Tutorial Sheets

Tutorial Sheet 1

Q1. Is the quadratic form

$$\boldsymbol{x}^T \begin{pmatrix} 1 & -7 \\ 1 & 1 \end{pmatrix} \boldsymbol{x}$$

positive definite, positive semidefinite, negative definite, negative semidefinite or indefinite?

Q2. Consider the quadratic form

$$f(x_1, x_2, x_3) = x_1^2 + x_2^2 + 5x_3^2 - 2\xi x_1 x_2 - 2x_1 x_3 + 4x_2 x_3.$$

Find the values of ξ for which the quadratic form is positive definite.

Q3. Show that for any matrix $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ and vector $\boldsymbol{b} \in \mathbb{R}^m$, the set (linear variety) $\{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}\}$ is convex.

Q4. Define the functions $f\colon \mathbb{R}^2 \to \mathbb{R}$ and $g\colon \mathbb{R} \to \mathbb{R}^2$ by $f(\boldsymbol{x}) = x_1^2/6 + x_2^2/4$, $\boldsymbol{g}(t) = [9t + 10, 6t - 5]^T$. Let $F\colon \mathbb{R} \to \mathbb{R}$ be given by $F(t) = f(g(t))$. Evaluate $\frac{dF}{dt}(t)$ using the chain rule.

Q5. Write down the Taylor series expansion of the following function about the given point x_0. Neglect the terms of order 3 or higher.

$$f(\boldsymbol{x}) = x_1 e^{-x_2} + x_2 + 1, \;\; \boldsymbol{x}_0 = [1, 0]^T.$$

Q6. Consider the problem

$$\begin{array}{ll} \text{minimize} & f(\boldsymbol{x}) \\ \text{subject to} & \boldsymbol{x} \in \Omega, \end{array}$$

where $f \in \mathcal{C}^2$. For each of the following specifications for Ω, $\boldsymbol{x}^*$ and f, determine if the given point $\boldsymbol{x}^*$ is: (i) definitely a local minimizer; (ii) definitely not a local minimizer; or (iii) possibly a local minimizer. Fully justify your answer.

(a) $f\colon \mathbb{R}^2 \to \mathbb{R}$, $\Omega = \{\boldsymbol{x} = [x_1, x_2]^T : x_1 \geq 1\}$, $\boldsymbol{x}^* = [1, 2]^T$, and gradient $\nabla f(\boldsymbol{x}^*) = [1, 1]^T$.

(b) $f\colon \mathbb{R}^2 \to \mathbb{R}$, $\Omega = \{\boldsymbol{x} = [x_1, x_2]^T : x_1 \geq 1, x_2 \geq 2\}$, $\boldsymbol{x}^* = [1, 2]^T$, and gradient $\nabla f(\boldsymbol{x}^*) = [1, 0]^T$.

(c) $f\colon \mathbb{R}^2 \to \mathbb{R}$, $\Omega = \{\boldsymbol{x} = [x_1, x_2]^T : x_1 \geq 0, x_2 \geq 0\}$, $\boldsymbol{x}^* = [1, 2]^T$, gradient $\nabla f(\boldsymbol{x}^*) = [0, 0]^T$, and Hessian $\boldsymbol{F}(\boldsymbol{x}^*) = \boldsymbol{I}$.

(d) $f\colon \mathbb{R}^2 \to \mathbb{R}$, $\Omega = \{\boldsymbol{x} = [x_1, x_2]^T : x_1 \geq 1, x_2 \geq 2\}$, $\boldsymbol{x}^* = [1, 2]^T$, gradient $\nabla f(\boldsymbol{x}^*) = [1, 0]^T$, and Hessian

$$\boldsymbol{F}(\boldsymbol{x}^*) = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Q7. Suppose that x^* is a local minimizer of f over Ω, and $\Omega \subset \Omega'$. Show that if x^* is an interior point of Ω, then x^* is a local minimizer of f over Ω'. Show that the same conclusion cannot be made if x^* is not an interior point of Ω.

Q8. Consider the problem

$$\begin{array}{ll} \text{minimize} & -x_2^2 \\ \text{subject to} & |x_2| \leq x_1^2 \\ & x_1 \geq 0, \end{array}$$

where $x_1, x_2 \in \mathbb{R}$.

(a) Does the point $[x_1, x_1]^T = \mathbf{0}$ satisfy the first-order necessary condition for a minimizer? That is, if f is the objective function, is it true that $\boldsymbol{d}^T \nabla f(\mathbf{0}) \geq 0$ for all feasible directions $\boldsymbol{d}$ at $\mathbf{0}$?

(b) Is the point $[x_1, x_1]^T = \mathbf{0}$ a local minimizer, a strict local minimizer, a local maximizer, a strict local maximizer, or none of the above?

Q9. Figure 1 shows a simplified model of a fetal heart monitoring system (the distances shown have been scaled down to make the calculations simpler). A heartbeat sensor is located at position x (see Figure 1).

The energy of the heartbeat signal measured by the sensor is the reciprocal of the squared distance from the source (baby's heart or mother's heart). Find the position of the sensor that maximizes the *signal-to-interference ratio*, which is the ratio of the signal energy from the baby's heart to the signal energy from the mother's heart.

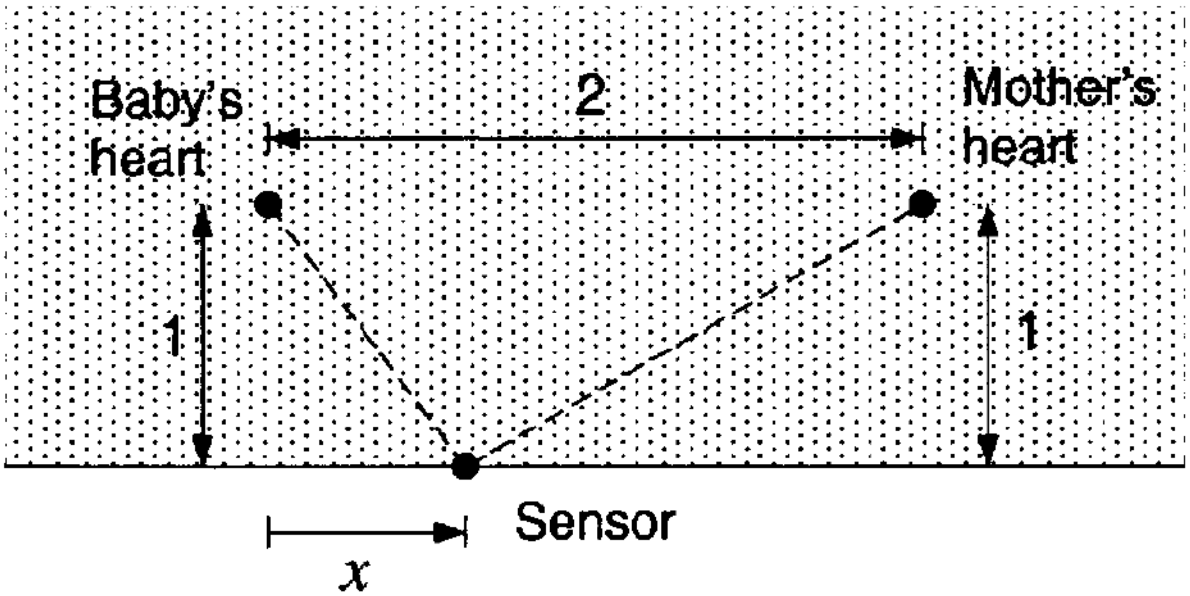

Figure 1: Simplified fetal heart monitoring system for Question 9. (SOURCE: Chap. 6 [CZ13])

Q10. Consider the problem:

$$\begin{array}{ll} \text{maximize} & c_1x_1 + c_2x_2 \\ \text{subject to} & x_1 + x_2 \leq 1 \\ & x_1, x_2 \geq 0, \end{array}$$

where c_1 and c_2 are constants such that $c_1 > c_2 \geq 0$. The above is a *linear programming* problem. Assuming that the problem has an optimal feasible solution, use the *First-Order Necessary Conditions* to show that the *unique* optimal feasible solution x^* is $[1, 0]^T$.

Hint: First show that x^* cannot lie in the interior of the constraint set. Then, show that x^* cannot lie on the line segments $L_1 = \{x : x_1 = 0, 0 \leq x_2 \leq 1\}$, $L_2 = \{x : 0 \leq x_1 < 1, x_2 = 0\}$, $L_3 = \{x : 0 \leq x_1 < 1, x_2 = 1 - x_1\}$.

Q11. *Line Fitting.* Let $[x_1, y_1]^T, \ldots, [x_n, y_n]^T$, $n \geq 2$, be points on the $\mathbb{R}^2$ plane (each $x_i, y_i \in \mathbb{R}$. We wish to find the straight line of "best fit" through these points ("best" in the sense that the average squared error is minimized); that is, we wish to find $a, b \in \mathbb{R}$ to minimize

$$f(a, b) = \frac{1}{n} \sum_{i=1}^{n} (ax_i + b - y_i)^2.$$

(a) Let

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$$
$$\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$
$$\overline{X^2} = \frac{1}{n}\sum_{i=1}^{n} x_i^2$$
$$\overline{Y^2} = \frac{1}{n}\sum_{i=1}^{n} y_i^2$$
$$\overline{XY} = \frac{1}{n}\sum_{i=1}^{n} x_i y_i.$$

Show that $f(a,b)$ can be written in the form $\boldsymbol{z}^T\boldsymbol{Q}\boldsymbol{z} - 2\boldsymbol{c}^T\boldsymbol{z} + d$, where $\boldsymbol{z} = [a,b]^T$, $\boldsymbol{Q} = \boldsymbol{Q}^T \in \mathbb{R}^{2\times 2}$, $\boldsymbol{c} \in \mathbb{R}^2$ and $d \in \mathbb{R}$, and find expressions for $\boldsymbol{Q}$, $\boldsymbol{c}$, and d in terms of $\overline{X}$, $\overline{Y}$, $\overline{X^2}$, $\overline{Y^2}$, and $\overline{XY}$.

(b) Assume that the x_i, $i = 1, \ldots, n$, are not all equal. Find the parameters a^* and b^* for the line of best fit in terms of $\overline{X}$, $\overline{Y}$, $\overline{X^2}$, $\overline{Y^2}$, and $\overline{XY}$. Show that the point $[a^*, b^*]^T$ is the only local minimizer of f.
Hint: $\overline{X^2} - (\overline{X})^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{X})^2$.

(c) Show that if a^* and b^* are the parameters of the line of best fit, then $\overline{Y} = a^*\overline{X} + b^*$ (and hence once we have computed a^*, we can compute b^* using the formula $b^* = \overline{Y} - a^*\overline{X}$).

Solution outlines

A1. Indefinite: if $f(\boldsymbol{x})$ is the given quadratic form, then observe that $f([1,0]^T) > 0$ and $f([1,1]^T) < 0$. Note that the *symmetric* matrix $\boldsymbol{Q}$ associated to the quadratic form is $\frac{1}{2}(\boldsymbol{M} + \boldsymbol{M}^T)$, where $\boldsymbol{M} = \begin{pmatrix} 1 & -7 \\ 1 & 1 \end{pmatrix}$. In particular, it is incorrect to try and apply Sylvestor's criterion to the matrix $\boldsymbol{M}$.

A2. The matrix $\boldsymbol{Q}$ in this case is $\begin{pmatrix} 1 & -\epsilon & -1 \\ -\epsilon & 1 & 2 \\ -1 & 2 & 5 \end{pmatrix}$.

The leading principal minors are 1, $1 - \epsilon^2$ and $-5\epsilon^2 + 4\epsilon$. We will have positive definiteness when all the leading principal minors are positive and this happens when $\epsilon \in (0, \frac{4}{5})$.

A3. Let $\boldsymbol{u}$ and $\boldsymbol{v} \in \mathbf{S} = \{\boldsymbol{x} : \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}\}$. Note that this implies $\boldsymbol{A}\boldsymbol{u} = \boldsymbol{b}$ and $\boldsymbol{A}\boldsymbol{v} = \boldsymbol{b}$. Let $\boldsymbol{z} = \alpha\boldsymbol{u} + (1-\alpha)\boldsymbol{v}$. To prove that $\mathbf{S}$ is convex, we just need to prove that $\boldsymbol{A}\boldsymbol{z} = \boldsymbol{b}$. Substitute the expression of $\boldsymbol{z}$ and simplify.

A4. $Df(\boldsymbol{x}) = [\frac{x_1}{3}, \frac{x_2}{2}]$
$\frac{d\boldsymbol{g}}{dt} = [9, 6]^T$
Chain Rule: $\frac{dF(t)}{dt} = Df(\boldsymbol{g}(t))\frac{d\boldsymbol{g}}{dt}(t) = 45t + 15.$

A5. Taylor series expansion is

$$f(\boldsymbol{x}) = f(\boldsymbol{x}_0) + Df(\boldsymbol{x}_0)(\boldsymbol{x} - \boldsymbol{x}_0) + \frac{(\boldsymbol{x} - \boldsymbol{x}_0)^T D^2 f(\boldsymbol{x}_0)(\boldsymbol{x} - \boldsymbol{x}_0)}{2} + \cdots$$

Compute $Df(\boldsymbol{x}_0)$ and $D^2 f(\boldsymbol{x}_0)$. Final answer is $1 + x_1 + x_2 - x_1x_2 + \frac{x_2^2}{2} + \cdots$

A6. (a) $\boldsymbol{x}^*$ is definitely not a local minimizer. Consider $\boldsymbol{d} = [1, -2]^T$. $\boldsymbol{d}$ is a feasible direction but FONC is not satisfied.

(b) In this case $\boldsymbol{x}^*$ satisfies FONC and hence is possibly a local minimizer but we can't be sure based on given information.

(c) In this case $\boldsymbol{x}^*$ satisfies SOSC and hence is definitely a local minimzer.

(d) In this case $\boldsymbol{x}^*$ is definitely not a local minimizer as the SONC conditions are not satisfied for $\boldsymbol{d} = [0, 1]^T$.

A7. Suppose $\boldsymbol{x}^*$ is an interior point of Ω. Then there exists $\epsilon' > 0$ such that $\{\boldsymbol{y} : \|\boldsymbol{y} - \boldsymbol{x}^*\| < \epsilon'\} \subset \Omega$. Also since $\boldsymbol{x}^*$ is a local minimizer of f over Ω there exists $\epsilon'' > 0$ such that $f(\boldsymbol{x}^*) \leq f(\boldsymbol{x})$ whenever $\boldsymbol{x} \in \{\boldsymbol{y} : \|\boldsymbol{y} - \boldsymbol{x}^*\| < \epsilon''\}$. Take $\epsilon = \min\{\epsilon', \epsilon''\}$. Then $\boldsymbol{x}^*$ is an interior point of Ω' and a minimizer of f over Ω'.
For the second part consider $\Omega = \{0\}$, $\Omega' = [-1, 1]$ and $f(x) = x$ as the counterexample.

A8. (a) In this problem the only feasible directions at $\mathbf{0}$ are of the form $\boldsymbol{d} = [d_1, 0]$. Hence, $\boldsymbol{d}^T \nabla f(\mathbf{0}) = 0$ for all feasible directions $\boldsymbol{d}$ at $\mathbf{0}$.

(b) The point $\mathbf{0}$ is not a strict local maximizer because for any $\boldsymbol{x}$ of the form $\boldsymbol{x} = [x_1, 0]$, we have $f(\boldsymbol{x}) = 0 = f(\mathbf{0})$, and there are such points in any neighborhood of $\mathbf{0}$.
The point 0 is not a local minimizer because for any point $\boldsymbol{x}$ of the form $\boldsymbol{x} = [x_1, x_1^2]$ with $x_1 > 0$, we have $f(\boldsymbol{x}) = -x_1^4 < 0$ and there are such points in any neighborhood of $\mathbf{0}$. Since $\mathbf{0}$ is not a local minimizer, it is also not a strict local minimizer.

A9. The signal-to-interference ratio when the sensor is placed at position x is $f(x) = \frac{1+(2-x)^2}{1+x^2}$. Compute the critical points of this function to find the optimal position. Final answer: $x^* = 1 - \sqrt{2}$.

A10. The objective function is $f(\boldsymbol{x}) = -(c_1 x_1 + c_2 x_2)$. Therefore, $\nabla f(x) = [-c_1, -c_2]^T \neq \mathbf{0}$ for all $\boldsymbol{x}$. Thus, by FONC, the optimal solution $\boldsymbol{x}^*$ cannot lie in the interior of the feasible set. Next, for all $\boldsymbol{x} \in L_1 \bigcup L_2$, $\boldsymbol{d} = [1, 1]^T$ is a feasible direction. Therefore, $\boldsymbol{d}^T \nabla f(\boldsymbol{x}) = -c_1 - c_2 < 0$. Hence, by FONC, the optimal solution $\boldsymbol{x^*}$ cannot lie in $L_1 \bigcup L_2$. Lastly, for all $\boldsymbol{x} \in L_3$, $\boldsymbol{d} = [1, -1]^T$ is a feasible direction. Therefore, $\boldsymbol{d}^T \nabla f(\boldsymbol{x}) = c_2 - c_1 < 0$. Hence, by FONC, the optimal solution $\boldsymbol{x}^*$ cannot lie in L_3. Therefore, by elimination, the unique optimal feasible solution must be $[1, 0]^T$.

A11. (a) Expanding $f(a, b)$ we get

$$f(a,b) = \begin{pmatrix} a & b \end{pmatrix} \begin{pmatrix} \frac{\sum_{i=1}^n x_i^2}{n} & \frac{\sum_{i=1}^n x_i}{n} \\ \frac{\sum_{i=1}^n x_i}{n} & 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} - 2 \begin{pmatrix} \frac{\sum_{i=1}^n x_i y_i}{n} & \frac{\sum_{i=1}^n y_i}{n} \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} + \frac{\sum_{i=1}^n y_i^2}{n}.$$

Identify $\boldsymbol{z}$, $\boldsymbol{Q}$, $\boldsymbol{c}$ and $\boldsymbol{d}$ from here.

(b) If point $\boldsymbol{z}^*$ is a solution then by FONC we get that $\boldsymbol{Q}\boldsymbol{z}^* = \boldsymbol{c}$ and since $\det(\boldsymbol{Q}) \neq 0$ we have $\boldsymbol{Q}$ non-singular and $\boldsymbol{z^*} = \boldsymbol{Q}^{-1}\boldsymbol{c}$.
Use SOSC to show that z^* is a strict local minimizer, and use this and FONC to argue that it must also be the only local minimizer.

(c) Substitute the values of a^* and b^* found previously into $a^*\overline{X} + b^*$ to show that it equals $\overline{Y}$.

Tutorial Sheet 2

Q1. Find the range of values of the parameter α for which the function

$$f(x_1, x_2, x_3) = 2x_1x_3 - x_1^2 - x_2^2 - 5x_3^2 + 2\alpha x_1 x_2 - 4x_2x_3$$

is concave.

Q2. Consider the function

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

where $\boldsymbol{Q} = \boldsymbol{Q}^T > 0$ and $\boldsymbol{x}, \boldsymbol{b} \in \mathbb{R}^n$. Define the function $\phi\colon \mathbb{R} \to \mathbb{R}$ by $\phi(\alpha) = f(\boldsymbol{x} + \alpha\boldsymbol{d})$, where $\boldsymbol{x}, \boldsymbol{d} \in \mathbb{R}^n$ are fixed vectors and $\boldsymbol{d} \neq \boldsymbol{0}$. Show that $\phi(\alpha)$ is a strictly convex quadratic function of α.

Q3. Show that $f(\boldsymbol{x}) = 2x_1x_2$ is a convex function on $\Omega = \{[a, ma]^T : a \in \mathbb{R}\}$, where m is any given nonnegative constant.

Q4. Suppose that the set $\Omega = \{\boldsymbol{x} : h(\boldsymbol{x}) = c\}$ is convex, where $h\colon \mathbb{R}^n \to \mathbb{R}$ and $c \in \mathbb{R}$. Show that h is convex and concave over Ω.

Q5. Suppose that we have a convex optimization problem on $\mathbb{R}^3$.

(a) Consider the following three feasible points: $[1, 0, 0]^T$, $[0, 1, 0]^T$, $[0, 0, 1]^T$. Suppose that all three have objective function value 1. What can you say about the objective function value of the point $(1/3)[1, 1, 1]^T$? Explain fully.

(b) Suppose we know that the three points in part (a) are global minimizers. What can you say about the point $(1/3)[1, 1, 1]^T$? Explain fully.

Q6. Let $\Omega \subset \mathbb{R}^n$. Suppose that $f, f_1, f_2\colon \Omega \to \mathbb{R}$ are convex functions, and $a \geq 0$ is any real number. Show that:

(a) af is convex;

(b) $f_1 + f_2$ is convex;

(c) $\max\{f_1, f_2\}$ is convex.

Note: The notation $\max\{f_1, f_2\}$ denotes a function from Ω to $\mathbb{R}$ such that for each $\boldsymbol{x} \in \Omega$, its value is the larger one between the numbers $f_1(\boldsymbol{x})$ and $f_2(\boldsymbol{x})$.

Q7. Let $\Omega \subset \mathbb{R}^n$ be an open convex set. Show that a symmetric matrix $\boldsymbol{Q} \in \mathbb{R}^{n\times n}$ is positive semidefinite if and only if for each $\boldsymbol{x}, \boldsymbol{y} \in \Omega$,

$$(\boldsymbol{x} - \boldsymbol{y})^T\boldsymbol{Q}(\boldsymbol{x} - \boldsymbol{y}) \geq 0.$$

Show that a similar result for positive definiteness holds if we replace "$\geq$" by "$>$" in the inequality above.

Q8. Consider the problem

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|^2 \\ \text{subject to} & x_1 + \cdots + x_n = 1 \\ & x_1, \ldots, x_n \geq 0. \end{array}$$

Is the problem a convex optimization problem? If yes, give a complete proof. If no, explain why not, giving complete explanations.

Q9. Consider the optimization problem

$$\begin{array}{ll} \text{minimize} & f(\boldsymbol{x}) \\ \text{subject to} & \boldsymbol{x} \in \Omega, \end{array}$$

where $f(\boldsymbol{x}) = x_1 x_2^2$, where $\boldsymbol{x} = [x_1, x_2]^T$, and $\Omega = \{\boldsymbol{x} \in \mathbb{R}^2 : x_1 = x_2,\ x_1 \geq 0\}$. Show that the problem is a convex optimization problem.

Q10. Consider the convex optimization problem

$$\begin{array}{ll} \text{minimize} & f(\boldsymbol{x}) \\ \text{subject to} & \boldsymbol{x} \in \Omega. \end{array}$$

Suppose that the points $\boldsymbol{y} \in \Omega$ and $\boldsymbol{z} \in \Omega$ are local minimizers. Determine the largest set of points $G \subset \Omega$ for which you can be sure that every point in G is a global minimizer.

Solution outlines

A1. $f(x_1, x_2, x_3)$ is concave is the same as $-f(x_1, x_2, x_3)$ is convex. Therefore question reduces to figuring out value of α for which $-f(x_1, x_2, x_3)$ is convex. Write $-f(x_1, x_2, x_3)$ as a quadratic form and use the quadratic form sufficient condition to determine values of α. $\boldsymbol{Q}$ is $\begin{pmatrix} 1 & -\alpha & -1 \\ -\alpha & 1 & 2 \\ -1 & 2 & 5 \end{pmatrix}$ and α can take values in the range $[0, \frac{4}{5}]$.

A2. $\frac{d^2\phi}{d\alpha^2} = \boldsymbol{d}^T \boldsymbol{Q} \boldsymbol{d} > 0$ as $\boldsymbol{Q} > 0$.

A3. Write $f(\boldsymbol{x})$ in the form $\boldsymbol{x}^T \boldsymbol{Q} \boldsymbol{x}$. Then use the quadratic form sufficient condition taking $\boldsymbol{x} = [a, ma]^T$ and $\boldsymbol{y} = [b, mb]^T$.

A4. For establishing convexity of h, verify the condition $h(\alpha\boldsymbol{x} + (1-\alpha)\boldsymbol{y}) \leq \alpha h(\boldsymbol{x}) + (1-\alpha)h(\boldsymbol{y})$. Proving h is concave is the same as proving $-h$ is convex. So verify the above condition for $-h$.

A5. (a) Let f be the objective function and Ω be the constrained set. Consider the set $\tau = \{\boldsymbol{x} \in \Omega : f(\boldsymbol{x}) \leq 1\}$. This contains all of the three points. Somehow write the point $\frac{1}{3}[1, 1, 1]^T$ as a linear combination of other three points. This and the fact that τ is convex enables us to conclude that the value of the objective function at $\frac{1}{3}[1, 1, 1]^T$ is ≤ 1.

(b) If all three points are global minimizers then so is the point $\frac{1}{3}[1, 1, 1]^T$.

A6. (a) Define $g(\boldsymbol{x}) = \alpha f(\boldsymbol{x})$. Use the convexity of f to prove the convexity of g.

(b) Define $g(\boldsymbol{x}) = f_1(\boldsymbol{x}) + f_2(\boldsymbol{x})$. Use the convexity of f_1 and f_2 to prove convexity of g. (*Hint:* "Use the convexity of f" means apply the definition of convexity on f).

(c) Define $\bar{f} = \max\{f_1, f_2, \ldots, f_n\}$. Consider $\bar{f}(\alpha\boldsymbol{x} + (1-\alpha)\boldsymbol{y})$ and use convexity of each f_i to prove the result.

A7. ($\Longrightarrow$) Follows by definition of positive definiteness.
($\Longleftarrow$) We have to show for given $\boldsymbol{d} \in \mathbb{R}^n$, $\boldsymbol{d}^T \boldsymbol{Q} \boldsymbol{d} \geq 0$. Fix $\boldsymbol{x} \in \Omega$. Since Ω is open there exists $\alpha \neq 0$ such that $\boldsymbol{y} = \boldsymbol{x} - \alpha\boldsymbol{d}$. Now use this in the given detail that $(\boldsymbol{y} - \boldsymbol{x})^T \boldsymbol{Q} (\boldsymbol{y} - \boldsymbol{x}) \geq 0$ to get the final answer.

A8. Yes, it is a convex optimization problem. If we write $\frac{1}{2}\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|^2 = \frac{1}{2}\boldsymbol{x}^T(\boldsymbol{A}^T\boldsymbol{A})\boldsymbol{x} - (\boldsymbol{b}^T\boldsymbol{A})\boldsymbol{x} +$ constant then we see that the Hessian $\boldsymbol{A}^T\boldsymbol{A}$ is positive semi-definite. Use the definition of convexity to prove that the constraint set is convex.

A9. Use the definition of convexity to prove that the function as well as the constraint set is convex.

A10. $G = \alpha\boldsymbol{y} + (1-\alpha)\boldsymbol{z}, 0 \leq \alpha \leq 1$.

Tutorial Sheet 3

Q1. A rock is accelerated to 3, 5 and 6 $\mathrm{m\,s^{-2}}$ by applying forces of 1, 2 and 3 N, respectively. Assuming that Newton's law $F = ma$ holds, where F is the force and a is the acceleration, estimate the mass m of the rock using the least-squares method.

Q2. A spring is stretched to lengths $L = 3,\ 4$ and 5 cm under applied forces of $F = 1,\ 2$ and 4 N, respectively. Assuming that Hooke's law $L = a + bF$ holds, estimate the normal length a and spring constant b using the least-squares approach.

Q3. Suppose that we perform an experiment to calculate the gravitational constant g as follows. We drop a ball from certain height and measure its distance from original point at certain time instants. The results of the experiment are shown in the following table.

Time (seconds)	1.00	2.00	3.00
Distance (meters)	5.00	19.5	44.0

The equation relating the distance s and time t at which s is measured is given by

$$s = \frac{1}{2}gt^2$$

Find a least-squares estimate of g using the experimental results from the table above.

Q4. Suppose that we wish to estimate the value of the resistance R of a resistor. Ohm's law states that if V is the voltage across the resistor and I is the current through the resistor, then $V = IR$. To estimate R, we apply a 1 A current through the resistor and measure the voltage across it. We perform the experiment on n voltage-measuring devices and obtain measurements of $V_1, V_2, \ldots, V_n$. Show that the least-squares estimate of R is simply the average of $V_1, V_2, \ldots, V_n$.

Q5. We are given two mixtures, A and B. Mixture A contains 30 % gold, 40 % silver, and 30 % platinum, whereas mixture B contains 10 % gold, 20 % silver, and 70 % platinum (all percentages of weight). We wish to determine the ratio of the weight of mixture A to the weight of mixture B such that we have as close as possible to a total of 5 oz t of gold, 3 oz t of silver, and 4 oz t of platinum. Formulate and solve the problem using the linear least-squares method.

Q6. Suppose that we take measurements of a sinusoidal signal $y(t) = \sin(\omega t + \theta)$ at times $t_1, t_2, \ldots, t_p$, and obtain values $y_1, y_2, \ldots, y_p$, where $-\pi/2 \le \omega t_i + \theta \le \pi/2$, $i = 1, 2, \ldots, p$, and the t_i are not all equal. We wish to determine the values of the frequency ω and phase θ.

(a) Express the problem as a system of linear equations.

(b) Find the least-squares estimate of ω and θ based on part (a). Use the following notation:

$$\begin{aligned}
\overline{T} &= \frac{1}{p}\sum_{i=1}^{p} t_i, \\
\overline{T^2} &= \frac{1}{p}\sum_{i=1}^{p} t_i^2, \\
\overline{TY} &= \frac{1}{p}\sum_{i=1}^{p} t_i \arcsin y_i, \\
\overline{Y} &= \frac{1}{p}\sum_{i=1}^{p} \arcsin y_i.
\end{aligned}$$

Q7. We are given a point $[x_0, y_0]^T \in \mathbb{R}^2$. Consider the straight line on the $\mathbb{R}^2$ plane given by the equation $y = mx$. Using a least-squares formulation, find the point on the straight line that is closest to the point $[x_0, y_0]$, where the measure of closeness is in terms of the Euclidean norm on $\mathbb{R}^2$.

Hint: The given line can be expressed as the range of the matrix $\boldsymbol{A} = [1, m]^T$.

Q8. Consider the affine function $f\colon \mathbb{R}^n \to \mathbb{R}$ of the form $f(\boldsymbol{x}) = \boldsymbol{a}^T\boldsymbol{x} + c$, where $\boldsymbol{a} \in \mathbb{R}^n$ and $c \in \mathbb{R}$.

(a) We are given a set of p pairs $(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2), \ldots, (\boldsymbol{x}_p, y_p)$, where $\boldsymbol{x}_i \in \mathbb{R}^n$ and $y_i \in \mathbb{R}$. We wish to find the affine function of the best fit to these points, where "best" is in the sense of minimizing the total square error

$$\sum_{i=1}^{n} (f(\boldsymbol{x}_i) - y_i)^2.$$

Formulate the above as an optimization problem of the form: minimize $\|\boldsymbol{A}\boldsymbol{z} - \boldsymbol{b}\|^2$ with respect to $\boldsymbol{z}$. Specify the dimensions of $\boldsymbol{A}$, $\boldsymbol{z}$ and $\boldsymbol{b}$.

(b) Suppose that the points satisfy

$$\boldsymbol{x}_1 + \boldsymbol{x}_2 + \cdots + \boldsymbol{x}_p = \boldsymbol{0}$$

and

$$y_1\boldsymbol{x}_1 + y_2\boldsymbol{x}_2 + \cdots + y_p\boldsymbol{x}_p = \boldsymbol{0}.$$

Find the affine function of best fit in this case, assuming that it exists and is unique.

Q9. Let $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, $\boldsymbol{b} \in \mathbb{R}^m$, $m \le n$, $\operatorname{rank} \boldsymbol{A} = m$, and $\boldsymbol{x}_0 \in \mathbb{R}^n$. Consider the problem

$$\begin{array}{ll} \text{minimize} & \|\boldsymbol{x} - \boldsymbol{x}_0\| \\ \text{subject to} & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}. \end{array}$$

Show that this problem has a unique solution given by

$$\boldsymbol{x}^* = \boldsymbol{A}^T(\boldsymbol{A}\boldsymbol{A}^T)^{-1}\boldsymbol{b} + (\boldsymbol{I}_n - \boldsymbol{A}^T(\boldsymbol{A}\boldsymbol{A}^T)^{-1}\boldsymbol{A})\boldsymbol{x}_0.$$

Q10. Given $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, $m \ge n$, $\operatorname{rank} \boldsymbol{A} = n$, and $\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_p \in \mathbb{R}^m$, consider the problem

$$\text{minimize} \quad \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}_1\|^2 + \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}_2\|^2 + \cdots + \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}_p\|^2.$$

Suppose that $\boldsymbol{x}_i^*$ is a solution to the problem

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

where $i = 1, 2, \ldots, p$. Write the solution to the problem in terms of $\boldsymbol{x}_1^*, \boldsymbol{x}_2^*, \ldots, \boldsymbol{x}_p^*$.

Q11. Let $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, $\boldsymbol{b} \in \mathbb{R}^m$, $m \le n$, and $\operatorname{rank} \boldsymbol{A} = m$. Show that $\boldsymbol{x}^* = \boldsymbol{A}^T(\boldsymbol{A}\boldsymbol{A}^T)^{-1}\boldsymbol{b}$ is the only vector in $\mathcal{R}(\boldsymbol{A}^T)$ satisfying $\boldsymbol{A}\boldsymbol{x}^* = \boldsymbol{b}$.

Solution outlines

A.1. $m^* = \frac{31}{70}$.

A.2. $[a^*, b^*]^T = [\frac{5}{2}, \frac{9}{14}]^T$.

A.3. $g = 9.776$.

A.4. $\boldsymbol{A} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$, $\boldsymbol{z} = [R]$, and $\boldsymbol{b} = [V_1, V_2, \ldots, V_n]^T$.

A.5. $\boldsymbol{A} = \begin{bmatrix} 0.3 & 0.1 \\ 0.4 & 0.2 \\ 0.3 & 0.7 \end{bmatrix}$, $\boldsymbol{z} = [x_1, x_2]^T$, and $\boldsymbol{b} = [5, 3, 4]^T$. Ratio $= 11$.

A.6. Write the equations as $\omega t_i + \theta = \arcsin y_i$. Now identify entities $\boldsymbol{A}$, $\boldsymbol{z}$, $\boldsymbol{b}$ and proceed with the least squares formula.

A.7. Using the given hint, construct the appropriate equation to minimize. The point closest is $[x^*, mx^*]^T$, where x^* is given by $\frac{x_0 + my_0}{1+m^2}$.

A.8. (a) The dimensions of $\boldsymbol{A}$, $\boldsymbol{z}$, and $\boldsymbol{b}$ are $p \times (n+1)$, $(n+1) \times 1$, and $p \times 1$, respectively.

(b) The affine function for best fit is $f(\boldsymbol{x}) = c$, where $c = \frac{1}{p}\sum_{i=1}^{p} y_i$.

A.9. Take $\boldsymbol{z} = \boldsymbol{x} - \boldsymbol{x}_0$. After appropriate changes to the constraint, the problem reduces to the following:

$$\begin{aligned} \text{minimize} \quad & \|\boldsymbol{z}\| \\ \text{subject to} \quad & \boldsymbol{A}\boldsymbol{z} = \boldsymbol{c}, \end{aligned}$$

where $\boldsymbol{c} = \boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}_0$. Now, using the appropriate result, one will get the solution.

A.10. Keep in mind that $\|\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}\|^2 = \boldsymbol{x}^T\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x} - 2\boldsymbol{x}^T\boldsymbol{A}^T\boldsymbol{b} + \|\boldsymbol{b}\|^2$. Use this in the expression and obtain $\boldsymbol{x}^* = \frac{1}{p}\sum_{i=1}^{p} \boldsymbol{x}_i^*$.

A.11. Let $\boldsymbol{y}^*$ be another vector in $\mathcal{R}(\boldsymbol{A}^T)$, which satisfies $\boldsymbol{A}\boldsymbol{y}^* = \boldsymbol{b}$, $\boldsymbol{y}^* \neq \boldsymbol{x}^*$. Since $\text{rank}(\boldsymbol{A}) = m$, this implies

$$\exists\ \boldsymbol{z} \in \mathbb{R}^m \text{ such that } \boldsymbol{y}^* = \boldsymbol{A}^T\boldsymbol{z} \tag{9.1}$$

$$\implies \boldsymbol{A}(\boldsymbol{A}^T\boldsymbol{z}) = \boldsymbol{b} \tag{9.2}$$

$$\text{We also know } \boldsymbol{A}(\boldsymbol{A}^T(\boldsymbol{A}\boldsymbol{A}^T)^{-1}\boldsymbol{b}) = \boldsymbol{b} \tag{9.3}$$

Subtracting (9.3) from (9.2) and substituting, we get the answer.

Tutorial Sheet 4

Q1. Consider the problem to fit the data (t_i, y_i), $i = 1, \ldots, 4$, with the model $m(\boldsymbol{x}, t_i) = e^{t_i x_1} + e^{t_i x_2}$, using nonlinear least squares. That is, minimize $f(\boldsymbol{x}) := \frac{1}{2}\boldsymbol{r}(\boldsymbol{x})^T\boldsymbol{r}(\boldsymbol{x}) = \frac{1}{2}\sum_{i=1}^{4} r_i(\boldsymbol{x})^2$, where $r_i(\boldsymbol{x}) = e^{t_i x_1} + e^{t_i x_2} - y_i$, $i = 1, \ldots, 4$. Compute

$$\nabla f(\boldsymbol{x}),\ J(\boldsymbol{x}),\ S(\boldsymbol{x}),\ \text{and } \nabla^2 f(\boldsymbol{x}).$$

Describe (i) Newton's method (ii) Gauss-Newton method (iii) Levenberg-Marquardt method to solve the minimization problem.

Q2. Let $\boldsymbol{r}\colon \mathbb{R}^4 \to \mathbb{R}^{20}$, $r_i(\boldsymbol{x}) = x_1 + x_2 e^{-(t_i + x_3^2)/x_4} - y_i$, $i = 1, \ldots, 20$, and $f(\boldsymbol{x}) := \frac{1}{2}\boldsymbol{r}(\boldsymbol{x})^T\boldsymbol{r}(\boldsymbol{x}) = \frac{1}{2}\sum_{i=1}^{20} r_i(\boldsymbol{x})^2$. Compute

$$\nabla f(\boldsymbol{x}),\ J(\boldsymbol{x}),\ S(\boldsymbol{x}),\ \text{and } \nabla^2 f(\boldsymbol{x}).$$

Describe (i) Newton's method (ii) Gauss-Newton method (iii) Levenberg-Marquardt method to solve the minimization problem.

Q3. Prove that $[1, 1]^T$ is the unique global minimizer of $f(\boldsymbol{x}) = 100(x_2 - x_1)^2 + (1 - x_1)^2$ and use two iterations of Newton's method to minimize $f(\boldsymbol{x})$ choosing the initial starting point as $[0, 0]^T$.

Q4. Apply Newton's method to minimize $f(x) = x^{4/3}$, $x \in \mathbb{R}$. Show that as long as the initial starting point is not chosen as 0, the algorithm does not converge (no matter how close we start to the minimizer 0).

Q5. Find the minima, maxima and saddle points of $f(x, y) = x^2 - 4xy + y^4$. Is the function coercive?

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 $\boldsymbol{Q}$.