Convex Optimization

Lecture 5 - Duality

Instructor: Yuanzhang Xiao

University of Hawaii at Manoa

Fall 2017

Today's Lecture

1. The Lagrange Dual Function
2. Lagrange Dual Function and Duality
3. Optimality Conditions
4. Problem Reformulation and Dual Problems

Outline

The Lagrangian

a general problem: (not necessarily convex)

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

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

Lagrangian: $L : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$

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

with $\mathrm{dom} L = \mathcal{D} \times \mathbb{R}^m \times \mathbb{R}^p$

- λ_i: Lagrangian multiplier associated with the ith inequality
- ν_i: Lagrangian multiplier associated with the ith equality
- λ, ν: Lagrangian multipliers or dual variables

The Lagrange Dual Function

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

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

important properties:

- g is concave in (λ, ν)
  – even when the original problem is not convex
- $g(\lambda, \nu) \leq p^\star$ for any $\lambda \geq 0$ and any ν
  $\to$ g provides lower bounds of optimal value $p^\star$

note:

- $g(\lambda, \nu)$ can be $-\infty$ for some (λ, ν)
- dual feasible (λ, ν): $\lambda \geq 0$ and $g(\lambda, \nu) > -\infty$

Examples - Least-Squares Solution of Linear Equations

least-square solution of linear equations:

$$\begin{aligned} \text{minimize} \quad & x^T x \\ \text{subject to} \quad & Ax = b \end{aligned}$$

find the dual function:

- Lagrangian $L(x, \nu) = x^T x + \nu^T (Ax - b)$
- solve the problem

$$\inf_{x \in \mathbb{R}} L(x, \nu) = x^T x + \nu^T (Ax - b)$$

  infimum obtained when $x = -\frac{1}{2} A^T \nu$
- the dual function is

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

Examples - Standard Form LP

standard form LP:

$$\begin{aligned} \text{minimize} \quad & c^T x \\ \text{subject to} \quad & Ax = b \\ & x \geq 0 \end{aligned}$$

find the dual function:

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

$$\inf_{x \in \mathbb{R}} L(x, \lambda, \nu) = -\nu^T b + \inf_x \left(c + A^T \nu - \lambda\right)^T x$$

- the dual function is

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

Examples - Two-Way Partitioning

(nonconvex) two-way partitioning:

$$\begin{array}{ll} \text{minimize} & x^T W x \\ \text{subject to} & x_i^2 = 1, \ i = 1, \dots, n \end{array}$$

with $W \in \mathbb{S}^n$.

find the dual function:

- Lagrangian

$$\begin{aligned} L(x, \nu) &= x^T W x + \sum_{i=1}^n \nu_i (x_i^2 - 1) \\ &= x^T \left(W + \mathrm{diag}(\nu) \right) x - 1^T \nu \end{aligned}$$

- the dual function is

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

Outline

The Dual Problem

(Lagrange) dual problem:

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

- find the best lower bound on $p^\star$
- always a convex optimization problem
- simplified by making implicit constraint $(\lambda, \nu) \in \mathrm{dom}\, g$ explicit

denote the optimal value of the dual problem by $d^\star$

Implicit Constraints – Examples

standard form LP: minimize $c^T x$ subject to $Ax = b, x \geq 0$

the dual function: $g(\lambda, \nu) = \begin{cases} -\nu^T b & c + A^T\nu - \lambda = 0 \\ -\infty & \text{otherwise} \end{cases}$

the dual problem:

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

equivalent problems:

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

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

Duality

weak duality: $d^\star \le p^\star$

- holds for any problem (convex or nonconvex)
- can be used to find lower bounds for difficult problems

strong duality: $d^\star = p^\star$

- does not hold in general
- (usually) holds for convex problems
- constraint qualifications: conditions under which strong duality holds for convex problems

Slater's Constraint Qualification

Slater's Constraint Qualification

strong duality holds for a convex optimization problem if there exists a strictly feasible point x, namely

$$x \in \text{int}\mathcal{D} \text{ such that } f_i(x) < 0, \; i = 1, \ldots, m, \; Ax = b$$

relaxation:
only feasibility (not strict feasibility) are needed for affine constraints

Examples - QCQP

QCQP:

$$\begin{array}{ll} \text{minimize} & (1/2)x^T P_0 x + q_0^T x + r_0 \\ \text{subject to} & (1/2)x^T P_i x + q_i^T x + r_i \leq 0, \;\; i = 1, \ldots, m \end{array}$$

with $P_0 \in \mathbb{S}^n_{++}$, and $P_i \in \mathbb{S}^n_{+}$, $i = 1, \ldots, m$

Lagrangian: $L(x, \lambda) = (1/2)x^T P(\lambda)x + q(\lambda)^T x + r(\lambda)$ with

$$P(\lambda) = P_0 + \sum_{i=1}^{m} \lambda_i P_i, \;\; q(\lambda) = q_0 + \sum_{i=1}^{m} \lambda_i q_i, \;\; r(\lambda) = r_0 + \sum_{i=1}^{m} \lambda_i r_i$$

dual function: $g(\lambda) = -(1/2)q(\lambda)^T P(\lambda)^{-1} q(\lambda) + r(\lambda)$

dual problem:

$$\begin{array}{ll} \text{maximize} & -(1/2)q(\lambda)^T P(\lambda)^{-1} q(\lambda) + r(\lambda) \\ \text{subject to} & \lambda \geq 0 \end{array}$$

Examples - Nonconvex Problem With Strong Duality

a nonconvex problem:

$$\begin{array}{ll} \text{minimize} & x^T A x + 2b^T x \\ \text{subject to} & x^T x \leq 1 \end{array}$$

with $A \in \mathbb{S}^n$ but $A \not\succeq 0$

Lagrangian: $L(x, \lambda) = x^T (A + \lambda I) x + 2b^T x - \lambda$

dual function:

$$g(\lambda) = \begin{cases} -b^T (A + \lambda I)^{-1} b - \lambda & A + \lambda I \succeq 0, \ b \in \mathcal{R}(A + \lambda I) \\ -\infty & \text{otherwise} \end{cases}$$

strong duality holds (see book chapter B.1)

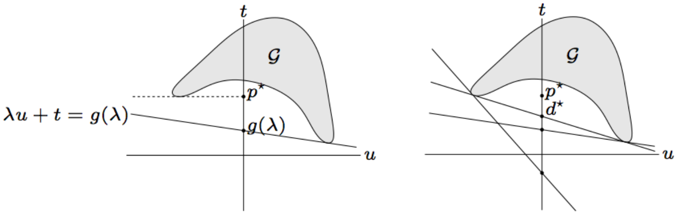
Geometric Interpretation – Weak Duality

a simple problem with one constraint:

$$\text{minimize } f_0(x) \quad \text{subject to } f_1(x) \leq 0$$

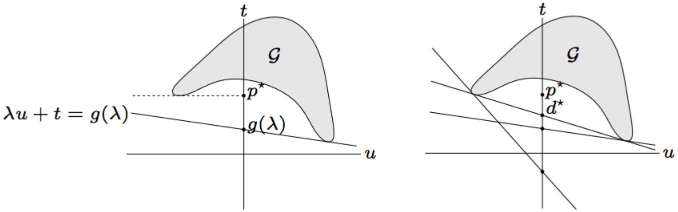
dual function:

$$g(\lambda) = \inf_{(u,t)\in\mathcal{G}} (\lambda, 1)^T(u,t), \text{ where } \mathcal{G} = \{(f_1(x), f_0(x)) \,|x \in \mathcal{D}\}$$



- $(\lambda, 1)^T(u,t) = g(\lambda)$ is a supporting hyperplane to $\mathcal{G}$
- $g(\lambda)$ is the intersection of t-axis and supporting hyperplane

Geometric Interpretation – Strong Duality



- supporting hyperplane $(\lambda, 1)^T(u, t) = g(\lambda)$ is not vertical
- for convex problems, $\mathcal{G}$ is convex
- there exists a supporting hyperplane at $(0, p^\star)$
- Slater's condition: exists $(u, t) \in \mathcal{G}$ with $u < 0 \Rightarrow$ supporting plane must be non-vertical

Outline

Duality Gap and Certificate of Suboptimality

a dual feasible (λ, ν) and a primal feasible x

duality gap:

$$f_0(x) - g(\lambda, \nu)$$

- duality gap is zero
  $\Rightarrow$ x is primal optimal and (λ, ν) is dual optimal

suboptimality certificate:

$$f_0(x) - p^\star \leq f_0(x) - g(\lambda, \nu)$$

- distance to the optimal value no greater than duality gap

Complementary Slackness

suppose that strong duality holds

$x^\star$ is primal optimal, and $(\lambda^\star, \nu^\star)$ is dual optimal

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

- $x^\star$ minimizes $L(x, \lambda^\star, \nu^\star)$
- complementary slackness: $\lambda_i^\star f_i(x^\star) = 0, \; i = 1, \ldots, m$

Karush-Kuhn-Tucker (KKT) Conditions

suppose that for a general problem with differentiable functions,

- $x^\star$ is primal optimal and $(\lambda^\star, \nu^\star)$ is dual optimal
- zero duality gap

we must have KKT conditions:

$$
\begin{aligned}
f_i(x^\star) &\leq 0, \quad i = 1, \ldots, m \\
h_i(x^\star) &= 0, \quad i = 1, \ldots, p \\
\lambda_i^\star &\geq 0, \quad i = 1, \ldots, m \\
\lambda_i^\star f_i(x^\star) &= 0, \quad i = 1, \ldots, m \\
\nabla f_0(x^\star) + \sum_{i=1}^{m} \lambda_i^\star \nabla f_i(x^\star) + \sum_{i=1}^{p} \nu_i^\star \nabla h_i(x^\star) &= 0
\end{aligned}
$$

for any problem where strong duality holds, any pair of primal and dual optimal points must satisfy KKT conditions

$\Rightarrow$ KKT conditions are necessary for optimality

KKT Conditions For Convex Problems

for convex problems, if $\tilde{x}$ and $(\tilde{\lambda}, \tilde{\nu})$ satisfy KKT conditions, then:

- complementary slackness $\Rightarrow f_0(\tilde{x}) = L(\tilde{x}, \tilde{\lambda}, \tilde{\nu})$
- convexity $\Rightarrow g(\tilde{\lambda}, \tilde{\nu}) = L(\tilde{x}, \tilde{\lambda}, \tilde{\nu})$

$\Rightarrow$ zero duality gap $\Rightarrow$ optimality

for convex problem where strong duality holds, KKT conditions are sufficient and necessary for optimality

Examples of Using KKT Conditions

equality constrained quadratic programming:

$$\begin{array}{ll} \text{minimize} & (1/2)x^T P x + q^T x + r \\ \text{subject to} & Ax = b \end{array}$$

with $P \in \mathbb{S}^n_{++}$

KKT conditions:

$$Ax^\star = b, \quad Px^\star + q + A^T \nu^\star = 0$$

KKT conditions rewritten:

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

optimization problem $\Leftrightarrow$ solving linear equations

Outline

Problem Reformulation Results in Different Dual Problems

problem reformulation $\Rightarrow$ easier or more useful dual problems

recall equivalent problem formulations (that preserve convexity):

- introduce new variables and equality constraints
- make implicit (resp. explicit) constraints explicit (resp. implicit)
- transformation of functions

Examples – Introducing New Variables and Equality Constraints

consider: minimize $f_0(Ax+b)$

- dual function: $g = \inf_x f_0(Ax+b) = p^*$
- finding dual function equivalent to solving primal problem

equivalent problem:

$$\begin{array}{ll} \text{minimize} & f_0(y) \\ \text{subject to} & y = Ax + b \end{array}$$

dual function:

$$g(\nu) = \begin{cases} \inf_y f_0(y) - \nu^T y + b^T \nu & A^T\nu = 0 \\ -\infty & \text{otherwise} \end{cases}$$

dual problem:

$$\begin{array}{ll} \text{maximize} & g(\nu) \\ \text{subject to} & A^T\nu = 0 \end{array}$$

Examples – Implicit Constraints

LP with box constraints: primal and dual problems

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

equivalent problem:

$$\begin{array}{ll} \text{minimize} & f_0(x) = \begin{cases} c^T x & -1 \leq x \leq 1 \\ \infty & \text{otherwise} \end{cases} \\ \text{subject to} & Ax = b \end{array}$$

dual function:

$$g(\nu) = \inf_{-1 \leq x \leq 1} c^T x + \nu^T (Ax - b) = -b^T \nu - \|A^T \nu + c\|_1$$

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