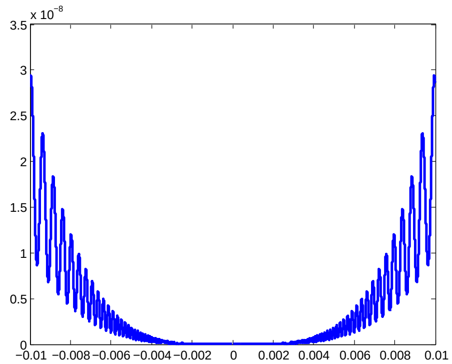
Solutions to Part A of Problem Sheet 1

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 neighourhood 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 $x_k = 1/(\pi k)$ for $k > 0$, 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, for m sufficiently large 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) 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 function $\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x}$ can be written as

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

so that the first derivative is

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

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

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

An interesting special case is when the (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:

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

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

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

Solution (1.3)

(a) The unit circle with respect to the ∞-norm is the square with corners $(\pm 1, \pm 1)^\top$.

(b) The trick in transforming the unconstrained problem

$$\text{minimize} \quad \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|_\infty \tag{5}$$

into a constrained linear programming problem is to characterise the ∞-norm as the solution of a minimization problem. In fact, for any set of numbers $x_1, \dots, x_n$,

$$\max_{1 \le i \le n} |x_i| = \min_{\forall i \colon |x_i| \le t} t.$$

Put simply, the *maximum* of a set of non-negative numbers is the *smallest* upper bound on these numbers. We can further replace the condition $|x_i| \le t$ by $-t \le x_i \le t$, so that the problem (5) becomes

$$\begin{aligned} \underset{(\boldsymbol{x},t)}{\text{minimize}} \quad & t \\ \text{subject to} \quad & -t \le \boldsymbol{a}_1^\top \boldsymbol{x} \le t \\ & \dots \\ & -t \le \boldsymbol{a}_m^\top \boldsymbol{x} \le t, \end{aligned} \tag{6}$$

where $\boldsymbol{a}_i^\top$ are the rows of the matrix $\boldsymbol{A}$. This problem can be brought into *standard form* by replacing each condition with the pair of conditions

$$\begin{aligned} \boldsymbol{a}_i^\top \boldsymbol{x} - t &\le 0 \\ -\boldsymbol{a}_i^\top \boldsymbol{x} - t &\le 0. \end{aligned}$$

The solution $\boldsymbol{x}$ of Problem (5) can be read off the solution $(\boldsymbol{x}, t)$ of Problem (6).