Friederike Horn & Bileam Scheuvens

Justify all your claims.

Exercise 1 (Extremal points)

Consider the function $f : \mathbb{R}^2 \to \mathbb{R}; (x, y) \mapsto x^3 + 1/3y^3 - 12x - y$

a) (x, y) is an extremal point if $\frac{\partial f}{\partial x} = 0$ and $\frac{\partial f}{\partial y} = 0$.

$$\frac{\partial f}{\partial x} = 3x^2 - 12$$
$$\frac{\partial f}{\partial y} = y^2 - 1$$

From this we can get four different extremal points:

$$(x_1, y_1) = (2, 1)$$
$$(x_2, y_2) = (-2, 1)$$
$$(x_3, y_3) = (2, -1)$$
$$(x_4, y_4) = (-2, -1)$$

In order to classify them we need to compute the Hessian:

$$H = \begin{pmatrix} 6x & 0 \\ 0 & 2y \end{pmatrix}.$$

We directly see that this is diagonal and thus the eigenvalues correspond to the diagonal entriess
We see that for (x_1, y_1) we have positive eigenvalues and therefore a strict local minimum. Likewise, for (x_4, y_4) we have negative eigenvalues and therefore a strict local maximum. For the other two points the Hessian matrix has a negative and a positive eigenvalue and thus it is indefinite and we have two saddle points.

b) (x, y) is a global maximum iff there exists no other (x', y') such that $f(x', y') > f(x, y))$ and likewise for global minimum.
In this case the function has no global maximum or minimum. E.g. $f(x_1, y_1) = 8+1/3-24-1 = -17.3$, but we find that $f(-3, 0) = -27 < -17.3$.
Similarly, $f(x_4, y_4) = -8 - 1/3 + 24 + 1 = 16.67$, but $f(3, 0) = 27 > 16.67 = f(x_4, y_4)$.

c) Consider $g : \mathbb{R}^3 \to \mathbb{R}, (x, y, z) \mapsto \alpha x^2 e^y + y^2 e^z + z^2 e^x$, with $\alpha \in \mathbb{R}$. The point $(0, 0, 0)$ is an extremal point (local, minimum, maximum or saddle point) iff the first derivative is zero i.e. $\nabla f = \mathbf{0}$.

$$\nabla f = \begin{pmatrix} \alpha 2xe^y + z^2e^x \\ \alpha x^2e^y + 2ye^z \\ y^2e^z + 2ze^x \end{pmatrix}_{(0,0)}$$
$$= \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

To differentiate between the different points we compute the Hessian:

$$H = \begin{pmatrix} \alpha 2e^y + z^2e^x & \alpha 2xe^y & 2ze^x \\ \alpha 2xe^y & x^2e^y + 2e^z & 2ye^z \\ 2ze^x & 2ye^z & y^2e^z + 2e^x \end{pmatrix}_{(0,0)}$$

$$\begin{pmatrix} \alpha 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}.$$

As this is again a diagonal matrix we can directly read out the eigenvalues from the diagonal entries and as two af them are greater than zero we can only have a local minimum or saddle point. If $\alpha < 0$ the matrix is indefinite and we therefore have a saddle point. If $a \geq 0$ we have a local minimum.

Exercise 2 (Derivatives)

a) The directional derivatives exist in point $(0,0,0)$ iff the $\lim_{x\to 0+} \frac{\partial f}{\partial x}_{y=0} = \lim_{x\to 0-} \frac{\partial f}{\partial x}_{y=0}$ and likewise for y. We therefore bult the two derivatives:

$$\frac{\partial f_+}{\partial x} =$$

Exercise 3 (Taylor Series)

Given the function $f : \mathbb{R}^2 \to \mathbb{R}; (x_1, x_2) \mapsto \frac{1}{1-x_1-x_2}$ find the Taylor series around point $(0,0)$ and the set over which it converges.

Solution

We first note for the derivatives:

$$\frac{\partial f^n}{\partial x_1{}^{\alpha} \partial x_2{}^{n-\alpha}} = n! \frac{1}{(1-x_1-x_2)^n}.$$

This is because the partial derivative in x_1 direction is equal to the derivative in x_2 direction and each derivative simply multiplies by the power of the denominator and then increases the power of the denominator by one.
For this reason the multivariate Taylor series (around the point $(0,0)$) can immediately be written out as:

$$T_f = \sum_{n=0}^{\infty} \sum_{\alpha=0}^{n} \frac{n!}{\alpha!(n-\alpha)!} x_1^{\alpha} x_2^{n-\alpha}.$$

This can be rewritten as:

$$T_f = \sum_{n=0}^{\infty} \sum_{\alpha=0}^{n} \binom{n}{\alpha} x_1^{\alpha} x_2^{n-\alpha} = \sum_{n=0}^{\infty} (x_1 + x_2)^n.$$

This series is the geometric series and converges iff $|x_1 + x_2| < 1$.

Exercise 4 (Matrix Cookbook)

Solution

a)

$$\frac{\partial a^t x}{\partial x} = \begin{pmatrix} \frac{\partial a^t x}{\partial x_1} \\ \frac{\partial a^t x}{\partial x_2} \\ \frac{\partial a^t x}{\partial x_n} \end{pmatrix}$$

With $a^t x = a_1 x_1 + a_2 x_2 + ... + a_n x_n$ we conclude:

$$\frac{\partial a^t x}{\partial x} = \begin{pmatrix} a_1 \\ a_2 \\ a_n \end{pmatrix} = a.$$

Secondly, $x^t A x = \sum_i x_i \sum_j A_i j x_j$ and thus the gradient is:

$$\frac{\partial x^t A x}{\partial x} = \begin{pmatrix} \frac{\partial x^t A x}{\partial x_1} \\ \frac{\partial x^t A x}{\partial x_2} \\ ... \\ \frac{\partial x^t A x}{\partial x_n} \end{pmatrix}$$

$$= \begin{pmatrix} \sum_j A_{1j} x_j + \sum_i A_{i1} x_i \\ \sum_j A_{2j} x_j + \sum_i A_{i2} x_i \\ ... \\ \sum_j A_{nj} x_j + \sum_i A_{in} x_i \end{pmatrix}$$

$$= \begin{pmatrix} \sum_j A_{1j} x_j + \sum_j A^T_{1j} x_j \\ \sum_j A_{2j} x_j + \sum_j A^T_{2j} x_j \\ ... \\ \sum_j A_{nj} x_j + \sum_j A^T_{nj} x_j \end{pmatrix}$$

where we renamed the $i \to j$ and used that $A_{ij} =$A$^T_{ji}$. But this simply corresponds to $(A + A^t)x$.

b) To minimize with respect to w we first write out the norm squared:

$$\frac{1}{n}||Xw - Y||^2 + \lambda||w||^2 = \frac{1}{n}(Xw - Y)^T(Xw - Y) + \lambda w^T w$$

$$= \frac{1}{n}(w^T X^T X w - w^T X^T Y - Y^T X w + Y^T Y) + \lambda w^T w = f.$$

Taking the derivative with respect to w we obtain with the results from (a):

$$\frac{\partial f}{\partial w} = \frac{1}{n}((X^T X + (X^T X)^T)w - 2Y^T X) + \lambda w.$$

Here we also used that we are in $\mathbb{R}$ and thus $w^T X^T y = y^T X w$.