Exercise sheet 2

You can return your solutions to questions 6 and 7 to get them marked. If so, please upload them on Moodle before Monday 28/02 at 12noon.

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. (Directional derivatives) Let $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ be convex. Let $x \in \textbf{int dom}(f)$.
   (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 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 $f'(x; v) \geq \langle g, v\rangle$ 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 with $\textbf{dom}(f) = \mathbb{R}^n$ 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) 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, $\max_{g\in\partial h(x)} \langle g, v\rangle = \max_{g\in A^*\partial f(Ax)} \langle g, v\rangle$.
   (iii) Conclude (hint: use the strict separating hyperplane theorem).

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

6. (*) 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 $1/\sqrt{k}$ (hint: start from the inequalities relating $\|x_{k+1} - x^*\|_2^2$ to $\|x_k - x^*\|_2^2$).

7. (*) Consider the following optimization problem for denoising a one-dimensional signal $b \in \mathbb{R}^n$:

$$\min_{x\in\mathbb{R}^n} \|x-b\|_2^2 + \gamma\|Dx\|_1$$

where $\gamma > 0$, and $D \in \mathbb{R}^{(n-1)\times n}$ is the finite-difference operator $Dx = [x_{i+1} - x_i]_{1\le i\le n-1}$.

After introducing the new variable $y = Dx$, compute the Lagrangian and the dual problem, and discuss algorithms to solve the dual problem as well as their convergence properties. Compare with the subgradient method applied to the original problem. Extra: implement the algorithms with b a piecewise constant signal corrupted by some Gaussian noise.

8. (a) Let $\mathbb{R}^n_+ = \{x \in \mathbb{R}^n : x_i \ge 0\ \forall i\}$. Compute the normal cone of $\mathbb{R}^n_+$ at any $x \in \mathbb{R}^n_+$.

(b) Consider the following linear program:

$$\min_{x\in\mathbb{R}^n} c^T x \quad \text{s.t.} \quad Ax = b, x \ge 0$$

where $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$ and $c \in \mathbb{R}^n$. Derive the KKT conditions for optimality. What is the dual optimization problem?

9. (a) Let $\mathbf{S}^n_+$ be the convex cone of real symmetric $n\times n$ matrices that are positive semidefinite. Compute the normal cone of $\mathbf{S}^n_+$ at any $X \in \mathbf{S}^n_+$. [The inner product for two real symmetric matrices $A, B \in \mathbf{S}^n$ is $\langle A, B\rangle = \operatorname{tr}(AB) = \sum_{ij} A_{ij}B_{ij}$.]

(b) Consider the following convex optimization problem, known as a *semidefinite program*:

$$\min_{X\in\mathbf{S}^n} \langle C, X\rangle \quad \text{s.t.} \quad A(X) = b, X \in \mathbf{S}^n_+$$

where $A : \mathbf{S}^n \to \mathbb{R}^m$ is a linear map, $b \in \mathbb{R}^m$ and $C \in \mathbf{S}^n$. Assuming that $\mathbf{int}(\mathbf{S}^n_+) \cap \{A(X) = b\}$ is nonempty, derive the KKT conditions for optimality. What is the dual optimization problem?

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.

11. (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

$$x_i \in \operatorname{span}\{g_0, \ldots, g_i\} \tag{2}$$

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 $\{x \in \mathbb{R}^n : \|x\|_2 \le R\}$ [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.