Duality, Complementary Slackness and KKT Conditions: an overview

Caio De Naday Hornhardt

October, 2022

1 Definition of Dual Problem

Consider the following optimization problem (referred to as the *primal* problem):

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & g(x) \leq 0 \\ & h(x) = 0, \end{aligned} \tag{1}$$

where $f\colon D \to \mathbb{R}$, $g\colon D \to \mathbb{R}^{k_1}$, $h\colon D \to \mathbb{R}^{k_2}$ and D is any set, which called the *domain* of the problem. For many problems, D is taken to be $\mathbb{R}^n$ or a open subset of $\mathbb{R}^n$, but it could also be a discrete set in what follows.

Remark. Even though we state optimization problems as finding the minimum or the maximum, in general, we can only guarantee the existence of the infimum or the supremum (allowing it to be $\pm\infty$).

To find the *dual problem*, we first consider a "relaxed" version of the primal. Given $\mu \in \mathbb{R}^{k_1}$ such that $\mu \geq 0$ and $\lambda \in \mathbb{R}^{k_2}$, we define

$$\ell_{\mu,\lambda}(x) \coloneqq f(x) + \mu^\top g(x) + \lambda^\top h(x) \tag{2}$$

We then solve the unrestricted problem:

$$\min \quad \ell_{\mu,\lambda}(x)\,.$$

If x is a feasible point for the primal problem, then

$$\ell_{\mu,\lambda}(x) = f(x) + \mu^\top g(x) + \lambda^\top h(x) \leq f(x)\,,$$

so the solutions of this unrestricted problem are better or at least as good as the solutions of the primal problem. Formally speaking, setting

$$d(\mu, \lambda) \coloneqq \inf_x \ell_{\mu,\lambda}(x)\,,$$

we have that

$$d(\mu, \lambda) \leq \inf_{x \text{ feasible}} f(x)\,, \tag{3}$$

i.e., $d(\mu, \lambda)$ is a lower bound for the optimal value of the primal problem.

The *dual problem* is to find the best lower bound that can achieved this way, i.e.,

$$\begin{aligned} \max \quad & d(\mu, \lambda) \\ \text{s.t.} \quad & \mu \geq 0 \,. \end{aligned} \tag{4}$$

There is only one condition ($\mu \geq 0$) in the dual problem, but we say that a point (μ, λ) is *dual feasible* if $\mu \geq 0$ and $d(\mu, \lambda) > -\infty$. This is because, frequently, the latter condition can be written explicitly with equations and inequalities.

Of course, (3) is equivalent to

$$\sup_{\mu \geq 0, \text{ any } \lambda} d(\mu, \lambda) \leq \inf_{x \text{ feasible}} f(x), \tag{5}$$

which is referred to as *weak duality*. When we have the equality on (5), we say that we have *strong duality*. In general, the difference between the right-hand side and the left-hand side is called *duality gap*.

2 Examples

Example 1. Consider the primal problem:

$$\begin{aligned} \min \quad & 2x + y \\ \text{s.t.} \quad & x^2 + y^2 \leq 1, \end{aligned}$$

where $x, y \in \mathbb{R}$. Given $\mu \geq 0$, we define

$$\ell_\mu(x, y) \coloneqq 2x + y + \mu(x^2 + y^2 - 1) \,.$$

For $\mu = 0$, $\inf \ell_\mu(x) = -\infty$, so 0 is not dual feasible. For other values of μ, we can derive ℓ_μ and get

$$\nabla \ell_\mu(x) = (2 + 2\mu x, 1 + 2\mu y),$$

so the $(-\frac{1}{\mu}, -\frac{1}{2\mu})$ is the only critical point. Computing the second derivatives we see that ℓ_μ is convex and, hence, $(-\frac{1}{\mu}, -\frac{1}{2\mu})$ is the point where the minimum is achieved. The minimum value is then

$$\begin{aligned} d(\mu) &= \ell_\mu \left(-\frac{1}{\mu}, -\frac{1}{2\mu} \right) \\ &= 2 \left(-\frac{1}{\mu} \right) + \left(-\frac{1}{2\mu} \right) + \mu \left(-\frac{1}{\mu} \right)^2 + \mu \left(-\frac{1}{2\mu} \right)^2 - \mu \\ &= -\frac{2}{\mu} - \frac{1}{2\mu} + \frac{1}{\mu} + \frac{1}{4\mu} - \mu \\ &= \frac{-8 - 2 + 4 + 1}{4\mu} - \mu = -\frac{5}{4\mu} - \mu \end{aligned}$$

To find the maximum value of $d(\mu)$, we derive d:

$$d'(\mu) = \frac{5}{4\mu^2} - 1,$$

hence the $\mu = \frac{\sqrt{5}}{2}$ is the only critical point for which $\mu > 0$. It is easy to see d' is positive before and negative after $\frac{\sqrt{5}}{2}$, so it is the maximum point. The maximum value is then

$$d\left(\frac{\sqrt{5}}{2}\right) = -\frac{5 \cdot 2}{4\sqrt{5}} - \frac{\sqrt{5}}{2} = -\sqrt{5}.$$

So far we found a lower bound for the primal problem. Is it a tight bound? Well, we have that the minimum of ℓ_μ is achieved at $(x, y) = (-\frac{1}{\mu}, -\frac{1}{2\mu})$, so a reasonable guess is to plug $\mu = \frac{\sqrt{5}}{2}$ into this, getting $(x, y) = (-\frac{2}{\sqrt{5}}, -\frac{1}{\sqrt{5}})$. We then have

$$x^2 + y^2 = \frac{4}{5} + \frac{1}{5} = 1,$$

so (x, y) is feasible for the primal problem. Also,

$$2x + y = -\frac{4}{\sqrt{5}} - \frac{1}{\sqrt{5}} = -\frac{5}{\sqrt{5}} = -\sqrt{5},$$

so the strong duality is achieved.

Example 2. Consider the primal problem

$$\begin{aligned} \max \quad & x \\ \text{s.t.} \quad & x^2 \leq 0\,. \end{aligned}$$

There is only one viable point, $x = 0$, hence it is the minimum point and the minimum value is 0. To find the dual problem, fix $\mu \geq 0$ and define

$$\ell_\mu(x) \coloneqq x + \mu x^2.$$

For $\mu = 0$, $\inf \ell_\mu(x) = -\infty$, so 0 is not dual feasible. For $\mu \neq 0$, we have

$$\ell'_\mu(x) = 1 + 2\mu x,$$

so the minimum point is $x = -\frac{1}{2\mu}$ and the minimum value is

$$d(\mu) \coloneqq \ell_\mu\left(-\frac{1}{2\mu}\right) = -\frac{1}{2\mu} + \mu\left(-\frac{1}{2\mu}\right)^2 = -\frac{1}{2\mu} + \frac{1}{4\mu} = -\frac{1}{4\mu}.$$

We now want to maximize d, but it does not have a maximum. Yet, $\sup d(\mu) = 0$ (which is the limit of $d(\mu)$ as $\mu \to +\infty$), so strong duality does hold.

Example 3. Consider the primal problem

$$\begin{aligned} \min \quad & x^2 \\ \text{s.t.} \quad & x^3 \leq -1 \,. \end{aligned}$$

It is easy to see that the point of minimum is $x = -1$, for which we obtain the minimum value of 1. To find the dual problem, we define

$$\ell_\mu(x) \coloneqq x^2 + \mu(x^3 - 1) \,,$$

for $\mu \geq 0$. If $\mu \neq 0$, then the $d(\mu) \coloneqq \inf_x \ell_\mu(x) = -\infty$. For $\mu = 0$, $d(\mu) = 0$. We then have that the duality gap is 1.

Example 4. The primal problem in Example 3 is clearly equivalent

$$\begin{aligned} \max \quad & x^2 \\ \text{s.t.} \quad & x \leq -1 \,. \end{aligned}$$

Yet, for this new problem, we have strong duality. To see that, define

$$\ell_\mu(x) \coloneqq x^2 + \mu(x + 1),$$

for $\mu \geq 0$. To find its minimum point, we solve

$$\ell'_\mu(x) = 2x + \mu = 0$$

getting $x = -\frac{\mu}{2}$. Then the minimum value of ℓ_μ is

$$d(\mu) \coloneqq \ell_\mu\left(-\frac{\mu}{2}\right) = \frac{\mu^2}{4} + \mu\left(\frac{2-\mu}{2}\right) = \frac{-\mu^2 + 4\mu}{2} \,.$$

To find the maximum of $d(\mu)$, we solve

$$d'(\mu) = \frac{-2\mu + 4}{2} = -\mu + 2 = 0 \,,$$

getting $\mu = 2$. Finally, we have that $d(2) = 1$, so strong duality holds.

3 Complementary Slackness

In Example 1, we did "reasonable guess" for the minimum point (x, y) for the primal problem: we took μ to be the maximum point for d and chose (x, y) to be the point that minimizes ℓ_μ. The guess worked out there, but would this strategy always work? After all, why would the minimum points of these two different functions be the same?

Let us revisit our argument in Section 1 establishing that weak duality always holds: for any primal feasible point x, any $\mu \in \mathbb{R}^{k_1}$ with $\mu \geq 0$, and any $\lambda \in \mathbb{R}^{k_2}$, we have that

$$\ell_{\mu,\lambda}(x) = f(x) + \mu^\top g(x) + \lambda^\top h(x) = f(x) + \mu^\top g(x) \tag{6}$$

and, since $\mu^\top g(x) \leq 0$, $\ell_{\mu,\lambda}(x) \leq f(x)$. Then, by definition of $d(\mu, \lambda)$,

$$d(\mu, \lambda) \leq \ell_{\mu,\lambda}(x) \leq f(x)\,. \tag{7}$$

If strong duality holds and x^* and (μ^*, λ^*) are optimal points for the primal and dual problems, respectively, Equation (7) implies that $d(\mu^*, \lambda^*) = \ell_{\mu^*,\lambda^*}(x^*)$, so $x*$ is a minimum point for ℓ_{μ^*,λ^*}, and also that $\ell_{\mu^*,\lambda^*}(x^*) = f(x^*)$. This justifies our "reasonable guess."

By Equation (6), $\ell_{\mu^*,\lambda^*}(x^*) = f(x^*)$ is equivalent to $(\mu^*)^\top g(x^*) = 0$. We refer to this condition as *complementary slackness*. Since $\mu^* \geq 0$ and $g(x^*) \leq 0$, it is equivalent to $\mu_i^* g_i(x^*) = 0$, for $1 \leq i \leq k_1$, where μ_i^* and $g_i(x^*)$ are the i-th components of the vectors μ^* and $g(x^*)$, respectively. In other words, if one of the (scalar) inequalities conditions in the primal problem is not tightly satisfied, then the corresponding inequality condition in the dual problem must be tightly satisfied, hence the name.

We summarize the discussion above in the following:

Proposition 5. *Suppose x^* and (μ^*, λ^*) are, respectively, optimal points for the primal and dual problems, as defined in Equations* (1) *and* (4)*, and strong duality holds. Then $(\mu^*)^\top g(x^*) = 0$ and x^* is a optimal point for the function ℓ_{μ^*,λ^*}, as defined in Equation* (2). □

Proposition 5 can be adapted to include cases where strong duality holds, but we do not have achieve a minimum or a maximum (like in Example 2). This version can also be useful in the view of iterative algorithms. To avoid confusion with coordinates, we will index our sequences with upper scripts.

Proposition 6. *Suppose that strong duality holds and consider sequences $(x^i)_{i\in\mathbb{N}}$ and $(\mu^i, \lambda^i)_{i\in\mathbb{N}}$ of feasible points in the primal and dual problems, respectively, such that both $f(x^i)$ and $d(\mu^i, \lambda^i)$ converge to the optimal value. Then $(\mu^i)^\top g(x^i)$ converges to* 0 *and $\ell_{\mu^i,\lambda^i}(x^i)$ converges to the optimal value.*

Proof. Given x and (μ, λ) primal and dual feasible points, set $\epsilon \coloneqq f(x) - d(\mu, \lambda)$. Then

$$\begin{aligned} f(x) &= d(\mu, \lambda) + \epsilon \\ &\leq \ell_{\mu,\lambda}(x) + \epsilon \\ &= f(x) + \mu^\top g(x) + \lambda^\top h(x) + \epsilon \\ &= f(x) + \mu^\top g(x) + \epsilon\,, \end{aligned}$$

so $-\mu^\top g(x) \leq \epsilon$. Since $\mu^\top g(x) \leq 0$, we have that $0 \leq -\mu^\top g(x) \leq \epsilon$. Taking $\{x^{(i)}\}_{i\in\mathbb{N}}$ and $\{(\mu^{(i)}, \lambda^{(i)})\}_{i\in\mathbb{N}}$ as above, we get a sequence $\epsilon_i \to 0$, concluding the proof. □

4 Optimality Certificates and KKT

It follows from the weak duality (5) that if we find feasible points x and (μ, λ) for the primal and dual problems, respectively, such that $f(x) = d(\mu, \lambda)$, then strong duality holds, and x and (μ, λ) are optimal points. This motivates the following:

Definition 7. Let x and (μ, λ) be primal and dual feasible points, respectively. We say that the triple (x, μ, λ) is a *optimality certificate* if $f(x) = d(\mu, \lambda)$.

It is important to notice that simply having that $\ell_{\mu,\lambda}(x) = f(x)$ does not imply that (x, μ, λ) is a optimality certificate. We could still have that $d(\mu, \lambda) < \ell_{\mu,\lambda}(x)$ in Equation (7) and, hence, not only (μ, λ) wouldn't be a optimal value for the dual problem, but we also wouldn't be able to guaranty that x is an optimal point for the primal problem.

Yet, $\ell_{\mu,\lambda}(x) = f(x)$ or, equivalently, complementary slackness, is a necessary condition for (x, μ, λ) to be an optimality certificate. Let us add it together to the conditions in Definition 7 to get a list necessary conditions:

(i) Primal feasibility, i.e., $g(x) \leq 0$ and $h(x) = 0$;

(ii) Dual feasibility, i.e., $\mu \geq 0$ and $d(\mu, \lambda) > -\infty$;

(iii) Complementary slackness, i.e., $\mu^\top g(x) = 0$.

We will now develop these conditions further for some specific cases. As usual for this topic in the literature, we will follow the pattern of listing all the equality conditions before the inequalities.

4.1 Linear Case

Consider the primal to be a linear problem in standard form:

$$\begin{aligned} \min \quad & c^\top x \\ \text{s.t.} \quad & Ax = b \\ & x \geq 0\,, \end{aligned}$$

where $x, c \in \mathbb{R}^n$, $b \in \mathbb{R}^{k_1}$ and A is an $k \times n$ matrix.

Given $\mu \in \mathbb{R}^n$, $\mu \geq 0$, and $\lambda \in \mathbb{R}^k$, we have that

$$\ell_{\mu,\lambda}(x) = c^\top x - \mu^\top x + \lambda^\top (b - Ax) = (c^\top - \mu^\top - \lambda^\top A)x + \lambda^\top b\,.$$

A point (μ, λ) is dual feasible if, and only if, $\mu \geq 0$ and $c^\top - \mu^\top - \lambda^\top A = 0$ and, hence, $\ell_{\mu,\lambda}$ is constant and equal to $\lambda^\top b$. In particular $d(\mu, \lambda) = \lambda^\top b$. The dual problem, then, can be written as

$$\begin{aligned} \max \quad & b^\top \lambda \\ \text{s.t.} \quad & A^\top \lambda + \mu = c \\ & \mu \geq 0\,. \end{aligned}$$

Since $\ell_{\mu,\lambda}$ is a constant function,

$$\begin{aligned} c^\top x = d(\mu,\lambda) &\iff c^\top x = \ell_{\mu,\lambda}(x) \\ &\iff c^\top x = c^\top x - \mu^\top x + \lambda^\top (b - Ax) \\ &\iff \mu^\top x = 0\,. \end{aligned}$$

We have the complementary slackness is necessary and sufficient!

Summarizing, a triple (x, μ, λ) is a optimality certificate if, and only if,

$$\begin{aligned} (i) &\quad Ax = b \\ (ii) &\quad A^\top \lambda + \mu = c \\ (iii) &\quad \mu^\top x = 0 \\ (iv) &\quad x \geq 0 \\ (v) &\quad \mu \geq 0\,. \end{aligned} \tag{8}$$

Remark. Generaly speaking, the *interior methods* for solving linear programming problems follow the strategy of solving (a slightly modified version of) the system given by $(i)-(iii)$ iteratively, while trying to keep strict inequalities in conditions (iv) and (v).

4.2 Differentiable Case

We now suppose all the functions involved in the primal problem to be differentiable and that the domain is an open set. If (x, μ, λ) be an optimality certificate, then $d(\mu, \lambda) = \ell_{\mu,\lambda}(x)$, so x is a point of minimum value of $\ell_{\mu,\lambda}$. It follows that

$$\nabla \ell_{\mu,\lambda}(x) = \nabla f(x) + \sum_{i=1}^{k_1} \mu_i \nabla g_i(x) + \sum_{i=1}^{k_2} \lambda_i \nabla h_i(x) = 0\,.$$

Therefore, we get the following set of necessary conditions for (x, μ, λ) to be a certificate:

$$\begin{aligned} (i) &\quad \nabla f(x) + \sum_{i=1}^{k_1} \mu_i \nabla g_i(x) + \sum_{i=1}^{k_2} \lambda_i \nabla h_i(x) = 0 \\ (ii) &\quad h(x) = 0 \\ (iii) &\quad \mu^\top g(x) = 0 \\ (iv) &\quad g(x) \leq 0 \\ (v) &\quad \mu \geq 0\,, \end{aligned} \tag{9}$$

which are called the *Karush–Kuhn–Tucker (KKT) conditions.*

Remark. We note that, for the linear case, condition $9(i)$ give us precisely condition $8(ii)$.

Note that, if the primal problem has no inequalities restrictions, than the KKT conditions reduce to the Lagrange multipliers condition.

While the strong duality may not hold globally for many problems, it may still hold if we take the domain to be a smaller open set. If this is the case, the KKT conditions will still hold for locally optimal points. The so called *constraint qualifications* are conditions that, if true, guaranty that KKT holds for locally optimal points.

4.3 Differentiable and Convex Case

We already know that the KKT conditions are necessary for a triple (x, μ, λ) to be a optimality certificate. Now let us assume the the primal problem is also convex, i.e., f and g are convex functions and h is a affine function.

Since $\mu \geq 0$ and h is affine, we have that $\ell_{\mu,\lambda}(x) = f(x) + \mu^\top g(x) + \lambda^\top h(x)$ is a convex function. For a differentiable convex function, if a point x is stationary (condition $9(i)$), then it must be a global minimum. In other words, $d(\mu, \lambda) = \ell_{\mu,\lambda}(x)$. If (x, μ, λ) also satisfies conditions the other KKT conditions, then

$$d(\mu, \lambda) = \ell_{\mu,\lambda}(x) = f(x) + \mu^\top g(x) + \lambda^\top h(x) = f(x).$$

We then conclude that (x, μ, λ) is a certificate, i.e., the KKT conditions are necessary and sufficient in this case.