1 The Dual Problems

The concept “duality”, just like accuracy and stability, frequently appears in different research areas of scientific computing. For example, among many other aspects, the dual of a given problem often offers a way to study the sensitivity of numerical methods of almost all kinds, including linear programs. Dual problems are not unfamiliar in our lectures, particularly the standard minimum form and the standard maximum form† are intentionally formulated as the dual to each other:

The standard minimum problem: Find a vector \( x \in \mathbb{R}^n \), such that:

\[
\min c^t x \quad \text{(1.1a)}
\]

is minimized subject to:

\[
A x \geq b, \quad \text{(1.1b)}
\]

and:

\[
x \geq 0. \quad \text{(1.1c)}
\]

The standard maximum problem: Find a vector \( y \in \mathbb{R}^m \), such that:

\[
\max b^t y \quad \text{(1.2a)}
\]

is maximized subject to:

\[
y^t A \leq c^t, \quad \text{(1.2b)}
\]

and:

\[
y \geq 0. \quad \text{(1.2c)}
\]

Attention should be paid to how the coefficients for the objective function of one problem become those for the constraints of the other problem, and vice versa. In addition, the matrix \( A \) in the inequality constraints is the same. The two problems (1.1) and (1.2) are called the symmetric form of duality.

To get a sense of why they are called dual problems, we have the weak duality lemma as follows:

**Lemma 1** (Weak Duality Lemma for Symmetric Duality). If \( x \) and \( y \) are feasible for (1.1) and (1.2), respectively, then \( c^t x \geq b^t y \).

†They are slightly different from what we had before for easier presentation here.
**Proof.** Because $x \geq 0$, we have:

\[ c^T x \geq (y^T A)x = y^T (Ax). \]

Since $y \geq 0$, this inequality can be continued:

\[ c^T x \geq y^T (Ax) \geq y^T b. \]

As a direct corollary, if we manage to find a feasible $x_0$ to the minimum problem and a feasible $y_0$ to the maximum problem, such that $c^T x_0 = b^T y_0$, it can be immediately claimed that both of them are optimal. The converse is also true, as we will see later in the duality theorem of linear programming; that is, there is no “gap” between the two optimal objective values of the dual problems. This is the foundation why we can use the ellipsoid method or the interior penalty method to solve the linear program, which finds any point in the polytope $\Omega$:

\[ \Omega = \{ z \in \mathbb{R}^M : z^T D \leq d, D \in \mathbb{R}^{M \times N} \}. \]

Indeed, suppose we want to solve the standard minimum problem, then by the duality theorem, an optimal solution $x$ is part of $(x,y)$ that satisfy the following linear constraints:

\[
\begin{align*}
  c^T x - b^T y & \leq 0, \\
  -Ax & \leq -b, \\
  A^T y & \leq c, \\
  -x & \leq 0, \\
  -y & \leq 0.
\end{align*}
\]

This polytope is defined by $M = m + n$ variables and $N = 1 + 2m + 2n$ linear inequality constraints.

So what is the dual problem for a linear program in the standard form?

The standard linear programming problem: Find a vector $x \in \mathbb{R}^n$, such that:

\[ c^T x \] (1.3a)

is minimized subject to:

\[ Ax = b \geq 0, \] (1.3b)

and:

\[ x \geq 0. \] (1.3c)

The first step is to put it in a standard minimum form as: Minimizing

\[ c^T x \]

such that

\[ Ax \geq b, \quad (-A)x \geq -b, \]

and $x \geq 0$. The corresponding dual maximum problem reads: Finding $u,v \in \mathbb{R}^n$ such that

\[ b^T u - b^T v \]
is maximized, subject to
\[
[u \ v]^t \begin{bmatrix} A \\ -A \end{bmatrix} = (u - v)^t A \leq c,
\]
and \(u, v \geq 0\). Noticing how the two non-negative vectors \(u\) and \(v\) always appear in a pair as \(u - v\) except for the non-negativity constraints; this is exactly the trick we used before to put a free variable into non-negative ones. Reverting this process, we have the following dual problem with free variables:

The linear program with free variables: Find a vector \(y \in \mathbb{R}^m\), such that:

\[
b^t y \tag{1.4a}
\]

is maximized subject to:

\[
y^t A \leq c^t. \tag{1.4b}
\]

The dual relation between \([1.3]\) and \([1.4]\) is called the asymmetric form of the duality of linear programs. Usually for two dual problems, if one is called primal, the other one is called the dual to the primal one. In this lecture, we usually identify the standard form as the primal one.

With the help of the asymmetric form of the duality, we have another interpretation of the dual problem and some deeper insight why the number of variables in the dual problem is the same as the number of constraints in the primal one. This is related to the saddle point problem: Given \(f(x,y)\) we want to solve for \(\min_x \max_y f(x,y)\). The saddle point problem itself is a rich area of researches. Let us denote the solution to the saddle point problem by \((x^*, y^*)\); to get an idea of how to find such a solution, we consider the simple objective function \(f(x, y) = (x - 3)^2 - (y - x)^2\). This function has neither minimum nor maximum; but if we fix \(x = x_0\) it is possible to solve the maximization problem \(\max_y f(x_0, y)\) to find the solution \(y^*(x_0) = x_0\). Then we can solve the minimization problem \(\min_x f(x^*, y(x))\) and find out that \(x^* = 3\). At this saddle point \((x^*, y^*) = (3, 3)\), the gradient of \(f\) is zero but this point is neither minimum nor maximum. Applications that lead to saddle point problems composes a long list, which includes mixed finite element analysis, fluid dynamics of incompressible flows, shape and topological optimization, and Kalman filtering in signal processing, to name a few.

Here we consider a saddle point problem that is created by using Lagrangian multipliers (the objective function is hence called the Lagrangian). The target is to find \(\min_x \max_\lambda f(x, \lambda)\), where:

\[
f(x, \lambda) = c^t x - \lambda^t (Ax - b),
\]

subject to the constraints \(x \geq 0\). Similar to the previous example, we can first fix \(x\) to find out the optimality condition for \(\lambda\), which reads:

\[
\nabla_\lambda f(x, \lambda) = 0, \quad \text{or} \quad Ax - b = 0.
\]

The second step of the saddle point problem is then exactly the linear program \([1.3]\).

Alternatively, we can find the solution in the following way: Supposing \((x^*, \lambda^*)\) is the solution, we must have \(\nabla_x f(x^*, \lambda^*) = c - A^t \lambda^* \geq 0\). The reason is that if there is any component satisfying \(c_j - a^t_j \lambda^* < 0\), we can increase \(x_j\) by a little bit so that the non-negativity constraints for \(x\) are satisfied.\[\)

\[\)

\[\)

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]

\[\]
still satisfied but the objective value $f(x, \lambda^*)$ is decreased. Similarly, we must have $c_j - a_j^T \lambda^* \leq 0$ if $x_j > 0$; because otherwise we can decrease $x_j$ by a little bit and decrease the objective value. To this end, the saddle point satisfies:

$$(\lambda^*)^T A \leq c^T, \quad (c - A\lambda^*) \circ x^* = 0.$$  \hspace{1cm} (1.5)

Here $\circ$ denotes the componentwise product between two vectors. This indicates $f(x^*, \lambda^*) = b^T \lambda^*$; hence $\lambda$ must maximize $b^T \lambda$ among all $\lambda$ that satisfies the first constraint. That is, $\lambda^*$ is the solution to the dual linear programming problem (1.4).

Let $x$ and $y$ be the solutions to the primal and the dual problems, respectively. Previous discussion indicates that they satisfies the condition $\text{(1.5)}_2$ where $\lambda$’s role is played by $y$. Hence any basic (non-zero) variable $x_j$ corresponds to an equality constraint for $y$: $c_j - a_j^T y = 0$; and a strict inequality constraint $c_j - a_j^T y > 0$ indicates a zero component $x_j$. Thus this condition is also known as the complementary slackness condition.

To close this section, it is noted that the dual problem is often used to obtain an alternative interpretation to the primal one, especially in economics and logistics. Let us consider the diet problem again: There are $m$ foods with the unit prices $c_1, \ldots, c_m$; and each food $F_j$ contains $a_{ij}$ units of the nutrient $N_i$, $1 \leq i \leq n$. In order to keep healthy, the daily intake of the nutrient $N_i$ must be no less than $b_i$ units. Then the problem of staying healthy at the minimum daily cost is to minimize:

$$c^T x,$$

where $x_i \geq 0$ is the units of food $F_j$ consumed per day, such that:

$$\sum_{j=1}^{m} a_{ij} x_j \geq b_i \quad \forall 1 \leq i \leq n.$$

This is a standard minimum problem.

If we take the manufacturer’s point of view, say that we can produce pills for each nutrient directly, what price shall we put on each unit of the nutrient so that we can maximize the profit? Suppose the price to be charged for each unit of $N_i$ is $y_i \geq 0$, then the purpose is of course to maximize:

$$b^T y,$$

because in this model, the only purpose for the customers is to stay healthy while keeping the cost as low as possible. So they won’t purchase more than necessary to stay healthy. However, in order to keep the market active, the nutrient prices must be able to compete with the foods on the market: The customer will choose to buy $F_j$ rather than the pills if it provides the same amount of nutrients at a lower cost! Hence the prices for the pills must satisfy:

$$\sum_{i=1}^{n} a_{ij} y_i \leq c_j \quad \forall 1 \leq j \leq m.$$

Thus the problem for the manufacturer is exactly the dual standard maximize problem!
2 The Duality Theorem

First, we note that Lemma 1 translates perfectly to asymmetric duality.

Lemma 2 (Weak Duality Lemma for Asymmetric Duality). If \( x \) and \( y \) are feasible for (1.3) and (1.4), respectively, then \( c^T x \geq b^T y \).

Proof. Because \( x \geq 0 \), we have \( c^T x \geq (y^T A)x = y^T (Ax) = y^T b \).

Similar as before, if \( x_0 \) and \( y_0 \) are feasible solutions to the two linear programs and \( c^T x_0 = b^T y_0 \), then they are also optimal. Now we establish the converse of this statement in the next theorem.

Theorem 2.1 (The Duality Theorem of Linear Programming). If either (1.3) or (1.4) has a finite optimal solution, so does the other and the corresponding values of the objective functions are equal. If either problem is unbound, the other problem has no feasible solution.

Proof. The second part of the theorem is actually a direct consequence of Lemma 2. Indeed, if the primal problem (resp., the dual problem) has at least one feasible solution, its objective value provides a finite upper bound (resp., lower bound) to the dual problem (resp., the primal problem); hence the other problem cannot be unbound.

Now we consider the first part and suppose the primal problem has an optimal solution. By the fundamental theorem of linear programming, it has an optimal feasible solution and without loss of generality\(^1\) we can suppose that it is given by:

\[
A = [B \ D], \quad x = \begin{bmatrix} x_B \\ x_D \end{bmatrix}; \quad x_B = B^{-1} b \geq 0 \quad \text{and} \quad x_D = 0.
\]

By the theory of the simplex method, the relative cost coefficient vector is non-negative:

\[
r_D = c_D^T - c_B^T B^{-1} D \geq 0.
\]

Now defining \( y = c_B^T B^{-1} \), we have:

\[
y^T A = c_B^T B^{-1} [B \ D] = [c_B^T \ c_B^T B^{-1} D] = [c_B^T \ c_D^T - r_D^T] \leq c^T.
\]

Hence \( y \) is feasible for the dual problem. Furthermore, \( y^T b = c_B^T B^{-1} b = c_B^T x_B = c^T x \); by the corollary of Lemma 2, \( y \) is optimal.

Lastly, suppose the dual problem has an optimal feasible solution, then the primal problem is not unbound; the target thusly reduces to show that the primal problem has at least one feasible solution. Indeed, if this is true, by the fundamental theorem of linear programming there is a basic feasible solution, and by the simplex method we can find an optimal solution since the problem is not unbound and reduce the situation to previous case.

To show that the primal problem has a feasible solution, we use the method of contradiction and suppose that this is not true. If the primal problem is infeasible, i.e., there is no \( x \geq 0 \) that satisfies \( Ax = b \), we adopt the fact that there exists \( y_0 \in \mathbb{R}^m \) such that:

\[
y_0^T A \leq 0 \quad \text{and} \quad b^T y_0 > 0.
\]

\(^1\)We did not assume rank \( A = m \) in the duality relation. Here if \( A \) is rank-deficient, we can always eliminate some constraints without changing the nature of the problem.
This result is known as the Farkas’ lemma, which states that given a vector \((b)\) and a convex cone (in this case the linear combination of columns of \(A\) with non-negative coefficients), then either the vector is in that cone (which indicates the feasibility of the primal problem) or there is a hyperplane separating the vector and the cone. In this last scenario, we pick \(y_0\) as the unit normal vector to this hyperplane.

Now because the dual problem has an optimal solution, it must be feasible. Suppose \(y\) is a feasible solution to the dual problem, then we have for all \(\epsilon \geq 0\):

\[
(y + \epsilon y_0)^t A \leq y^t A \leq c,
\]

hence \(y + \epsilon y_0\) is feasible for all such \(\epsilon\). However, \(b^t(y + \epsilon y_0) = b^t y + \epsilon (b^t y_0) \to \infty\) as \(\epsilon \to \infty\) and the dual problem is unbound, contradiction that it has an optimal solution.

As a final remark, we relate the dual problem to the sensitivity analysis of the linear program. Suppose \(x = [x_B \ x_D]\) is an optimal basic feasible solution to the primal problem, and we want to know how the objective value changes if the right hand side \(b\) changes a little. When the change is very small, we typically expect that the set of basic solutions remain the same so that the change in the objective value can be expressed as:

\[
c^t_B \delta x_B = c^t_B B^{-1} \delta b = y^t \delta b,
\]

where \(\delta b\) is the perturbation and \(\delta x_B = B^{-1} \delta b\) is the corresponding change in the optimal solution. Hence we see that the components of the solution to the dual problem describe the rate of change in the objective value of the primal problem w.r.t. each right hand side of the equality constraints.

### 3 Concluding Remarks

SIAM has selected a top-10 list for the algorithms in the past century (SIAM News, Volume 33, Number 4):

- (1946) The Monte Carlo method (or the Metropolis algorithm).
- (1947) The simplex method.
- (1950) The Krylov subspace iteration methods.
- (1951) Decomposition approach to matrix computation (formalized by Householder!).
- (1957) Fortran optimizing compiler.
- (1962) Quicksort.
- (1965) The fast Fourier transform (FFT).

We have covered four of them in this class, and three of which concern the matrix analysis. This is not surprising, as matrix indeed plays a major role in the modern scientific computing.