Lecture 9: Duality

goals:
- intro to Lagrange duality
- systematic way to obtain related problem that gives lower bounds on optimal value
    - interpretation as "certificates" that prove performance limits
- weak + strong duality
- sensitivity
- examples

The Lagrangian Function

Recall standard form optimization problem (possibly non-convex)

$$\boxed{\begin{array}{lll} \text{minimize} & f(x) & \\ \text{subject to} & g_i(x) \leq 0 & i=1,\dots,m \\ & h_i(x) = 0 & i=1,\dots,p \end{array}} \qquad \text{(primal)}$$

with (primal) variable $x \in \mathbb{R}^n$, domain D, optimal value p^*

Lagrangian function: $L : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$ $\quad (\text{dom } L = D \times \mathbb{R}^m \times \mathbb{R}^p)$

$$L(x, \lambda, \nu) = f(x) + \sum_{i=1}^m \lambda_i g_i(x) + \sum_{i=1}^p \nu_i h_i(x) \qquad \lambda_i \geq 0$$

- λ_i : Lagrange multiplier associated with $g_i(x) \leq 0$
- ν_i : Lagrange multiplier associated with $h_i(x) = 0$

- Lagrangian is weighted sum of objective and constraint functions

<u>Lagrange Dual Function</u>: $g : \mathbb{R}^m \times \mathbb{R}^p \rightarrow \mathbb{R}$

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

$$= \inf_{x \in D} \left(f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{i=1}^{p} \nu_i h_i(x) \right)$$

<u>Note</u>: ① g is always <u>concave</u>, regardless of whether primal problem is convex

- g is <u>pointwise</u> <u>infimum</u> of a family of affine functions

② g can be <u>extended</u> <u>valued</u>

- can be $-\infty$ for some λ, ν

Lower bound property

If $\lambda \succeq 0$, then $g(\lambda, \nu) \le p^*$

<u>proof</u>: for any feasible $\tilde{x}$ and $\lambda \succeq 0$, we have

$$f(\tilde{x}) \ge L(\tilde{x}, \lambda, \nu) \ge \inf_{x \in D} L(x, \lambda, \nu) = g(\lambda, \nu)$$

minimizing f over feasible $\tilde{x}$ gives $p^* \ge g(\lambda, \nu)$ ∎

Lagrange Dual Problem

$$\boxed{\begin{array}{ll} \text{maximize} & g(\lambda, \nu) \\ \text{subject to} & \lambda \succeq 0 \end{array}} \qquad \text{(dual)}$$

- gives best lower bound on p^* obtained from Lagrange dual function
- always a convex optimization problem (concave maximization) even when primal problem is not
- denote optimal value d^*; we have $d^* \le p^*$ (more later)
- (λ, ν) called dual feasible if $\lambda \succeq 0$ and $(\lambda, \nu) \in \mathrm{dom}(g)$
  - can often impose constraint $(\lambda, \nu) \in \mathrm{dom}(g)$ explicitly in dual problem

Ex Dual of a Linear Program (LP)

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

"standard form"

<u>dual function</u>:

- Lagrangian: $L(x, \lambda, \nu) = c^T x + \nu^T (Ax - b) - \lambda^T x$

$$= -b^T \nu + (c + A^T \nu - \lambda)^T x$$

- Since L is affine on x

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

i.e. g is <u>linear</u> on affine domain $\{(\lambda, \nu) \mid A^T \nu - \lambda + c = 0\}$

<u>dual problem(s)</u>:

$$\begin{array}{ll} \text{maximize} & g(\lambda, \nu) \\ \text{subject to} & \lambda \succeq 0 \end{array} \iff \begin{array}{ll} \text{maximize} & -b^T \nu \\ \text{subject to} & A^T \nu - \lambda + c = 0 \\ & \lambda \succeq 0 \end{array} \iff \begin{array}{ll} \text{maximize} & -b^T \nu \\ \text{subject to} & \\ & A^T \nu + c \succeq 0 \end{array}$$

"inequality form"

- lower bound: $p^* \geq -b^T \nu$ if $A^T \nu + c \succeq 0$
- best lower bound $p^* \geq d^*$
- in fact, for LPs $p^* = d^*$ except when both primal and dual are infeasible (then $p^* = +\infty$, $d^* = -\infty$)

Ex Dual of a Quadratic Program (QP)

$$\begin{array}{ll} \text{minimize} & x^T P x \\ \text{subject to} & Ax \preceq b \end{array} \qquad (\text{assume } P \in S^n_{++})$$

<u>dual</u> <u>function</u>

$$g(\lambda) = \inf_x \overbrace{\left(x^T P x + \lambda^T (Ax - b) \right)}^{L(x,\lambda)}$$

$$\frac{\partial L}{\partial x} = 2Px + A^T \lambda = 0 \implies x^* = -\frac{1}{2} P^{-1} A^T \lambda$$

$$= \frac{1}{4} \lambda^T A P^{-1} P P^{-1} A^T \lambda - \frac{1}{2} \lambda^T A P^{-1} A^T \lambda - b^T \lambda$$

$$= -\frac{1}{4} \lambda^T A P^{-1} A^T \lambda - b^T \lambda$$

<u>dual</u> <u>problem</u>

$$\begin{array}{ll} \text{maximize} & -\frac{1}{4} \lambda^T A P^{-1} A \lambda - b^T \lambda \\ \text{subject to} & \lambda \succeq 0 \end{array}$$

- also a QP

Ex Equality constrained norm minimization

$$\begin{array}{ll} \text{minimize} & \|x\| \qquad \text{(any norm)} \\ \text{subject to} & Ax = b \end{array}$$

dual function

$$g(\nu) = \inf_x \overbrace{\left(\|x\| + \nu^T(b - Ax)\right)}^{L(x,\nu)} = \inf_x \left(\|x\| - \nu^T A x\right) + b^T \nu$$

$$\inf_x \left(\|x\| - y^T x\right) = \begin{cases} 0 & \|y\|_* \le 1 \\ -\infty & \text{otherwise} \end{cases}$$

$$= \begin{cases} b^T \nu & \|A^T \nu\|_* \le 1 \\ -\infty & \text{otherwise} \end{cases}$$

dual problem

$$\begin{array}{ll} \text{maximize} & b^T \nu \\ \text{subject to} & \|A^T \nu\|_* \le 1 \end{array}$$

Ex Dual of a Mixed-Integer Linear Program (MILP)

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax \le b \\ & x \in \{-1, 1\}^n \end{array}$$

dual function

$$g(\lambda) = \inf_{x_i \in \{-1,1\}} \left(c^T x + \lambda^T (Ax - b) \right) = \inf_{x_i \in \{-1,1\}} \left((c^T + \lambda^T A) x \right) - b^T \lambda$$

$$\inf_{x_i \in \{-1,1\}} z^T x = -\sup_{x_i \in \{-1,1\}} -z^T x = -\sup_{\|x\|_\infty \le 1} -z^T x = -\|z\|_1$$

$$= -\|A^T \lambda + c\|_1 - b^T \lambda$$

dual problem

$$\begin{array}{ll} \text{maximize} & -\|A^T \lambda + c\|_1 - b^T \lambda \\ \text{subject to} & \lambda \ge 0 \end{array}$$

- dual of an ILP is (equivalent to) an LP! (w/o integers)

Weak and Strong Duality

- **Weak Duality**
  - always true that $d^* \le p^*$
  - can be used to find nontrivial lower bounds for difficult (non-convex) problems
  - dual feasible points give suboptimality certificates for any primal feasible points
    - "primal-dual" algorithms not only give primal solutions, but also (dual) certificates proving distance to optimality

- **Strong Duality**
  - sometimes true that $d^* = p^*$
  - usually holds for convex problems (except in pathological cases)
  - usually does not hold for non-convex problems
    - but sometimes does! in very special cases
  - conditions that guarantee strong duality in convex problems are called constraint qualifications (CQs)

Slater's Constraint Qualification

- most widely known of many possible CQs
- strong duality holds for a convex problem

$$\begin{array}{lll} \text{minimize} & f(x) & \\ \text{subject to} & g_i(x) \le 0 & i=1,\dots,m \\ & Ax = b & \end{array}$$

if it is strictly feasible, i.e.

$$\exists\, x \in \text{int dom}(f) : \begin{array}{ll} g_i(x) < 0 & i=1,\dots,m \\ Ax = b & \end{array}$$

- usually satisfied by practical convex problems
- only sufficient! Slater $\Longrightarrow$ strong duality
  - but $\exists$ problems where Slater fails yet strong duality holds
- stronger version: only nonlinear constraints must be strictly feasible
- also guarantees that the dual optimal value is obtained (if $p^* > -\infty$)

Complementary Slackness

- Assume strong duality holds, with x^* primal optimal and (λ^*, ν^*) dual optimal, i.e.

$$g(\lambda^*, \nu^*) = d^* = p^* = f(x^*)$$

Then

$$\begin{aligned} f(x^*) = g(\lambda^*, \nu^*) &= \inf_{x \in D} \left(f(x) + \sum_{i=1}^m \lambda_i^* g_i(x) + \sum_{i=1}^p \nu_i^* h_i(x) \right) \\ &\le f(x^*) + \sum_{i=1}^m \lambda_i^* g_i(x^*) + \sum_{i=1}^p \nu_i^* h_i(x^*) \\ &\le f(x^*) \qquad \text{(since } x^* \text{ and } (\lambda^*, \nu^*) \text{ are feasible)} \end{aligned}$$

$$\Longrightarrow f(x^*) = f(x^*) + \sum_{i=1}^m \lambda_i^* g_i(x^*) + \sum_{i=1}^p \nu_i^* h_i(x^*)$$

But since $h_i(x^*) = 0$ and $g_i(x^*) \le 0$

$$\boxed{\lambda_i^* g_i(x^*) = 0 \qquad \forall i = 1, \dots, m}$$

This property is called complementary slackness

i.e. $g_i(x^*) < 0 \Longrightarrow \lambda_i^* = 0$ and $\lambda_i^* > 0 \Longrightarrow g_i(x) = 0$

i.e. λ^* and $\begin{bmatrix} g_1(x^*) \\ \vdots \\ g_m(x^*) \end{bmatrix}$ have complementary sparsity patterns

The Karush - Kuhn - Tucker (KKT) Conditions

- assume objective & constraint functions differentiable, but problem not necessarily convex
- <u>necessary</u> conditions for optimality and strong duality that generalize previous gradient conditions to constrained problems

① Primal feasibility

$$g_i(x^*) \leq 0 \qquad i = 1, \dots, m$$
$$h_i(x^*) = 0 \qquad i = 1, \dots, p$$

② Dual feasibility

$$\lambda_i^* \geq 0 \qquad i = 1, \dots, m$$

③ Complementary slackness

$$\lambda_i^* g_i(x^*) = 0 \qquad i = 1, \dots, m$$

④ Stationarity

$$\nabla L_x(x^*, \lambda^*, \nu^*) = \nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla g_i(x^*) + \sum_{i=1}^{p} \nu_i^* \nabla h_i(x^*) = 0$$

- (x^*, λ^*, ν^*) optimal $\Longrightarrow$ (x^*, λ^*, ν^*) satisfy KKT (only necessary)
  - c.f. 1st order cond. for unconstrained prob $\nabla f(x^*) = 0$

- For convex optimization problems

① If $\tilde{x}, \tilde{\lambda}, \tilde{\nu}$ satisfy KKT, then they are optimal, and strong duality holds

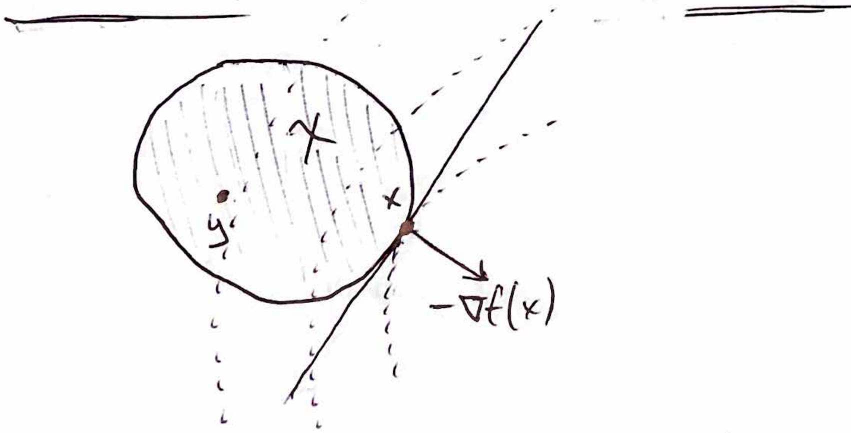
- from complementary slackness: $f(\tilde{x}) = L(\tilde{x}, \tilde{\lambda}, \tilde{\nu})$
- from stationarity and convexity: $g(\tilde{\lambda}, \tilde{\nu}) = L(\tilde{x}, \tilde{\lambda}, \tilde{\nu})$
  - L convex and $\nabla_x L(\tilde{x}) = 0 \implies \tilde{x}$ minimizes L
- hence $f(\tilde{x}) = g(\tilde{\lambda}, \tilde{\nu}) = p^* = d^*$

② If Slater's CQ holds, then

- x^* optimal $\iff \exists (\lambda^*, \nu^*)$ satisfying KKT

Geometrically, for a convex problem w/ differentiable objective, x is optimal iff it is feasible and

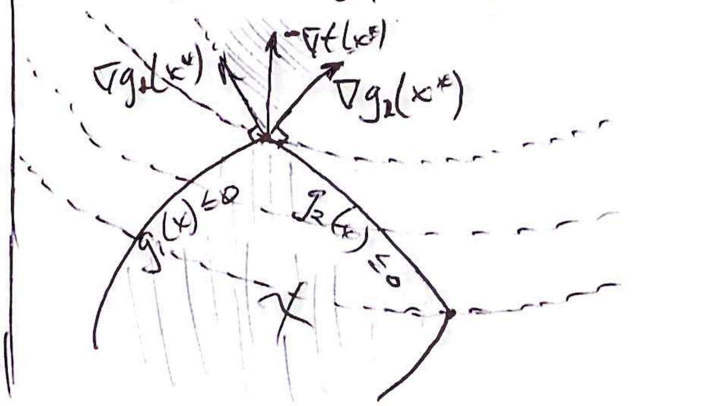
$$\nabla f(x)^T (y - x) \geq 0 \quad \forall \text{ feasible } y$$



- gradient can be non-zero as long as all other feasible points are "uphill" from optimum

- Assume inequality constraints only
- From stationarity

$$-\nabla f(x) = \sum_{i=1}^{m} \lambda_i \nabla g_i(x)$$



Ex KKT Conditions for a QP

$$\begin{aligned} \text{minimize} \quad & \tfrac{1}{2} x^T Q x + q^T x \qquad Q \succeq 0 \\ \text{subject to} \quad & Ax = b \\ & x \succeq 0 \end{aligned}$$

- Lagrangian: $L(x, \lambda, \nu) = \frac{1}{2} x^T Q x + q^T x + \nu^T (Ax - b) - \lambda^T x$

- KKT:

$$\left.\begin{aligned} Ax &= b \\ x &\succeq 0 \end{aligned}\right\} \quad \text{primal feas.}$$

$$\lambda \succeq 0 \qquad \text{dual feas.}$$

$$x_i \lambda_i = 0 \qquad \text{complementarity}$$

$$\nabla_x L = Qx + A^T \nu - \lambda + q = 0 \qquad \text{stationarity}$$

- algorithms for constrained optimization essentially search for points that satisfy KKT

Perturbation and Sensitivity Analysis

- optimization problem and its dual

minimize $f(x)$

subject to $g_i(x) \le 0 \quad i=1,\dots,m$

$h_i(x) = 0 \quad i=1,\dots,p$

maximize $g(\lambda, \upsilon)$

subject to $\lambda \succeq 0$

- perturbed problem and its dual

minimize $f(x)$

subject to $g_i(x) \le u_i \quad i=1,\dots,m$

$h_i(x) = v_i \quad i=1,\dots,p$

maximize $g(\lambda, \upsilon) - u^T\lambda - v^T\upsilon$

subject to $\lambda \succeq 0$

- u, v are parameters perturbing constraints
  - u_i loosens or tightens g_i
  - v_i shifts target value of h_i

- want to study sensitivity of optimal value $p^*(u, v)$ to changes in u, v
  - info from unperturbed problem and its dual?

Global Sensitivity

- assume strong duality holds for unperturbed prob., let λ^*, ν^* be dual optimal
- apply weak duality to perturbed problem:

$$p^*(u,v) \geq g(\lambda^*, \nu^*) - u^T\lambda^* - v^T\nu^*$$

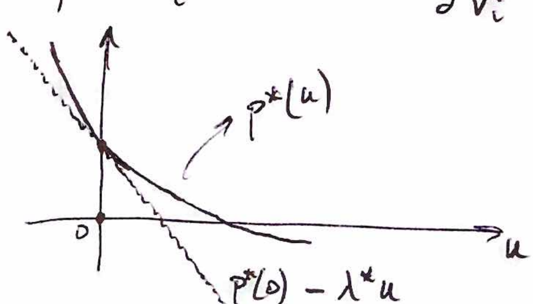
$$= p^*(0,0) - u^T\lambda^* - v^T\nu^*$$

- sensitivity interpretation
  - λ_i^* large $\Longrightarrow$ p^* increases greatly if constraint i tightened ($u_i < 0$)
  - λ_i^* small $\Longrightarrow$ p^* decreases little if constraint i loosened ($u_i > 0$)
  - ν_i^* large and positive $\Longrightarrow$ p^* increases greatly if $v_i < 0$
  - ν_i^* " " negative $\Longrightarrow$ p^* " " if $v_i > 0$
  - ν_i^* small and positive $\Longrightarrow$ p^* decreases little if $v_i > 0$
  - ν_i^* " " negative $\Longrightarrow$ p^* " " if $v_i < 0$

Local Sensitivity: if $p^*(u,v)$ differentiable at $(0,0)$, then

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

Ex $p^*(u)$ for prob. w/ one inequality constraint



Duality and Problem Reformulations

- equivalent formulations of a problem can lead to very different duals
- primal reformulation can be very useful when dual is difficult to derive (or uninteresting)
    - lots of possible reformulations

Ex Box-constrained LP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b \\ & -1 \le x \le 1 \end{array} \quad \Longleftrightarrow \quad \begin{array}{ll} \text{maximize} & -b^T \nu - 1^T \lambda_1 - 1^T \lambda_2 \\ \text{subject to} & c + A^T \nu + \lambda_1 - \lambda_2 = 0 \\ & \lambda_1 \ge 0, \ \lambda_2 \ge 0 \end{array}$$

Reformulate box constraints implicitly

$$\begin{array}{ll} \text{minimize} & f(x) = \begin{cases} c^T x & -1 \le x \le 1 \\ +\infty & \text{otherwise} \end{cases} \\ \text{subject to} & Ax = b \end{array}$$

$(\Longleftrightarrow \|x\|_\infty \le 1)$

Dual function

$$g(\nu) = \inf_{\|x\|_\infty \le 1} \left(c^T x + \nu^T (Ax - b) \right)$$

$$= -b^T \nu - \|A^T \nu + c\|_1$$

Dual problem:

$$\text{maximize} \quad -b^T \nu - \|A^T \nu + c\|_1$$

(no constraints!)

Duality with generalized (cone) inequalities

$$\boxed{\begin{array}{lll} \text{minimize} & f(x) & \\ \text{subject to} & g_i(x) \preceq_{K_i} 0 & i=1,\dots,m \\ & h_i(x) = 0 & i=1,\dots,p \end{array}}$$

- $\preceq_{K_i}$ is generalized (cone) inequality on $\mathbb{R}^{k_i}$ or S^{k_i}
- everything in previous discussion holds, but we use inner product when forming Lagrangian, and multipliers of generalized inequalities constrained to dual cone
  - $L(x, \lambda_1, \dots, \lambda_m, \nu) = f(x) + \sum_{i=1}^{m} \langle \lambda_i, f_i(x) \rangle + \sum_{i=1}^{p} \nu_i h_i(x)$
  - lower bound holds when $\lambda_i \in K_i^*$
- dual problem: $$\boxed{\begin{array}{lll} \text{maximize} & g(\lambda_1, \dots, \lambda_m, \nu) & \\ \text{subject to} & \lambda_i \succeq_{K_i^*} 0 & i=1,\dots,m \end{array}}$$

- weak duality: $d^* \leq p^*$ always
- stron duality: $d^* = p^*$ for convex problems w/ CQ (e.g. Slater)

Ex Dual of an SDP (in standard form)

$$\begin{aligned} &\text{minimize} && \operatorname{trace}(CX) \\ &\text{subject to} && \operatorname{trace}(A_iX) = b_i \quad i=1,\dots,m \\ & && X \succeq 0 \end{aligned}$$

Lagrangian:

$$L(X,\nu,\Lambda) = \operatorname{trace}(CX) + \sum_{i=1}^{m} \nu_i(\operatorname{trace}(A_iX) - b_i) - \underbrace{\operatorname{trace}(\Lambda X)}_{\langle \Lambda, X\rangle}$$

$$= \operatorname{trace}\left(\left(C + \sum_{i=1}^{m} \nu_i A_i - \Lambda\right)X\right) - \nu^T b$$

Dual function:

$$g(\nu,\Lambda) = \inf_{X\in S^n} L(X,\nu,\Lambda) = \begin{cases} -\nu^T b & C + \sum_{i=1}^{m} \nu_i A_i - \Lambda = 0 \\ -\infty & \text{otherwise} \end{cases}$$

Dual problem:

$$\begin{aligned} &\text{maximize} && -\nu^T b \\ &\text{subject to} && C + \nu_1 A_1 + \dots + \nu_m A_m - \Lambda = 0 \\ & && \Lambda \succeq 0 \end{aligned} \iff \begin{aligned} &\text{maximize} && -\nu^T b \\ &\text{subject to} && \\ & && C + \nu_1 A_1 + \dots + \nu_m A_m \succeq 0 \end{aligned}$$

- nicely mirrors duality for LPs

Ex Dual of an SDP (in inequality form)

$$\boxed{\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & x_1 A_1 + \cdots + x_n A_n \preceq B \end{array}} \qquad A_i, B \in S^k$$

Lagrangian:

$$L(x, \Lambda) = c^T x + \overbrace{\operatorname{trace}\left(\Lambda\left(x_1 A_1 + \cdots + x_n A_n - B\right)\right)}^{\langle \Lambda,\, x_1 A_1 + \cdots + x_n A_n - B \rangle}$$

$$= c^T x + x_1 \operatorname{trace}(\Lambda A_1) + \cdots + x_n \operatorname{trace}(\Lambda A_n) - \operatorname{trace}(\Lambda B)$$

Dual function:

$$g(\Lambda) = \inf_x L(x, \Lambda) = \begin{cases} -\operatorname{trace}(B\Lambda) & c_i + \operatorname{trace}(A_i \Lambda) \quad i = 1, \ldots, n \\ -\infty & \text{otherwise} \end{cases}$$

Dual problem:

$$\boxed{\begin{array}{lll} \text{maximize} & -\operatorname{trace}(B\Lambda) & \\ \text{subject to} & c_i + \operatorname{trace}(A_i \Lambda) = 0 & i = 1, \ldots, n \\ & \Lambda \succeq 0 & \end{array}}$$