Convex Optimization Problem cont'

Lecture 6 for 18660/18460: Optimization

Guannan Qu

Jan 29, 2026

Admin Stuff

- Quiz for Lecture 6 released, due Feb 3 before lecture

Recall

Convex programs

$$\min_{x \in \mathbb{R}^n} f(x)$$

$$Ax = b$$

$$g_1(x) \le 0$$

$$\vdots$$

$$g_k(x) \le 0$$

Feasibility

$$\min_x f(x)$$

$$x \in C$$

feasible means C is non-empty

Recall

Optimality

$$\min_{x} f(x) \quad \text{s.t. } x \in C$$

x^* is optimal solution if

$f(x^*) \leq f(x), \ \forall x \in C$

$f(x^*)$ optimal value.

Optimality condition

Unconstrained: $\min_{x} f(x)$

x^* is optimal solution if and only if $\nabla f(x^*) = 0$

$f(x^* - t\nabla f(x^*)) < f(x^*)$ if $\nabla f(x^*) \neq 0$, for small t

First Order Optimality Condition

Theorem. For **unconstrained** convex optimization problem with differentiable objective $f(x)$, x^* is an optimal solution if and only if $\nabla f(x^*) = 0$.

Proof. $\Longrightarrow$: suppose x^* is an optimal solution, want to prove $\nabla f(x^*) = 0$

Assume $\nabla f(x^*) \neq 0$, we want to show $\exists \tilde{x}$ s.t. $f(\tilde{x}) < f(x^*)$

$g(t) = f(x^* - t \cdot \nabla f(x^*))$

$g'(t) = \nabla f(x^* - t \nabla f(x^*))^T \cdot [-\nabla f(x^*)]$

$g'(0) = \nabla f(x^*)^T [-\nabla f(x^*)] = -\|\nabla f(x^*)\|_2^2 < 0$

Must exist $\tilde{t} > 0$ s.t. $g(\tilde{t}) < g(0) \Leftrightarrow f(\underbrace{x^* - \tilde{t} \nabla f(x^*)}_{\tilde{x}}) < f(x^*)$

$\Rightarrow \nabla f(x^*) = 0$

First Order Optimality Condition

Theorem. For **unconstrained** convex optimization problem with differentiable objective $f(x)$, x^* is an optimal solution if and only if $\nabla f(x^*) = 0$.

Proof. $\Leftarrow$: suppose $\nabla f(x^*) = 0$, want to prove x^* is an optimal solution

By 1-st order Condition for convexity

$\forall x$. $\quad f(x) \geq f(x^*) + \underbrace{(\nabla f(x^*))}_{=0}{}^T (x - x^*)$

$= f(x^*)$

$\Rightarrow x^*$ is optimizer

First Order Optimality Condition

Theorem. For **unconstrained** convex optimization problem with differentiable objective $f(x)$, x^* is an optimal solution if and only if $\nabla f(x^*) = 0$.

Summary of key proof idea:

- $f\big(x^* - t\nabla f(x^*)\big) < f(x^*)$ for small t, non-zero $\nabla f(x^*)$
  - Negative gradient direction is a “descent” direction!

- $\nabla f(x^*) = 0$ eliminates the possibility of obtaining lower cost by “descent”.

First Order Optimality Condition: Examples

Quadratic cost functions:

$$\min_{x \in \mathbb{R}^n} f(x) = \frac{1}{2} x^\top A x + b^\top x + c$$

Calculate Gradient $\quad \nabla f(x) = Ax + b$

Optimal x^* Satisfies $\quad \nabla f(x^*) = Ax^* + b = 0$

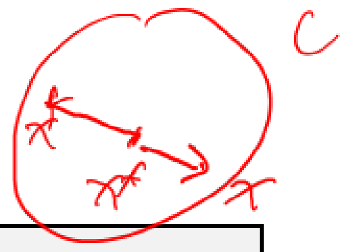
For general cost f, we can't solve $\nabla f(x^*) = 0$ directly. But the gradient is useful and form the basis of gradient based methods, which we will cover in lecture 9.

First Order Optimality Condition

What about constrained convex program?

Idea: the condition shall eliminate the possibility of obtaining lower cost by descent.

First Order Optimality Condition (Constrained Case)



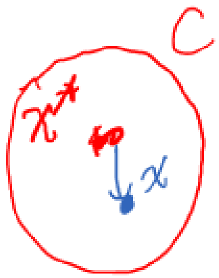
Theorem. Given a convex program $\min_{x \in C} f(x)$ with differentiable objective $f(x)$, x^* is an optimal solution if and only if

$$\nabla f(x^*)^{\mathsf{T}}(x - x^*) \geq 0, \forall x \in C$$

$\varepsilon > 0$

$$\left.\begin{array}{l} \varepsilon a \geq 0 \\ -\varepsilon a \geq 0 \end{array}\right\} \Rightarrow a = 0$$

If x^* lies in the interior of C, the condition is equivalent to $\nabla f(x^*) = 0$



choose x s.t. $x = x^* + \varepsilon [0 \cdots 0, 1, 0, \cdots, 0]^T$

ith entry

Since x^* is in interior, $\exists\, \varepsilon > 0$, s.t. $x \in C$

$$\left.\begin{array}{l} \nabla f(x^*)^T [x - x^*] = \varepsilon \cdot [\nabla f(x^*)]_i \geq 0 \\ \text{Do the same for } (-\varepsilon),\ (-\varepsilon)[\nabla f(x^*)]_i \geq 0 \end{array}\right\} \Rightarrow [\nabla f(x^*)]_i = 0$$

Do the same for all $i \Rightarrow \nabla f(x^*) = 0$

First Order Optimality Condition (Constrained Case)

$a^T b \geq 0$

Theorem. Given a convex program $\min_{x \in C} f(x)$ with differentiable objective $f(x)$, x^* is an optimal solution if and only if

$$\nabla f(x^*)^\top (x - x^*) \geq 0, \forall x \in C$$

If x^* lies on the boundary of C, the condition means gradient direction is "aligned" with the set C.

$-t\nabla f(x^*)$

outside constraint set

First Order Optimality Condition (Constrained Case)

Theorem. Given a convex program $\min_{x \in C} f(x)$ with differentiable objective $f(x)$, x^* is an optimal solution if and only if

$$\nabla f(x^*)^\top (x - x^*) \geq 0, \forall x \in C$$

Proof. $\Longrightarrow$: suppose x^* is an optimal solution, want to prove $\nabla f(x^*)^\top (x - x^*) \geq 0, \forall x \in C$

We use contradiction, Suppose $\exists x \in C$ s.t. $[\nabla f(x^*)]^\top (x - x^*) < 0$

Define $g(t) = f((1-t)x^* + tx)$

$g'(t) = (x - x^*)^\top \nabla f((1-t)x^* + tx)$

$g'(0) = (x - x^*)^\top \nabla f(x^*) < 0$

$\exists$ small $t^* \in [0,1)$ s.t. $g(t^*) < g(0)$

$f((1-t^*)x^* + t^* x) < f(x^*) \Rightarrow x^*$ can't be optimizer

$\Rightarrow$ Contradiction

First Order Optimality Condition (Constrained Case)

Theorem. Given a convex program $\min_{x \in C} f(x)$ with differentiable objective $f(x)$, x^* is an optimal solution if and only if

$$\nabla f(x^*)^\top (x - x^*) \geq 0, \forall x \in C$$

Proof. $\Leftarrow$: suppose $\nabla f(x^*)^\top (x - x^*) \geq 0, \forall x \in C$, we want to prove x^* is an optimizer

Apply first order condition for convexity

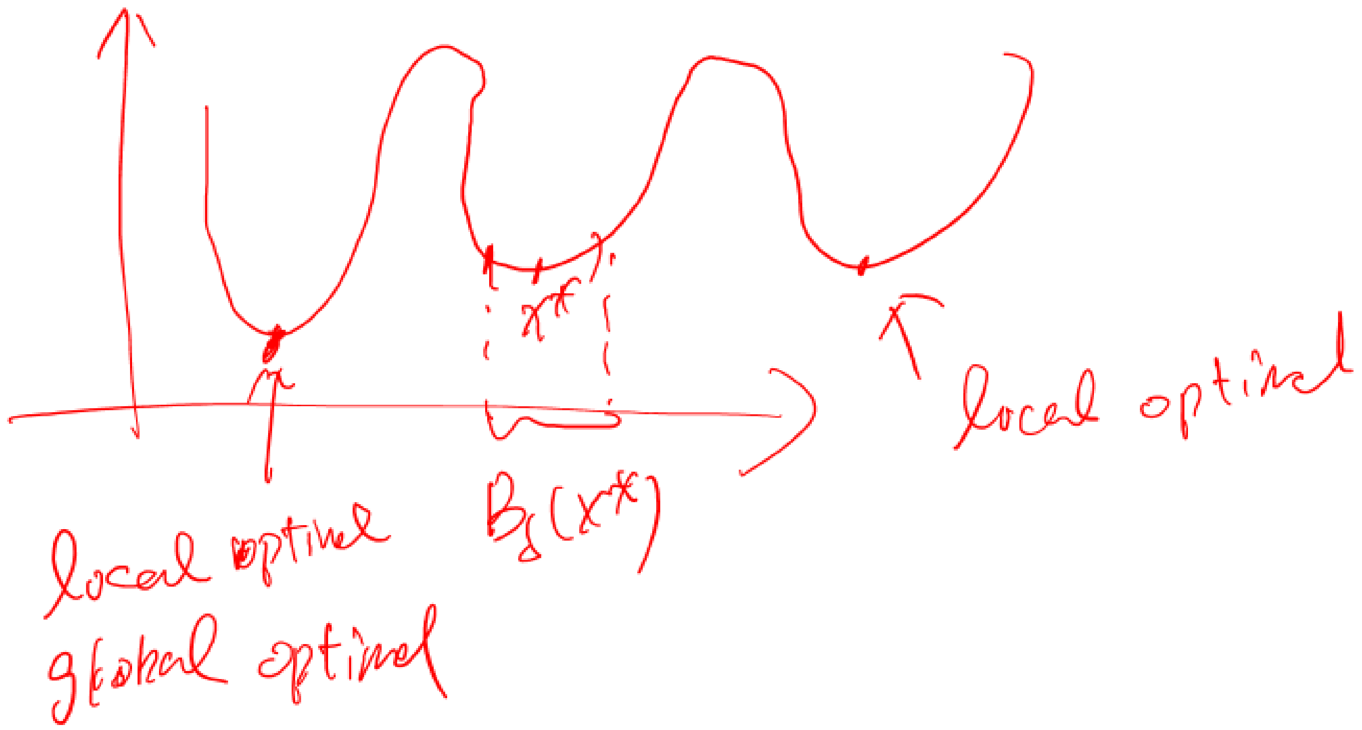
$$f(x) \geq f(x^*) + \underbrace{[\nabla f(x^*)]^\top (x - x^*)}_{\geq 0, \forall x \in C}$$

$$\geq f(x^*), \forall x \in C$$

Global Optimality vs Local Optimality

not necessarily convex

Definition. For optimization problem $\min_{x \in C} f(x)$, x^* is a local optimal solution if there exits $\delta > 0$ s.t. for any feasible point $x \in C \cap B_\delta(x^*)$ where $B_\delta(x^*) = \{x: \|x - x^*\| \leq \delta\}$, we have $f(x^*) \leq f(x)$.



Global Optimality vs Local Optimality

Theorem. For convex optimization, a local optimal solution is a global optimal solution.

Proof. Exercise in HW2.

Quiz Results

Select all that is true.

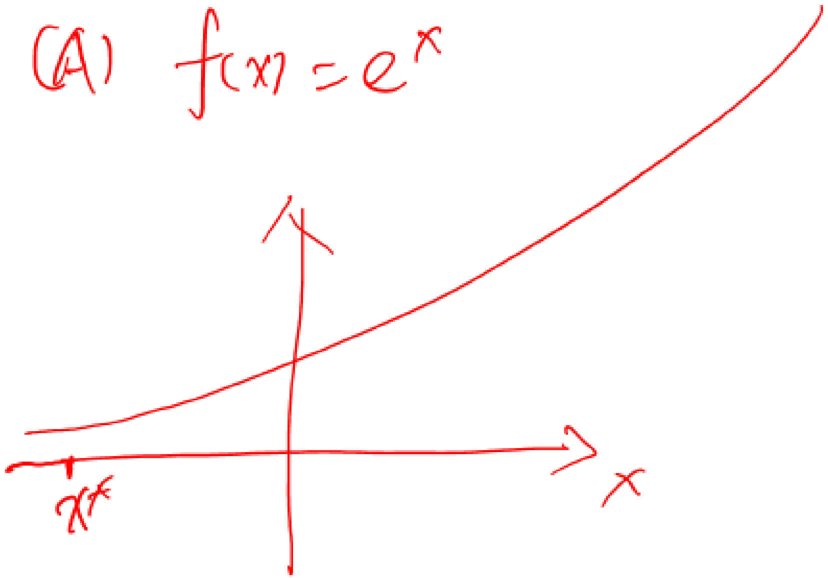
(A) For a convex optimization problem, if the cost function is lower bounded (i.e. $\exists c \in \mathbb{R}$ s.t. $f(x) \geq c, \forall x$), then an optimal solution must exist.

(B) For an unconstrained convex optimization problem with a differentiable objective $f(x)$, suppose x^* satisfies $\nabla f(x^*) = 0$, then x^* is an optimizer of the optimization problem.

(C) The following is a convex optimization problem

$$\min_{x_1, x_2 \in (0, +\infty)} x_1^2 x_2^3$$
$$\text{s.t. } x_1 x_2 + x_1 x_2^2 \leq 1$$

(A) $f(x) = e^x$



(C) $f(x) = x_1^2 x_2^3$

$$\nabla f(x) = \begin{bmatrix} 2x_1 x_2^3 \\ 3x_1^2 x_2^2 \end{bmatrix}$$

$$\nabla^2 f(x) = \begin{bmatrix} 2x_2^3 & 6x_1 x_2^2 \\ 6x_1 x_2^2 & 6x_1^2 x_2 \end{bmatrix}$$

$x_1 = x_2 = 1$

$$\nabla^2 f(x) = \begin{bmatrix} 2 & 6 \\ 6 & 6 \end{bmatrix}$$

NOT PSD

Canonical Forms and Equivalent Transforms

- Linear program
- Quadratic program
- Second-order cone program (Lecture 7)
- Semi-definite program (Lecture 7)

Note: these are “classical” canonical forms of convex optimization and are by no means exhaustive. Many optimization problems do not fall under these forms.

- Equivalent transforms

Linear Program

Both the cost and the constraints are affine

affine objective

$$\min_{x \in \mathbb{R}^n} c^\top x + d$$

$$\text{s.t.} \quad Gx \leq h \quad \in \text{affine ineq.}$$

$$Ax = b \quad \in \text{affine eq.}$$

polyhedron

First introduced by **Kantorovich** in the late 1930s and **Dantzig** in the 1940s

Example: Diets

Grad student diet problem: choose quantities $x_1, \dots, x_n$ of n foods

- one unit of food j costs c_j, contains amount a_{ij} of nutrient i
- healthy diet requires nutrient i in quantity at least b_i

Goal: find cheapest healthy diet.

$$\min_{x_1, \dots, x_n} \; x_1 c_1 + x_2 c_2 + \cdots + x_n c_n \Leftarrow \text{affine}$$

$$\left.\begin{array}{l} x_j \geq 0 \quad \forall j = 1, \dots, n \\ \underbrace{\text{Amount of Nutrient } i}_{a_{i1} x_1 + a_{i2} x_2 + \cdots + a_{in} x_n} \geq b_i \end{array}\right\} \text{affine.}$$

Quadratic Program

- Quadratic objective with affine constraints

P needs to be psd.

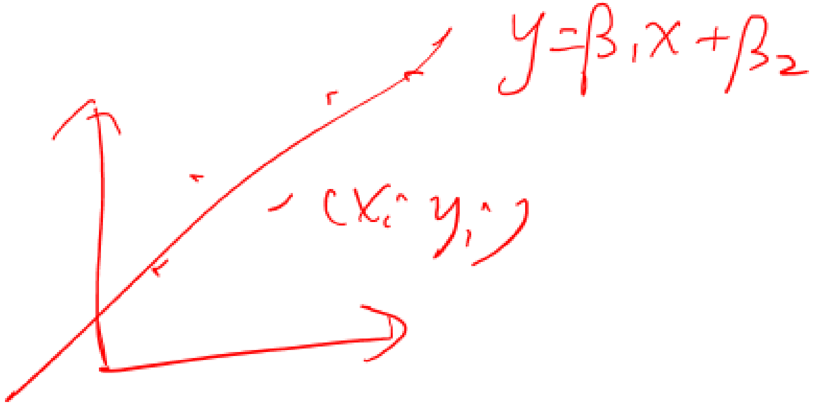
$$\min_x \frac{1}{2} x^\top P x + c^\top x + d$$

$$\text{s.t.} \quad \left.\begin{aligned} Gx &\le h \\ Ax &= b \end{aligned}\right\} \text{affine}$$

Example

- Least squares

$$\min_{\beta_1 \beta_2} \sum_{i=1}^{n} (y_i - \beta_1 x_i - \beta_2)^2$$



Example: LP with random cost

- LP with cost as $c^\top x + d$ but c, d are independent random variables
- Mean of the cost $E(c^T x + d) = (\underbrace{Ec}_{\bar{c}})^T x + (\underbrace{Ed}_{\bar{d}})$
- Variance of cost

$$\mathrm{Var}(c^T x + d) = \mathrm{Var}(c^T x) + \mathrm{Var}(d)$$

$$= E(c^T x - \bar{c}^T x)^2 + \mathrm{Var}(d)$$

$$= E((c-\bar{c})^T x)^2 + \mathrm{Var}(d)$$

$$= x^T (\underbrace{E(c-\bar{c})(c-\bar{c})^T}_{\Sigma}) x + \mathrm{Var}(d)$$

$$= x^T \Sigma x + \mathrm{Var}(d)$$

$$\Sigma = E(c-\bar{c})(c-\bar{c})^T = \mathrm{Cov}(c)$$

Example: LP with random cost

- LP with cost as $c^\top x + d$ but c, d are independent random variables
- Mean of the cost
- Variance of cost

$$\min_{x\in\mathbb{R}^n} \mathrm{Mean}(c^\top x + d) + \gamma \mathrm{Var}(c^\top x + d)$$

$$\text{s.t.} \quad Gx \le h$$
$$Ax = b$$

$$\min_{x\in\mathbb{R}^n} \bar{c}^\top x + \bar{d} + \gamma x^\top \Sigma x + \gamma \mathrm{Var}(d)$$

$$\text{s.t.} \quad Gx \le h$$
$$Ax = b$$

LP is a subclass of QP

LP

$$\min_{x \in \mathbb{R}^n} c^\top x + d$$
$$\text{s.t.} \quad Gx \leq h$$
$$Ax = b$$

$\subseteq$

QP

$P = 0$

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} x^\top P x + c^\top x + d$$
$$\text{s.t.} \quad Gx \leq h$$
$$Ax = b$$

Summary so far:

- Linear program
- Quadratic program

Next: equivalent transforms

Equivalent Transforms

Two problems are (informally) equivalent if the solution of one is readily obtained from the solution of the other, and vice-versa.

Equivalent transformations are useful because they can allow us to:

- Transform optimization problems to more “familiar” forms
- Provide new perspectives that were not obvious from the original formulation

Epigraph Forms

$$\min_{x\in\mathbb{R}^n} f(x)$$
$$\text{s.t.}\quad \left.\begin{aligned} Gx &\le h \\ Ax &= b \end{aligned}\right\}\ \text{affine}$$

where $\underbrace{f(x) = \max(c_1^\top x + d_1, c_2^\top x + d_2)}_{}$ convex

We know this is a convex program. But is this a linear program?

Epigraph Forms

(A)

$$\min_{x\in\mathbb{R}^n} f(x)$$
$$\text{s.t.} \quad x \in C$$

(B)

$$\min_{x\in\mathbb{R}^n, t\in\mathbb{R}} t$$
$$\text{s.t.} \quad x \in C$$
$$f(x) \le t$$

Claim: if x^* is solution to (A), then $(x^*, \underbrace{f(x^*)}_{t^*})$ solves (B)

Pf: Assume (x^*, t^*) is NOT solution to (B)

Then there $\exists$ (x', t') s.t. $x' \in C$, AND $t' < t^*$

$f(x') \le t' < t^* = f(x^*)$ We find a better solution to (A) than x^* Contradiction!

Claim if (x^*, t^*) solves (B) $\Rightarrow$ x^* solves (A)

Epigraph Forms

$$f(x) = \max(c_1^\top x + d_1, c_2^\top x + d_2)$$

$$\begin{aligned} \min_{x\in\mathbb{R}^n} \ & f(x) \\ \text{s.t.} \ & Gx \le h \\ & Ax = b \end{aligned}$$

$$\begin{aligned} \min_{x\in\mathbb{R}^n, t\in\mathbb{R}} \ & t \\ \text{s.t.} \ & Gx \le h \\ & Ax = b \\ & \underline{f(x)} \le t \end{aligned}$$

$\max(c_1^\top x + d_1, c_2^\top x + d_2) \le t$

$$\begin{aligned} \min_{x\in\mathbb{R}^n, t\in\mathbb{R}} \ & t \\ \text{s.t.} \ & Gx \le h \\ & Ax = b \\ & c_1^\top x + d_1 \le t \\ & c_2^\top x + d_2 \le t \end{aligned}$$

Eliminating equality constraints

$$\min_x f(x)$$

$$g_i(x) \leq 0, i \in 1,2, \dots, k$$

$$Ax + b = 0$$

$$\min_z f(Fz + x_0)$$

$$g_i(Fz + x_0) \leq 0, i \in 1,2, \dots, k$$

where columns of F forms null space of A

Change of Variables

If $\phi: \mathbb{R}^n \to \mathbb{R}^n$ is one-to-one, and its image covers feasible set C, then we can change variables in an optimization problem:

$x = \phi(z)$

$$\min_{x \in C} f(x) \quad \Leftrightarrow \quad \min_{z \in \phi^{-1}(C)} f(\phi(z))$$

- Could turn non-convex problems into convex problems!

$a = [2, 3]$

$x_1 = e^{y_1} \quad x_2 = e^{y_2}$

$$\min_{x_1, x_2 \in (0, +\infty)} x_1^2 x_2^3$$
$$\text{s.t. } x_1 x_2 + x_1 x_2^2 \leq 1$$

Geometric Programing

$\min_{y_1 y_2} e^{2y_1 + 3y_2} \longrightarrow e^{a^T y}$ convex

$e^{y_1 + y_2} + e^{y_1 + 2y_2} \leq 1$

convex

Summary so far

Canonical forms

- Linear program
- Quadratic program

Equivalent transforms

- Epigraph forms, eliminating equality constraints, change of variables

Next lecture:

- Second-order cone programming
- Semi-definite programming
- Examples of optimization