Continuous Maths HT 2019: Problem Sheet 1

Derivatives, Taylor's theorem, 1-dimensional optimization

1.1 Using the quotient and chain rules, prove equation (1.2) in the lecture notes: if f and g are differentiable on $\mathbb{R}$ and $k \in \mathbb{N}$ then

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{f}{g^k}\right) = \frac{g\frac{\mathrm{d}f}{\mathrm{d}x} - kf\frac{\mathrm{d}g}{\mathrm{d}x}}{g^{k+1}}.$$

Is the result still true for $k = 0$? What about integers $k < 0$, or non-integer k?

Now consider differentiable multivariate functions $f, g : \mathbb{R}^n \to \mathbb{R}$. What is $\frac{\mathrm{d}}{\mathrm{d}\boldsymbol{x}}\left(\frac{f}{g^k}\right)$? How does your proof differ from the univariate case $n = 1$?

1.2 Let A be a symmetric n-by-n matrix and define, for $\boldsymbol{x} \neq \mathbf{0}$,

$$f(\boldsymbol{x}) = \frac{\boldsymbol{x}^T\mathbf{A}\boldsymbol{x}}{\boldsymbol{x}^T\boldsymbol{x}}.$$

Compute $\frac{\mathrm{d}f}{\mathrm{d}\boldsymbol{x}}$, and show that this is zero if and only if x is an eigenvector of A.

1.3 In machine learning we sometimes wish to use the function $f(\boldsymbol{x}) = \max x_i$ (where the components of $\boldsymbol{x}$ are $x_1, \ldots, x_n$), but this function is not differentiable, making numerical optimization algorithms more difficult. Instead, f is sometimes replaced by the function l, where

$$l(\boldsymbol{x}) = \ln\Big(\sum_{i=1}^{n} \exp(x_i)\Big).$$

l is sometimes called the log-sum-exp function.

(a) Show that $\max x_i \leq l(\boldsymbol{x}) \leq \max x_i + \ln n$. (This means that $l(\boldsymbol{x})$ is a reasonable approximation for $\max x_i$ as long as the x_i are not small in relation to $\ln n$).

(b) Show that l is differentiable, then find $\frac{\partial l}{\partial x_j}$, $\frac{\partial^2 l}{\partial x_j^2}$, and $\frac{\partial^2 l}{\partial x_k x_j}$.

1.4 Let $H : [0, 1] \to \mathbb{R}$ be the *binary entropy function*,

$$H(x) = -x \log_2 x - (1 - x)\log_2(1 - x), \quad H(0) = H(1) = 0.$$

(a) Compute the first four derivatives of H.

(b) Write down the third-order Taylor expansion of H, with fourth-order (Lagrange) error term, about $x_0 = \frac{1}{2}$.

(c) Why is there no Taylor expansion for $H(x)$ at $x = 0$?

1.5 Let $f : \mathbb{R} \to \mathbb{R}$ be $f(x) = e^x + e^{-x}$.

(a) Compute $\frac{\mathrm{d}^k f}{\mathrm{d}x^k}$ for each integer k.

(b) Write down the order-k Taylor polynomial for $f(x)$, around the point $x_0 = 1$, and the (Lagrange) remainder term $e_{k+1}(x)$.

(c) If $x \in (0, 1)$, show that

$$0 < e_{k+1}(x) \leq \frac{(e + \frac{1}{e})}{(k+1)!}$$

for k odd, and find equivalent bounds for the case when k is even.

(d) Write a short program to find the smallest k for which the k-order polynomial is an accurate approximation to $f(x)$ with error no more than 10^{-15}. What will the sign of the error be?

1.6 Let $\mathbf{A}$ be a fixed symmetric matrix, $\boldsymbol{b}$ a fixed vector, and c a constant. Define

$$f(\boldsymbol{x}) = \sin(\boldsymbol{x}^T\mathbf{A}\boldsymbol{x} + \boldsymbol{x}^T\boldsymbol{b} + c).$$

(a) Compute $\frac{\mathrm{d}f}{\mathrm{d}\boldsymbol{x}}$ and $\mathbf{H}(f)$.

(b) Show that the second-order Taylor approximation to f, about $\boldsymbol{x_0} = \mathbf{0}$, is

$$f_2(\boldsymbol{x}) = \sin c + (\cos c)\boldsymbol{x}^T\boldsymbol{b} + \boldsymbol{x}^T\big((\cos c)\mathbf{A} - \tfrac{\sin c}{2}\boldsymbol{b}\boldsymbol{b}^T\big)\boldsymbol{x} + e_3,$$

where e_3 is the remainder term.

1.7 Let $a, b \in \mathbb{R}$, and $f : \mathbb{R} \to \mathbb{R}$ be

$$f(x) = 3x^4 - 4(a+b)x^3 + 6abx^2.$$

Verify that f has a stationary point at $x = 0$, and classify it as a local maximum, local minimum, or stationary point of inflection, depending on the values of a and b.

If $x = 0$ is a local maximum, for what values of a and b is it the global maximum? If $x = 0$ is a local minimum, for what values of a and b is it the global minimum?