# Exercise 3.6

Sunday, 10 September 2023 20:27

1. Let $x_i$ be decision variables indicating which distribution units should be build, with $1 \leq i \leq n$; the value of variable $x_i$ is 0 if distribution unit $u_i$ should not be built, and 1 if distribution unit $u_i$ should be built. Furthermore, let $y_{i,j}$ be decision variables indicating which houses should be hooked up to which distribution units; that is, $y_{i,j} = 1$ if and only if house $j$ should be connected to distribution unit $i$, for $1 \leq i \leq n$ and $1 \leq j \leq m$. (Otherwise, $y_{i,j} = 0$.)
   Then, the 0/1-LP is as follows:
   Minimize $\sum_{1\leq i\leq n}\left((x_i \cdot f_i) + \sum_{1\leq j\leq m}(y_{i,j}\cdot g_{i,j})\right)$
   Subject to:
   - $\sum_{i\in\{1\leq i\leq n \mid u_i\in U(h_j)\}} y_{i,j} \geq 1$ for all $1 \leq j \leq m$
   - $n \cdot x_i - \sum_{1\leq j\leq n} y_{i,j} \geq 0$ for $1 \leq i \leq n$
   - $x_i \in \{0,1\}$ for $1 \leq i \leq n$
   - $y_{i,j} \in \{0,1\}$ for $1 \leq i \leq n$ and $1 \leq j \leq m$

   We ignore the fact that $g_{i,j}$ is undefined for 'incompatible' combinations of distribution points and houses.
   or define $g_{i,j} = \infty$
2. ~~Take $\tau = \frac{1}{4}$ as threshold in the rounding scheme.~~ notes: always round up (i.e. threshold > 0)
   use that $x$ must be at least $\frac{1}{n}$ (if d-point is built for one connection)
   $y_{i,j}$ is at least $\frac{1}{m}$ for all $i,j$

(i) See above for the formulation. Roughly, this program can be explained as follows:

The term $x_i \cdot f_i$ gives the cost of building distribution unit $i$, if and only if it is built.

The term $\sum_{1\leq j\leq m} y_{i,j} \cdot g_{i,j}$ gives the cost of building connections between houses and the distribution point $i$, only if said connections are built. Thus, the sum of these terms gives the total cost associated with distribution point $i$; summing over all distribution points then gives the total cost, which we want to minimize.

The constraints can be explained as follows:

1. The first constraint states that every house should have at least one connection built.
2. The second constraint states that whenever a connection between house $j$ and distribution point $i$ has been built, distribution point $i$ should also be built. Note: since building connections without building the corresponding distribution point only leads to increased costs while not helping to satisfy any constraints, we can state that whenever the connection needs to be built as part of a solution, the distribution point corresponding to it needs to be built as well.

3/4. These are typical 0/1 - constraints.

(ii) The algorithm could be as follows:

Power - Costs $(U, H, \{f_1, \dots, f_n\}, \{g_{1,1}, \dots, g_{n,m}\})$:

1. $n \leftarrow |U|$
2. $m \leftarrow |H|$
3. Solve the relaxed linear program corresponding to the given problem:

   Minimize $\sum_{1\leq i\leq n}\left((x_i \cdot f_i) + \sum_{1\leq j\leq m}(y_{i,j}\cdot g_{i,j})\right)$
   Subject to:
   - $\sum_{i\in\{1\leq i\leq n \mid u_i\in U(h_j)\}} y_{i,j} \geq 1$ for all $1 \leq j \leq m$
   - $n \cdot x_i - \sum_{1\leq j\leq n} y_{i,j} \geq 0$ for $1 \leq i \leq n$ alternatively: $y_{i,j} \leq x_i$ for all $(u_i, h_j) \in U \times H$
   - $x_i \in \{0,1\}$ for $1 \leq i \leq n$
   - $y_{i,j} \in \{0,1\}$ for $1 \leq i \leq n$ and $1 \leq j \leq m$

4. $solU \leftarrow \{u_i \in U \mid x_i \geq \cancel{\frac{1}{n^2}}\}$ using alternative constraint: $\frac{1}{4}$ # set of distribution points to build
5. $solC \leftarrow \{(u_i, h_j) \in U \times H \mid y_{i,j} \geq \frac{1}{n}\}$ # set of connections to build
   $\frac{1}{n} \rightarrow \frac{1}{4}$, since we know there are at most 4 viable elements (by the assumption in the question, i.e. $|U(h_j)| \leq 4$)
6. return $(solU, solC)$

Proof of correctness

To see why this algorithm returns a valid solution, we prove the following statements:

1. If a connection between a house $h_j$ and a distribution point $u_i$ is built, then the distribution point $u_i$ is built.
2. All houses are connected to at least one distribution point.

If both of these statements hold, then, by combining these statements, we find that all houses are connected to at least one distribution point which has been built. This ensures the solution is valid.

Hence, it remains to be shown that the two statements hold. We first prove the second statement. This statement follows from the first constraint in the linear program; by the first constraint, we have that the sum over all the decision variables $y_{i,j}$ *(which determine whether a connection between a house $h_j$ and a distribution point $u_i$ is built)* must be at least 1. Given that the sum can range over at most $n$ distribution points (when $U = U(h_j)$), it must be that the average value of the decision variables is at least $\frac{1}{n}$. But this also implies that at least one of the variables must have a value of at least $\frac{1}{n}$, i.e. that at least one connection between the given house $h_j$ and some distribution point is built (since at least one connection will then be included in `solC`). Since this holds for all houses, we have that the second statement holds.
To see why the first statement holds, we consider the second and fourth constraints. From the fourth constraint, we see that the decision variables $y_{i,j}$ cannot be negative; this holds even in the relaxed LP. From line 5 of the algorithm, we know that a connection between house $h_j$ and distribution point $u_i$ is built if and only if the value of the decision variable $y_{i,j}$ is at least $\frac{1}{n}$. This means that the sum over all decision variables $y_{i,j}$ in the second constraint is at least $\frac{1}{n}$ whenever the distribtion point $u_i$ needs to be built (i.e. because some house $h_j$ wants to connect to $u_i$). Now, we note that the second constraint states that $n \cdot x_i$ should be at least as large as the sum of the decision variables $y_{i,j}$. In other words, if the sum over the decision variables is at least $\frac{1}{n}$, then we need to have that $n \cdot x_i \geq \frac{1}{n} \Rightarrow x_i \geq \frac{1}{n^2}$. But then, by line 4 of the algorithm, we have that the distribution point $u_i$ is included in the set of distribution points which need to be built. This means that, whenever a connection between a house $h_j$ and a distribution point $u_i$ is built, then the distribution point $u_i$ is built as well. This proves the first statement.

Now that we have proven the first and second statement, we can conclude that the algorithm (by the reasoning given above) must give a valid solution.

Proof of bound on approximation ratio

We have that $OPT \geq OPT_{relaxed} = \sum_{1\leq i\leq n}\left((x_i f_i) + \sum_{1\leq j\leq m}(y_{i,j}\, g_{i,j})\right)$

Furthermore, we have that Power - Costs $(\dots)$

$$
\begin{aligned}
&= \sum_{i\in\{1\leq i\leq n \mid u_i\in solU\}} f_i + \sum_{(i,j)\in\{(i,j)\in\{i \mid 1\leq i\leq n\}\times\{j \mid 1\leq j\leq m\} \mid (u_i,h_j)\in solC\}} g_{i,j} \\
&\leq \sum_{i\in\{1\leq i\leq n \mid u_i\in solU\}} f_i\, n x_i + \sum_{(i,j)\in\{(i,j)\in\{i \mid 1\leq i\leq n\}\times\{j \mid 1\leq j\leq m\} \mid (u_i,h_j)\in solC\}} g_{i,j}\, n^2 y_{i,j} \\
&\leq \sum_{1\leq i\leq n} f_i\, n x_i + \sum_{(i,j)\in\{(i,j)\in\{i \mid 1\leq i\leq n\}\times\{j \mid 1\leq j\leq m\}\}} g_{i,j}\, n^2 y_{i,j} \\
&= n \sum_{1\leq i\leq n} f_i x_i + n^2 \sum_{(i,j)\in\{(i,j)\in\{i \mid 1\leq i\leq n\}\times\{j \mid 1\leq j\leq m\}\}} g_{i,j}\, y_{i,j} \\
&\leq n^2 \cdot \left(\sum_{1\leq i\leq n} f_i x_i + \sum_{(i,j)\in\{(i,j)\in\{i \mid 1\leq i\leq n\}\times\{j \mid 1\leq j\leq m\}\}} g_{i,j}\, y_{i,j}\right) \\
&= n^2 \left(\left(\sum_{1\leq i\leq n}(x_i f_i)\right) + \left(\sum_{1\leq i\leq n}\sum_{1\leq j\leq m}(y_{i,j}\, g_{i,j})\right)\right) \\
&= n^2 \left(\sum_{1\leq i\leq n}\left((x_i f_i) + \sum_{1\leq j\leq m}(y_{i,j}\, g_{i,j})\right)\right) \\
&= n^2\, OPT_{relaxed} \\
&\leq n^2\, OPT
\end{aligned}
$$

Hence, Power - Costs is an $n^2$-approximation algorithm.