CS5487 - Geometric Interpretation of Duality

Antoni Chan
Department of Computer Science
City University of Hong Kong

1 Primal and dual problems

Consider the following optimization problem with inequality constraints:

$$f^* = \min_x f(x) \quad \text{s.t. } g(x) \geq 0. \tag{1}$$

The Lagrangian of this problem is

$$L(x, \lambda) = f(x) - \lambda g(x), \tag{2}$$

where $\lambda \geq 0$ is the Lagrange multiplier. At the solution of (1) the following KKT conditions hold:

1. $g(x) \geq 0$,
2. $\lambda \geq 0$,
3. $\lambda g(x) = 0$.

The corresponding dual problem for (1) is

$$q^* = \max_{\lambda \geq 0} q(\lambda) \tag{3}$$

where the dual function is

$$q(\lambda) = \min_x L(x, \lambda) = \min_x f(x) - \lambda g(x). \tag{4}$$

2 Geometric interpretation of the dual problem

Now we will look at a geometric interpretation of the dual problem. First, we can rewrite $L(x, \lambda)$ by defining

$$w = \begin{bmatrix} -\lambda \\ 1 \end{bmatrix}, \quad z(x) = \begin{bmatrix} g(x) \\ f(x) \end{bmatrix}, \quad \Rightarrow \quad L(x, \lambda) = w^T z(x) \tag{5}$$

We will consider the optimization problem in the 2D space of $z = \begin{bmatrix} g \\ f \end{bmatrix}$.

For now, assume that the optimization problem is *convex*, i.e., $f(x)$ is a convex function and $g(x) \geq 0$ is a convex region.

2.1 Feasible region

Define the set $\mathcal{Z} = \{ \begin{bmatrix} g(x) \\ f(x) \end{bmatrix} | \forall x \}$ as the set of all possible z vectors, i.e., all possible pairs of $(g(x), f(x))$, for all possible x. The region $\mathcal{Z}$ is called the *feasible region* since it encompasses all possible values of $(g(x), f(x))$. Note that here we are not keeping track of which x gave us the particular z vector.

As for the optimal f^*, there are two possible scenarios:

(a) The equality constraint is inactive, i.e., $g(x^*) = g^* > 0$, and the optimal solution is at the point $z^* = \begin{bmatrix} g^* \\ f^* \end{bmatrix}$. See Figure 1(a).

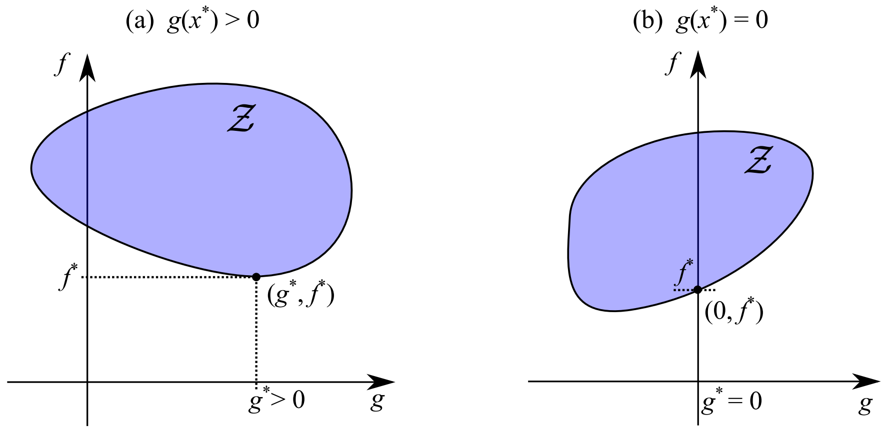

Figure 1: Feasible regions $\mathcal{Z}$ and optimal solution when (a) equality constraint is inactive $g(x^*) > 0$, and (b) equality constraint is active $g(x^*) = 0$.

(b) The equality constraint is active, i.e., $g(x^*) = g^* = 0$, and hence the optimal solution is at the point $z^* = \begin{bmatrix} 0 \\ f^* \end{bmatrix}$. See Figure 1(b).

2.2 Supporting hyperplane

Now consider the vector $w = \begin{bmatrix} -\lambda \\ 1 \end{bmatrix}$. When $\lambda = 0$ then the vector w is pointing up (in the positive f direction). When $\lambda > 0$, then w is pointing into the second quadrant (up and to the left). See Figure 2(a).

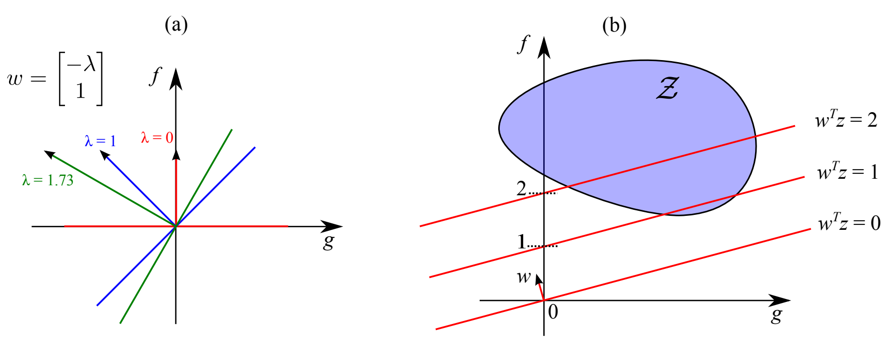

Figure 2: (a) Examples of the vector w and corresponding hyperplanes; (b) Parallel lines defined by $w^T z = b$ for different f-intercept values b.

Selecting a value for λ gives us a w, which defines a corresponding linear function $w^T z$. Note that $w^T z = 0$ defines a line perpendicular to w and passing through the origin. Likewise, $w^T z = b$ defines a line passing through the point $z = \begin{bmatrix} 0 \\ b \end{bmatrix}$, i.e., the point where the line intercepts the f-axis. See Figure 2(b).

Now consider the dual objective function $q(\lambda) = \min_x w^T z(x)$. For a given λ, we need to find the optimal x that minimizes the linear function $w^T z(x)$. This is equivalent to minimizing over the z vectors in the set $\mathcal{Z}$,

$$q(\lambda) = \min_x w^T z(x) = \min_{z \in \mathcal{Z}} w^T z. \tag{6}$$

Geometrically, for a given w (i.e., λ), this corresponds to finding the smallest b such that the line $w^T z = b$ passes through the region $\mathcal{Z}$. The b is minimized when the optimal line $w^T z = b$ is just touching (tangent) to the region $\mathcal{Z}$. This is called a *supporting hyperplane* of the region $\mathcal{Z}$. See Figure 3(a).

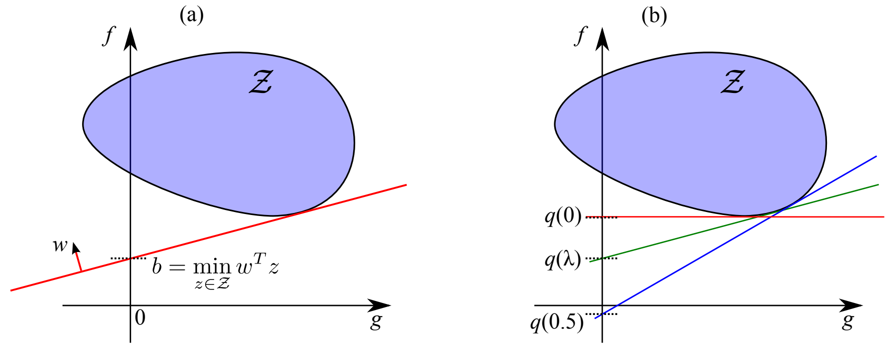

Figure 3: (a) Supporting hyperplane $w^T z = b$ for vector w is tangent to the $\mathcal{Z}$. (b) Supporting hyperplanes for different values of λ. The f-intercepts of the supporting hyperplanes are the values of $q(\lambda)$.

Note that b is the value of the dual objective function, $q(\lambda) = b = \min_{z \in \mathcal{Z}} w^T z$. For different values of λ, we get different supporting hyperplanes, and the corresponding values of $q(\lambda)$ are the f-intercept values. See Figure 3(b).

2.3 Geometric interpretation

The dual problem is to find the maximum of $q^* = \max_{\lambda \geq 0} q(\lambda)$. Geometrically, this corresponds to finding the supporting hyperplane w that has the largest f-intercept value. There are two cases of interest:

(a) The optimal supporting hyperplane is horizontal and tangent to $\mathcal{Z}$ at the point (g^*, f^*). This corresponds to the case when the equality constraint is inactive, i.e., $g^* > 0$ and $\lambda = 0$. See Figure 4(a).

(b) the optimal supporting hyperplane is tangent to the point $(0, f^*)$ of the region $\mathcal{Z}$. This corresponds to the case when the equality constraint is active, i.e., $g^* = 0$ and $\lambda > 0$. See Figure 4(b).

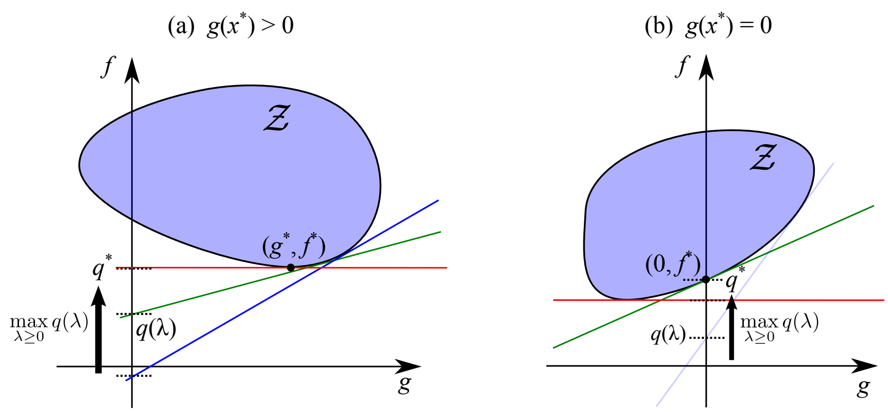

Figure 4: (a) Optimal supporting hyperplane is horizontal and tangent to the point (g^*, f^*), and $g^* > 0$. (b) Optimal supporting hyperplane is tangent to the point $(0, f^*)$, i.e., $g^* = 0$.

Note that in both the above cases, $q^* = f^*$, and hence solving the dual problem gives us the solution to the primal problem. By the Strong Duality Theorem, the equivalence of the dual and primal problems is guaranteed when the primal problem is convex (as we have assumed so far).

2.4 Duality gap

In general, by the Weak Duality Theroem, the dual solution will always lower bound the primal solution, $q^* \leq f^*$. When it the problem is not convex, then the f-intercept q^* for the optimal supporting hyperplane may not reach the true minimum f^*, as seen in the example in Figure 5(a), which results in a *duality gap*. On the other hand, not all non-convex problems have a duality gap, e.g., see Figure 5(b).

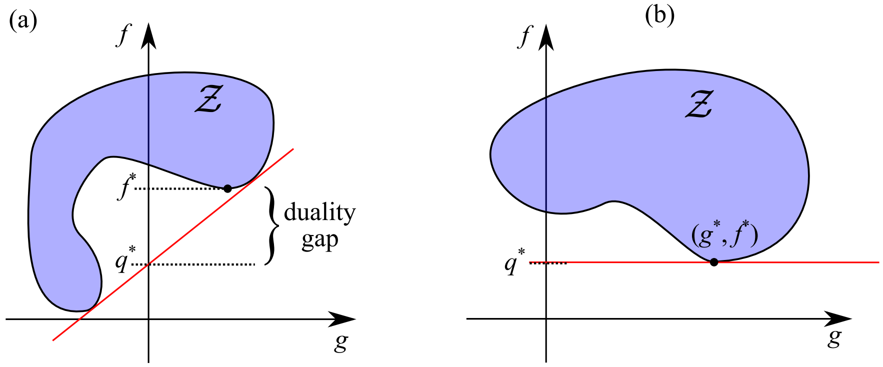

Figure 5: Examples of non-convex problems: (a) has a *duality gap* and $q^* < f^*$. (b) does not have a duality gap ($q^* = f^*$), although the problem is non-convex.