CME 307 / MS&E 311: Optimization

Quadratic optimization

Professor Udell

Management Science and Engineering, Stanford

February 28, 2024

Questions from last time

- why require the matrix Q in equality-constrained QP to be psd?
- when would you use the second-order condition to prove convexity?
- invexity is confusing. (luckily, also unimportant!)

Outline

Quadratic optimization

Quadratic approximations

Quadratic optimization

a **quadratic optimization** problem is written as

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}\|Ax - b\|^2 := f_0(x) \\ \text{variable} & x \in \mathbf{R}^n \end{array}$$

where

- $A \in \mathbf{R}^{m \times n}$: matrix
- $b \in \mathbf{R}^m$: vector

how to solve?

Quadratic optimization

a **quadratic optimization** problem is written as

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}\|Ax - b\|^2 := f_0(x) \\ \text{variable} & x \in \mathbf{R}^n \end{array}$$

where

- $A \in \mathbf{R}^{m \times n}$: matrix
- $b \in \mathbf{R}^m$: vector

how to solve? take gradient and set to 0:

$$\nabla f_0(x) = A^T(Ax - b) = 0$$

$\Longrightarrow$ linear system solvers also solve quadratic problems

Symmetric positive semidefinite matrices

Definition

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

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

$$Q \in \mathbf{S}^n_+ \iff Q \succeq 0 \iff Q = Q^T, \ \lambda_{\min}(Q) \geq 0$$

Symmetric positive semidefinite matrices

Definition

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

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

$$Q \in \mathbf{S}^n_+ \iff Q \succeq 0 \iff Q = Q^T, \; \lambda_{\min}(Q) \geq 0$$

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

Symmetric positive semidefinite matrices

Definition

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

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

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

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

why care about psd matrices Q?

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

Quadratic program

an equality constrained **quadratic program** is written as

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}x^T Qx + c^T x \\ \text{subject to} & Ax = b \\ \text{variable} & x \in \mathbf{R}^n \end{array}$$

where

- $Q \in \mathbf{R}^{n \times n}$: symmetric positive semidefinite matrix
- $c \in \mathbf{R}^n$: vector

how to solve?

Quadratic program

an equality constrained **quadratic program** is written as

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}x^TQx + c^Tx \\ \text{subject to} & Ax = b \\ \text{variable} & x \in \mathbf{R}^n \end{array}$$

where

- $Q \in \mathbf{R}^{n \times n}$: symmetric positive semidefinite matrix
- $c \in \mathbf{R}^n$: vector

how to solve? reduce to quadratic optimization problem:

- (explicit) form solution set $\{x : Ax = b\} = \{x_0 + Vz \mid z \in \mathbf{R}^{n-m}\}$ by computing a solution $Ax_0 = b$ and a basis V for the null space of A
- (implicit) use duality to recast problem as larger linear (KKT) system

Quadratic program

an equality constrained **quadratic program** is written as

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}x^TQx + c^Tx \\ \text{subject to} & Ax = b \\ \text{variable} & x \in \mathbf{R}^n \end{array}$$

where

- $Q \in \mathbf{R}^{n \times n}$: symmetric positive semidefinite matrix
- $c \in \mathbf{R}^n$: vector

how to solve? reduce to quadratic optimization problem:

- (explicit) form solution set $\{x : Ax = b\} = \{x_0 + Vz \mid z \in \mathbf{R}^{n-m}\}$ by computing a solution $Ax_0 = b$ and a basis V for the null space of A
- (implicit) use duality to recast problem as larger linear (KKT) system
- inequality constraints: harder. http://www.cs.cornell.edu/courses/cs4220/2017sp/lec/2017-04-28.pdf has details.

Solving equality-constrained quadratic program

$x^\star \in \mathbf{R}^n$ solves the equality-constrained quadratic program

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}x^TQx + c^Tx \\ \text{subject to} & Ax = b \\ \text{variable} & x \in \mathbf{R}^n \end{array}$$

$\iff$ there exists $\lambda^\star \in \mathbf{R}^m$ such that

$$\begin{bmatrix} Q & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x^\star \\ \lambda^\star \end{bmatrix} = \begin{bmatrix} -c \\ b \end{bmatrix}$$

Solving equality-constrained quadratic program

$x^\star \in \mathbf{R}^n$ solves the equality-constrained quadratic program

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}x^TQx + c^Tx \\ \text{subject to} & Ax = b \\ \text{variable} & x \in \mathbf{R}^n \end{array}$$

$\iff$ there exists $\lambda^\star \in \mathbf{R}^m$ such that

$$\begin{bmatrix} Q & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x^\star \\ \lambda^\star \end{bmatrix} = \begin{bmatrix} -c \\ b \end{bmatrix}$$

proof: form Lagrangian

$$\mathcal{L}(x, \lambda) = \frac{1}{2}x^TQx + c^Tx + \lambda^T(Ax - b)$$

and solve for $\bar{x}$, $\bar{\lambda}$ so that $\nabla\mathcal{L}(\bar{x}, \bar{\lambda}) = 0$.

- $\frac{1}{2}\bar{x}^TQ\bar{x} + c^T\bar{x}$ provides an upper bound on $p^\star$. (why?)
- $\frac{1}{2}\bar{x}^TQ\bar{x} + c^T\bar{x}$ provides a lower bound on $p^\star$. (why?)

Quadratic program: application

Markowitz portfolio optimization problem:

$$\begin{array}{ll} \text{minimize} & \gamma x^T \Sigma x - \mu^T x \\ \text{subject to} & \sum_i x_i = 1 \\ & Ax = 0 \\ \text{variable} & x \in \mathbf{R}^n \end{array}$$

where

- $\Sigma \in \mathbf{R}^{n \times n}$: asset covariance matrix
- $\mu \in \mathbf{R}^n$: asset return vector
- $\gamma \in \mathbf{R}$: risk aversion parameter
- rows of $A \in \mathbf{R}^{m \times n}$ correspond to other portfolios
  - ensures new portfolio is independent, *e.g.*, of market returns

Quadratic program: application

control system design problem:

$$x^+ = Ax + Bu$$

- $x \in \mathbf{R}^n$: state (e.g., position, velocity)
- $u \in \mathbf{R}^m$: control (e.g., force, torque)

$$\begin{array}{ll} \text{minimize} & \sum_{t=1}^{T} x_t^T Q x_t + u_t^T R u_t \\ \text{subject to} & x_{t+1} = Ax_t + Bu_t, \quad t = 0, \ldots, T-1 \\ & x_0 = x^{\text{init}} \end{array}$$

Outline

Quadratic approximation

Suppose $f : \mathbf{R} \to \mathbf{R}$ is twice differentiable. For any $x \in \mathbf{R}$, approximate f about x:

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

If f is a quadratic function, $\nabla^2 f(x) = H$ is constant.

Quadratic approximation

Suppose $f : \mathbf{R} \to \mathbf{R}$ is twice differentiable. For any $x \in \mathbf{R}$, approximate f about x:

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

If f is a quadratic function, $\nabla^2 f(x) = H$ is constant.

Quadratic approximations are useful because quadratics are easy to minimize:

$$\begin{aligned} y^\star &= \operatorname*{argmin}_y f(x) + \nabla f(x)^T (y-x) + \frac{1}{2}(y-x)^T H (y-x) \\ &\implies \nabla f(x) + H(y^\star - x) = 0 \\ y^\star &= x - H^{-1}(\nabla f(x)). \end{aligned}$$

Quadratic approximation

Suppose $f : \mathbf{R} \to \mathbf{R}$ is twice differentiable. For any $x \in \mathbf{R}$, approximate f about x:

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

If f is a quadratic function, $\nabla^2 f(x) = H$ is constant.

Quadratic approximations are useful because quadratics are easy to minimize:

$$\begin{aligned} y^\star &= \operatorname*{argmin}_y f(x) + \nabla f(x)^T(y-x) + \frac{1}{2}(y-x)^T H(y-x) \\ &\implies \nabla f(x) + H(y^\star - x) = 0 \\ y^\star &= x - H^{-1}(\nabla f(x)). \end{aligned}$$

If we approximate the Hessian of f by $H = \frac{1}{t}I$ for some $t > 0$ and choose x^+ to minimize the quadratic approximation, we obtain the **gradient descent** update with step size t:

$$x^+ = x + -t\nabla f(x)$$

Quadratic upper bound

Definition (Smooth)

A function $f : \mathbf{R} \to \mathbf{R}$ is L-**smooth** if for all $x, y \in \mathbf{R}$,

$$f(y) \leq f(x) + \nabla f(x)^T (y - x) + \frac{L}{2} \|y - x\|^2.$$

Equivalently, assuming the derivatives exist,

- the operator $\frac{1}{L} \nabla f$ is L-**Lipschitz continuous**:

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

- $\nabla^2 f(x) \preceq LI$ for all $x \in \mathbf{dom}\, f$.

Quadratic upper bound

Definition (Smooth)

A function $f : \mathbf{R} \to \mathbf{R}$ is L-**smooth** if for all $x, y \in \mathbf{R}$,

$$f(y) \leq f(x) + \nabla f(x)^T (y - x) + \frac{L}{2}\|y - x\|^2.$$

Equivalently, assuming the derivatives exist,

- the operator $\frac{1}{L}\nabla f$ is L-**Lipschitz continuous**:

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

- $\nabla^2 f(x) \preceq LI$ for all $x \in \textbf{dom}\, f$.

Q: For $A \succeq 0$, the quadratic function $f(x) = \frac{1}{2}x^T Ax$ is ?-smooth

Quadratic upper bound

Definition (Smooth)

A function $f : \mathbf{R} \to \mathbf{R}$ is L-**smooth** if for all $x, y \in \mathbf{R}$,

$$f(y) \leq f(x) + \nabla f(x)^T (y - x) + \frac{L}{2} \|y - x\|^2.$$

Equivalently, assuming the derivatives exist,

- the operator $\frac{1}{L} \nabla f$ is L-**Lipschitz continuous**:

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

- $\nabla^2 f(x) \preceq LI$ for all $x \in \mathbf{dom}\, f$.

Q: For $A \succeq 0$, the quadratic function $f(x) = \frac{1}{2} x^T A x$ is ?-smooth
A: $\lambda_{\max}(A)$-smooth

Quadratic lower bound

Definition (Strongly convex)

A function $f : \mathbf{R} \to \mathbf{R}$ is μ-**strongly convex** if for all $x, y \in \mathbf{R}$,

$$f(y) \geq f(x) + \nabla f(x)^T (y - x) + \frac{\mu}{2} \|y - x\|^2.$$

Equivalently, assuming the derivatives exist,

- the operator $\frac{1}{\mu} \nabla f$ is μ-**coercive**:

$$\|\nabla f(y) - \nabla f(x)\| \geq \mu \|y - x\|$$

- $\nabla^2 f(x) \succeq \mu I$ for all $x \in \textbf{dom}\, f$.

Quadratic lower bound

Definition (Strongly convex)

A function $f : \mathbf{R} \to \mathbf{R}$ is μ-**strongly convex** if for all $x, y \in \mathbf{R}$,

$$f(y) \geq f(x) + \nabla f(x)^T(y - x) + \frac{\mu}{2}\|y - x\|^2.$$

Equivalently, assuming the derivatives exist,

- the operator $\frac{1}{\mu}\nabla f$ is μ-**coercive**:

$$\|\nabla f(y) - \nabla f(x)\| \geq \mu\|y - x\|$$

- $\nabla^2 f(x) \succeq \mu I$ for all $x \in \mathbf{dom}\, f$.

Q: For $A \succeq 0$, the quadratic function $f(x) = \frac{1}{2}x^T Ax$ is ?-strongly convex

Quadratic lower bound

Definition (Strongly convex)

A function $f : \mathbf{R} \to \mathbf{R}$ is μ-**strongly convex** if for all $x, y \in \mathbf{R}$,

$$f(y) \geq f(x) + \nabla f(x)^T (y - x) + \frac{\mu}{2}\|y - x\|^2.$$

Equivalently, assuming the derivatives exist,

- the operator $\frac{1}{\mu}\nabla f$ is μ-**coercive**:

$$\|\nabla f(y) - \nabla f(x)\| \geq \mu \|y - x\|$$

- $\nabla^2 f(x) \succeq \mu I$ for all $x \in \mathbf{dom}\, f$.

Q: For $A \succeq 0$, the quadratic function $f(x) = \frac{1}{2}x^T A x$ is ?-strongly convex
A: $\lambda_{\min}(A)$-strongly convex

Contrast to strict convexity

Definition (Strictly convex)

A function $f : \mathbf{R} \to \mathbf{R}$ is **strictly convex** if for all $x, y \in \mathbf{R}$,

$$f(y) > f(x) + \nabla f(x)^T (y - x)$$

intuitively, the function has no flat spots.

Contrast to strict convexity

Definition (Strictly convex)

A function $f : \mathbf{R} \to \mathbf{R}$ is **strictly convex** if for all $x, y \in \mathbf{R}$,

$$f(y) > f(x) + \nabla f(x)^T (y - x)$$

intuitively, the function has no flat spots.

Q: Give an example of a function that is strictly convex but not strongly convex.

Some important functions

for $A \in \mathbf{R}^{m \times n}$, $b \in \mathbf{R}^m$, $x \in \mathbf{R}^n$,

- **Quadratic loss.** $\|Ax - b\|^2$
- **Logistic loss.** $f(x) = \sum_{i=1}^m \log(1 + \exp\left(b_i a_i^T x\right))$
  where a_i is ith row of A

Some important functions

for $A \in \mathbf{R}^{m \times n}$, $b \in \mathbf{R}^m$, $x \in \mathbf{R}^n$,

- **Quadratic loss.** $\|Ax - b\|^2$
- **Logistic loss.** $f(x) = \sum_{i=1}^m \log(1 + \exp\left(b_i a_i^T x\right))$
  where a_i is ith row of A

Q: Which of these are smooth? Under what conditions?

Some important functions

for $A \in \mathbf{R}^{m \times n}$, $b \in \mathbf{R}^m$, $x \in \mathbf{R}^n$,

- **Quadratic loss.** $\|Ax - b\|^2$
- **Logistic loss.** $f(x) = \sum_{i=1}^m \log(1 + \exp\left(b_i a_i^T x\right))$
  where a_i is ith row of A

Q: Which of these are smooth? Under what conditions?
A: Both.

Some important functions

for $A \in \mathbf{R}^{m \times n}$, $b \in \mathbf{R}^m$, $x \in \mathbf{R}^n$,

- **Quadratic loss.** $\|Ax - b\|^2$
- **Logistic loss.** $f(x) = \sum_{i=1}^m \log(1 + \exp\left(b_i a_i^T x\right))$
  where a_i is ith row of A

Q: Which of these are smooth? Under what conditions?
A: Both.
Q: Which of these are strongly convex? Under what conditions?

Some important functions

for $A \in \mathbf{R}^{m \times n}$, $b \in \mathbf{R}^m$, $x \in \mathbf{R}^n$,

- **Quadratic loss.** $\|Ax - b\|^2$
- **Logistic loss.** $f(x) = \sum_{i=1}^m \log(1 + \exp\left(b_i a_i^T x\right))$
  where a_i is ith row of A

Q: Which of these are smooth? Under what conditions?
A: Both.
Q: Which of these are strongly convex? Under what conditions?
A: Quadratic loss is strongly convex if A is rank n. Logistic loss is strongly convex on a compact domain if A is rank n.

Optimizing the upper bound

start at $x^{(0)}$. suppose f is L-smooth, so for all $y \in \mathbf{R}$,

$$f(y) \leq f(x^{(0)}) + \nabla f(x)^T (y - x^{(0)}) + \frac{L}{2}\|y - x^{(0)}\|^2$$

let's choose next iterate $x^{(1)}$ to minimize this upper bound:

$$\begin{aligned} x^{(1)} &= \operatorname*{argmin}_y f(x) + \nabla f(x)^T (y - x) + \frac{L}{2}\|y - x\|^2 \\ &\implies \nabla f(x^{(0)}) + L(x^{(1)} - x^{(0)}) = 0 \\ x^{(1)} &= x^{(0)} - \frac{1}{L}\nabla f(x^{(0)}) \end{aligned}$$

Optimizing the upper bound

start at $x^{(0)}$. suppose f is L-smooth, so for all $y \in \mathbf{R}$,

$$f(y) \leq f(x^{(0)}) + \nabla f(x)^T (y - x^{(0)}) + \frac{L}{2}\|y - x^{(0)}\|^2$$

let's choose next iterate $x^{(1)}$ to minimize this upper bound:

$$\begin{aligned} x^{(1)} &= \operatorname*{argmin}_y f(x) + \nabla f(x)^T (y - x) + \frac{L}{2}\|y - x\|^2 \\ &\implies \nabla f(x^{(0)}) + L(x^{(1)} - x^{(0)}) = 0 \\ x^{(1)} &= x^{(0)} - \frac{1}{L}\nabla f(x^{(0)}) \end{aligned}$$

- **gradient descent** update with step size $t = \frac{1}{L}$
- lower bound ensures true optimum can't be too far away...