COMP 9602: Convex Optimization

Duality (III)

Dr. C. Wu

Department of Computer Science
The University of Hong Kong

Next

- Optimality conditions
  - Complementary slackness
  - KKT conditions

Complementary slackness

assume strong duality holds, $x^\star$ is primal optimal, $(\lambda^\star, \nu^\star)$ is dual optimal

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

hence, the two inequalities hold with equality

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

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

Complementary slackness for LP

- Primal and dual LPs $(x \in R^n, y \in R^m, A : m \times n)$

Primal: $\min c^T x$

subject to:

$$Ax \succeq b$$

Canonical form LP $x \succeq 0$

Dual: $\max b^T y$

subject to:

$$A^T y \preceq c$$

$$y \succeq 0$$

- Complementary slackness theorem for LP

A pair of feasible solutions $x \in R^n$ and $y \in R^m$ for primal and dual LP problems is optimal if and only if

$$y_i(b_i - (Ax)_i) = 0, \forall i = 1, \ldots, m$$

$$x_j(c_j - (A^T y)_j) = 0, \forall j = 1, \ldots, n$$

$$\Updownarrow$$

$$(Ax)_i = b_i, \text{ or } (Ax)_i > b_i \ \& \ y_i = 0, \forall i = 1, \ldots, m$$

$$(A^T y)_j = c_j, \text{ or } (A^T y)_j < c_j \ \& \ x_j = 0, \forall j = 1, \ldots, n$$

Karush-Kuhn-Tucker (KKT) conditions

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

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

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

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

- For any optimization problem with differentiable objective and constraint functions, if strong duality holds, any pair of primal and dual optimal points x^*, λ^*, ν^* must satisfy the KKT conditions

KKT conditions for convex programs

- For a convex program, any points $\tilde{x}, \tilde{\lambda}, \tilde{\nu}$ that satisfy the KKT conditions are primal and dual optimal, and have zero duality gap

=> If a convex program with differentiable objective and constraint functions satisfies Slater's conditions, then KKT condition is necessary and sufficient for optimality

- recall that Slater implies strong duality

Importance of KKT conditions

- KKT conditions are very important in optimization
  - possibly solve the KKT conditions to derive (analytical) solution for primal/dual problem
  - many algorithms for convex optimization can be interpreted as methods for solving the KKT conditions

A simple example

$$\min x_1^2 + 2x_2^2$$

subject to:

$$x_1 + x_2 \geq 3$$

To solve this, try to find (x^*, λ^*) that satisfy the KKT conditions

Water-filling examples

$$\min -\sum_{i=1}^{k} \log(1 + \frac{p_i}{N_i})$$

subject to:

$$\sum_{i=1}^{k} p_i \leq P$$

$$p_i \geq 0, \forall i = 1, \ldots, k$$

Another: Example 5.2, pp. 245, textbook

Water-filling examples: generalization

$$\min f(p)$$

subject to:

$$\sum_{i=1}^{k} p_i \leq P$$

$$p_i \geq 0, \forall i = 1, \ldots, k$$

where f(p) is convex and twice differentiable ($\bigtriangledown^2 f(p) \succ 0$), $\frac{\partial f(\mathbf{p})}{\partial p_i} < 0$ and invertible, $f(p)$ is separable in terms of each p_i ,

the same water-filling algorithm applies:

maximally allocate resource to pi with the current smallest marginal utility ($\frac{\partial f(\mathbf{p})}{\partial p_i}$,that increases with pi), until all resource is used up.

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

Example I

- Introducing new variables and equality constraints

minimize $\quad f_0(Ax+b)$

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

reformulated problem and its dual

$$\begin{array}{ll} \text{minimize} & f_0(y) \\ \text{subject to} & Ax+b-y=0 \end{array} \qquad\qquad \begin{array}{ll} \text{maximize} & b^T\nu - f_0^*(\nu) \\ \text{subject to} & A^T\nu = 0 \end{array}$$

Example:

$$\text{minimize } \|Ax-b\| \quad \Longleftrightarrow \quad \begin{array}{ll} \text{minimize} & \|y\| \\ \text{subject to} & y = Ax-b \end{array}$$

conjugate of norm on page 93, textbook

Example II

- Making explicit constraints implicit

LP with box constraints: primal and dual problem

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

reformulation with box constraints made implicit

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

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

Partial Lagrangian relaxation

- One can take Lagrangian with respect to only a subset of the constraints
  - e.g.,

$$\min f_0(x)$$

s.t.: $$f_i(x) \le 0, i = 1, 2$$

Lagrangian $$L(x, \lambda_1, \lambda_2) = f_0(x) + \lambda_1 f_1(x) + \lambda_2 f_2(x)$$

Lagrange dual $$g(\lambda_1, \lambda_2) = \min_x L(x, \lambda_1, \lambda_2)$$

Or

(partial) Lagrangian $$\tilde{L}(x, \lambda) = f_0(x) + \lambda f_1(x)$$

(partial) Lagrange dual $$\tilde{g}(\lambda) = \min_x \tilde{L}(x, \lambda) \text{ s.t. } f_2(x) \le 0$$

If convex problem and Slater's condition is satisfied:

$$\max_{\lambda \succeq 0} \tilde{g}(\lambda) = \min_{f_i(x) \le 0} f_0(x)$$

The above property extensible to general case, when the partial lagrangian is derived by relaxing any subset of the constraints of a convex program

Partial Lagrangian relaxation (cont'd)

- Example:
$$\min \sum_{(i,j)\in E} c_{ij}x_{ij}$$

subject to:

$$\sum_{j:(i,j)\in E} x_{ij} - \sum_{j:(j,i)\in E} x_{ji} = \begin{cases} v, \text{for } i = s, \\ 0, \text{for all } i \in V - \{s,t\}, \\ -v, \text{for } i = t, \end{cases} \qquad (1)$$

$$0 \le x_{ij} \le u_{ij}, \forall (i,j) \in E \qquad (2)$$

$$\sum_{j:(i,j)\in E} x_{ij} \le O_i, \forall i \in V \qquad \lambda_i$$

$$\sum_{j:(j,i)\in E} x_{ji} \le I_i, \forall i \in V \qquad \mu_i$$

partial Lagrangian

$$\widetilde{L}(\mathrm{x},\lambda,\mu) = \sum_{(i,j)\in E} c_{ij}x_{ij} + \sum_{i\in V} \lambda_i \left(\sum_{j:(i,j)\in E} x_{ij} - O_i\right)$$
$$+ \sum_{i\in V} \mu_i \left(\sum_{j:(j,i)\in E} x_{ji} - I_i\right)$$

Partial Lagrangian relaxation (cont'd)

partial Lagrange dual:

$$\widetilde{g}(\lambda,\mu) = \min_{\mathrm{x}} \widetilde{L}(x,\lambda,\mu)$$

s.t.: (1) (2)

dual problem:

$$\max \quad \widetilde{g}(\lambda,\mu)$$

s.t.:

$$\lambda \succeq 0$$
$$\mu \succeq 0$$

Example III

- Transform objective function

$$\text{minimize} \quad \| Ax - b \|$$

reformulation

$$\text{minimize} \ \frac{1}{2} \| y \|^2$$

$$\text{subject to} \quad y = Ax - b$$

dual problem on page 257, textbook

Generalized inequality

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

where $K_i \subseteq R^{k_i}$ are proper cones

Definitions for Lagrange dual are parallel to scalar case:

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

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

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

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

Generalized inequality

dual problem

$$\begin{array}{ll} \text{maximize} & g(\lambda_1, \ldots, \lambda_m, \nu) \\ \text{subject to} & \lambda_i \succeq_{K_i^*} 0, \quad i = 1, \ldots, m \end{array}$$

Properties:

- **lower bound property:** if $\lambda_i \succeq_{K_i^*} 0$, then $g(\lambda_1, \ldots, \lambda_m, \nu) \leq p^\star$
- weak duality: $p^\star \geq d^\star$ always
- strong duality: $p^\star = d^\star$ for convex problem with constraint qualification (for example, Slater's: primal problem is strictly feasible)
- similarly complementary slackness, KKT conditions also hold. (pp. 267, textbook)

Example

- Semidefinite program

primal SDP $(F_i, G \in \mathbf{S}^k)$

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & x_1 F_1 + \cdots + x_n F_n \preceq G \end{array}$$

- Lagrange multiplier is matrix $Z \in \mathbf{S}^k$, define $< A, B >= \mathbf{tr}(AB)$
- Lagrangian $L(x, Z) = c^T x + \mathbf{tr}\left(Z(x_1 F_1 + \cdots + x_n F_n - G)\right)$
- dual function

$$g(Z) = \inf_x L(x, Z) = \begin{cases} -\mathbf{tr}(GZ) & \mathbf{tr}(F_i Z) + c_i = 0, \quad i = 1, \ldots, n \\ -\infty & \text{otherwise} \end{cases}$$

dual problem

$$\begin{array}{ll} \text{maximize} & -\mathbf{tr}(GZ) \\ \text{subject to} & Z \succeq 0, \quad \mathbf{tr}(F_i Z) + c_i = 0, \quad i = 1, \ldots, n \end{array}$$

$p^\star = d^\star$ if primal SDP is strictly feasible ($\exists x$ with $x_1 F_1 + \cdots + x_n F_n \prec G$)

- Reference
  - Chapter 5.5, 5.7, 5.9, Convex Optimization.
- Acknowledgement
  - Some materials are extracted from the slides created by Prof. Stephen Boyd for the textbook
  - Some materials are extracted from the lecture notes of Convex Optimization by Prof. Wei Yu at the University of Toronto