CME 307 / MS&E 311: Optimization

Optimality conditions and convexity

Professor Udell

Management Science and Engineering
Stanford

January 29, 2024

Questions from last class

- clarify proof of strong duality
- how many iterations of branch and bound?
- how to use duality to solve a problem? when to stop?
- duality for problems with inequality constraints?

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

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

how to solve?

Unconstrained smooth optimization

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

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

how to solve? approximate as a quadratic problem

$$f(x) \approx f(x_0) + \nabla f(x_0)^T (x - x_0) + \frac{1}{2}(x - x_0)^T H(x_0)(x - x_0)$$

and find solution x_{quad} to the quadratic problem.
then set $x_0 \leftarrow x_{\text{quad}}$ and repeat.

Nonlinear optimization

optimization problem: nonlinear form

$$
\begin{array}{ll}
\text{minimize} & f_0(x) \\
\text{subject to} & f_i(x) \leq b_i, \quad i = 1, \dots, m_1 \\
& h(x) = 0 \\
\text{variable} & x \in \mathbf{R}^n
\end{array}
$$

- $x = (x_1, \dots, x_n)$: optimization variables
- $f_0 : \mathbf{R}^n \to \mathbf{R}$: objective function
- $f_i : \mathbf{R}^n \to \mathbf{R}$, $i = 1, \dots, m$: constraint functions

special case: **unconstrained optimization**

Example: process control

You are the process engineer for a desalination plant that produces drinking water. The plant has a variety of knobs, collected in vector x, that you can turn to control the process. These control, *e.g.*, how much water is pumped into the plant, how much pressure is used to force the water through filters, and how much of each chemical is added to the water.

- $f_0(x)$: cost of water produced
- $f_i(x)$: level of each measured impurity in the water
- b_i: maximum allowable level of each impurity

Given a setting of the knobs, you can observe the cost of water produced and the levels of impurities.

What is the optimal setting of the knobs?

Oracles

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

- **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"'

Nonlinear optimization: how to solve?

depends on the oracle:

- first- or second-order: approximate by a sequence of quadratic problems
- zero-order: harder, lots of methods
  - simulated annealing
  - Bayesian optimization
  - pseudo-higher-order methods, *e.g.*, compute approximate gradient

Outline

Solution of an optimization problem

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

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

First order optimality condition

Theorem

If $x^\star \in \mathbf{R}^n$ is a local minimizer of a differentiable function $f : \mathbf{R}^n \to \mathbf{R}$, 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.

Second order optimality condition

Theorem

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

Second order optimality condition

Theorem

If $x^\star \in \mathbf{R}^n$ is a local minimizer of a twice differentiable function $f : \mathbf{R}^n \to \mathbf{R}$, 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.

Outline

Convex sets

Definition

A set $S \subseteq \mathbf{R}^n$ 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 $S \subseteq \mathbf{R}^n$ 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, half moon

Convex functions

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

Convex functions

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

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

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

Convex functions

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

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

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

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

Convex functions

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

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

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

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

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

Convex functions

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

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

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

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

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

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

$$\lambda_{\min}(\nabla^2 f(x)) \geq 0 \qquad \text{for all} x \in \mathbf{R}^n$$

Convex functions

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

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

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

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

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

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

$$\lambda_{\min}(\nabla^2 f(x)) \geq 0 \qquad \text{for all} x \in \mathbf{R}^n$$

Q: Which of these are convex?
quadratic, abs, pwl, step, jump, logistic, logistic loss

Convex optimization

an optimization problem is convex if:

- **Geometrically:** the feasible set and the epigraph of the objective are convex
- **NLP:** the objective and inequality constraints are convex functions, and the equality constraints are affine

Convex optimization

an optimization problem is convex if:

- **Geometrically:** the feasible set and the epigraph of the objective are convex
- **NLP:** the objective and inequality constraints are convex functions, and the equality constraints are affine

why convex optimization?

- relatively complete theory
- efficient solvers
- conceptual tools that generalize

duality, stopping conditions, . . .

Convex optimization

an optimization problem is convex if:

- **Geometrically:** the feasible set and the epigraph of the objective are convex
- **NLP:** the objective and inequality constraints are convex functions, and the equality constraints are affine

why convex optimization?

- relatively complete theory
- efficient solvers
- conceptual tools that generalize

duality, stopping conditions, . . .

- a function f is concave if $-f$ is convex
- concave maximization results in a **convex** optimization problem

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: 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.

First-order condition

Definition

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

First-order condition

Definition

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

Q: Can a global minimum have a non-zero gradient?

First-order condition

Definition

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

Q: Can a global minimum have a non-zero gradient?
A: No.

First-order condition

Definition

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

Q: Can a global minimum have a non-zero gradient?
A: No.
Q: Is a stationary point always a global minimum?

First-order condition

Definition

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

Q: Can a global minimum have a non-zero gradient?
A: No.
Q: Is a stationary point always a global minimum?
A: No.

First-order condition

Definition

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

Q: Can a global minimum have a non-zero gradient?
A: No.
Q: Is a stationary point always a global minimum?
A: No.
Q: . . . for convex functions?

First-order condition

Definition

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

Q: Can a global minimum have a non-zero gradient?
A: No.
Q: Is a stationary point always a global minimum?
A: No.
Q: . . . for convex functions?
A: Yes.

First-order condition

Definition

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

Q: Can a global minimum have a non-zero gradient?
A: No.
Q: Is a stationary point always a global minimum?
A: No.
Q: . . . for convex functions?
A: Yes.

$\nabla f(x^\star) = 0$ is the **first-order (necessary) condition** for optimality.

Invex function

Definition

A function $f : \mathbf{R}^n \to \mathbf{R}$ is **invex** if for some vector-valued function $\eta : \mathbf{R}^n \times \mathbf{R}^n \to \mathbf{R}^n$,

$$f(x) - f(u) \geq \eta(x, u)^\top \nabla f(u) \qquad \forall u \in \mathbf{R}^n,\ x \in \mathbf{dom}\, f$$

Theorem (Craven and Glover, Ben-Israel and Mond)

A function is invex iff every stationary point is a global minimum.