CME 307: Optimization CME 307 / MS&E 311 / OIT 676

Lecture 7: Optimality conditions and convexity

Fall 2025 *Prof. Udell*

1 Optimality conditions

What does it mean to solve an optimization problem? We will study unconstrained and then (briefly) constrained smooth optimization, focusing on precise definitions and the conditions that characterize optimal solutions.

Definition 1.1 (Global, local, isolated local, unique minimizers)**.** Let $f : D \subseteq \mathbb{R}^n \to \mathbb{R}$. A point $x^\star \in D$ is a

- *global minimizer* if $f(x) \geq f(x^\star)$ for all $x \in D$,
- *local minimizer* if there is a neighborhood N of $x^\star$ such that $f(x) \geq f(x^\star)$ for all $x \in N$,
- *isolated local minimizer* if the neighborhood N contains no other local minimizers,
- *unique minimizer* if it is the *only* global minimizer.

Example 1.2. A few examples to illustrate the definitions:

- For $f(x) = x^4$ on $\mathbb{R}$, $x^\star = 0$ is a global minimizer (hence also local) that is isolated.
- For $f(x) = 1 - \cos(x)$ on $\mathbb{R}$, $x = 0$ is an isolated local minimizer, but not a global minimizer.
- For $f(x) = 0$, $x = 0$ is a global minimizer, but not isolated.

Here we will formalize and prove the two central results that underlie most algorithms: the *first-* and *second-order* optimality conditions. We first study optimality for a differentiable, unconstrained objective, and then consider more general constrained problems.

1.1 First-order necessary condition

Definition 1.3 (Stationary point)**.** A point $x^\star$ is *stationary* if $\nabla f(x^\star) = 0$.

Theorem 1.4 (First-order necessary condition)**.** *If $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable and $x^\star$ is a local minimizer, then $\nabla f(x^\star) = 0$.*

Proof. Let $d \in \mathbb{R}^n$ be arbitrary, and define $\phi(\alpha) = f(x^\star + \alpha d)$. Since $x^\star$ is a local minimizer, $\alpha = 0$ is a minimizer of ϕ on some interval around 0, so $\phi'(0) = 0$. But $\phi'(0) = \nabla f(x^\star)^T d$ by the chain rule. Because this holds for all d, we must have $\nabla f(x^\star) = 0$. □

Remark 1.5 (Geometric picture). At a local minimizer of a differentiable function $f : \mathbb{R}^n \to \mathbb{R}$, all directional derivatives are 0 to first order; the tangent hyperplane given by the first-order Taylor expansion touches but does not cross below the graph of f near $x^\star$.

Gotcha 1.6 (Stationary $\neq$ minimizer). Stationarity is necessary but not sufficient for optimality. Saddle points and local maxima also satisfy $\nabla f = 0$. For example, $f(x) = x^3$ has $\nabla f(0) = 0$ but $x = 0$ is not a local minimizer. A second-order test is needed.

1.2 Symmetric positive semidefinite (PSD) matrices

We will need the following definition to generalize the first-order condition.

Definition 1.7 (PSD/SPD)**.** A symmetric matrix $Q \in \mathbb{R}^{n\times n}$ is *positive semidefinite* (PSD), written $Q \succeq 0$, if $x^T Q x \geq 0$ for all $x \in \mathbb{R}^n$; it is *positive definite* (PD), written $Q \succ 0$, if $x^T Q x > 0$ for all $x \neq 0$. Equivalently: $Q = Q^T$ and $\lambda_{\min}(Q) \geq 0$ (PSD) or > 0 (PD).

PSD matrices arise throughout least squares, quadratic models, and curvature tests; e.g., if $f(x) = \frac{1}{2}x^T Q x + c^T x$, then $\nabla^2 f(x) \equiv Q$ and f is convex $\iff Q \succeq 0$.

Example 1.8 (Quadratic bowls)**.** If $Q = U\Lambda U^T$ with $\Lambda = \mathrm{conv}(\lambda_i)$, then $x^T Q x = \sum_i \lambda_i \, (u_i^T x)^2 \geq 0$ iff $\lambda_i \geq 0$ for all i. Level sets $x^T Q x = \text{const}$ are ellipsoids when $Q \succ 0$.

Remark 1.9 (Why PSD matters here). The SONC/SOSC are nothing but statements about the quadratic form $d^T H d$ at a stationary point. Testing optimality reduces to *testing PSD or PD of the Hessian.*

Gotcha 1.10. The Hessian is always symmetric. However, elsewhere in optimization and linear algebra, it is important to emphasize the distinction between symmetric and possibly non-symmetric positive definite matrices. Confusingly, some authors use SPD to mean "symmetric positive definite". So SPD means $\succ 0$, while PSD means $\succeq 0$, and both are symmetric.

1.3 Second-order necessary and sufficient conditions

Definition 1.11 (Hessian and quadratic form)**.** If f is twice differentiable, its Hessian at x is $H(x) := \nabla^2 f(x)$.

For small d, Taylor's theorem gives

$$f(x + d) = f(x) + \nabla f(x)^T d + \tfrac{1}{2} d^T H(x) d + o(\|d\|^2).$$

Theorem 1.12 (Second-order necessary condition (SONC))**.** *If f is twice differentiable and $x^\star$ is a local minimizer, then*

$$\nabla f(x^\star) = 0 \qquad \textit{and} \qquad \nabla^2 f(x^\star) \succeq 0.$$

Proof. We already have $\nabla f(x^\star) = 0$ by Theorem 1.4. By Taylor's theorem, for small d,

$$f(x^\star + d) = f(x^\star) + \tfrac{1}{2}\, d^T \nabla^2 f(x^\star)\, d + o(\|d\|^2).$$

If $\nabla^2 f(x^\star)$ had a direction d with $d^T \nabla^2 f(x^\star) d < 0$, then $f(x^\star + \alpha d) < f(x^\star)$ for sufficiently small $\alpha > 0$, contradicting local minimality. Thus $\nabla^2 f(x^\star) \succeq 0$. □

Theorem 1.13 (Second-order sufficient condition (SOSC))**.** *If $\nabla f(x^\star) = 0$ and $\nabla^2 f(x^\star) \succ 0$, then $x^\star$ is a* strict *local minimizer.*

Proof. Positive definiteness implies there exists $c > 0$ and $\delta > 0$ with $d^T \nabla^2 f(x^\star) d \geq c\|d\|^2$ for $\|d\| < \delta$. The same Taylor expansion gives, for $\|d\|$ small, $f(x^\star + d) \geq f(x^\star) + \frac{c}{2}\|d\|^2 + o(\|d\|^2) > f(x^\star)$ for $d \neq 0$. □

Example 1.14 (Flat extrema)**.** In 1D, $f(x) = x^4$ has $f'(0) = 0$, $f''(0) = 0$ but $x = 0$ is a (strict) local minimum. The SONC holds ($f''(0) \geq 0$) but SOSC does not (not > 0). Higher-order terms decide.

1.4 Worked examples

Example 1.15 (A saddle point)**.** $f(x_1, x_2) = x_1^2 - x_2^2$ has $\nabla f(0,0) = 0$ but $H = \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix}$ is indefinite, so $(0,0)$ is a saddle; not a local minimizer.

Exercise. For $f(x) = \frac{1}{2}\|Ax - b\|_2^2$ with $A \in \mathbb{R}^{m \times n}$: (i) Find the stationary condition; (ii) characterize when the solution is unique; (iii) relate your answer to $A^T A \succeq 0$.

1.5 Consequences for algorithms

- **Termination certificates.** If an algorithm returns x with $\|\nabla f(x)\|$ small and $H(x) \succeq -\epsilon I$, then x is an approximate second-order stationary point.
- **Newton/Quasi-Newton.** Newton steps rely on local quadratic models; convergence hinges on the value of the smallest eigenvalue of $H(x)$ near the solution (or regularization when H is indefinite).
- **Convexity.** We will see that for a convex function, any stationary point is automatically a global minimizer; in that setting, it suffices to check the FONC to determine global optimality!

Summary

- **FONC (necessary):** local min $\Rightarrow \nabla f = 0$.
- **SONC (necessary):** local min $\Rightarrow \nabla^2 f \succeq 0$.
- **SOSC (sufficient):** $\nabla f = 0$ and $\nabla^2 f \succ 0 \Rightarrow$ strict local min.

2 Convex Analysis

Convex analysis supplies the language and tools that make convex optimization so powerful: clean global optimality guarantees, robust modeling rules (epigraphs, sublevel sets, pointwise maxima, compositions), and the first-order theory for *nondifferentiable* functions via supporting hyperplanes and subgradients.

Convex sets

Definition 2.1 (Convex set)**.** A set $S \subseteq \mathbb{R}^n$ is *convex* if it contains every chord: for all $w, v \in S$ and $\theta \in [0,1]$,

$$\theta w + (1-\theta)v \in S.$$

Equivalently, any convex combination of points in S remains in S.

Geometrically, the line segment between any two points of S lies inside S.

Proposition 2.2 (Basic closure rules for sets)**.** *If S, T are convex, then so are:*

(a) $S \cap T$ (intersection),

(b) $S + T := \{s + t \mid s \in S,\ t \in T\}$ (Minkowski sum),

(c) $\{x \mid \exists\, y$ with $(x, y) \in S\}$ (projection of a convex set).

Proof sketches. (a) Intersections inherit the chord property. (b) For $\theta \in [0,1]$ and $s_i \in S$, $t_i \in T$, $\theta(s_1 + t_1) + (1-\theta)(s_2 + t_2) = [\theta s_1 + (1-\theta)s_2] + [\theta t_1 + (1-\theta)t_2] \in S + T$. (c) If x_1, x_2 are in the projected set, then there exist $(x_i, y_i) \in S$ for $i = 1, 2$. The line segment between (x_1, y_1) and (x_2, y_2) is contained in S, and its projection onto the x-space is the line segment between x_1 and x_2. □

Consequence for modeling. Eliminating variables (projections), combining uncertainty sets (Minkowski sums), and intersecting constraint sets preserve convexity of the feasible set. These operations are building blocks for constructing convex feasible sets.

Convex functions: four equivalent views

Definition 2.3 (Convex function)**.** A function $f : \mathbb{R}^n \to \mathbb{R}$ is *convex* if, for all $w, v \in \mathbb{R}^n$ and $\theta \in [0,1]$,

$$f(\theta w + (1-\theta)v) \ \le \ \theta f(w) + (1-\theta) f(v).$$

This is the *chord test*.

Proposition 2.4 (Equivalent characterizations)**.** *For $f : \mathbb{R}^n \to \mathbb{R}$, the following are equivalent:*

(a) (Chord test) $f(\theta w + (1-\theta)v) \le \theta f(w) + (1-\theta) f(v)$ for all w, v, $\theta \in [0,1]$.

(b) (Epigraph) $\operatorname{epi}(f) := \{(x,t) \mid t \ge f(x)\}$ is a convex subset of $\mathbb{R}^{n+1}$.

(c) (First-order condition, C^1) (if f is differentiable) $f(v) \ge f(w) + \nabla f(w)^\top (v - w)$ for all $w, v \in \mathbb{R}^n$.

(d) (Second-order condition, C^2) (if f is twice differentiable) $\lambda_{\min}(\nabla^2 f(x)) \ge 0$ for every $x \in \mathbb{R}^n$.

Proof of the proposition. We show $(a) \Leftrightarrow (b)$ for any f, $(a) \Leftrightarrow (c)$ when $f \in C^1$, and $(a) \Leftrightarrow (d)$ when $f \in C^2$.

(a) $\Rightarrow$ (b). Let $(x_1, t_1), (x_2, t_2) \in \operatorname{epi}(f)$, i.e., $t_i \ge f(x_i)$. For $\theta \in [0,1]$, set $x_\theta = \theta x_1 + (1-\theta) x_2$ and $t_\theta = \theta t_1 + (1-\theta) t_2$. By convexity (the chord test),

$$f(x_\theta) \ \le \ \theta f(x_1) + (1-\theta) f(x_2) \ \le \ \theta t_1 + (1-\theta) t_2 \ = \ t_\theta,$$

so $(x_\theta, t_\theta) \in \operatorname{epi}(f)$. Thus $\operatorname{epi}(f)$ is convex.

(b) $\Rightarrow$ (a). Fix w, v and $\theta \in [0,1]$. Since $(w, f(w))$ and $(v, f(v))$ belong to $\operatorname{epi}(f)$ and $\operatorname{epi}(f)$ is convex, the convex combination

$$\big(\theta w + (1-\theta) v,\ \theta f(w) + (1-\theta) f(v)\big)$$

also lies in $\operatorname{epi}(f)$. By the definition of epigraph,

$$f(\theta w + (1-\theta)v) \ \le \ \theta f(w) + (1-\theta) f(v),$$

which is the chord test.

Assume $f \in C^1$: (a) $\Rightarrow$ (c). Fix w, v and define $\phi(\alpha) := f\big(w + \alpha(v - w)\big)$ for $\alpha \in [0,1]$. We will show that the chord condition (a) implies the supporting line at $\alpha = 0$ underestimates ϕ:

$$\phi(1) \ \ge \ \phi(0) + \phi'(0) \cdot (1 - 0) \quad \Rightarrow \quad f(v) \ \ge \ f(w) + \nabla f(w)^\top (v - w),$$

since $\phi'(0) = \nabla f(w)^\top (v - w)$. To show this, let's argue by contradiction. If the inequality fails, then there exists $\epsilon > 0$ such that

$$\phi(\epsilon) \ < \ \phi(0) + \phi'(0) \cdot \epsilon.$$

Draw a line segment between $(0, \phi(0))$ and $(\epsilon, \phi(\epsilon))$; its slope is less than $\phi'(0)$, so it crosses the tangent line at some $\alpha \in (0, \epsilon)$. This implies that the chord between α and 0 lies below ϕ at some point, contradicting (a). Thus the supporting line inequality holds.

Assume $f \in C^1$: (c) $\Rightarrow$ (a). Let $x_\theta = \theta w + (1-\theta)v$. Apply (c) at the point x_θ with the test points w and v:

$$f(w) \geq f(x_\theta) + \nabla f(x_\theta)^\top (w - x_\theta),$$
$$f(v) \geq f(x_\theta) + \nabla f(x_\theta)^\top (v - x_\theta).$$

Multiply the first inequality by θ and the second by $(1-\theta)$ and add:

$$\theta f(w) + (1-\theta) f(v) \;\geq\; f(x_\theta) + \nabla f(x_\theta)^\top \big(\theta(w - x_\theta) + (1-\theta)(v - x_\theta)\big) = f(x_\theta),$$

because $\theta w + (1-\theta)v = x_\theta$. Rearranging yields the chord test.

Assume $f \in C^2$: (a) $\Rightarrow$ (d). Fix x and direction d. The univariate restriction $\phi(\alpha) := f(x + \alpha d)$ is convex by (a), hence $\phi''(0) \geq 0$. But $\phi''(0) = d^\top \nabla^2 f(x)\, d$, so $d^\top \nabla^2 f(x)\, d \geq 0$ for all d, i.e., $\nabla^2 f(x) \succeq 0$.

Assume $f \in C^2$: (d) $\Rightarrow$ (a). Fix w, v and set $d := v - w$. Consider $\phi(\alpha) := f(w + \alpha d)$ on $[0,1]$. By (d), $\phi''(\alpha) = d^\top \nabla^2 f(w + \alpha d)\, d \geq 0$, so ϕ is convex on $[0,1]$. Thus for $\theta \in [0,1]$,

$$f(\theta w + (1-\theta)v) = \phi(1-\theta) \;\leq\; \theta\, \phi(0) + (1-\theta)\, \phi(1) = \theta f(w) + (1-\theta) f(v),$$

the chord test.

Geometric intuition. (a)$\Leftrightarrow$(b): convexity of f says the straight-line chord of its graph lies *above* the graph; in (x,t)-space this is exactly "the epigraph is convex." (a)$\Leftrightarrow$(c): for a smooth convex f, every tangent hyperplane is a global underestimator (supports the epigraph); conversely, if all tangents lie below the graph, chords lie above it. (a)$\Leftrightarrow$(d): PSD Hessian means nonnegative curvature in every direction; along any line, f restricts to a convex univariate function, hence satisfies the chord test. □

Gotcha 2.5 (PSD at a single point is not enough). The condition $\nabla^2 f(x^\star) \succeq 0$ at one point is a *local* necessary condition for a local minimum, not a global convexity certificate; convexity via (d) requires $\nabla^2 f(x) \succeq 0$ *for all* x when $f \in C^2$.

Examples of convex and nonconvex functions

Example 2.6. Decide convexity and justify briefly:

(a) $f(x) = x^2$ on $\mathbb{R}$ (convex).

(b) $f(x) = |x|$ on $\mathbb{R}$ (convex, nondifferentiable at 0).

(c) $f(x) = x^\top A x$ on $\mathbb{R}^n$ with $A \succeq 0$ (convex by PSD quadratic form).

(d) $f(x) = x^\top A x$ with A indefinite (nonconvex; saddle directions).

(e) $f(x) = (x-1)(x-3)(x-5)$ (nonconvex; changing curvature).

(f) $f(x) = 1/x$ on $x > 0$ (convex on its domain; domain matters).

(g) Jump: $f(x) = \mathbf{1}_{\{x \geq 0\}}$ (not convex; not lower semicontinuous).

(h) Extended-value barrier:

$$f(x) = \begin{cases} 0 & x \in [-1, 1], \\ \infty & \text{else.} \end{cases}$$

encodes a hard interval constraint.

It is generally easiest to check convexity of quadratics using the second-order condition (d). For univariate functions, the chord test (a) is often easiest. The epigraph definition provides a nice proof that the extended-value barrier is convex: the epigraph is the (unbounded) rectangle $\{(x, t) \mid x \in [-1, 1], t \geq 0\}$, a convex set.

2.1 Operations preserving convexity of functions

Proposition 2.7 (Closure rules for functions)**.** *If $f, g : \mathbb{R}^n \to \mathbb{R}$ are convex, then so are:*

(a) cf *for* $c \geq 0$*,*

(b) $x \mapsto f(Ax + b)$ *for affine* $Ax + b$*,*

(c) $f + g$*,*

(d) $\max\{f, g\}$ *(pointwise maximum).*

Moreover, the composition $f \circ g$ is convex if g is convex and f is convex and (elementwise) nondecreasing. For C^2 scalar functions,

$$(f \circ g)''(x) \;=\; f''(g(x))\,(g'(x))^2 \;+\; f'(g(x))\,g''(x) \;\geq 0.$$

These rules allow us to build complex convex functions from simple ones, and to prove convexity of many functions we encounter in practice without resorting to the chord test.

Gotcha 2.8 (Composition pitfalls). If the outer function is convex but *decreasing*, convexity may be destroyed (the $f'(g(x))\,g''(x)$ term can flip sign).

Proof. We give short, self-contained arguments; when convenient we use the epigraph calculus $\operatorname{epi}(h) = \{(x, t) \mid t \geq h(x)\}$, together with the facts that (i) intersections and affine preimages of convex sets are convex, and (ii) projections of convex sets are convex.

(a) Positive scaling. For $c \geq 0$ and any $x, y, \theta \in [0, 1]$,

$$(cf)(\theta x + (1-\theta)y) = c\,f(\theta x + (1-\theta)y) \leq c\big(\theta f(x) + (1-\theta)f(y)\big) = \theta(cf)(x) + (1-\theta)(cf)(y).$$

Equivalently, $\operatorname{epi}(cf) = \{(x, t) \mid t \geq cf(x)\}$ is convex because it is the image of $\operatorname{epi}(f)$ under the linear map $(x, s) \mapsto (x, cs)$ with $c \geq 0$.

(b) Affine precomposition. Let $T(x) = Ax + b$ be affine. Using the chord test,

$$f(T(\theta x + (1-\theta)y)) = f\big(\theta T(x) + (1-\theta)T(y)\big) \le \theta f(T(x)) + (1-\theta) f(T(y)).$$

Epigraphically,

$$\mathrm{epi}(f \circ T) = \{(x,t) \mid (T(x), t) \in \mathrm{epi}(f)\}$$

is the affine preimage of the convex set $\mathrm{epi}(f)$, hence convex.

(c) Sum. By the chord test for f and g,

$$(f+g)(\theta x + (1-\theta)y) \le \theta f(x) + (1-\theta) f(y) + \theta g(x) + (1-\theta) g(y) = \theta (f+g)(x) + (1-\theta)(f+g)(y).$$

Epigraphically, one can also write

$$\mathrm{epi}(f+g) = \{(x,t) \mid \exists\, s,u:\ \ s \ge f(x),\ u \ge g(x),\ s+u \le t\},$$

a projection of the convex set $\{(x,s,u,t) \mid (x,s) \in \mathrm{epi}(f),\ (x,u) \in \mathrm{epi}(g),\ s+u \le t\}$; hence convex.

(d) Pointwise maximum. For any x, y, θ,

$$\begin{aligned} \max\{f,g\}(\theta x + (1-\theta)y) &\le \max\{\theta f(x) + (1-\theta) f(y),\ \theta g(x) + (1-\theta) g(y)\} \\ &\le \theta \max\{f(x), g(x)\} + (1-\theta) \max\{f(y), g(y)\}. \end{aligned}$$

Equivalently, $\mathrm{epi}(\max\{f,g\}) = \mathrm{epi}(f) \cap \mathrm{epi}(g)$, the intersection of convex sets, hence convex.

(e) Composition with a convex, nondecreasing outer map. Assume $g : \mathbb{R}^n \to \mathbb{R}$ is convex and $f : \mathbb{R} \to \mathbb{R}$ is convex and nondecreasing. For any x, y, θ,

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

by convexity of g, so by monotonicity of f,

$$f\big(g(\theta x + (1-\theta)y)\big) \ \le\ f\big(\theta g(x) + (1-\theta) g(y)\big) \ \le\ \theta f(g(x)) + (1-\theta) f(g(y)),$$

the last inequality by convexity of f on $\mathbb{R}$. Thus $f \circ g$ is convex.

Second-derivative check (scalar C^2 case). If $g : \mathbb{R} \to \mathbb{R}$ is convex ($g'' \ge 0$) and $f : \mathbb{R} \to \mathbb{R}$ is convex and nondecreasing ($f'' \ge 0$, $f' \ge 0$), then by the chain rule

$$(f \circ g)''(x) = f''(g(x))\,(g'(x))^2 + f'(g(x))\,g''(x) \ \ge\ 0,$$

so $f \circ g$ is convex. □

Jensen's inequality

Theorem 2.9 (Jensen). *If $f : \mathbb{R}^n \to \mathbb{R}$ is convex and X is a random variable, then*

$$f(\mathbb{E}[X]) \ \le\ \mathbb{E}[f(X)].$$

In the discrete case, this is the multi-point chord test for convex combinations.

2.2 Sublevel sets and quasiconvexity

Definition 2.10 (Sublevel set)**.** For $t \in \mathbb{R}$, the *sublevel set* of f at level t is

$$S_t := \{\, x \in \mathbb{R}^n \mid f(x) \le t \,\}.$$

Proposition 2.11 (Convex sublevel sets)**.** *If f is convex, then S_t is convex for every t.*

Proof. If $x, y \in S_t$, then by convexity $f(\theta x + (1-\theta)y) \le \theta f(x) + (1-\theta) f(y) \le t$. ☐

For an even more elegant proof, first consider the intersection of the epigraph of f with the halfspace $\{(x, s) \mid s \le t\}$, which is convex. Projecting this intersection onto the x-space yields S_t, which is therefore convex.

Definition 2.12 (Quasiconvex function)**.** A function f is *quasiconvex* if all its sublevel sets are convex.

Quasiconvexity is strictly weaker than convexity: every convex function is quasiconvex, but not vice versa. For example, $f(x) = \max\{x, 0\}$ is convex, while $f(x) = \min\{x, 0\}$ is quasiconvex but not convex. One-dimensional quasiconvex functions are precisely those that are monotonic or unimodal (decreasing then increasing).

Quasiconvexity is still a useful property for optimization, as local minima are still global minima.

Gotcha 2.13 (What you lose when only quasiconvex)*.* Quasiconvexity does *not* imply Jensen, global linear underestimators, or a subgradient calculus as rich as for convex functions.

2.3 Supporting hyperplanes for convex sets

Supporting hyperplanes generalize the first-order condition for smooth convex functions to nonsmooth convex functions and sets. They provide an alternative definition of convexity that forms a critical building block for convex duality.

Definition 2.14 (Supporting hyperplane)**.** A hyperplane $H = \{y \in \mathbb{R}^n \mid a^\top y = b\}$ *supports* a set S at $x \in S$ if $a^\top x = b$ and $a^\top y \ge b$ for all $y \in S$.

A differentiable convex function f has a supporting hyperplane at every point $(x, f(x))$ of its graph, given by the tangent plane $\{(y, t) \mid t = \nabla f(x)^\top (y - x) + f(x)\}$. More generally, the epigraph of any convex function has a supporting hyperplane at every boundary point $(x, f(x))$ with $x \in \operatorname{relint} \operatorname{dom} f$.

We will state but not prove the following fundamental theorem.

Theorem 2.15 (Existence of supports)**.** *Every nonempty convex set has a supporting hyperplane at every boundary point.*

A partial converse also holds: if a closed set with nonempty interior admits a supporting hyperplane at every boundary point, then it is convex.

2.4 Convex functions, epigraphs, and subgradients

Theorem 2.16 (Supporting hyperplanes $\Leftrightarrow$ convexity of f)**.** *f is convex if and only if, for every $x \in \operatorname{relint}\operatorname{dom} f$, the epigraph admits a supporting hyperplane at $(x, f(x))$. That is, there exists $g \in \mathbb{R}^n$ with*

$$f(y) \;\geq\; f(x) + g^\top (y - x) \quad \forall\, y.$$

This theorem generalizes the differentiable first-order condition to nondifferentiable functions, and gives a new way to assess whether a nondifferentiable function is convex.

Definition 2.17 (Subgradient and subdifferential)**.** A *subgradient* of f at x is any g satisfying $f(y) \geq f(x) + g^\top (y - x)$ for all y; the set of all subgradients is the *subdifferential* $\partial f(x)$.

Key facts. g is a subgradient $\iff$ the affine function $y \mapsto f(x) + g^\top (y - x)$ is a global underestimator of f; equivalently, $(g, -1)$ supports the epigraph at $(x, f(x))$. If f is convex and differentiable at x, then $\partial f(x) = \{\nabla f(x)\}$.

Example 2.18 (Pointwise maximum)**.** Let $f = \max\{f_1, f_2\}$ with f_1, f_2 convex and C^1. If $f_1(x) \neq f_2(x)$, then f is differentiable at x and $\nabla f(x) = \nabla f_i(x)$ for whichever function i is largest. If $f_1(x) = f_2(x)$, then

$$\partial f(x) = \operatorname{conv}\{\nabla f_1(x), \nabla f_2(x)\}.$$

The kink where the active function switches is precisely where f is nonsmooth and the subgradient is non-unique.

Proposition 2.19 (Subdifferential properties)**.** *For any extended-real f,*

$$\partial f(x) \;=\; \{\, g \mid f(y) \geq f(x) + g^\top (y - x) \;\forall y \,\}$$

is a closed, convex (possibly empty) set; $\partial f(x) = \varnothing$ if $f(x) = \infty$. If f is convex, then $\partial f(x) \neq \varnothing$ for $x \in \operatorname{relint}\operatorname{dom} f$; if f is convex and differentiable at x, $\partial f(x) = \{\nabla f(x)\}$, and conversely if $\partial f(x) = \{g\}$, then f is differentiable at x with $\nabla f(x) = g$.

Notice the subdifferential is *set-valued*!

Example 2.20 (ℓ_1 norm in 1D)**.** For $f(x) = |x|$,

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

This is the prototypical nonsmooth convex function; the fan of supporting lines at $x = 0$ has slopes in $[-1, 1]$.

2.5 Global optimality for convex objectives

Theorem 2.21 (Local $\Rightarrow$ global)**.** *If $x^\star$ is a local minimizer of a convex function f, then $x^\star$ is a global minimizer.*

Proof. If $x^\star$ is a local minimizer, there exists $\epsilon > 0$ such that $f(x^\star) \le f(y)$ for all y with $\|y - x^\star\| < \epsilon$. For any z, the chord test with $y = x^\star + \epsilon(z - x^\star)/\|z - x^\star\|$ gives

$$f(z) \;\ge\; \frac{\|z - x^\star\|}{\epsilon} f(y) + \left(1 - \frac{\|z - x^\star\|}{\epsilon}\right) f(x^\star) \;\ge\; f(x^\star).$$

□

Corollary 2.22 (First-order certificate)**.** *If f is convex and differentiable and $\nabla f(x^\star) = 0$, then $x^\star$ is a global minimizer. In general, $0 \in \partial f(x^\star)$ certifies global optimality for convex (possibly nonsmooth) f.*

Proof. If $\nabla f(x^\star) = 0$, then by the first-order condition for convexity,

$$f(y) \;\ge\; f(x^\star) + \nabla f(x^\star)^\top (y - x^\star) = f(x^\star) \quad \forall y,$$

so $x^\star$ is a global minimizer. If $0 \in \partial f(x^\star)$, then $f(y) \ge f(x^\star)$ for all y by the definition of subgradient. □

Why this matters for optimization. For convex f, any stationary point is a global minimizer, so local optimization methods (e.g., gradient descent) can find (and certify) global solutions.

2.6 Exercises

Exercise. Show that the projection of a polyhedron is a polyhedron, and use this to argue that projecting a convex feasible region yields a convex feasible region.

Exercise. Prove the composition rule: if g is convex and f is convex and nondecreasing, then $f \circ g$ is convex. Give a counterexample when f is convex but *decreasing*.

Exercise. Compute $\partial \|x\|_1$ at a point $x \in \mathbb{R}^n$. (Hint: use separability and the 1D formula for $|x|$.)

Exercise. Let $f = \max\{f_1, f_2\}$ with convex C^1 functions f_1, f_2. Derive $\partial f(x)$ in the cases $f_1(x) \neq f_2(x)$ and $f_1(x) = f_2(x)$.

3 Convex optimization

This section transitions from convex *analysis* (objects and properties) to *convex optimization problems*: mathematical programs whose geometry and calculus yield *global* guarantees, practical stopping criteria, and efficient algorithms.

3.1 What is a convex optimization problem?

Definition 3.1 (Convex optimization problem)**.** An optimization problem is *convex* if the feasible set is convex and the epigraph of the objective is convex.

For a problem written in the *nonlinear programming (NLP) form*

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \le 0, \quad i = 1, \dots, m, \\ & h_j(x) = 0, \quad j = 1, \dots, p, \\ \text{variable} & x \in \mathbb{R}^n, \end{array}$$

the problem is convex if and only if

1. the objective f_0 and inequality constraint functions are *convex* with f_0, f_i convex and the equality constraints *affine*.

Maximizing a *concave* function subject to convex constraints is also a convex optimization problem: simply minimize the negative of the concave function.

Remark 3.2 (Why convex optimization?). Convex optimization enjoys a relatively complete theory, efficient solvers, and LP-like tools such as duality and stopping conditions. Convexity is an important *analytic* tool, as it allows us to reason about global optimality and stability. It is an important *algorithmic* tool, as convex problems can be solved reliably and efficiently at scale. It is also a powerful *modeling* tool, as many problems can be modeled or approximated as convex programs, which allows us to leverage the analytic and algorithmic benefits.

Epigraph modeling. Many objectives can be modeled through an auxiliary variable t and an epigraph constraint:

$$\min_x f(x) \quad \Longleftrightarrow \quad \min_{x,t} \; t \;\; \text{s.t.} \;\; f(x) \le t,$$

preserving convexity and often simplifying compositions and pointwise maxima.

Gotcha 3.3 (Nonlinear equalities destroy convexity). Only *affine* equalities preserve convexity in the NLP view; general nonlinear equalities typically break convexity of the feasible set. For example, $x^2 + y^2 = 1$ describes a nonconvex set (the unit circle).