Exercise sheet 2

You can return your solutions to questions 7 and 8 to get them marked. If so, please upload them on Moodle before Tuesday 21/02 at 5pm.

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

2. (Directional derivatives) Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be convex. Let $x \in \mathbf{int\,dom}(f)$.
   (i) Show that the directional derivative of f
   $$f'(x;h) := \lim_{t\to 0^+} \frac{f(x+th) - f(x)}{t}$$
   is well-defined and finite for any h, even if f is not differentiable at x. [*Hint: show that the* $\lim_{t\to 0^+}$ *can be replaced by* $\inf_{t\to 0^+}$.]
   (ii) Show that $f'(x;h)$ is homogeneous in h, i.e., $f'(x;\lambda h) = \lambda f'(x;h)$ for all $\lambda > 0$. Show that $f'(x;h)$ is convex in h.
   (iii) Let g be a subgradient for $v \mapsto f'(x;v)$ at $v = h$. Show that $f'(x;h) = \langle g, h\rangle$ and that $f'(x;v) \ge \langle g, v\rangle$ for all v. Deduce from the latter that $g \in \partial f(x)$.
   (iv) Deduce from the above that $f'(x;h) = \max_{g\in\partial f(x)} \langle g, h\rangle$.

3. (Subgradient calculus) Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function defined on the whole of $\mathbb{R}^n$, and let $A : \mathbb{R}^m \to \mathbb{R}^n$ be a linear map. Let $h(x) = f(Ax)$. The goal of this exercise is to show that $\partial h(x) = A^* \partial f(Ax)$.
   (i) Let $x \in \mathbb{R}^n$. Show that $h'(x;v) = f'(Ax;Av)$, where $h'(x;v)$ is the directional derivative defined in the previous exercise.
   (ii) Deduce that for any v, $\max_{g\in\partial h(x)} \langle g, v\rangle = \max_{g\in A^*\partial f(Ax)} \langle g, v\rangle$.
   (iv) Conclude.

4. Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a convex function and $x \in \mathbf{int\,dom}\, f$ such that $\partial f(x)$ is a singleton, namely $\partial f(x) = \{g\}$. Using Question 2(iv), show that f is differentiable at x, i.e.,
   $$\frac{f(x+h) - f(x) - \langle g, h\rangle}{\|h\|} \to 0 \quad \text{as} \quad h \to 0.$$

5. 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 $b_1, \dots, b_m \in \mathbb{R}$. Given $x \in \mathbb{R}^n$ let $I(x) = \big\{i \in \{1,\dots,m\} : a_i^T x + b_i = f(x)\big\}$. Show that the subdifferential of f at x is given by
   $$\partial f(x) = \mathbf{conv}\,\{a_i : i \in I(x)\} \tag{1}$$
   where $\mathbf{conv}(X)$ denotes the *convex hull* of X.

6. Compute the normal cones $N_C(x)$ for $x \in C$ for the following convex sets
   (a) $C = L = \{x \in \mathbb{R}^n : Ax = b\}$, where $A \in \mathbb{R}^{m\times n}$ with $m < n$ (subspace)
   (b) $C = \mathbb{R}^n_+ = \{x \in \mathbb{R}^n : x_i \ge 0\ \forall i = 1,\dots,n\}$
   (c) $C = \mathbf{S}^n_+ = \{X \in \mathbf{S}^n : X \succeq 0\}$ ($n \times n$ real symmetric positive semidefinite matrices)

7. (*) (a) Let $f : \mathbb{R}^n \to \mathbb{R}$ differentiable convex function and C a closed convex set. Show that x^* is a minimizer of $f(x)$ over $x \in C$ if, and only if, $-\nabla f(x^*) \in N_C(x^*)$
(b) Use the subgradient calculus rules to show that x^* is a solution of the following linear program

$$\min \ \langle c, x \rangle \ \text{ s.t. } \ x \geq 0, \ Ax = b$$

(where $A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m, c \in \mathbb{R}^n$) if, and only if, there exist $s^* \geq 0$ and $z^* \in \mathbb{R}^m$ such that

$$\begin{cases} x^* \geq 0, Ax^* = b \\ s^* \geq 0, A^T z^* + s^* = c \\ x_i^* s_i^* = 0 \ \ \forall i = 1, \ldots, n. \end{cases}$$

8. (*) Show that the subgradient method with step size $t_i = (f(x_i) - f^*)/\|g_i\|_2^2$ (known as *Polyak step size*) gives iterates $f_{\text{best},k}$ that converge to f^* at the rate $1/\sqrt{k}$

9. (Lower complexity bound for the subgradient method) 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

$$x_{i+1} \in \text{span}\{g_0, \ldots, g_i\} \tag{2}$$

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 (2) 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 (2), 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 $\|x_0 - x^*\|_2$ centered at x_0.

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