Lecture 8: Feb 14, 2013

1 The Assignment problem and the Hungarian Algorithm

1.1 The problem statement

The project-selection problem is as follows.
Input:

- A set of n workers: $W = \{1, 2, .., n\}$
- A set of n tasks: $T = \{1, 2, , n\}$
- For each pair of worker i and task j: t_{ij} = time (or cost) for worker i to finish task j

Objective: To minimize the total time (or cost) for finishing all n tasks
Constraint: Each worker is assigned to do one task; Each task is assigned to one worker

1.2 The Algorithm

0. Set up a cost table. Assume all costs t_{ij} are ≥ 0.
1. For each row $i = 1, \ldots, n$, let α_i = the smallest entry in row i. Update:

$$t_{ij} \leftarrow t_{ij} - \alpha_i$$

   for all entries (i, j) in row i.
2. For each col $j = 1, \ldots, n$, let β_j = the smallest entry in column j. Update:

$$t_{ij} \leftarrow t_{ij} - \beta_j$$

   for all entries (i, j) in column j.
3. Check if there is all-zero-assignment.
   - If there is one, this assignment is optimal.
   - If there isn't one: Find a zero-cver containing fewer than n lines. Then, let δ = the smallest entry not covered by any lines. Update the entries of the cost table as follows:
     $t_{ij} \leftarrow t_{ij} - \delta$ for each entry (i, j) not covered by any lines
     $t_{ij} \leftarrow t_{ij}$ for each entry (i, j) covered by exactly one line
     $t_{ij} \leftarrow t_{ij} + \delta$ for each entry (i, j) covered by exactly two lines

   (Repeat step 3 until there is an all-zero-assignment.)

1.3 Claims and proofs

In this section, we will denote an feasible solution by a vector x that has n^2 component:

$$x = (x_{11}, x_{12}, \dots, x_{1n}, x_{21}, \dots, x_{2n}, \dots, x_{nn}).$$

Observation. If $t_{ij} \geq 0$ for all pairs (i, j) and if we have an assignment $x = \{x_{ij}, \forall i = 1, \dots, n; j = 1, \dots, n\}$ of zero total cost, then x is an optimal assignment. Moreover, each pairing in this assignment must have zero cost. (I.e. if $x_{ij} = 1$, then $t_{ij} = 0$.)

Claim 1. Suppose x denote an optimal assignment to an input with edge costs $\{t_{ij}, \forall i = 1, \dots, n; j = 1, \dots, n\}$. Suppose that we subtract α_i from each entry of row i (or β_j from each entry of column j). (I.e., subtract α_i from the cost of each edge adjacent to worker-node i, or β_j from the cost of each edge adjacent to task-node j, respectively.)

Then x is still optimal for the modified input.

Proof. Let us denote the total cost of the assignment x with $v(x) = \sum_{i=1}^n \sum_{j=1}^n x_{ij} t_{ij}$. We assume that x is an optimal assignment, so for any other feasible assignment y, we know that the total cost of the assignment y, namely $v(y)$ must be greater than $v(x)$:

$$v(x) \leq v(y).$$

Now, consider the cost table after we subtract α_1 from each entry in row 1 (for simplicity of notation, we subtract from row 1 in this proof, but it can be from any row or any column). Let $\hat{t}_{ij}$ denote the (i, j)-entry of the modified table, where

$$\begin{array}{rcll} \hat{t}_{1j} & = & t_{1j} - \alpha_1 & \forall j = 1, \dots, n \\ \hat{t}_{ij} & = & t_{ij} & \forall i = 2, \dots, n; \forall j = 1, \dots, n. \end{array}$$

Then, for any feasible solution y, the total cost of the assignment y with respect to the new costs $\hat{t}_{ij}$, which we will denote with $\hat{v}(y)$, is

$$\begin{aligned} \hat{v}(y) &= \sum_{i=1}^n \sum_{j=1}^n y_{ij} \hat{t}_{ij} \\ &= \sum_{j=1}^n y_{1j} \hat{t}_{1j} + \sum_{i=2}^n \sum_{j=1}^n y_{ij} \hat{t}_{ij} \\ &= \sum_{j=1}^n y_{1j} (t_{1j} - \alpha_1) + \sum_{i=2}^n \sum_{j=1}^n y_{ij} \hat{t}_{ij} \\ &= \sum_{j=1}^n y_{1j} t_{1j} + \sum_{i=2}^n \sum_{j=1}^n y_{ij} \hat{t}_{ij} - \sum_{j=1}^n y_{1j} \alpha_1 \\ &= \left(\sum_{i=1}^n \sum_{j=1}^n y_{ij} t_{ij} \right) - \sum_{j=1}^n y_{1j} \alpha_1 \\ &= v(y) - \sum_{j=1}^n y_{1j} \alpha_1 \\ &= v(y) - \alpha_1, \end{aligned}$$

where the last equality is because among $y_{11}, y_{12}, \ldots, y_{1n}$, exactly one of them takes the value 1 while the rest are zeros.

So, for any feasible solution y, $\hat{v}(y) = v(y) - \alpha_1$. That is, the total cost of assignment y with respect to the new cost table is α_1 less than the total cost of the same assignment y with respect to the original cost table.

This is also true for the optimal solution x:

$$\hat{v}(x) = v(x) - \alpha_1.$$

So, comparing x to any feasible solution y:

$$\begin{array}{rrcl} & v(x) & \leq & v(y) \\ \Leftrightarrow & v(x) - \alpha_1 & \leq & v(y) - \alpha_1 \\ \Leftrightarrow & \hat{v}(x) & \leq & \hat{v}(y), \end{array}$$

which implies that x is still optimal after each entry of the first row is reduced by α_1. □

Claim 2.

Proof. □