Exercise sheet 2

1. Let $f : \mathbb{R}^n \to \mathbb{R}$ convex. Show that f is G-Lipschitz (with respect to the ℓ_2 norm) iff $\|g\|_2 \leq G$ for all $g \in \partial f(x)$ for all $x \in \mathbb{R}^n$.

2. Polyak step size for subgradient method: show that the subgradient method with step size $t_i = (f(x_i) - f^*)/\|g_i\|_2^2$ gives iterates $f_{\text{best},k}$ that converge to f^* at the rate $1/\sqrt{k}$ (hint: start from the inequalities relating $\|x_{k+1} - x^*\|_2^2$ to $\|x_k - x^*\|_2^2$).

3. To minimize a nonsmooth function f on a convex set C, the projected subgradient method proceeds as follows: $x_{k+1} = P_C(x_k - t_k g_k)$ where P_C is the Euclidean projection on C, and $g_k \in \partial f(x_k)$. Analyze the convergence of the projected subgradient descent.

4. Compute conjugates of following functions
   (a) $f(x) = \sum_{i=1}^n x_i \log x_i$
   (b) $f(x) = -\sum_{i=1}^n \log x_i$
   (c) $f(X) = -\log\det X$ where X is an $n \times n$ real symmetric positive definite matrix

5. Show that for any closed convex function f we have $\text{prox}_f(x) + \text{prox}_{f^*}(x) = x$ for all x (Moreau's identity).

6. Let $f(x) = \max_{i=1}^m a_i^T x + b_i$. Show that $f(x) = h^*(Ax + b)$ where $h(y)$ is the indicator function of the simplex $\{y \in \mathbb{R}^m : y_i \geq 0 \ \forall i = 1, \ldots, m \text{ and } \sum y_i = 1\}$. Find an expression for $f_\mu(x) = (h + \mu d)^*(Ax + b)$ where $d(y) = \sum_{i=1}^m y_i \log y_i - \log m$.

7. Implement the subgradient method to minimize $\|Ax - b\|_1$ where A and b are generated at random. Experiment with different choices of step size. Compare with the smoothing method of Nesterov.

8. In this exercise we prove a lower complexity bound for nonsmooth convex optimization. Consider an algorithm that starts at $x_0 = 0$ and such that when applied to a function f, the $(i+1)$'th iterate satisfies
$$x_i \in \text{span}\{g_0, \ldots, g_i\} \tag{1}$$
where $g_0 \in \partial f(x_0) = \partial f(0), \ldots, g_i \in \partial f(x_i)$.
   (a) Consider the function
   $$f(x) = \max_{i=1,\ldots,n} x_i + \frac{1}{2}\|x\|_2^2$$
   with $x \in \mathbb{R}^n$. Compute $\partial f(x)$ for any x.
   (b) Compute $f^* = \min_{x \in \mathbb{R}^n} f(x)$ and find a minimizer x^*.
   (c) Show that f is $(1 + R)$-Lipschitz on the Euclidean ball $\{x \in \mathbb{R}^n : \|x\|_2 \leq R\}$ [Hint: consider $\|g\|_2$ for $g \in \partial f(x)$.]
   (d) A first-order oracle for f gives, for any $x \in \mathbb{R}^n$, an element $g \in \partial f(x)$. Show that one can design a specific first-order oracle for f ensuring that x_i satisfying (1) is always supported on the first i components only (i.e., the components $i + 1, \ldots, n$ are zero).

(e) Set $n = k + 1$. Show that for any algorithm satisfying (1), the following holds:

$$\frac{f_{\text{best},k} - f^*}{G\|x_0 - x^*\|_2} \geq \frac{c}{\sqrt{k+1}}$$

for a constant $c > 0$, where $f_{\text{best},k} = \min\{f(x_0), \ldots, f(x_k)\}$ and G is the Lipschitz constant of f on the Euclidean ball of radius $\|x_0 - x^*\|_2$.