13 Set Cover via Dual Fitting

In this chapter we will introduce the method of dual fitting, which helps analyze combinatorial algorithms using LP-duality theory. Using this method, we will present an alternative analysis of the natural greedy algorithm (Algorithm 2.2) for the set cover problem (Problem 2.1). Recall that in Section 2.1 we deferred giving the lower bounding method on which this algorithm was based. We will provide the answer below. The power of this approach will become apparent when we show the ease with which it extends to solving several generalizations of the set cover problem (see Section 13.2).

The method of dual fitting can be described as follows, assuming a minimization problem: The basic algorithm is combinatorial – in the case of set cover it is in fact the simple greedy algorithm. Using the linear programming relaxation of the problem and its dual, one shows that the primal integral solution found by the algorithm is fully paid for by the dual computed; however, the dual is infeasible. By fully paid for we mean that the objective function value of the primal solution found is at most the objective function value of the dual computed. The main step in the analysis consists of dividing the dual by a suitable factor and showing that the shrunk dual is feasible, i.e., it fits into the given instance. The shrunk dual is then a lower bound on OPT, and the factor is the approximation guarantee of the algorithm.

13.1 Dual-fitting-based analysis for the greedy set cover algorithm

To formulate the set cover problem as an integer program, let us assign a variable $x_S$ for each set $S \in \mathcal{S}$, which is allowed 0/1 values. This variable will be set to 1 iff set $S$ is picked in the set cover. Clearly, the constraint is that for each element $e \in U$ we want that at least one of the sets containing it be picked.

\[
\begin{align*}
\text{minimize} & \quad \sum_{S \in \mathcal{S}} c(S)x_S \\
\text{subject to} & \quad \sum_{S: e \in S} x_S \geq 1, \quad e \in U
\end{align*}
\]  

(13.1)
The LP-relaxation of this integer program is obtained by letting the domain of variables \( x_S \) be \( 1 \geq x_S \geq 0 \). Since the upper bound on \( x_S \) is redundant, we get the following LP. A solution to this LP can be viewed as a fractional set cover.

\[
\begin{align*}
\text{minimize} & \quad \sum_{S \in \mathcal{S}} c(S)x_S \\
\text{subject to} & \quad \sum_{S: e \in S} x_S \geq 1, \quad e \in U \\
& \quad x_S \geq 0, \quad S \in \mathcal{S}
\end{align*}
\] (13.2)

**Example 13.1** Let us give a simple example to show that a fractional set cover may be cheaper than the optimal integral set cover. Let \( U = \{e, f, g\} \) and the specified sets be \( S_1 = \{e, f\} \), \( S_2 = \{f, g\} \), \( S_3 = \{e, g\} \), each of unit cost. An integral cover must pick two of the sets for a cost of 2. On the other hand, picking each set to the extent of 1/2 gives a fractional cover of cost 3/2. \( \square \)

Introducing a variable \( y_e \) corresponding to each element \( e \in U \), we obtain the dual program.

\[
\begin{align*}
\text{maximize} & \quad \sum_{e \in U} y_e \\
\text{subject to} & \quad \sum_{e: e \in S} y_e \leq c(S), \quad S \in \mathcal{S} \\
& \quad y_e \geq 0, \quad e \in U
\end{align*}
\] (13.3)

Intuitively, why is LP (13.3) the dual of LP (13.2)? In our experience, this is not the right question to be asked. As stated in Section 12.1, there is a purely mechanical procedure for obtaining the dual of a linear program. Once the dual is obtained, one can devise intuitive, and possibly physically meaningful, ways of thinking about it. Using this mechanical procedure, one can obtain the dual of a complex linear program in a fairly straightforward manner. Indeed, the LP-duality-based approach derives its wide applicability from this fact.

An intuitive way of thinking about LP (13.3) is that it is packing "stuff" into elements, trying to maximize the total amount packed, subject to the constraint that no set is overpacked. A set is said to be overpacked if the total amount packed into its elements exceeds the cost of the set. Whenever the coefficients in the constraint matrix, objective function, and right-hand side are all nonnegative, the minimization LP is called a covering LP and