gamma_k} x_k \right).$$

This can be rewritten as

$$x_{k+1} = \frac{1}{\gamma_k} \left(Q + \frac{1}{\gamma_k} I \right)^{-1} x_k - \left(Q + \frac{1}{\gamma_k} I \right)^{-1} c. \tag{2.1}$$

Hence, $A = \frac{1}{\gamma_k} \left(Q + \frac{1}{\gamma_k} I \right)^{-1}$ and $b = -\left(Q + \frac{1}{\gamma_k} I \right)^{-1} c$.

[6 marks]

c) The fixed point $\bar{x}$ of equation (2.1) is such that

$$\bar{x} = \frac{1}{\gamma_k} \left(Q + \frac{1}{\gamma_k} I \right)^{-1} \bar{x} - \left(Q + \frac{1}{\gamma_k} I \right)^{-1} c.$$

Multiplying on left by $\left(Q + \frac{1}{\gamma_k} I \right)$, yields

$$\left(Q + \frac{1}{\gamma_k} I \right) \bar{x} = \frac{1}{\gamma_k} \bar{x} - c,$$

which gives

$$\bar{x} = -Q^{-1} c,$$

i.e. the global minimizer of the quadratic function.

[2 marks]

d) i) Trivially

$$A = \frac{1}{\gamma_k} \left(Q + \frac{1}{\gamma_k} I \right)^{-1} = (\gamma_k)^{-1} \left(Q + \frac{1}{\gamma_k} I \right)^{-1} = (\gamma_k Q + I)^{-1}.$$

Consider now $A'A < I$. Multiplying on both sides for the matrix $(A'A)^{-1}$ (which exists because $A'A$ is positive definite) yields

$$I < (\gamma_k Q + I)(\gamma_k Q + I)'$$

and

$$\gamma_k^2 Q' Q + \gamma_k (Q' + Q) > 0,$$

which is satisfied given the positivity of Q and γ_k.

[2 marks]

ii) We add and subtract x^* and Ax^* to equation (2.1), obtaining

$$x_{k+1} - x^* = A(x_k - x^*) + b - x^* + Ax^*.$$

Defining $\tilde{b} = b - x^* + Ax^*$, it remains to prove that $\tilde{b} = 0$. Multiplying the expression of $\tilde{b}$ on the left-hand side by $\left(Q + \frac{1}{\gamma_k} I\right)$ yields

$$\begin{aligned}\left(Q+\frac{1}{\gamma_k}I\right)\bar{b} &= \left(Q+\frac{1}{\gamma_k}I\right)\left[\frac{1}{\gamma_k}\left(Q+\frac{1}{\gamma_k}I\right)^{-1}x^* - \left(Q+\frac{1}{\gamma_k}I\right)^{-1}c - x^*\right] \\ &= -c - \left(Q+\frac{1}{\gamma_k}I\right)x^* + \frac{1}{\gamma_k}Ix^* = -c - Qx^*.\end{aligned}$$

The claim is proved substituting the minimizer $x^* = -Q^{-1}c$ in the last equation.

[4 marks]

iii) The equation

$$x_{k+1} - x^* = A(x_k - x^*)$$

is a linear difference equation in which all the eigenvalues of the dynamic matrix A have modulus strictly smaller than one. Hence, the state $x_k - x^*$ converges globally to zero, *i.e.* x_k converges to the optimal solution x^*. The greater the value of γ_k, the smaller the modulus of the eigenvalues of A. Thus, increasing γ_k corresponds to a faster convergence of the algorithm.

[4 marks]

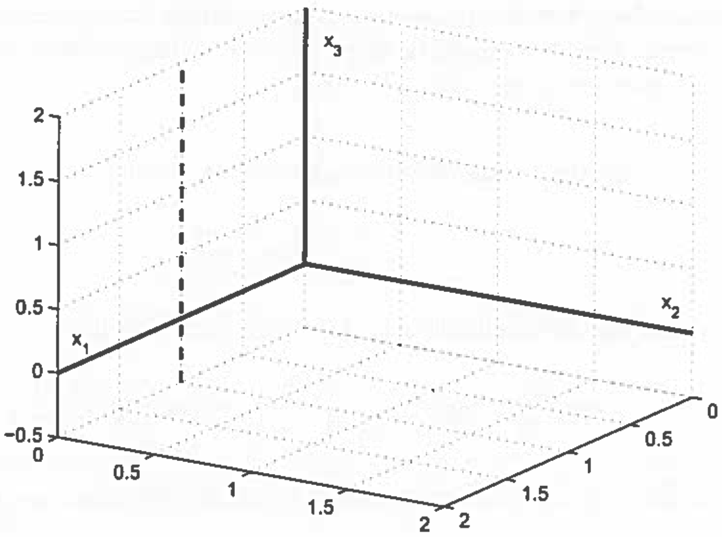

Figure 3.1 Admissible set (dashed line).

3. a) From the first constraint we obtain $x_1 = 1$. Substituting this in the second constraint yields $x_2^2 = 0$. The admissible set $S := \{(1,0,x_3) : x_3 \in \mathbb{R}\}$ is shown in Fig. 3.1. To verify that all the points are non-regular points we check if the gradients of the active constraints are linearly independent. Note however that

$$\text{rank}\begin{bmatrix} 1 & 0 & 0 \\ 2x_1 & 2x_2 & 0 \end{bmatrix}_S = \text{rank}\begin{bmatrix} 1 & 0 & 0 \\ 2 & 0 & 0 \end{bmatrix} = 1,$$

thus all the points of the admissible set are non-regular.

[2 marks]

b) From Fig. 3.1 we note that x_1 and x_2 are fixed. The optimization problem reduces to the minimization of x_3^2, with $x_3 \in \mathbb{R}$. The minimizer is $x_3 = 0$, and the minimum is 1.

[2 marks]

c) The Lagrangian function is

$$L(x,\lambda,\rho) = x_1 + x_2 + x_3^2 + \lambda_1(x_1 - 1) + \lambda_2(x_1^2 + x_2^2 - 1).$$

The first order necessary conditions of optimality are

$$\frac{\partial L}{\partial x_1} = 1 + \lambda_1 + 2\lambda_2 x_1 = 0, \tag{3.1}$$

$$\frac{\partial L}{\partial x_2} = 1 + 2\lambda_2 x_2 = 0, \tag{3.2}$$

$$\frac{\partial L}{\partial x_3} = 2x_3 = 0, \tag{3.3}$$

$$x_1 - 1 = 0, \tag{3.4}$$

$$x_1^2 + x_2^2 - 1 = 0. \tag{3.5}$$

From the last two equations it follows that $x_1 = 1$ and $x_2 = 0$. Substituting this last relation in equation (3.2), yields $\frac{\partial L}{\partial x_2} = 1 = 0$, which does not give any solution. Hence, the first order necessary condition does not give any candidate solution. Note that the first order necessary condition can be used only for regular points. Since this minimization problem has no regular points, the condition cannot give candidate optimal solutions.

[6 marks]

i) Consider the constraints (other selections are possible)

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

Note that the admissible set is left unchanged. Note that

$$
\text{rank}\,\frac{\partial g}{\partial x} = \text{rank}\begin{bmatrix} 1 & 0 & 0 \\ 2x_1 & 1 & 0 \end{bmatrix}_S = \text{rank}\begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \end{bmatrix} = 2.
$$

Hence, all the points of the admissible set are regular.

[2 marks]

ii) The Lagrangian function is

$$
L(x,\lambda) = x_1 + x_2 + x_3^2 + \lambda_1(x_1 - 1) + \lambda_2(x_1^2 + x_2 - 1).
$$

The first order necessary conditions of optimality are

$$
\frac{\partial L}{\partial x_1} = 1 + \lambda_1 + 2\lambda_2 x_1 = 0, \tag{3.6}
$$

$$
\frac{\partial L}{\partial x_2} = 1 + \lambda_2 = 0, \tag{3.7}
$$

$$
\frac{\partial L}{\partial x_3} = 2x_3 = 0, \tag{3.8}
$$

$$
x_1 - 1 = 0, \tag{3.9}
$$

$$
x_1^2 + x_2 - 1 = 0. \tag{3.10}
$$

The unique solution of this set of equations can be easily computed, namely $x_1 = 1$, $x_2 = 0$, $x_3 = 0$, $\lambda_1 = 1$ and $\lambda_2 = -1$.

[4 marks]

iii) The second order sufficient condition of optimality is

$$
s'\nabla^2_{xx}L(x^*,\lambda^*)s = s'\begin{bmatrix} -2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 2 \end{bmatrix} s > 0, \tag{3.11}
$$

for all $s \neq 0$ such that

$$
0 = \frac{\partial g(x^*)}{\partial x}s = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \end{bmatrix} s. \tag{3.12}
$$

Condition (3.12) yields

$$
s = \begin{bmatrix} 0 \\ 0 \\ \alpha \end{bmatrix},
$$

with $\alpha \neq 0$. Hence, condition (3.11) is $2\alpha^2 > 0$, thus the point $(1,0,0)$ is a solution of the considered optimization problem.

[4 marks]

4. a) The Lagrangian function is

$$
\begin{aligned}
L(x,\rho) = {} & 9x_1^2 - 54x_1 + 13x_2^2 - 78x_2 + \rho_1(x_1-4) + \rho_2(x_2-6) \\
& + \rho_3(3x_1+2x_2-18) + \rho_4(-x_1) + \rho_5(-x_2).
\end{aligned}
$$

The necessary conditions of optimality are

$$0 = \frac{\partial L}{\partial x_1} = \rho_1 + 3\rho_3 - \rho_4 + 18x_1 - 54,$$

$$0 = \frac{\partial L}{\partial x_2} = \rho_2 + 2\rho_3 - \rho_5 + 26x_2 - 78,$$

$$x_1 - 4 \leq 0, \qquad x_2 - 6 \leq 0, \qquad 3x_1 + 2x_2 - 18 \leq 0, \qquad -x_1) \leq 0, \qquad -x_2 \leq 0,$$

$$\rho_1(x_1-4) = 0, \qquad \rho_2(x_2-6) = 0, \qquad \rho_3(3x_1+2x_2-18) = 0, \qquad \rho_4(-x_1) = 0, \qquad \rho_5(-x_2),$$

$$\rho_1 \geq 0, \qquad \rho_2 \geq 0, \qquad \rho_3 \geq 0, \qquad \rho_4 \geq 0, \qquad \rho_5 \geq 0.$$

[4 marks]

b) Since at least three multipliers have to be simultaneously zero, one can use the complementarity conditions to identify candidate optimal solutions. For example, setting $\rho_1 = \rho_2 = \rho_3 = 0$ yields

$$x_1 = 0 \qquad x_2 = 0 \qquad \rho_4 = -54 \quad \rho_5 = -78,$$

which does not satisfy the necessary conditions of optimality. The only selection of multipliers which gives a candidate optimal solution is $\rho_1 = \rho_2 = \rho_3 = \rho_4 = \rho_5 = 0$, yielding

$$x_1 = x_2 = 3.$$

Note that since all multipliers are zero then the optimal solution is an interior point of the admissible set, there is no active constraint. [8 marks]

c) One has essentially to repeat the same computation as in part b) looking for all negative multipliers, or changing the sign of f. Using the first approach one has the following candidate optimal solutions (all unspecified multipliers are zero):

$$x_1 = 0 \quad x_2 = 0 \quad \rho_4 = -54 \quad \rho_5 = -78$$

$$x_1 = 4 \quad x_2 = 0 \quad \rho_1 = -18 \quad \rho_5 = -78$$

$$x_1 = 3 \quad x_2 = 0 \quad \rho_5 = -78$$

$$x_1 = 0 \quad x_2 = 3 \quad \rho_4 = -54$$

$$x_1 = 4 \quad x_2 = 3 \quad \rho_1 = -18$$

$$x_1 = 3 \quad x_2 = 3.$$

Comparing the values of the function at the candidate points shows that the point $(x_1, x_2) = (0,0)$ is the global maximizer.