Lecture 4. Convex Optimization and Duality

Ivan Papusha

CDS270–2: Mathematical Methods in Control and System Engineering

April 20, 2015

Logistics

- hw3 due this **Wed, Apr 22**
  - do an easy problem or CYOA
- hw2 solutions posted online
- **Wed lecture only: 2–2:55pm (243 Annenberg)**
- continue reading: lmibook, Ch 1–2

Convexity

convex set. a set $\mathcal{C}$ is convex if $x, y \in \mathcal{C}$ implies

$$\theta x + (1-\theta)y \in \mathcal{C}$$

for all $\theta \in [0, 1]$.

convex function. a function $f : \mathbf{R}^n \to \mathbf{R}$ is convex if its epigraph

$$\mathbf{epi}(f) = \{(x, t) \mid x \in \mathbf{dom}(f), f(x) \leq t\} \subseteq \mathbf{R}^n \times \mathbf{R}$$

is a convex set, or equivalently if

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

for all $\theta \in [0, 1]$ and $x, y \in \mathbf{dom}(f)$.

concave function. g is concave if $-g$ is convex.

Why convexity?

Given a (proper) convex function $f : \mathbf{R}^n \to \mathbf{R}$

- for every $x \in \mathbf{dom}(f)$, there exists a subgradient $g \in \mathbf{R}^n$, which defines a global affine underestimator of f at x,

$$f(y) \geq f(x) + g^T(y - x), \quad \text{for all } y \in \mathbf{R}^n$$

- every local minimum is a global minimum (effective algorithms)
- calculus of convex functions

Composition rule

Define the composition

$$f(x) = h(g_1(x), g_2(x), \ldots, g_k(x)),$$

where $h : \mathbf{R}^k \to \mathbf{R}$ is convex, and $g_i : \mathbf{R}^n \to \mathbf{R}$. Suppose that for each i, one of the following holds:

- h is nondecreasing in the ith argument, and g_i is convex
- h is nonincreasing in the ith argument, and g_i is concave
- g_i is affine

Then the function f is convex.

Example

$$f(x, y) = \left\| \begin{bmatrix} x + y \\ y \end{bmatrix} \right\|_2 + \frac{(x-2)^2}{y}$$

- $+ : \mathbf{R} \times \mathbf{R} \to \mathbf{R}$ is nondecreasing in both arguments (affine)
- $\| \cdot \|_2 : \mathbf{R}^n \to \mathbf{R}$ is nondecreasing in all arguments (convex)
- $g(z_1, z_2) = z_1^2 / z_2$ is convex in (z_1, z_2) for $z_2 > 0$, and nonincreasing in z_2

f is convex over $(x, y) \in \mathbf{R} \times \mathbf{R}_{++}$

Standard form convex problem

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & h_i(x) = 0, \quad i = 1, \ldots, p \end{array}$$

- variable is $x \in \mathbf{R}^n$
- domain of definition is $\mathcal{D} = \bigcap_{i=0}^m \mathbf{dom}(f_i) \cap \bigcap_{i=1}^p \mathbf{dom}(h_i) \subseteq \mathbf{R}^n$
- f_0 is objective
- if $f_0(x) \equiv 0$, then problem is a *feasibility* problem
- $f_i : \mathbf{R}^n \to \mathbf{R}$ are convex, $i = 0, \ldots, m$
- $h_i : \mathbf{R}^n \to \mathbf{R}$ are affine, $i = 1, \ldots, p$
- $x^\star$ is *an* optimizing point (if it exists)
- optimal (primal) value is

$$p^\star \triangleq \begin{cases} f(x^\star), & \text{if feasible and } x^\star \text{ exists} \\ \infty, & \text{otherwise} \end{cases}$$

Lagrangian

$$L(x, \lambda, \nu) \triangleq f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x)$$

- real-valued function defined for
  - $x \in \mathcal{D}$
  - $\lambda_i \geq 0$, $i = 1, \ldots, m$
  - $\nu_i \in \mathbf{R}$, $i = 1, \ldots, p$
- under-approximation property: if x is feasible, and $\lambda_i \geq 0$,

$$\begin{aligned} L(x, \lambda, \nu) &= f_0(x) + \underbrace{\sum_{i=1}^{m} \lambda_i f_i(x)}_{\leq 0} + \underbrace{\sum_{i=1}^{p} \nu_i h_i(x)}_{=0} \\ &\leq f_0(x) \end{aligned}$$

Dual function

$$\begin{aligned} g(\lambda, \nu) &\overset{\Delta}{=} \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) \\ &= \inf_{x \in \mathcal{D}} \left\{ f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x) \right\} \\ &\leq \inf_{x \in \mathcal{D}} f_0(x) \\ &= p^\star \end{aligned}$$

- dual function is lower bound on optimal value
- best (largest) lower bound:

$$d^\star \overset{\Delta}{=} \sup_{\lambda \succeq 0, \nu \in \mathbf{R}^p} g(\lambda, \nu)$$

Primal and dual problems

primal:

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \dots, m \\ & h_i(x) = 0, \quad i = 1, \dots, p \end{array}$$

dual:

$$\begin{array}{ll} \text{maximize} & g(\lambda, \nu) \\ \text{subject to} & \lambda_i \geq 0, \quad i = 1, \dots, m \end{array}$$

- weak duality: $d^\star \leq p^\star$ always obtains

$$\begin{aligned} g(\lambda, \nu) &\leq L(x, \lambda, \nu), \qquad \text{for all } x \in \mathcal{D} \\ &\leq L(x^\star, \lambda, \nu) \\ &\leq f_0(x^\star) = p^\star \qquad \text{(then take supremum)} \end{aligned}$$

- strong duality: $d^\star = p^\star$ holds with a *constraint qualification*
- *Slater's condition:* suppose primal problem is convex, and $f_1, \dots, f_k$ are affine, then strong duality holds if there exists an x

$$\begin{aligned} f_i(x) &\leq 0, \quad i = 1, \dots, k \\ f_i(x) &< 0, \quad i = k+1, \dots, m \\ h_i(x) &= 0, \quad i = 1, \dots, p \end{aligned}$$

KKT conditions

Let $f_0, \dots, f_m$, $h_1, \dots, h_p$ be differentiable, $x^\star$ and $(\lambda^\star, \nu^\star)$ be any primal and dual optimal points, $p^\star = d^\star$, then these points necessarily satisfy

1. primal feasibility:

$$f_i(x^\star) \le 0, \quad i = 1, \dots, m$$
$$h_i(x^\star) = 0, \quad i = 1, \dots, p$$

2. dual feasibility:

$$\lambda_i^\star \ge 0, \quad i = 1, \dots, m$$

3. complementary slackness:

$$\lambda_i^\star f_i(x^\star) = 0, \quad i = 1, \dots, m$$

4. stationarity of Lagrangian:

$$\nabla f_0(x^\star) + \sum_{i=1}^{m} \lambda_i^\star \nabla f_i(x^\star) + \sum_{i=1}^{p} \nu_i^\star \nabla h_i(x^\star) = 0$$

KKT results

$$\left.\begin{cases} f_i, h_i \text{ differentiable} \\ x^\star, (\lambda^\star, \nu^\star) \text{ primal-dual optimal} \\ p^\star = d^\star \end{cases}\right\} \Longrightarrow \text{KKT holds}$$

$$\left.\begin{cases} f_i, h_i \text{ differentiable} \\ f_i \text{ convex}, h_i \text{ affine} \\ x^\star, (\lambda^\star, \nu^\star) \text{ satsifies KKT} \\ \text{Slater's condition holds} \end{cases}\right\} \Longrightarrow \left\{\begin{array}{l} x^\star, (\lambda^\star, \nu^\star) \text{ primal-dual optimal} \\ p^\star = d^\star \\ \text{dual optimum attained} \end{array}\right\}$$

for (much) more, see R. T. Rockafellar *Convex Analysis*

Valid convex optimization problems

objective.

- minimize $\{$ convex function $\}$
- maximize $\{$ concave function $\}$

constraints.

- $\{$ convex function $\} \leq \{$ concave function $\}$
- $\{$ concave function $\} \geq \{$ convex function $\}$
- $\{$ affine function $\} = \{$ affine function $\}$

Example: linear program (LP)

The standard form LP is

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b, \quad x \succeq 0 \end{array}$$

with variable $x \in \mathbf{R}^n$

- Lagrangian:

$$\begin{aligned} L(x, \lambda, \nu) &= c^T x - \lambda^T x + \nu^T (Ax - b) \\ &= (A^T \nu + c - \lambda)^T x - \nu^T b \end{aligned}$$

- dual function:

$$\begin{aligned} g(\lambda, \nu) &= \inf_x \left\{ (A^T \nu + c - \lambda)^T x - \nu^T b \right\} \\ &= \begin{cases} -\nu^T b & \text{if } A^T \nu + c - \lambda = 0 \\ -\infty & \text{otherwise} \end{cases} \end{aligned}$$

- related dual problem:

$$\begin{array}{ll} \text{maximize} & -\nu^T b \\ \text{subject to} & A^T \nu + c \succeq 0 \end{array}$$

Example: quadratic program (QP) w/ equality constraints

Consider the quadratic program

$$\begin{array}{ll} \text{minimize} & x^T P x + c^T x \\ \text{subject to} & Ax = b \end{array}$$

with $x \in \mathbf{R}^n$ a variable, and $P = P^T \succeq 0$

- if $b \notin \textbf{range}(A)$, then primal is infeasible

$$\begin{aligned} L(x, \nu) &= x^T P x + c^T x + \nu^T (Ax - b) \\ &= x^T P x + (A^T \nu + c)^T x - \nu^T b \end{aligned}$$

- taking a gradients gives necessary conditions for optimality

$$\nabla L(x^\star, \nu^\star) = (P + P^T) x^\star + (A^T \nu^\star + c) = 0$$

- optimal primal and dual variables are solutions (when they exist) to

$$\begin{bmatrix} P + P^T & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x^\star \\ \nu^\star \end{bmatrix} = \begin{bmatrix} -c \\ b \end{bmatrix}$$

Example: open loop control of a vehicle network

$$
\begin{array}{ll}
\text{minimize} & \sum_{k=0}^{T}\sum_{i=1}^{N} \|u^{(i)}(k)\|_2^2 + \mu \sum_{k=0}^{T}\sum_{i,j=1}^{N} \|x^{(i)}(k) - x^{(j)}(k) - r_{ij}\|_2^2 \\
\text{subject to} & x^{(i)}(k+1) = A^{(i)}x^{(i)}(k) + B^{(i)}u^{(i)}(k) + c^{(i)}(k), \\
& \qquad i = 1,\ldots,N, \quad k = 0,\ldots,T-1, \\
& x^{(i)}(0) = z^{(i)}, \quad i = 1,\ldots,N \\
& \frac{1}{N}\sum_{i=1}^{N} x^{(i)}(T) = w
\end{array}
$$

- N vehicles, each with state $x^{(i)} \in \mathbf{R}^2$ and input $u^{(i)} \in \mathbf{R}$
- minimize total fuel
- penalize deviation from prescribed geometry $r_{ij} \in \mathbf{R}^2$
- each vehicle obeys discrete-time affine dynamics

$$
x^{(i)}(k+1) = A^{(i)}x^{(i)}(k) + B^{(i)}u^{(i)}(k) + c^{(i)}(k)
$$

- initial condition $z^{(i)} \in \mathbf{R}^2$, final average position $w \in \mathbf{R}^2$

Example: geometric program (GP)

posynomial. a function $f : \mathbf{R}^n \to \mathbf{R}$ of the form

$$f(x_1, \ldots, x_n) = \sum_{k=1}^{t} c_k x_1^{a_{1k}} x_2^{a_{2k}} \cdots x_n^{a_{nk}},$$

where $c_j \geq 0$ and $a_{ij} \in \mathbf{R}$, *e.g.*, $0.7 + 2x_1/x_3^2 + x_2^{0.3}$

monomial. a posynomial with one term ($t = 1$), *e.g.*, $2.3(x_1/x_2)^{0.5}$

geometric program. an optimization problem of the form

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 1, \quad i = 1, \ldots, m \\ & g_i(x) = 1, \quad i = 1, \ldots, p \\ & x_i > 0, \quad i = 1, \ldots, n \end{array}$$

where f_i are posynomials and g_i are monomials

Exponential form of GP

Standard form GP

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 1, \quad i = 1, \ldots, m \\ & g_i(x) = 1, \quad i = 1, \ldots, p \\ & x_i > 0, \quad i = 1, \ldots, n \end{array}$$

with variable $x \in \mathbf{R}^n$

- define new variables $y_i = \log x_i$ and $b_k = \log c_k$

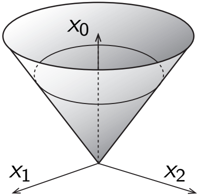
$$h(y) = \log\left(f\left(e^{y_1}, \ldots, e^{y_n}\right)\right) = \log\left(\sum_{k=1}^{t} e^{a_k^T y + b_k}\right), \quad a_k = \begin{bmatrix} a_{1k} \\ \vdots \\ a_{nk} \end{bmatrix}$$

- equivalent convex formulation in variable $y \in \mathbf{R}^n$

$$\begin{array}{ll} \text{minimize} & \log(f_0(e^{y_1}, \ldots, e^{y_n})) \\ \text{subject to} & \log(f_i(e^{y_1}, \ldots, e^{y_n})) \leq 0, \quad i = 1, \ldots, m \\ & \log(g_i(e^{y_1}, \ldots, e^{y_n})) = 0, \quad i = 1, \ldots, p \end{array}$$

Example: second order cone program (SOCP)

second order cone. a subset $\mathcal{Q}^n$ of $\mathbf{R}^n$ given by



$$\mathcal{Q}^n = \{(x_0, x_1) \in \mathbf{R} \times \mathbf{R}^{n-1} \mid \|x_1\|_2 \leq x_0\}$$

second order cone program. an optimization problem of the form

$$\begin{array}{ll} \text{minimize} & f^T x \\ \text{subject to} & \|A_i x + b_i\|_2 \leq c_i^T x + d_i, \quad i = 1, \ldots, m \end{array}$$

- a point x is feasible if and only if

$$\begin{bmatrix} c_i^T \\ A_i^T \end{bmatrix} x + \begin{bmatrix} d_i \\ b_i \end{bmatrix} \in \mathcal{Q}^{n_i} \qquad (n_i = 1 + \text{number of rows of } A_i)$$

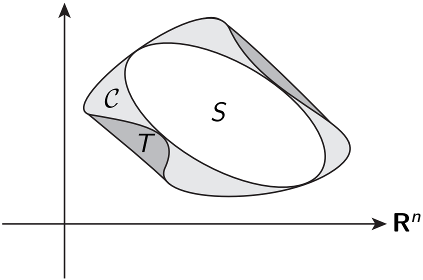
Relaxation and restriction

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & x \in \mathcal{C} \end{array}$$

suppose $S \subseteq \mathcal{C} \subseteq T$, then

$$\inf_{x \in T} f(x) \leq \inf_{x \in \mathcal{C}} f(x) \leq \inf_{x \in S} f(x)$$

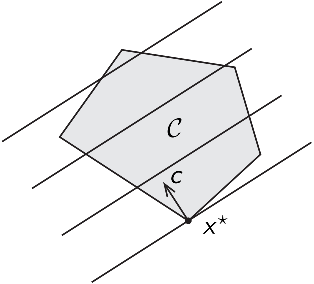
- $\mathcal{C}$ is original feasible set
- T is *relaxation*
- S is *restriction*



common practice if $\mathcal{C}$ is not convex
or is too complicated to describe

Affine function over convex set

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & x \in \mathcal{C} \end{array}$$



- if $\mathcal{C}$ is compact and objective is affine, an optimal point exists on the boundary $\partial\mathcal{C}$ of the feasible set
- also works for *maximization* of a *convex* objective
- LP: $\mathcal{C}$ is a polyhedron, $x^\star$ can be a vertex
- SDP: $\mathcal{C}$ is a slice of $\mathbf{S}_+^n$

main idea behind *many* practical algorithms (simplex. . .)

Conic optimization

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & \mathcal{A}(x) = b \\ & x \in \mathcal{K} \end{array}$$

- generalization of LP, SOCP, SDP
- variable is $x \in \mathbf{R}^n$
- f_0 is convex objective, often affine
- domain is a convex cone $\mathcal{K}$
- affine constraints $\mathcal{A} : \mathbf{R}^n \to \mathcal{K}$

Epigraph trick

- can arrange for objective to be linear by introducing an extra variable

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & x \in \mathcal{C} \end{array}$$

$$\Downarrow$$

$$\begin{array}{ll} \text{minimize} & \gamma \\ \text{subject to} & f(x) \leq \gamma \\ & x \in \mathcal{C} \end{array}$$

- new variable is (x, γ)
- if **epi**(f) and $\mathcal{C}$ are cone representable, the result is a conic program

Cone representations: SDP

sets. a convex set $\mathcal{C} \subseteq \mathbf{R}^n$ is SDP representable if there exists an affine mapping $\mathcal{A} : \mathbf{R}^n \times \mathbf{R}^m \to \mathbf{S}^p$ such that

$$x \in \mathcal{C} \quad \Longleftrightarrow \quad \exists u \in \mathbf{R}^m,\ \mathcal{A}(x, u) \succeq 0.$$

functions. a convex function $f : \mathbf{R}^n \to \mathbf{R}$ is SDP representable if its epigraph $\mathbf{epi}(f)$ is an SDP representable set.

Examples: SDP representable functions

- *absolute value:* $f(x) = |x|$

$$|x| \leq t \quad \Longleftrightarrow \quad \begin{bmatrix} x+t & 0 \\ 0 & -x+t \end{bmatrix} \succeq 0$$

- *euclidean norm:* $f(x) = \|x\|_2$

$$\|x\|_2 \leq t \quad \Longleftrightarrow \quad t^2 - x^T x \geq 0 \quad \Longleftrightarrow \quad \begin{bmatrix} t & x^T \\ x & tI \end{bmatrix} \succeq 0$$

- *largest eigenvalue:* $f(X) = \lambda_{\max}(X)$

$$\lambda_{\max}(X) \leq t \quad \Longleftrightarrow \quad tI - X \succeq 0$$

- *sum of k largest eigenvalues*: $f(X) = \lambda_1(X) + \cdots + \lambda_k(X)$

$$f(X) \leq t \quad \Longleftrightarrow \quad \exists Z = Z^T \text{ and } s \in \mathbf{R} \text{ with } \left\{ \begin{array}{r} t - ks - \mathbf{Tr}(Z) \geq 0 \\ Z \succeq 0 \\ Z - X + sI \succeq 0 \end{array} \right\}$$