7 Subgradients

Many optimization problems that arise in practice involve nonsmooth functions, such as $\|x\|_1$, $\|x\|_\infty$, or in general $\max\{f_1(x), \ldots, f_m(x)\}$. In this lecture we give a brief overview of the tools from convex analysis needed to study such optimization problems. The main concept we study in this lecture is that of a *subgradient*.

Definition 7.1. Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ and $x \in \mathbf{dom}(f)$. We say that g is a *subgradient* of f at x if for any $y \in \mathbb{R}^n$,

$$f(y) \geq f(x) + \langle g, y - x \rangle .$$

The set of all subgradients of f at x is denoted $\partial f(x)$, and is called the *subdifferential* of f at x.

Remark that x^* is a minimizer of f if, and only if, $0 \in \partial f(x)$.

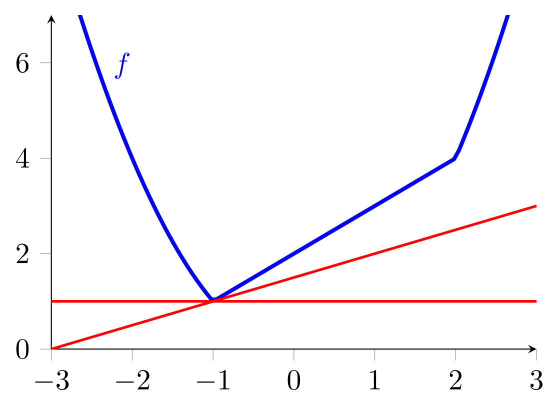

Figure 1: Subgradients of a convex function.

Clearly if f is convex and differentiable at x, then $\nabla f(x)$ is a subgradient of f at x. The theorem below shows that subgradients always exist for convex functions, even if f is not differentiable.

Theorem 7.1. *Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be convex. Then*
(i) $\partial f(x)$ is nonempty for all $x \in \mathbf{int}\,\mathbf{dom}(f)$
(ii) $\partial f(x)$ is closed and convex for all x. For $x \in \mathbf{int}\,\mathbf{dom}(f)$, $\partial f(x)$ is bounded.
(iii) $\partial f(x)$ is a singleton if, and only if, f is differentiable at x.

Proof. (i) We apply the supporting hyperplane theorem to

$$\mathbf{epi}(f) = \{(x,t) \in \mathbb{R}^n \times \mathbb{R} : f(x) \leq t\} \subset \mathbb{R}^{n+1}.$$

Since $(x, f(x)) \in \mathbf{bd}\,\mathbf{epi}(f)$ [here $\mathbf{bd}\,C = \mathbf{cl}\,C \backslash \mathbf{int}\,C$ is the boundary of C] we can find a supporting hyperplane, i.e., a vector $a = (a^1, a^2) \in \mathbb{R}^n \times \mathbb{R}$ and a scalar b such that $\left\langle a^1, x \right\rangle + a^2 f(x) = b$ and $\left\langle a^1, y \right\rangle + a^2 t \geq b$ for all $(y,t) \in \mathbf{epi}(f)$. Since t can be made arbitrarily large, it must be that $a^2 \geq 0$. Since $x \in \mathbf{int}\,\mathbf{dom}(f)$, $a^2 \neq 0$ (if $a^2 = 0$ then we get a supporting hyperplane to $\mathbf{dom}(f)$ at x). Dividing by a^2 we can assume $a^2 = 1$, so that $\left\langle a^1, x \right\rangle + f(x) = b$ and $\left\langle a^1, y \right\rangle + f(y) \geq b$ for all $y \in \mathbf{dom}(f)$, i.e., $f(y) \geq f(x) + \langle g, y - x \rangle$ where $g = -a^1$, i.e., $g \in \partial f(x)$.

(ii) $\partial f(x) = \{g \in \mathbb{R}^n : f(y) \geq f(x) + \langle g, y - x\rangle\}$ is an intersection of closed halfspaces and so is closed and convex. If $x \in \textbf{int}\,\textbf{dom}(f)$, then for some $\epsilon > 0$, $B(x, \epsilon) \subset \textbf{dom}(f)$. If $g \in \partial f(x)$, then by letting $h = \epsilon g/\|g\|_2$ we have $f(x + h) \geq f(x) + \langle g, h\rangle = f(x) + \epsilon\|g\|_2$ which implies that $\|g\|_2 \leq \frac{1}{\epsilon} \max_{y\in B(x,\epsilon)}(f(y) - f(x)) < \infty$.

(iii) If f is differentiable at x, then we know from the results seen in Lecture 2 that $\nabla f(x) \in \partial f(x)$. Also if $g \in \partial f(x)$ then for any direction h we have

$$f(x) + t\,\langle\nabla f(x), h\rangle + o(t) = f(x + th) \geq f(x) + t\,\langle g, h\rangle\,.$$

Simplifying, this yields $\langle\nabla f(x) - g, h\rangle \geq 0$. This has to hold for all h, and so necessarily $g = \nabla f(x)$. We have thus shown that if f is differentiable at x, then $\partial f(x) = \{\nabla f(x)\}$.

We omit the proof of the converse here (see Exercise sheet 2). □

Example: normal cones Let $C \subset \mathbb{R}^n$ be a closed convex set, and let

$$I_C(x) = \begin{cases} 0 & \text{if } x \in C \\ +\infty & \text{else} \end{cases}$$

be its indicator function. What is $\partial I_C(x)$ for $x \in C$? By definition, we have $g \in \partial I_C(x)$ if, and only if, $0 \geq \langle g, y - x\rangle$ for all $y \in C$. If $x \in \textbf{int}\, C$, then clearly this implies that $g = 0$. However, this is not the case if $g \in C \setminus \textbf{int}\, C$. In fact, the set $\partial I_C(x) = \{g : \langle g, x\rangle \geq \langle g, y\rangle\ \forall y \in C\}$ is known as the *normal cone* of C at x, and denoted $N_C(x)$.

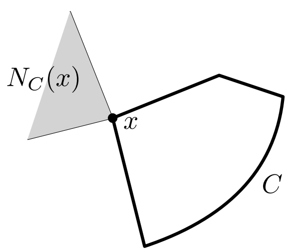

Figure 2: Normal cone

7.1 Subgradient calculus

If $f : \mathbb{R}^n \to \mathbb{R}$ is a differentiable function, and $h(x) = f(Ax)$ where $A \in \mathbb{R}^{n\times m}$, then it is immediate to verify that $\nabla h(x) = A^*\nabla f(Ax)$, where A^* is the adjoint (transpose) of A. Also if f_1, f_2 are two differentiable functions, then $\nabla(f_1 + f_2)(x) = \nabla f_1(x) + \nabla f_2(x)$. These relations also hold in general for the subgradient of convex functions; however the proof is not immediate and relies on duality theory.

Theorem 7.2. *Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a convex function.*

(i) If $h(x) = f(Ax)$, where $A \in \mathbb{R}^{n\times m}$, such that[1] $\textbf{im}(A) \cap \textbf{int}\,\textbf{dom}(f) \neq \emptyset$, *then $\partial h(x) = A^*\partial f(Ax)$ for all x.*

(ii) If f_1, f_2 are two convex functions, such that[2] $\textbf{int}\,\textbf{dom}\, f_1 \cap \textbf{int}\,\textbf{dom}\, f_2 \neq \emptyset$, *then $\partial(f_1 + f_2)(x) = \partial f_1(x) + \partial f_2(x)$ for all x, where the right-hand side is the Minkowski sum of sets $A + B = \{a + b : a \in A, b \in B\}$.*

[1] If f is polyhedral (i.e., $\textbf{epi}(f)$ is a convex set defined using a finite number of linear inequalities), this assumption can be relaxed to $\textbf{im}(A) \cap \textbf{dom}(f) \neq \emptyset$.

[2] If f_1 is polyhedral (i.e., $\textbf{epi}(f_1)$ is a convex set defined using a finite number of linear inequalities), this assumption can be relaxed to $\textbf{dom}\, f_1 \cap \textbf{int}\,\textbf{dom}\, f_2 \neq \emptyset$. If f_2 is also polyhedral, then we just need $\textbf{dom}\, f_1 \cap \textbf{dom}\, f_2 \neq \emptyset$.

(iii) Let $(f_\alpha)_{\alpha\in\mathcal{A}}$ be a finite collection of convex functions, and let $f(x) = \max_{\alpha\in\mathcal{A}} f_\alpha(x)$. Then for any $x \in \textbf{int dom}\, f$,

$$\partial f(x) = \textbf{conv} \cup_{\alpha\in\mathcal{A}(x)} \partial f_\alpha(x). \tag{1}$$

where $\mathcal{A}(x) = \{\alpha \in \mathcal{A} : f_\alpha(x) = f(x)\}$, and where **conv** *denotes the convex hull.*
More generally, (1) holds if $\mathcal{A}$ is a compact set, and $f_\alpha(x)$ depends continuously on α.

Proof. (i) The inclusion $\supset$ is easy to verify: If $g \in \partial f(Ax)$, then for any y we have

$$h(y) = f(Ay) \geq f(Ax) + \langle g, Ay - Ax\rangle = f(Ax) + \langle A^*g, y - x\rangle = h(x) + \langle A^*g, y - x\rangle$$

which shows that $A^*g \in \partial h(x)$. The reverse inclusion $\subseteq$ is omitted here (see Exercise sheet 2 for a special case, and see [Roc15, Theorem 23.9] for the general case).

(ii) Let $F : \mathbb{R}^{2n} \to \overline{\mathbb{R}}$ defined by $F(x_1, x_2) = f_1(x_1) + f_2(x_2)$. It easy to check that $\partial F(x_1, x_2) = \partial f_1(x_1) \times \partial f_2(x_2)$. Let $A : \mathbb{R}^n \to \mathbb{R}^{2n}$ be the linear map $Ax = (x, x)$ whose adjoint is $A^*(x_1, x_2) = x_1 + x_2$. Then $f(x) = F(x, x) = F(Ax)$ and so, by (i), $\partial f(x) = A^*\partial F(Ax) = \partial f_1(x) + \partial f_2(x)$.

(iii) The inclusion $\supset$ is easy to check. We omit the proof of the reverse inclusion. (See [HUL13, VI.4.4, p.266], see also Exercise sheet 2 for a special case). □

For more on subgradients, and subdifferentials, see [SB18].

References

[HUL13] Jean-Baptiste Hiriart-Urruty and Claude Lemaréchal. *Convex analysis and minimization algorithms I: Fundamentals*, volume 305. Springer science & business media, 2013. 3

[Roc15] Ralph Tyrell Rockafellar. *Convex analysis*. Princeton university press, 2015. 3

[SB18] L. Vandenberghe S. Boyd, J. Duchi. Subgradients, 2018. Available at `http://stanford.edu/class/ee364b/lectures/subgradients_notes.pdf`. 3