Topic 15: Subgradients and Directional Derivatives

15.1 Directional derivatives and the subdifferential

The following definition adopts the notation of Fenchel [1] and Rockafellar [3]. Phelps [2] uses the notation $d^+(x)(v)$.

15.1.1 Definition *For $f\colon \boldsymbol{R}^{\mathrm{m}} \to \boldsymbol{R}^{\sharp}$, define the* ***one-sided directional derivative of f at the point x in the direction v*** *by*

$$f'(x;v) = \lim_{\lambda \downarrow 0} \frac{f(x+\lambda v) - f(x)}{\lambda},$$

allowing the values ∞ and $-\infty$, provided the limit exists.

In Example 14.1.6, $f'(0;1) = -\infty$, that is, the graph of the function becomes arbitrarily steep as we approach the boundary. This is the only way superdifferentiability fails. I prove it in Corollary 15.1.5 below.

15.1.2 Lemma *Let f be a proper convex function, let x belong to* $\operatorname{dom} f$*, let v belong to X, and let $0 < \mu < \lambda$. Then the difference quotients satisfy*

$$\frac{f(x+\mu v) - f(x)}{\mu} \leqslant \frac{f(x+\lambda v) - f(x)}{\lambda}.$$

In particular, $\lim_{\lambda \downarrow 0} \frac{f(x+\lambda v) - f(x)}{\lambda}$ exists in $\boldsymbol{R}^{\sharp}$.

Proof: Write

$$x + \mu v = \frac{\mu}{\lambda}(x + \lambda v) + \frac{\lambda - \mu}{\lambda} x,$$

a convex combination, so convexity of f yields

$$\begin{aligned} f(x+\mu v) &\leqslant \frac{\mu}{\lambda} f(x+\lambda v) + \frac{\lambda-\mu}{\lambda} f(x) \\ &= f(x) + \frac{\mu}{\lambda}\Big(f(x+\lambda v) - f(x)\Big). \end{aligned}$$

Subtract $f(x)$ from both sides and divide by $\mu > 0$ to get desired inequality. ∎

Remarkably, if f is subdifferentiable at x, then this limit is finite. To see this, rewrite the subgradient inequality

$$f(y) \geqslant f(x) + p \cdot (y - x)$$

as

$$p \cdot v \leqslant \frac{f(x+\lambda v) - f(x)}{\lambda}, \quad \text{where } y = x + \lambda v.$$

In this case, the difference quotient is bounded below by $p \cdot v$ for any $p \in \partial f(x)$, so the limit is finite.

We now show that $f'(x;\cdot)$ is a positively homogeneous convex (i.e., sublinear) function.

15.1.3 Theorem *Let f be a proper convex function on $\boldsymbol{R}^{\mathrm{m}}$. The directional derivative mapping $v \mapsto f'(x;v)$ from $\boldsymbol{R}^{\mathrm{m}}$ into $\boldsymbol{R}^{\sharp}$ satisfies the following properties.*

a. *The function $v \mapsto f'(x;v)$ is a positively homogeneous convex function (that is, sublinear) and its effective domain is a convex cone.*

b. *If f is continuous at $x \in \operatorname{dom} f$, then $v \mapsto f'(x;v)$ is continuous and finite-valued.*

Proof: (a.): It is easy to see that the function $v \mapsto f'(x;v)$ is homogeneous, as

$$\frac{f(x+\lambda\alpha v) - f(x)}{\lambda} = \alpha \frac{f(x+\lambda\alpha v) - f(x)}{\alpha\lambda},$$

and so $f'(x;\alpha v) = \alpha f'(x;v)$. This also shows that the effective domain is a cone.

For convexity, observe that

$$\begin{aligned}\frac{f(x+\lambda(\alpha v+(1-\alpha)w) - f(x)}{\lambda} &= \frac{f\big(\alpha(x+\lambda v)+(1-\alpha)(x+\lambda w)\big) - f(x)}{\lambda} \\ &\leqslant \frac{\alpha f(x+\lambda v)+(1-\alpha)f(x+\lambda w)) - f(x)}{\lambda} \\ &= \alpha\frac{f(x+\lambda v) - f(x)}{\lambda} + (1-\alpha)\frac{f(x+\lambda w) - f(x)}{\lambda},\end{aligned}$$

and letting $\lambda \downarrow 0$ yields $f'\Big(x;(\alpha v+(1-\alpha)w\Big) \leqslant \alpha d f'(x;v) + (1-\alpha)f'(x;w)$.

(b): By Lemma 6.1.3, we have

$$|f(x+\lambda v) - f(x)| \leqslant \lambda \max\{f(x+v) - f(x), f(x-v) - f(x)\}$$

for $0 < \lambda \leqslant 1$. So let $\varepsilon > 0$ be given. If f is continuous at x, there exists $\delta > 0$ such that $\|v\| < \delta$ implies $|(f(x \pm v) - f(x)| < \varepsilon$. We thus have

$$|f'(x;v)| \leqslant \frac{|f(x+\lambda v) - f(x)|}{\lambda} \leqslant \max\{f(x+v) - f(x), f(x-v) - f(x)\} < \varepsilon.$$

(Why?) That is, $f'(x;\cdot)$ is bounded on $B_\delta(0)$. By homogeneity, $f'(x;v)$ is finite for all v, so its domain is $\boldsymbol{R}^{\mathrm{m}}$, and by Theorem 6.3.3, it is continuous. ∎

Given a point x in a convex set C, the set of directions v into C at x is

$$P_C(x) = \{v \in \boldsymbol{R}^{\mathrm{m}} : (\,\exists \varepsilon > 0\,)\ [\,x + \varepsilon v \in C\,]\}$$

is a convex cone, but not necessarily a closed cone. (Think of this set for a point on the boundary of a disk—it is an open half space together with zero.) The set $x + P_C(x)$, a cone with vertex x, is what Fenchel [1, p. 41] calls the **projecting cone** of C from x.

There is an intimate relation between one-sided directional derivatives and the superdifferential, cf. Fenchel [1, Property 29, p. 81] or Rockafellar [3, Theorem 23.2, p. 216]. We start with the following extension of Theorem 14.1.9.

15.1.4 Lemma *Let f be a concave function on $\boldsymbol{R}^{\mathrm{m}}$ and let f be finite at x. Then for every $y \in \boldsymbol{R}^{\mathrm{m}}$,*

$$f(x) + f'(x; y - x) \geqslant f(y).$$

(If f is convex the inequality is reversed.)

Proof: If $y \notin \operatorname{dom} f$, then $f(y) = -\infty$, so the conclusion follows. If y belongs to the effective domain, then by concavity

$$\frac{f\Big(x + \lambda(y - x)\Big) - f(x)}{\lambda} \geqslant f(y) - f(x).$$

Letting $\lambda \downarrow 0$, the left hand side converges to $f'(x; y - x)$, which may be $+\infty$. ∎

Geometrically, this says that the hypograph of $y \mapsto f(x) + f'(x; y - x)$ includes the hypograph of f. We can use this to complete the description of subdifferentiability of f. The following result may be partially found in Fenchel [1, Property 31, p. 84] and more explicitly in Rockafellar [3, Theorem 23.3, p. 216] (which are stated for convex functions).

15.1.5 Corollary *Let f be a proper concave function on $\boldsymbol{R}^{\mathrm{m}}$, and let $x \in \operatorname{dom} f$. If $f'(x; v) < \infty$ for some v such that $x + v \in \operatorname{ri} \operatorname{dom} f$, then f is superdifferentiable at x.*

Proof: Let v satisfy $x + v \in \operatorname{ri} \operatorname{dom} f$ and $f'(x; v) < \infty$. Then, as in the proof of Proposition 14.1.5 there is $(p, -1)$ supporting the hypograph of $f'(x; \cdot)$ at the point $\Big(v, f'(x; v)\Big)$. That is,

$$p \cdot v - f'(x; v) \leqslant p \cdot u - f'(x, u) \quad \text{for all } u \in \operatorname{dom} f'(x; \cdot). \tag{1}$$

Taking $u = 0$ implies $p \cdot v - f'(x; v) \leqslant 0$. Taking $u = \lambda v$ for $\lambda > 0$ large implies $p \cdot v - f'(x; v) \geqslant 0$. Thus $f'(x; v) = p \cdot v$. Then (1) becomes

$$p(u) \geqslant f'(x; u) \quad \text{for all } u \in \operatorname{dom} f'(x; \cdot).$$

Adding $f(x)$ to both sides and applying Lemma 15.1.4, we get the supergradient inequality

$$f(x) + p(u) \geqslant f(x) + f'(x, u) \geqslant f(x + u)$$

for $u \in \operatorname{dom} f'(x; \cdot) = P_{\operatorname{dom} f}(x)$. For any u not in this set, $f(x + \lambda u) = -\infty$ for $\lambda > 0$ and the supergradient inequality holds trivially. Thus p is a supergradient of f at x. ∎

The next lemma asserts that for concave functions, the directional derivative mapping at x is the cost function of the superdifferential at x.

15.1.6 Lemma (Support function of the superdifferential) *Let f be a concave function on $\boldsymbol{R}^{\mathrm{m}}$, and let $f(x)$ be finite. Then*

$$p \in \partial f(x) \iff (\forall v \in \boldsymbol{R}^{\mathrm{m}}) \, [\, p(v) \geqslant f'(x; v) \,].$$

Let f be a convex function on $\boldsymbol{R}^{\mathrm{m}}$, and let $f(x)$ be finite. Then

$$p \in \partial f(x) \iff (\forall v \in \boldsymbol{R}^{\mathrm{m}}) \, [\, p(v) \leqslant f'(x; v) \,].$$

Proof for the concave case:: $(\Longrightarrow)$ Let $p \in \partial f(x)$. By the supergradient inequality, for any $v \in \boldsymbol{R}^{\mathrm{m}}$,

$$f(x) + p(\lambda v) \geqslant f(x + \lambda v)$$

We may subtract the finite value $f(x)$ from the right hand side, even if $x + \lambda v \notin \operatorname{dom} f$. Thus

$$p(\lambda v) \geqslant f(x + \lambda v) - f(x).$$

Dividing by $\lambda > 0$ and letting $\lambda \downarrow 0$ gives

$$p(v) \geqslant f'(x; v)$$

.

$(\Longleftarrow)$ If $p \notin \partial f(x)$, then there is some v such that the supergradient inequality is violated, that is,

$$f(x) + p(v) < f(x + v). \tag{2}$$

Since $f(x+v) = -\infty$ if $x+v \notin \operatorname{dom} f$, we conclude $x+v \in \operatorname{dom} f$. By concavity, for $0 < \lambda \leqslant 1$,

$$f(x + \lambda v) \geqslant f(x) + \lambda \Big[f(x + v) - f(x) \Big]$$

or

$$\frac{f(x + \lambda v) - f(x)}{\lambda} \geqslant f(x + v) - f(x),$$

so by (2)

$$\frac{f(x + \lambda v) - f(x)}{\lambda} \geqslant f(x + v) - f(x) > p(v),$$

so taking limits gives $f'(x; v) > p(v)$. The conclusion now follows by contraposition. ∎

The next result may be found in Rockafellar [3, Theorem 23.2, p. 216].

15.1.7 Corollary *Let f be a concave function on $\boldsymbol{R}^{\mathrm{m}}$, and let $f(x)$ be finite. Then the closure of the directional derivative at x (as a concave function of the direction) is the cost function of the superdifferential at x. That is,*

$$\operatorname{cl} f'(x;\cdot) = c_{\partial f(x)}(\cdot).$$

Proof: Since $h\colon v \mapsto f'(x;v)$ is concave and homogeneous, by Theorem 21.4.1,

$$\operatorname{cl} h = c_A, \text{ where } A = \Big\{p : (\forall v)\ )\ [\,p(v) \geqslant h(v)\,]\Big\}.$$

By Lemma 15.1.6, $A = \partial f(x)$. ∎

15.2 Supergradient of a support function

If the infimum of p is actually achieved at a point in A, we can say more. By Theorem 9.1.2 we might as well assume that A is closed and convex.

15.2.1 Theorem *Let A be a closed convex set. Then x is a supergradient of the cost function c_A at p if and only if x belongs to A and minimizes p over A. In other words,*

$$\partial c_A(p) = \{x \in A : p \cdot x = c_A(p)\}.$$

Proof: Recall that the supergradient inequality for this case is

$$c_A(p) + x \cdot (q - p) \geqslant c_A(q) \qquad \text{for all } q.$$

($\Longrightarrow$) I first claim that if x does not belong to A, it is not a supergradient of c_A at p. For if $x \notin A$, then by Theorem 9.1.2 there is some q for which $q \cdot x < c_A(q)$. Thus for $\lambda > 0$ large enough, $\lambda q \cdot x < c_A(\lambda q) + \big(p \cdot x - c_A(p)\big)$. Rearranging terms violates the supergradient inequality applied to λq. Therefore, by contraposition, if x is a supergradient of the support function c_A at p, then x belongs to A.

So let x be a supergradient of c_A at p. Setting $q = 0$ in the supergradient inequality, we conclude that $c_A(p) \geqslant p \cdot x$. But x belongs to A, so x minimizes p over A, and $c_A(p) = p \cdot x$.

In other words, $\partial c_A(p) \subset \{x \in A : p \cdot x = c_A(p)\}$

($\Longleftarrow$) Suppose now that x belongs to A and $p \cdot x = c_A(p)$, that is, x minimizes p over A. By the definition of c_A, for any $q \in \boldsymbol{R}^{\mathrm{m}}$, $q \cdot x \geqslant c_A(q)$. Now add $c_A(p) - p \cdot x = 0$ to the left-hand side of the inequality to obtain the supergradient inequality.

Thus $\{x \in A : p \cdot x = c_A(p)\} \subset \partial c_A(p)$, completing the proof. ∎

15.2.2 Corollary *Let A be a closed convex set. Suppose x belongs to A and strictly minimizes p over A. Then c_A is differentiable at p and*

$$c_A'(p) = x.$$

out of order!

Proof: This follows from Theorem 15.2.1 and Theorem 18.3.1. ∎

15.2.3 Example Let's look at $A = \{(x_1, x_2) \in \boldsymbol{R}^2_{++} : x_1 x_2 \geqslant 1\}$. This is a closed convex set and its support function is easily calculated: If $p \notin \boldsymbol{R}^2_{+}$, then $c_A(p) = -\infty$. For $p \geqq 0$, it not hard to see that $c_A(p) = 2\sqrt{p_1 p_2}$, which has no supergradient when $p_1 = 0$ or $p_2 = 0$.

(To see this, consider first the case $p \geqq 0$. The Lagrangean for the minimization problem is $p_1 x_1 + p_2 x_2 + \lambda(1 - x_1 x_2)$. By the Lagrange Multiplier Theorem, the first order conditions are $p_1 - \lambda x_1^* = 0$ and $x_2^* - \lambda q_2 = 0$. Thus $x_1^* x_2^* = \frac{p_1 p_2}{\lambda^2}$, so $\lambda = \sqrt{p_1 p_2}$. Thus $x_1^* = \sqrt{\frac{p_1}{p_2}}$ and $x_2^* = \sqrt{\frac{p_2}{p_1}}$ and $c_A(p) = p_1 x_1^* + p_2 x_2^* = 2\sqrt{p_1 p_2}$.

Now suppose some $p_i < 0$. For instance, suppose $p_2 < 0$. Then $p \cdot (\varepsilon, \frac{1}{\varepsilon}) \to -\infty$ as $\varepsilon \to 0$, so $c_A(p) = -\infty$.) □

References

[1] W. Fenchel. 1953. Convex cones, sets, and functions. Lecture notes, Princeton University, Department of Mathematics. From notes taken by D. W. Blackett, Spring 1951.

[2] R. R. Phelps. 1993. *Convex functions, monotone operators and differentiability*, 2d. ed. Number 1364 in Lecture Notes in Mathematics. Berlin: Springer–Verlag.

[3] R. T. Rockafellar. 1970. *Convex analysis*. Number 28 in Princeton Mathematical Series. Princeton: Princeton University Press.