The University of Texas at Austin
Department of Electrical and Computer Engineering

EE381K: Convex Optimization — Fall 2019

Lecture 19

Aryan Mokhtari Wednesday, November 6, 2019.

Goal: In this lecture, we talk about weak and strong duality, Slater's condition, and the KKT conditions.

1 Weak and strong duality

Weak duality implies that $d^* \leq p^*$, which follows directly from the fact that the dual function always provides a lower bound for p^*. This condition always holds.

Strong duality means $d^* = p^*$, i.e., the optimal duality gap $p^* - d^*$ is zero. This condition does not hold in general. For *convex optimization problems*, it usually (not always) holds.

Conditions that guarantee strong duality in convex problems are called constraint qualifications.

2 Slater's condition

A simple way to check if strong duality holds for a convex program is Slater's condition. The Slater's theorem states that for the general convex program of the form

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

strong duality holds if it is strictly feasible, i.e.,

$$\exists\ \mathbf{x} \qquad \text{such that} \quad f_i(\mathbf{x}) < 0, \quad i = 1, \ldots, m, \qquad \mathbf{A}\mathbf{x} = \mathbf{b}.$$

Such a point is called strictly feasible.
If some of the inequality constraint functions $f_1, \ldots, f_k$ are affine the strong duality holds under weaker condition:

$$\exists\ \mathbf{x} \qquad \text{such that} \quad f_i(\mathbf{x}) \leq 0, \quad i = 1, \ldots, k, \qquad f_i(\mathbf{x}) < 0, \quad i = k+1, \ldots, m, \qquad \mathbf{A}\mathbf{x} = \mathbf{b}.$$

The affine inequalities do not need to hold with strict inequality!

2.1 Examples

Least-squares solution for linear equations: For the following problem

$$\begin{aligned} \min: \quad & \mathbf{x}^\top \mathbf{x} \\ \text{s.t.}: \quad & \mathbf{A}\mathbf{x} = \mathbf{b}. \end{aligned}$$

we showed that the dual problem is

$$\max : -\frac{1}{4}\boldsymbol{\nu}^\top \mathbf{A}\mathbf{A}^\top \boldsymbol{\nu} - \mathbf{b}^\top \boldsymbol{\nu}$$

In this case, according to Slater's condition strong duality holds if the primal problem is feasible, i.e., there exists a point $\mathbf{x}$ such that $\mathbf{A}\mathbf{x} = \mathbf{b}$.

Minimum volume covering ellipsoid: For this problem which is defined as

$$\begin{aligned} \min : \quad & f_0(\mathbf{X}) := \log(\det(\mathbf{X}^{-1}))) \\ \text{s.t.} : \quad & \mathbf{a}_i^\top \mathbf{X}\mathbf{a}_i \leq 1, \quad i = 1, \ldots, m. \end{aligned}$$

and its dual is given by

$$\begin{aligned} \max : \quad & g(\boldsymbol{\lambda}) = n - \boldsymbol{\lambda}^\top \mathbf{1} + \log\left(\det\left(\sum_{i=1}^m \lambda_i \mathbf{a}_i \mathbf{a}_i^\top\right)\right) \\ \text{s.t.} : \quad & \boldsymbol{\lambda} \geq \mathbf{0}, \qquad \sum_{i=1}^m \lambda_i \mathbf{a}_i \mathbf{a}_i^\top \succ \mathbf{0}. \end{aligned}$$

In this case, the inequality constraints for the primal problem are all affine functions and therefore based on weak Slater's condition strong duality holds if the primal problem is feasible, i.e., there exists $\mathbf{X} \in \mathbf{S}_{++}^n$ such that $\mathbf{a}_i^\top \mathbf{X}\mathbf{a}_i \leq 1, \quad i = 1, \ldots, m$. This is always satisfied and strong duality always holds.

A nonconvex problem with strong duality: Consider the following quadratic program

$$\begin{aligned} \min : \quad & \mathbf{x}^\top \mathbf{A}\mathbf{x} + 2\mathbf{b}^\top \mathbf{x} \\ \text{s.t.} : \quad & \mathbf{x}^\top \mathbf{x} \leq 1 \end{aligned}$$

where $\mathbf{A}$ is symmetric ($\mathbf{A} \in \mathbf{S}^n$) but not positive semidefinite. In this case, the objective function is not convex. The Lagrangian of this problem is given by

$$\begin{aligned} L(\mathbf{x}, \boldsymbol{\lambda}) &= \mathbf{x}^\top \mathbf{A}\mathbf{x} + 2\mathbf{b}^\top \mathbf{x} + \lambda(\mathbf{x}^\top \mathbf{x} - 1) \\ &= \mathbf{x}^\top (\mathbf{A} + \lambda \mathbf{I})\mathbf{x} + 2\mathbf{b}^\top \mathbf{x} - \lambda \end{aligned}$$

and therefore the dual function is

$$g(\lambda) = \inf_{\mathbf{x}} L(\mathbf{x}, \lambda) = \begin{cases} -\mathbf{b}^\top (\mathbf{A} + \lambda \mathbf{I})^\dagger \mathbf{b} - \lambda & \text{if } \ \mathbf{A} + \lambda \mathbf{I} \succeq \mathbf{0} \text{ and } \mathbf{b} \in \mathcal{R}(\mathbf{A} + \lambda \mathbf{I}) \\ -\infty & \text{otherwise} \end{cases}$$

where $(\mathbf{A} + \lambda \mathbf{I})^\dagger$ is the pseudo-inverse of $\mathbf{A} + \lambda \mathbf{I}$. In this case, the dual problem is

$$\begin{aligned} \max : \quad & -\mathbf{b}^\top (\mathbf{A} + \lambda \mathbf{I})^\dagger \mathbf{b} - \lambda \\ \text{s.t.} : \quad & \mathbf{A} + \lambda \mathbf{I} \succeq \mathbf{0}, \quad \mathbf{b} \in \mathcal{R}(\mathbf{A} + \lambda \mathbf{I}), \quad \lambda \geq \mathbf{0}. \end{aligned}$$

It can be shown that this is a convex problem. In fact, the duality gap of this problem is zero. The proof is available in Appendix B.4 of the textbook (uses Theorem of Alternatives).

3 Complementary slackness

Suppose we face a problem for which **strong duality** holds. In this case, if $\mathbf{x}^*$ primal optimal and $(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$ is dual optimal then we have

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

Hence, the inequalities are are equalities and we can conclude that

1. $\mathbf{x}^*$ minimizes the Lagrangian for optimal dual variables, i.e., $\mathbf{x}^*$ minimizes $L(\mathbf{x}, \boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$
2. complementary slackness holds, i.e., $\lambda_i^* f_i(\mathbf{x}^*) = 0$ for $i = 1, \ldots, m$. Indeed, this condition implies that
$$\lambda_i^* > 0 \;\Rightarrow\; f_i(\mathbf{x}^*) = 0, \qquad f_i(\mathbf{x}^*) < 0 \;\Rightarrow\; \lambda_i = 0.$$

4 Karush-Kuhn-Tucker (KKT) conditions

For a problem with differentiable functions (possibly nonconvex), the following four conditions are called KKT conditions:

1. primal constraints $f_i(\mathbf{x}) \leq 0$ for $i = 1, \ldots, m$ and $h_i(\mathbf{x}) = 0$ for $i = 1, \ldots, p$ are satisfied.
2. $\lambda_i \geq 0$ for $i = 1, \ldots, m$.
3. complementary slackness holds, i.e., $\lambda_i f_i(\mathbf{x}) = 0$ for $i = 1, \ldots, m$.
4. gradient of the Lagrangian with respect to $\mathbf{x}$ is zero:
$$\nabla f_0(\mathbf{x}) + \sum_{i=1}^m \lambda_i \nabla f_i(\mathbf{x}) + \sum_{i=1}^p \nu_i \nabla h_i(\mathbf{x}) = \mathbf{0}.$$

Theorem 1 (KKT conditions under strong duality)**.** *Consider a general optimization problem for which strong duality holds. If $\tilde{\mathbf{x}}$ is primal optimal and $(\tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\nu}})$ is dual optimal, then $\tilde{\mathbf{x}}$, $\tilde{\boldsymbol{\lambda}}$, and $\tilde{\boldsymbol{\nu}}$ satisfy the KKT conditions.*

Proof. Since $\tilde{\mathbf{x}}$ is primal optimal we know it is feasible and therefore condition 1 is satisfied. Since $(\tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\nu}})$ is dual optimal, $\tilde{\boldsymbol{\lambda}}$ is dual feasible and condition 2 is satisfied. Since strong duality holds, we know that $\tilde{\mathbf{x}}$ minimizes the Lagrangian for optimal dual variables, i.e., $\tilde{\mathbf{x}}$ minimizes $L(\mathbf{x}, \tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\nu}})$ and therefore its gradient should be zero (first-order optimality condition) and therefore condition 4 is satisfied. Again, since strong duality holds complementary slackness holds and condition 3 is satisfied. Hence, if strong duality holds and $\tilde{\mathbf{x}}$, $\tilde{\boldsymbol{\lambda}}$, and $\tilde{\boldsymbol{\nu}}$ are optimal, they must satisfy KKT conditions. ∎

Remark 1. *This result shows that* ***under strong duality****, satisfying the KKT conditions is necessary for optimality (even for problems that are not convex).*

Theorem 2 (KKT conditions for convex problems)**.** *Consider a general convex optimization problem. If* $\tilde{\mathbf{x}}$, $\tilde{\boldsymbol{\lambda}}$, *and* $\tilde{\boldsymbol{\nu}}$ *satisfy the KKT conditions, then* $\tilde{\mathbf{x}}$ *is primal optimal and* $(\tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\nu}})$ *is dual optimal, and the duality gap for these points is zero.*

Proof. Suppose that $\tilde{\mathbf{x}}$, $\tilde{\boldsymbol{\lambda}}$, and $\tilde{\boldsymbol{\nu}}$ satisfy the KKT conditions. Then, according to the first condition $\tilde{\mathbf{x}}$ is primal feasible and according to the second condition $\tilde{\boldsymbol{\lambda}}$ is dual feasible.
Further, the second condition also implies that the Lagrangian $L(\mathbf{x}, \tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\nu}})$ is convex in $\mathbf{x}$, cause it can be written as a sum of finite number of convex functions ($\lambda_i \geq 0$ and therefore $\sum_{i=1}^{m} \tilde{\lambda}_i f_i(\mathbf{x})$ is a convex function) and affine functions (in convex optimization $h_i(\tilde{\mathbf{x}})$ are affine functions).
Note that since $L(\mathbf{x}, \tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\nu}})$ is convex, by setting its gradient to zero we obtain its optimal solution. Hence, the last KKT condition implies that $\tilde{\mathbf{x}}$ is the minimizer of $L(\mathbf{x}, \tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\nu}})$, as $\nabla_{\mathbf{x}} L(\tilde{\mathbf{x}}, \tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\nu}}) = \mathbf{0}$.
Considering these observations we can write that

$$\begin{aligned} g(\tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\nu}}) &= \inf_{\mathbf{x}} \Big(L(\mathbf{x}, \tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\nu}}) \Big) \\ &= L(\tilde{\mathbf{x}}, \tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\nu}}) \\ &= f_0(\tilde{\mathbf{x}}) + \sum_{i=1}^{m} \tilde{\lambda}_i f_i(\tilde{\mathbf{x}}) + \sum_{i=1}^{p} \tilde{\nu}_i h_i(\tilde{\mathbf{x}}) \\ &= f_0(\tilde{\mathbf{x}}) \end{aligned}$$

where the last equality follows from the fact that $\tilde{\mathbf{x}}$ is primal feasible and $h_i(\tilde{\mathbf{x}}) = 0$ (based on the first KKT condition) and the fact that complementary slackness holds $\tilde{\lambda}_i f_i(\tilde{\mathbf{x}}) = 0$ (based on the third KKT condition).
Hence, the feasible points $\tilde{\mathbf{x}}$, $\tilde{\boldsymbol{\lambda}}$, and $\tilde{\boldsymbol{\nu}}$ which satisfy the KKT conditions have zero duality gap and therefore they are primal dual optimal. □

Remark 2. *Theorem 2 shows that satisfying the KKT conditions is a sufficient condition for optimality in* ***convex programs****. In the following theorem, we show that for convex programs that satisfy Slater's condition, satisfying the KKT conditions is necessary and sufficient for optimality.*

Theorem 3 (KKT conditions for convex problems under Slater's condition)**.** *Consider a general convex optimization problem which satisfies Slater's condition. Then,* $\tilde{\mathbf{x}}$, $\tilde{\boldsymbol{\lambda}}$, *and* $\tilde{\boldsymbol{\nu}}$ *satisfy the KKT conditions if and only if* $\tilde{\mathbf{x}}$ *is primal optimal and* $(\tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\nu}})$ *is dual optimal, and the duality gap for these points is zero.*

Proof. According to Theorem 2, when $\tilde{\mathbf{x}}$, $\tilde{\boldsymbol{\lambda}}$, and $\tilde{\boldsymbol{\nu}}$ satisfy the KKT conditions, then they are primal dual optimal with zero duality gap. Hence, satisfying KKT implies optimality.
When Slater's condition holds for convex programs then we have strong duality. We further know from Theorem 1 when we have strong duality optimal solutions satisfy the KKT conditions. Hence, optimality implies satisfying KKT. □

Example 1. *Consider the following convex optimization problem*

$$\text{minimize} \;\; \frac{1}{2}\mathbf{x}^\top \mathbf{P}\mathbf{x} + \mathbf{q}^\top \mathbf{x} + r, \qquad \text{subject to} \;\; \mathbf{A}\mathbf{x} = \mathbf{b},$$

where $\mathbf{P} \in \mathbf{S}_{+}^{n}$. *The KKT conditions for this problem are*

$$\mathbf{A}\mathbf{x}^* = \mathbf{b}, \qquad \mathbf{P}\mathbf{x}^* + \mathbf{q} + \mathbf{A}^\top \boldsymbol{\nu}^* = \mathbf{0}$$

which can be written as the solution of the following system of equations

$$\begin{bmatrix} \mathbf{P} & \mathbf{A}^\top \\ \mathbf{A} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{x}^* \\ \boldsymbol{\nu}^* \end{bmatrix} = \begin{bmatrix} -\mathbf{q} \\ \mathbf{b} \end{bmatrix}$$