CSE 203B: Convex Optimization Week 3 Discuss Session

Xinyuan Wang

01/23/2019

Contents

- Related Mathematics (Ref. Appendix A4)
  - Function properties
  - Gradient and chain rule
- Convex functions (Ref. Chap.3)
  - Jensen’s inequality and its extensions
  - First order condition
  - Second order condition

Functions and Related Properties

- For $f: A \to B$ we say that f is a function on the set $\mathrm{dom}\, f \subseteq A$ into the set B.
  - $f: R^n \to R^m$ means that the function maps some n-vectors into m-vectors; the $\mathrm{dom}\, f$ is a subset of R^n but not necessarily covers the whole space.
- Continuity
  - A function $f: R^n \to R^m$ is ***continuous*** at $x \in \mathrm{dom}\, f$ if for all $\epsilon > 0$ there exists a δ such that

$$y \in \mathrm{dom}\, f, \qquad \left|\left|y - x\right|\right|_2 \leq \delta \Rightarrow \left|\left|f(y) - f(x)\right|\right|_2 \leq \epsilon$$

- Closed function
  - A function $f: R^n \to R$ is ***closed*** if for each $\alpha \in R$, the sublevel set

$$\{x \in dom\, f \mid f(x) \leq \alpha\}$$

    is closed.
  - Open and closed set: a set C is *open* if $C = \mathbf{int}\, C$; a set $C \in R^n$ is *closed* if its complement $R^n \backslash C$ is open.

Derivative and Gradient

- Suppose $f: R^n \to R^m$ and $x \in \operatorname{int}\operatorname{dom} f$. The function f is differentiable at x if there exists a matrix $Df(x) \in R^{m\times n}$ that satisfies

$$\lim_{z\in\operatorname{dom} f, z\neq x, z\to x} \frac{||f(z) - f(x) - Df(x)(z-x)||}{||z-x||_2} = 0.$$

- First-order approximation

We call $Df(x)$ the ***derivative (Jacobian)*** of f at x where

$$Df(x)_{ij} = \frac{\partial f_i(x)}{\partial x_j}, \qquad i = 1, \dots, m, \qquad j = 1, \dots, n.$$

- The function f is ***differentiable*** if $\operatorname{dom} f$ is open, and differentiable at $\forall x \in \operatorname{dom} f$.
- Gradient: transpose of derivative (row -> column)

$$\nabla f(x) = Df(x)^T$$

first-order approximation $\hat{f}(x) = f(x) + \nabla f(x)^T(z-x)$

Derivative and Gradient

- Suppose $f: R^n \to R$ and second-order differentiable

Gradient: $\nabla f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix}$, Hessian $\nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$

second-order approximation

$$\hat{f}(x) = f(x) + \nabla f(x)^T (z - x) + \frac{1}{2}(z - x)^T \nabla^2 f(x)(z - x)$$

Example: calculate the gradient and Hessian of a quadratic function $f(x) = \frac{1}{2} x^T P x + q^T x + r$, *where* $P \in S^n, q \in R^n, r \in R$.

derivative $Df(x) = x^T P + q^T$, *gradient* $\nabla f(x) = Px + q$

Hessian $\nabla^2 f(x) = D\nabla f(x) = P$

Chain Rule

- Suppose $f: R^n \to R^m$ is differentiable at $x \in \text{int dom } f$ and $g: R^m \to R^p$ is differentiable at $f(x) \in \text{int dom } g$. Define the composition $h: R^n \to R^p$ $h(x) = g\big(f(x)\big)$. Then h is differentiable at x

$$Dh(x) = Dg\big(f(x)\big)Df(x)$$

Example: prove the convexity of $g(x) = f(Ax + b)$ with $dom\ g = \{x | Ax + b \in dom\ f\}$. It is a composition with affine function discussed in class.

By the chain rule: $Dg(x) = Df(Ax + b)A$

the gradient $\nabla g(x) = A^T \nabla f(Ax + b)$

the Hessian $\nabla^2 g(x) = D\nabla g(x) = A^T \nabla^2 f(Ax + b)A$

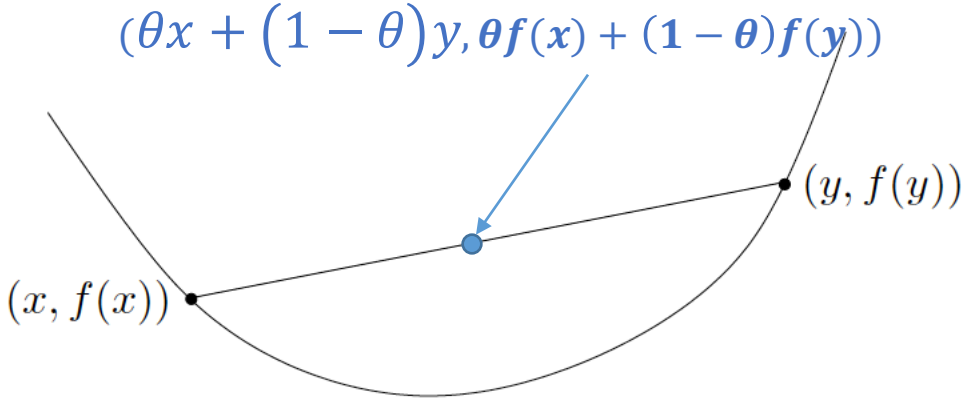
Definition of Convex Functions

- A function $f: R^n \rightarrow R$ is convex if $\mathrm{dom}\ f$ is a convex set and if for all $x, y \in \mathrm{dom}\ f$ and $0 \leq \theta \leq 1$

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

sometimes called Jensen's inequality

- Review the proof in class: necessary and sufficiency
- Strict convexity: $f(\theta x + (1-\theta)y) < \theta f(x) + (1-\theta) f(y), x \neq y, 0 < \theta < 1$
- Concave functions: $-f$ is convex



Example: convexity of functions with definition

3.4 [RV73, page 15] Show that a continuous function $f : \mathbf{R}^n \to \mathbf{R}$ is convex if and only if for every line segment, its average value on the segment is less than or equal to the average of its values at the endpoints of the segment: For every $x,\ y \in \mathbf{R}^n$,

$$\int_0^1 f(x + \lambda(y - x))\, d\lambda \leq \frac{f(x) + f(y)}{2}.$$

i. Necessary: suppose f is convex, then f satisfies the basic inequality

$$f\big(x + \lambda(y - x)\big) \leq f(x) + \lambda(f(y) - f(x))$$

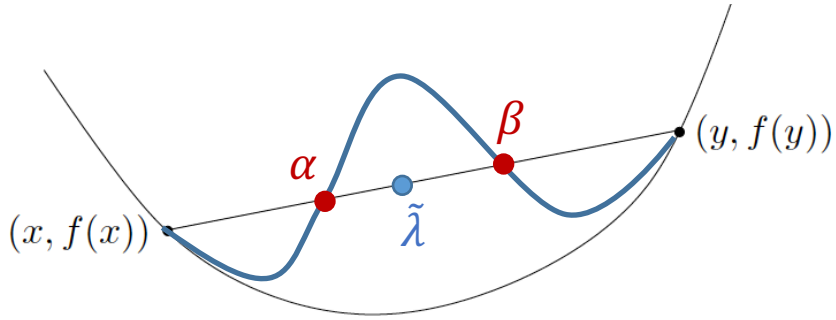
for $0 \leq \lambda \leq 1$. Integrating both sides from 0 to 1 on λ

$$\int_0^1 f\big(x + \lambda(y - x)\big) d\lambda \leq \int_0^1 f(x) + \lambda\big(f(y) - f(x)\big)\, d\lambda = \frac{f(x) + f(y)}{2}$$

ii. Sufficiency: suppose f is not convex, then there exists $0 \leq \tilde{\lambda} \leq 1$ and

$$f\left(x + \tilde{\lambda}(y - x)\right) > f(x) + \tilde{\lambda}(f(y) - f(x))$$

there exist $\alpha, \beta \in [0,1]$ so that in the interval $[\alpha, \beta]$ the above inequality always holds, which contradicts the given inequality.

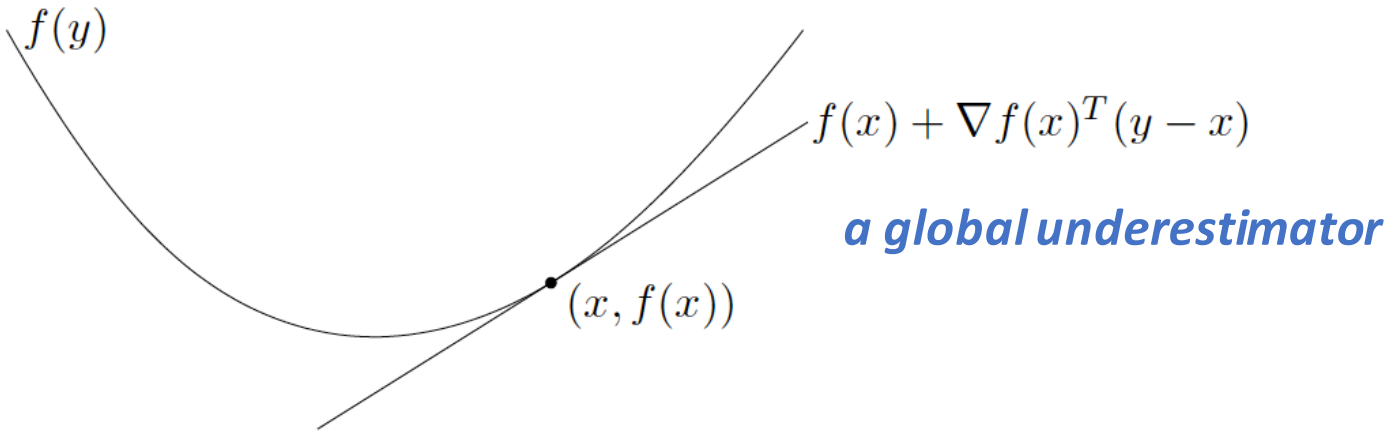


First-order Condition

- Suppose f is differentiable ($dom\ f$ is open and ∇f exists at $\forall x \in dom\ f$), then f is convex iff $\mathrm{dom}\ f$ is convex and for all $x, y \in dom\ f$

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

- Review the proof in class: necessary and sufficiency
- Strict convexity: $f(y) > f(x) + \nabla f(x)^T (y - x), x \neq y$
- Concave functions: $f(y) \leq f(x) + \nabla f(x)^T (y - x)$



a global underestimator

Second Order Condition

- Suppose f is twice differentiable ($dom\ f$ is open and its Hessian exists at $\forall x \in dom\ f$), then f is convex iff $\mathrm{dom}\ f$ is convex and for all $x, y \in dom\ f$

$$\nabla^2 f(x) \succcurlyeq 0 \text{ (positive semidefinite)}$$

- Review the proof in class: necessary and sufficiency
- Strict convexity: $\nabla^2 f(x) \succ 0$
- Concave functions: $\nabla^2 f(x) \preccurlyeq 0$

Example of Convex Functions

- Quadratic over linear function

$$f(x,y) = \frac{x^2}{y}, \text{for } y > 0$$

Its gradient $\nabla f(x) = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix} = \begin{bmatrix} \frac{2x}{y} \\ \frac{-x^2}{y^2} \end{bmatrix}$

Hessian $\nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix} = \frac{2}{y^3} \begin{bmatrix} y^2 & -xy \\ -xy & x^2 \end{bmatrix} \succcurlyeq 0 \Rightarrow$ convex

Positive semidefinite? $\begin{bmatrix} y \\ -x \end{bmatrix} \begin{bmatrix} y \\ -x \end{bmatrix}^{\mathrm{T}}$, for any $u \in R^2, u^T(vv^T)u = (v^T u)^T(v^T u) = ||v^T u||_2^2 \geq 0.$

Restriction of a convex function to a line

- A function $f: R^n \to R$ is convex if and only if the function $g: R \to R$,

$$g(t) = f(x + tv), \qquad \mathrm{dom}\ g = \{t |\ x + tv \in \mathrm{dom}\ f\}$$

  is a convex on its domain for $\forall x \in \mathrm{dom}\ f, v \in R^n$.

- The property can be useful to check the convexity of a function

Example: Prove $f(X) = \log\det X$, $\mathrm{dom}\ f = S_{++}^n$ is concave.

Restriction of a convex function to a line

Example: Prove $f(X) = \log\det X$, $\mathrm{dom}\ f = S_{++}^n$ is concave.

Consider an arbitrary line $X = Z + tV$, where $Z \in S_{++}^n, V \in S^n$. Define $g(t) = f(Z + tV)$ and restrict g to the interval values of t for $Z + tV \succ 0$. We have

$$g(t) = \log\det(Z + tV) = \log\det(Z^{\frac{1}{2}}(I + tZ^{-\frac{1}{2}}VZ^{-\frac{1}{2}})Z^{1/2})$$

$$= \sum_{i=1}^{n} \log(1 + t\lambda_i) + \log\det Z$$

- The property of eigenvalues $\det A = \prod_{i=1}^{n} \lambda_i$

where λ_i are the eigenvalues of $Z^{-\frac{1}{2}}VZ^{-\frac{1}{2}}$. So we have

$$g'(t) = \sum_{i=1}^{n} \frac{\lambda_i}{1 + t\lambda_i}, \qquad g''(t) = -\sum_{i=1}^{n} \frac{\lambda_i^2}{(1 + t\lambda_i)^2} \leq 0$$

$f(X)$ is concave. For more practice, see Exercise 3.18.

Extension of Jensen's Inequality

- The basic inequality can be extended to convex combinations of more than two points.

$$f(\theta_1 x_1 + \cdots + \theta_k x_k) \leq \theta_1 f(x_1) + \cdots + \theta_k f(x_k)$$

if f is convex $x_1, \ldots, x_k \in \mathrm{dom}\, f, \theta_1, \ldots, \theta_k \geq 0\ and\ \theta_1 + \cdots + \theta_k = 1.$

- Consider infinite sums and integral

$$f\left(\int_S p(x)x dx\right) \leq \int_S p(x) f(x) dx$$

where $p(x) \geq 0\ on\ S \subseteq dom\ f, \int_S p(x)dx = 1.$

➡ *Example with expectation from class* $f(\boldsymbol{E}x) \leq \boldsymbol{E}f(x)$

Operations that preserve convexity

Practical methods for establishing convexity of a function

- Verify definition (often simplified by restricting to a line)
- For twice differentiable functions, show $\nabla^2 f(x) \succcurlyeq 0$
- Show that f is obtained from simple convex functions by operations that preserve convexity (Ref. Chap. 3.2)
  - Nonnegative weighted sum
  - Composition with affine function
  - Pointwise maximum and supremum
  - Composition
  - Minimization
  - Perspective