Exercise sheet 1

1. Let $f(x) = \max_{i=1,\dots,m}(a_i^T x + b_i)$ where $a_1, \dots, a_m \in \mathbb{R}^n$ and $b_1, \dots, b_m \in \mathbb{R}$. Explain why $f(x)$ is convex. Given $x \in \mathbb{R}^n$ let $I(x) = \left\{ i \in \{1, \dots, m\} : a_i^T x + b_i = f(x) \right\}$. Show that the subdifferential of f at x is given by
$$\partial f(x) = \operatorname{conv}\{a_i : i \in I(x)\}$$
where $\operatorname{conv}(X)$ denotes the *convex hull* of a set X.

2. Prove that the following functions are convex on their domain:
   (a) $f(x) = \|Ax - b\|_2^2$ where $x \in \mathbb{R}^n$
   (b) $f(x) = \log(\sum_{i=1}^n e^{x_i})$ where $x \in \mathbb{R}^n$
   (c) $f(x) =$ sum of k largest components of x, where $x \in \mathbb{R}^n$ and $k \in \{1, \dots, n\}$. (for example, $f(x) = \max_{i=1,\dots,n} x_i$ when $k = 1$, and $f(x) = x_1 + \cdots + x_n$ when $k = n$.)
   (d) $f(X) =$ largest eigenvalue of X (X real symmetric $n \times n$ matrix)
   (e) $f(X) = -\log\det X$ where X is a symmetric positive definite matrix
   (f) $f(x, y) = \sum_{i=1}^n x_i \log(x_i/y_i)$ where $x, y \in \mathbb{R}_+^n$

   Also specify which functions are smooth, in which case provide an expression for the gradient. For nonsmooth functions provide an expression for a subgradient (if possible, compute the full subdifferential $\partial f(x)$)

3. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function. Show that f is m-strongly convex iff $f - (m/2)\|x\|_2^2$ is convex. Show that ∇f is L-Lipschitz iff $(L/2)\|x\|_2^2 - f$ is convex.

4. Prove that the gradient method, with the following backtracking line search, converges at the rate $O(1/k)$: at each iteration k, initialize t_k to 1 and keep updating $t_k \leftarrow \beta t_k$ (where $\beta \in (0, 1)$) until $f(x_k - t_k \nabla f(x_k)) \le f(x_k) - (1/2)t_k \|\nabla f(x_k)\|_2^2$.

5. Consider the problem of minimizing a convex function $f(x)$ on a closed convex set C, i.e., we want to compute $\min_{x \in C} f(x)$. The *projected gradient method* works as follows: starting from $x_0 \in C$, let $x_{k+1} = P_C(x_k - t_k \nabla f(x_k))$ where P_C is the Euclidean projection on C defined by
$$P_C(x) = \operatorname*{argmin}_{y \in C} \|y - x\|_2^2.$$
By adapting the convergence proof of the gradient method seen in lecture, show that the projected gradient method converges with a rate $O(1/k)$ when ∇f is assumed L-Lipschitz, and the step size t_k is fixed $t_k = t \in (0, 1/L]$.

6. Implement the gradient method and fast gradient method to minimize the following convex function (logistic regression loss)
$$f(x) = \sum_{i=1}^N \log\left[1 + \exp(y_i a_i^T x)\right]$$
where $a_1, \dots, a_N \in \mathbb{R}^n$ and $y_1, \dots, y_N \in \{-1, +1\}$ are randomly generated. Take $N = 50$ and $n = 30$. Plot $f(x_k) - f^*$ as a function of k. Comment.