Fundamentals of Linear Algebra and Optimization
Weak and Strong Duality

Jean Gallier and Jocelyn Quaintance

CIS Department
University of Pennsylvania

jean@cis.upenn.edu

November 23, 2020

Dual Bounds Primal Problem (P)

Another important property of the dual function G is that it provides a *lower bound* on the value of the objective function J.

Dual Bounds Primal Problem (P)

Another important property of the dual function G is that it provides a *lower bound* on the value of the objective function J.

Indeed, we have

$$G(\mu) \leq L(u, \mu) \leq J(u) \quad \text{for all } u \in U \text{ and all } \mu \in \mathbb{R}^m_+, \qquad (\dagger)$$

since $\mu \geq 0$ and $\varphi_i(u) \leq 0$ for $i = 1, \ldots, m$, so

$$G(\mu) = \inf_{v \in \Omega} L(v, \mu) \leq L(u, \mu) = J(u) + \sum_{i=1}^{m} \mu_i \varphi_i(u) \leq J(u).$$

Weak Duality

If the Primal Problem (P) has a minimum denoted p^* and the Dual Problem (D) has a maximum denoted d^*, then the above inequality implies that

$$d^* \leq p^* \qquad (\dagger_w)$$

known as *weak duality*.

Weak Duality Restated

Equivalently, for every optimal solution λ^* of the dual problem and every optimal solution u^* of the primal problem, we have

$$G(\lambda^*) \leq J(u^*). \qquad (\dagger_{w'})$$

Weak Duality Restated

Equivalently, for every optimal solution λ^* of the dual problem and every optimal solution u^* of the primal problem, we have

$$G(\lambda^*) \leq J(u^*). \qquad (\dagger_{W'})$$

In particular, if $p^* = -\infty$, which means that the primal problem is unbounded below, then the dual problem is unfeasible.

Weak Duality Restated

Equivalently, for every optimal solution λ^* of the dual problem and every optimal solution u^* of the primal problem, we have

$$G(\lambda^*) \leq J(u^*). \qquad (\dagger_{W'})$$

In particular, if $p^* = -\infty$, which means that the primal problem is unbounded below, then the dual problem is unfeasible.

Conversely, if $d^* = +\infty$, which means that the dual problem is unbounded above, then the primal problem is unfeasible.

Strong Duality

Definition. The difference $p^* - d^* \geq 0$ is called the *optimal duality gap*. If the duality gap is zero, that is, $p^* = d^*$, then we say that *strong duality* holds.

Strong Duality

Definition. The difference $p^* - d^* \geq 0$ is called the *optimal duality gap*. If the duality gap is zero, that is, $p^* = d^*$, then we say that *strong duality* holds.

Even when the duality gap is strictly positive, the inequality $(\dagger_w)$ can be helpful to find a lower bound on the optimal value of a primal problem that is difficult to solve, since the dual problem is *always* convex.

Strong Duality

Definition. The difference $p^* - d^* \geq 0$ is called the *optimal duality gap*. If the duality gap is zero, that is, $p^* = d^*$, then we say that *strong duality* holds.

Even when the duality gap is strictly positive, the inequality $(\dagger_w)$ can be helpful to find a lower bound on the optimal value of a primal problem that is difficult to solve, since the dual problem is *always* convex.

If the primal problem and the dual problem are feasible and if the optimal values p^* and d^* are finite and $p^* = d^*$ (no duality gap), then the complementary slackness conditions hold for the inequality constraints.

Complementary Slackness Conditions

Proposition (*complementary slackness*). Given the Minimization Problem (P)

$$\begin{aligned} &\text{minimize} \quad J(v) \\ &\text{subject to} \quad \varphi_i(v) \leq 0, \quad i = 1, \ldots, m, \end{aligned}$$

and its Dual Problem (D)

$$\begin{aligned} &\text{maximize} \quad G(\mu) \\ &\text{subject to} \quad \mu \in \mathbb{R}^m_+, \end{aligned}$$

Complementary Slackness Conditions

if both (P) and (D) are feasible, $u \in U$ is an optimal solution of (P), $\lambda \in \mathbb{R}^m_+$ is an optimal solution of (D), and $J(u) = G(\lambda)$, then

$$\sum_{i=1}^{m} \lambda_i \varphi_i(u) = 0.$$

In other words, *if the constraint φ_i is inactive at u, then $\lambda_i = 0$*.

Weak Duality for Linear Programming

Going back to the example of the last lesson, we see that weak duality says that for any feasible solution u of the Primal Problem (P), that is, some $u \in \mathbb{R}^n$ such that

$$Au \leq b, \;\; u \geq 0,$$

and for any feasible solution $\mu \in \mathbb{R}^m$ of the Dual Problem (D_1), that is,

$$A^\top \mu \geq -c, \;\; \mu \geq 0,$$

we have

$$-b^\top \mu \leq c^\top u.$$

Weak Duality for Linear Programming

Actually, if u and λ are *optimal*, then it can be shown that *strong duality* holds, namely $-b^\top \mu = c^\top u$, but the proof of this fact is nontrivial.

Duality Gap

The following theorem establishes a link between the solutions of the Primal Problem (P) and those of the Dual Problem (D). It also gives *sufficient conditions* for the *duality gap to be zero*.

Duality Gap Theorem

Theorem. Consider the Minimization Problem (P):

$$\begin{aligned} &\text{minimize} \quad J(v) \\ &\text{subject to} \quad \varphi_i(v) \leq 0, \quad i = 1, \ldots, m, \end{aligned}$$

where the functions J and φ_i are defined on some open subset Ω of a finite-dimensional Euclidean vector space V (more generally, a real Hilbert space V).

Duality Gap Theorem

(1) Suppose the functions $\varphi_i \colon \Omega \to \mathbb{R}$ are continuous, and that for every $\mu \in \mathbb{R}^m_+$, the Problem (P_μ):

$$\begin{aligned} &\text{minimize} && L(v, \mu) \\ &\text{subject to} && v \in \Omega, \end{aligned}$$

has a *unique solution* u_μ, so that

$$L(u_\mu, \mu) = \inf_{v \in \Omega} L(v, \mu) = G(\mu),$$

Duality Gap Theorem

and the function $\mu \mapsto u_\mu$ is continuous (on $\mathbb{R}^m_+$).

Duality Gap Theorem

and the function $\mu \mapsto u_\mu$ is continuous (on $\mathbb{R}^m_+$). Then the function G is differentiable for all $\mu \in \mathbb{R}^m_+$, and

$$G'_\mu(\xi) = \sum_{i=1}^{m} \xi_i \varphi_i(u_\mu) \quad \text{for all } \xi \in \mathbb{R}^m.$$

Duality Gap Theorem

and the function $\mu \mapsto u_\mu$ is continuous (on $\mathbb{R}^m_+$). Then the function G is differentiable for all $\mu \in \mathbb{R}^m_+$, and

$$G'_\mu(\xi) = \sum_{i=1}^{m} \xi_i \varphi_i(u_\mu) \quad \text{for all } \xi \in \mathbb{R}^m.$$

If λ is any solution of Problem (D):

$$\begin{aligned} &\text{maximize} && G(\mu) \\ &\text{subject to} && \mu \in \mathbb{R}^m_+, \end{aligned}$$

then the solution u_λ of the corresponding Problem (P_λ) is a solution of Problem (P).

Duality Gap Theorem

(2) Assume Problem (P) has some solution $u \in U$, and that Ω is convex (open), the functions φ_i $(1 \leq i \leq m)$ and J are convex and differentiable at u, and that the constraints are qualified. Then Problem (D) has a solution $\lambda \in \mathbb{R}^m_+$, and $J(u) = G(\lambda)$; that is, *the duality gap is zero*.

Duality Gap Theorem

Informally, in Part (1) of the preceding theorem, the hypotheses say that if $G(\mu)$ can be "computed nicely," in the sense that there is a *unique minimizer* u_μ of $L(v, \mu)$ (with $v \in \Omega$) such that $G(\mu) = L(u_\mu, \mu)$, and if a maximizer λ of $G(\mu)$ (with $\mu \in \mathbb{R}^m_+$) can be determined, then u_λ yields the minimum value of J, that is, $p^* = J(u_\lambda)$.

Duality Gap Theorem

Informally, in Part (1) of the preceding theorem, the hypotheses say that if $G(\mu)$ can be "computed nicely," in the sense that there is a *unique minimizer* u_μ of $L(v, \mu)$ (with $v \in \Omega$) such that $G(\mu) = L(u_\mu, \mu)$, and if a maximizer λ of $G(\mu)$ (with $\mu \in \mathbb{R}^m_+$) can be determined, then u_λ yields the minimum value of J, that is, $p^* = J(u_\lambda)$.

If the constraints are qualified and if the functions J and φ_i are convex and differentiable, then since the KKT conditions hold, the duality gap is zero; that is,

$$G(\lambda) = L(u_\lambda, \lambda) = J(u_\lambda).$$

Duality Gap of a Linear Program

Example. Going back to the example of the previous lesson where we considered the Linear Program (P)

$$\begin{aligned} &\text{minimize} \quad c^\top v \\ &\text{subject to} \quad Av \leq b, \quad v \geq 0, \end{aligned}$$

with A an $m \times n$ matrix, the Lagrangian $L(v, \mu, \nu)$ is given by

$$L(v, \mu, \nu) = -b^\top \mu + (c + A^\top \mu - \nu)^\top v,$$

and we found that the dual function $G(\mu, \nu) = \inf_{v \in \mathbb{R}^n} L(v, \mu, \nu)$ is given for all $\mu \geq 0$ and $\nu \geq 0$ by

$$G(\mu, \nu) = \begin{cases} -b^\top \mu & \text{if } A^\top \mu - \nu + c = 0, \\ -\infty & \text{otherwise.} \end{cases}$$

Duality Gap of a Linear Program

The hypotheses of Part (1) certainly *fail* since there are infinitely $u_{\mu,\nu} \in \mathbb{R}^n$ such that $G(\mu,\nu) = \inf_{v\in\mathbb{R}^n} L(v,\mu,\nu) = L(u_{\mu,\nu},\mu,\nu)$.

Duality Gap of a Linear Program

The hypotheses of Part (1) certainly *fail* since there are infinitely $u_{\mu,\nu} \in \mathbb{R}^n$ such that $G(\mu, \nu) = \inf_{v \in \mathbb{R}^n} L(v, \mu, \nu) = L(u_{\mu,\nu}, \mu, \nu)$.

Therefore, the dual function G is no help in finding a solution of the Primal Problem (P).

Duality Gap of a Linear Program

The hypotheses of Part (1) certainly *fail* since there are infinitely $u_{\mu,\nu} \in \mathbb{R}^n$ such that $G(\mu, \nu) = \inf_{v \in \mathbb{R}^n} L(v, \mu, \nu) = L(u_{\mu,\nu}, \mu, \nu)$.

Therefore, the dual function G is no help in finding a solution of the Primal Problem (P).

As we saw earlier, if we consider the modified dual Problem (D_1), then strong duality holds, but this *does not* follow from the preceding theorem, and a different proof is required.

Duality Gap of a Linear Program

Thus, we have the somewhat counter-intuitive situation that the *general* theory of Lagrange duality does not apply, at least directly, to linear programming, a fact that is not sufficiently emphasized in many expositions. *A separate treatment of duality is required*.

Duality Gap of a Linear Program

Thus, we have the somewhat counter-intuitive situation that the *general* theory of Lagrange duality does not apply, at least directly, to linear programming, a fact that is not sufficiently emphasized in many expositions. *A separate treatment of duality is required*.

Unlike the case of linear programming, which needs a separate treatment, the preceding theorem applies to the optimization problem involving a convex quadratic objective function and a set of affine inequality constraints.

Duality Gap of a Linear Program

Thus, we have the somewhat counter-intuitive situation that the *general* theory of Lagrange duality does not apply, at least directly, to linear programming, a fact that is not sufficiently emphasized in many expositions. *A separate treatment of duality is required*.

Unlike the case of linear programming, which needs a separate treatment, the preceding theorem applies to the optimization problem involving a convex quadratic objective function and a set of affine inequality constraints.

So in some sense, convex quadratic programming is simpler than linear programming!

Duality and Quadratic Optimization

Example. Consider the quadratic objective function

$$J(v) = \frac{1}{2} v^\top A v - v^\top b,$$

where A is an $n \times n$ matrix which is *symmetric positive definite*, $b \in \mathbb{R}^n$, and the constraints are affine inequality constraints of the form

$$Cv \leq d,$$

where C is an $m \times n$ matrix and $d \in \mathbb{R}^m$. For the time being, we do not assume that C has rank m.

Duality and Quadratic Optimization

Since A is symmetric positive definite, J is *strictly convex*.

Duality and Quadratic Optimization

Since A is symmetric positive definite, J is *strictly convex*.

The Lagrangian of this quadratic optimization problem is given by

$$\begin{aligned} L(v, \mu) &= \frac{1}{2} v^\top A v - v^\top b + (Cv - d)^\top \mu \\ &= \frac{1}{2} v^\top A v - v^\top (b - C^\top \mu) - \mu^\top d. \end{aligned}$$

Duality and Quadratic Optimization

Since A is symmetric positive definite, the function $v \mapsto L(v, \mu)$ has a *unique minimum* obtained for the solution u_μ of the linear system

$$Av = b - C^\top \mu;$$

that is,

$$u_\mu = A^{-1}(b - C^\top \mu).$$

Duality and Quadratic Optimization

This shows that the Problem (P_μ) has a *unique* solution which depends continuously on μ. Then for *any* solution λ of the dual problem, $u_\lambda = A^{-1}(b - C^\top \lambda)$ is an optimal solution of the primal problem.

Duality and Quadratic Optimization

This shows that the Problem (P_μ) has a *unique* solution which depends continuously on μ. Then for *any* solution λ of the dual problem, $u_\lambda = A^{-1}(b - C^\top \lambda)$ is an optimal solution of the primal problem.

We find that $G(\mu)$ is given by

$$G(\mu) = -\frac{1}{2}\mu^\top CA^{-1}C^\top \mu + \mu^\top (CA^{-1}b - d) - \frac{1}{2}b^\top A^{-1}b.$$

Duality and Quadratic Optimization

Since A is symmetric positive definite, the matrix $CA^{-1}C^{\top}$ is symmetric positive semidefinite. It can be shown that $CA^{-1}C^{\top}$ is symmetric positive definite iff C has rank m.

Duality and Quadratic Optimization

Since A is symmetric positive definite, the matrix $CA^{-1}C^{\top}$ is symmetric positive semidefinite. It can be shown that $CA^{-1}C^{\top}$ is symmetric positive definite iff C has rank m.

In this case it can be shown that *if* the inequalities $Cx \leq d$ have a solution, *then* the primal problem has a unique solution.

Duality and Quadratic Optimization

As a consequence, by Part (2) of the duality gap theorem, the function $-G(\mu)$ always has a minimum, which is unique if C has rank m. The fact that $-G(\mu)$ has a minimum is not obvious when C has rank $< m$, since in this case $CA^{-1}C^{\top}$ is not invertible.

Duality and Quadratic Optimization

As a consequence, by Part (2) of the duality gap theorem, the function $-G(\mu)$ always has a minimum, which is unique if C has rank m. The fact that $-G(\mu)$ has a minimum is not obvious when C has rank $< m$, since in this case $CA^{-1}C^\top$ is not invertible.

We also verify easily that the gradient of G is given by

$$\nabla G_\mu = Cu_\mu - d = -CA^{-1}C^\top \mu + CA^{-1}b - d.$$

Observe that since $CA^{-1}C^\top$ is symmetric positive semidefinite, $-G(\mu)$ is convex.

Duality and Quadratic Optimization

Therefore, if C has rank m, a solution of Problem (P) is obtained by finding the unique solution λ of the equation

$$-CA^{-1}C^{\top}\mu + CA^{-1}b - d = 0,$$

and then the minimum u_λ of Problem (P) is given by

$$u_\lambda = A^{-1}(b - C^{\top}\lambda).$$

Duality and Quadratic Optimization

Therefore, if C has rank m, a solution of Problem (P) is obtained by finding the unique solution λ of the equation

$$-CA^{-1}C^{\top}\mu + CA^{-1}b - d = 0,$$

and then the minimum u_{λ} of Problem (P) is given by

$$u_{\lambda} = A^{-1}(b - C^{\top}\lambda).$$

If C has rank $< m$, then we can find $\lambda \geq 0$ using a method called ADMM.