Chapter A: Convex Analysis

Ernest K. Ryu

MATH 273A: Optimization Theory
University of California, Los Angeles
Department of Mathematics

Last edited: October 31, 2025

Line segment

Given $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^n$,

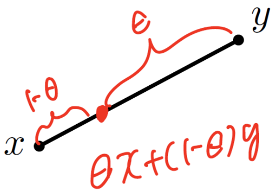
$$\theta x + (1 - \theta) y$$

is a point in between x and y if $\theta \in [0, 1]$.

The set of all points between a given $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^n$

$$\{\theta x + (1 - \theta) y \mid \theta \in [0, 1]\}$$

is called the *line segment* between x and y



Update figure

Convex combinations

Given $x_1, \ldots, x_k \in \mathbb{R}^n$,

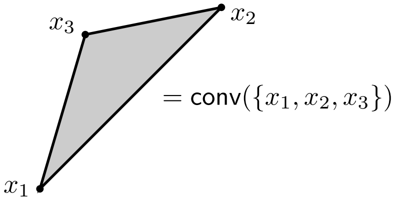
$$\theta_1 x_1 + \cdots + \theta_k x_k$$

is called a *convex combination* or a *weighted average* of $x_1, \ldots, x_k$ if $\theta_1, \ldots, \theta_k \geq 0$ and $\theta_1 + \cdots + \theta_k = 1$.

Given $x_1, \ldots, x_k \in \mathbb{R}^n$, the set of all convex combinations

$$\mathrm{conv}(\{x_1, \ldots, x_k\}) = \{\theta_1 x_1 + \cdots + \theta_k x_k \mid \theta_1, \ldots, \theta_k \geq 0,\ \theta_1 + \cdots + \theta_k = 1\}$$

is called the *convex hull* of $x_1, \ldots, x_k$.

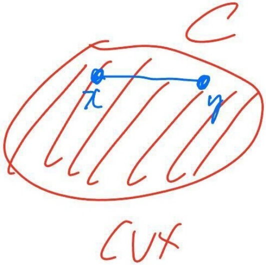


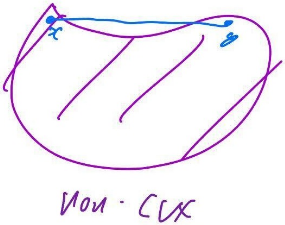
Convex sets

We say a set $C \subseteq \mathbb{R}^n$ is *convex* if

$$\theta x + (1-\theta)y \in C, \qquad \forall\, x, y \in C,\, \theta \in (0,1).$$

In other words, C is convex if $x, y \in C$ implies the line segment connecting x and y is wholly contained in C.



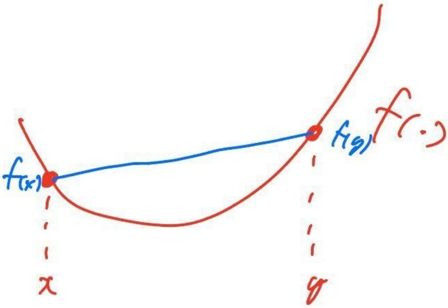


Convex functions

We say a function $f\colon \mathbb{R}^n \to \mathbb{R}$ is *convex* if

$$f(\theta x + (1-\theta)y) \le \theta f(x) + (1-\theta) f(y), \qquad \forall\, x, y \in \mathbb{R}^n,\ \theta \in [0,1].$$

I.e., f is convex if the chord (line segment) connecting $(x, f(x))$ and $(y, f(y))$ lies above the graph of f.



We say $f\colon \mathbb{R}^n \to \mathbb{R}$ is *concave* if $-f$ is convex.

Strictly convex functions

Recall that $f\colon \mathbb{R}^n \to \mathbb{R}$ is convex if

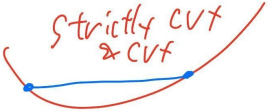
$$f(\theta x + (1-\theta)y) \le \theta f(x) + (1-\theta) f(y), \qquad \forall\, x, y \in C,\; x \ne y,\; \theta \in (0,1).$$

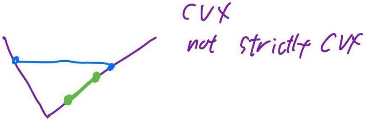
(Our prior definition of convexity is equivalent to this.)

We say $f\colon \mathbb{R}^n \to \mathbb{R}$ is *strictly convex* if

$$f(\theta x + (1-\theta)y) < \theta f(x) + (1-\theta) f(y), \qquad \forall\, x, y \in C,\; x \ne y,\; \theta \in (0,1).$$

I.e., f is strictly convex if the chord connecting $(x, f(x))$ and $(y, f(y))$ lies strictly above the graph of f (excluding the endpoints).



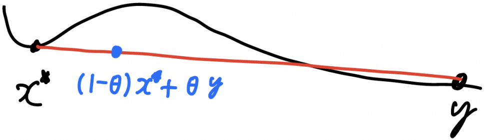


No bad local minima for cvx. functions

Theorem.
Let f be convex. Then any local minimizer is a global minimizer.

Thus, when we minimize convex functions, we never get stuck at bad local minima because there aren't any bad local minima.

Illustration of proof. Let $x_\star$ be a local minimizer. Assume for contradiction that $x_\star$ is not a global minimum.



Draw a contradiction because the chord is below the graph for $\theta \approx 0$.

No bad local minima for cvx. functions

Proof. Let $x_\star \in \mathbb{R}^n$ be a local minimizer of f. Assume for contradiction that there is $y \in \mathbb{R}^n$ such that $f(y) < f(x_\star)$, i.e., assume for contradiction that $x_\star$ is not a global minimizer. By convexity,

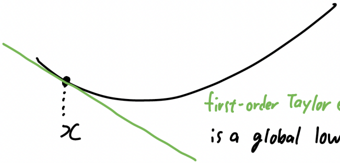
$$f((1-\theta)x_\star + \theta y) \leq (1-\theta)f(x_\star) + \theta f(y) < f(x_\star)$$

for any $\theta \in (0,1)$, even for θ very close to 0. However, $x_\star$ is a local minimizer, so $f((1-\theta)x_\star + \theta y) \geq f(x_\star)$ for θ sufficiently close to 0, and we have a contradiction. Thus we conclude that such y cannot exist, i.e., $x_\star$ is a global minimizer. $\square$

Gradient provides global lower bound for cvx. functions

Theorem.
Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex. Assume f is differentiable at $x \in \mathbb{R}^n$. Then,

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle, \qquad \forall\, y \in \mathbb{R}^n.$$



first-order Taylor expansion of f about x
is a global lower bound of f.

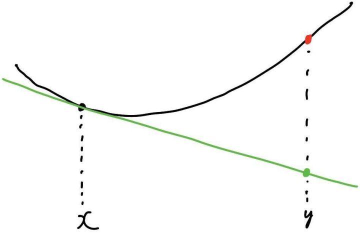
Gradient provides global lower bound for cvx. functions

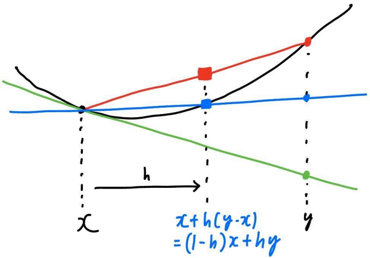
Theorem.

Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex. Assume f is differentiable at $x \in \mathbb{R}^n$. Then,

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle, \qquad \forall\, y \in \mathbb{R}^n.$$

Illustration of proof.





Q) Why is ● ≤ ● ?

- Green line is the limit of blue line as $h \to 0$.
- By convexity, ■ ≤ ■.
- Since and are similar triangles, $f(\cdot) \leq f(\cdot)$.
- Finally, ● $= \lim_{h \to 0}$ ● ≤ ●.

Gradient provides global lower bound for cvx. functions

Theorem.

Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be convex. Assume f is differentiable at x. Then,

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle, \qquad \forall\, y \in \mathbb{R}^n.$$

Proof. By convexity,

$$f(x + \theta(y - x)) \leq (1 - \theta) f(x) + \theta f(y), \qquad \forall\, \theta \in (0, 1).$$

Reorganizing, we get

$$f(y) \geq f(x) + \frac{f(x + \theta(y - x)) - f(x)}{\theta}, \qquad \forall\, \theta \in (0, 1).$$

By taking $\theta \to 0$, we get the directional derivative of f at x in direction $(y - x)$ and arrive at the desired inequality. $\square$

Gradient provides global lower bound for cvx. functions

The inequality

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

is called the *convexity inequality*.

In fact, the convexity inequality can be thought of as a defining property of convexity, rather than a consequence of convexity. In particular, a differentiable $f\colon \mathbb{R}^n \to \mathbb{R}$ is convex if and only if it satisfies the convexity inequality everywhere.

No bad stationary point for cvx. functions

Corollary.

Let $f\colon \mathbb{R}^n \to \mathbb{R}$ *be convex. If* f *is differentiable at* x *and* $\nabla f(x) = 0$*, then* $x \in \operatorname{argmin} f$*.*

Proof. By the convexity inequality, $f(y) \geq f(x)$ for all $y \in \mathbb{R}^n$. $\square$

Basic calculus of convex sets and functions

Theorem.
The intersection of convex sets is convex.

Theorem.
A nonnegative combination of convex functions is convex.

Theorem.
A sublevel set of a convex function is convex.

Basic calculus of convex sets and functions

Theorem.
The intersection of convex sets is convex.

So if $A \subseteq \mathbb{R}^n$ and $B \subseteq \mathbb{R}^n$ are convex sets, then $A \cap B$ is convex.

- The intersection can be arbitrary, i.e., the intersection can be over countably or uncountably infinite convex sets.
- To clarify, an empty set is defined to be a convex set, and the intersection of convex sets can be empty.

Basic calculus of convex sets and functions

Theorem.
A nonnegative combination of convex functions is convex.

I.e., if $\alpha_1, \ldots, \alpha_k$ are nonnegative scalars and $f_1, \ldots, f_k$ are convex functions, then $\alpha_1 f_1 + \cdots + \alpha_k f_k$ is convex.

- If f is convex, then αf is convex and $-\alpha f$ is concave if $\alpha \geq 0$.
- Often, one shows that an f is convex by arguing that $f = g + h$ and showing that g and h are convex.

Basic calculus of convex sets and functions

Theorem.
A sublevel set of a convex function is convex.

For any $f\colon \mathbb{R}^n \to \mathbb{R}$ and $\alpha \in \mathbb{R}$, the α-sublevel set of f is defined as

$$\{x \mid f(x) \leq \alpha\} \subseteq \mathbb{R}^n,$$

which is the set of x attaining function value better than α.

- In particular, this implies that the set of minimizers of a convex function is convex, i.e., if f is convex, then $\operatorname{argmin} f$ is convex.
- Often, one shows that a set is convex by showing that it is a sublevel set of a convex function.

Convexity via monotonicity

For differentiable f, convexity is monotonicity of f'.

Theorem.

A differentiable univariate function $f : \mathbb{R} \to \mathbb{R}$ *is convex if and only if* f' *is non-decreasing.*

(To clarify, convex functions need not be differentiable.)

Convexity via monotonicity

For differentiable f, convexity is monotonicity of f'.

Theorem.

A differentiable univariate function $f : \mathbb{R} \to \mathbb{R}$ is convex if and only if f' is non-decreasing.

Proof. ($\Rightarrow$) Assume f is convex. Then, by the convexity inequality,

$$f(y) \geq f(x) + f'(x)(y - x)$$
$$f(x) \geq f(y) + f'(y)(x - y)$$

for all $x, y \in \mathbb{R}$. Adding the two, we get

$$(f'(x) - f'(y))(x - y) \geq 0,$$

which implies $f'(x) \geq f'(y)$ if $x > y$, i.e. f' is non-decreasing.

Convexity via monotonicity

($\Leftarrow$) Assume $f' \colon \mathbb{R} \to \mathbb{R}$ is non-decreasing. Let $x \leq y$ and $z = \theta x + (1-\theta)y$ with $\theta \in [0,1]$. So, $x \leq z \leq y$. Then, we can show the convexity inequalities about z:

$$\begin{aligned} f(y) - f(z) = \int_z^y f'(t)\, dt &\geq \int_z^y f'(z)\, dt = f'(z)(y-z) \\ &= f'(z)\theta(y-x) \\ f(x) - f(z) = -\int_x^z f'(t)\, dt &\geq -\int_x^z f'(z)\, dt = f'(z)(x-z) \\ &= f'(z)(1-\theta)(x-y). \end{aligned}$$

Now we can combine these convexity inequalities to obtain the definition of convexity: Multiplying the first inequality by $(1-\theta)$ and the second my θ and adding them gives us

$$\theta f(x) + (1-\theta) f(y) - f(z) \geq 0.$$

$\square$

Convexity via curvature

For twice-differentiable f, convexity is positive (nonnegative) curvature.

Theorem.

A twice-differentiable univariate function $f : \mathbb{R} \to \mathbb{R}$ *is convex if and only if* $f''(x) \geq 0$ *for all* $x \in \mathbb{R}$.

Proof. From the previous theorem, f is convex if and only if f' is non-decreasing. Since f' is assumed to be differentiable, this holds if and only if $f'' \geq 0$. □

Positive semidefinite matrices

We say a matrix $A \in \mathbb{R}^{n \times n}$ is symmetric positive semidefinite and write

$$A \succeq 0$$

if A is symmetric and all eigenvalues of A are nonnegative, i.e., if $A = A^\intercal$ and $\lambda_{\min}(A) \geq 0$.

Lemma.

Let $A \in \mathbb{R}^{n \times n}$ and $A = A^\intercal$. Then, $A \succeq 0$ if and only if

$$v^\intercal A v \geq 0, \qquad \forall\, v \in \mathbb{R}^n.$$

Proof follows from the spectral theorem.

Convexity on lines

Lemma.
Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be convex and $x, v \in \mathbb{R}^n$. Then, $g(t) = f(x + tv)$ is convex for any $v \in \mathbb{R}^n$.

Lemma.
Let $f\colon \mathbb{R}^n \to \mathbb{R}$. If $g(t) = f(x + tv)$ is convex $\forall\, x, v \in \mathbb{R}^n$, then f is cvx.

In other words, to certify that f is convex, it is enough to check the convexity of f restricted to lines.

Proofs are straightforward from the definition of convexity.

Convexity via curvature

For multivariate convex functions, the curvature condition is given by the eigenvalues of the Hessian.

Theorem.

A twice continuously differentiable multivariate function $f\colon \mathbb{R}^n \to \mathbb{R}$ *is convex if and only if* $\nabla^2 f(x) \succeq 0$ *for all* $x \in \mathbb{R}^n$*.*

Proof. (By Schwartz's theorem, $\nabla^2 f(x) \in \mathbb{R}^{n\times n}$ is symmetric.)

Assume f is convex. For any $x, v \in \mathbb{R}^n$, let $g(t) = f(x + tv)$. By the chain rule, and since $g\colon \mathbb{R} \to \mathbb{R}$ is convex and twice-differentiable,

$$g''(0) = v^\intercal \nabla^2 f(x) v \geq 0.$$

Since this holds for all $v \in \mathbb{R}^n$, we conclude $\nabla^2 f(x) \succeq 0$.

Conversely, assume $\nabla^2 f(x) \succeq 0$ for all $x \in \mathbb{R}^n$. For any $x, v \in \mathbb{R}^n$, let $g(t) = f(x + tv)$. By the chain rule and $\nabla^2 f(\cdot) \succeq 0$,

$$g''(t) = v^\intercal \nabla^2 f(x + tv) v \geq 0.$$

Then, $g'' \geq 0$ implies g is convex, which, in turn, implies f is convex. $\square$

Affine functions are convex

A function $f\colon \mathbb{R}^n \to \mathbb{R}$ is *affine*

$$f(x) = \langle a, x\rangle + b$$

for some $a \in \mathbb{R}^n$ and $b \in \mathbb{R}$.

Strictly speaking, a function is *linear* if it is affine with $b = 0$.

Theorem.

An affine function is convex.

Proof 1. An affine function has 0 curvature, which is nonnegative. □

Proof 2. If f is affine,

$$\begin{aligned} f(\theta x + (1-\theta)y) &= \langle a, \theta x + (1-\theta)y\rangle + b \\ &= \theta\langle a, x\rangle + \theta b + (1-\theta)\langle a, y\rangle + (1-\theta)b \\ &= \theta f(x) + (1-\theta)f(y). \end{aligned}$$

□

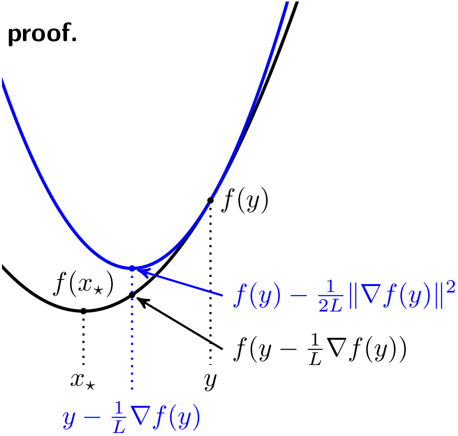
Cocoercivity inequality for smooth convex functions

Lemma.

Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be L-smooth. For any $x_\star \in \operatorname{argmin} f$,

$$f(y) - \frac{1}{2L}\|\nabla f(y)\|^2 \geq f(x_\star), \qquad \forall\, y \in \mathbb{R}^n.$$

Illustration of proof.



Cocoercivity inequality for smooth convex functions

Lemma.

Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be L-smooth. For any $x_\star \in \operatorname{argmin} f$,

$$f(y) - \frac{1}{2L}\|\nabla f(y)\|^2 \geq f(x_\star), \qquad \forall\, y \in \mathbb{R}^n.$$

Proof. By $x_\star \in \operatorname{argmin} f$ and the L-smoothness lemma,

$$f(x_\star) \leq f(y+\delta) \leq f(y) + \langle \nabla f(y), \delta \rangle + \frac{L}{2}\|\delta\|^2, \qquad \forall \delta \in \mathbb{R}^n.$$

Let $\delta = -\frac{1}{L}\nabla f(y)$ (which minimizes the RHS) to get

$$f(x_\star) \leq f(y - \tfrac{1}{L}\nabla f(y)) \leq f(y) - \frac{1}{2L}\|\nabla f(y)\|^2.$$

$\square$

Cocoercivity inequality for smooth convex functions

Lemma.

Let $f : \mathbb{R}^n \to \mathbb{R}$ be L-smooth. For any $x_\star \in \operatorname{argmin} f$,

$$f(y) - \frac{1}{2L}\|\nabla f(y)\|^2 \geq f(x_\star), \qquad \forall\, y \in \mathbb{R}^n.$$

Interpretation 1: This strengthens the simple inequality $f(y) \geq f(x_\star)$.

Interpretation 2: The suboptimality of y, measured by $f(y) - f(x_\star)$, is larger than $\frac{1}{2L}\|\nabla f(y)\|^2$.

- If $\|\nabla f(y)\|^2$ is large, then the suboptimality is necessarily large.
- If $\|\nabla f(y)\|^2$ is small, are we assured that the suboptimality is small?

Interpretation 3: $\frac{1}{2L}\|\nabla f(y)\|^2$ is the guaranteed progress (descent) to be made by taking a gradient descent step $y \mapsto y - \frac{1}{L}\nabla f(y)$.

Cocoercivity inequality for smooth convex functions

Theorem.
Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be convex and L-smooth. Then,

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \frac{1}{2L} \|\nabla f(y) - \nabla f(x)\|^2, \qquad \forall\, x, y \in \mathbb{R}^n.$$

Proof. Let

$$g(y) = f(y) - \langle \nabla f(x), y \rangle.$$

Note that g is convex, g is L-smooth, $\nabla g(y) = \nabla f(y) - \nabla f(x)$, and $x \in \operatorname{argmin} g$. (The argument $x \in \operatorname{argmin} g$ uses convexity.) Finally, apply the previous lemma to g. $\square$

Cocoercivity inequality for smooth convex functions

This inequality

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \frac{1}{2L} \|\nabla f(y) - \nabla f(x)\|^2$$

is called the *cocoercivity inequality* for smooth convex functions.

Note that this is stronger than the convexity inequality

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

which holds for differentiable convex functions.

The cocoercivity inequality is fundamental when analyzing smooth convex functions.

In fact, it can be shown that the cocoercivity inequality implies the L-smoothness inequality. Therefore,

$$[L\text{-smooth \& convexity inequalities}] \quad \Leftrightarrow \quad [\text{cocoercivity inequality}]$$

Strong convexity

A function $f\colon \mathbb{R}^n \to \mathbb{R}$ is μ-*strongly convex* is $f(x) - \frac{\mu}{2}\|x\|^2$ is convex. (Strongly convex functions need not be differentiable.)

Lemma.

Strong convexity implies convexity.

Proof. If f is strongly convex, then

$$\left(f(x) - \frac{\mu}{2}\|x\|^2\right) + \frac{\mu}{2}\|x\|^2 = f(x)$$

is convex, since a sum of convex functions is convex. $\square$

Lemma.

Let $x_0 \in \mathbb{R}^n$. Then, $f\colon \mathbb{R}^n \to \mathbb{R}$ is strongly convex if and only if $f(x) - \frac{\mu}{2}\|x - x_0\|^2$ is convex.

Proof. Note

$$f(x) - \frac{\mu}{2}\|x - x_0\|^2 = f(x) - \frac{\mu}{2}\|x\|^2 - \underbrace{\mu\langle x, x_0\rangle + \frac{\mu}{2}\|x_0\|^2}_{\text{affine}}.$$

Adding or subtracting an affine function does not affect convexity. $\square$

Strong convexity: First-order characterization

Theorem.

Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be μ-strongly convex and differentiable. Then,

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \frac{\mu}{2}\|x - y\|^2, \qquad \forall\, x, y \in \mathbb{R}^n.$$

This is called the *strong convexity inequality.*

Proof. Let $g(y) = f(y) - \frac{\mu}{2}\|y - x\|^2$. The convexity inequality on g is

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

which is

$$f(y) - \tfrac{\mu}{2}\|y - x\|^2 \geq f(x) + \langle \nabla f(x), y - x \rangle.$$

Reorganizing, we conclude the result. $\square$

(The converse is also true: If $f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \frac{\mu}{2}\|x - y\|^2$ holds for all $x, y \in \mathbb{R}^n$, then f is μ-strongly convex.)

Strong convexity: Second-order characterization

Theorem.

Let $f : \mathbb{R}^n \to \mathbb{R}$ twice continuously differentiable. Then, f is μ-strongly convex if and only if $\nabla^2 f(x) \succeq \mu I$ for all $x \in \mathbb{R}^n$.

Proof. f is μ-strongly convex if and only if $g(x) = f(x) - \frac{\mu}{2}\|x\|^2$ is convex, which in turn holds if and only if

$$\nabla^2 g(x) \succeq 0.$$

Conclude with $\nabla^2 g(x) = \nabla^2 f(x) - \mu I$. □

Polyak–Łojasiewicz inequality

Lemma.

Let $f\colon \mathbb{R}^n \to \mathbb{R}$ *be* μ*-strongly convex, differentiable, and* $x_\star \in \operatorname{argmin} f$*. Then,*

$$f(y) - \frac{1}{2\mu}\|\nabla f(y)\|^2 \le f(x_\star), \qquad \forall\, x \in \mathbb{R}^n.$$

Proof. By μ-strong convexity,

$$\begin{aligned} f(x) &\ge f(y) + \langle \nabla f(y), x - y\rangle + \frac{\mu}{2}\|x - y\|^2 \\ &\ge \inf_{x\in\mathbb{R}^n} \left\{ f(y) + \langle \nabla f(y), x - y\rangle + \frac{\mu}{2}\|x - y\|^2 \right\} = f(y) - \frac{1}{2\mu}\|\nabla f(y)\|^2. \end{aligned}$$

(Infimum is attained at $x = y - \frac{1}{\mu}\nabla f(y)$.) Plugging $x = x_\star$ into the LHS, we arrive at the conclusion. $\square$

This is called the Polyak–Łojasiewicz (PL) inequality. Strong convexity implies PL. However, the converse is not true.

Polyak–Łojasiewicz inequality

Lemma.

Let $f : \mathbb{R}^n \to \mathbb{R}$ be μ-strongly convex, differentiable, and $x_\star \in \operatorname{argmin} f$. Then,

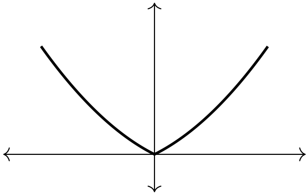
$$f(y) - \frac{1}{2\mu}\|\nabla f(y)\|^2 \leq f(x_\star), \qquad \forall\, x \in \mathbb{R}^n.$$

Interpretation: The suboptimality of y, measured by $f(y) - f(x_\star)$, is smaller than $\frac{1}{2\mu}\|\nabla f(y)\|^2$.

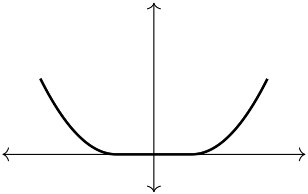
- If $\|\nabla f(y)\|^2$ is large, are we assured that the suboptimality is large?
- If $\|\nabla f(y)\|^2$ is small, then the suboptimality is necessarily small.

Strong convexity and smoothness

Informally speaking, μ-strongly convex functions have upward curvature of at least μ, and L-smooth convex functions have upward curvature of no more than L. We can think of nondifferentiable points as points with infinite curvature.



Strongly convex but not smooth



Smooth but not strongly convex.

(In fact, strong convexity and smoothness are dual properties:
[f is μ-strongly convex] $\Leftrightarrow$ [f^* is $(1/\mu)$-smooth].)

Strictly convex functions

We say a function $f\colon \mathbb{R}^n \to \mathbb{R}$ is *strictly convex* if

$$f(\theta x+(1-\theta)y) < \theta f(x)+(1-\theta)f(y), \qquad \forall\, x, y \in \mathbb{R}^n,\, x \neq y,\, \theta \in (0,1).$$

I.e., f is convex if the chord connecting $(x, f(x))$ and $(y, f(y))$ lies *strictly* above the graph of f except at the endpoints.

XXX Figure example XXX

Strictly convex functions

Lemma.

Strong convexity implies strict convexity.

Proof. Homework. □

Lemma.

Let $f : \mathbb{R}^n \to \mathbb{R}$ be strictly convex and $C \subseteq \mathbb{R}^n$ nonempty convex. Then,

$$\begin{array}{ll} \underset{x \in \mathbb{R}^n}{\textit{minimize}} & f(x) \\ \textit{subject to} & x \in C \end{array}$$

has at most one solution.

Proof. Assume for contradiction that $x_\star$ and $y_\star$ are distinct solutions. Then,

$$f(\underbrace{\theta x_\star + (1-\theta) y_\star}_{\in C \text{ by convexity}}) < \theta f(x_\star) + (1-\theta) f(y_\star) = \inf f, \qquad \forall\, \theta \in (0,1),$$

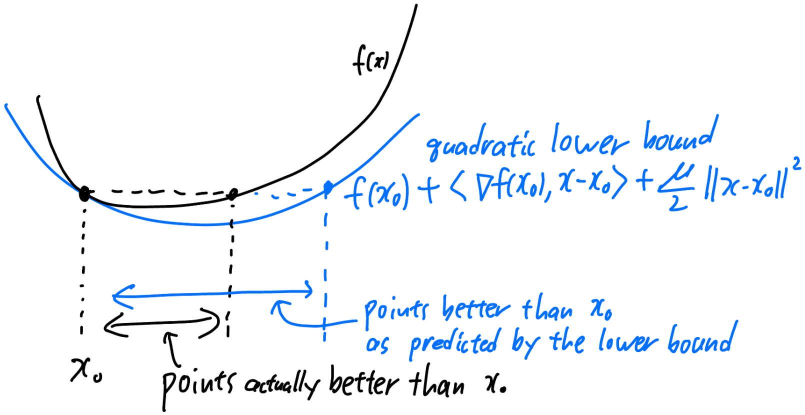
which is a contradiction. □

Minimizers of strongly convex functions

Lemma.

Let $f : \mathbb{R}^n \to \mathbb{R}$ be strongly convex. Then f has exactly one minimizer. (Minimizer exists and is unique).

Illustration of proof.



Minimizers of strongly convex functions

Lemma.

Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be strongly convex. Then f has exactly one minimizer. (Minimizer exists and is unique).

Proof. Uniqueness follows from strict convexity. Remains to show existence. Assume f is diff. and let $x_0 \in \mathbb{R}^n$. Then,

$$\begin{aligned} f(x) &\geq f(x_0) + \langle \nabla f(x_0), x - x_0 \rangle + \frac{\mu}{2}\|x - x_0\|^2 \\ &= f(x_0) + \frac{\mu}{2}\Big\|x - \big(x_0 - \tfrac{\nabla f(x_0)}{\mu}\big)\Big\|^2 - \tfrac{\mu}{2}\Big\|\tfrac{\nabla f(x_0)}{\mu}\Big\|^2 \end{aligned}$$

for all $x \in \mathbb{R}^n$. Therefore,

$$\underbrace{\{x \mid f(x) \leq f(x_0)\}}_{\text{closed (pre-image of cont. fn.)}} \subset \underbrace{B\Big(x_0 - \tfrac{\nabla f(x_0)}{\mu}, \Big\|\tfrac{\nabla f(x_0)}{\mu}\Big\|\Big)}_{\text{ball, bounded}}$$

and

$$\operatorname*{argmin}_{x \in \mathbb{R}^n} f(x) = \operatorname*{argmin}_{x : f(x) \leq f(x_0)} f(x).$$

Since the RHS is a minimization of a continuous function over a compact set, the minimum is attained, i.e., a minimizer exists.

When f is non-differentiable, the same argument works with a subgradient. Continuity of the convex function f will be shown later. $\square$

Smooth strongly convex functions

Lemma.

Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be μ-strongly convex and L-smooth. Then $\mu \leq L$.

Proof. Let $x \neq y$. By μ-strong convexity,

$$f(y) \geq f(x) + \langle \nabla f(x), y - x\rangle + \tfrac{\mu}{2}\|x - y\|^2$$
$$f(x) \geq f(y) + \langle \nabla f(y), x - y\rangle + \tfrac{\mu}{2}\|x - y\|^2$$

and adding jthese two we have

$$\langle \nabla f(x) - \nabla f(y), x - y\rangle \geq \mu\|x - y\|^2.$$

By Cauchy–Schwartz,

$$\mu\|x - y\|^2 \leq \langle \nabla f(x) - \nabla f(y), x - y\rangle \leq \|\nabla f(x) - \nabla f(y)\|\|x - y\|$$

and

$$\mu\|x - y\| \leq \|\nabla f(x) - \nabla f(y)\|.$$

By L-smoothness,

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|,$$

so $\mu \leq L$. □

Projection onto convex sets

Projection[1] of $p \in \mathbb{R}^n$ onto C is the point within C that is closest to p. Is this notion well-defined?

Theorem.
Let $C \subseteq \mathbb{R}^n$ be a nonempty closed convex set and let $p \in \mathbb{R}^n$. Then

$$\Pi_C(p) = \underset{x \in C}{\operatorname{argmin}} \|x - p\|,$$

where $\|\cdot\|$ is the standard Euclidean norm, uniquely exists.

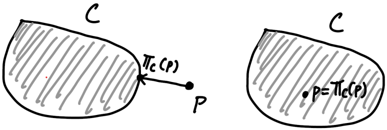


Illustration when C is nonempty closed convex. (Setting of the theorem)

[1] In linear algebra, our notion of projection corresponds to orthogonal projections but not oblique projections.

Projection onto convex sets

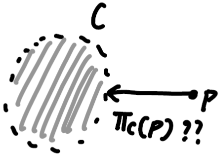


Illustration when C is open. The projection is not attained.

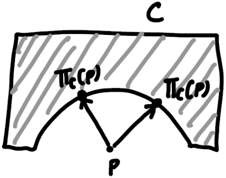


Illustration when C is not convex. Projection may not be unique.

Projection onto convex sets

Proof. Clearly,

$$\Pi_C(p) = \operatorname*{argmin}_{x \in C} \|x - p\| = \operatorname*{argmin}_{x \in C} \|x - p\|^2.$$

Since $\|x - p\|^2$ is a strictly convex function of f, a minimizer, if exists, must be unique. (So, there are 0 or 1 minimizers.)

Let $\{x_k\}_k$ be a sequence such that

$$\|x_k - p\| \to \inf_{x \in C} \|x - p\|.$$

Since $\{x_k\}_k$ is bounded, it has a convergent subsequence $x_{k_j} \to x_\infty \in C$ by the Bolzano–Weierstrass theorem and closedness of C. By continuity of $\|x - p\|^2$ as a function of x, we conclude

$$\|x_\infty - p\|^2 = \inf_{x \in C} \|x - p\|^2,$$

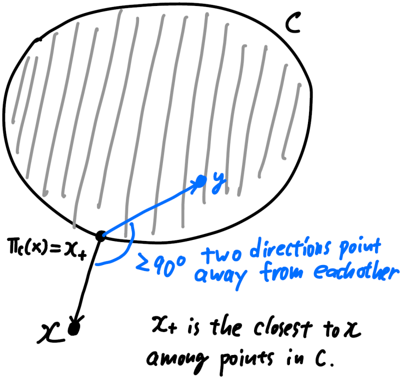
i.e., x_∞ is a minimizer. (So, there are more than 0 minimizers.) $\square$

Projection theorem

Theorem.
Let $C \subseteq \mathbb{R}^n$ be a nonempty closed convex set. Then, $x_+ = \Pi_C(x)$ if and only if

$$\langle y - x_+, x - x_+ \rangle \le 0, \qquad \forall\, y \in C.$$

(Also called the Bourbaki–Cheney–Goldstein inequality.)
Illustration of proof.



Projection theorem

Theorem.

Let $C \subseteq \mathbb{R}^n$ be a nonempty closed convex set. Then, $x_+ = \Pi_C(x)$ if and only if $x_+ \in C$ and

$$\langle y - x_+, x - x_+ \rangle \leq 0, \qquad \forall\, y \in C.$$

Proof. ($\Rightarrow$) Assume $x_+ = \operatorname{argmin}_{z \in C} \|z - x\|^2$ and let $y \in C$. Then,

$$\|y - x\|^2 \geq \|x_+ - x\|^2, \qquad \forall\, y \in C.$$

Since $x_+ + \theta(y - x_+) \in C$ for $\theta \in (0, 1]$ by convexity of C,

$$\|\underbrace{x_+ + \theta(y - x_+)}_{\text{from } x_+ \text{ move towards } y} - x\|^2 \geq \|x_+ - x\|^2.$$

Reorganizing the terms, we get

$$\theta^2 \|y - x_+\|^2 + 2\theta \langle y - x_+, x_+ - x \rangle \geq 0.$$

Dividing by θ and letting $\theta \to 0$, we conclude

$$\langle y - x_+, x_+ - x \rangle \geq 0 \quad \Rightarrow \quad \langle y - x_+, x - x_+ \rangle \leq 0.$$

Projection theorem

Theorem.

Let $C \subseteq \mathbb{R}^n$ be a nonempty closed convex set. Then, $x_+ = \Pi_C(x)$ if and only if $x_+ \in C$ and

$$\langle y - x_+, x - x_+ \rangle \le 0, \qquad \forall\, y \in C.$$

Proof. $(\neg \Rightarrow \neg)$ Assume $x_+ \neq \Pi_C(x)$, which means either (i) x_+ is not even in C or (ii) $x_+ \in C$ but isn't the closest to x. In case (i), we are done. In case (ii), $x_+ \in C$ and there is a $y \in C$ such that

$$\|y - x\|^2 < \|x_+ - x\|^2.$$

By convexity of $\|\cdot - x\|^2$,

$$\|\underbrace{x_+ + \theta(y - x_+)}_{\text{from } x_+ \text{ move towards } y} - x\|^2 \le (1-\theta)\|x_+ - x\|^2 + \theta\|y - x\|^2 < \|x_+ - x\|^2$$

for $\theta \in (0,1)$. (I.e., $x_+ + \theta(y - x_+)$ is closer to x.) Reorganizing terms,

$$\theta^2\|y - x_+\|^2 + 2\theta\langle y - x_+, x_+ - x\rangle < 0.$$

Dividing by θ and letting $\theta \to 0$, we conclude

$$\langle y - x_+, x_+ - x\rangle < 0 \quad \Rightarrow \quad \langle y - x_+, x - x_+\rangle > 0. \qquad \square$$

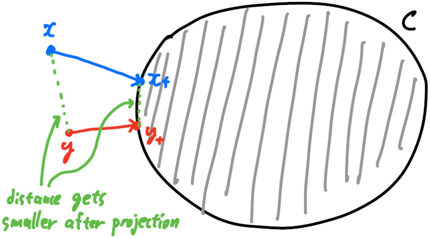
Projection is nonexpansive

Theorem.

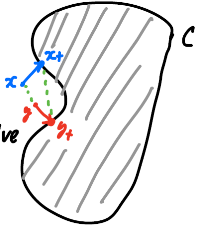
Let $C \subseteq \mathbb{R}^n$ be a nonempty closed convex set. Then $\Pi_C : \mathbb{R}^n \to \mathbb{R}^n$ is a nonexpansive operator.

In other words, if $x_+ = \Pi_C(x)$ and $y_+ = \Pi_C(y)$, then

$$\|x_+ - y_+\| \le \|x - y\|.$$



If C is non-convex
Π_C may not be nonexpansive



Projection is nonexpansive

Theorem.
Let $C \subseteq \mathbb{R}^n$ be a nonempty closed convex set. Then $\Pi_C \colon \mathbb{R}^n \to \mathbb{R}^n$ is a nonexpansive operator.

Proof. Let $x, y \in \mathbb{R}^n$, $x_+ = \Pi_C(x)$, and $y_+ = \Pi_C(y)$. By the projection theorem,

$$\begin{aligned} \langle y_+ - x_+, x - x_+ \rangle &\le 0 \\ \langle x_+ - y_+, y - y_+ \rangle &\le 0. \end{aligned}$$

Summing these two inequalities, we get

$$\langle x_+ - y_+, x_+ - y_+ \rangle \le \langle x_+ - y_+, x - y \rangle.$$

Using Cauchy–Schwartz, we get

$$\|x_+ - y_+\|^2 \le \langle x_+ - y_+, x - y \rangle \le \|x_+ - y_+\| \|x - y\|.$$

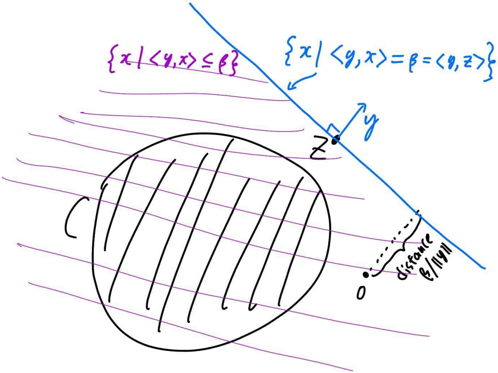
Dividing by $\|x_+ - y_+\|$ (when nonzero), we conclude the statement. $\square$

Separating hyperplane theorem

Theorem.

Let $C \subset \mathbb{R}^n$ be a nonempty closed convex set, and let $z \in \mathbb{R}^n$. If $z \notin C$, then there is a $(y, \beta) \in \mathbb{R}^n \times \mathbb{R}$ such that $y \neq 0$ and

$$\begin{aligned} \langle y, x\rangle &\leq \beta, \qquad \forall\, x \in C \\ \langle y, z\rangle &= \beta. \end{aligned}$$



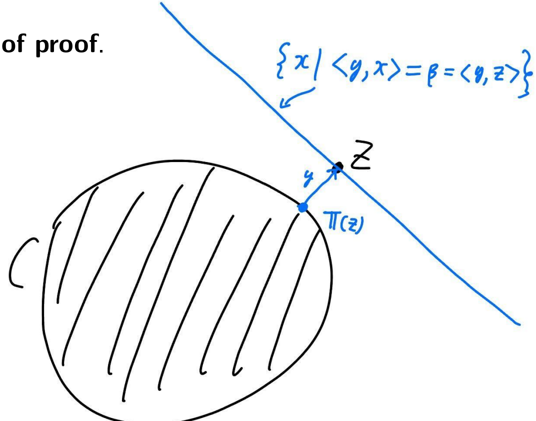
Separating hyperplane theorem

Theorem.

Let $C \subset \mathbb{R}^n$ be a nonempty closed convex set, and let $z \in \mathbb{R}^n$. If $z \notin C$, then there is a $(y, \beta) \in \mathbb{R}^n \times \mathbb{R}$ such that $y \neq 0$ and

$$\begin{aligned} \langle y, x \rangle &\leq \beta, \qquad \forall\, x \in C \\ \langle y, z \rangle &= \beta. \end{aligned}$$

Illustration of proof.



Separating hyperplane theorem

Theorem.
Let $C \subset \mathbb{R}^n$ be a nonempty closed convex set, and let $z \in \mathbb{R}^n$. If $z \notin C$, then there is a $(y, \beta) \in \mathbb{R}^n \times \mathbb{R}$ such that $y \neq 0$ and

$$\begin{aligned} \langle y, x \rangle &\leq \beta, \qquad \forall\, x \in C \\ \langle y, z \rangle &= \beta. \end{aligned}$$

Proof. Let $y = z - \Pi_C(z)$. Note, $y \neq 0$, since $z \notin C$. By the projection theorem,

$$\langle x - \Pi_C(z), y \rangle \leq 0, \qquad \forall\, x \in C.$$

If we let $\beta = \langle \Pi_C(z), y \rangle$, then

$$\langle y, x \rangle \beta, \qquad \forall\, x \in C.$$

$\square$

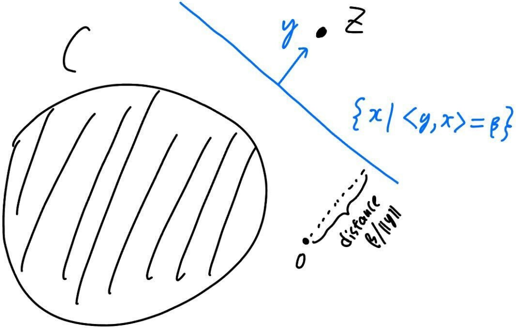
Strict separating hyperplane theorem

The result can be strengthened such that the separation is strict.

Theorem.

Let $C \subset \mathbb{R}^n$ be a nonempty closed convex set, and let $z \in \mathbb{R}^n$. If $z \notin C$, then there is a $(y, \beta) \in \mathbb{R}^n \times \mathbb{R}$ such that $y \neq 0$ and

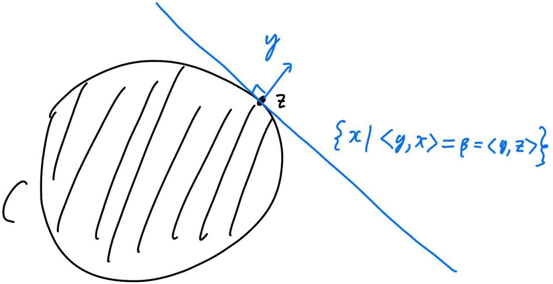
$$\begin{aligned} \langle y, x\rangle &< \beta, \qquad \forall\, x \in C \\ \langle y, z\rangle &> \beta. \end{aligned}$$



Supporting hyperplane theorem

Theorem.
Let $C \subset \mathbb{R}^n$ be a nonempty closed convex set and let $z \in \partial C$. (∂C is the boundary of C.) Then, there is a $(y, \beta) \in \mathbb{R}^n \times \mathbb{R}$ such that $y \neq 0$ and

$$\begin{aligned} \langle y, x\rangle &\leq \beta, \qquad \forall\, x \in C \\ \langle y, z\rangle &= \beta. \end{aligned}$$



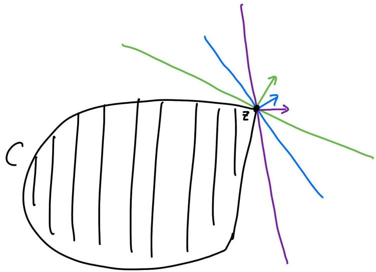
Supporting hyperplane theorem

Sometimes, it is convenient to eliminate β.

Theorem.
Let $C \subset \mathbb{R}^n$ be a nonempty closed convex set and let $z \in \partial C$. (∂C is the boundary of C.) Then, there is a non-zero $y \in \mathbb{R}^n$ such that

$$\langle y, x \rangle \le \underbrace{\langle y, z \rangle}_{=\beta}, \qquad \forall\, x \in C.$$

Also, the supporting hyperplane may not be unique if z is at a "corner point" of C.



Supporting hyperplane theorem

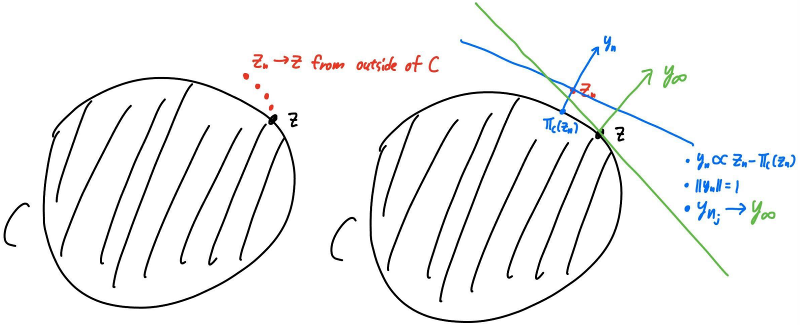
Write ∂C to denote the boundary of C, i.e., ∂C is the set of points in the closure of C not belonging to the interior of C.

Theorem.

Let $C \subset \mathbb{R}^n$ be a nonempty closed convex set and let $z \in \partial C$. Then, there is a non-zero $y \in \mathbb{R}^n$ such that

$$\langle y, x \rangle \leq \langle y, z \rangle, \qquad \forall\, x \in C.$$

Illustration of proof.



Supporting hyperplane theorem

Theorem.

Let $C \subset \mathbb{R}^n$ be a nonempty closed convex set and let $z \in \partial C$. (∂C is the boundary of C.) Then, there is a non-zero $y \in \mathbb{R}^n$ such that

$$\langle y, x\rangle \leq \langle y, z\rangle, \qquad \forall\, x \in C.$$

Proof. For any $\varepsilon > 0$, it must be that $B(z, \varepsilon) \not\subset C$, since $z \in \partial C$. Choose $z_n \in B(z, 1/2^n)\backslash C$ for $n \in \mathbb{N}$, so $z_n \to z$. Let

$$y_n = \frac{z_n - \Pi_C(z_n)}{\|z_n - \Pi_C(z_n)\|},$$

where we note $\Pi_C(z_n) \neq z_n$ since $z_n \neq C$. Since $\{y_n\}_{n\in\mathbb{N}}$ is a sequence on the unit ball in $\mathbb{R}^n$ (which is compact), it has a convergent subsequence with a limit, which we denote $y_{n_j} \to y_\infty$.

By the projection theorem,

$$\langle x - \Pi_C(z_n), z_n - \Pi_C(z_n)\rangle \leq 0, \qquad \forall\, x \in C,$$

so

$$\left\langle x - \Pi_C(z_n), \underbrace{\frac{z_n - \Pi_C(z_n)}{\|z_n - \Pi_C(z_n)\|}}_{=y_n} \right\rangle \leq 0, \qquad \forall\, x \in C.$$

Therefore,

$$\langle y_n, x\rangle \leq \langle y_n, \Pi_C(z_n)\rangle, \qquad \forall\, x \in C.$$

Taking the limit on $y_{n_j} \to y_\infty$,

$$\langle y_\infty, x\rangle \leq \langle y_\infty, \lim_{n\to\infty} \Pi_C(z_n)\rangle \overset{\text{(i)}}{=} \langle y_\infty, \Pi_C(z)\rangle \overset{\text{(ii)}}{=} \langle y_\infty, z\rangle, \qquad \forall\, x \in C,$$

where (i) follows from fact that Π_C is nonexpansive (that is, 1-Lipschitz continuous) and hence continuous, and (ii) follows from the fact that $z \in C$, so $\Pi_C(z) = z$. □

Supporting hyperplane theorem II

Theorem.
Let $C \subset \mathbb{R}^n$ be a nonempty convex set (not necessarily closed) and let $z \in \partial C$. Then, there is a non-zero $y \in \mathbb{R}^n$ such that

$$\langle y, x \rangle \leq \langle y, z \rangle, \qquad \forall\, x \in C.$$

Proof. The supporting hyperplane theorem with the nonempty closed convex set $\overline{C}$ guarantees a non-zero $y \in \mathbb{R}^n$ such that

$$\langle y, x \rangle \leq \langle y, z \rangle, \qquad \forall\, x \in \overline{C}.$$

This requirement is also satisfied with C in place of $\overline{C}$. □

Subgradient

Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be convex (but not necessarily differentiable). We say $g \in \mathbb{R}^n$ is a *subgradient* of convex f at x if

$$f(y) \geq f(x) + \langle g, y - x\rangle \qquad \forall\, y \in \mathbb{R}^n.$$

The *subdifferential* of convex f at x is

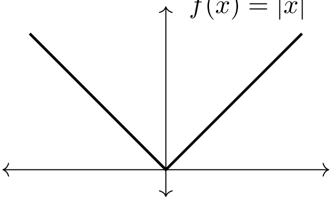
$$\partial f(x) = \{g \in \mathbb{R}^n \,|\, f(y) \geq f(x) + \langle g, y - x\rangle, \forall\, y \in \mathbb{R}^n\},$$

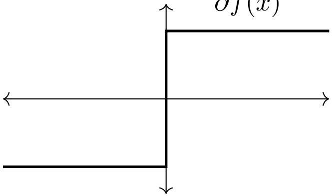
i.e., $\partial f(x) = \{\text{subgradients of } f \text{ at } x\}$.

We have already established that $\nabla f(x) \in \partial f(x)$ if f is differentiable at x (convexity inequality), but convex functions can be non-differentiable. Nevertheless, a subgradient always exists.

Subdifferential example

The absolute value function is differentiable everywhere except at 0.

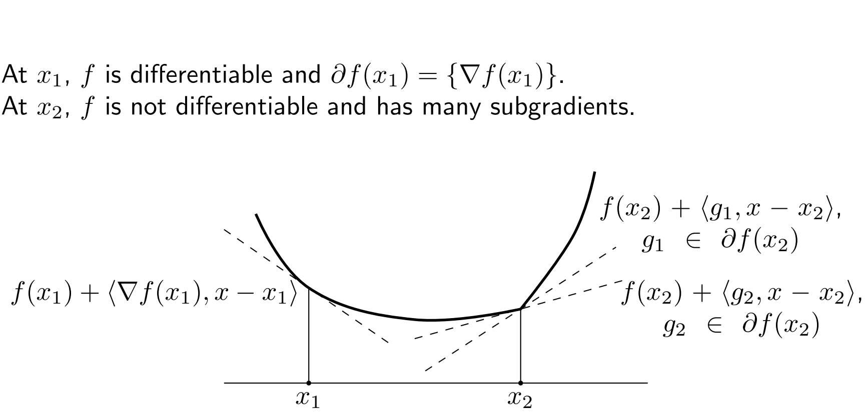




$$\partial f(x) = \begin{cases} \{-1\} & \text{for } x < 0 \\ [-1, 1] & \text{for } x = 0 \\ \{+1\} & \text{for } x > 0 \end{cases}$$

Subdifferential example

At x_1, f is differentiable and $\partial f(x_1) = \{\nabla f(x_1)\}$.
At x_2, f is not differentiable and has many subgradients.

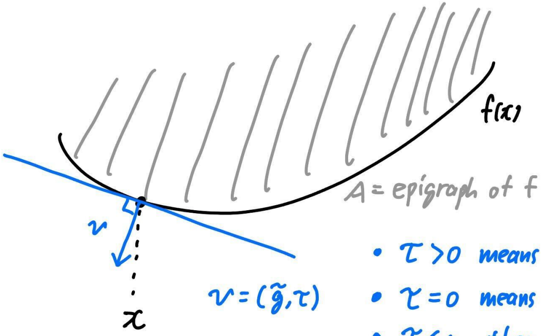


Existence of a subgradient

Theorem.

Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex. For any $x \in \mathbb{R}^n$, then exists a subgradient of f at x, i.e., $\partial f(x) \neq \emptyset$.

Proof.



- $\tau > 0$ means v points up ✗
- $\tau = 0$ means hyperplane is vertical ✗
- $\tau < 0$, then $g = \tilde{g}/\tau \in \partial f(x)$. ✓

Existence of a subgradient

Theorem.

Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be convex. For any $x \in \mathbb{R}^n$, then exists a subgradient of f at x, i.e., $\partial f(x) \neq \emptyset$.

Proof.

$$A = \{(x,t) \,|\, f(x) \leq t,\, x \in \mathbb{R}^n,\, t \in \mathbb{R}\} \subset \mathbb{R}^{n+1}.$$

By construction, $(x, f(x)) \in \partial A$. By the supporting hyperplane theorem, there is a $v = (\tilde{g}, \tau) \in \mathbb{R}^{n+1}$ such that $v \neq 0$ and

$$\langle v, u\rangle = \tilde{g}^\intercal y + \tau s \leq \tilde{g}^\intercal x + \tau f(x) = \langle v, (x, f(x))\rangle, \quad \forall\, u = (y,s) \in A.$$

We argue that $\tau < 0$. Indeed, if $\tau > 0$, we can take $s \to \infty$ and draw a contradiction. If $\tau = 0$, then $\tilde{g}^\intercal y \leq \tilde{g}^\intercal x$. Since $\tilde{g} \neq 0$ (since $v \neq 0$), we draw a contradiction with $y = \alpha \tilde{g}$ and $\alpha \to \infty$.
If $\tau < 0$, let $g = \tilde{g}/\tau$ to get

$$-g^\intercal y + s \geq -g^\intercal x + f(x), \qquad \forall\, (y,s) \in A.$$

Plugging in $s = f(y)$, we conclude

$$f(y) \geq f(x) + g^\intercal (y - x), \qquad \forall\, y \in \mathbb{R}^n.$$

□

Uniqueness of subgradient $\Leftrightarrow$ differentiability

Theorem.
Let $f : \mathbb{R}^n \to \mathbb{R}$ *be convex. Then* $\partial f(x) = \{g\}$ *if and only if* f *is differentiable at* x *and* $g = \nabla f(x)$.

We shall come back to this proof later.

Minimizers as zeros of subdifferentials

Lemma.

Let $f : \mathbb{R}^n \to \mathbb{R}$ *be convex. Then,*

$$x_\star \in \operatorname{argmin} f \quad \Leftrightarrow \quad 0 \in \partial f(x_\star).$$

Proof. $x_\star$ minimizes f if and only if

$$f(y) \geq f(x_\star) + \underbrace{\langle 0, y - x_\star \rangle}_{=0}, \qquad \forall\, y \in \mathbb{R}^n.$$

By definition of subgradients, this holds if and only if $0 \in \partial f(x_\star)$. $\square$

Subdifferential sum rule

Lemma.

Let $f : \mathbb{R}^n \to \mathbb{R}$ *be convex. Let*

$$\tilde{f}(x) = f(x) + \langle u, x \rangle + b$$

for $u \in \mathbb{R}^n$ *and* $b \in \mathbb{R}$. *Then,*

$$\partial \tilde{f}(x) = \partial f(x) + u = \{g + u \,|\, g \in \partial f(x)\}.$$

Proof. Exercise. $\square$

Tilting subdifferentials

Lemma.

Let $f\colon \mathbb{R}^n \to \mathbb{R}$ *be convex. Let*

$$\tilde{f}(y) = f(y) - \langle g, x_0 \rangle + b$$

for some $g \in \partial f(x_0)$ *and* $b \in \mathbb{R}$*. Then,* $x_0 \in \operatorname{argmin} \tilde{f}$*.*

Proof.

$$\partial \tilde{f}(x_0) \ni \underbrace{g}_{\in \partial f(x_0)} - g = 0.$$

□

Continuity of univariate convex functions

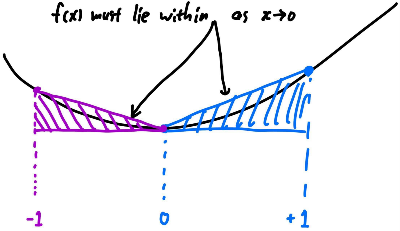
Theorem.
Let $f: \mathbb{R} \to \mathbb{R}$ be convex. Then, f is continuous.

Proof. W.L.O.G. consider continuity at $x = 0$, since continuity of f and $\tilde{f}(x) = f(x - y)$ are equivalent. W.L.O.G. assume f is minimized at $x = 0$ with $f(0) = 0$, since otherwise we can consider

$$\tilde{f}(x) = f(x) - f(0) - gx,$$

where $g \in \partial f(0)$.

Illustration of the remaining argument:



Continuity of univariate convex functions

Theorem.

Let $f : \mathbb{R} \to \mathbb{R}$ be convex. Then, f is continuous.

Proof. W.L.O.G. consider continuity at $x = 0$, since continuity of f and $\tilde{f}(x) = f(x - y)$ are equivalent. W.L.O.G. assume f is minimized at $x = 0$ with $f(0) = 0$, since otherwise we can consider

$$\tilde{f}(x) = f(x) - f(0) - gx, \qquad f \in \partial f(x).$$

For any $\varepsilon \in [0, 1]$,

$$0 \overset{\text{(i)}}{\leq} f(\varepsilon) \overset{\text{(ii)}}{\leq} \varepsilon f(1)$$

where (i) follows from 0 being a global minimizer and (ii) follows from the convexity inequality between input points 0 and 1. Therefore,

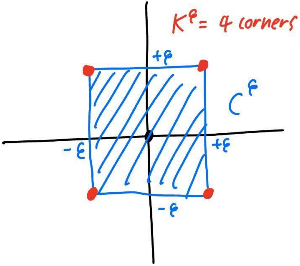
$$0 \leq \liminf_{x \to 0} f(x) \leq \limsup_{x \to 0} f(x) \leq \varepsilon \max\{f(+1), f(-1)\}$$

for any $\varepsilon > 0$. By the squeeze theorem, we conclude $\lim_{x \to 0} f(x) = 0$, i.e., continuity. $\square$

Lemma: Convex fn. are maximized at extreme points

For any $\varepsilon > 0$, let

$$K^\varepsilon = \{(\pm\varepsilon, \ldots, \pm\varepsilon) \in \mathbb{R}^n\}, \qquad \text{(So } |K| = 2^n\text{)}$$
$$C^\varepsilon = \{x \in \mathbb{R}^n \mid \|x\|_\infty \leq \varepsilon\}.$$



Lemma.
Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex. Then,

$$\sup_{x \in C^\varepsilon} f(x) = \max_{x \in K^\varepsilon} f(x).$$

I.e., maximizers are attained on K^ε. (If a point in $C^\varepsilon \backslash K^\varepsilon$ is a maximizer, there is another maximizer in K^ε attaining the same function value.)

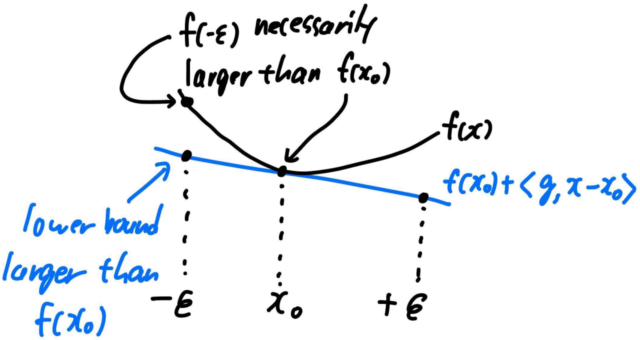
Lemma: Convex fn. are maximized at extreme points

Proof. First, we prove the claim for univariate functions. Let $f : \mathbb{R} \to \mathbb{R}$ and assume there is an $x_\circ \in (-\varepsilon, +\varepsilon)$ such that

$$f(x_\circ) > \max_{x=\pm\varepsilon} f(x)$$

Then, there is a subgradient $g \in \partial f(x_\circ)$ such that

$$f(x) \geq f(x_\circ) + g(x - x_\circ).$$



Lemma: Convex fn. are maximized at extreme points

Proof. First, we prove the claim for univariate functions. Let $f\colon \mathbb{R} \to \mathbb{R}$ and assume there is an $x_\circ \in (-\varepsilon, +\varepsilon)$ such that

$$f(x_\circ) > \max_{x=\pm\varepsilon} f(x)$$

Then, there is a subgradient $g \in \partial f(x_\circ)$ such that

$$f(x) \geq f(x_\circ) + g(x - x_\circ).$$

If $g \geq 0$, then $x = +\varepsilon$ maximizes the RHS and if $g < 0$ then $x = -\varepsilon$ maximizes the RHS. Therefore, LHS and RHS over $x = \pm\varepsilon$, we get

$$\max_{x=\pm\varepsilon} f(x) \geq \max_{x=\pm\varepsilon} \left\{ f(x) \geq f(x_\circ) + g(x - x_\circ) \right\}$$

$$\geq f(x_\circ) \overset{?}{>} \max_{x=\pm\varepsilon} f(x).$$

So we have a contradiction, and we are forced to conclude that such an $x_\circ$ does not exist, i.e., the maximum of f over $[-\varepsilon, +\varepsilon]$ is always attained on the endpoints $\pm\varepsilon$.

(Although the maximizer of f may not be unique and may contain points in $(-\varepsilon, +\varepsilon)$, an endpoint, $-\varepsilon$ or $+\varepsilon$, will still be a maximizer.)

Lemma: Convex fn. are maximized at extreme points

Next, consider the general case in $\mathbb{R}^n$. Assume for contradiction that there is a $x_\circ \in C^\varepsilon$ such that

$$f(x_\circ) > \max_{x \in K^\varepsilon} f(x).$$

So $x_\circ \notin K^\varepsilon$, and there must be a coordinate $i \in \{1, \ldots, n\}$ such that $(x_\circ)_i \in (-\varepsilon, +\varepsilon)$. Then, the univariate function

$$f\big((x_\circ)_1, \ldots, (x_\circ)_{i-1}, \delta, (x_\circ)_{i+1}, \ldots, (x_\circ)_n\big)$$

attains a maximizer at $\delta = \pm\varepsilon$. By modifying the i-th coordinate of $x_\circ$ to $\pm\varepsilon$, we can only improve (increase) the function value.

Repeating this process at most n times, we get a point in K^ε with function value not smaller than the original $f(x_\circ)$. This contradicts the assumption $f(x_\circ) > \max_{x \in K^\varepsilon} f(x)$, and we are forced to conclude

$$\sup_{x \in C^\varepsilon} f(x) = \max_{x \in K^\varepsilon} f(x).$$

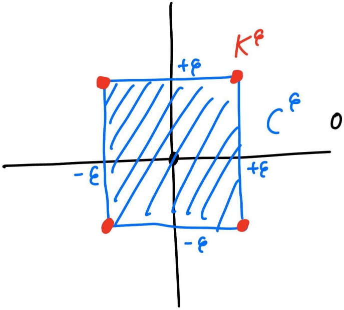
□

Continuity of multivariate convex functions

Theorem.

Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex. Then, f is continuous.

Proof. W.L.O.G. consider continuity at $x = 0$, and assume $0 \in \operatorname{argmin} f$ and $0 = \min f$.



Squeez Theorem

$$0 = f(0) \le f(x) \le \underbrace{\max_{x \in K^\varepsilon} f(x)}_{\to 0 \text{ as } \varepsilon \to 0}$$

for $x \in C^\varepsilon$

Continuity of multivariate convex functions

Theorem.

Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be convex. Then, f is continuous.

Proof. W.L.O.G. consider continuity at $x = 0$, and assume $0 \in \operatorname{argmin} f$ and $0 = \min f$. Consider K^ε and C^ε as previously defined.

Let $\{x^{(1)}, \ldots x^{(2^n)}\} = K^\varepsilon$. Then,

$$0 = f(0) \leq f(x) \leq \max_{j=1,\ldots,2^n} f(x^{(j)}), \qquad \forall\, x \in C^\varepsilon.$$

Since univariate convex functions are continuous,

$$\lim_{\varepsilon \to 0} \max_{j=1,\ldots,2^n} f(x^{(j)}) = \max_{j=1,\ldots,2^n} \lim_{\varepsilon \to 0} f(x^{(j)}) = f(0) = 0.$$

Therefore,

$$0 \leq \inf_{x \in C^\varepsilon} f(x) \leq \sup_{x \in C^\varepsilon} f(x) = \max_{x \in K^\varepsilon} f(x) \to 0$$

as $\varepsilon \to 0$, and we conclude continuity. $\square$

Jensen's inequality

Theorem.
Let $X \in \mathbb{R}^n$ *be a random variable such that* $\mathbb{E}[X] \in \mathbb{R}^n$ *is well defined, and let* $\varphi\colon \mathbb{R}^n \to \mathbb{R}$ *be convex. Then,*

$$\varphi\big(\mathbb{E}[X]\big) \leq \mathbb{E}[\varphi(X)].$$

Proof. Let $g \in \partial\varphi(\mathbb{E}[X])$. Then,

$$\varphi(X) \geq \varphi(\mathbb{E}[X]) + \langle g, X - \mathbb{E}[X] \rangle.$$

Taking expectations on both sides completes the proof. $\square$

Lipschitz continuity $\Leftrightarrow$ bounded subgradients

We say $f\colon \mathbb{R}^n \to \mathbb{R}$ is G-Lipschitz if

$$|f(y) - f(x)| \le G\|y - x\| \qquad \forall\, x, y \in \mathbb{R}^n.$$

Theorem.

Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be convex. Then f is G-Lipschitz if and only if

$$\|g\| \le G \qquad \forall\, g \in \partial f(x),\, x \in \mathbb{R}^n.$$

Proof. ($\Rightarrow$) Let $x \in \mathbb{R}^n$ and $g \in \partial f(x)$. For any $u \in \mathbb{R}^n$,

$$f(x) + \langle g, u\rangle \le f(x+u) \le f(x) + G\|u\|$$

by the subgradient inequality and G-Lipschitz continuity.Therefore,

$$\langle g, u\rangle \;\le\; G\|u\| \qquad \forall\, u \in \mathbb{R}^n.$$

Taking the supremum over all unit vectors u (i.e., $\|u\| = 1$) yields

$$\|g\| = \sup_{\|u\|=1} \langle g, u\rangle \;\le\; G.$$

Lipschitz continuity $\Leftrightarrow$ bounded subgradients

Theorem.

Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be convex. Then f is G-Lipschitz if and only if

$$\|g\| \le G \qquad \forall\, g \in \partial f(x),\, x \in \mathbb{R}^n.$$

($\Leftarrow$) Let $x, y \in \mathbb{R}^n$, $g_x \in \partial f(x)$, and $g_y \in \partial f(y)$. The subgradient and Cauchy–Schwartz inequalities yield

$$f(y) \ge f(x) + \langle g_x,\, y - x\rangle \ge f(x) - \|g_x\|\,\|y - x\| \ge f(x) - G\|y - x\|$$

and

$$f(x) \ge f(y) + \langle g_y,\, x - y\rangle \ge f(y) - \|g_y\|\,\|x - y\| \ge f(y) - G\|x - y\|.$$

Combined, we get

$$-G\|y - x\| \le f(y) - f(x) \le G\|y - x\|.$$

$\square$

Closed graph theorem for ∂f

Theorem.
Let $f\colon \mathbb{R}^n \to \mathbb{R}$ *be convex. If* $x_k \to x$, $g_k \in \partial f(x_k)$, *and* $g_k \to g$, *then* $g \in \partial f(x)$.

Proof. For all $k \in \mathbb{N}$ and $y \in \mathbb{R}^n$,

$$f(y) \geq f(x_k) + \langle g_k, y - x_k \rangle.$$

Taking $k \to \infty$ and using continuity of f,

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

Since this holds for all $y \in \mathbb{R}^n$, we conclude $g \in \partial f(x)$. □

This is equivalent to saying that

$$\{(x, g) \mid x \in \mathbb{R}^n,\, g \in \partial f(x)\} \subset \mathbb{R}^n \times \mathbb{R}^n,$$

which is referred to as the *graph* of ∂f, is a closed set.

Boundedness of subgradients on compact sets

Lemma.

Let $f: \mathbb{R}^n \to \mathbb{R}$ be convex and let $K \subset \mathbb{R}^n$ be compact. Then,

$$\bigcup_{x \in K} \partial f(x) = \{g \mid g \in \partial f(x),\, x \in K\}$$

is bounded.

Proof. Let $B = \{u \in \mathbb{R}^n : \|u\| \leq 1\}$. By continuity of f and compactness of $K \times B$, the quantity

$$M = \max_{(x,u) \in K \times B} \big(f(x+u) - f(x)\big) < \infty.$$

For any $x \in K$, $g \in \partial f(x)$, and $u \in B$, the subgradient inequality gives

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

Taking the supremum over $u \in B$ yields

$$M \geq \sup_{\|u\| \leq 1} \langle g, u \rangle = \|g\|.$$

□

Local Lipschitz continuity and differentiability a.e.

We say $f : \mathbb{R}^n \to \mathbb{R}$ is locally Lipschitz if for every compact $K \subset \mathbb{R}^n$, there is an $L_K < \infty$ such that

$$|f(y) - f(x)| \leq L_K \|y - x\| \qquad \forall\, x, y \in K.$$

Theorem.

If $f : \mathbb{R}^n \to \mathbb{R}$ is convex, then it is locally Lipschitz.

Proof. By the previous lemma, there is an L_K such that

$$M = \sup\{\|g\| \,|\, g \in \partial f(x),\, x \in K\} < \infty.$$

Then, the boundedness of subgradients in K implies $|f(y) - f(x)| \leq M\|y - x\|$ by the same argument as before. $\square$

Corollary.

If $f : \mathbb{R}^n \to \mathbb{R}$ is convex, it is differentiable almost everywhere.

Proof. Follows from local Lipschitz and Rademacher's theorem. $\square$

Uniqueness of subgradient $\Leftrightarrow$ differentiability

Theorem.

Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be convex. Then $\partial f(x) = \{g\}$ if and only if f is differentiable at x and $g = \nabla f(x)$.

Proof. ($\Rightarrow$) Assume $\partial f(x) = \{g\}$. For any $y \in \mathbb{R}^n$ choose $g_y \in \partial f(y)$. The subgradient inequalities at x and at y give the sandwich

$$\langle g,\, y - x\rangle \;\le\; f(y) - f(x) \;\le\; \langle g_y,\, y - x\rangle.$$

By local boundedness of subgradients, any sequence $y_k \to x$ has a subsequence $y_{k_j} \to x$ such that $g_{y_{k_j}} \to \tilde{g}$. The closed graph theorem then implies $\tilde{g} \in \partial f(x) = \{g\}$. So every cluster point equals g, i.e. $g_y \to g$ as $y \to x$.

From the sandwich inequality and Cauchy–Schwartz, we have

$$0 \le f(y) - f(x) - \langle g,\, y - x\rangle \le \underbrace{\|g_y - g\|}_{\to 0}\,\|y - x\| = o(\|y - x\|)$$

as $y \to x$, so f is (Fréchet) differentiable at x with $\nabla f(x) = g$.

Uniqueness of subgradient $\Leftrightarrow$ differentiability

Theorem.

Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex. Then $\partial f(x) = \{g\}$ if and only if f is differentiable at x and $g = \nabla f(x)$.

($\Leftarrow$) Assume f is differentiable at x. Then, the convexity inequality gives $\nabla f(x) \in \partial f(x)$. If $h \in \partial f(x)$ as well, then for any unit v and $\varepsilon > 0$,

$$f(x) + \varepsilon\langle \nabla f(x), v\rangle + o(\varepsilon) = f(x + \varepsilon v) \ \geq \ f(x) + \varepsilon\langle h, v\rangle$$

Divide by ε and let $\varepsilon \to 0^+$ to get $\langle h, v\rangle \leq \langle \nabla f(x), v\rangle$. Repeating the same argument with $-v$ yields $\langle h, v\rangle \geq \langle \nabla f(x), v\rangle$, so $\langle h, v\rangle = \langle \nabla f(x), v\rangle$ for all $v \in \mathbb{R}^n$. Hence $h = \nabla f(x)$ and $\partial f(x) = \{\nabla f(x)\}$. $\square$