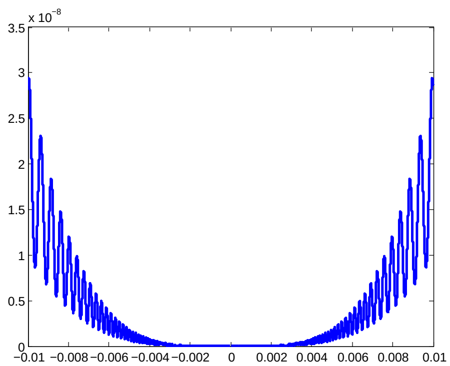
Solutions to Part A of Problem Sheet 1

Solution (1.1)

(a) The function $f(x) = x^4$ has a strict minimum at $x = 0$, but the second derivative satisfies $f''(0) = 0$.

(b) We construct a function that has a strict minimizer x^*, but such that every open neighbourhood U of x^* contains other local minimizers. One such function is

$$f(x) = \begin{cases} x^4(\cos(1/x) + 2) & \text{for } x \neq 0 \\ 0 & \text{for } x = 0. \end{cases}$$



We explain the construction of this function:

1. Start out with $g(x) = \cos(1/x) + 2$ for $x \neq 0$ and $g(0) = 1$. This function has minimizers $x_0 = 0$ and $x_k = 1/(\pi(2k+1)$ for $k \geq 0$, with values $g(x_k) = 1$ at all minimizers. Therefore, any open interval around 0 contains (infinitely many) local minimizers x_k other than $x_0 = 0$.
2. Multipy x^4 to the function: $f(x) = x^4 g(x)$. This ensures that $f(0) = 0$ and $f(x) > 0$ for $x \neq 0$. There are still local minima in every neighbourhood of 0. To see this, compute the derivative

$$f'(x) = x^2(4x\cos(1/x) + \sin(1/x) + 8x). \tag{1}$$

Set $z_m = 1/(\pi/2 + m\pi)$ for $m > 0$. Since $\sin(1/z_m) = \sin(\pi/2 + m\pi) = 1$ for m even and -1 for m odd, and for m sufficiently large the contribution of the other terms is negligible (as the z_m become arbitrary small), the derivative (1) changes signs between successive z_m. Since $f'(x)$ is continuous, it has roots between any z_m and z_{m+1} for large enough m, and these correspond to maxima and minima of f.

The function is in $C^2(\mathbb{R})$. For $x \neq 0$ this is clear, and to verify this at $x = 0$, one shows that the right and left limits as $x \to 0$ of $f'(x)$ and $f''(x)$ coincide (they are in fact 0).

Note the subtle point that one minimizer x^* can have local minimizers that are arbitrary close: while each open interval I surrounding x^* has another local minimizer $\tilde{x}$, every such $\tilde{x}$ has an interval $\tilde{I}$ surrounding it where this $\tilde{x}$ is the only minimizer!

Solution (1.2) (a) We want to show that the function

$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}^\top \boldsymbol{x} + c \tag{2}$$

is convex if and only $\boldsymbol{A}$ is positive semidefinite. To see this, we compute the partial derivatives and the Hessian of f. The parts $\boldsymbol{b}^\top \boldsymbol{x}$ and c disappear when computing second derivatives. The vector $\boldsymbol{A}\boldsymbol{x}$ has entries $(\boldsymbol{A}\boldsymbol{x})_i = \sum_{j=1}^n a_{ij}x_j$ for $1 \leq i \leq n$, where a_{ij} are the entries of $\boldsymbol{A}$. Therefore, the function $\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x}$ can be written as

$$\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} = \boldsymbol{x}^\top(\boldsymbol{A}\boldsymbol{x}) = \sum_{i=1}^n x_i(\boldsymbol{A}\boldsymbol{x})_i = \sum_{i=1}^m \sum_{j=1}^n a_{ij}x_ix_j,$$

so that the first derivative is

$$\frac{\partial f}{\partial x_i} = \frac{1}{2}\sum_{\substack{j=1\\ j\neq i}}^n (a_{ij} + a_{ji})x_j + a_{ii}x_i + b_i = \sum_{j=1}^n a_{ij}x_j + b_i,$$

where we used the symmetry of $\boldsymbol{A}$ (i.e., $a_{ij} = a_{ji}$). The gradient and Hessian are therefore just given by

$$\nabla f(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}, \quad \nabla^2 f(\boldsymbol{x}) = \boldsymbol{A}.$$

(b) An interesting special case is when the quadratic function (2) arises in the form

$$f(\boldsymbol{x}) = \frac{1}{2}\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|_2^2. \tag{3}$$

The quadratic system then has the form

$$\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|_2^2 = (\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b})^\top(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}) = \boldsymbol{x}^\top \boldsymbol{A}^\top \boldsymbol{A}\boldsymbol{x} - 2\boldsymbol{b}^\top \boldsymbol{A}\boldsymbol{x} + \|\boldsymbol{b}\|_2^2. \tag{4}$$

The matrix $\boldsymbol{A}^\top \boldsymbol{A}$ is always symmetric and positive semidefinite:

$$(\boldsymbol{A}^\top \boldsymbol{A})^\top = \boldsymbol{A}^\top(\boldsymbol{A}^\top)^\top = \boldsymbol{A}^\top \boldsymbol{A}, \quad \text{and} \quad \boldsymbol{x}^\top \boldsymbol{A}^\top \boldsymbol{A}\boldsymbol{x} = \|\boldsymbol{A}\boldsymbol{x}\|_2^2 \geq 0,$$

so that the function (3) is convex. From (4) we also see that the derivative of (3) is

$$\boldsymbol{A}^\top \boldsymbol{A}\boldsymbol{x} - \boldsymbol{A}^\top \boldsymbol{b} = \boldsymbol{A}^\top(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}).$$

Note that the matrix $\boldsymbol{A}$ in (2) is not the same as the matrix $\boldsymbol{A}$ in (3)!

If $m \geq n$ and $\boldsymbol{A}$ has rank n, then $\boldsymbol{x}^\top \boldsymbol{A}^\top \boldsymbol{A}\boldsymbol{x} = \|\boldsymbol{A}\boldsymbol{x}\|_2^2 > 0$ unless $\boldsymbol{x} = \boldsymbol{0}$ (otherwise $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$ would imply that the columns of $\boldsymbol{A}$ are linearly dependent), so the function is strictly convex in this case. The rank of $\boldsymbol{A}^\top \boldsymbol{A} \in \mathbb{R}^{n\times n}$ is n, so this matrix is invertible. Setting the derivative to zero gives

$$\boldsymbol{A}^\top \boldsymbol{A}\boldsymbol{x} - \boldsymbol{A}^\top \boldsymbol{b} = \boldsymbol{0} \Rightarrow \boldsymbol{x} = (\boldsymbol{A}^\top \boldsymbol{A})^{-1}\boldsymbol{A}^\top \boldsymbol{b}.$$

The matrix $\boldsymbol{A}^\dagger := (\boldsymbol{A}^\top \boldsymbol{A})^{-1}\boldsymbol{A}^\top$ is called a *pseudo-inverse* of $\boldsymbol{A}$. It is not technically the inverse, since $\boldsymbol{A}$ may not be quadratic, by it still satisfies $\boldsymbol{A}^\dagger \boldsymbol{A} = \boldsymbol{1}$.

Solution (1.3) The general procedure is as follows: we first make an educated guess as to whether the function could be convex or not. If we think it is not convex, then it is enough to find a *counterexample*: find points in S for which the line segment joining them is not completely contained in S. If we think it is convex, then we can show that for any two points the line segment joining them is in S.

(a) This set is not convex: take $\boldsymbol{x} = (1,0,0)^\top$ and $\boldsymbol{y} = (-1,0,0)^\top$, then $\frac{1}{2}\boldsymbol{x} + \frac{1}{2}\boldsymbol{y} = \boldsymbol{0} \notin S$.

(b) This set is convex: if $\boldsymbol{x}, \boldsymbol{y} \in S$, then $1 \leq x_1 - x_2 < 2$ and $1 \leq y_1 - y_2 < 2$, and

$$\lambda x_1 + (1-\lambda)y_1 - \lambda x_2 - (1-\lambda)y_2 = \lambda(x_1 - x_2) + (1-\lambda)(y_1 - y_2) < \lambda 2 + (1-\lambda)2 = 2,$$

with the same argument giving the lower bound.

(c) This set is convex. In fact, S is the unit ball of the 1-norm

$$\|\boldsymbol{x}\|_1 = \sum_{i=1}^n |x_i|.$$

Given $\boldsymbol{x}, \boldsymbol{y} \in S$,

$$\|\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}\|_1 \leq \lambda\|\boldsymbol{x}\|_1 + (1-\lambda)\|\boldsymbol{y}\|_1 \leq \lambda \cdot 1 + (1-\lambda)\cdot 1 = 1.$$

(d) This set is convex. Here, one needs to show that convex combinations preserve symmetry and positive definiteness of a matrix. The symmetry is clear. As for the positive definiteness, let $\boldsymbol{x} \neq \boldsymbol{0}$ be given. Then

$$\boldsymbol{x}^\top(\lambda \boldsymbol{A} + (1-\lambda)\boldsymbol{B})\boldsymbol{x} = \lambda \boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} + (1-\lambda)\boldsymbol{x}^\top \boldsymbol{B}\boldsymbol{x} \geq 0,$$

which shows that positive definiteness is also preserved.

Solution (1.4)

(a) This function is not convex. There are various ways of deriving this. For example, one can verify that the Hessian, or second derivative, is $-1/x^2$, which is not positive semidefinite.

Alternatively, one can also prove the statement using a pedestrian approach. We have to show that there are points $y \neq x$ and $\lambda \in [0,1]$ such that

$$\log(\lambda x + (1-\lambda)y) > \lambda \log(x) + (1-\lambda)\log(y).$$

Let's choose $y = 1$. Then what needs to be shown is that for the points $\boldsymbol{p}_1 = (1,0)$ and $\boldsymbol{p}_2 = (x, \log(x))$, the line joining $\boldsymbol{p}_1$ and $\boldsymbol{p}_2$ lies *below* the curve $(t, \log(t))$ between 1 and x. The line is given by the equation

$$\ell(t) = \frac{\log(x)}{x-1}(t-1).$$

Evaluating this, for example, at $x = 2$ and $t = 1.5$, one sees that $\ell(t) > \log(t)$, which is enough evidence that $\log(t)$ is not convex. With a little more effort one can deduce that the function is actually concave.

(b) The function $f(x) = x^4$ is convex, as we will verify using Theorem 2.10. First, note that the derivative $4x^3$ is an increasing function with x. Given two points (x, x^4) and (y, y^4) with $y > x$, the line connecting them has slope $(y^4 - x^4)/(y - x)$. By the mean value theorem, there exists a $z \in (x, y)$ such that

$$\frac{y^4 - x^4}{y - x} = f'(z) = 4z^3 \geq 4x^3.$$

Rearranging this inequality, we get

$$f(y) - f(x) = y^4 - x^4 \geq 4x^3(y-x) = f'(x)(y-x),$$

which is precisely the criterium for convexity in Theorem 2.10(1).

(c) Using Theorem 2.10(2), we compute the Hessian as

$$\nabla^2 f(\boldsymbol{x}) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

This matrix is positive semidefinite on $\mathbb{R}^2_{++}$, since for all $\boldsymbol{x} \in \mathbb{R}^2_{++}$ we have

$$\boldsymbol{x}^\top \nabla^2 f(\boldsymbol{x}) \boldsymbol{x} = 2x_1 x_2 > 0.$$

It follows that the function $f(\boldsymbol{x}) = x_1 x_2$ is convex.

(d) The Hessian matrix of $f(\boldsymbol{x}) = x_1/x_2$ is

$$\nabla^2 f(\boldsymbol{x}) = \begin{pmatrix} 0 & -\frac{1}{x_2^2} \\ -\frac{1}{x_2^2} & 2\frac{x_1}{x_2^3} \end{pmatrix}.$$

This matrix is not positive semidefinite for all valid values of $\boldsymbol{x}$ (take for example $\boldsymbol{x} = (1,1)^\top$, which leads to a negative eigenvalue).

(e) The function $e^x - 1$ is convex, as is easily seen using Theorem 2.10(2) by computing the second derivative.

(f) The function $f(\boldsymbol{x}) = \max_i x_i$ is convex. Here, we can't use the criteria from Theorem 2.10 since the function is not differentiable, so we have to verify convexity directly:

$$\max_i \lambda x_i + (1-\lambda) y_i \leq \lambda \max_i x_i + (1-\lambda) \max_i y_i.$$