Tutorial Sheet 5

Q1. Compute two iterations for the minimization of

$$f(x_1, x_2) = x_1 + \frac{1}{2}x_2 + \frac{1}{2}x_1^2 + x_2^2 + 3$$

using (i) steepest descent method (ii) Newton's method with starting point $\boldsymbol{x}^{(0)} = \mathbf{0}$. Determine the optimal solution analytically. Compare the rates of convergences.

Q2. The fixed-step-size gradient algorithm defined by

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} - \alpha \nabla f(\boldsymbol{x}^{(k)})$$

is known to converge iff $0 < \alpha < 2/\lambda_{\max}(\boldsymbol{Q})$. Find the largest ranges of values of α for which the minimization algorithm is globally convergent if:

(i) $f(x_1, x_2) = 3(x_1^2 + x_2^2) + 4x_1x_2 + 5x_1 + 6x_2 + 7$;

(ii) $f(x_1, x_2) = 1 + 2x_1 + 3(x_1^2 + x_2^2) + 4x_1x_2$;

(iii) $f(x_1, x_2) = \boldsymbol{x}^T \begin{bmatrix} 3 & 3 \\ 1 & 3 \end{bmatrix} \boldsymbol{x} + [16, 23]\, \boldsymbol{x} + \pi^2$.

Q3. Let $f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}^T\boldsymbol{x} + c$ where A is SPD. If $\boldsymbol{x}^{(0)}$ is such that $\boldsymbol{x}^{(0)} - \boldsymbol{x}^*$ is an eigenvector of $\boldsymbol{A}$, then show that the steepest descent method converges in one step.

Q4. Consider the function $f\colon \mathbb{R}^2 \to \mathbb{R}$ given by

$$f(\boldsymbol{x}) = \frac{3}{2}(x_1^2 + x_2^2) + (1 + a)x_1x_2 - (x_1 + x_2) + b,$$

where a and b are some unknown real-valued parameters.

(a) Write the function f in the usual multivariable quadratic form.

(b) Find the largest set of values a and b such that the unique global minimizer of f exists, and write down the minimizer (in terms of the parameters a and b).

(c) Consider the following algorithm:

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} - \frac{2}{5}\nabla f(\boldsymbol{x}^{(k)}).$$

Find the largest set of values of a and b for which this algorithm converges to the global minimizer of f for any initial point $\boldsymbol{x}^{(0)}$.

Q5. Consider the function $f\colon \mathbb{R} \to \mathbb{R}$ given by $f(x) = x^3 - x$. Suppose that we use a fixed-step-size algorithm $x^{(k+1)} = x^{(k)} - \alpha f'(x^{(k)})$ to find a local minimizer of f. Find the largest range of values of α such that the algorithm is locally convergent (i.e., for all x_0 sufficiently close to a local minimizer x^*, we have $x^{(k)} \to x^*$).

Q6. Consider the optimization problem

$$\text{minimize} \quad \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|^2,$$

where $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, $m \geq n$, and $\boldsymbol{b} \in \mathbb{R}^m$.

(a) Show that the objective function for this problem is a quadratic function, and write down the gradient and Hessian of this quadratic.

(b) Write down the fixed-step-size gradient algorithm for solving this optimization problem.

(c) Suppose that

$$\boldsymbol{A} = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$$

Find the largest range of values for α such that the algorithm in part **Q6.**b converges to the solution of the problem.

Q7. Consider a function $\boldsymbol{f}\colon \mathbb{R}^n \to \mathbb{R}^n$ given by $\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x}+\boldsymbol{b}$, where $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ and $\boldsymbol{b} \in \mathbb{R}^n$. Suppose that $\boldsymbol{A}$ is invertible and $\boldsymbol{x}^*$ is the zero of $\boldsymbol{f}$ [i.e., $\boldsymbol{f}(\boldsymbol{x}^*) = \boldsymbol{0}$]. We wish to compute $\boldsymbol{x}^*$ using the iterative algorithm

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} - \alpha \boldsymbol{f}(\boldsymbol{x}^{(k)}),$$

where $\alpha \in \mathbb{R}$, $\alpha > 0$. We say that the algorithm is *globally monotone* if for any $\boldsymbol{x}^{(0)}$, $\|\boldsymbol{x}^{(k+1)} - \boldsymbol{x}^*\| \leq \|\boldsymbol{x}^{(k)} - \boldsymbol{x}^*\|$ for all k.

(a) Assume that all the eigenvalues of $\boldsymbol{A}$ are real. Show that a necessary condition for the algorithm above to be *globally monotone* is that all the eigenvalues of $\boldsymbol{A}$ are nonnegative.

Hint: Use contraposition.

(b) Suppose that

$$\boldsymbol{A} = \begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix}, \qquad \boldsymbol{b} = \begin{bmatrix} 3 \\ -1 \end{bmatrix}.$$

Find the largest range of values of α for which the algorithm is *globally convergent* (i.e., $\boldsymbol{x}^{(k)} \to \boldsymbol{x}^*$ for all $\boldsymbol{x}^{(0)}$).

Q8. Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be given by $f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} - \boldsymbol{x}^T\boldsymbol{b}$, where $\boldsymbol{b} \in \mathbb{R}^n$ and $\boldsymbol{Q}$ is a real symmetric positive definite $n \times n$ matrix. Suppose that we apply the steepest descent method to this function, with $\boldsymbol{x}^{(0)} \neq \boldsymbol{Q}^{-1}\boldsymbol{b}$. Show that the method converges in one step, that is $\boldsymbol{x}^{(1)} = \boldsymbol{Q}^{-1}\boldsymbol{b}$, if and only if $\boldsymbol{x}^{(0)}$ is chosen such that $\boldsymbol{g}^{(0)} = \boldsymbol{Q}\boldsymbol{x}^{(0)} - \boldsymbol{b}$ is an eigenvector of $\boldsymbol{Q}$.

Solution outlines

A1. Express f in the form $f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} - \boldsymbol{b}^T\boldsymbol{x} + c$ for a symmetric matrix $\boldsymbol{Q}$. Verify that $\boldsymbol{Q}$ turns out to be positive definite.

(i) Apply the formula for the method of steepest descent

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} - \alpha_k \boldsymbol{g}^{(k)}$$

using the explicit formula for α_k given as

$$\alpha_k = \frac{\left(\boldsymbol{g}^{(k)}\right)^T \boldsymbol{g}^{(k)}}{\left(\boldsymbol{g}^{(k)}\right)^T \boldsymbol{Q}\boldsymbol{g}^{(k)}},$$

where $\boldsymbol{g}^{(k)} = \nabla f(\boldsymbol{x}^{(k)})$.

You should get the step sizes as $\alpha_0 = \frac{5}{6}$ and $\alpha_1 = \frac{5}{9}$. The result of two iterations should be $\boldsymbol{x}^{(2)} = [-\frac{25}{27}, -\frac{25}{108}]^T$.

(ii) Use the formula for Newton's method

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} - \boldsymbol{F}(\boldsymbol{x}^{(k)})^{-1}\boldsymbol{g}^{(k)},$$

where $\boldsymbol{g}^{(k)} = \nabla f(\boldsymbol{x}^{(k)})$.

The result of two iterations should be $\boldsymbol{x}^{(2)} = [-1, -\frac{1}{4}]^T$.

The optimal solution can be found analytically by solving $\boldsymbol{Q}\boldsymbol{x}^* = \boldsymbol{b}$. (Justify!) Hence, $\boldsymbol{x}^* = [-1, -\frac{1}{4}]^T$.

A2. In each case, write f in the "standard form" $f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} - \boldsymbol{b}^T\boldsymbol{x} + c$ for a symmetric matrix $\boldsymbol{Q}$, and compute the eigenvalues of $\boldsymbol{Q}$.

(i) $0 < \alpha < \frac{1}{5}$.
(ii) $0 < \alpha < \frac{1}{5}$.
(iii) $0 < \alpha < \frac{1}{5}$.

A3. $f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}^T\boldsymbol{x} + c$.

Given: $\boldsymbol{A}(\boldsymbol{x}^{(0)} - \boldsymbol{x}^*) = \lambda(\boldsymbol{x}^{(0)} - \boldsymbol{x}^*)$ for some $\lambda > 0$.

Steepest descent algorithm: $\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} - \alpha_k \nabla f(\boldsymbol{x}^{(k)})$.

To prove: $\boldsymbol{x}^{(1)} = \boldsymbol{x}^*$.

Proof: Note that $\boldsymbol{g}^{(0)} = \boldsymbol{A}\boldsymbol{x}^{(0)} - \boldsymbol{b}$ and $\boldsymbol{0} = \boldsymbol{A}\boldsymbol{x}^* - \boldsymbol{b}$. (Why? Justify!) So,

$$\begin{aligned}
\boldsymbol{x}^{(1)} &= \boldsymbol{x}^{(0)} - \frac{(\boldsymbol{g}^{(0)})^T\boldsymbol{g}^{(0)}}{(\boldsymbol{g}^{(0)})^T\boldsymbol{A}\boldsymbol{g}^{(0)}}\boldsymbol{g}^{(0)} \\
&= \boldsymbol{x}^{(0)} - \frac{(\boldsymbol{g}^{(0)})^T\boldsymbol{g}^{(0)}}{(\boldsymbol{g}^{(0)})^T\boldsymbol{A}[\boldsymbol{A}\boldsymbol{x}^{(0)} - \boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}^* + \boldsymbol{b}]}[\boldsymbol{A}\boldsymbol{x}^{(0)} - \boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}^* + \boldsymbol{b}] \quad (\text{subtract } \boldsymbol{0} = \boldsymbol{A}\boldsymbol{x}^* - \boldsymbol{b} \text{ in } \boldsymbol{g}^{(0)}) \\
&= \boldsymbol{x}^{(0)} - \frac{(\boldsymbol{g}^{(0)})^T\boldsymbol{g}^{(0)}}{\lambda(\boldsymbol{g}^{(0)})^T\boldsymbol{A}[\boldsymbol{x}^{(0)} - \boldsymbol{x}^*]}[\boldsymbol{A}\boldsymbol{x}^{(0)} - \boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}^* + \boldsymbol{b}] \\
&= \boldsymbol{x}^{(0)} - \frac{(\boldsymbol{g}^{(0)})^T\boldsymbol{g}^{(0)}}{\lambda(\boldsymbol{g}^{(0)})^T[\boldsymbol{A}\boldsymbol{x}^{(0)} - \boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}^* + \boldsymbol{b}]}[\boldsymbol{A}\boldsymbol{x}^{(0)} - \boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}^* + \boldsymbol{b}] \quad (\text{add and subtract } \boldsymbol{b}) \\
&= \boldsymbol{x}^{(0)} - \frac{1}{\lambda}\boldsymbol{A}[\boldsymbol{x}^{(0)} - \boldsymbol{x}^*] \quad (\text{subtract } \boldsymbol{0} = \boldsymbol{A}\boldsymbol{x}^* - \boldsymbol{b} \text{ in } \boldsymbol{g}^{(0)}) \\
&= \boldsymbol{x}^{(0)} - [\boldsymbol{x}^{(0)} - \boldsymbol{x}^*]
\end{aligned}$$

A4. (a)

$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T \begin{bmatrix} 3 & 1+a \\ 1+a & 3 \end{bmatrix} \boldsymbol{x} - \boldsymbol{x}^T \begin{bmatrix} 1 \\ 1 \end{bmatrix} + b.$$

(b) Unique global minimizer exists iff Hessian is positive definite. Use Sylvester's criterion for positive definiteness to find the required range of a and b. Final answer: $a \in (-4, 2)$ and $b \in (-\infty, \infty)$.

(c) This is a fixed-step-size gradient algorithm, so the algorithm is globally convergent iff $\frac{2}{5} < \frac{2}{\lambda_{\max}(\boldsymbol{Q})}$. Compute the eigenvalues of $\boldsymbol{Q}$ to find the required range of a and b. Final answer: $a \in (-3, 1)$ and $b \in (-\infty, \infty)$.

A5. Firstly, show that the only local minimizer of f is $x^* = 1/\sqrt{3}$. Then, note that to check for "local convergence", we may linearize the given fixed-step-size algorithm (since x_0 is assumed to be sufficiently close to x^*) as:

$$x^{(k+1)} = x^{(k)} - \alpha f''(x^*)(x^{(k)} - x^*).$$

Now, notice that this is nothing but the fixed-step-size algorithm applied to a quadratic with second derivative $f''(x^*)$. (Think about this!)

Since $f''(x^*) = 2\sqrt{3}$, the algorithm is locally convergent for $0 < \alpha < \frac{2}{2\sqrt{3}}$.

A6. (a) *Hint:* $\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|^2 = (\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b})^T(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b})$. Final answers: $\nabla f(\boldsymbol{x}) = 2(\boldsymbol{A}^T\boldsymbol{A})\boldsymbol{x} - 2(\boldsymbol{A}^T\boldsymbol{b})$ and $\boldsymbol{F}(\boldsymbol{x}) = 2(\boldsymbol{A}^T\boldsymbol{A})$.

(b) $x^{(k+1)} = x^{(k)} - 2\alpha\boldsymbol{A}^T(\boldsymbol{A}\boldsymbol{x}^{(k)} - \boldsymbol{b})$.

(c) $0 < \alpha < \frac{2}{\lambda_{\max}(\boldsymbol{F})} = \frac{1}{4}$.

A7. (a) Suppose $\boldsymbol{A}\boldsymbol{v} = \lambda\boldsymbol{v}$ for some $\boldsymbol{v} \in \mathbb{R}^n$ and $\lambda < 0$. Choose $\boldsymbol{x}^{(0)} = \boldsymbol{x}^* + \boldsymbol{v}$. Then,

$$\boldsymbol{x}^{(1)} = \boldsymbol{x}^* + \boldsymbol{v} - \alpha(\boldsymbol{A}(\boldsymbol{x}^* + \boldsymbol{v}) + \boldsymbol{b}) \implies \boldsymbol{x}^1 - \boldsymbol{x}^* = (1 - \alpha\lambda)(\boldsymbol{x}^{(0)} - \boldsymbol{x}^*).$$

Since $1 - \alpha\lambda > 1$, the algorithm is not globally monotone.

(b) Observe that the given iterative algorithm is identical to a fixed-step-size algorithm for a quadratic whose Hessian is $\boldsymbol{A}$. Hence, the required range of α is $0 < \alpha < \frac{2}{\lambda_{\max}(\boldsymbol{A})} = \frac{2}{5}$.

A8. Given: $\boldsymbol{x}^{(0)} \neq \boldsymbol{Q}^{-1}\boldsymbol{b}$.

To prove: $\boldsymbol{x}^{(1)} = \boldsymbol{Q}^{-1}\boldsymbol{b} \iff \boldsymbol{Q}\boldsymbol{g}^{(0)} = \lambda\boldsymbol{g}^{(0)}$ for some $\lambda \in \mathbb{R}$.

Proof of ($\implies$):

$$\begin{aligned}
& \boldsymbol{x}^{(1)} = \boldsymbol{x}^{(0)} - \alpha_0\boldsymbol{g}^{(0)} \\
\implies \quad & \boldsymbol{Q}^{-1}\boldsymbol{b} = \boldsymbol{x}^{(0)} - \alpha_0\boldsymbol{g}^{(0)} \\
\implies \quad & \boldsymbol{b} = \boldsymbol{Q}\boldsymbol{x}^{\boldsymbol{(0)}} - \alpha_0\boldsymbol{Q}\boldsymbol{g}^{(0)} \\
\implies \quad & \alpha_0\boldsymbol{Q}\boldsymbol{g}^{(0)} = \boldsymbol{g}^{(0)} \\
\implies \quad & \boldsymbol{Q}\boldsymbol{g}^{(0)} = \frac{1}{\alpha_0}\boldsymbol{g}^{(0)}.
\end{aligned}$$

Call $\frac{1}{\alpha_0} = \lambda$. (Why is $\alpha_0 \neq 0$? Justify!)

Proof of ($\impliedby$):

$$\begin{aligned}
\boldsymbol{x}^{(1)} &= \boldsymbol{x}^{(0)} - \alpha_0\boldsymbol{g}^{(0)} \\
&= \boldsymbol{x}^{(0)} - \frac{(\boldsymbol{g}^{(0)})^T\boldsymbol{g}^{(0)}}{(\boldsymbol{g}^{(0)})^T\boldsymbol{Q}\boldsymbol{g}^{(0)}}\boldsymbol{g}^{(0)} \\
&= \boldsymbol{x}^{(0)} - \frac{(\boldsymbol{g}^{(0)})^T\boldsymbol{g}^{(0)}}{\lambda(\boldsymbol{g}^{(0)})^T\boldsymbol{g}^{(0)}}\boldsymbol{g}^{(0)} \\
&= \boldsymbol{x}^{\boldsymbol{(0)}} - \frac{1}{\lambda}\boldsymbol{g}^{(0)} \\
&= \boldsymbol{x}^{\boldsymbol{(0)}} - \boldsymbol{Q}^{-1}\boldsymbol{g}^{(0)}.
\end{aligned}$$

Tutorial Sheet 6

Q1. Let $\{\boldsymbol{x}^{(k)}\}$ be the sequence generated by Newton's method for minimizing a given objective function $f(\boldsymbol{x})$. Show that if the Hessian $\boldsymbol{F}(\boldsymbol{x}^{(k)}) > 0$ and $\boldsymbol{g}^{(k)} = \nabla f(\boldsymbol{x}^{(k)}) \neq \boldsymbol{0}$, then the search direction

$$\boldsymbol{d}^{(k)} = -\boldsymbol{F}(\boldsymbol{x}^{(k)})^{-1}\boldsymbol{g}^{(k)} = \boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)}$$

from $\boldsymbol{x}^{(k)}$ to $\boldsymbol{x}^{(k+1)}$ is a descent direction for f in the sense that there exists an $\bar{\alpha} > 0$ such that for all $\alpha \in (0, \bar{\alpha}]$,

$$f(\boldsymbol{x}^{(k)} + \alpha\boldsymbol{d}^{(k)}) < f(\boldsymbol{x}^{(k)}).$$

Q2. Show that in the conjugate gradient algorithm,

$$\boldsymbol{g}^{(k+1)^T}\boldsymbol{d}^{(i)} = 0$$

for all k, $0 \leq k \leq n-1$, and $0 \leq i \leq k$.

Q3. Represent the function

$$f(x_1, x_2) = \frac{5}{2}x_1^2 + x_2^2 - 3x_1x_2 - x_2 - 7$$

in the form $f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} - \boldsymbol{x}^T\boldsymbol{b} + c$. Then use the *conjugate gradient algorithm* to construct a vector $\boldsymbol{d}^{(1)}$ that is $\boldsymbol{Q}$-conjugate with $\boldsymbol{d}^{(0)} = \nabla f(\boldsymbol{x}^{(0)})$, where $\boldsymbol{x}^{(0)} = \boldsymbol{0}$.

Q4. Let $f(\boldsymbol{x})$, $\boldsymbol{x} = [x_1, x_2]^T \in \mathbb{R}^2$, be given by

$$f(\boldsymbol{x}) = \frac{5}{2}x_1^2 + \frac{1}{2}x_2^2 + 2x_1x_2 - 3x_1 - x_2.$$

a. Express $f(\boldsymbol{x})$ in the form of $f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} - \boldsymbol{x}^T\boldsymbol{b}$.

b. Find the minimizer of f using the conjugate gradient algorithm. Use a starting point of $\boldsymbol{x}^{(0)} = [0, 0]^T$.

c. Calculate the minimizer of f analytically from $\boldsymbol{Q}$ and $\boldsymbol{b}$, and check it with your answer in part b..

Q5. Given $f\colon \mathbb{R}^n \to \mathbb{R}$, $f \in \mathcal{C}^1$, consider the algorithm

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \alpha_k\boldsymbol{d}^{(k)},$$

where $\boldsymbol{d}^{(1)}, \boldsymbol{d}^{(2)}, \ldots$ are vectors in $\mathbb{R}^n$, and $\alpha_k \geq 0$ is chosen to minimize $f(\boldsymbol{x}^{(k)} + \alpha\boldsymbol{d}^{(k)})$; that is,

$$\alpha_k = \arg\min_{\alpha \geq 0} f(\boldsymbol{x}^{(k)} + \alpha\boldsymbol{d}^{(k)}).$$

Note that the general algorithm encompasses almost all algorithms that we discussed in this part, including the steepest descent, Newton, conjugate gradient, and quasi-Newton algorithms.

Let $\boldsymbol{g}^{(k)} = \nabla f(\boldsymbol{x}^{(k)})$, and assume that $\boldsymbol{d}^{(k)^T}\boldsymbol{g}^{(k)} < 0$.

a. Show that $\boldsymbol{d}^{(k)}$ is a descent direction for f in the sense that there exists $\bar{\alpha} > 0$ such that for all $\alpha \in (0, \bar{\alpha}]$,

$$f(\boldsymbol{x}^{(k)} + \alpha\boldsymbol{d}^{(k)}) < f(\boldsymbol{x}^{(k)}).$$

b. Show that $\alpha_k > 0$.

c. Show that $\boldsymbol{d}^{(k)^T}\boldsymbol{g}^{(k+1)} = 0$.

d. Show that the following algorithms all satisfy the condition $\boldsymbol{d}^{(k)^T}\boldsymbol{g}^{(k)} < 0$, if $\boldsymbol{g}^{(k)} \neq \boldsymbol{0}$:

1. Steepest descent algorithm.
2. Newton's method, assuming that the Hessian is positive definite.
3. Conjugate gradient algorithm.

4. Quasi-Newton algorithm, assuming that $\boldsymbol{H}_k > 0$.

e. For the case where $f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} - \boldsymbol{x}^T\boldsymbol{b}$, with $\boldsymbol{Q} = \boldsymbol{Q}^T > 0$, derive an expression for α_k in terms of $\boldsymbol{Q}$, $\boldsymbol{d}^{(k)}$, and $\boldsymbol{g}^{(k)}$.

Q6. Minimize the function

$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \boldsymbol{x} - \boldsymbol{x}^T \begin{bmatrix} 1 \\ -1 \end{bmatrix} + 7$$

using a quasi-Newton method with the starting point $\boldsymbol{x}^{(0)} = \boldsymbol{0}$.

Solution outlines

For **A1**–**A4**, see the handwritten slides titled "Solution outlines to Tutorial 6".

A5. (a) $\phi'(\alpha) = \boldsymbol{d}^{(k)^T}\nabla f(\boldsymbol{x}^{(k)} + \alpha\boldsymbol{d}^{(k)})$.
Hence, $\phi'(0) = \boldsymbol{d}^{(k)^T}\boldsymbol{g}^{(k)}$.
Since ϕ' is continuous, if $\boldsymbol{d}^{(k)^T}\boldsymbol{g}^{(k)} < 0$, then there exists $\bar{\alpha} > 0$ such that for all $\alpha \in (0, \bar{\alpha}]$, $\phi(\alpha) < \phi(0)$.

(b) From part (a), $\phi(\alpha) < \phi(0)$ for all $\alpha \in (0, \bar{\alpha}]$.
Hence, $\alpha_k = \arg\min \phi(\alpha) \neq 0$, implying that $\alpha_k > 0$.

(c) $\boldsymbol{d}^{(k)^T}\boldsymbol{g}^{(k+1)} = \boldsymbol{d}^{(k)^T}\nabla f(\boldsymbol{x}^{(k)} + \alpha_k\boldsymbol{d}^{(k)}) = \phi'(\alpha_k)$. Since $\alpha_k > 0$, we have $\phi'(\alpha_k) = 0$.

(d) i) We have $\boldsymbol{d}^{(k)} = -\boldsymbol{g}^{(k)}$. Hence $\boldsymbol{d}^{(k)^T}\boldsymbol{g}^k = -\|\boldsymbol{g}^{(k)}\|^2$. If $\boldsymbol{g}^{(k)} \neq \boldsymbol{0}$, then $\|\boldsymbol{g}^{(k)}\|^2 > 0$, and hence proved.

ii) We have $\boldsymbol{d}^{(k)} = -\boldsymbol{F}(\boldsymbol{x}^{(k)})^{-1}\boldsymbol{g}^{(k)}$. Since $\boldsymbol{F}(\boldsymbol{x}^{(k)}) > 0$, we have $\boldsymbol{F}(\boldsymbol{x}^{(k)})^{-1} > 0$. Therefore, $\boldsymbol{d}^{(k)^T}\boldsymbol{g}^k = -\boldsymbol{g}^{(k)^T}\boldsymbol{F}(\boldsymbol{x}^{(k)})^{-1}\boldsymbol{g}^{(k)} < 0$ if $\boldsymbol{g}^{(k)} \neq \boldsymbol{0}$.

iii) We have $\boldsymbol{d}^{(k)} = -\boldsymbol{g}^{(k)} + \beta_{k-1}\boldsymbol{d}^{(k-1)}$.
Hence, $\boldsymbol{d}^{(k)^T}\boldsymbol{g}^k = -\|\boldsymbol{g}^{(k)}\|^2 + \beta_{k-1}\boldsymbol{d}^{(k-1)}\boldsymbol{g}^{(k)^T}$.
By part (c), $\boldsymbol{d}^{(k-1)^T}\boldsymbol{g}^{(k)} = 0$. Hence, if $\boldsymbol{g}^{(k)} \neq \boldsymbol{0}$ and $\|\boldsymbol{g}^{(k)}\| > 0$, then $\boldsymbol{d}^{(k)^T}\boldsymbol{g}^k < 0$.

iv) We have $\boldsymbol{d}^{(k)} = -\boldsymbol{H}_k\boldsymbol{g}^{(k)}$. Therefore, if $\boldsymbol{H} > 0$ and $\boldsymbol{g}^{(k)} \neq \boldsymbol{0}$, then $\boldsymbol{d}^{(k)^T}\boldsymbol{g}^k < 0$.

(e) Using the equation $\nabla f(\boldsymbol{x}) = \boldsymbol{Q}\boldsymbol{x} - \boldsymbol{b}$, we get

$$\begin{aligned} \boldsymbol{d}^{(k)^T}\boldsymbol{g}^{k+1} &= \boldsymbol{d}^{(k)^T}(\boldsymbol{Q}\boldsymbol{x}^{(k+1)} - \boldsymbol{b}) \\ &= \boldsymbol{d}^{(k)^T}(\boldsymbol{Q}(\boldsymbol{x}^{(k)} + \alpha_k\boldsymbol{d}^{(k)}) - \boldsymbol{b}) \\ &= \alpha\boldsymbol{d}^{(k)^T}\boldsymbol{Q}\boldsymbol{d}^{(k)} + \boldsymbol{d}^{(k)^T}\boldsymbol{g}^{(k)}. \end{aligned}$$

Now use part (c) to get the value of α.

A6. Using the rank-one correction method, we obtain the following steps:

$$\nabla f(\boldsymbol{x}^{(0)}) = \boldsymbol{Q}\boldsymbol{x}^{(0)} - \boldsymbol{b} = [-1, 1]^T$$

$$\boldsymbol{d}^{(0)} = -\boldsymbol{H}_0 \boldsymbol{g}^{(0)} = [1, -1]^T$$

$$\alpha_0 = -\frac{\boldsymbol{g}^{(0)T}\boldsymbol{d}^{(0)}}{\boldsymbol{d}^{(0)T}\boldsymbol{Q}\boldsymbol{d}^{(0)}} = \frac{2}{3}$$

$$\boldsymbol{x}^{(1)} = \boldsymbol{x}^{(0)} + \alpha_0 \boldsymbol{d}^{(0)} = [\tfrac{2}{3}, -\tfrac{2}{3}]^T$$

$$\nabla f(\boldsymbol{x}^{(1)} = [\tfrac{-1}{3}, \tfrac{-1}{3}]^T$$

$$\boldsymbol{H}_1 = \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{2} \end{pmatrix}$$

$$\boldsymbol{d}^{(1)} = [\tfrac{1}{3}, \tfrac{1}{6}]^T$$

$$\alpha_1 = 1$$

$$\boldsymbol{x}^{(2)} = \boldsymbol{x}^* = [1, \tfrac{-1}{2}]^T$$

(Try to also apply a rank-two correction method instead, and compare with the above.)

Tutorial Sheet 7

Q1. Show that in the BFGS method when one chooses $\boldsymbol{u} = \Delta\boldsymbol{g}^{(k)}$ and $\boldsymbol{v} = \boldsymbol{B}_k\Delta\boldsymbol{x}^{(k)}$, the vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ are linearly independent.

Q2. Verify that

$$\boldsymbol{B}_{k+1}^{-1} = \left(\boldsymbol{I} - \frac{(\Delta\boldsymbol{x}^{(k)})(\Delta\boldsymbol{g}^{(k)})^T}{(\Delta\boldsymbol{g}^{(k)})^T(\Delta\boldsymbol{x}^{(k)})}\right)\boldsymbol{B}_k^{-1}\left(\boldsymbol{I} - \frac{(\Delta\boldsymbol{g}^{(k)})(\Delta\boldsymbol{x}^{(k)})^T}{(\Delta\boldsymbol{g}^{(k)})^T(\Delta\boldsymbol{x}^{(k)})}\right) + \frac{(\Delta\boldsymbol{x}^{(k)})(\Delta\boldsymbol{x}^{(k)})^T}{(\Delta\boldsymbol{g}^{(k)})^T(\Delta\boldsymbol{x}^{(k)})}$$

by plugging the values (here, the 'k' is removed for notational convenience)

$$\begin{aligned}
\boldsymbol{A} &= \boldsymbol{B} \\
\boldsymbol{U} &= \begin{bmatrix}\boldsymbol{B}\Delta\boldsymbol{x} & \Delta\boldsymbol{g}\end{bmatrix} \\
\boldsymbol{C} &= \begin{bmatrix}\frac{-1}{(\Delta\boldsymbol{x})^T\boldsymbol{B}(\Delta\boldsymbol{x})} & 0 \\ 0 & \frac{1}{(\Delta\boldsymbol{g})^T\Delta\boldsymbol{x}}\end{bmatrix} \\
\boldsymbol{V} &= \begin{bmatrix}(\Delta\boldsymbol{x})^T\boldsymbol{B} \\ (\Delta\boldsymbol{g})^T\end{bmatrix}
\end{aligned}$$

into the Woodbury formula

$$(\boldsymbol{A}+\boldsymbol{U}\boldsymbol{C}\boldsymbol{V})^{-1} = \boldsymbol{A}^{-1} - \boldsymbol{A}^{-1}\boldsymbol{U}(\boldsymbol{C}^{-1}+\boldsymbol{V}\boldsymbol{A}^{-1}\boldsymbol{U})^{-1}\boldsymbol{V}\boldsymbol{A}^{-1}.$$

Q3. Show that rank-two updates preserve positive definiteness. *Hint:* Write the expression for $\boldsymbol{B}_{k+1}^{-1}$ in **Q2.** as $\boldsymbol{B}_{k+1}^{-1} = \boldsymbol{P}+\boldsymbol{Q}$. Then, assuming that $\boldsymbol{B}_k^{-1} > 0$, one needs to show that $\boldsymbol{B}_{k+1}^{-1} > 0$. Use the definition of positive definiteness to analyze the expressions $\boldsymbol{z}^T\boldsymbol{P}\boldsymbol{z}$ and $\boldsymbol{z}^T\boldsymbol{Q}\boldsymbol{z}$ separately, and put together all this information to deduce the positive definiteness of $\boldsymbol{B}_{k+1}^{-1}$.

Q4. Show that in the BFGS algorithm applied to a quadratic with Hessian $\boldsymbol{Q} = \boldsymbol{Q}^T$, we have $\boldsymbol{B}_{k+1}\Delta\boldsymbol{x}^{(i)} = \Delta\boldsymbol{g}^{(i)}$ for all $0 \le i \le k$. *Hint:* Use induction.

Q5. Use the BFGS method to minimize

$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} - \boldsymbol{x}^T\boldsymbol{b} + \pi^2,$$

where

$$\boldsymbol{Q} = \begin{bmatrix}5 & -3 \\ -3 & 2\end{bmatrix}, \qquad \boldsymbol{b} = \begin{bmatrix}0 \\ 1\end{bmatrix}.$$

Take $\boldsymbol{H}_0 = \boldsymbol{I}_2$ and $\boldsymbol{x}^{(0)} = [0,0]^T$. Verify that $\boldsymbol{H}_2 = \boldsymbol{Q}^{-1}$.

Q6. Find local extremizers for the following optimization problems:

(a)

$$\begin{aligned}
\text{Minimize} \quad & x_1^2 + 2x_1x_2 + 3x_2^2 + 4x_1 + 5x_2 + 6x_3 \\
\text{subject to} \quad & x_1 + 2x_2 = 3 \\
& 4x_1 + 5x_3 = 6.
\end{aligned}$$

(b)

$$\begin{aligned}
\text{Maximize} \quad & 4x_1 + x_2^2 \\
\text{subject to} \quad & x_1^2 + x_2^2 = 9.
\end{aligned}$$

(c)

$$\begin{aligned} \text{Maximize} \quad & x_1 x_2 \\ \text{subject to} \quad & x_1^2 + 4x_2^2 = 1. \end{aligned}$$

Q7. Find minimizers and maximizers of the function

$$f(\boldsymbol{x}) = (\boldsymbol{a}^T \boldsymbol{x})(\boldsymbol{b}^T \boldsymbol{x}), \quad \boldsymbol{x} \in \mathbb{R}^3,$$

subject to

$$\begin{aligned} x_1 + x_2 &= 0 \\ x_2 + x_3 &= 0, \end{aligned}$$

where

$$\boldsymbol{a} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \text{ and } \boldsymbol{b} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}.$$

Q8. We wish to construct a closed box with minimum surface area that encloses a volume V cubic feet, where $V > 0$.

(a) Let a, b, and c denote the dimensions of the box with minimum surface area (with volume V). Derive the Lagrangian condition that must be satisfied by a, b, and c.

(b) What does it mean for a point $\boldsymbol{x}^*$ to be a *regular* point in this problem?

(c) Find a, b, and c.

(d) Does the point $\boldsymbol{x}^* = [a, b, c]^T$ found in **Q8.**c satisfy the second-order sufficient condition?

Q9. Consider the problem

$$\begin{aligned} \text{minimize} \quad & x_1 x_2 - 2x_1, \qquad x_1, x_2 \in \mathbb{R} \\ \text{subject to} \quad & x_1^2 - x_2^2 = 0. \end{aligned}$$

(a) Apply Lagrange's theorem directly to the problem to show that if a solution exists, then it must be either $[1, 1]^T$ or $[-1, 1]^T$.

(b) Use the second-order necessary conditions to show that $[-1, 1]^T$ cannot possibly be the solution.

(c) Use the second-order sufficient conditions to show that $[1, 1]^T$ is a strict local minimizer.

Solution outlines

A4. Follow the proof of Theorem 11.3 in [CZ13] given on pages 203–204, taking $\boldsymbol{H}_{k+1} = \boldsymbol{B}_{k+1}^{-1}$.

A5. This is Example 11.4 in [CZ13] given on pages 209–211.

A6. (a) $l(\boldsymbol{x}, \boldsymbol{\lambda}) = f(\boldsymbol{x}) + \boldsymbol{\lambda}^T \boldsymbol{h}(\boldsymbol{x})$ is the lagrangian. We now find critical points by solving following equations:

$$\begin{aligned} \nabla f(\boldsymbol{x}) + \boldsymbol{\lambda}^T \nabla \boldsymbol{h}(\boldsymbol{x}) &= \boldsymbol{0} \\ \boldsymbol{h}(\boldsymbol{x}) &= \boldsymbol{0}^T. \end{aligned}$$

The unique solution is $\boldsymbol{x}^* = [\frac{16}{5}, \frac{-1}{10}, \frac{-34}{25}]^T$ and $\boldsymbol{\lambda}^* = [\frac{-27}{5}, \frac{-6}{5}]^T$.

The Hessian of the lagrangian is: $\begin{pmatrix} 2 & 2 & 0 \\ 2 & 6 & 0 \\ 0 & 0 & 0 \end{pmatrix}$.

$T(\boldsymbol{x}^*) = \{a[\frac{-5}{4}, \frac{5}{8}, 1]^T : a \in \mathbb{R}\}$. $\therefore$ $\boldsymbol{y}^T \boldsymbol{L}(\boldsymbol{x}^*, \boldsymbol{\lambda})\boldsymbol{y} = \frac{75}{32}a^2 > 0$ for all $a \neq 0$. Hence, $\boldsymbol{x}^*$ is a local minimizer.

(b) Proceed exactly as in the previous problem. The four points satisfying the Lagrange condition are:

$$\begin{aligned}\boldsymbol{x}^{(1)} &= [3,0]^T, \lambda^{(1)} = \tfrac{-2}{3};\\ \boldsymbol{x}^{(2)} &= [-3,0]^T, \lambda^{(2)} = \tfrac{2}{3};\\ \boldsymbol{x}^{(3)} &= [2,\sqrt{5}]^T, \lambda^{(3)} = -1;\\ \boldsymbol{x}^{(4)} &= [2,-\sqrt{5}]^T, \lambda^{(4)} = -1.\end{aligned}$$

For the first point, we have $\boldsymbol{L}(\boldsymbol{x}^{(1)},\lambda^{(1)}) = \begin{pmatrix} \frac{-4}{3} & 0 \\ 0 & \frac{2}{3}\end{pmatrix}$.
$T(\boldsymbol{x}^{(1)}) = \{a[0,1]^T : a \in \mathbb{R}\}$. $\therefore$ $\boldsymbol{y}^T\boldsymbol{L}(\boldsymbol{x}^{(1)},\lambda^{(1)})\boldsymbol{y} > 0$ for all nonzero $\boldsymbol{y} \in T(\boldsymbol{x}^{(1)})$.
Similarly, do for the remaining 3 points:
$\boldsymbol{x}^{(2)}$ is a strict local minimizer;
$\boldsymbol{x}^{(3)}$ is a strict local maximizer;
$\boldsymbol{x}^{(4)}$ is a strict local maximizer.

(c) The four points satisfying the Lagrange condition are:

$$\begin{aligned}\boldsymbol{x}^{(1)} &= [\tfrac{1}{\sqrt{2}}, -\tfrac{1}{2\sqrt{2}}]^T, \lambda^{(1)} = \tfrac{1}{4}\\ \boldsymbol{x}^{(2)} &= [\tfrac{-1}{\sqrt{2}}, \tfrac{1}{2\sqrt{2}}]^T, \lambda^{(2)} = \tfrac{1}{4}\\ \boldsymbol{x}^{(3)} &= [\tfrac{1}{\sqrt{2}}, \tfrac{1}{2\sqrt{2}}]^T, \lambda^{(3)} = \tfrac{-1}{4}\\ \boldsymbol{x}^{(4)} &= [\tfrac{-1}{\sqrt{2}}, -\tfrac{1}{2\sqrt{2}}]^T, \lambda^{(4)} = \tfrac{-1}{4}\end{aligned}$$

After calculations of SOSC we realize that first two points are strict local maximizers and last two are strict local minimizers.

A7. $l(\boldsymbol{x},\boldsymbol{\lambda}) = \boldsymbol{a}^T\boldsymbol{x}\boldsymbol{b}^T\boldsymbol{x} + \lambda_1(x_1+x_2) + \lambda_2(x_2+x_3)$

$$\nabla f(\boldsymbol{x}) + \boldsymbol{\lambda}^T\boldsymbol{h}(\boldsymbol{x}) = \begin{pmatrix} x_2+\lambda_1 \\ x_1+x_3+\lambda_1+\lambda_2 \\ x_2+\lambda_2\end{pmatrix} = \begin{pmatrix}0\\0\\0\end{pmatrix}.$$

$$\boldsymbol{h}(\boldsymbol{x}) = \begin{pmatrix} x_1+x_2 \\ x_2+x_3\end{pmatrix} = \begin{pmatrix}0\\0\end{pmatrix}.$$

$\boldsymbol{x}^* = 0$, $\boldsymbol{\lambda}^* = 0$ satisfy the Lagrange, FONC conditions.

The Hessian of the lagrangian is: $\begin{pmatrix}0&1&0\\1&0&1\\0&1&0\end{pmatrix}$

Tangent space: $\{\boldsymbol{y} : \boldsymbol{y} = a[1,-1,1]^T\}$.
$\boldsymbol{y}^T\boldsymbol{L}(\boldsymbol{x}^*,\boldsymbol{\lambda})\boldsymbol{y} = -4a^2 < 0$. Hence, the critical point satisfies SOSC and is a strict local minimizer.

A8. (a) Let x_1, x_2, x_3 be the dimensions of the box. The problem is :

$$\begin{aligned}\text{minimize} \quad & 2(x_1x_2 + x_2x_3 + x_1x_3\\ \text{subject to} \quad & x_1x_2x_3 = V\end{aligned}$$

The dimensions of the box with minimum surface area say [a,b,c] satisfies:

$$\begin{aligned}2(b+c) + \lambda bc &= 0\\ 2(a+c) + \lambda ac &= 0\\ 2(a+b) + \lambda ab &= 0\\ abc - V &= 0\end{aligned}$$

where $\lambda \in \mathbb{R}$.

(b) $a, b, c \neq 0$ (Why?)

(c) Multiply the first equation by a and second equation by b, and subtracting the first from the second, we obtain $c(a-b) = 0$ implying $a = b$. By a similar procedure we can conclude that $a = b = c = V^{\frac{1}{3}}$ and $\lambda = -4V^{\frac{-1}{3}}$.

(d) $\boldsymbol{L(x^*,}\lambda) = -2\begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$.

$T(\boldsymbol{x}^*) = (\boldsymbol{y} : y_3 = -(y_1 + y_2))$

$\boldsymbol{y^T L(x^*, \lambda) y} > 0$. Hence SOSC is satisfied.

A9. (a) Denote the solution by $[x_1^*, x_2^*]$. The Lagrange condition for this problem has the form:

$$\begin{aligned} x_2^* - 2 + 2\lambda^* x_1^* &= 0 \\ x_1^* - 2\lambda^* x_2^* &= 0 \\ (x_1^*)^2 - (x_2^*) &= 0. \end{aligned}$$

Note that $x_1^*, x_2^* \neq 0$.

Combining the first and second equations we obtain $\lambda^* = \frac{2-x_2^*}{2x_1^*} = \frac{x_1^*}{2x_2^*}$.

Hence, $x_2^* = 1$, and hence $x_1^* = 1$. Thus, the only two points satisfying Lagrange condition are $[1,1]^T$, $[-1,1]^T$.

(b) $\boldsymbol{y^T} L(\boldsymbol{x^*, \lambda^*})\boldsymbol{y} = -2a^2 < 0$. Hence, not a local minimizer.

(c) $\boldsymbol{y^T} L(\boldsymbol{x^*, \lambda^*})\boldsymbol{y} = 2a^2 > 0$ Hence it is a strict local minimizer.

Tutorial Sheet 8

Q1. Consider the problem

$$\begin{aligned} &\text{maximize} && ax_1 + bx_2, \qquad x_1, x_2 \in \mathbb{R} \\ &\text{subject to} && x_1^2 + x_2^2 = 2, \end{aligned}$$

where $a, b \in \mathbb{R}$. Show that if $[1, 1]^T$ is a solution to the problem, then $a = b$.

Q2. Let $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, $m \leq n$, $\operatorname{rank}(\boldsymbol{A}) = m$, and $\boldsymbol{x}_0 \in \mathbb{R}^n$. Let $\boldsymbol{x}^*$ be the point on the nullspace of $\boldsymbol{A}$ that is closest to $\boldsymbol{x}_0$ (in the sense of Euclidean norm).

(a) Show that $\boldsymbol{x}^*$ is orthogonal to $\boldsymbol{x}^* - \boldsymbol{x}_0$.

(b) Find a formula for $\boldsymbol{x}^*$ in terms of $\boldsymbol{A}$ and $\boldsymbol{x}_0$.

Q3. Consider the problem

$$\begin{aligned} &\text{minimize} && \frac{1}{2}\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|^2 \\ &\text{subject to} && \boldsymbol{C}\boldsymbol{x} = \boldsymbol{d}, \end{aligned}$$

where $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, $m > n$, $\boldsymbol{C} \in \mathbb{R}^{p\times n}$, $p < n$, and both $\boldsymbol{A}$ and $\boldsymbol{C}$ are of full rank. We wish to find an expression for the solution (in terms of $\boldsymbol{A}$, $\boldsymbol{b}$, $\boldsymbol{C}$, and $\boldsymbol{d}$).

Apply Lagrange's theorem to solve this problem.

Q4. Find local extremizers for:

(a) $x_1^2 + x_2^2 - 2x_1 - 10x_2 + 26$ subject to $\frac{1}{5}x_2 - x_1^2 \leq 0$, $5x_1 + \frac{1}{2}x_2 \leq 5$.

(b) $x_1^2 + x_2^2$ subject to $x_1 \geq 0$, $x_2 \geq 0$, $x_1 + x_2 \geq 5$.

(c) $x_1^2 + 6x_1x_2 - 4x_1 - 2x_2$ subject to $x_1^2 + 2x_2 \leq 1$, $2x_1 - 2x_2 \leq 1$.

Q5. Consider the problem

$$\begin{aligned} &\text{minimize} && x_2 \\ &\text{subject to} && x_2 \geq -(x_1 - 1)^2 + 3. \end{aligned}$$

(a) Find all points satisfying the KKT condition for the problem.

(b) For each point $\boldsymbol{x}^*$ in **Q5.**a, find $T(\boldsymbol{x}^*)$, $N(\boldsymbol{x}^*)$, and $\tilde{T}(\boldsymbol{x}^*, \mu^*)$.

(c) Find the subset of points from **Q5.**a that satisfy the second-order necessary condition.

Q6. Consider the problem of optimizing (either minimizing or maximizing) $(x_1 - 2)^2 + (x_2 - 1)^2$ subject to

$$\begin{aligned} x_2 - x_1^2 &\geq 0 \\ 2 - x_1 - x_2 &\geq 0 \\ x_1 &\geq 0. \end{aligned}$$

The point $\boldsymbol{x}^* = \boldsymbol{0}$ satisfies the KKT conditions.

(a) Does $\boldsymbol{x}^*$ satisfy the FONC for minimization or maximization? What are the KKT multipliers?

(b) Does $\boldsymbol{x}^*$ satisfy the SOSC? Carefully justify your answer.

Q7. Let $g : \mathbb{R}^n \to \mathbb{R}$ and $\boldsymbol{x}_0 \in \mathbb{R}^n$ be given, where $g(\boldsymbol{x}_0) > 0$. Consider the problem

$$\begin{aligned} &\text{minimize} && \frac{1}{2}\|\boldsymbol{x} - \boldsymbol{x}_0\|^2 \\ &\text{subject to} && g(\boldsymbol{x}) \leq 0. \end{aligned}$$

Suppose that $\boldsymbol{x}^*$ is a solution to the problem and $g \in \mathcal{C}^1$. Use the KKT theorem to decide which of the following equations/inequalities hold:

(i) $g(\boldsymbol{x}^*) < 0$.

(ii) $g(\boldsymbol{x}^*) = 0$.

(iii) $(\boldsymbol{x}^* - \boldsymbol{x}_0)^T \nabla g(\boldsymbol{x}^*) < 0$.

(iv) $(\boldsymbol{x}^* - \boldsymbol{x}_0)^T \nabla g(\boldsymbol{x}^*) = 0$.

(v) $(\boldsymbol{x}^* - \boldsymbol{x}_0)^T \nabla g(\boldsymbol{x}^*) > 0$.

Q8. Consider the constraint set $S = \{\boldsymbol{x} : \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}, \boldsymbol{g}(\boldsymbol{x}) \leq \boldsymbol{0}\}$. Let $\boldsymbol{x}^* \in S$ be a regular local minimizer of f over S and $J(\boldsymbol{x}^*)$ the index set of active inequality constraints. Show that $\boldsymbol{x}^*$ is also a regular local minimizer of f over the set $S' = \{\boldsymbol{x} : \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}, g_j(\boldsymbol{x}) = 0, j \in J(\boldsymbol{x}^*)\}$.

Hints:

(1) If g_i is not an active inequality constraint, then what can you say about $g_i(\boldsymbol{x}^*)$? By the continuity of g_i, can you say something about $g_i(\boldsymbol{x})$ for $\boldsymbol{x}$ close enough to $\boldsymbol{x}^*$?

(2) Let B be a ball of radius $\epsilon > 0$ centered at $\boldsymbol{x}^*$, where the choice of ϵ comes from the previous hint. Consider the constraint set $S_1 = \{\boldsymbol{x} : \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}, g_j(\boldsymbol{x}) \leq 0, j \in J(\boldsymbol{x}^*)\}$. Show that $S \cap B = S_1 \cap B$. (To show this, you need to prove that $S \cap B \subseteq S_1 \cap B$ as well as $S \cap B \supseteq S_1 \cap B$.) This means that "locally" near $\boldsymbol{x}^*$ the constraint set S looks the same as the constraint set S_1.

(3) Is $\boldsymbol{x}^*$ a regular local minimizer of f on S_1? Argue carefully.

(4) Note that $S' \subseteq S_1$ and $\boldsymbol{x}^* \in S'$. Conclude.

Q9. Consider the problem

$$\begin{aligned} \text{minimize} \quad & \boldsymbol{c}^T \boldsymbol{x} + 8 \\ \text{subject to} \quad & \frac{1}{2}\|\boldsymbol{x}\|^2 \leq 1, \end{aligned}$$

where $\boldsymbol{c} \in \mathbb{R}^n$, $\boldsymbol{c} \neq \boldsymbol{0}$. Suppose that $\boldsymbol{x}^* = \alpha \boldsymbol{e}$ is a solution to the problem, where $\alpha \in \mathbb{R}$ and $\boldsymbol{e} = [1, \ldots, 1]^T$, and the corresponding objective value is 4.

(a) Show that $\|\boldsymbol{x}^*\|^2 = 2$.

(b) Find α and $\boldsymbol{c}$ (they may depend on n).

Q10. Let $f : \mathbb{R}^n \to \mathbb{R}$, $f \in \mathcal{C}^1$, be a convex function on the set of feasible points

$$\Omega = \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}\},$$

where $\boldsymbol{h} : \mathbb{R}^n \to \mathbb{R}^m$, $\boldsymbol{h} \in \mathcal{C}^1$, and Ω is convex. Suppose that there exists $\boldsymbol{x}^* \in \Omega$ and $\boldsymbol{\lambda}^* \in \mathbb{R}^m$ such that

$$Df(\boldsymbol{x}^*) + \boldsymbol{\lambda}^{*T} D\boldsymbol{h}(\boldsymbol{x}^*) = \boldsymbol{0}^T.$$

Show that $\boldsymbol{x}^*$ is a global minimizer of f over Ω.

Hints:

(1) Recall that for a convex function $f : \Omega \to \mathbb{R}$, we have $f(\boldsymbol{x}) \geq f(\boldsymbol{y}) + Df(\boldsymbol{y})(\boldsymbol{x} - \boldsymbol{y})$ for all $\boldsymbol{x}, \boldsymbol{y} \in \Omega$.
Apply this to the above scenario when $\boldsymbol{y} = \boldsymbol{x}^*$, and express the Df term in terms of $D\boldsymbol{h}$.

(2) To examine the $D\boldsymbol{h}$ term more closely, you need to use the definition of directional derivative. So, start by writing down the expression $\boldsymbol{h}(\boldsymbol{x}^* + \alpha(\boldsymbol{x} - \boldsymbol{x}^*)) - \boldsymbol{h}(\boldsymbol{x}^*)$.
Use the convexity and definition of Ω to say something about this expression.
Then, left-multiply by $\boldsymbol{\lambda}^{*T}$, divide by α, and take the limit as $\alpha \to 0$. Conclude.

Q11. Let $f : \mathbb{R}^n \to \mathbb{R}$, $f \in \mathcal{C}^1$, be a convex function on the set of feasible points

$$\Omega = \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}, \boldsymbol{g}(\boldsymbol{x}) \leq \boldsymbol{0}\},$$

where $\boldsymbol{h} : \mathbb{R}^n \to \mathbb{R}^m$, $\boldsymbol{g} : \mathbb{R}^n \to \mathbb{R}^p$, $\boldsymbol{h}, \boldsymbol{g} \in \mathcal{C}^1$, and Ω is convex. Suppose that there exist $\boldsymbol{x}^* \in \Omega$, $\boldsymbol{\lambda}^* \in \mathbb{R}^m$, and $\boldsymbol{\mu}^* \in \mathbb{R}^p$, such that

1. $\boldsymbol{\mu}^* \geq \boldsymbol{0}$.
2. $Df(\boldsymbol{x}^*) + \boldsymbol{\lambda}^{*T} D\boldsymbol{h}(\boldsymbol{x}^*) + \boldsymbol{\mu}^{*T} D\boldsymbol{g}(\boldsymbol{x}^*) = \boldsymbol{0}^T$.
3. $\boldsymbol{\mu}^{*T}\boldsymbol{g}(\boldsymbol{x}^*) = 0$.

Show that $\boldsymbol{x}^*$ is a global minimizer of f over Ω.

Hints:

(1) Use the convexity of f similarly to the previous exercise. This time, you should be able to rewrite the Df expression in terms of $D\boldsymbol{h}$ and $D\boldsymbol{g}$.

(2) Similar to the previous exercise, show that the expression involving $D\boldsymbol{h}$ evaluates to zero.

(3) Similar to the method for handling the $D\boldsymbol{h}$ expression, can you write down the relevant expression that needs to be examined in order to examine the $D\boldsymbol{g}$ expression? Use the convexity of Ω and the third condition given in the problem. Conclude.

Solution outlines

A1. Verify that $[1, 1]^T$ is a regular point, so that one can apply the Lagrange multiplier theorem at this point to deduce that $a = b$.

A2. Observe that $\boldsymbol{x}^*$ is a solution to the problem

$$\begin{aligned} \text{minimize} \quad & \|\boldsymbol{x} - \boldsymbol{x}_0\|^2 \\ \text{subject to} \quad & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}. \end{aligned}$$

(a) Since $\operatorname{rank}(\boldsymbol{A}) = m$, every feasible point is also regular. Write down the Lagrange multiplier condition for $\boldsymbol{x}^*$. Multiply by $\boldsymbol{x}^*$ appropriately. Conclude.

(b) The Lagrange multiplier condition implies that $\boldsymbol{x}^* - \boldsymbol{x}_0 = \boldsymbol{A}^T\boldsymbol{\lambda}^*$. Multiply on the left by $\boldsymbol{A}$, and conclude. You should get $\boldsymbol{\lambda}^* = -(\boldsymbol{A}\boldsymbol{A}^T)^{-1}\boldsymbol{A}\boldsymbol{x}_0$ and $\boldsymbol{x}^* = (\boldsymbol{I}_n - \boldsymbol{A}^T(\boldsymbol{A}\boldsymbol{A}^T)^{-1}\boldsymbol{A})\boldsymbol{x}_0$.

A3. The Lagrange condition is:

$$\begin{aligned} (\boldsymbol{A}\boldsymbol{x}^* - \boldsymbol{b})^T\boldsymbol{A} + \boldsymbol{\lambda}^{*T}\boldsymbol{C} &= \boldsymbol{0}^T \\ \boldsymbol{C}\boldsymbol{x}^* &= \boldsymbol{d}. \end{aligned}$$

The first equation gives

$$\boldsymbol{x}^* = (\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T\boldsymbol{b} - (\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{C}^T\boldsymbol{\lambda}^*.$$

One needs to eliminate $\boldsymbol{\lambda}$. So, multiply on the left by $\boldsymbol{C}$ and use the second equation to get

$$\boldsymbol{d} = \boldsymbol{C}(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T\boldsymbol{b} - \boldsymbol{C}(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{C}^T\boldsymbol{\lambda}^*$$

Use this to find an expression for $\boldsymbol{\lambda}^*$ in terms of $\boldsymbol{A}$, $\boldsymbol{b}$, $\boldsymbol{C}$, and $\boldsymbol{d}$ only. Substitute that back into the expression for $\boldsymbol{x}^*$ to get your final answer.

A4. (a) The KKT conditions for the minimization problem are:

$$\begin{aligned} 2x_1 - 2 - 2\mu_1 x_1 + 5\mu_2 &= 0 \\ 2x_2 - 10 + \frac{1}{5}\mu_1 + \frac{1}{2}\mu_2 &= 0 \\ \mu_1\left(\frac{1}{5}x_2 - x_1^2\right) + \mu_2\left(5x_1 + \frac{1}{2}x_2 - 5\right) &= 0 \\ \boldsymbol{\mu} &\geq 0. \end{aligned}$$

We now have four cases for μ_1 and μ_2.

- Case I: $\mu_1 = \mu_2 = 0$. Then, $\boldsymbol{x}_{\mathrm{I}} = [1, 5]^T$ is the only critical point, but it is not feasible.
- Case II: $\mu_1 > 0$, $\mu_2 = 0$. The complementarity slack condition implies that $x_2 = 5x_1^2$. Using this, the KKT conditions simplify to:

$$
\begin{aligned}
x_1(1-\mu_1) &= 1 \\
50x_1^2 - 50 + \mu_1 &= 0.
\end{aligned}
$$

  Solve for μ_1 and then x_1. You should get:

$$\boldsymbol{\mu}_{\mathrm{II}}^+ = [26 + 5\sqrt{23}, 0]^T, \ \boldsymbol{x}_{\mathrm{II}}^+ = [\tfrac{-5+\sqrt{23}}{10}, \tfrac{24-5\sqrt{23}}{10}]^T$$

  and

$$\boldsymbol{\mu}_{\mathrm{II}}^- = [26 - 5\sqrt{23}, 0]^T, \ \boldsymbol{x}_{\mathbf{II}}^- = [\tfrac{-5-\sqrt{23}}{10}, \tfrac{24+5\sqrt{23}}{10}]^T.$$

  Both points are feasible.
- Case III: $\mu_1 = 0$, $\mu_2 > 0$. The complementarity slack condition implies that $x_2 = 10 - 10x_1$. Using this, the KKT conditions simplify to:

$$
\begin{aligned}
2x_1 - 2 + 5\mu_2 &= 0 \\
-20x_1 + 10 + \frac{1}{5}\mu_2 &= 0.
\end{aligned}
$$

  Solve for μ_2 and then x_1. You should get:

$$\boldsymbol{\mu}_{\mathrm{III}} = [0, \tfrac{50}{251}]^T, \ \boldsymbol{x}_{\mathrm{III}} = [\tfrac{126}{251}, \tfrac{1250}{251}]^T.$$

  However, this point is not feasible.
- Case IV: $\mu_1, \mu_2 > 0$. The complementarity slack condition implies that:

$$
\begin{aligned}
\frac{1}{5}x_2 - x_1^2 &= 0 \\
5x_1 + \frac{1}{2}x_2 - 5 &= 0.
\end{aligned}
$$

  Solving for x_1 and x_2, we get:

$$\boldsymbol{x}_{\mathrm{IV}}^+ = [-1-\sqrt{3}, 20+10\sqrt{3}]^T, \ \boldsymbol{x}_{\mathrm{IV}}^- = [-1+\sqrt{3}, 20-10\sqrt{3}]^T.$$

  Using these, solve for μ_1 and μ_2. You should get:

$$\boldsymbol{\mu}_{\mathrm{IV}}^+ = [101 + \tfrac{152}{\sqrt{3}}, \tfrac{-502}{5} - \tfrac{904}{5\sqrt{3}}]^T, \ \boldsymbol{\mu}_{\mathrm{IV}}^- = [101 - \tfrac{152}{\sqrt{3}}, \tfrac{-502}{5} + \tfrac{904}{5\sqrt{3}}]^T.$$

  Since $\boldsymbol{\mu}_{\mathrm{IV}}^+$ does not satisfy the dual feasibility condition, we eliminate $\boldsymbol{x}_{\mathrm{IV}}^+$ from consideration, and only retain $\boldsymbol{x}_{\mathrm{IV}}^-$, which is feasible.

Finally, check that the three candidate points $\boldsymbol{x}_{\mathrm{II}}^{\pm}$ and $\boldsymbol{x}_{\mathrm{IV}}^-$ are all regular.

What about for the corresponding maximization problem? The only difference is that the dual feasibility condition changes to $\boldsymbol{\mu}^* \leq 0$. While three of the four cases will correspondingly change, this does not result in any change in the final calculations. Since the KKT multipliers found for the feasible points do not satisfy $\boldsymbol{\mu}^* \leq 0$, there are no new points to consider.

So, we now have to apply the second order Lagrange conditions. The Hessian of the Lagrangian here is:

$$\boldsymbol{L}(\boldsymbol{x}^*, \boldsymbol{\mu}^*) = \begin{bmatrix} 2-2\mu_1 & 0 \\ 0 & 2 \end{bmatrix}.$$

Since there are no immediate conclusions that can be drawn from this for any of the three points, let's compute the tangent spaces at each point to apply the SONC. We have

$$J(\boldsymbol{x}_{\mathrm{II}}^{\pm}) = \{j : g_j(\boldsymbol{x}_{\mathrm{II}}^{\pm}) = 0\} = \{1\},$$

so,

$$T(\boldsymbol{x}_{\mathrm{II}}^{\pm}) = \{\boldsymbol{y} \in \mathbb{R}^2 : Dg_1(\boldsymbol{x}_{\mathrm{II}}^{\pm})\boldsymbol{y} = 0\}.$$

Solving, we get

$$T(\boldsymbol{x}_{\mathrm{II}}^{\pm}) = \{a[1, -5 \pm \sqrt{23}]^T : a \in \mathbb{R}\}.$$

Similarly,

$$J(\boldsymbol{x}_{\mathrm{IV}}^{-}) = \{j : g_j(\boldsymbol{x}_{\mathrm{IV}}^{-}) = 0\} = \{1, 2\},$$

so,

$$T(\boldsymbol{x}_{\mathrm{IV}}^{-}) = \{\boldsymbol{y} \in \mathbb{R}^2 : D\boldsymbol{g}(\boldsymbol{x}_{\mathrm{IV}}^{-})\boldsymbol{y} = \boldsymbol{0}\}.$$

Solving, we get

$$T(\boldsymbol{x}_{\mathrm{IV}}^{-}) = \{\boldsymbol{0}\}.$$

Now, check that for $\boldsymbol{y} \in T(\boldsymbol{x}_{\mathrm{II}}^{+})$, we have

$$\boldsymbol{y}^T \boldsymbol{L}(\boldsymbol{x}_{\mathrm{II}}^{+}, \boldsymbol{\mu}_{\mathrm{II}}^{+})\boldsymbol{y} = a^2[2 - 2(26 + 5\sqrt{23}) + 2(-5 + \sqrt{23})^2] = -2a^2[15\sqrt{23} - 23],$$

which is < 0 when $a \neq 0$ (for instance, take $a = 1$). Hence, $\boldsymbol{x}_{\mathrm{II}}^{+}$ does not satisfy the SONC.
Similarly, check that for $\boldsymbol{y} \in T(\boldsymbol{x}_{\mathrm{II}}^{-})$, we have

$$\boldsymbol{y}^T \boldsymbol{L}(\boldsymbol{x}_{\mathrm{II}}^{-}, \boldsymbol{\mu}_{\mathrm{II}}^{-})\boldsymbol{y} = a^2[2 - 2(26 - 5\sqrt{23}) + 2(-5 - \sqrt{23})^2] = 2a^2[15\sqrt{23} + 23],$$

which is always ≥ 0. Hence, $\boldsymbol{x}_{\mathrm{II}}^{-}$ satisfies the SONC.
Lastly, notice that since $T(\boldsymbol{x}_{\mathrm{IV}}^{-}) = \{\boldsymbol{0}\}$, it is true that $\boldsymbol{y}^T \boldsymbol{L}(\boldsymbol{x}_{\mathrm{IV}}^{-}, \boldsymbol{\mu}_{\mathrm{IV}}^{-})\boldsymbol{y} \leq 0$ for all $\boldsymbol{y} \in T(\boldsymbol{x}_{\mathrm{IV}}^{-})$, and so $\boldsymbol{x}_{\mathrm{IV}}^{-}$ also satisfies the SONC.
Next, to check the SOSC condition for $\boldsymbol{x}_{\mathrm{II}}^{-}$ and $\boldsymbol{x}_{\mathrm{IV}}^{-}$, we need to compute $\tilde{T}(\boldsymbol{x}^*, \boldsymbol{\mu}^*)$. We have

$$\tilde{J}(\boldsymbol{x}_{\mathrm{II}}^{-}, \boldsymbol{\mu}_{\mathrm{II}}^{-}) = \{j : g_j(\boldsymbol{x}_{\mathrm{II}}^{-}) = 0, \mu_j > 0\} = \{1\} = J(\boldsymbol{x}_{\mathrm{II}}^{-}),$$

so,

$$\tilde{T}(\boldsymbol{x}_{\mathrm{II}}^{-}, \boldsymbol{\mu}_{\mathrm{II}}^{-}) = T(\boldsymbol{x}_{\mathrm{II}}^{-}).$$

Hence, for all nonzero $y \in \tilde{T}(\boldsymbol{x}_{\mathrm{II}}^{-}, \boldsymbol{\mu}_{\mathrm{II}}^{-})$,

$$\boldsymbol{y}^T \boldsymbol{L}(\boldsymbol{x}_{\mathrm{II}}^{-}, \boldsymbol{\mu}_{\mathrm{II}}^{-})\boldsymbol{y} > 0.$$

Hence, $\boldsymbol{x}_{\mathrm{II}}^{-}$ satisfies the SOSC and is a strict local minimizer.
Similarly, we have

$$\tilde{J}(\boldsymbol{x}_{\mathrm{IV}}^{-}, \boldsymbol{\mu}_{\mathrm{IV}}^{-}) = \{j : g_j(\boldsymbol{x}_{\mathrm{IV}}^{-}) = 0, \mu_j > 0\} = \{1, 2\} = J(\boldsymbol{x}_{\mathrm{IV}}^{-}),$$

so,

$$\tilde{T}(\boldsymbol{x}_{\mathrm{IV}}^{-}, \boldsymbol{\mu}_{\mathrm{IV}}^{-}) = T(\boldsymbol{x}_{\mathrm{IV}}^{-}).$$

Since there is no nonzero $y \in \tilde{T}(\boldsymbol{x}_{\mathrm{IV}}^{-}, \boldsymbol{\mu}_{\mathrm{IV}}^{-})$, it is vacuously true that

$$\boldsymbol{y}^T \boldsymbol{L}(\boldsymbol{x}_{\mathrm{IV}}^{-}, \boldsymbol{\mu}_{\mathrm{IV}}^{-})\boldsymbol{y} > 0$$

for all nonzero $\boldsymbol{y} \in \tilde{T}(\boldsymbol{x}_{\mathrm{IV}}^{-}, \boldsymbol{\mu}_{\mathrm{IV}}^{-})$. Hence, $\boldsymbol{x}_{\mathrm{IV}}^{-}$ satisfies the SOSC and is a strict local minimizer.

(b) The KKT conditions for the minimization problem are:

$$\begin{aligned}
2x_1 - \mu_1 - \mu_3 &= 0 \\
2x_2 - \mu_2 - \mu_3 &= 0 \\
-x_1 &\leq 0 \\
-x_2 &\leq 0 \\
-x_1 - x_2 + 5 &\leq 0 \\
-\mu_1 x_1 - \mu_2 x_2 + \mu_3(-x_2 - x_2 + 5) &= 0 \\
\boldsymbol{\mu} &\geq 0.
\end{aligned}$$

We now have eight cases for μ_1, μ_2, and μ_3.

- Case I: $\mu_1 = \mu_2 = \mu_3 = 0$. Then, $\boldsymbol{x}_{\text{I}} = [0, 0]^T$ is the only critical point, but it is not feasible.
- Case II: $\mu_1 > 0$, $\mu_2 = \mu_3 = 0$. Then, $\boldsymbol{x}_{\text{II}} = [0, 0]^T$ is the only critical point, but it is not feasible.
- Case III: $\mu_2 > 0$, $\mu_1 = \mu_3 = 0$. Then, $\boldsymbol{x}_{\text{III}} = [0, 0]^T$ is the only critical point, but it is not feasible.
- Case IV: $\mu_3 > 0$, $\mu_1 = \mu_2 = 0$. Then, $\boldsymbol{\mu}_{\text{IV}} = [0, 0, 5]^T$ and $\boldsymbol{x}_{\text{IV}} = [\frac{5}{2}, \frac{5}{2}]^T$. This point is feasible.
- Case V: $\mu_1, \mu_2 > 0$, $\mu_3 = 0$. Then, $\boldsymbol{x}_{\text{V}} = [0, 0]^T$ is the only critical point, but it is not feasible.
- Case VI: $\mu_1, \mu_3 > 0$, $\mu_2 = 0$. Then, $\boldsymbol{x}_{\text{VI}} = [0, 5]^T$, $\boldsymbol{\mu}_{\text{VI}} = [-10, 0, 10]^T$. Since the dual feasibility condition is not satisfied, we eliminate $\boldsymbol{x}_{\text{VI}}$ from consideration.
- Case VII: $\mu_2, \mu_3 > 0$, $\mu_1 = 0$. Then, $\boldsymbol{x}_{\text{VII}} = [5, 0]^T$, $\boldsymbol{\mu}_{\text{VII}} = [0, -10, 10]^T$. Since the dual feasibility condition is not satisfied, we eliminate $\boldsymbol{x}_{\text{VII}}$ from consideration.
- Case VIII: $\mu_1, \mu_2, \mu_3 > 0$. Then, $\boldsymbol{x}_{\text{VIII}} = [0, 0]^T$ is the only critical point, but it is not feasible.

Finally, check that the candidate point $\boldsymbol{x}_{\text{IV}}$ is regular.

What about for the corresponding maximization problem? The only difference is that the dual feasibility condition changes to $\boldsymbol{\mu}^* \leq 0$. While seven of the eight cases will correspondingly change, this does not result in any change in the final calculations. Since the KKT multipliers found for the feasible points do not satisfy $\boldsymbol{\mu}^* \leq 0$, there are no new points to consider.

So, we now have to apply the second order Lagrange conditions. The Hessian of the Lagrangian here is:

$$\boldsymbol{L}(\boldsymbol{x}^*, \boldsymbol{\mu}^*) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} > 0.$$

Hence, $\boldsymbol{x}_{\text{IV}}$ is a strict local minimizer.

(c) The KKT conditions for the minimization problem are:

$$\begin{aligned} 2x_1 + 6x_2 - 4 + 2\mu_1 x_1 + 2\mu_2 &= 0 \\ 6x_1 - 2 + 2\mu_1 - 2\mu_2 &= 0 \\ x_1^2 + 2x_2 - 1 &\leq 0 \\ 2x_1 - 2x_2 - 1 &\leq 0 \\ \mu_1(x_1^2 + 2x_2 - 1) + \mu_2(2x_1 - 2x_2 - 1) &= 0 \\ \boldsymbol{\mu}^* &\geq 0. \end{aligned}$$

We now have four cases for μ_1 and μ_2.

- Case I: $\mu_1 = \mu_2 = 0$. Then, $\boldsymbol{x}_{\text{I}} = [\frac{1}{3}, \frac{5}{9}]^T$ is the only critical point, but it is not feasible.
- Case II: $\mu_1 > 0$, $\mu_2 = 0$. The complementarity slack condition implies that $x_2 = \frac{1}{2}(1 - x_1^2)$. Using this, the KKT conditions simplify to:

$$\begin{aligned} 1 - 3x_1 &= \mu_1 \\ 9x_1^2 - 4x_1 + 1 &= 0. \end{aligned}$$

  Since the quadratic in x_1 has no real roots, there are no critical points from this case.
- Case III: $\mu_1 = 0$, $\mu_2 > 0$. The complementarity slack condition implies that $x_2 = x_1 - \frac{1}{2}$. Using this, the KKT conditions simplify to:

$$\begin{aligned} 3x_1 - 1 &= \mu_2 \\ 14x_1 - 9 &= 0. \end{aligned}$$

Solving, you should get:

$$\boldsymbol{x}_{\text{III}} = [\tfrac{9}{14}, \tfrac{1}{7}]^T, \ \boldsymbol{\mu}_{\text{III}} = [0, \tfrac{13}{14}]^T.$$

This point is feasible.

- Case IV: $\mu_1, \mu_2 > 0$. The complementarity slack condition implies that:

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

  Solving for x_1 and x_2, we get:

$$\boldsymbol{x}_{\text{IV}}^{+} = [-1+\sqrt{3}, \tfrac{-3}{2}+\sqrt{3}]^T, \ \boldsymbol{x}_{\text{IV}}^{-} = [-1-\sqrt{3}, \tfrac{-3}{2}-\sqrt{3}]^T.$$

  Using these, solve for μ_1 and μ_2. You should get:

$$\boldsymbol{\mu}_{\text{IV}}^{+} = [\tfrac{-14\sqrt{3}+23}{2\sqrt{3}}, \tfrac{-22\sqrt{3}+41}{2\sqrt{3}}]^T, \ \boldsymbol{\mu}_{\text{IV}}^{-} = [\tfrac{-14\sqrt{3}-23}{2\sqrt{3}}, \tfrac{-22\sqrt{3}-41}{2\sqrt{3}}]^T.$$

  Since neither KKT multiplier satisfies the dual feasibility condition, we eliminate both points from consideration.

Finally, check that the candidate point $\boldsymbol{x}_{\text{III}}$ is regular.

What about the corresponding maximization problem? The only difference is that the dual feasibility condition changes to $\boldsymbol{\mu}^* \leq 0$. While three of the four cases will correspondingly change, this does not result in any change in the final calculations. Since the KKT multiplier $\boldsymbol{\mu}_{\text{IV}}^{-}$ satisfies $\boldsymbol{\mu}^* \leq 0$, the point $\boldsymbol{x}_{\text{IV}}^{-}$ is a candidate maximizer. We also check that the $\boldsymbol{x}_{\text{IV}}^{-}$ is regular.

So, we now have to apply the second order Lagrange condition. The Hessian of the Lagrangian here is:

$$\boldsymbol{L}(\boldsymbol{x}^*, \boldsymbol{\mu}^*) = \begin{bmatrix} 2+2\mu_1 & 6 \\ 6 & 0 \end{bmatrix}.$$

Since $\boldsymbol{L}(\boldsymbol{x}_{\text{IV}}^{-}, \boldsymbol{\mu}_{\text{IV}}^{-}) < 0$, the SOSC holds for $\boldsymbol{x}_{\text{IV}}^{-}$. Hence, $\boldsymbol{x}_{\text{IV}}^{-}$ is a strict local maximizer.

For the point $\boldsymbol{x}_{\text{III}}$, no immediate conclusions that can be drawn by looking at

$$\boldsymbol{L}(\boldsymbol{x}_{\text{III}}, \boldsymbol{\mu}_{\text{III}}) = \begin{bmatrix} 2 & 6 \\ 6 & 0 \end{bmatrix},$$

so let's compute the tangent space at $\boldsymbol{x}_{\text{III}}$ to apply the SONC. We have

$$J(\boldsymbol{x}_{\text{III}}) = \{j : g_j(\boldsymbol{x}_{\text{III}}) = 0\} = \{2\},$$

so,

$$T(\boldsymbol{x}_{\text{III}}) = \{\boldsymbol{y} \in \mathbb{R}^2 : Dg_2(\boldsymbol{x}_{\text{III}})\boldsymbol{y} = 0\}.$$

Solving, we get

$$T(\boldsymbol{x}_{\text{III}}) = \{a[1,1]^T : a \in \mathbb{R}\}.$$

Now, check that for $\boldsymbol{y} \in T(\boldsymbol{x}_{\text{III}})$, we have

$$\boldsymbol{y}^T \boldsymbol{L}(\boldsymbol{x}_{\text{III}}, \boldsymbol{\mu}_{\text{III}})\boldsymbol{y} = 14a^2,$$

which is always ≥ 0. Hence, $\boldsymbol{x}_{\text{III}}$ satisfies the SONC.

Next, to check the SOSC condition for $\boldsymbol{x}_{\text{III}}$, we need to compute $\tilde{T}(\boldsymbol{x}_{\text{III}}, \boldsymbol{\mu}_{\text{III}})$. We have

$$\tilde{J}(\boldsymbol{x}_{\text{III}}, \boldsymbol{\mu}_{\text{III}}) = \{j : g_j(\boldsymbol{x}_{\text{III}}) = 0, \mu_j > 0\} = \{2\} = J(\boldsymbol{x}_{\text{III}}),$$

so,

$$\tilde{T}(\boldsymbol{x}_{\text{III}}, \boldsymbol{\mu}_{\text{III}}) = T(\boldsymbol{x}_{\text{III}}).$$

Hence, for all nonzero $\boldsymbol{y} \in \tilde{T}(\boldsymbol{x}_{\text{III}}, \boldsymbol{\mu}_{\text{III}})$,

$$\boldsymbol{y}^T \boldsymbol{L}(\boldsymbol{x}_{\text{III}}, \boldsymbol{\mu}_{\text{III}})\boldsymbol{y} > 0.$$

Hence, $\boldsymbol{x}_{\text{III}}$ satisfies the SOSC and is a strict local minimizer.

A5. (a) The only point satisfying the KKT condition is $\boldsymbol{x}^* = [1, 3]^T$ with KKT multiplier $\mu^* = 1$.

(b) The constraint $g(\boldsymbol{x})$ is active at $\boldsymbol{x}^*$, so

$$T(\boldsymbol{x}^*) = \{\boldsymbol{y} \in \mathbb{R}^2 : Dg(\boldsymbol{x}^*)\boldsymbol{y} = 0\} = \{a[1, 0]^T : a \in \mathbb{R}\},$$

and

$$N(\boldsymbol{x}^*) = \{\boldsymbol{y} : \boldsymbol{y} = Dg(\boldsymbol{x}^*)^T z,\ z \in \mathbb{R}\} = \{a[0, 1]^T : a \in \mathbb{R}\}.$$

Since $\mu^* > 0$, $\tilde{J}(\boldsymbol{x}^*, \mu^*) = J(\boldsymbol{x}^*)$ in this case. So, $\tilde{T}(\boldsymbol{x}^*, \mu^*) = T(\boldsymbol{x}^*) = \{a[1, 0]^T : a \in \mathbb{R}\}$.

(c)

$$\boldsymbol{L}(\boldsymbol{x}^*, \mu^*) = \begin{bmatrix} -2 & 0 \\ 0 & 0 \end{bmatrix}$$

which is negative semidefinite. So, we cannot conclude anything immediately, and need to check the sign of $\boldsymbol{y}^T \boldsymbol{L}(\boldsymbol{x}^*, \mu^*)\boldsymbol{y}$ for an arbitrary $\boldsymbol{y} \in T(\boldsymbol{x}^*)$ to see if $\boldsymbol{x}^*$ satisfies the SONC. In this case, we get for $\boldsymbol{y} = a[1, 0]^T$,

$$\boldsymbol{y}^T \boldsymbol{L}(\boldsymbol{x}^*, \mu^*)\boldsymbol{y} = -2a^2$$

which is < 0 when $a \neq 0$ (for instance, when $a = 1$). Thus, $\boldsymbol{x}^*$ does not satisfy the SONC, and is not a local minimizer.

A6. (a) Write down the KKT conditions and find the KKT multipliers for the point $\boldsymbol{x}^* = \boldsymbol{0}$. You should get $\mu_1 = -2$, $\mu_2 = 0$, and $\mu_3 = -4$. Hence, $\boldsymbol{x}^*$ satisfies the FONC for maximization, but not for minimization.

(b) The Hessian of the Lagrangian in this case is:

$$\boldsymbol{L}(\boldsymbol{x}^*, \boldsymbol{\mu}^*) = \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix},$$

so we cannot conclude anything immediately. To check whether $\boldsymbol{x}^*$ satisfies the SOSC, we need to find $\tilde{T}(\boldsymbol{x}^*, \boldsymbol{\mu}^*)$. Here, we have $\tilde{J}(\boldsymbol{x}^*, \boldsymbol{\mu}^*) = \{1, 3\}$, so

$$\tilde{T}(\boldsymbol{x}^*, \boldsymbol{\mu}^*) = \{\boldsymbol{y} \in \mathbb{R}^2 : Dg_1(\boldsymbol{x}^*)\boldsymbol{y} = 0,\ Dg_3(\boldsymbol{x}^*)\boldsymbol{y} = 0\}.$$

Solving, we get

$$\tilde{T}(\boldsymbol{x}^*, \boldsymbol{\mu}^*) = \{\boldsymbol{0}\}.$$

Hence, the SOSC is vacuously satisfied by $\boldsymbol{x}^*$, and $\boldsymbol{x}^*$ is a strict local maximizer.

A7. The KKT conditions are:

$$\begin{aligned} (\boldsymbol{x}^* - \boldsymbol{x}_0) + \mu^* \nabla g(\boldsymbol{x}^*) &= 0 \\ \mu^* g(\boldsymbol{x}^*) &= 0. \end{aligned}$$

Multiply on the left by $(\boldsymbol{x}^* - \boldsymbol{x}_0)^T$ to get

$$\|\boldsymbol{x}^* - \boldsymbol{x}_0\|^2 + \mu^*(\boldsymbol{x}^* - \boldsymbol{x}_0)^T \nabla g(\boldsymbol{x}^*) = 0.$$

Since $\|\boldsymbol{x}^* - \boldsymbol{x}_0\|^2 > 0$ (why?) and $\mu^* \geq 0$, condition (iii) must be true, with $\mu^* > 0$. Hence, condition (ii) is also true. These are the only ones that hold.

A8. (1) By definition of $J(\boldsymbol{x}^*)$, we have $g_i(\boldsymbol{x}^*) < 0$ for all $i \notin J(\boldsymbol{x}^*)$. Since by assumption g_i is continuous for all i, there exists $\epsilon > 0$ such that $g_i(\boldsymbol{x}) < 0$ for all $i \in J(\boldsymbol{x}^*)$ and all $\boldsymbol{x}$ such that $\|\boldsymbol{x} - \boldsymbol{x}^*\| < \epsilon$.

(2) Clearly, $S \cap B \subseteq S_1 \cap B$. To show that $S_1 \cap B \subseteq S \cap B$, suppose that $\boldsymbol{x} \in S_1 \cap B$. Then, by definition of S_1 and B, we have $\boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}$, $g_j(\boldsymbol{x}) \leq 0$ for all $j \in J(\boldsymbol{x}^*)$, and $g_i(\boldsymbol{x}^*) < 0$ for all $i \notin J(\boldsymbol{x}^*)$. Hence, $x \in S \cap B$.

(3) Since $\boldsymbol{x}^*$ is a local minimizer of f over S, and $S \cap B \subseteq S$, $\boldsymbol{x}^*$ is also a local minimizer of f over $S \cap B = S_1 \cap B$. Hence, we conclude that $\boldsymbol{x}^*$ is a regular local minimizer of f on S_1.

(4) Since $S' \subseteq S_1$ and $\boldsymbol{x}^* \in S'$, we conclude that $\boldsymbol{x}^*$ is a regular local minimizer of f on S'.

A9. (a) Here, $\nabla f(\boldsymbol{x}) = \boldsymbol{c}$ and $\nabla g(\boldsymbol{x}) - \boldsymbol{x}$. Note that $\boldsymbol{x}^* \neq \boldsymbol{0}$ (why?), and therefore it is regular. By the KKT Theorem, there exists $\mu^* \geq 0$ such that $\boldsymbol{c} = -\mu^* \boldsymbol{x}^*$ and $\mu^* g(\boldsymbol{x}^*) = 0$. Since $\boldsymbol{c} \neq \boldsymbol{0}$, we must have $\mu^* \neq 0$. Hence, $g(\boldsymbol{x}^*) = 0$, so $\frac{1}{2}\|\boldsymbol{x}^*\|^2 - 1 = 0$.

(b) Since $\|\boldsymbol{e}\|^2 = n$ and $\alpha^2 \|\boldsymbol{e}\|^2 = 2$ from the previous part, we have $\alpha = (2.n)^{1/2}$.

To find $\boldsymbol{c}$, use $f(\boldsymbol{x}^*) = 4$ and $\boldsymbol{c} = \mu^* \boldsymbol{x}^*$ to find μ^* and then $\boldsymbol{c}$. You should get $\mu^* = -2$ and $\boldsymbol{c} = (-2(2/n)^{1/2})\boldsymbol{e}$.

A10. See Theorem 22.8 in [CZ13]. Also see Lecture 24 slides.

A11. See Theorem 22.9 in [CZ13]. Also see Lecture 24 slides.

Tutorial Sheet 9

Q1. Consider the following standard form LP problem:

$$\begin{aligned} \text{minimize} \quad & 2x_1 - x_2 - x_3 \\ \text{subject to} \quad & 3x_1 + x_2 + x_4 = 4 \\ & 6x_1 + 2x_2 + x_3 + x_4 = 5 \\ & x_1, x_2, x_3, x_4 \geq 0. \end{aligned}$$

(a) Write down the $\boldsymbol{A}$, $\boldsymbol{b}$, and $\boldsymbol{c}$ matrices/vectors for the problem.

(b) Consider the basis consisting of the third and fourth columns of $\boldsymbol{A}$, ordered according to $[\boldsymbol{a}_4, \boldsymbol{a}_3]$. Compute the canonical tableau corresponding to this basis.

(c) Write down the basic feasible solution corresponding to the basis above, and its objective function value.

(d) Write down the values of the reduced cost coefficients (for all the variables) corresponding to the basis.

(e) Is the basic feasible solution in part (c) an optimal feasible solution? If yes, explain why. If not, determine which element of the canonical tableau to pivot about so that the new basic feasible solution will have a lower objective function value.

Q2. Use the simplex method to solve the following linear program:

$$\begin{aligned} \text{maximize} \quad & x_1 + x_2 + 3x_3 \\ \text{subject to} \quad & x_1 + x_3 = 1 \\ & x_2 + x_3 = 2 \\ & x_1, x_2, x_3 \geq 0. \end{aligned}$$

Q3. Consider the linear program

$$\begin{aligned} \text{maximize} \quad & 2x_1 + x_2 \\ \text{subject to} \quad & 0 \leq x_1 \leq 5 \\ & 0 \leq x_2 \leq 7 \\ & x_1 + x_2 \leq 9. \end{aligned}$$

Convert the problem to standard form and solve it using the simplex method.

Q4. Consider the problem

$$\begin{aligned} \text{maximize} \quad & -x_1 - 2x_2 \\ \text{subject to} \quad & x_1 \geq 0 \\ & x_2 \geq 1. \end{aligned}$$

(a) Convert the problem into a standard form linear programming problem.

(b) Use the simplex method to compute the solution to this problem and the value of the objective function at the optimal solution of the problem.

Solution outlines

A1. (a)

$$A = \begin{bmatrix} 3 & 1 & 0 & 1 \\ 6 & 2 & 1 & 1 \end{bmatrix}, \quad b = \begin{bmatrix} 4 \\ 5 \end{bmatrix}, \quad c = [2, -1, -1, 0].$$

(b) Pivoting about $(1,4)$ and $(2,3)$, we obtain

$$\begin{array}{ccccc} 3 & 1 & 0 & 1 & 4 \\ 3 & 1 & 1 & 0 & 1 \\ 5 & 0 & 0 & 0 & 1 \end{array}$$

(c) $\boldsymbol{x} = [0,0,1,4]^T$, $\boldsymbol{x}^T\boldsymbol{c} = -1$.

(d) $[r_1, r_2, r_3, r_4] = [5,0,0,0]$

(e) Since reduced cost solution is ≥ 0, feasible solution in part (c) is optimal.

A2. Write down the problem in standard form and form its tableau. Performing the necessary row operations to bring it into canonical form, you can get:

$$\begin{array}{cccc} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 2 \\ 0 & 0 & -1 & 3 \end{array}$$

Pivot about the $(1,3)$th element, and complete the solution. The optimal solution is $[0,1,1]^T$ and optimal cost is 4.

A3. Again, bring the system into the standard form, as:

$$\begin{array}{ll} \text{minimize} & -2x_1 - x_2 \\ \text{subject to} & \begin{array}{l} x_1 + x_3 = 5 \\ x_2 + x_4 = 7 \\ x_1 + x_2 + x_5 = 9 \\ x_1, \ldots, x_5 \geq 0. \end{array} \end{array}$$

Tableau:

$$\begin{array}{cccccc} 1 & 0 & 1 & 0 & 0 & 5 \\ 0 & 1 & 0 & 1 & 0 & 7 \\ 1 & 1 & 0 & 0 & 1 & 9 \\ -2 & -1 & 0 & 0 & 0 & 0 \end{array}$$

Pivot about the $(1,1)$th element. After row operations,

$$\begin{array}{cccccc} 1 & 0 & 1 & 0 & 0 & 5 \\ 0 & 1 & 0 & 1 & 0 & 7 \\ 0 & 1 & -1 & 0 & 1 & 4 \\ 0 & -1 & 2 & 0 & 0 & 10 \end{array}$$

Pivot about the $(3,2)$th element. After appropriate row operations:

$$\begin{array}{cccccc} 1 & 0 & 1 & 0 & 0 & 5 \\ 0 & 0 & 1 & 1 & -1 & 3 \\ 0 & 1 & -1 & 0 & 1 & 4 \\ 0 & 0 & 1 & 0 & 1 & 14 \end{array}$$

Therefore solution is $[5,4,0,3,0]^T$ and optimal cost is 14.

A4. (a) In standard form:

$$\begin{array}{ll} \text{minimize} & x_1 + 2x_2 \\ \text{subject to} & x_2 - x_3 = 1 \\ & x_1, x_2, x_3 \geq 1. \end{array}$$

(b) Final answer: The optimal solution to the original problem is $[0,1]^T$ with objective function value -2.