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

Optimality conditions and convexity

Professor Udell

Management Science and Engineering
Stanford

October 13, 2025

Outline

Constrained vs unconstrained optimization

constrained optimization

- examples: scheduling, routing, packing, logistics, scheduling, control
- what's hard: finding a feasible point

unconstrained optimization

- examples: data fitting, statistical/machine learning
- what's hard: reducing the objective

both are necessary for real-world problems!

Unconstrained smooth optimization

for $f : \mathbb{R}^n \to \mathbb{R}$ ctsly differentiable,

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{variable} & x \in \mathbb{R}^n \end{array}$$

examples:

- least squares
- logistic regression
- neural network training (with smooth activation like tanh, ELU, GeLU, ...)
- ...

Oracles

an optimization **oracle** is your interface for accessing the problem data:
e.g., an oracle for $f : \mathbb{R}^n \to \mathbb{R}$ can evaluate for any $x \in \mathbb{R}^n$:

- **zero-order:** $f_0(x)$
- **first-order:** $f_0(x)$ and $\nabla f_0(x)$
- **second-order:** $f_0(x)$, $\nabla f_0(x)$, and $\nabla^2 f_0(x)$

why oracles?

- can optimize real systems based on observed output (not just models)
- can use and extend old or complex but trusted code (*e.g.*, NASA, PDE simulations, ...)
- can prove lower bounds on the oracle complexity of a problem class

source: Nesterov 2004 "Introductory Lectures on Convex Optimization"

Outline

Constrained and unconstrained optimization

Optimality conditions

Convex analysis

Convex optimization

Solution of an optimization problem

$$\text{minimize} \quad f(x)$$

for $f : \mathcal{D} \to \mathbb{R}$. $x^\star$ is a

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

Solution of an optimization problem

$$\text{minimize} \quad f(x)$$

for $f : \mathcal{D} \to \mathbb{R}$. $x^\star$ is a

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

pictures!

First order optimality condition

Theorem

If $x^\star \in \mathbb{R}^n$ *is a local minimizer of a differentiable function* $f : \mathbb{R}^n \to \mathbb{R}$, *then* $\nabla f(x^\star) = 0$.

First order optimality condition

Theorem

If $x^\star \in \mathbb{R}^n$ is a local minimizer of a differentiable function $f : \mathbb{R}^n \to \mathbb{R}$, then $\nabla f(x^\star) = 0$.

proof: suppose by contradiction that $\nabla f(x^\star) \neq 0$. consider points of the form $x_\alpha = x^\star - \alpha \nabla f(x^\star)$ for $\alpha > 0$. by definition of the gradient,

$$\lim_{\alpha \to 0} \frac{f(x_\alpha) - f(x^\star)}{\alpha} = -\nabla f(x^\star)^\top \nabla f(x^\star) = -\|\nabla f(x^\star)\|^2 < 0$$

so for any sufficiently small $\alpha > 0$, we have $f(x_\alpha) < f(x^\star)$, which contradicts the fact that $x^\star$ is a local minimizer.

First order optimality condition

Theorem

If $x^\star \in \mathbb{R}^n$ is a local minimizer of a differentiable function $f : \mathbb{R}^n \to \mathbb{R}$, then $\nabla f(x^\star) = 0$.

proof: suppose by contradiction that $\nabla f(x^\star) \neq 0$. consider points of the form $x_\alpha = x^\star - \alpha \nabla f(x^\star)$ for $\alpha > 0$. by definition of the gradient,

$$\lim_{\alpha \to 0} \frac{f(x_\alpha) - f(x^\star)}{\alpha} = -\nabla f(x^\star)^\top \nabla f(x^\star) = -\|\nabla f(x^\star)\|^2 < 0$$

so for any sufficiently small $\alpha > 0$, we have $f(x_\alpha) < f(x^\star)$, which contradicts the fact that $x^\star$ is a local minimizer.

Definition

$x^\star \in \mathbb{R}^n$ is a **stationary point** of a differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ if $\nabla f(x^\star) = 0$.

is a stationary point always a local minimizer?

First order optimality condition

Theorem

If $x^\star \in \mathbb{R}^n$ is a local minimizer of a differentiable function $f : \mathbb{R}^n \to \mathbb{R}$, then $\nabla f(x^\star) = 0$.

proof: suppose by contradiction that $\nabla f(x^\star) \neq 0$. consider points of the form $x_\alpha = x^\star - \alpha \nabla f(x^\star)$ for $\alpha > 0$. by definition of the gradient,

$$\lim_{\alpha \to 0} \frac{f(x_\alpha) - f(x^\star)}{\alpha} = -\nabla f(x^\star)^\top \nabla f(x^\star) = -\|\nabla f(x^\star)\|^2 < 0$$

so for any sufficiently small $\alpha > 0$, we have $f(x_\alpha) < f(x^\star)$, which contradicts the fact that $x^\star$ is a local minimizer.

Definition

$x^\star \in \mathbb{R}^n$ is a **stationary point** of a differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ if $\nabla f(x^\star) = 0$.

is a stationary point always a local minimizer? no! saddle points, local maximizers.

Second order optimality condition

Theorem

If $x^\star \in \mathbb{R}^n$ *is a local minimizer of a twice differentiable function* $f : \mathbb{R}^n \to \mathbb{R}$*, then* $\nabla^2 f(x^\star) \succeq 0$.

Second order optimality condition

Theorem

If $x^\star \in \mathbb{R}^n$ is a local minimizer of a twice differentiable function $f : \mathbb{R}^n \to \mathbb{R}$, then $\nabla^2 f(x^\star) \succeq 0$.

proof: similar to the previous proof. use the fact that the second order approximation

$$f(x_\alpha) \approx f(x^\star) + \nabla f(x^\star)^\top (x_\alpha - x^\star) + \frac{1}{2}(x_\alpha - x^\star)^\top \nabla^2 f(x^\star)(x_\alpha - x^\star)$$

is accurate locally to show a contradiction unless $\nabla^2 f(x^\star) \succeq 0$: if not, there is a direction v such that $v^T \nabla^2 f(x^\star) v < 0$. then $f(x + \alpha v) < f(x^\star)$ for α arbitrarily small, which contradicts the fact that $x^\star$ is a local minimizer.

Symmetric positive semidefinite matrices

Definition

a symmetric matrix $Q \in \mathbb{R}^{n \times n}$ is **positive semidefinite** (psd) if $x^T Q x \geq 0$ for all $x \in \mathbb{R}^n$.

these matrices are so important that there are many ways to write them! for $Q \in \mathbb{R}^{n \times n}$,

$$Q \in \mathbf{S}_+^n \iff Q \succeq 0 \iff Q = Q^T,\ \lambda_{\min}(Q) \geq 0 \iff v^T Q v \geq 0 \quad \forall v \in \mathbb{R}^n$$

Symmetric positive semidefinite matrices

Definition

a symmetric matrix $Q \in \mathbb{R}^{n \times n}$ is **positive semidefinite** (psd) if $x^T Q x \geq 0$ for all $x \in \mathbb{R}^n$.

these matrices are so important that there are many ways to write them! for $Q \in \mathbb{R}^{n \times n}$,

$$Q \in \mathbf{S}^n_+ \iff Q \succeq 0 \iff Q = Q^T,\ \lambda_{\min}(Q) \geq 0 \iff v^T Q v \geq 0 \quad \forall v \in \mathbb{R}^n$$

$Q \in \mathbf{S}^n_{++}$ is **symmetric positive definite** (spd) ($Q \succ 0$) if $x^T Q x > 0$ for all $x \neq 0$.

Symmetric positive semidefinite matrices

Definition

a symmetric matrix $Q \in \mathbb{R}^{n \times n}$ is **positive semidefinite** (psd) if $x^T Qx \geq 0$ for all $x \in \mathbb{R}^n$.

these matrices are so important that there are many ways to write them! for $Q \in \mathbb{R}^{n \times n}$,

$$Q \in \mathbf{S}^n_+ \iff Q \succeq 0 \iff Q = Q^T, \ \lambda_{\min}(Q) \geq 0 \iff v^T Qv \geq 0 \quad \forall v \in \mathbb{R}^n$$

$Q \in \mathbf{S}^n_{++}$ is **symmetric positive definite** (spd) ($Q \succ 0$) if $x^T Qx > 0$ for all $x \neq 0$.

why care about psd matrices Q?

- least-squares objective is quadratic with psd Hessian $A^T A$
- level sets of $x^T Qx$ are (bounded) ellipsoids if $Q \succ 0$
- the quadratic form $x^T Qx$ is a metric iff $Q \succ 0$
- eigenvalue decomp and svd coincide for psd matrices

Outline

Convex sets

Definition

A set $S \subseteq \mathbb{R}^n$ is convex if it contains every chord: for all $\theta \in [0, 1]$, w, $v \in S$,

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

Convex sets

Definition

A set $S \subseteq \mathbb{R}^n$ is convex if it contains every chord: for all $\theta \in [0, 1]$, w, $v \in S$,

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

Q: Which of these are convex?
ellipsoid, crescent moon, . . .

Operations that preserve convexity

if $S \subseteq \mathbb{R}^n$ and $T \subseteq \mathbb{R}^n$ are convex, then so are:

- intersection: $S \cap T$
- sum: $S + T = \{s + t \mid s \in S, \; t \in T\}$
- projection: $\{x : (x, y) \in S\}$

Convex functions

a function $f : \mathbb{R}^n \to \mathbb{R}$ is convex iff

Convex functions

a function $f : \mathbb{R}^n \to \mathbb{R}$ is convex iff

- **Chords.** it never lies above its chord: $\forall \theta \in [0, 1]$, $w, v \in \mathbb{R}^n$

$$f(\theta w + (1 - \theta)v) \leq \theta f(w) + (1 - \theta)f(v)$$

Convex functions

a function $f : \mathbb{R}^n \to \mathbb{R}$ is convex iff

- **Chords.** it never lies above its chord: $\forall \theta \in [0,1]$, $w, v \in \mathbb{R}^n$

$$f(\theta w + (1-\theta)v) \leq \theta f(w) + (1-\theta) f(v)$$

- **Epigraph.** $\textbf{epi}(f) = \{(x,t) : \; t \geq f(x)\}$ is convex

Convex functions

a function $f : \mathbb{R}^n \to \mathbb{R}$ is convex iff

- **Chords.** it never lies above its chord: $\forall \theta \in [0, 1]$, $w, v \in \mathbb{R}^n$

$$f(\theta w + (1-\theta)v) \leq \theta f(w) + (1-\theta) f(v)$$

- **Epigraph.** $\mathbf{epi}(f) = \{(x, t) : \; t \geq f(x)\}$ is convex
- **First order condition.** if f is differentiable,

$$f(v) \geq f(w) + \nabla f(w)^\top (v - w), \qquad \forall w, v \in \mathbb{R}^n$$

Convex functions

a function $f : \mathbb{R}^n \to \mathbb{R}$ is convex iff

- **Chords.** it never lies above its chord: $\forall \theta \in [0,1]$, $w, v \in \mathbb{R}^n$

$$f(\theta w + (1-\theta)v) \leq \theta f(w) + (1-\theta) f(v)$$

- **Epigraph.** $\textbf{epi}(f) = \{(x,t) : \; t \geq f(x)\}$ is convex
- **First order condition.** if f is differentiable,

$$f(v) \geq f(w) + \nabla f(w)^\top (v - w), \qquad \forall w, v \in \mathbb{R}^n$$

- **Second order condition.** If f is twice differentiable, its Hessian is always psd:

$$\lambda_{\min}(\nabla^2 f(x)) \geq 0, \qquad \forall x \in \mathbb{R}^n$$

Convexity examples

Q: Which of these functions are convex?

- quadratic function $f(x) = x^2$ for $x \in \mathbb{R}$
- absolute value function $f(x) = |x|$ for $x \in \mathbb{R}$
- quadratic function $f(x) = x^T Ax$, $x \in \mathbb{R}^n$, $A \succeq 0$
- quadratic function $f(x) = x^T Ax$, A indefinite
- rollercoaster function (cubic) $f(x) = (x-1)(x-3)(x-5)$
- hyperbolic function $f(x) = 1/x$ for $x > 0$
- jump function $f(x) = 1$ if $x \geq 0$, $f(x) = 0$ otherwise
- jump to infinity function $f(x) = 1$ if $x \in [-1, 1]$, $f(x) = \infty$ otherwise

Operations that preserve convexity

if $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}$ are convex, then so are:

- cf for $c \geq 0$
- $f(Ax + b)$ for $A \in \mathbb{R}^{n \times m}$, $b \in \mathbb{R}^n$
- $f + g$
- $\max\{f, g\}$

Operations that preserve convexity

if $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}$ are convex, then so are:

- cf for $c \geq 0$
- $f(Ax + b)$ for $A \in \mathbb{R}^{n \times m}$, $b \in \mathbb{R}^n$
- $f + g$
- $\max\{f, g\}$

Q: Pick one and assume f and g are twice-differentiable. What is the easiest way to prove convexity?

Operations that preserve convexity

if $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}$ are convex, then so are:

- cf for $c \geq 0$
- $f(Ax + b)$ for $A \in \mathbb{R}^{n \times m}$, $b \in \mathbb{R}^n$
- $f + g$
- $\max\{f, g\}$

Q: Pick one and assume f and g are twice-differentiable. What is the easiest way to prove convexity?
most general rule:

$$f \circ g(x) = f(g(x)) \text{ is convex if } g \text{ is convex and } f \text{ is convex and nondecreasing}$$

since

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

Jensen's inequality

Jensen's inequality generalizes the chord condition to a distribution of points:

Theorem

If $f : \mathbb{R}^n \to \mathbb{R}$ is convex and X is a random variable, then

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

Sublevel set

Definition

The **sublevel set** of a function $f : \mathbb{R}^n \to \mathbb{R}$ at level t is

$$S_t = \{x \in \mathbb{R}^n \mid f(x) \leq t\}$$

Sublevel set

Definition

The **sublevel set** of a function $f : \mathbb{R}^n \to \mathbb{R}$ at level t is

$$S_t = \{x \in \mathbb{R}^n \mid f(x) \leq t\}$$

Theorem

A convex function $f : \mathbb{R}^n \to \mathbb{R}$ has convex sublevel sets.

Sublevel set

Definition

The **sublevel set** of a function $f : \mathbb{R}^n \to \mathbb{R}$ at level t is

$$S_t = \{x \in \mathbb{R}^n \mid f(x) \leq t\}$$

Theorem

A convex function $f : \mathbb{R}^n \to \mathbb{R}$ has convex sublevel sets.

proof: Jensen's inequality. if $x, y \in S_t$, then for $\theta \in [0, 1]$,

$$f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta)f(y) \leq \theta t + (1-\theta)t = t$$

so $\theta x + (1-\theta)y \in S_t$.

Quasiconvexity

converse is not true: a function can have all sublevel sets convex, and still be non-convex.

Definition

A function $f : \mathbb{R}^n \to \mathbb{R}$ is **quasiconvex** if its sublevel sets are convex.

examples of functions that are quasiconvex but not convex?

Supporting hyperplane

Definition

A **supporting hyperplane** to a set $S \subseteq \mathbb{R}^n$ at a point $x \in S$ is a hyperplane that touches S at x and lies entirely on one side of S:

$$H = \{y \in \mathbb{R}^n \mid a^\top y = b\} \text{ supports } S \text{ at } x \text{ if} \qquad \begin{aligned} a^\top x &= b \\ a^\top y &\geq b \quad \forall y \in S \end{aligned}$$

Supporting hyperplane

Definition

A **supporting hyperplane** to a set $S \subseteq \mathbb{R}^n$ at a point $x \in S$ is a hyperplane that touches S at x and lies entirely on one side of S:

$$H = \{y \in \mathbb{R}^n \mid a^\top y = b\} \text{ supports } S \text{ at } x \text{ if } \qquad \begin{aligned} a^\top x &= b \\ a^\top y &\geq b \quad \forall y \in S \end{aligned}$$

Theorem (Supporting hyperplane)

Any nonempty convex set has a supporting hyperplane at every boundary point.

Supporting hyperplane condition for convexity

Theorem (Partial converse)

If a closed set with nonempty interior has a supporting hyperplane at every boundary point, then it is convex.

Supporting hyperplane condition for convexity

Theorem (Partial converse)

If a closed set with nonempty interior has a supporting hyperplane at every boundary point, then it is convex.

Theorem

A function $f : \mathbb{R}^n \to \mathbb{R}$ is convex $\iff$ for all $x \in \textbf{relint dom}\, f$, the epigraph of f has a supporting hyperplane at $(x, f(x))$: for some $g \in \mathbb{R}^n$,

$$f(y) \geq f(x) + g^\top (y - x) \quad \forall y \in \mathbb{R}^n$$

Supporting hyperplane condition for convexity

Theorem (Partial converse)

If a closed set with nonempty interior has a supporting hyperplane at every boundary point, then it is convex.

Theorem

*A function $f : \mathbb{R}^n \to \mathbb{R}$ is convex $\iff$ for all $x \in$ **relint dom** f, the epigraph of f has a supporting hyperplane at $(x, f(x))$: for some $g \in \mathbb{R}^n$,*

$$f(y) \geq f(x) + g^\top (y - x) \quad \forall y \in \mathbb{R}^n$$

generalizes first-order condition for convexity to non-differentiable functions!

Supporting hyperplane condition for convexity

Theorem (Partial converse)

If a closed set with nonempty interior has a supporting hyperplane at every boundary point, then it is convex.

Theorem

*A function $f : \mathbb{R}^n \to \mathbb{R}$ is convex $\iff$ for all $x \in$ **relint dom** f, the epigraph of f has a supporting hyperplane at $(x, f(x))$: for some $g \in \mathbb{R}^n$,*

$$f(y) \geq f(x) + g^\top (y - x) \quad \forall y \in \mathbb{R}^n$$

generalizes first-order condition for convexity to non-differentiable functions!

Definition

A vector $g \in \mathbb{R}^n$ is a **subgradient** of $f : \mathbb{R}^n \to \mathbb{R}$ at $x \in \mathbb{R}^n$ if $f(y) \geq f(x) + g^\top (y - x)$ for all $y \in \mathbb{R}^n$.

Example: subgradients

$f = \max\{f_1, f_2\}$, with f_1, f_2 convex and differentiable

Q: Where is the function f differentiable? Where is the subgradient unique?

Subdifferential

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

$$\partial f(x) = \{g : f(y) \geq f(x) + g^T(y - x) \quad \forall y\}$$

Subdifferential

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

$$\partial f(x) = \{g : f(y) \geq f(x) + g^T(y - x) \quad \forall y\}$$

for any f,

- $\partial f(x)$ is a closed convex set (can be empty)
- $\partial f(x) = \emptyset$ if $f(x) = \infty$

proof: use the definition

Subdifferential

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

$$\partial f(x) = \{g : f(y) \geq f(x) + g^T(y - x) \quad \forall y\}$$

for any f,

- $\partial f(x)$ is a closed convex set (can be empty)
- $\partial f(x) = \emptyset$ if $f(x) = \infty$

proof: use the definition

if f is convex,

- $\partial f(x)$ is nonempty, for $x \in$ **relint dom** f
- $\partial f(x) = \{\nabla f(x)\}$, if f is differentiable at x
- if $\partial f(x) = \{g\}$, then f is differentiable at x and $g = \nabla f(x)$

Outline

Convex optimization

an optimization problem is convex if:

- **Geometrically:** the feasible set and the epigraph of the objective are convex

for example, a nonlinear minimization is convex if the objective and inequality constraints are convex functions, and the equality constraints are affine

$$
\begin{array}{ll}
\text{minimize} & f_0(x) \\
\text{subject to} & f_i(x) \leq b_i, \quad i = 1, \ldots, m_1 \\
& Ax = b_2 \\
\text{variable} & x \in \mathbb{R}^n
\end{array}
$$

Convex optimization

an optimization problem is convex if:

- **Geometrically:** the feasible set and the epigraph of the objective are convex

for example, a nonlinear minimization is convex if the objective and inequality constraints are convex functions, and the equality constraints are affine

$$
\begin{array}{ll}
\text{minimize} & f_0(x) \\
\text{subject to} & f_i(x) \leq b_i, \quad i = 1, \ldots, m_1 \\
 & Ax = b_2 \\
\text{variable} & x \in \mathbb{R}^n
\end{array}
$$

concave functions:

- a function f is concave if $-f$ is convex
- concave maximization $\implies$ a convex optimization problem

Why care about convex optimization?

- local optimality $\Rightarrow$ global optimality
- efficient solvers
- conceptual tools that generalize linear programming: duality, stopping conditions, ...

Local minima are global for convex functions

Theorem

If $x^\star$ is a local minimizer of a convex function f, then $x^\star$ is a global minimizer.

Local minima are global for convex functions

Theorem

If $x^\star$ is a local minimizer of a convex function f, then $x^\star$ is a global minimizer.

proof?

Local minima are global for convex functions

Theorem

If $x^\star$ is a local minimizer of a convex function f, then $x^\star$ is a global minimizer.

proof? suppose by contradiction that another point x' is a global minimizer, with $f(x') < f(x^\star)$. draw the chord between x' and $x^\star$. since the chord lies above f, every convex combination $x = \theta x^\star + (1-\theta)x'$ of x' and $x^\star$ for $\theta \in (0,1)$ has a value $f(x) < f(x^\star)$. this is true even for $x \to x^\star$, contradicting our assumption that $x^\star$ is a local minimizer.

Corollary

Corollary

If f is convex and differentiable and $\nabla f(x^\star) = 0$, then $x^\star$ is a global minimizer.

Corollary

Corollary

If f is convex and differentiable and $\nabla f(x^\star) = 0$, then $x^\star$ is a global minimizer.

Q: Is a global minimizer of a convex function always unique?

Corollary

Corollary

If f is convex and differentiable and $\nabla f(x^\star) = 0$, then $x^\star$ is a global minimizer.

Q: Is a global minimizer of a convex function always unique?
A: No. Picture.