ECE236B Discussion 6

Xin Jiang

Department of Electrical and Computer Engineering, UCLA

ECE236B Convex Optimization (Spring 2018)

May 11, 2018

Problem setup

standard form problem

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

the domain of primal problem $\mathcal{D} = \cap_{i=0}^{m} \mathbf{dom}\, f_i \cap \cap_{i=1}^{p} \mathbf{dom}\, h_i$

Lagrangian $L\colon \mathbf{R}^m \times \mathbf{R}^m \times \mathbf{R}^p \to \mathbf{R}$, with $\mathbf{dom}\, L = \mathcal{D} \times \mathbf{R}^m \times \mathbf{R}^p$:

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

Lagrange dual function: the infimum of $L(x, \lambda, \nu)$

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

Deriving the dual and weak duality

common tricks on deriving the dual

- introduce new variables and equality constraints
- make explicit constraints implicit, or vice versa
- transform objective or constraint functions

weak duality

$$d^\star \leq p^\star$$

- dual feasible: $\lambda \succeq 0$ and $(\lambda, \nu) \in \mathbf{dom}\, g$
- weak duality always holds, even for nonconvex problems
- weak duality can be meaningless when $g(\lambda, \nu) = -\infty$

Two-way partition example

$$\begin{array}{ll} \text{minimize} & x^T W x \\ \text{subject to} & x_i^2 = 1 \end{array} \quad \text{(primal)} \qquad \begin{array}{ll} \text{maximize} & g(\nu) = -\mathbf{1}^T \nu \\ \text{subject to} & W + \mathbf{diag}(\nu) \succeq 0 \end{array} \quad \text{(dual)}$$

$$\begin{array}{ll} \text{minimize} & x^T W x \\ \text{subject to} & \sum x_i^2 = n \end{array} \quad \text{(relax)} \qquad \begin{array}{ll} \text{minimize} & \mathbf{tr}(WX) \\ \text{subject to} & X_{ii} = 1, \ X \succeq 0 \end{array} \quad \text{(dual 2)}$$

$$\begin{array}{ll} \text{minimize} & \mathbf{tr}(WX) \\ \text{subject to} & X_{ii} = 1 \\ & X \succeq 0, \ \mathbf{rank}\, X = 1 \end{array} \quad \text{(primal SDP)}$$

- $p^\star$, $d^\star$, $q^\star$ are the optimal values for (primal), (dual), and (relax), respectively
- $\tilde{\nu} = -\lambda_{\min}(W) \cdot \mathbf{1}$ is feasible for (dual), with $g(\tilde{\nu}) = n\lambda_{\min}(W)$

$$p^\star \geq d^\star \geq g(\tilde{\nu}) = q^\star$$

- (dual 2) is a relaxation of (primal) or (primal SDP)

Strong duality and Slater's condition

now consider a convex optimization problem

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

- strong duality: $p^\star = d^\star$
- convex problems do not always have strong duality, and problems with strong duality are not always convex problems
- convex problems + constraint qualification $\Longrightarrow$ strong duality
- Slater's theorem: for a convex problem, if there exists x with $f_i(x) < 0$, then
  1. strong duality holds, *i.e.*, $p^\star = d^\star$;
  2. furthermore, if $p^\star > -\infty$, the dual optimum is attained, *i.e.*, there exists $(\lambda^\star, \nu^\star)$ with $g(\lambda^\star, \nu^\star) = d^\star = p^\star$

Example: convex problem vs. strong duality

a convex problem without strong duality (T5.21)

$$\begin{array}{ll} \text{minimize} & e^{-x} \\ \text{subject to} & f_1(x,y) \le 0 \end{array}$$

$$\begin{array}{ll} \text{maximize} & 0 \\ \text{subject to} & \lambda \ge 0 \end{array}$$

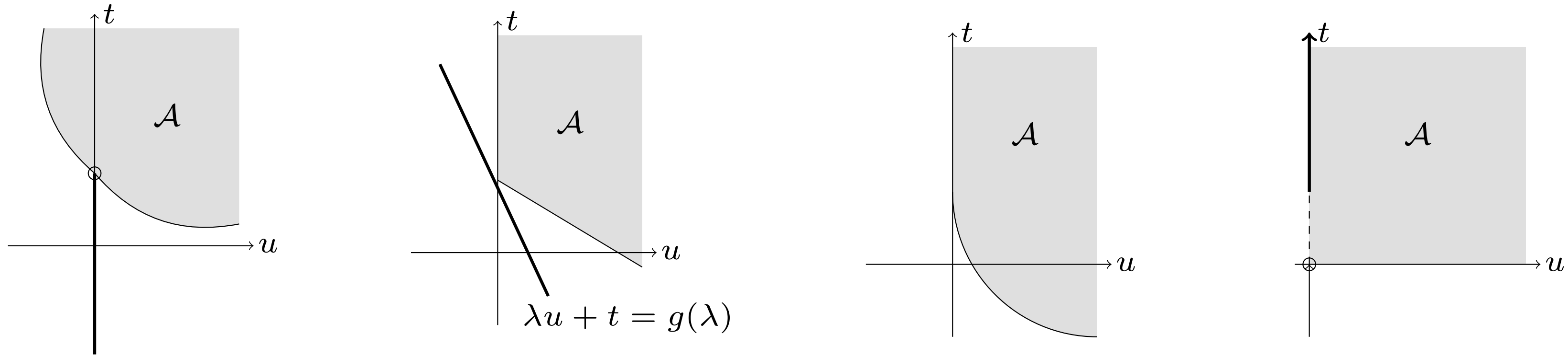
where $f_1(x,y) = \begin{cases} x^2/y \le 0, & y > 0 \\ \infty, & \text{otherwise} \end{cases}$

a nonconvex problem with strong duality

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

$$\begin{array}{ll} \text{maximize} & -t - \lambda \\ \text{subject to} & \begin{bmatrix} A + \lambda I & b \\ b^T & t \end{bmatrix} \succeq 0 \end{array}$$

Geometric interpretation



$$\mathcal{A} = \{(u,t) \in \mathbf{R}^2 \mid f_1(x) \le u,\ f_0(x) \le t \text{ for some } x \in \mathcal{D}\}$$

- for convex problems, $\mathcal{A}$ is convex; there is a supporting hyperplane at $(0, p^\star)$
- Slater's condition: if there exists a pair of $(\tilde{u}, \tilde{t}) \in \mathcal{A}$ with $\tilde{u} < 0$, then supporting hyperplanes at $(0, p^\star)$ must be non-vertical

Slater's condition	$\surd$	$\times$	$\times$	$\times$
strong duality	$\surd$	$\surd$	$\surd$	$\times$
dual optimal attained	$\surd$	$\surd$	$\times$	$\surd$

Karush–Kuhn–Tucker conditions

if the pair $(\tilde{x}, \tilde{\lambda}, \tilde{\nu})$ satisfies KKT conditions, then it means

1° primal constraints: $f_i(\tilde{x}) \leq 0,\ i = 1, \ldots, m,\ h_i(\tilde{x}) = 0,\ i = 1, \ldots, p$

2° dual constraints: $\tilde{\lambda} \succeq 0$

3° complementary slackness: $\tilde{\lambda}_i f_i(\tilde{x}) = 0,\ i = 1, \ldots, m$

4°(a) $\tilde{x}$ minimizes $L(x, \tilde{\lambda}, \tilde{\nu})$ over x

(b) gradient of Lagrangian with respect to x vanishes

$$\nabla f_0(\tilde{x}) + \sum_{i=1}^{m} \tilde{\lambda}_i \nabla f_i(\tilde{x}) + \sum_{i=1}^{p} \tilde{\nu}_i h_i(\tilde{x}) = 0$$

primal and dual optimal, strong duality $\iff$ KKT condition with 4(a)

if the problem is convex, differentiable, and satisfies Slater's condition

primal and dual optimal $\iff$ KKT condition with 4(b)

Example: KKT condition vs. optimality

KKT (with 4b) is not sufficient for optimality in a nonconvex problem (T5.29)

$$\begin{array}{ll} \text{minimize} & -3x_1^2 + x_2^2 + 2x_3^2 + 2(x_1 + x_2 + x_3) \\ \text{subject to} & x_1^2 + x_2^2 + x_3^2 = 1 \end{array}$$

KKT condition:

$$\frac{1}{(-3+\nu)^2} + \frac{1}{(1+\nu)^2} + \frac{1}{(2+\nu)^2} = 1$$

this is also an example of solving the problem via optimality conditions

KKT is not necessary for optimality when Slater's condition fails (T5.26)

$$\begin{array}{ll} \text{minimize} & x_1^2 + x_2^2 \\ \text{subject to} & (x_1 - 1)^2 + (x_2 - 1)^2 \le 1 \\ & (x_1 - 1)^2 + (x_2 + 1)^2 \le 1 \end{array}$$

Slater's condition fails, but strong duality holds and dual optimal is not attained

Example: projection onto ℓ_1-norm ball

(A4.5) derive the dual problem and describe an efficient method for solving it

$$\begin{array}{ll} \text{minimize} & (1/2)\|x-a\|_2^2 \\ \text{subject to} & \|x\|_1 \leq 1 \end{array}$$

the Lagrangian is

$$L(x,\lambda) = \frac{1}{2}\|x-a\|_2^2 + \lambda(\|x\|_1 - 1) = \frac{1}{2}\sum_{k=1}^{n}((x_k - a_k)^2 + 2\lambda|x_k|) - \lambda$$

for a given λ, the minimizer $\tilde{x}$ has elements:

$$\tilde{x}_k = \max\{|a_k| - \lambda^\star, 0\} = \begin{cases} a_k - \lambda & a_k \geq \lambda \\ 0 & |a_k| \leq \lambda \\ a_k + \lambda & a_k \leq -\lambda \end{cases}$$

Example: projection onto ℓ_1-norm ball

$(x^\star, \lambda^\star, \nu^\star)$ are primal–dual optimal if and only if they satisfy the KKT conditions:

1. primal feasibility: $\|x^\star\|_1 \leq 1$
2. dual feasibility: $\lambda^\star \geq 0$
3. complementary slackness: $\lambda(1 - \|x\|_1) = 0$
4. $x^\star$ is the minimizer of $L(x, \lambda^\star)$:

$$x_k^\star = \begin{cases} a_k - \lambda^\star & a_k \geq \lambda^\star \\ 0 & |a_k| \leq \lambda^\star \\ a_k + \lambda^\star & a_k \leq -\lambda^\star \end{cases}$$

Example: projection onto ℓ_1-norm ball

- $\lambda^\star = 0$ only if $\|a\|_1 \leq 1$
- $\lambda^\star > 0$ only if it satisfies (assuming $|a_1| \leq |a_2| \leq \cdots \leq |a_n|$)

$$\|x^\star\|_1 = \sum_{k=1}^{n} \max\{|a_k| - \lambda^\star, 0\} = 1$$