Lecture: Duality

http://bicmr.pku.edu.cn/~wenzw/opt-2016-fall.html

Acknowledgement: this slides is based on Prof. Lieven Vandenberghe's lecture notes

Introduction

- Lagrange dual problem
- weak and strong duality
- geometric interpretation
- optimality conditions
- perturbation and sensitivity analysis
- examples
- generalized inequalities

Lagrangian

standard form problem (not necessarily convex)

$$\begin{array}{ll} \min & f_0(x) \\ \text{s.t.} & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & h_i(x) = 0, \quad i = 1, \ldots, p \end{array}$$

variable $x \in \mathbb{R}^n$, domain $\mathcal{D}$, optimal value p^*

Lagrangian: $L : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$, with $\mathrm{dom}\, L = \mathcal{D} \times \mathbb{R}^m \times \mathbb{R}^p$,

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

- weighted sum of objective and constraint functions
- λ_i is Lagrange multiplier associated with $f_i(x) \leq 0$
- ν_i is Lagrange multiplier associated with $h_i(x) = 0$

Lagrange dual function

Lagrange dual function: $g : \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$,

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

g is concave, can be $-\infty$ for some λ, ν

lower bound property: if $\lambda \succeq 0$, then $g(\lambda, \nu) \leq p^*$

proof: if $\tilde{x}$ is feasible and $\lambda \succeq 0$, then

$$f_0(\tilde{x}) \geq L(\tilde{x}, \lambda, \nu) \geq \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) = g(\lambda, \nu)$$

minimizing over all feasible $\tilde{x}$ gives $p^* \geq g(\lambda, \nu)$

Least-norm solution of linear equations

$$\begin{array}{ll} \min & x^T x \\ \text{s.t.} & Ax = b \end{array}$$

dual function

- Lagrangian is $L(x, \nu) = x^T x + \nu^T (Ax - b)$
- to minimize L over x, set gradient equal to zero:

$$\nabla_x L(x, \nu) = 2x + A^T \nu = 0 \quad \Longrightarrow \quad x = -(1/2)A^T \nu$$

- plug in L to obtain g:

$$g(\nu) = L((-1/2)A^T \nu, \nu) = -\frac{1}{4}\nu^T A A^T \nu - b^T \nu$$

a concave function of ν

lower bound property: $p^* \geq -(1/4)\nu^T A A^T \nu - b^T \nu$ for all ν

Standard form LP

$$\begin{array}{ll} \min & c^T x \\ \text{s.t.} & Ax = b, \quad x \geq 0 \end{array}$$

dual function

- Lagrangian is

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

- L is affine in x, hence

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

g is linear on affine domain $\{(\lambda, \nu) | A^T \nu - \lambda + c = 0\}$, hence concave

lower bound property: $p^* \geq -b^T \nu$ if $A^T \nu + c \geq 0$

Equality constrained norm minimization

$$\begin{array}{ll} \min & \|x\| \\ \text{s.t.} & Ax = b \end{array}$$

dual function

$$g(\nu) = \inf_x (\|x\| - \nu^T Ax + b^T \nu) = \begin{cases} b^T \nu & \|A^T \nu\|_* \leq 1 \\ -\infty & \text{otherwise} \end{cases}$$

where $\|v\|_* = \sup_{\|u\| \leq 1} u^T v$ is dual norm of $\| \cdot \|$

proof: follows from $\inf_x (\|x\| - y^T x) = 0$ if $\|y\|_* \leq 1$, $-\infty$ otherwise

- if $\|y\|_* \leq 1$, then $x - y^T x \geq 0$ for all x, with equality if $x = 0$
- if $\|y\|_* > 1$, choose $x = tu$ where $\|u\| \leq 1$, $u^T y = \|y\|_* > 1$:

$$\|x\| - y^T x = t(\|u\| - \|y\|_*) \to -\infty \quad \text{as } t \to \infty$$

lower bound property: $p^* \geq b^T \nu$ if $\|A^T \nu\|_* \leq 1$

Two-way partitioning

$$\begin{array}{ll} \min & x^T W x \\ \text{s.t.} & x_i^2 = 1, \quad i = 1, \ldots, n \end{array}$$

- a nonconvex problem; feasible set contains 2^n discrete points
- interpretation: partition $\{1, \ldots, n\}$ in two sets; W_{ij} is cost of assigning i, j to the same set; $-W_{ij}$ is cost of assigning to different sets

dual function

$$g(\nu) = \inf_x (x^T W x + \sum_i \nu_i (x_i^2 - 1)) = \inf_x x^T (W + \mathrm{diag}(\nu)) x - \mathbf{1}^T \nu$$

$$= \begin{cases} -\mathbf{1}^T \nu & W + \mathrm{diag}(\nu) \succeq 0 \\ -\infty & \text{otherwise} \end{cases}$$

lower bound property: $p^* \geq -\mathbf{1}^T \nu$ if $W + \mathrm{diag}(\nu) \succeq 0$
example: $\nu = -\lambda_{\min}(W)\mathbf{1}$ gives bound $p^* \geq n\lambda_{\min}(W)$

Lagrange dual and conjugate function

$$\begin{array}{ll} \min & f_0(x) \\ \text{s.t.} & Ax \leq b, \quad Cx = d \end{array}$$

dual function

$$\begin{aligned} g(\lambda, \nu) &= \inf_{x \in \mathrm{dom}\, f_0} \left(f_0(x) + (A^T\lambda + C^T\nu)^T x - b^T\lambda - d^T\nu \right) \\ &= -f_0^*(-A^T\lambda - C^T\nu) - b^T\lambda - d^T\nu \end{aligned}$$

- recall definition of conjugate $f^*(y) = \sup_{x \in \mathrm{dom}\, f}(y^T x - f(x))$
- simplifies derivation of dual if conjugate of f_0 is known

example: entropy maximization

$$f_0(x) = \sum_{i=1}^n x_i \log x_i, \qquad f_0^*(y) = \sum_{i=1}^n e^{y_i - 1}$$

The dual problem

Lagrange dual problem

$$\begin{array}{ll} \max & g(\lambda, \nu) \\ \text{s.t.} & \lambda \succeq 0 \end{array}$$

- finds best lower bound on p^*, obtained from Lagrange dual function
- a convex optimization problem; optimal value denoted d^*
- λ, ν are dual feasible if $\lambda \succeq 0$, $(\lambda, \nu) \in \mathbf{dom}\, g$
- often simplified by making implicit constraint $(\lambda, \nu) \in \mathbf{dom}\, g$ explicit

example: standard form LP and its dual (page 5-5)

$$\begin{array}{ll} \min & c^T x \\ \text{s.t.} & Ax = b \\ & x \succeq 0 \end{array} \qquad \begin{array}{ll} \max & -b^T \nu \\ \text{s.t.} & A^T \nu + c \succeq 0 \end{array}$$

Weak and strong duality

weak duality: $d^* \leq p^*$

- always holds (for convex and nonconvex problems)
- can be used to find nontrivial lower bounds for difficult problems
  for example, solving the SDP

$$\begin{array}{ll} \max & -\mathbf{1}^T \nu \\ \text{s.t.} & W + \operatorname{diag}(\nu) \succeq 0 \end{array}$$

  gives a lower bound for the two-way partitioning problem on page 5-7

strong duality: $d^* = p^*$

- does not hold in general
- (usually) holds for convex problems
- conditions that guarantee strong duality in convex problems are called **constraint qualifications**

Slater's constraint qualification

strong duality holds for a convex problem

$$\begin{aligned} \min \quad & f_0(x) \\ \text{s.t.} \quad & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & Ax = b \end{aligned}$$

if it is strictly feasible, *i.e.*,

$$\exists x \in \mathrm{int}\, \mathcal{D}: \qquad f_i(x) < 0, \quad i = 1, \ldots, m, \quad Ax = b$$

- also guarantees that the dual optimum is attained (if $p^* > -\infty$)
- can be sharpened: *e.g.*, can replace $\mathrm{int}\, \mathcal{D}$ with $\mathrm{relint}\, \mathcal{D}$ (interior relative to affine hull); linear inequalities do not need to hold with strict inequality,...
- there exist many other types of constraint qualifications

Inequality form LP

primal problem

$$\begin{array}{ll} \min & c^T x \\ \text{s.t.} & Ax \le b \end{array}$$

dual function

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

dual problem

$$\begin{array}{ll} \max & -b^T\lambda \\ \text{s.t.} & A^T\lambda + c = 0, \quad \lambda \ge 0 \end{array}$$

- from Slater's condition: $p^* = d^*$ if $A\tilde{x} < b$ for some $\tilde{x}$
- in fact, $p^* = d^*$ except when primal and dual are infeasible

Quadratic program

primal problem (assume $P \in \mathbb{S}^n_{++}$)

$$\begin{array}{ll} \min & x^T P x \\ \text{s.t.} & Ax \leq b \end{array}$$

dual function

$$g(\lambda) = \inf_x \left(x^T P x + \lambda^T (Ax - b) \right) = -\frac{1}{4} \lambda^T A P^{-1} A^T \lambda - b^T \lambda$$

dual problem

$$\begin{array}{ll} \max & -(1/4) \lambda^T A P^{-1} A^T \lambda - b^T \lambda \\ \text{s.t.} & \lambda \geq 0 \end{array}$$

- from Slater's condition: $p^* = d^*$ if $A\tilde{x} < b$ for some $\tilde{x}$
- in fact, $p^* = d^*$ always

A nonconvex problem with strong duality

$$\begin{array}{ll} \min & x^T A x + 2b^T x \\ \text{s.t.} & x^T x \leq 1 \end{array}$$

$A \not\succeq 0$, hence nonconvex

dual function: $g(\lambda) = \inf_x (x^T(A + \lambda I)x + 2b^T x - \lambda)$

- unbounded below if $A + \lambda I \not\succeq 0$ or if $A + \lambda I \succeq 0$ and $b \notin \mathcal{R}(A + \lambda I)$
- minimized by $x = -(A + \lambda I)^\dagger b$ otherwise: $g(\lambda) = -b^T(A + \lambda I)^\dagger b - \lambda$

dual problem and equivalent SDP:

$$\begin{array}{ll} \max & -b^T(A + \lambda I)^\dagger b - \lambda \\ \text{s.t.} & A + \lambda I \succeq 0 \\ & b \in \mathcal{R}(A + \lambda I) \end{array}$$

$$\begin{array}{ll} \max & -t - \lambda \\ \text{s.t.} & \begin{bmatrix} A + \lambda I & b \\ b^T & t \end{bmatrix} \succeq 0 \end{array}$$

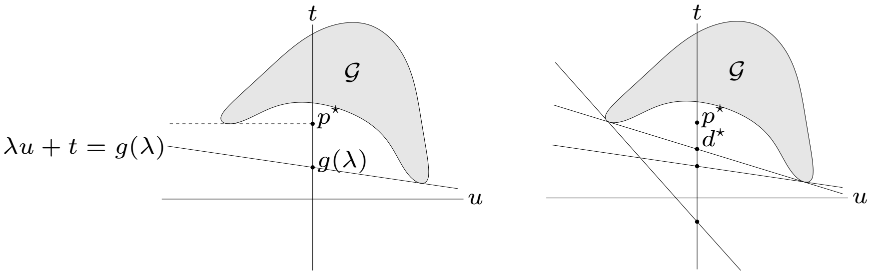
strong duality although primal problem is not convex (not easy to show)

Geometric interpretation

for simplicity, consider problem with one constraint $f_1(x) \leq 0$

interpretation of dual function:

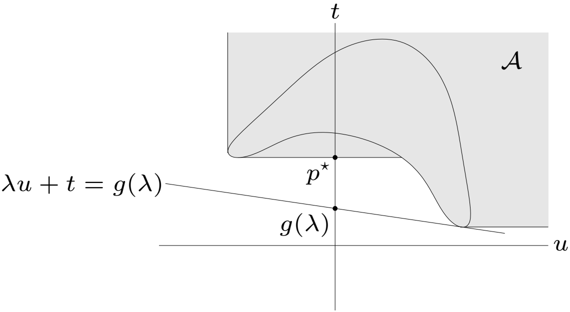
$$g(\lambda) = \inf_{(u,t)\in\mathcal{G}} (t + \lambda u), \qquad \text{where } \mathcal{G} = \{(f_1(x), f_0(x)) | x \in \mathcal{D}\}$$



- $\lambda u + t = g(\lambda)$ is (non-vertical) supporting hyperplane to $\mathcal{G}$
- hyperplane intersects t-axis at $t = g(\lambda)$

epigraph variation: same interpretation if $\mathcal{G}$ is replaced with

$$\mathcal{A} = \{(u,t) \mid f_1(x) \leq u, f_0(x) \leq t \text{ for some } x \in \mathcal{D}\}$$



strong duality

- holds if there is a non-vertical supporting hyperplane to $\mathcal{A}$ at $(0, p^*)$
- for convex problem, $\mathcal{A}$ is convex, hence has supp. hyperplane at $(0, p^*)$
- Slater's condition: if there exist $(\tilde{u}, \tilde{t}) \in \mathcal{A}$ with $\tilde{u} < 0$, then supporting hyperplanes at $(0, p^*)$ must be non-vertical

Complementary slackness

assume strong duality holds, x^* is primal optimal, (λ^*, ν^*) is dual optimal

$$\begin{aligned} f_0(x^*) = g(\lambda^*, \nu^*) &= \inf_x \left(f_0(x) + \sum_{i=1}^m \lambda_i^* f_i(x) + \sum_{i=1}^p \nu_i^* h_i(x) \right) \\ &\leq f_0(x^*) + \sum_{i=1}^m \lambda_i^* f_i(x^*) + \sum_{i=1}^p \nu_i^* h_i(x^*) \\ &\leq f_0(x^*) \end{aligned}$$

hence, the two inequalities hold with equality

- x^* minimizes $L(x, \lambda^*, \nu^*)$
- $\lambda_i^* f_i(x^*) = 0$ for $i = 1, \ldots, m$ (known as complementary slackness):

$$\lambda_i^* > 0 \Longrightarrow f_i(x^*) = 0, \qquad f_i(x^*) < 0 \Longrightarrow \lambda_i^* = 0$$

Karush-Kuhn-Tucker (KKT) conditions

the following four conditions are called KKT conditions (for a problem with differentiable f_i, h_i):

1. primal constraints: $f_i(x) \leq 0, i = 1, ..., m,\ h_i(x) = 0, i = 1, ..., p$
2. dual constraints: $\lambda \succeq 0$
3. complementary slackness: $\lambda_i f_i(x) = 0, i = 1, ..., m$
4. gradient of Lagrangian with respect to x vanishes:

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

from page 5-17: if strong duality holds and x, λ , ν are optimal, then they must satisfy the KKT conditions

KKT conditions for convex problem

if $\tilde{x}$,$\tilde{\lambda}$,$\tilde{\nu}$ satisfy KKT for a convex problem, then they are optimal:

- from complementary slackness: $f_0(\tilde{x}) = L(\tilde{x}, \tilde{\lambda}, \tilde{\nu})$
- from 4th condition (and convexity): $g(\tilde{\lambda}, \tilde{\nu}) = L(\tilde{x}, \tilde{\lambda}, \tilde{\nu})$

hence, $f_0(\tilde{x}) = g(\tilde{\lambda}, \tilde{\nu})$

if **Slater's condition** is satisfied:

x is optimal if and only if there exist λ , ν that satisfy KKT conditions

- recall that Slater implies strong duality, and dual optimum is attained
- generalizes optimality condition $\nabla f_0(x) = 0$ for unconstrained problem

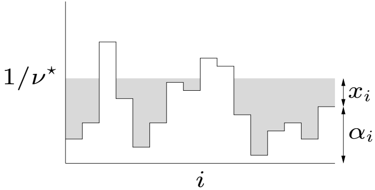
example:water-filling (assume $\alpha_i > 0$)

$$\begin{array}{ll} \min & -\sum_{i=1}^{n} \log(x_i + \alpha_i) \\ \text{s.t.} & x \geq 0, \quad \mathbf{1}^T x = 1 \end{array}$$

x is optimal iff $x \geq 0$, $\mathbf{1}^T x = 1$, and there exist $\lambda \in \mathbb{R}^n$, $\nu \in \mathbb{R}$ such that

$$\lambda \geq 0, \qquad \lambda_i x_i = 0, \qquad \frac{1}{x_i + \alpha_i} + \lambda_i = \nu$$

- if $\nu < 1/\alpha_i$: $\lambda_i = 0$ and $x_i = 1/\nu - \alpha_i$
- if $\nu \geq 1/\alpha_i$: $\lambda_i = \nu - 1/\alpha_i$ and $x_i = 0$
- determine ν from $\mathbf{1}^T x = \sum_{i=1}^{n} \max\{0, 1/\nu - \alpha_i\} = 1$



interpretation

- n patches; level of patch i is at height α_i
- flood area with unit amount of water
- resulting level is $1/\nu^*$

Perturbation and sensitivity analysis

(unperturbed) optimization problem and its dual

$$\begin{array}{lll} \min & f_0(x) & \\ \text{s.t.} & f_i(x) \leq 0, & i = 1, \ldots, m \\ & h_i(x) = 0, & i = 1, \ldots, p \end{array} \qquad \begin{array}{ll} \max & g(\lambda, \nu) \\ \text{s.t.} & \lambda \geq 0 \end{array}$$

perturbed problem and its dual

$$\begin{array}{lll} \min & f_0(x) & \\ \text{s.t.} & f_i(x) \leq u_i, & i = 1, \ldots, m \\ & h_i(x) = v_i, & i = 1, \ldots, p \end{array} \qquad \begin{array}{ll} \max & g(\lambda, \nu) - u^T \lambda - v^T \nu \\ \text{s.t.} & \lambda \geq 0 \end{array}$$

- x is primal variable; u, v are parameters
- $p^*(u, v)$ is optimal value as a function of u, v
- we are interested in information about $p^*(u, v)$ that we can obtain from the solution of the unperturbed problem and its dual

global sensitivity result

assume strong duality holds for unperturbed problem, and that λ^*, ν^* are dual optimal for unperturbed problem

apply weak duality to perturbed problem:

$$\begin{aligned} p^*(u,v) &\geq g(\lambda^*,\nu^*) - u^T\lambda^* - v^T\nu^* \\ &= p^*(0,0) - u^T\lambda^* - v^T\nu^* \end{aligned}$$

sensitivity interpretation

- if λ^* large: p^* increases greatly if we tighten constraint i ($u_i < 0$)
- if λ^* small: p^* does not decrease much if we loosen constraint i ($u_i > 0$)
- if ν^* large and positive: p^* increases greatly if we take $v_i < 0$;
  if ν^* large and negative: p^* increases greatly if we take $v_i > 0$
- if ν^* small and positive: p^* does not decrease much if we take $v_i > 0$;
  if ν^* small and negative: p^* does not decrease much if we take $v_i < 0$

local sensitivity: if (in addition) $p^*(u,v)$ is differentiable at $(0,0)$, then

$$\lambda_i^* = -\frac{\partial p^*(0,0)}{\partial u_i}, \qquad \nu_i^* = -\frac{\partial p^*(0,0)}{\partial v_i}$$

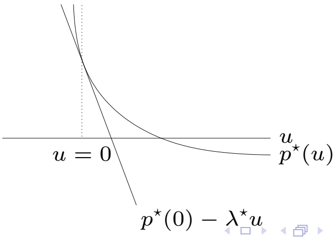
proof (for λ_i^*): from global sensitivity result,

$$\frac{\partial p^*(0,0)}{\partial u_i} = \lim_{t \searrow 0} \frac{p^*(te_i,0) - p^*(0,0)}{t} \geq -\lambda_i^*$$

$$\frac{\partial p^*(0,0)}{\partial u_i} = \lim_{t \nearrow 0} \frac{p^*(te_i,0) - p^*(0,0)}{t} \leq -\lambda_i^*$$

hence, equality

$p^*(u)$ for a problem with one (inequality) constraint:



Duality and problem reformulations

- equivalent formulations of a problem can lead to very different duals
- reformulating the primal problem can be useful when the dual is difficult to derive, or uninteresting

common reformulations

- introduce new variables and equality constraints
- make explicit constraints implicit or vice-versa
- transform objective or constraint functions
  $e.g.$, replace $f_0(x)$ by $\phi(f_0(x))$ with ϕ convex, increasing

Introducing new variables and equality constraints

$$\min \quad f_0(Ax+b)$$

- dual function is constant: $g = \inf_x L(x) = \inf_x f_0(Ax+b) = p^*$
- we have strong duality, but dual is quite useless

reformulated problem and its dual

$$\begin{array}{ll} \min & f_0(y) \\ \text{s.t.} & Ax+b-y=0 \end{array} \qquad \begin{array}{ll} \max & b^T\nu - f_0^*(\nu) \\ \text{s.t.} & A^T\nu = 0 \end{array}$$

dual function follows from

$$\begin{aligned} g(\nu) &= \inf_{x,y}(f_0(y) - \nu^T y + \nu^T Ax + b^T\nu) \\ &= \begin{cases} -f_0^*(\nu) + b^T\nu & A^T\nu = 0 \\ -\infty & \text{otherwise} \end{cases} \end{aligned}$$

norm approximation problem: $\min \|Ax - b\|$

$$\begin{array}{ll} \min & \|y\| \\ \text{s.t.} & y = Ax - b \end{array}$$

can look up conjugate of $\|\cdot\|$, or derive dual directly

$$\begin{aligned} g(\nu) &= \inf_{x,y}(\|y\| + \nu^T y - \nu^T Ax + b^T \nu) \\ &= \begin{cases} b^T \nu + \inf_y(\|y\| + \nu^T y) & A^T \nu = 0 \\ -\infty & \text{otherwise} \end{cases} \\ &= \begin{cases} b^T \nu & A^T \nu = 0, \quad \|\nu\|_* \le 1 \\ -\infty & \text{otherwise} \end{cases} \end{aligned}$$

dual of norm approximation problem

$$\begin{array}{ll} \max & b^T \nu \\ \text{s.t.} & A^T \nu = 0, \quad \|\nu\|_* \le 1 \end{array}$$

LP with box constraints: primal and dual problem

$$\begin{array}{ll} \min & c^T x \\ \text{s.t.} & Ax = b \\ & -\mathbf{1} \preceq x \preceq \mathbf{1} \end{array} \qquad \begin{array}{ll} \max & -b^T \nu - \mathbf{1}^T \lambda_1 - \mathbf{1}^T \lambda_2 \\ \text{s.t.} & c + A^T \nu + \lambda_1 - \lambda_2 = 0 \\ & \lambda_1 \succeq 0, \quad \lambda_2 \succeq 0 \end{array}$$

reformulation with box constraints made implicit

$$\begin{array}{ll} \min & f_0(x) = \begin{cases} c^T x & -\mathbf{1} \preceq x \preceq \mathbf{1} \\ -\infty & \text{otherwise} \end{cases} \\ \text{s.t.} & Ax = b \end{array}$$

dual function

$$\begin{aligned} g(\nu) &= \inf_{-\mathbf{1} \preceq x \preceq \mathbf{1}} (c^T x + \nu^T (Ax - b)) \\ &= -b^T \nu - \|A^T \nu + c\|_1 \end{aligned}$$

dual problem: $\max \ -b^T \nu - \|A^T \nu + c\|_1$

Problems with generalized inequalities

$$\begin{aligned} \min \quad & f_0(x) \\ \text{s.t.} \quad & f_i(x) \preceq_{K_i} 0, \quad i = 1, ..., m \\ & h_i(x) = 0, \quad i = 1, ..., p \end{aligned}$$

$\preceq_{K_i}$ is generalized inequality on $\mathbb{R}^{k_i}$

definitions are parallel to scalar case:

- Lagrange multiplier for $f_i(x) \preceq_{K_i} 0$ is vector $\lambda_i \in \mathbb{R}^{k_i}$
- Lagrangian $L : \mathbb{R}^n \times \mathbb{R}^{k_1} \times \cdots \times \mathbb{R}^{k_m} \times \mathbb{R}^p \to \mathbb{R}$, is defined as

$$L(x, \lambda_1, \ldots, \lambda_m, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i^T f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x)$$

- dual function $g : \mathbb{R}^{k_1} \times \cdots \times \mathbb{R}^{k_m} \times \mathbb{R}^p \to \mathbb{R}$, is defined as

$$g(\lambda_1, \ldots, \lambda_m, \nu) = \inf_{x \in \mathcal{D}} L(x, \lambda_1, \ldots, \lambda_m, \nu)$$

lower bound property: if $\lambda_i \succeq_{K_i^*} 0$, then $g(\lambda_1, ..., \lambda_m, \nu) \le p^*$

proof: if $\tilde{x}$ is feasible and $\lambda \succeq_{K_i^*} 0$, then

$$\begin{aligned} f_0(\tilde{x}) &\ge f_0(\tilde{x}) + \sum_{i=1}^{m} \lambda_i^T f_i(\tilde{x}) + \sum_{i=1}^{p} \nu_i h_i(\tilde{x}) \\ &\ge \inf_{x \in \mathcal{D}} L(x, \lambda_1, ..., \lambda_m, \nu) \\ &= g(\lambda_1, ..., \lambda_m, \nu) \end{aligned}$$

minimizing over all feasible $\tilde{x}$ gives $p^* \ge g(\lambda_1, ..., \lambda_m, \nu)$

dual problem

$$\begin{array}{ll} \max & g(\lambda_1, ..., \lambda_m, \nu) \\ \text{s.t.} & \lambda_i \succeq_{K_i^*} 0, \quad i = 1, ..., m \end{array}$$

- weak duality: $p^* \ge d^*$ always
- strong duality: $p^* = d^*$ for convex problem with constraint qualification
  (for example, Slater's: primal problem is strictly feasible)

Semidefinite program

primal SDP ($A_i, C \in \mathbb{S}^n$)

$$\begin{array}{ll} \min & b^T y \\ \text{s.t.} & y_1 A_1 + \cdots + y_m A_m \preceq C \end{array}$$

- Lagrange multiplier is matrix $Z \in \mathbb{S}^n$
- Lagrangian $L(y, Z) = b^T y + \mathrm{tr}(Z(y_1 A_1 + \cdots + y_m A_m - C))$
- dual function

$$g(Z) = \inf_y L(y, Z) = \begin{cases} -\mathrm{tr}(CZ) & \mathrm{tr}(A_i Z) + b_i = 0, \quad i = 1, \ldots, m \\ -\infty & \text{otherwise} \end{cases}$$

dual SDP

$$\begin{array}{ll} \max & -\mathrm{tr}(CZ) \\ \text{s.t.} & Z \succeq 0, \mathrm{tr}(A_i Z) + b_i = 0, \quad i = 1, \ldots, m \end{array}$$

$p^* = d^*$ if primal SDP is strictly feasible ($\exists y$ with $y_1 A_1 + \cdots + y_m A_m \prec C$)

LP Duality

Strong duality: If a LP has an optimal solution, so does its dual, and their objective fun. are equal.

dual \ primal	finite	unbounded	infeasible
finite	$\surd$	$\times$	$\times$
unbounded	$\times$	$\times$	$\surd$
infeasible	$\times$	$\surd$	$\surd$

- If $p^* = -\infty$, then $d^* \leq p^* = -\infty$, hence dual is infeasible
- If $d^* = +\infty$, then $+\infty = d^* \leq p^*$, hence primal is infeasible
- $$\begin{array}{lll} \min & x_1 + 2x_2 \\ \text{s.t.} & x_1 + x_2 = 1 \\ & 2x_1 + 2x_2 = 3 \end{array} \qquad \begin{array}{ll} \max & p_1 + 3p_2 \\ \text{s.t.} & p_1 + 2p_2 = 1 \\ & p_1 + 2p_2 = 2 \end{array}$$

SOCP/SDP Duality

$$\textbf{(P)} \quad \min \quad c^\top x \quad \text{s.t.} \quad Ax = b, x_{\mathcal{Q}} \succeq 0$$

$$\textbf{(D)} \quad \max \quad b^\top y \quad \text{s.t.} \quad A^\top y + s = c, s_{\mathcal{Q}} \succeq 0$$

$$\begin{aligned} \textbf{(P)} \quad \min \quad & \langle C, X \rangle \\ \text{s.t.} \quad & \langle A_1, X \rangle = b_1 \\ & \dots \\ & \langle A_m, X \rangle = b_m \\ & X \succeq 0 \end{aligned}$$

$$\begin{aligned} \textbf{(D)} \quad \max \quad & b^\top y \\ \text{s.t.} \quad & \sum_i y_i A_i + S = C \\ & S \succeq 0 \end{aligned}$$

Strong duality

- If $p^* > -\infty$, (P) is **strictly** feasible, then (D) is feasible and $p^* = d^*$
- If $d^* < +\infty$, (D) is **strictly** feasible, then (P) is feasible and $p^* = d^*$
- If (P) and (D) has **strictly** feasible solutions, then both have optimal solutions.

Failure of SOCP Duality

$$\begin{array}{lll} \inf & (1,-1,0)x \\ \text{s.t.} & (0,0,1)x = 1 \\ & x_{\mathcal{Q}} \succeq 0 \end{array} \qquad \begin{array}{ll} \sup & y \\ \text{s.t.} & (0,0,1)^\top y + z = (1,-1,0)^\top \\ & z_{\mathcal{Q}} \succeq 0 \end{array}$$

- primal: $\min\ x_0 - x_1,$ s.t. $x_0 \geq \sqrt{x_1^2 + 1}$; It holds $x_0 - x_1 > 0$ and $x_0 - x_1 \to 0$ if $x_0 = \sqrt{x_1^2 + 1} \to \infty$. Hence, $p^* = 0$, no finite solution
- dual: $\sup\ y$ s.t. $1 \geq \sqrt{1 + y^2}$. Hence, $y = 0$

$p^* = d^*$ but primal is not attainable.

Failure of SDP Duality

Consider

$$\begin{aligned} \min \quad & \left\langle \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, X \right\rangle \\ \text{s.t.} \quad & \left\langle \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, X \right\rangle = 0 \\ & \left\langle \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 2 \end{pmatrix}, X \right\rangle = 2 \\ & X \succeq 0 \end{aligned}$$

$$\begin{aligned} \max \quad & 2y_2 \\ \text{s.t.} \quad & \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} y_1 + \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 2 \end{pmatrix} y_2 \preceq \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \end{aligned}$$

- primal: $X^* = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$, $p^* = 1$
- dual: $y^* = (0, 0)$. Hence, $d^* = 0$

Both problems have finite optimal values, but $p^* \neq d^*$