Convex optimization exercises - GD

Subgradients

Gradient Descent can work even if gradients do not always exist, by using so-called *subgradients* instead. For a convex f, g_x is a subgradient of f at x if $f(y) \geq \langle g_x, y - x \rangle$ for all y. We use $\partial f(x)$ to denote the set of all subgradients at x (called *subdifferential*). We showed in the lecture that $\partial f(x) \neq \emptyset$, except possibly when x lies on the boundary of $dom(f)$. The next few exercises explore subgradients in more detail.

Exercise 1. *Show that if f is differentiable at x, then $\partial f(x) = \{\nabla f(x)\}$.*
Note: *The converse is also true for convex f and x in the interior of $dom(f)$, but we will not prove it.*

Exercise 2. *Show that $\partial f(x)$ is always a convex set.*

Exercise 3. *Show that if f is L-Lipschitz, then for any $x \in dom(f)$ and $g_x \in \partial f(x)$, we have $||g_x|| \leq L$.*

It can be shown that for x in the interior of both $dom(f_1)$ and $dom(f_2)$ we have $\partial(f_1 + f_2)(x) = \partial f_1(x) + \partial f_2(x)$. It can also be shown (with analogous assumptions) that for $f(x) = \max(f_1, \ldots, f_k)(x)$ we have $\partial f(x) = \text{conv}(\bigcup_{i: f_i(x) = f(x)} \partial f_i(x))$. Proofs are a bit harder than you might expect.

Exercise 4. *Consider the function $f(x) = \gamma \max x_i + \frac{\alpha}{2}||x||_2$. What is $\partial f(x)$? (this function appears in the lecture a few times)*

Exercise 5. *Show that for a convex f, x^* is a global minimizer of f iff $0 \in \partial f$.*

Exercise 6. *Use the previous exercise to characterize the global optimum points of $f(x) = \frac{1}{2}||Ax - b||_2^2 + \lambda||x||_1$.*