# On subdifferential calculus *

Tieyong Zeng
zeng@math.cuhk.edu.hk
Feb, 2020

**Definition 2.30** *Let* $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ *be a convex function and let* $\bar{x} \in \operatorname{dom} f$. *An element* $v \in \mathbb{R}^n$ *is called a* SUBGRADIENT *of* $f$ *at* $\bar{x}$ *if*

$$\langle v, x - \bar{x} \rangle \leq f(x) - f(\bar{x}) \ \text{for all} \ x \in \mathbb{R}^n . \tag{2.13}$$

*The collection of all the subgradients of* $f$ *at* $\bar{x}$ *is called the* SUBDIFFERENTIAL *of the function at this point and is denoted by* $\partial f(\bar{x})$.

# Subdifferential

the *subdifferential* $\partial f(x)$ of $f$ at $x$ is the set of all subgradients:

$$\partial f(x) = \{g \mid g^T(y - x) \leq f(y) - f(x), \ \forall y \in \operatorname{dom} f\}$$

## Properties

- $\partial f(x)$ is a closed convex set (possibly empty)

  this follows from the definition: $\partial f(x)$ is an intersection of halfspaces

- if $x \in \operatorname{int} \operatorname{dom} f$ then $\partial f(x)$ is nonempty and bounded

  proof on next two pages

*Proof:* we show that $\partial f(x)$ is nonempty when $x \in \mathrm{int}\,\mathrm{dom}\, f$

- $(x, f(x))$ is in the boundary of the convex set $\mathrm{epi}\, f$

- therefore there exists a supporting hyperplane to $\mathrm{epi}\, f$ at $(x, f(x))$:

$$\exists (a,b) \neq 0, \qquad \begin{bmatrix} a \\ b \end{bmatrix}^T \left( \begin{bmatrix} y \\ t \end{bmatrix} - \begin{bmatrix} x \\ f(x) \end{bmatrix} \right) \leq 0 \qquad \forall (y,t) \in \mathrm{epi}\, f$$

- $b > 0$ gives a contradiction as $t \to \infty$

- $b = 0$ gives a contradiction for $y = x + \epsilon a$ with small $\epsilon > 0$

- therefore $b < 0$ and $g = \dfrac{1}{|b|} a$ is a subgradient of $f$ at $x$

*Proof:* $\partial f(x)$ is bounded when $x \in \mathbf{int\,dom}\, f$

- for small $r > 0$, define a set of $2n$ points

$$B = \{x \pm re_k \mid k = 1, \ldots, n\} \subset \mathbf{dom}\, f$$

  and define $M = \max_{y \in B} f(y) < \infty$

- for every $g \in \partial f(x)$, there is a point $y \in B$ with

$$r\|g\|_\infty = g^T(y - x)$$

  (choose an index $k$ with $|g_k| = \|g\|_\infty$, and take $y = x + r\,\mathrm{sign}(g_k)e_k$)

- since $g$ is a subgradient, this implies that

$$f(x) + r\|g\|_\infty = f(x) + g^T(y - x) \le f(y) \le M$$

- we conclude that $\partial f(x)$ is bounded:

$$\|g\|_\infty \le \frac{M - f(x)}{r} \quad \text{for all } g \in \partial f(x)$$

**Definition 2.34** *We say that* $f : \mathbb{R}^n \to \mathbb{R}$ *is (Fréchet)* DIFFERENTIABLE *at* $\bar{x} \in \operatorname{int}(\operatorname{dom} f)$ *if there exists an element* $v \in \mathbb{R}^n$ *such that*

$$\lim_{x \to \bar{x}} \frac{f(x) - f(\bar{x}) - \langle v, x - \bar{x} \rangle}{\|x - \bar{x}\|} = 0.$$

*In this case the element* $v$ *is uniquely defined and is denoted by* $\nabla f(\bar{x}) := v$.

**Proposition 2.35** *Let* $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ *be convex and let* $\bar{x} \in \operatorname{dom} f$. *Then* $f$ *attains its local/global minimum at* $\bar{x}$ *if and only if* $0 \in \partial f(\bar{x})$.

***Proof.*** Suppose that $f$ attains its global minimum at $\bar{x}$. Then

$$f(\bar{x}) \leq f(x) \quad \text{for all} \quad x \in \mathbb{R}^n,$$

which can be rewritten as

$$0 = \langle 0, x - \bar{x} \rangle \leq f(x) - f(\bar{x}) \quad \text{for all} \quad x \in \mathbb{R}^n \,.$$

The definition of the subdifferential shows that this is equivalent to $0 \in \partial f(\bar{x})$. □

Now we show that the subdifferential (2.13) is indeed a singleton for differentiable functions reducing to the classical derivative/gradient at the reference point and clarifying the notion of differentiability in the case of convex functions.

**Proposition 2.36** *Let* $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ *be convex and differentiable at* $\bar{x} \in \text{int}(\text{dom}\, f)$. *Then we have* $\partial f(\bar{x}) = \{\nabla f(\bar{x})\}$ *and*

$$\langle \nabla f(\bar{x}), x - \bar{x} \rangle \le f(x) - f(\bar{x}) \quad \textit{for all} \quad x \in \mathbb{R}^n . \tag{2.17}$$

**Proposition 2.36** *Let* $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ *be convex and differentiable at* $\bar{x} \in \text{int}(\text{dom } f)$. *Then we have* $\partial f(\bar{x}) = \{\nabla f(\bar{x})\}$ *and*

$$\langle \nabla f(\bar{x}), x - \bar{x} \rangle \le f(x) - f(\bar{x}) \ \textit{for all} \ \ x \in \mathbb{R}^n . \tag{2.17}$$

***Proof.*** It follows from the differentiability of $f$ at $\bar{x}$ that for any $\epsilon > 0$ there is $\delta > 0$ with

$$-\epsilon\|x - \bar{x}\| \le f(x) - f(\bar{x}) - \langle \nabla f(\bar{x}), x - \bar{x} \rangle \le \epsilon\|x - \bar{x}\| \ \text{ whenever } \ \|x - \bar{x}\| < \delta. \tag{2.18}$$

Consider further the convex function

$$\varphi(x) := f(x) - f(\bar{x}) - \langle \nabla f(\bar{x}), x - \bar{x} \rangle + \epsilon\|x - \bar{x}\|, \quad x \in \mathbb{R}^n,$$

and observe that $\varphi(x) \ge \varphi(\bar{x}) = 0$ for all $x \in I\!B(\bar{x}; \delta)$. The convexity of $\varphi$ ensures that $\varphi(x) \ge \varphi(\bar{x})$ for all $x \in \mathbb{R}^n$. Thus

$$\langle \nabla f(\bar{x}), x - \bar{x} \rangle \le f(x) - f(\bar{x}) + \epsilon\|x - \bar{x}\| \ \text{ whenever } \ x \in \mathbb{R}^n,$$

which yields (2.17) by letting $\epsilon \downarrow 0$.

It follows from (2.17) that $\nabla f(\bar{x}) \in \partial f(\bar{x})$. Picking now $v \in \partial f(\bar{x})$, we get

$$\langle v, x - \bar{x} \rangle \le f(x) - f(\bar{x}).$$

Then the second part of (2.18) gives us that

$$\langle v - \nabla f(\bar{x}), x - \bar{x} \rangle \le \epsilon\|x - \bar{x}\| \ \text{ whenever } \ \|x - \bar{x}\| < \delta.$$

Finally, we observe that $\|v - \nabla f(\bar{x})\| \le \epsilon$, which yields $v = \nabla f(\bar{x})$ since $\epsilon > 0$ was chosen arbitrarily. Thus $\partial f(\bar{x}) = \{\nabla f(\bar{x})\}$. □

**Example 2.38** Let $p(x) := \|x\|$ be the Euclidean norm function on $\mathbb{R}^n$. Then we have

$$\partial p(x) = \begin{cases} I\!B & \text{if } x = 0, \\ \left\{\dfrac{x}{\|x\|}\right\} & \text{otherwise.} \end{cases}$$

To verify this, observe first that the Euclidean norm function $p$ is differentiable at any nonzero point with $\nabla p(x) = \dfrac{x}{\|x\|}$ as $x \neq 0$. It remains to calculate its subdifferential at $x = 0$. To proceed by definition (2.13), we have that $v \in \partial p(0)$ if and only if

$$\langle v, x \rangle = \langle v, x - 0 \rangle \leq p(x) - p(0) = \|x\| \text{ for all } x \in \mathbb{R}^n .$$

Letting $x = v$ gives us $\langle v, v \rangle \leq \|v\|$, which implies that $\|v\| \leq 1$, i.e., $v \in I\!B$. Now take $v \in I\!B$ and deduce from the Cauchy-Schwarz inequality that

$$\langle v, x - 0 \rangle = \langle v, x \rangle \leq \|v\| \cdot \|x\| \leq \|x\| = p(x) - p(0) \text{ for all } x \in \mathbb{R}^n$$

and thus $v \in \partial p(0)$, which shows that $\partial p(0) = I\!B$.

**Theorem 2.40** Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a differentiable function on its domain $D$, which is an open convex set. Then $f$ is convex if and only if

$$\langle \nabla f(u), x - u \rangle \leq f(x) - f(u) \text{ for all } x, u \in D. \tag{2.21}$$

***Proof.*** The "only if" part follows from Proposition 2.36. To justify the converse, suppose that (2.21) holds and then fix any $x_1, x_2 \in D$ and $t \in (0, 1)$. Denoting $x_t := tx_1 + (1 - t)x_2$, we have $x_t \in D$ by the convexity of $D$. Then

$$\langle \nabla f(x_t), x_1 - x_t \rangle \leq f(x_1) - f(x_t), \quad \langle \nabla f(x_t), x_2 - x_t \rangle \leq f(x_2) - f(x_t).$$

It follows furthermore that

$$\begin{aligned} &t\langle \nabla f(x_t), x_1 - x_t \rangle \leq tf(x_1) - tf(x_t) \text{ and} \\ &(1-t)\langle \nabla f(x_t), x_2 - x_t \rangle \leq (1-t)f(x_2) - (1-t)f(x_t). \end{aligned}$$

Summing up these inequalities, we arrive at

$$0 \leq tf(x_1) + (1-t)f(x_2) - f(x_t),$$

which ensures that $f(x_t) \leq tf(x_1) + (1-t)f(x_2)$, and so verifies the convexity of $f$. □

*Moreau-Rockafellar theorem*

**Corollary 2.45** *Let* $f_i : \mathbb{R}^n \to \overline{\mathbb{R}}$ *for* $i = 1, 2$ *be convex functions such that there exists* $u \in \operatorname{dom} f_1 \cap \operatorname{dom} f_2$ *for which* $f_1$ *is continuous at* $u$ *or* $f_2$ *is continuous at* $u$. *Then*

$$\partial(f_1 + f_2)(x) = \partial f_1(x) + \partial f_2(x) \tag{2.28}$$

*whenever* $x \in \operatorname{dom} f_1 \cap \operatorname{dom} f_2$. *Consequently, if both functions* $f_i$ *are finite-valued on* $\mathbb{R}^n$, *then the sum rule* (2.28) *holds for all* $x \in \mathbb{R}^n$.

**Theorem 2.9 (Moreau-Rockafellar)** *Let* $f, g : \mathbb{R}^n \to (-\infty, +\infty]$ *be convex functions. Then for every* $x_0 \in \mathbb{R}^n$

$$\partial f(x_0) + \partial g(x_0) \subset \partial(f+g)(x_0).$$

*Moreover, suppose that* $\operatorname{int} \operatorname{dom} f \cap \operatorname{dom} g \neq \emptyset$*. Then for every* $x_0 \in \mathbb{R}^n$ *also*

$$\partial(f+g)(x_0) \subset \partial f(x_0) + \partial g(x_0).$$

PROOF. The proof of the first part is elementary: Let $\xi_1 \in \partial f(x_0)$ and $\xi_2 \in \partial g(x_0)$. Then for all $x \in \mathbb{R}^n$

$$f(x) \geq f(x_0) + \xi_1^t(x - x_0), \; g(x) \geq g(x_0) + \xi_2^t(x - x_0),$$

so addition gives $f(x) + g(x) \geq f(x_0) + g(x_0) + (\xi_1 + \xi_2)^t(x - x_0)$. Hence $\xi_1 + \xi_2 \in \partial(f + g)(x_0)$.

To prove the second part, let $\xi \in \partial(f + g)(x_0)$. First, observe that $f(x_0) = +\infty$ implies $(f + g)(x_0) = +\infty$, whence $f + g \equiv +\infty$, which is impossible by $\xi \in \partial(f + g)(x_0)$. Likewise, $g(x_0) = +\infty$ is impossible. Hence, from now on we know that both $f(x_0)$ and $g(x_0)$ belong to $\mathbb{R}$. We form the following two sets in $\mathbb{R}^{n+1}$.

$$\Lambda_f := \{(x - x_0, y) \in \mathbb{R}^n \times \mathbb{R} : y > f(x) - f(x_0) - \xi^t(x - x_0)\}$$

$$\Lambda_g := \{(x - x_0, y) : -y \geq g(x) - g(x_0)\}.$$

$$\Lambda_f := \{(x - x_0, y) \in \mathbb{R}^n \times \mathbb{R} : y > f(x) - f(x_0) - \xi^t(x - x_0)\}$$

$$\Lambda_g := \{(x - x_0, y) : -y \geq g(x) - g(x_0)\}.$$

Observe that both sets are nonempty and convex (see Exercise 2.8), and that $\Lambda_f \cap \Lambda_g = \emptyset$ (the latter follows from $\xi \in \partial(f + g)(x_0)$). Hence, by the set-set-separation Theorem A.4, there exists $(\xi_0, \mu) \in \mathbb{R}^{n+1}$ and $\alpha \in \mathbb{R}$, $(\xi_0, \mu) \neq (0, 0)$, such that

$$\xi_0^t(x - x_0) + \mu y \leq \alpha \text{ for all } (x, y) \text{ with } y > f(x) - f(x_0) - \xi^t(x - x_0),$$

$$\xi_0^t(x - x_0) + \mu y \geq \alpha \text{ for all } (x, y) \text{ with } -y \geq g(x) - g(x_0).$$

By $(0, 0) \in \Lambda_g$ we get $\alpha \leq 0$. But also $(0, \epsilon) \in \Lambda_f$ for every $\epsilon > 0$, and this gives $\mu\epsilon \leq \alpha$, so $\mu \leq 0$ (take $\epsilon = 1$). In the limit, for $\epsilon \to 0$, we find $\alpha \geq 0$. Hence $\alpha = 0$ and $\mu \leq 0$. We now claim that $\mu = 0$ is impossible. Indeed, if one had $\mu = 0$, then the first of the above two inequalities would give

$$\xi_0^t(x - x_0) \leq 0 \text{ for all } (x, y) \text{ with } y > f(x) - f(x_0) - \xi^t(x - x_0),$$

which is equivalent to

$$\xi_0^t(x - x_0) \leq 0 \text{ for all } x \in \text{dom } f$$

(simply note that when $f(x) < +\infty$ one can always achieve $y > f(x) - f(x_0) - \xi^t(x - x_0)$ by choosing $y$ sufficiently large). Likewise, the second inequality would give

$$\xi_0^t(x - x_0) \geq 0 \text{ for all } x \in \text{dom } g.$$

In particular, for $\tilde{x}$ as above this would imply $\xi_0^t(\tilde{x} - x_0) = 0$. But since $\tilde{x}$ lies in the interior of dom $f$ (so for some $\delta > 0$ the ball $N_\delta(\tilde{x})$ belongs to dom $f$), the preceding would imply

$$\xi_0^t u = \xi_0^t(\tilde{x} + u - x_0) \leq 0 \text{ for all } u \in N_\delta(0).$$

Clearly, this would give $\xi_0 = 0$ (take $u := \delta\xi_0/2$), which would be in contradiction to $(\xi_0, \mu) \neq (0, 0)$. Hence, we conclude $\mu < 0$. Dividing the separation inequalities by $-\mu$ and setting $\bar{\xi}_0 := -\xi_0/\mu$, this results in

$$\bar{\xi}_0^t(x - x_0) \leq y \text{ for all } (x, y) \text{ with } y > f(x) - f(x_0) - \xi^t(x - x_0),$$

$$\bar{\xi}_0^t(x - x_0) \geq y \text{ for all } (x, y) \text{ with } -y \geq g(x) - g(x_0).$$

The last inequality gives $-\bar{\xi}_0 \in \partial g(x_0)$ (set $y := g(x_0) - g(x)$) and the one but last inequality gives $\xi + \bar{\xi}_0 \in \partial f(x_0)$ (take $y := f(x) - f(x_0) - \xi^t(x - x_0) + \epsilon$ and let $\epsilon \downarrow 0$). Since $\xi = (\xi + \bar{\xi}_0) - \bar{\xi}_0$, this finishes the proof. QED

As a precursor to the Karush-Kuhn-Tucker theorem, we have now the following application of the Moreau-Rockafellar theorem.

**Theorem 2.10** *Let* $f : \mathbb{R}^n \to \mathbb{R}$ *be a convex function and let* $S \subset \mathbb{R}^n$ *be a nonempty convex set. Consider the optimization problem*

$$(P) \quad \inf_{x \in S} f(x).$$

*Then* $\bar{x} \in S$ *is an optimal solution of* $(P)$ *if and only if there exists a subgradient* $\bar{\xi} \in \partial f(\bar{x})$ *such that*

$$\bar{\xi}^t (x - \bar{x}) \geq 0 \text{ for all } x \in S. \tag{1}$$

Proof. Recall from Definition 2.3 that $\chi_S$ is the indicator function of $S$. Now let $\bar{x} \in S$ be arbitrary. Then the following is trivial: $\bar{x}$ is an optimal solution of $(P)$ if and only if

$$0 \in \partial(f + \chi_S)(\bar{x}).$$

By the Moreau-Rockafellar Theorem 2.9, we have

$$\partial(f + \chi_S)(\bar{x}) = \partial f(\bar{x}) + \partial\chi_S(\bar{x}).$$

To see that its conditions hold, observe that dom $f = \mathbb{R}^n$ and dom $\chi_S = S$. So it follows that $\bar{x}$ is an optimal solution of $(P)$ if and only if $0 \in \partial f(\bar{x}) + \partial\chi_S(\bar{x})$. By the definition of the sum of two sets this means that $\bar{x}$ is an optimal solution of $(P)$ if and only if $0 = \bar{\xi} + \bar{\xi}'$ for some $\bar{\xi} \in \partial f(\bar{x})$ and $\bar{\xi}' \in \partial\chi_S(\bar{x})$. Of course, the former means $\bar{\xi}' = -\bar{\xi}$, so $-\bar{\xi} \in \partial\chi_S(\bar{x})$, which is equivalent to

$$\chi_S(x) \geq \chi_S(\bar{x}) + (-\bar{\xi})^t(x - \bar{x}) \text{ for all } x \in \mathbb{R}^n,$$

i.e., to (1). QED

**Definition 2.13** The *directional derivative* of a convex function $f : \mathbb{R}^n \to (-\infty, +\infty]$ at the point $x_0 \in \mathrm{dom} f$ in the direction $d \in \mathbb{R}^n$ is defined as

$$f'(x_0; d) := \lim_{\lambda \downarrow 0} \frac{f(x_0 + \lambda d) - f(x_0)}{\lambda}.$$

The above limit is a well-defined number in $[-\infty, +\infty]$. This follows from the following proposition (why?), which shows that the difference quotients of a convex functions possess a monotonicity property:

**Proposition 2.14** *Let $f : \mathbb{R}^n \to (-\infty, +\infty]$ be a convex function and let $x_0$ be a point in* dom$f$*. Then for every direction $d \in \mathbb{R}^n$ and every $\lambda_1, \lambda_2 \in \mathbb{R}$ such that $\lambda_2 > \lambda_1 > 0$ we have*

$$\frac{f(x_0 + \lambda_1 d) - f(x_0)}{\lambda_1} \leq \frac{f(x_0 + \lambda_2 d) - f(x_0)}{\lambda_2}$$

PROOF. Note that

$$x_0 + \lambda_1 d = \frac{\lambda_1}{\lambda_2}(x_0 + \lambda_2 d) + (1 - \frac{\lambda_1}{\lambda_2})x_0.$$

So by convexity of $f$

$$f(x_0 + \lambda_1 d) \leq \frac{\lambda_1}{\lambda_2} f(x_0 + \lambda_2 d) + (1 - \frac{\lambda_1}{\lambda_2}) f(x_0).$$

Simple algebra shows that this is equivalent to the desired inequality. QED

**Theorem 2.15** *Let $f : \mathbb{R}^n \to (-\infty, +\infty]$ be a convex function and let $x_0$ be a point in* int dom $f$*. Then*

$$f'(x_0; d) = \sup_{\xi \in \partial f(x_0)} \xi^t d \text{ for every } d \in \mathbb{R}^n.$$

Proof of Theorem 2.15. By Proposition 2.14

$$q(d) := f'(x_0; d) := \lim_{\lambda \downarrow 0} \frac{f(x_0 + \lambda d) - f(x_0)}{\lambda} = \inf_{\lambda > 0} \frac{f(x_0 + \lambda d) - f(x_0)}{\lambda}.$$

Since the pointwise limit of a sequence of convex functions is convex, it follows that $q : \mathbb{R}^n \to \mathbb{R}$ is convex (by the infimum expression for $q(d)$ the fact that $x_0 \in \text{int dom } f$ implies automatically $q(d) < +\infty$ for every $d$; also, $q(d) > -\infty$ for every $d$, because of the nonemptiness part of Lemma 2.16). Hence, $q$ is continuous at every point $d \in \mathbb{R}^n$ (apply the continuity part of Lemma 2.16). So by the Fenchel-Moreau theorem (Theorem B.5 in the Appendix) we have for every $d$

$$q(d) = q^{**}(d) := \sup_{\xi \in \mathbb{R}^n} [d^t \xi - q^*(\xi)].$$

Let us calculate $q^*$. For any $\xi \in \mathbb{R}^n$ we have

$$q^*(\xi) := \sup_{d \in \mathbb{R}^n} [\xi^t d - q(d)] = \sup_{d, \lambda > 0} [\xi^t d - \frac{f(x_0 + \lambda d) - f(x_0)}{\lambda}] = \sup_{\lambda > 0} \sup_{d} [\xi^t d - \frac{f(x_0 + \lambda d) - f(x_0)}{\lambda}]$$

by the above infimum expression for $q(d)$. Fix $\lambda > 0$; then $z := x_0 + \lambda d$ runs through all of $\mathbb{R}^n$ as $d$ runs through $\mathbb{R}^n$. Hence

$$\sup_d [\xi^t d - \frac{f(x_0 + \lambda d) - f(x_0)}{\lambda}] = \frac{f(x_0) - \xi^t x_0 + \sup_z [\xi^t z - f(z)]}{\lambda}.$$

Clearly, this gives

$$q^*(\xi) = \sup_{\lambda > 0} \frac{f(x_0) - \xi^t x_0 + f^*(\xi)}{\lambda} = \begin{cases} 0 & \text{if } \xi \in \partial f(x_0) \\ +\infty & \text{otherwise} \end{cases}$$

where we use Proposition B.4($v$). Observe that in terms of the indicator function of the subdifferential this can be rewritten as $q^* = \chi_{\partial f(x_0)}$. Now that $q^*$ has been calculated, we conclude from the above that for every $d \in \mathbb{R}^n$

$$f'(x_0; d) = q(d) = q^{**}(d) = \chi^*_{\partial f(x_0)}(d) = \sup_{\xi \in \partial f(x_0)} \xi^t d,$$

which proves the result. QED

**Proposition 2.54** *Let* $f_i \colon \mathbb{R}^n \to \overline{\mathbb{R}}$, $i = 1, \ldots, m$, *be convex functions. Take any point* $\bar{x} \in \bigcap_{i=1}^m \operatorname{dom} f_i$ *and assume that each* $f_i$ *is continuous at* $\bar{x}$. *Then we have the maximum rule*

$$\partial\big(\max f_i\big)(\bar{x}) = \operatorname{co} \bigcup_{i \in I(\bar{x})} \partial f_i(\bar{x}).$$

**Theorem 2.17 (Dubovitskii-Milyutin)** *Let* $f_1, \cdots, f_m : \mathbb{R}^n \to (-\infty, +\infty]$ *be convex functions and let* $x_0$ *be a point in* $\cap_{i=1}^m \mathrm{int\ dom}\, f_i$. *Let* $f : \mathbb{R}^n \to (-\infty, +\infty]$ *be given by*

$$f(x) := \max_{1 \leq i \leq m} f_i(x)$$

*and let* $I(x_0)$ *be the (nonempty) set of all* $i \in \{1, \cdots, m\}$ *for which* $f_i(x_0) = f(x_0)$. *Then*

$$\partial f(x_0) = \mathrm{co}\ \cup_{i \in I(x_0)} \partial f_i(x_0).$$

PROOF. For our convenience we write $I := I(x_0)$. To begin with, observe that $\xi \in \partial f_i(x_0)$ easily implies $\xi \in \partial f(x_0)$ for each $i \in I$. Since $\partial f(x_0)$ is evidently convex, the inclusion "$\supset$" follows with ease. To prove the opposite inclusion, let $\xi_0$ be arbitrary in $\partial f(x_0)$. If $\xi_0$ were not to belong to the compact set co $\cup_{i \in I} \partial f_i(x_0)$, then we could separate strictly (note that each set $\partial f_i(x_0)$ is both closed and compact (exercise)): by Theorem A.2 there would exist $d \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$ such that

$$\xi_0^t d > \alpha \geq \max_{i \in I} \sup_{\xi \in \partial f_i(x_0)} \xi^t d = \max_{i \in I} f_i'(x_0; d),$$

where the final identity follows from Theorem 2.15. But now observe that

$$f'(x_0; d) := \lim_{\lambda \downarrow 0} \max_{i \in I} \frac{f_i(x_0 + \lambda d) - f_i(x_0)}{\lambda} = \max_{i \in I} \lim_{\lambda \downarrow 0} \frac{f_i(x_0 + \lambda d) - f_i(x_0)}{\lambda} = \max_{i \in I} f_i'(x_0; d),$$

so the above gives $\xi_0^t d > f'(x_0; d)$. On the other hand, by $\xi_0 \in \partial f(x_0)$ it follows that $f(x_0 + \lambda d) \geq f(x_0) + \lambda \xi_0^t d$ for every $\lambda > 0$, whence $f'(x_0; d) \geq \xi_0^t d$. We thus have arrived at a contradiction. So the inclusion "$\subset$" must hold as well. QED

# Directional derivative

**Definition** (for general $f$): the *directional derivative* of $f$ at $x$ in the direction $y$ is

$$\begin{aligned} f'(x;y) &= \lim_{\alpha \searrow 0} \frac{f(x+\alpha y) - f(x)}{\alpha} \\ &= \lim_{t \to \infty} \left( t(f(x + \frac{1}{t}y) - t f(x) \right) \end{aligned}$$

(if the limit exists)

- $f'(x;y)$ is the right derivative of $g(\alpha) = f(x + \alpha y)$ at $\alpha = 0$
- $f'(x;y)$ is homogeneous in $y$:

$$f'(x;\lambda y) = \lambda f'(x;y) \quad \text{for } \lambda \geq 0$$

## Directional derivative of a convex function

**Equivalent definition** (for convex $f$): replace $\lim$ with $\inf$

$$
\begin{aligned}
f'(x;y) &= \inf_{\alpha>0} \frac{f(x+\alpha y) - f(x)}{\alpha} \\
&= \inf_{t>0} \left( t f(x + \frac{1}{t} y) - t f(x) \right)
\end{aligned}
$$

*Proof*

- the function $h(y) = f(x+y) - f(x)$ is convex in $y$, with $h(0) = 0$
- its perspective $t h(y/t)$ is nonincreasing in $t$ (ECE236B ex. A2.5); hence

$$
f'(x;y) = \lim_{t\to\infty} t h(y/t) = \inf_{t>0} t h(y/t)
$$

## Properties

consequences of the expressions (for convex $f$)

$$\begin{aligned} f'(x;y) &= \inf_{\alpha>0} \frac{f(x+\alpha y)-f(x)}{\alpha} \\ &= \inf_{t>0} \left( t f(x+\frac{1}{t}y) - t f(x) \right) \end{aligned}$$

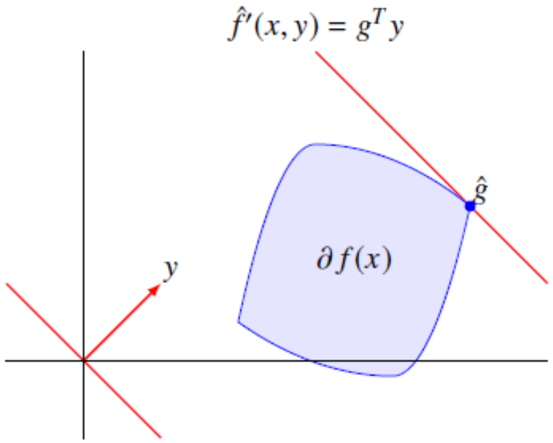
- $f'(x;y)$ is convex in $y$ (partial minimization of a convex function in $y$, $t$)
- $f'(x;y)$ defines a lower bound on $f$ in the direction $y$:

$$f(x+\alpha y) \geq f(x) + \alpha f'(x;y) \quad \text{for all } \alpha \geq 0$$

# Directional derivative and subgradients

for convex $f$ and $x \in \operatorname{int} \operatorname{dom} f$

$$f'(x; y) = \sup_{g \in \partial f(x)} g^T y$$




$f'(x; y)$ is *support function* of $\partial f(x)$

- generalizes $f'(x; y) = \nabla f(x)^T y$ for differentiable functions
- implies that $f'(x; y)$ exists for all $x \in \operatorname{int} \operatorname{dom} f$, all $y$ (see page 2.4)

*Proof:* if $g \in \partial f(x)$ then from page 2.29

$$f'(x;y) \geq \inf_{\alpha>0} \frac{f(x) + \alpha g^T y - f(x)}{\alpha} = g^T y$$

it remains to show that $f'(x;y) = \hat{g}^T y$ for at least one $\hat{g} \in \partial f(x)$

- $f'(x;y)$ is convex in $y$ with domain $\mathbf{R}^n$, hence subdifferentiable at all $y$
- let $\hat{g}$ be a subgradient of $f'(x;y)$ at $y$: then for all $v$, $\lambda \geq 0$,

$$\lambda f'(x;v) = f'(x;\lambda v) \geq f'(x;y) + \hat{g}^T(\lambda v - y)$$

- taking $\lambda \to \infty$ shows that $f'(x;v) \geq \hat{g}^T v$; from the lower bound on page 2.30,

$$f(x+v) \geq f(x) + f'(x;v) \geq f(x) + \hat{g}^T v \quad \text{for all } v$$

  hence $\hat{g} \in \partial f(x)$

- taking $\lambda = 0$ we see that $f'(x;y) \leq \hat{g}^T y$