PRINCETON UNIVERSITY
ORFE

Instructor:
Amir Ali Ahmadi
Fall 2016

TAs:
B. El Khadir,
G. Hall,
Z. Li,
K. Wang,
J. Ye,
J. Zhang

This lecture:

- More on convex optimization
- Minima of convex problems
- First and second order characterizations of convex functions
- Least squares revisited
- Strict convexity and uniqueness of optimal solutions
- Optimality condition for convex problems

Recall from the last lecture that a convex optimization problem is a problem of the form:

$$\begin{array}{lll} \text{min.} & f(x) & \\ \text{s.t.} & g_i(x) \leq 0 & i = 1, \dots, m \\ & h_i(x) = 0 & i = 1, \dots, k \end{array}$$

where

- Each h_i is affine: $h_i(x) = a_i{}^T x - b_i$

- $f, g_i, \dots, g_m$ are convex: $f(\lambda x + (1-\lambda)y \leq \lambda f(x) + (1-\lambda) f(y)$

  $\forall\, x, y \in dom(f), \forall\, \lambda \in [0,1]$

  Similarly for the $g_i's$.

- Today we start off by proving results that explain why we give special attention to convex optimization problems.
  - In a convex problem, every local minimum is automatically a global minimum. (This is true even for the more abstract definition of a convex optimization problem from the last lecture that only required the feasible set to be a convex set.)
  - In the unconstrained case, every stationary point (i.e., zero of the gradient) is automatically a global minimum.
- We will also see new characterizations for convex functions that make the task of checking convexity somewhat easier, though in general checking convexity can be a very difficult problem [AOPT13].

Let's recall the definition of local and global minima and generalize them to the constrained setting.

$$\begin{array}{lll} \text{min.} & f(x) & \\ \text{s.t.} & x \in \Omega & (e.g.,\ \Omega = \{x \mid g_i(x) \leq 0, h_j(x) = 0\}) \end{array}$$

Definition: A point $x^* \in \mathbb{R}^n$ is

- *feasible*, if $x^* \in \Omega$; $i.e., g_i(x^*) \leq 0, \forall i, h_j(x^*) = 0, \forall j$
- a *local minimum*, if feasible, and if $\exists\ \delta > 0$ s.t.

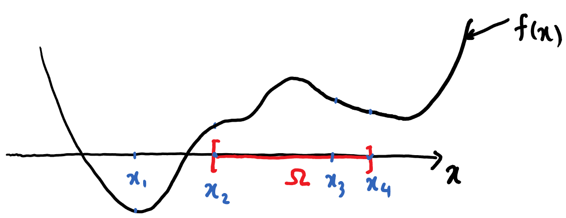
$$f(x^*) \leq f(x),\ \ \forall\ x\ \text{ s.t. } x \in \Omega \text{ and } ||x - x^*|| \leq \delta$$

- a *strict local minimum*, if feasible, and if $\exists\ \delta > 0$ s.t.

$$f(x^*) < f(x),\ \ \forall\ x \neq x^*\ \text{ s.t. } x \in \Omega\ \text{ and } ||x - x^*|| \leq \delta$$

- a *global minimum*, if feasible, and if $f(x^*) \leq f(x), \forall\ x \in \Omega$
- a *strict global minimum*, if feasible, and if $f(x^*) < f(x), \forall\ x \in \Omega, x \neq x^*$

- x_1: not feasible
- x_2: strict global minimum
- x_3: feasible
- x_4: strict local minimum



Our next few theorems show the nice features of convex problems in terms of inferring global properties from local ones.

Theorem. Consider an optimization problem

$$\begin{array}{ll} \text{min.} & f(x) \\ \text{s.t.} & x \in \Omega \end{array}$$

where f is a convex function and Ω is a convex set. Then, every local minimum is also a global minimum.

Proof.

Let x be a local minium. Suppose for the sake of contradiction that x is not a global minimum.

$$\Rightarrow \exists\, y \in \Omega, s.t.\ f(y) < f(x).$$

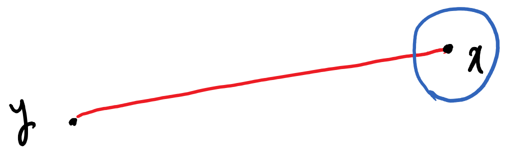
But $x \in \Omega, y \in \Omega,\ \Omega$ convex $\Rightarrow \lambda x + (1-\lambda)y \in \Omega, \forall \lambda \in [0,1]$

and f convex $\Rightarrow f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y)$

$$< \lambda f(x) + (1-\lambda)f(x) = f(x), \forall \lambda \in [0,1].$$

As $\lambda \to 1$, $(\lambda x + (1-\lambda)y) \to x$. So there are points arbitrarity close to x with a better objective value than x. This contradicts local optimality of x. □

Intuition:

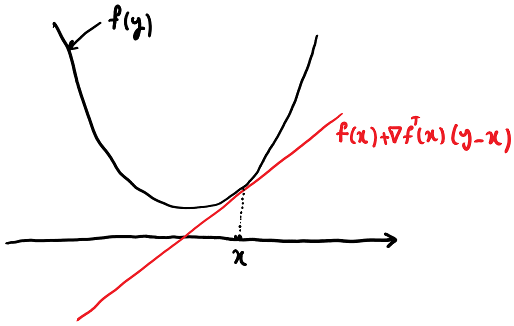


$$f(y) < f(x) \Rightarrow f(\textit{all line}) < f(x)$$

First and second order characterization of convex functions

Theorem. Suppose $f: \mathbb{R}^n \rightarrow \mathbb{R}$ is twice differentiable over its domain. Then, the following are equivalent:

(i) f is convex.
(ii) $f(y) \geq f(x) + \nabla f^T(x)(y-x), \qquad \forall\, x, y \in dom(f)$.
(iii) $\nabla^2 f(x) \succcurlyeq 0, \qquad \forall\, x \in dom(f)$ (i.e., the Hessian is psd $\forall x \in dom(f)$).



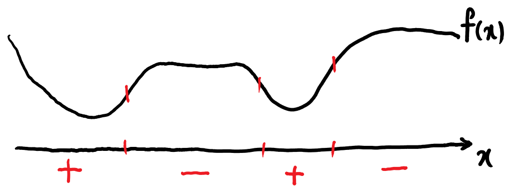
Interpretation:

(i) The first order Taylor expansion at any point is a global under estimator of the function

(ii) Function has *nonnegative curvature* everywhere:

"It curves up".

- In one dimension: $f''(x) \geq 0, \forall\, x \in dom(f)$



Sign of the curvature (sign of f'')

We prove $(i) \Leftrightarrow (ii)$. For $(ii) \Leftrightarrow (iii)$, see, e.g., Theorem 22.5 of [CZ13].

Proof: ([Tit13])

(i) $\Rightarrow$ (ii) If f convex, by definition

$$f(\lambda y + (1-\lambda)x) \leq \lambda f(y) + (1-\lambda) f(x), \quad \forall \lambda \in [0,1],\ x,y \in \text{dom}(f).$$

After rewriting, we have

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

$$\Rightarrow \quad f(y) - f(x) \geq \frac{f(x+\lambda(y-x)) - f(x)}{\lambda}, \quad \forall \lambda \in [0,1].$$

As $\lambda \searrow 0$, we get

$$f(y) - f(x) \geq \nabla f^T(x)(y-x). \qquad ①$$

(ii) $\Rightarrow$ (i) Suppose ① holds $\forall x,y \in \text{dom}(f)$.

Take any $x,y \in \text{dom}(f)$ and let $z = \lambda x + (1-\lambda) y$.

We have

$$f(x) \geq f(z) + \nabla f^T(z)(x-z) \qquad ②$$

$$f(y) \geq f(z) + \nabla f^T(z)(y-z). \qquad ③$$

Multiplying ② by λ, ③ by $(1-\lambda)$ and adding, we get

$$\lambda f(x) + (1-\lambda) f(y) \geq f(z) + \nabla f^T(z)\underbrace{(\lambda x + (1-\lambda)y - z)}_{0}$$

$$= f(z)$$

$$= f(\lambda x + (1-\lambda) y). \qquad \square$$

Corollary. Consider an unconstrained optimization problem:

$$\begin{aligned} &\min.\ f(x) \\ &\text{s.t. } x \in \mathbb{R}^n, \end{aligned}$$

where f is convex and differentiable. Then, any point $\bar{x}$ that satisfies $\nabla f(\bar{x}) = 0$, is a global minimum.

Proof. From the first order characterization of convexity

we have $\quad f(y) \geq f(x) + \nabla f^T(x)(y-x) \quad \forall x, y$

In particular, $\quad f(y) \geq f(\bar{x}) + \nabla f^T(\bar{x})(y-\bar{x}) \quad \forall\, y$

Since $\nabla f(\bar{x}) = 0$, we get

$$f(y) \geq f(\bar{x}) \qquad \forall y. \qquad \square$$

Remark 1. Recall that $\nabla f(x) = 0$ is always a necessary condition for local optimality in an unconstrained problem. The theorem says that for convex problems $\nabla f(x) = 0$ is not only necessary, but also sufficient for local *and global* optimality.

Remark 2. Recall that in absence of convexity, $\nabla f(x) = 0$ is not sufficient even for local optimality (e.g., think of $f(x) = x^3$ and $\bar{x} = 0$).

Remark 3. Recall that another necessary condition for (unconstrained) local optimality of a point x was: $\nabla^2 f(x) \succcurlyeq 0$.

- Note that a convex function automatically passes this test.

Quadratic functions revisited

- Let $f(x) = x^T Ax + bx + c$ (A symmetric)
- When is f convex?
  - Let's use the second order test:

    $\nabla^2 f(x) = 2A$

    $\Rightarrow f$ convex $\Leftrightarrow A \succcurlyeq 0$ (independent of b, c)
- Consider the unconstrained optimization problem

  $$\min_{x} \quad x^T Ax + bx + c$$

  - $A \not\succcurlyeq 0$ (f not convex) $\Rightarrow$ unbounded below (why?)

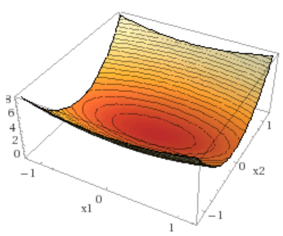
    Let $\bar{x}$ be an eigenvector, with a negative eigenvalue λ.

    $A\bar{x} = \lambda\bar{x} \Rightarrow \bar{x}^T A\bar{x} = \lambda\bar{x}^T\bar{x} < 0$.

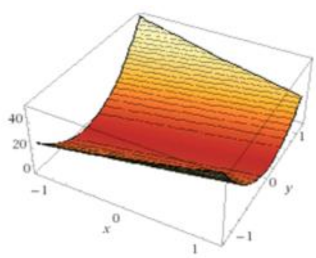
    $f(\alpha\bar{x}) = \alpha^2 \underbrace{\bar{x}^T A\bar{x}}_{<0} + \alpha b\bar{x} + c$

    $f(\alpha\bar{x}) \to -\infty$ as $\alpha \to \infty$.
  - $A \succ 0 \Rightarrow$ convex (in fact, $\Rightarrow f$ strictly convex as we see next)

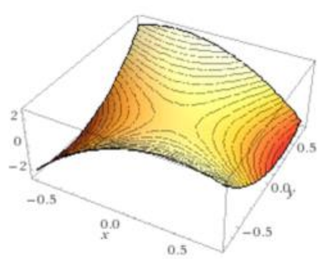
    Unique solution: $x^* = -\frac{1}{2}A^{-1}b$ (why?)
  - $A \succcurlyeq 0 \;\Rightarrow f$ convex. Optimal value may or may not be bounded, and there could be many optimal solutions.



$A \succ 0$



$A \succcurlyeq 0$
(but $A \not\succ 0$)



$A \not\succcurlyeq 0$

Least squares, revisited.

Given: A $m \times n$ matrix (Assume columns of A are linearly independent)

b $m \times 1$ vector

Solve: $\min_{x} \|Ax-b\|^2$

Let $f(x) = \|Ax-b\|^2 = (Ax-b)^T(Ax-b)$

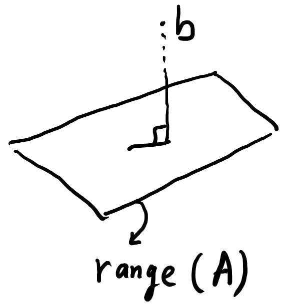
$= x^T A^T A x - 2x^T A^T b + b^T b.$

Strictly Convex! (b/c $A^TA \succ 0$)

$\nabla f(x) = 2A^TAx - 2A^Tb$

$\nabla f(x) = 0 \Rightarrow A^TAx = A^Tb$ ← Called "Normal Equations"

$\Rightarrow \boxed{x = (A^TA)^{-1}A^Tb}$



A^TA is invertible b/c its null space is just the origin:

$A^TAx = 0 \Rightarrow x^TA^TAx = 0 \Rightarrow (Ax)^T(Ax) = 0 \Rightarrow \|Ax\|^2 = 0 \Rightarrow Ax = 0$

$\Rightarrow x = 0.$

↑ Columns of A linearly independent.

$\nabla^2 f(x) = 2A^TA \succ 0$ (b/c $x^TA^TAx = \|Ax\|^2 \geq 0$ and $= 0 \Leftrightarrow x = 0$)

$\Rightarrow x = (A^TA)^{-1}A^Tb$ is a strict local minimum. □

This is the best conclusion we could make before without knowing that f is convex. Now that we know f is (strictly) convex, we immediately know that the solution $x = (A^TA)^{-1}A^Tb$ is a (strict) global minimum.

Characterization of Strict Convexity

Recall that we say $f: \mathbb{R}^n \to \mathbb{R}$ is strictly convex, if $\forall x, y, x \neq y, \forall \lambda \in (0,1)$,

$$f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda) f(y).$$

- f strictly convex $\Rightarrow$ f convex (obvious from the definition)
- f convex $\not\Rightarrow$ f strictly convex

  e.g., $f(x) = x \quad (x \in \mathbb{R})$

- Second order sufficient condition:

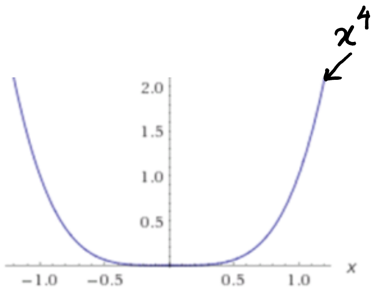
  $\nabla^2 f(x) \succ 0 \; \forall x \in \Omega \Rightarrow f$ strictly convex on Ω

- Converse not true:

  $f(x) = x^4 \quad (x \in \mathbb{R})$

  f is strictly convex (why?).

  But $f''(0) = 0 \quad$ (check)



- First order characterization:

  f strictly convex on $\Omega \subseteq \mathbb{R}^n$

  $\Leftrightarrow$

  $f(y) > f(x) + \nabla f^T(x)(y - x), \; \forall x, y \in \Omega, x \neq y$

- One of the main uses of strict convexity is to ensure uniqueness of optimal solutions. We see this next.

Strict Convexity and Uniqueness of Optimal Solutions

Theorem. Consider an optimization problem

$$\begin{aligned} &\text{min. } f(x) \\ &\text{s.t. } x \in \Omega, \end{aligned}$$

where $f: \mathbb{R}^n \to \mathbb{R}$ is *strictly convex* on Ω and Ω is a convex set. Then, the (optimal) solution is unique (assuming it exists).

Proof. Suppose there were two optimal solutions $x, y \in \mathbb{R}^n$. This means that $x, y \in \Omega$ and

$$f(x) = f(y) \leq f(z), \forall z \in \Omega \qquad (1)$$

But consider $z = \frac{x+y}{2}$. By convexity of Ω, we have $z \in \Omega$. By strict convexity,

we have $f(z) = f\left(\frac{x+y}{2}\right)$

$$< \frac{1}{2} f(x) + \frac{1}{2} f(y)$$

$$= \frac{1}{2} f(x) + \frac{1}{2} f(x) = f(x).$$

But this contradicts (1). □

Practice: for each function below, determine whether it is convex, strictly convex, or neither.

- $f(x) = (x_1 - 3x_2)^2$
- $f(x) = (x_1 - 3x_2)^2 + (x_1 - 2x_2)^2$
- $f(x) = (x_1 - 3x_2)^2 + (x_1 - 2x_2)^2 + x_1^3$
- $f(x) = |x| \ \ (x \in \mathbb{R})$
- $f(x) = ||x|| \ \ (x \in \mathbb{R}^n)$

An Optimality Condition for Convex Problems

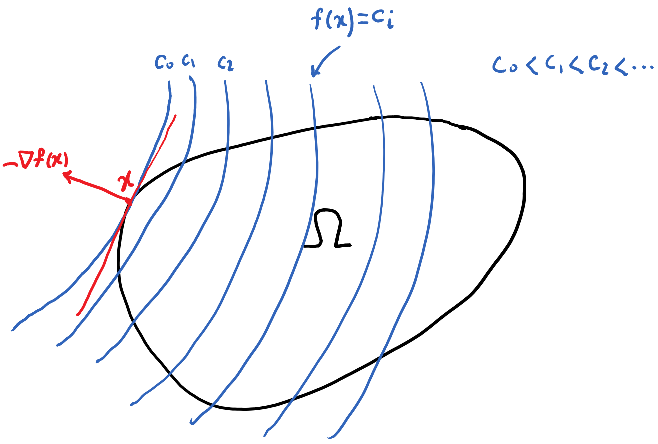
Theorem. Consider an optimization problem

$$\min f(x)$$
$$\text{s.t. } x \in \Omega,$$

where $f: \mathbb{R}^n \to \mathbb{R}$ is convex and differentiable and Ω is a convex. Then, a point x is optimal if and only if $x \in \Omega$ and

$$\nabla f(x)^T(y - x) \geq 0, \forall y \in \Omega.$$

- **What does this mean?**
  - If you move from x towards any feasible y, you will increase f locally.
  - $-\nabla f(x)$ (assuming it is nonzero) serves as a hyperplane that "supports" the feasible set Ω at x (see figure below).



- The necessity of the condition holds independent of convexity of f.
- Convexity is used in establishing sufficiency.
- If $\Omega = \mathbb{R}^n$, can you see how the condition above reduces to our first order unconstriained optimality condition $\nabla f(x) = 0$?
  - Hint: take $y = x - \nabla f(x)$.

Proof.

(Sufficiency)

Suppose $x \in \Omega$

satisfies $\nabla f^T(x)(y-x) \geq 0, \ \forall y \in \Omega.$ ①

By the first order characterization of convexity we have:

$$f(y) \geq f(x) + \nabla f^T(x)(y-x), \quad \forall y \in \Omega \quad ②$$

① + ② $\Rightarrow$ $f(y) \geq f(x), \ \forall y \in \Omega.$

$\Rightarrow$ x is optimal.

(necessity)

Suppose x is optimal, but for some $y \in \Omega$ we had

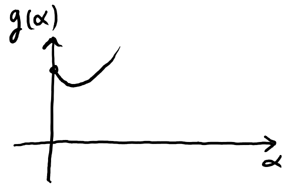
$$\nabla f^T(x)(y-x) < 0.$$

Consider $g(\alpha) := f(x + \alpha(y-x)).$

Because Ω is convex, $\forall \alpha, \ x + \alpha(y-x) \in \Omega.$

Observe that $g'(\alpha) = (y-x)^T \nabla f(x + \alpha(x-y)).$

$\Rightarrow$ $g'(0) = (y-x)^T \nabla f(x) < 0.$



$\Rightarrow \exists \ \delta > 0$ s.t. $g(\alpha) < g(0) \quad \forall \alpha \in (0, \delta).$

$\Rightarrow f(x + \alpha(y-x)) < f(x) \quad \forall \alpha \in (0, \delta).$

But this contradicts

optimality of x. □

Notes:

- Further reading for this lecture can include the first few pages of Chapters 2,3,4 of [BV04]. Your [CZ13] book defines convex sets in Section 4.3. Convex optimization appears in Chapter 22. The relevant sections are 22.1-22.3.

References:

- [AOPT13] A.A. Ahmadi, A. Olshevsky, P.A. Parrilo, and J.N. Tsitsiklis. NP-hardness of checking convexity of quartic forms and related problems. *Mathematical Programming,* 2013. http://web.mit.edu/~a_a_a/Public/Publications/convexity_nphard.pdf

- [BV04] S. Boyd and L. Vandenberghe. Convex Optimization. Cambridge University Press, 2004. http://stanford.edu/~boyd/cvxbook/

- [CZ13] E.K.P. Chong and S.H. Zak. An Introduction to Optimization. Fourth edition. Wiley, 2013.

- [Tit13] A.L. Tits. Lecture notes on optimal control. University of Maryland, 2013.