1 Introduction

For a formal introduction to the pipage rounding framework, refer to the reading notes from [2]. In these notes, we are only going to focus on reformulating section 2 of [1] as the more commonly studied max cover problem.

The proof of the pipage rounding algorithm are a little more efficient here than in [2]. Indeed, we show that the number of integral variables of the fractional vector we wish to make integral increases by at least 1 at every iteration, rather than every $n$ iteration, where $n$ is the number of elements in the universe set $N$. We also show a better finite-sample approximation ratio than $(1 - 1/e)$, which was obtained in the general case for matroid constraints.

Since we know max cover can be solved within $1 - (1 - p)^p$, where $p$ is the number of sets we are allowed to take, does pipage rounding improve the approximation factor? Maybe. In fact, we will show that pipage rounding gets us within $1 - (1 - k)^k$ of the optimum, where $k$ is the largest number of sets any item in $N$ is covered by. Both converge to the magical $(1 - 1/e)$-factor, but depending on the nature of the problem, pipage rounding can improve on the greedy algorithm. For example, in the case where no item in $N$ is covered by more than 2 sets, pipage rounding achieves a $3/4$-approximation ratio, whereas this is only the case when we are allowed to pick two sets for the greedy algorithm.

2 Formulation

We want to maximize a coverage function. There is a set of $N$ items in the universe, and a series of sets $S$ on $N$. Our objective is to choose $p$ sets, such that the sum of the weights of the elements covered is maximized. This can be expressed in the following integer program where $z_j$ is the indicator variable for the $j^{th}$ element being covered, $w_j$ is the weight of that element, and $x_i$ is the indicator variable for the $i^{th}$ set being picked.

\[
\begin{align*}
\text{max} & \quad \sum_{j=1}^{m} w_j z_j \\
\text{s.t.} & \quad \forall j, \sum_{i: j \in S_i} x_i \geq z_j \quad (1) \\
& \quad \sum_{i} x_i = p \quad (2) \\
& \quad \forall i, x_i \in \{0, 1\} \quad (3) \\
& \quad \forall j, 0 \leq z_j \leq 1 \quad (4)
\end{align*}
\]

We can define the function

\[
F(x) := \sum_{j=1}^{m} w_j \left(1 - \prod_{i: j \in S_i} (1 - x_i)\right)
\]

and reformulate this LP as:

\[
\begin{align*}
\text{max} & \quad F(x) \\
\text{s.t.} & \quad \sum_{i} x_i = p \quad (5) \\
& \quad \forall i, x_i \in \{0, 1\} \quad (6)
\end{align*}
\]
By relaxing (6) to \( y \in [0, 1] \), the function \( F \) defined above is exactly the multilinear relaxation of \( f \). We are going to define a particular extension to \( f \):

\[
\tilde{f}(x) = \sum_{j=1}^{m} w_j \min\{1, \sum_{i:j \in S_i} x_i\}
\]

3 Algorithm

The general framework of pipage rounding is the same:

- Find \( \hat{y} \) which maximises \( \tilde{f} \) under the relaxed max cover constraints.
- Transform \( \hat{y} \) into an integral vector \( \hat{y} \), such that \( F(\hat{y}) \geq F(\tilde{f}) \)

We will show that we have this series of inequalities:

\[
f(\hat{y}) = F(\hat{y}) \geq F(\tilde{f}) \geq (1 - 1/e)F(\tilde{f}) \geq (1 - 1/e)OPT
\]

The algorithm is very similar to before. Suppose that we are given a fractional vector \( y \), which optimizes \( \tilde{f} \) under (5) and a relaxed (6). \( y \) cannot have exactly one fractional variable, otherwise (5) would not be matched exactly. If \( y \) is not integral, then we can find two coordinates \( i \) and \( j \), which are fractional. Note that exchanging mass from one to the other, as long as we verify \( 0 \leq y_i \leq 1 \) and \( 0 \leq y_j \leq 1 \) (which are not matched exactly), does not violate any constraint. Let \( \epsilon^+ > 0 \) (resp. \( \epsilon^- < 0 \)) be the exchange of mass which makes \( y_j \) integral, (resp \( y_i \) integral).

Since \( F \) is the multilinear relaxation of \( f \), it is cross-convex, and \( \phi : t \mapsto F(\hat{y} + t(e_i - e_j)) \) reaches its maximum at one of the two endpoints of segment \([\epsilon^-, \epsilon^+]\). If \( \phi(\epsilon^+) > \phi(\epsilon^-) \), we let \( \tilde{y} \leftarrow \tilde{y} + \epsilon^+(e_i - e_j) \), and vice-versa otherwise. We repeat the process until \( \tilde{y} \) has no more fractional variables. We let \( \hat{y} \) be the output of this pipage rounding procedure. At every iteration, we gain an additional integral variable. At the end of at most \( n \) iterations, we have integral vector \( \hat{y} \) such that \( F(\hat{y}) > F(\tilde{f}) \)

4 Theoretical Guarantees

Let \( k := \max_j |S : j \in S| \) be the largest number of sets an item in \( N \) is covered by. We are now going to prove that:

\[\forall y \in [0, 1]^N, F(y) \geq (1 - (1-k)^k)\tilde{f}(y)\]

From the arithmetic-geometric inequality,

\[
\sqrt[\prod_{i=1}^{k}(1 - y_i)} \leq \frac{1}{k} \sum_{i=1}^{k}(1 - y_i)
\]

which implies that:

\[
1 - \prod_{i=1}^{k}(1 - y_i) \geq 1 - \left(1 - \frac{\sum_{i=1}^{k} y_i}{k}\right)^k
\]

Let \( z := \min(1, \sum_{i=1}^{k} y_i) \). Since \( \phi : z \mapsto 1 - (1 - \frac{z}{k})^k \) is monotone increasing, we have that:

\[
1 - \prod_{i=1}^{k}(1 - y_i) \geq 1 - \left(1 - \frac{z}{k}\right)^k
\]

Note that \( \phi \) is concave on the segment \([0, 1]\), that \( \phi(0) = 0 \) and \( \phi(1) = 1 - (1 - 1/k)^k \). We finally obtain:

\[
g(z) \geq \left(1 - (1 - 1/k)^k\right) \cdot z
\]

which implies that:

\[
1 - \prod_{i=1}^{k}(1 - y_i) \geq \left(1 - (1 - \frac{1}{k})^k\right) \min(1, \sum_{i=1}^{k} y_i)
\]

and therefore that: \( F(y) \geq (1 - (1/k)^k)\tilde{f}(y) \)
5 Lower Bound

We now demonstrate the integrality gap can be arbitrarily close to $1 - (1 - 1/k)^k$). Set $n = kp$, $\forall j, w_j = 1$. Suppose there are $n$ sets and $\binom{n}{k}$ items in $N$, such that each item is covered by exactly $k$ sets of the same size $\binom{k}{n}$. In other words, you can think of a bipartite graph, where one side (items) corresponds to all the subsets of size $k$ of the other side (sets). By symmetry, choosing any $p$ sets amongst the $n$ covers $\binom{n}{k} - \binom{n-p}{k}$ items. Choosing every set with fraction $1/k$ covers all items with probability 1, i.e. $\binom{n}{k}$ items. This does not violate any constraint since $n/k = p$. Standard algebraic manipulations show that:

$$\frac{\binom{n}{k} - \binom{n-p}{k}}{\binom{n}{k}} \leq 1 - \left(1 - \frac{1}{k} - \frac{k + 1}{n}\right)^k$$

$$\rightarrow_{n\rightarrow+\infty} 1 - (1 - \frac{1}{k})^k$$ (8)

References
