Randomized Pipage Rounding for Matroid Polytopes and Applications

Chandra Chekuri*   Jan Vondrák†

September 23, 2009

Abstract

We present concentration bounds for linear functions of random variables arising from the pipage rounding procedure on matroid polytopes.

As an application, we give a $(1 - 1/e - \epsilon)$-approximation algorithm for the problem of maximizing a monotone submodular function subject to 1 matroid and k linear constraints, for any constant $k \geq 1$ and $\epsilon > 0$. This generalizes the result for k linear constraints by Kulik et al. [11]. We also give the same result for a super-constant number k of "loose" linear constraints, where the right-hand side dominates the matrix entries by an $\Omega(\epsilon^{-2} \log k)$ factor.

As another application, we present a general result on minimax packing problems that involve a matroid base constraint. An example is the multi-path routing problem with integer demands for pairs of vertices; the goal is to minimize congestion. We give an $O(\log m/ \log\log m)$-approximation for the general problem $\min\{\lambda : \exists x \in \{0,1\}^N, x \in B(\mathcal{M}), Ax \leq \lambda b\}$ where m is the number of packing constraints.

1 Introduction

Pipage rounding is a procedure which aims to convert a fractional solution of an optimization problem into an integral one, through a sequence of simple updates. Unlike other rounding techniques that are typically used for linear programs, pipage rounding is flexible enough that it can be used even with non-linear objective functions. The analysis of pipage rounding relies on certain convexity properties of the objective function which make it possible to compare the value of the fractional solution to the value of the rounded one. However, interesting applications can be obtained even with linear objective functions.

1.1 Background

Pipage rounding was introduced by Ageev and Sviridenko [1], who used it for rounding fractional solutions in the bipartite matching polytope. They used an LP to obtain a fractional solution to a certain problem, but the rounding procedure was based on an auxiliar (non-linear) objective. The auxiliary objective $F(x)$ was defined in such a way that $F(x)$ would always increase or stay constant throughout the rounding procedure. A comparison between $F(x)$ and the original objective yields an approximation guarantee.

*Dept. of Computer Science, Univ. of Illinois, Urbana, IL 61801. Partially supported by NSF grant CCF-0728782. E-mail: chekuri@cs.illinois.edu

†IBM Almaden Research Center, San Jose, CA 95120. E-mail: jvondrak@us.ibm.com

Srinivasan [23] and Gandhi et al. [8] considered variants of dependent randomized rounding similar to pipage rounding. In this case, each rounding step is random and oblivious (independent of the objective function). Randomization at this stage does not seem necessary, but it has certain advantages - namely, the rounded solution has certain *negative correlation* properties, implying concentration bounds which can be used to deal with additional constraints.

Calinescu et al. [2] adapted the pipage rounding technique to problems involving a matroid constraint rather than bipartite matchings. Moreover, they showed that the necessary convexity properties are satisfied whenever the auxiliary function $F(x)$ is a *multilinear extension of a submodular function*. This turned out to be crucial for further developments on submodular maximization problems - in particular an optimal $(1-1/e)$-approximation for maximizing a monotone submodular function subject to a matroid constraint [24, 3], and a $(1-1/e-\epsilon)$-approximation for maximizing a monotone submodular function subject to a constant number of linear constraints [11]. In this paper, we consider a common generalization of these two problems.

The pipage rounding technique as presented in [2] is a deterministic procedure (apart from the evaluation of $F(x)$ which uses random sampling). However, it can be randomized similarly to Srinivasan's work [23], and this is the variant presented in [3]. This variant starts with a fractional solution in the matroid base polytope, $y \in B(\mathcal{M})$, and produces a random base $B \in \mathcal{M}$ such that $\mathbf{E}[f(B)] \geq F(y)$. A further rounding stage is needed in case the starting point is inside the matroid polytope $P(\mathcal{M})$ rather than the matroid base polytope $B(\mathcal{M})$; pipage rounding has been extended to this case in [25]. We summarize the known properties of pipage rounding in the matroid polytope in the following lemma [3, 25].

Lemma 1.1. *For any starting point $y \in P(\mathcal{M})$, the pipage rounding technique produces a random set S independent in $\mathcal{M}$, such that for any submodular function $f(S)$ and its multilinear extension $F(y)$, $\mathbf{E}[f(S)] \geq F(y)$.*

In addition, if the starting point is in the matroid base polytope $B(\mathcal{M})$, the rounded solution S is a random base of $\mathcal{M}$.

While randomized pipage rounding gives an approximation guarantee only in expectation, it has the added benefit that costly evaluations of $F(x)$ are not needed at all in the rounding stage. Further benefits arise from the correlation properties of the rounded solution. This is the focus of this paper.

1.2 Our work

In this paper, we revisit the technique of randomized pipage rounding in the context of matroid polytopes. In particular, we are interested in the correlation properties of the integral solution obtained by randomized pipage rounding. Our first result is that the 0/1 random variables $X_1, \ldots, X_n$ associated with the rounded solution are *negatively correlated*.

Theorem 1.2. *Let $(x_1, \ldots, x_n) \in P(\mathcal{M})$ be a fractional solution in the matroid polytope and $(X_1, \ldots, X_n) \in \{0,1\}^n$ the integral solution obtained by randomized pipage rounding. Then $\mathbf{E}[X_i] = x_i$, and for any $T \subseteq [n]$,*

- $\mathbf{E}[\prod_{i \in T} X_i] \leq \prod_{i \in T} x_i$,
- $\mathbf{E}[\prod_{i \in T}(1 - X_i)] \leq \prod_{i \in T}(1 - x_i)$.

This yields Chernoff-type concentration bounds for any linear function of $X_1, \dots, X_n$, as proved by Panconesi and Srinivasan [17]. We refer to Theorem 3.1 in [8], which together with Theorem 1.2 implies the following.

Corollary 1.3. *Let $a_i \in [0,1]$ and $X = \sum a_i X_i$, where $(X_1, \dots, X_n)$ are obtained by randomized pipage rounding from a fractional vector $(x_1, \dots, x_n) \in P(\mathcal{M})$.*

- *If $\delta \geq 0$ and $\mu \geq \mathbf{E}[X] = \sum a_i x_i$, then*

$$\Pr[X \geq (1+\delta)\mu] \leq \left(\frac{e^\delta}{(1+\delta)^{1+\delta}}\right)^\mu,$$

- *If $\delta \in [0,1]$, and $\mu \leq \mathbf{E}[X] = \sum a_i x_i$, then*

$$\Pr[X \leq (1-\delta)\mu] \leq e^{-\mu\delta^2/2}.$$

It is well-known that for $\delta \in [0,1]$, the first bound can be simplified as follows:

$$\Pr[X \geq (1+\delta)\mu] \leq e^{-\mu\delta^2/3}.$$

In particular, these bounds hold for $X = \sum_{i \in S} X_i$ where S is an arbitrary subset of the variables. We remark that in contrast, when randomized pipage rounding is performed on bipartite graphs, negative correlation holds only for subsets of edges incident to a fixed vertex [8].

More generally, we can consider concentration properties for a submodular function $f(S)$, where S is the outcome of a certain random process. Equivalently, we can also write $f(S) = f(X_1, X_2, \dots, X_n)$ where $X_i \in \{0,1\}$ is a random variable indicating whether $i \in S$. First, we consider the scenario where $X_1, \dots, X_n$ are independent random variables. We prove that in this case, Chernoff-type bounds hold for $f(X_1, X_2, \dots, X_n)$ just like they would for a linear function.

Theorem 1.4. *Let $f : \{0,1\}^n \to \mathbb{R}_+$ be a monotone submodular function with marginal values in $[0,1]$. Let $X_1, \dots, X_n$ be independent random variables in $\{0,1\}$. Let $\mu = \mathbf{E}[f(X_1, X_2, \dots, X_n)]$. Then for any $\delta > 0$,*

- $\Pr[f(X_1, \dots, X_n) \geq (1+\delta)\mu] \leq \left(\frac{e^\delta}{(1+\delta)^{1+\delta}}\right)^\mu$.
- $\Pr[f(X_1, \dots, X_n) \leq (1-\delta)\mu] \leq e^{-\mu\delta^2/2}$.

A natural question is whether the concentration bounds for submodular functions also hold for dependent random variables $X_1, \dots, X_n$ arising from pipage rounding. Currently we do not know how to prove the above bounds in this case. We remark that Theorem 1.4 can be used to simplify some previous results for submodular maximization under linear constraints, where variables are rounded independently [11, 12]. However, this concentration bound is not necessary for these applications, and we do not actually use Theorem 1.4 in our applications either. We believe that it might be useful for future applications.

Applications.

- *Submodular maximization subject to* 1 *matroid and* k *linear constraints.* We consider the problem of maximizing a submodular function subject to a matroid constraint and a given set of linear packing constraints. More formally, the problem is $\max f(x), Ax \leq b, x \in P(\mathcal{M}), x \in \{0,1\}^n$ where $f : 2^N \to \mathbb{R}$ is a monotone submodular function, $\mathcal{M}$ is a matroid on N, A is a $k \times n$ non-negative matrix and b is a $1 \times n$ non-negative matrix. N is the ground set of f and $n = |N|$. For any fixed $\epsilon > 0$, we obtain a $(1 - 1/e - \epsilon)$ approximation in two settings. First, when k is a fixed constant independent of n. Second when the constraints are sufficiently "loose": $b_i = \Omega(\epsilon^{-2} \log k) \cdot A_{ij}$ for all i, j. Note that the approximation in both cases is optimal up to the arbitrarily small ϵ (even for 1 matroid or 1 linear constraint [16, 6]), and generalizes the previously known results in the special cases of 1 matroid [3] and a fixed number of k linear constraints [11].

- *Minimax Integer Programs subject to a matroid constraint.* Let $\mathcal{M}$ be a matroid on a ground set N. Let $B(\mathcal{M})$ be the base polytope of $\mathcal{M}$. We consider the problem $\min \lambda, Ax \leq \lambda b, x \in B(\mathcal{M}), x \in \{0,1\}^N$ where A is an $m \times N$ non-negative matrix and $b \in \mathbb{R}^n$. We give an $O(\log m / \log \log m)$ approximation for this problem, and a similar result for the min-cost version (with given packing constraints and element costs). This generalizes earlier results on minimax integer programs which were considered in the context of routing and partitioning problems [19, 14, 22, 23, 8]; the underlying matroid in these setting is the partition matroid. Several of the applications in [23, 8] are more naturally viewed as pipage rounding subject to a matroid constraint. We elaborate on this in Section 5.

The rest of the paper is organized as follows. In Section 2, we present the necessary definitions. In Section 3, we prove the property of negative correlations for randomized pipage rounding. In Section 4, we present our algorithm for maximizing a monotone submodular function subject to 1 matroid and k linear constraints. In Section 5, we present our results on minimax integer programs. In Appendix A, we give a complete description of randomized pipage rounding. In Appendix B, we present our concentration bounds for submodular functions.

2 Preliminaries

Matroid polytopes. Given a matroid $\mathcal{M} = (N, \mathcal{I})$, the matroid polytope is the convex hull of vectors corresponding to independent sets in $\mathcal{M}$:

$$P(\mathcal{M}) = \text{conv}\{\mathbf{1}_I : I \in \mathcal{I}\} = \{x \geq 0 : \forall S; \sum_{i \in S} x_i \leq r(S)\}$$

where $r(S)$ is the rank function of the matroid $\mathcal{M}$ [5]. Let $\mathcal{B}$ denote the collection of *bases* of $\mathcal{M}$, i.e. independent sets of maximum cardinality. We also work with the matroid base polytope, which is the convex hull of vertices corresponding to bases:

$$B(\mathcal{M}) = \text{conv}\{\mathbf{1}_B : B \in \mathcal{B}\} = P(\mathcal{M}) \cap \{x : \sum_{i \in N} x_i = r(N)\}.$$

Submodular functions. A function $f : 2^N \to \mathbb{R}$ is submodular if for any $A, B \subseteq N$,

$$f(A \cup B) + f(A \cap B) \leq f(A) + f(B).$$

In addition, f is monotone if $f(S) \le f(T)$ whenever $S \subseteq T$. We denote by $f_A(i) = f(A+i) - f(A)$ the *marginal value* of i with respect to A.

An important concept in recent work on submodular functions [2, 24, 3, 11, 12, 25] is the *multilinear extension* of a submodular function:

$$F(x) = \sum_{S \subseteq N} f(S) \prod_{i \in S} x_i \prod_{i \in N \setminus S} (1 - x_i).$$

Pipage rounding. The pipage rounding technique serves to round a fractional solution of the problem

$$\max\{F(x) : x \in P(\mathcal{M})\}$$

to an integral one. In its randomized version, it is entirely oblivious of the objective function $F(x)$ and produces a random vertex of $P(\mathcal{M})$, with a distribution depending only on the starting point $x \in P(\mathcal{M})$. If the starting point is in the matroid base polytope $B(\mathcal{M})$, the rounded solution is a (random) base of $\mathcal{M}$.

We already mentioned Lemma 1.1 [3, 25] which is crucial for the purpose of maximizing submodular functions subject to a matroid constraint: it says that the expected value of the rounded solution is $\mathbf{E}[f(S)] \ge F(x)$. We give a complete description of the pipage rounding technique in the appendix.

3 Negative correlation for pipage rounding

In this section, we prove Theorem 1.2 which states that the random variables corresponding to the rounded solution are *negatively correlated*. The proof follows the same lines as [8] in the case of bipartite graphs. The intuitive reason for negative correlation is that whenever a pair of variables is being modified, their sum remains constant. Hence, knowing that one variable is high can only make the expectation of another variable lower.

Proof. Let $X_{i,t}$ denote the value of X_i after t steps of randomized pipage rounding. We use the following properties of the pipage rounding technique (see Appendix A):

- In each step, at most two variables are modified.
- Given $X_{i,t}$, its expectation in the next step remains the same: $\mathbf{E}[X_{i,t+1} \mid X_{i,t}] = X_{i,t}$.
- If two variables X_i, X_j are modified, it happens in such a way that their sum is preserved with probability 1: $X_{i,t+1} + X_{j,t+1} = X_{i,t} + X_{j,t}$.

We are interested in the quantity $Y_t = \prod_{i \in S} X_{i,t}$. At the beginning of the process, we have $X_{i,0} = x_i$ and $Y_0 = \prod_{i \in S} x_i$. The main claim is that for each t, we have $\mathbf{E}[Y_{t+1} | Y_t] \le Y_t$.

Let us condition on a particular configuration of variables at time t, $(X_{1,t}, \dots, X_{n,t})$. We consider three cases:

- If no variable X_i, $i \in S$, is modified in step t, we have $Y_{t+1} = \prod_{i \in S} X_{i,t+1} = \prod_{i \in S} X_{i,t}$.
- If exactly one variable X_i, $i \in S$, is modified in step t, then we use the property that $\mathbf{E}[X_{i,t+1} \mid X_{i,t}] = X_{i,t}$:

$$\mathbf{E}[Y_{t+1} \mid X_{1,t}, \dots, X_{n,t}] = \mathbf{E}[X_{i,t+1} \mid X_{i,t}] \cdot \prod_{j \in S \setminus \{i\}} X_{j,t} = \prod_{j \in S} X_{j,t}.$$

- If two variables X_i, X_j, $i, j \in S$, are modified in step t, we use the property that their sum is preserved: $X_{i,t+1} + X_{j,t+1} = X_{i,t} + X_{j,t}$. This also implies that

$$\mathbf{E}[(X_{i,t+1} + X_{j,t+1})^2 \mid X_{i,t}, X_{j,t}] = (X_{i,t} + X_{j,t})^2. \tag{1}$$

  On the other hand, the value of each variable is preserved in expectation. Applying this to their difference, we get $\mathbf{E}[X_{i,t+1} - X_{j,t+1} \mid X_{i,t}, X_{j,t}] = X_{i,t} - X_{j,t}$. Since $\mathbf{E}[Z^2] \geq (\mathbf{E}[Z])^2$ holds for any random variable, we get

$$\mathbf{E}[(X_{i,t+1} - X_{j,t+1})^2 \mid X_{i,t}, X_{j,t}] \geq (X_{i,t} - X_{j,t})^2. \tag{2}$$

  Combining (1) and (2), and using the formula $XY = \frac{1}{4}((X+Y)^2 - (X-Y)^2)$, we get

$$\mathbf{E}[X_{i,t+1} X_{j,t+1} \mid X_{i,t}, X_{j,t}] \leq X_{i,t} X_{j,t}.$$

  Therefore,

$$\mathbf{E}[Y_{t+1} \mid X_{1,t}, \ldots, X_{n,t}] = \mathbf{E}[X_{i,t+1} X_{j,t+1} \mid X_{i,t}, X_{j,t}] \cdot \prod_{k \in S \setminus \{i,j\}} X_{k,t} \leq \prod_{k \in S} X_{k,t}.$$

Therefore, in every case $\mathbf{E}[Y_{t+1} \mid X_{1,t}, \ldots, X_{n,t}] \leq \prod_{k \in S} X_{k,t}$. By taking expectation over all configurations at time t with a fixed value of $\prod_{k \in S} X_{k,t} = Y_t$, we obtain $\mathbf{E}[Y_{t+1}] \leq \mathbf{E}[Y_t]$. Consequently, $\mathbf{E}[Y_t] \leq \mathbf{E}[Y_{t-1}] \leq \ldots \leq \mathbf{E}[Y_1] \leq Y_0$ and at the end of the procedure (after a maximum number of $t^* = n^2$ steps), we get

$$\mathbf{E}[\prod_{i \in S} X_i] = \mathbf{E}[Y_{t^*}] \leq Y_0 = \prod_{i \in S} x_i.$$

The same proof applies when we replace X_i by $1 - X_i$, because the properties of pipage rounding that we used hold for $1 - X_i$ as well. $\square$

As we mentioned in Section 1, this implies strong concentration bounds for linear functions of the variables $X_1, \ldots, X_n$ (Corollary 1.3).

4 Submodular maximization subject to 1 matroid and k linear constraints

In this section, we present an algorithm for the problem of maximizing a monotone submodular function subject to 1 matroid and k linear ("knapsack") constraints.

Problem definition. *Given a monotone submodular function $f : 2^N \to \mathbb{R}_+$ (by a value oracle), and a matroid $\mathcal{M} = (N, \mathcal{I})$ (by an independence oracle). For each $i \in N$, we have k parameters c_{ij}, $1 \leq j \leq k$. A set $S \subseteq N$ is feasible if $S \in \mathcal{I}$ and $\sum_{i \in S} c_{ij} \leq 1$ for each $1 \leq j \leq k$.*

Kulik et al. gave a $(1 - 1/e - \epsilon)$-approximation for the same problem with a constant number of linear constraints, but *without* the matroid constraint [11]. Gupta, Nagarajan and Ravi [9] show that a knapsack constraint can in a technical sense be simulated in a black-box fashion by a collection of partition matroid constraints. Using their reduction and known results on submodular set function maximization subject to matroid constraints [7, 13], they obtain a $1/(p+q+1)$ approximation with p knapsacks and q matroids for any $q \geq 1$ and fixed $p \geq 1$ (or $1/(p+q+\epsilon)$ for any fixed $p \geq 1, q \geq 2$ and $\epsilon > 0$).

4.1 Constant number of knapsack constraints

We consider first 1 matroid and a constant number k of linear constraints, in which case each linear constraint is thought of as a "knapsack" constraint. We show a $(1-1/e-\epsilon)$-approximation in this case, building upon the algorithm of Kulik, Shachnai and Tamir [11], which works for k knapsack constraints (without a matroid constraint). The basic idea is that we can add the knapsack constraints to the multilinear optimization problem

$$\max\{F(x) : x \in P(\mathcal{M})\}$$

which is used to achieve a $(1-1/e)$-approximation for 1 matroid constraint [3]. Using standard techniques (partial enumeration), we get rid of all items of large value or size, and then scale down the constraints a little bit, so that we have some room for overflow in the rounding stage. We can still solve the multilinear optimization problem within a factor of $1-1/e$ and then round the fractional solution using pipage rounding. Using the fact that randomized pipage rounding makes the size in each knapsack strongly concentrated, we conclude that our solution is feasible with constant probability.

Algorithm.

- Assume $0 < \epsilon < \min\{1/k^2, 0.001\}$. Enumerate all sets A of at most $1/\epsilon^4$ items which form a feasible solution. (We are trying to guess the most valuable items in the optimal solution under a greedy ordering.) For each candidate set A, repeat the following.

- Let $\mathcal{M}' = \mathcal{M}/A$ be the matroid where A has been contracted. For each $1 \leq j \leq k$, let $C_j = 1 - \sum_{i \in A} c_{ij}$ be the remaining capacity in knapsack j. Let B be the set of items $i \notin A$ such that $c_{ij} > k\epsilon^3 C_j$ for some j (the item is too big in some knapsack). Throw away all the items in B.

- We consider a reduced problem on the item set $N \setminus (A \cup B)$, with the matroid constraint $\mathcal{M}'$, knapsack capacities C_j, and objective function $g(S) = f(A \cup S) - f(A)$. Define a polytope

$$P' = \left\{x \in P(\mathcal{M}') : \forall j; \sum c_{ij} x_i \leq C_j\right\} \tag{3}$$

  where $P(\mathcal{M}')$ is the matroid polytope of $\mathcal{M}'$. We solve (approximately) the following optimization problem:

$$\max\left\{G(x) : x \in (1-\epsilon)P'\right\} \tag{4}$$

  where $G(x) = \mathbf{E}[g(\hat{x})]$ is the multilinear extension of $g(S)$. Since linear functions can be optimized over P' in polynomial time, we can use the continuous greedy algorithm [24] to find a fractional solution x^* within a factor of $1-1/e$ of optimal.

- Given a fractional solution x^*, we apply randomized pipage rounding to x^* with respect to the matroid polytope $P(\mathcal{M}')$. Call the resulting set R_A. Among all candidate sets A such that $A \cup R_A$ is feasible, return the one maximizing $f(A \cup R_A)$.

We remark that the value of this algorithm (unlike the $(1-1/e)$-approximation for 1 matroid constraint) is purely theoretical, as it relies on enumeration of a huge (constant) number of elements.

Theorem 4.1. *The algorithm above returns a solution of expected value at least* $(1-1/e-3\epsilon)OPT$.

Proof. Consider an optimum solution O, i.e. $OPT = f(O)$. Order the elements of O greedily by decreasing marginal values, and let $A \subseteq O$ be the elements whose marginal value is at least $\epsilon^4 OPT$. There can be at most $1/\epsilon^4$ such elements, and so the algorithm will consider them as one of the candidate sets. We assume in the following that this is the set A chosen by the algorithm.

We consider the reduced instance, where $\mathcal{M}' = \mathcal{M}/A$ and the knapsack capacities are $C_j = 1 - \sum_{i \in A} c_{ij}$. $O \setminus A$ is a feasible solution for this instance and we have $g(O \setminus A) = f_A(O \setminus A) = OPT - f(A)$. We know that in $O \setminus A$, there are no items of marginal value more than the last item in A. In particular, $f_A(i) \leq \epsilon^4 OPT$ for all $i \in O \setminus A$. We throw away the set $B \subseteq N \setminus A$ of items whose size in some knapsack is more then $k\epsilon^3 C_j$. In $O \setminus A$, there can be at most $1/(k\epsilon^3)$ such items for each knapsack, i.e. $1/\epsilon^3$ items in total. Since their marginal values with respect to A are bounded by $\epsilon^4 OPT$, these items together have value $g(O \cap B) = f_A(O \cap B) \leq \epsilon OPT$. $O' = O \setminus (A \cup B)$ is still a feasible set for the reduced problem, and using submodularity, its value is

$$g(O') = G((O \setminus A) \setminus (O \cap B)) \geq g(O \setminus A) - g(O \cap B) \geq OPT - f(A) - \epsilon OPT.$$

Now consider the multilinear problem (4). Note that the indicator vector $\mathbf{1}_{O'}$ is feasible in P', and hence $(1-\epsilon)\mathbf{1}_{O'}$ is feasible in $(1-\epsilon)P'$. Using the concavity of $G(x)$ along the line from the origin to $\mathbf{1}_{O'}$, we have $G((1-\epsilon)\mathbf{1}_{O'}) \geq (1-\epsilon)g(O') \geq (1-2\epsilon)OPT - f(A)$. Using the continuous greedy algorithm [24], we find a fractional solution x^* of value

$$G(x^*) \geq (1 - 1/e)G((1-\epsilon)\mathbf{1}_{O'}) \geq (1 - 1/e - 2\epsilon)OPT - f(A).$$

Finally, we apply randomized pipage rounding to x^* and call the resulting set R. By the construction of pipage rounding, R is independent in $\mathcal{M}'$ with probability 1. However, R might violate some of the knapsack constraints.

Consider a fixed knapsack constraint, $\sum_{i \in S} c_{ij} \leq C_j$. Our fractional solution x^* satisfies $\sum c_{ij} x_i^* \leq (1-\epsilon)C_j$. Also, we know that all sizes in the reduced instance are bounded by $c_{ij} \leq k\epsilon^3 C_j$. By scaling, $c'_{ij} = c_{ij}/(k\epsilon^3 C_j)$, we can apply Corollary 1.3 with $\mu = (1-\epsilon)/(k\epsilon^3)$:

$$\Pr[\sum_{i \in R} c_{ij} > C_j] \leq \Pr[\sum_{i \in R} c'_{ij} > (1+\epsilon)\mu] \leq e^{-\mu\epsilon^2/3} < e^{-1/4k\epsilon}.$$

By the union bound,

$$\Pr[\exists j; \sum_{i \in R} c_{ij} > C_j] < ke^{-1/4k\epsilon}.$$

Thus with constant probability, arbitrarily close to 1 for $\epsilon \to 0$, all knapsack constraints are satisfied by R. We also know from Lemma 1.1 that $\mathbf{E}[g(R)] \geq G(x^*) \geq (1-1/e-2\epsilon)OPT - f(A)$. This implies $\mathbf{E}[f(A \cup R)] \geq f(A) + \mathbf{E}[g(R)] \geq (1 - 1/e - 2\epsilon)OPT$. However, we are not done yet, because the value of $f(A \cup R)$ is correlated with the event that $A \cup R$ is feasible, and hence the expectation of $f(A \cup R)$ *conditioned on being feasible* might be too small. Here, we use a trick from the paper by Kulik et al. [11], which relates the value of $g(R)$ to the amount of overflow on the knapsack constraints, and shows that the contribution of infeasible sets cannot be too large.

Let us denote by $\mathcal{F}_1$ the event that $\mathbf{1}_R \in P'$, and by $\mathcal{F}_\ell$ (for $\ell \geq 2$) the event that $\mathbf{1}_R \in \ell P' \setminus (\ell-1)P'$, i.e. the rounded solution R is feasible for $\ell P'$ but not for $(\ell-1)P'$. Obviously, exactly one of the events $\mathcal{F}_\ell$ occurs. By the law of conditional probabilities,

$$\mathbf{E}[g(R)] = \sum_{\ell=1}^{\infty} \mathbf{E}[g(R) \mid \mathcal{F}_\ell] \Pr[\mathcal{F}_\ell]. \tag{5}$$

We already estimated that $\Pr[\mathcal{F}_1] \geq 1 - ke^{-1/4k\epsilon}$ and $\Pr[\mathcal{F}_2] \leq ke^{-1/4k\epsilon}$. Let us estimate the probabilities for $\ell \geq 3$. Corollary 1.3 for $\delta = \ell - 2$, $\mu = (1-\epsilon)/(k\epsilon^3)$ gives

$$\Pr[\sum_{i \in R} c_{ij} > (\ell-1)C_j] \leq \Pr[\sum_{i \in R} c'_{ij} > (1+\delta)\mu] \leq \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{\mu} = \left(\frac{e^{\ell-2}}{(\ell-1)^{\ell-1}}\right)^{\mu}.$$

This probability decays very rapidly with ℓ. It can be verified that for all $\ell \geq 3$, it is upper-bounded by $e^{-\ell\mu/9}$. Using the union bound and plugging in $\mu = (1-\epsilon)/(k\epsilon^3) \geq 0.9/(k\epsilon^3)$, we can write for any $\ell \geq 3$

$$\Pr[\mathcal{F}_\ell] \leq \Pr[\exists j; \sum_{i \in R} c_{ij} > (\ell-1)C_j] \leq ke^{-\ell\mu/9} \leq ke^{-\ell/10k\epsilon^3}.$$

Recall that $\mathbf{E}[g(R)] \geq G(x^*)$. By the concavity of $G(x)$ along rays through the origin, we have $\max\{G(x) : x \in \ell P'\} \leq \frac{\ell}{1-\epsilon} \max\{G(x) : x \in (1-\epsilon)P'\} \leq \frac{\ell}{(1-\epsilon)(1-1/e)} G(x^*) \leq 2\ell G(x^*)$. I.e., $\mathbf{E}[g(R) \mid \mathcal{F}_\ell] \leq 2\ell G(x^*)$. Plugging our bounds into (5), we get

$$\begin{aligned}
G(x^*) &\leq \mathbf{E}[g(R) \mid \mathcal{F}_1] \Pr[\mathcal{F}_1] + 4G(x^*) \Pr[\mathcal{F}_2] + \sum_{\ell=3}^{\infty} 2\ell G(x^*) \Pr[\mathcal{F}_\ell] \\
&\leq \mathbf{E}[g(R) \mid \mathcal{F}_1] \Pr[\mathcal{F}_1] + 4G(x^*) \cdot ke^{-1/4k\epsilon} + \sum_{\ell=3}^{\infty} 2\ell G(x^*) \cdot ke^{-\ell/10k\epsilon^3} \\
&\leq \mathbf{E}[g(R) \mid \mathcal{F}_1] \Pr[\mathcal{F}_1] + 6G(x^*) \cdot ke^{-1/4k\epsilon}
\end{aligned}$$

using the formula $\sum_{\ell=3}^{\infty} \ell q^\ell = \frac{3q^3}{1-q} + \frac{q^4}{(1-q)^2}$ with $q = e^{-1/10k\epsilon^3}$ (a very small positive number), and replacing this formula by the much larger expression $q^{2.5\epsilon^2} = e^{-1/4k\epsilon}$. For $\epsilon < \min\{1/k^2, 0.001\}$, it holds that $6ke^{-1/4k\epsilon} \leq 6\epsilon^{-1/2} e^{-\epsilon^{-1/2}/4} \leq \epsilon$ and hence

$$\mathbf{E}[g(R) \mid \mathcal{F}_1] \Pr[\mathcal{F}_1] \geq (1-\epsilon)G(x^*) \geq (1-\epsilon)((1-1/e-2\epsilon)OPT - f(A)) \geq (1-1/e-3\epsilon)OPT - f(A).$$

Recall that $A \cup R$ is feasible for the original problem iff $\mathbf{1}_R \in P'$ which is exactly the event $\mathcal{F}_1$. We can conclude that

$$\mathbf{E}[f(A \cup R) \mid \mathcal{F}_1] \Pr[\mathcal{F}_1] = f(A) + \mathbf{E}[g(R) \mid \mathcal{F}_1] \Pr[\mathcal{F}_1] \geq (1 - 1/e - 3\epsilon)OPT.$$

This means that even conditioned on $A \cup R$ being feasible, the expected value of $f(A \cup R)$ is at least $(1 - 1/e - 3\epsilon)OPT$. $\square$

4.2 Loose packing constraints

In this section we consider the case when the number of linear packing constraints is not a fixed constant. The notation we use in this case is that of a packing integer program:

$$\max\{f(x) : x \in P(\mathcal{M}), Ax \leq b, x \in \{0,1\}^n\}.$$

Here $f : 2^N \to \mathbb{R}$ is a monotone submodular function with $n = |N|$, $\mathcal{M} = (N, \mathcal{I})$ is a matroid, $A \in \mathbb{R}_+^{k \times n}$ is a non-negative matrix and $b \in \mathbb{R}_+^k$ is a non-negative vector. This problem has been studied extensively when $f(x)$ is a linear function, in other words $f(x) = w^T x$ for some non-negative weight vector $w \in \mathbb{R}^n$. Even this case with A, b having only $0, 1$ entries captures the maximum independent set problem in graphs and hence is NP-hard to approximate to within an $n^{1-\epsilon}$-factor for any fixed $\epsilon > 0$. For this reason a variety of restrictions on A, b have been studied.

We consider the case when the constraints are sufficiently loose, i.e. the right-hand side b is significantly larger than entries in A: in particular, we assume $b_i \geq c \log k \cdot \max_j A_{i,j}$ for $1 \leq i \leq k$. In this case, we propose a straightforward algorithm which works as follows.

Algorithm.

- Let $\epsilon = \sqrt{6/c}$. Solve (approximately) the following optimization problem:

  $$\max\{F(x) : x \in (1-\epsilon)P\}$$

  where $F(x) = \mathbf{E}[f(\hat{x})]$ is the multilinear extension of $f(S)$, and

  $$P = \{x \in P(\mathcal{M}) \mid \forall i; \sum A_{ij} x_j \leq b_i\}.$$

  Since linear functions can be optimized over P in polynomial time, we can use the continuous greedy algorithm [24] to find a fractional solution x^* within a factor of $1 - 1/e$ of optimal.

- Apply randomized pipage rounding to x^* with respect to the matroid polytope $P(\mathcal{M})$. If the resulting solution R satisfies the packing constraints, return R; otherwise, fail.

Theorem 4.2. *Assume that $A \in \mathbb{R}^{k \times n}$ and $b \in \mathbb{R}^k$ such that $b_i \geq A_{ij} c \log k$ for all i, j and some constant $c = 6/\epsilon^2$. Then the algorithm above gives a $(1 - 1/e - O(\epsilon))$-approximation with high probability.*

We remark that it is NP-hard to achieve a better than $(1-1/e)$-approximation even when $k = 1$ and the constraint is very loose ($A_{ij} = 1$ and $b_i \to \infty$) [6].

Proof. The proof is similar to that of Theorem 4.1, but simpler. We only highlight the main differences.

In the first stage we obtain a fractional solution such that $F(x^*) \geq (1-\epsilon)(1-1/e)OPT$. Randomized pipage rounding yields a random solution R which satisfies the matroid constraint. It remains to check the packing constraints. For each i, we have

$$\mathbf{E}[\sum_{j \in R} A_{ij}] = \sum_{j \in R} A_{ij} x_j^* \leq (1-\epsilon) b_i.$$

The variables X_j are negatively correlated and by Corollary 1.3 with $\delta = \epsilon = \sqrt{6/c}$ and $\mu = c \log k$,

$$\Pr[\sum_{j \in R} A_{ij} > b_i] < e^{-\delta^2 \mu/3} = \frac{1}{k^2}.$$

By the union bound, all packing constraints are satisfied with probability at least $1 - 1/k$. We assume here that $k = \omega(1)$. By employing a trick similar to Theorem 4.1, we can also conclude that the expected value of the solution conditioned on being feasible is at least $(1 - 1/e - O(\epsilon))OPT$. $\square$

5 Minimax integer programs with a matroid constraint

Minimax integer programs are motivated by applications to routing and partitioning. The setup is as follows; we follow [22]. We have boolean variables $x_{i,j}$ for $i \in [n]$ and $j \in [\ell_i]$ for integers $\ell_1, \ldots, \ell_n$. Let $N = \sum_{i \in [n]} \ell_i$ The goal is to minimize λ subject to:

- equality constraints: $\forall i \in [n], \sum_{j \in [\ell_i]} x_{i,j} = 1$
- a system of linear inequalities $Ax \leq \lambda \mathbf{1}$ where $A \in [0,1]^{m \times N}$
- integrality constraints: $x_{i,j} \in \{0,1\}$ for all i, j.

The variables $x_{i,j}$, $j \in [\ell_i]$ for each $i \in [n]$ capture the fact that exactly one option amongst the ℓ_i options in group i should be chosen. A canonical example is the congestion minimization problem for integral routings in graphs where for each i, the $x_{i,j}$ variables represent the different paths for routing the flow of a pair (s_i, t_i) and the matrix A encodes the capacity constraints of the edges. A natural approach is to solve the natural LP relaxation for the above problem and then apply randomized rounding by choosing independently for each i exactly one $j \in [\ell_i]$ where the probability of choosing $j \in [\ell_i]$ is exactly equal to $x_{i,j}$. This follows the randomized rounding method of Raghavan and Thompson for congestion minimization [19] and one obtains an $O(\log m / \log \log m)$ approximation with respect to the fractional solution. Using Lovász Local Lemma (and complicated derandomization) it is possible to obtain an improved bound of $O(\log d / \log \log d)$ [14, 22] where d is the maximum number of non-zero entries in any column of A. This refined bound has various applications.

Interestingly, the above problem becomes non-trivial if we make a slight change to the equality constraints. Suppose for each $i \in [n]$ we now have an equality constraint of the form $\sum_{j \in [\ell_i]} x_{i,j} = k_i$ where k_i is an integer. For routing, this corresponds to a requirement of k_i paths for pair (s_i, t_i). We call this the *low congestion multi-path routing problem.* Now the standard randomized rounding doesn't quite work. Srinivasan [23], motivated by this generalized routing problem, developed dependent randomized rouding and used the negative correlation properties of this rounding to obtain an $O(\log m / \log \log m)$ approximation. This was further generalized in [8] as randomized versions of pipage rounding in the context of other applications.

5.1 Congestion minimization under a matroid base constraint

Here we show that randomized pipage rounding in matroids allows a clean generalization of the type of constraints considered in several applications in [23, 8]. Let $\mathcal{M}$ be a matroid on a ground set N. Let $B(\mathcal{M})$ be the base polytope of $\mathcal{M}$. We consider the problem

$$\min \left\{ \lambda : \exists x \in \{0,1\}^N, x \in B(\mathcal{M}), Ax \leq \lambda \mathbf{1} \right\}$$

where $A \in [0,1]^{m \times N}$. We observe that the previous problem with the variables partitioned into groups and equality constraints can be cast naturally as a special case of this matroid constraint problem; the equality constraints simply correspond to a partition matroid on the ground set of all variables $x_{i,j}$.

However, our framework is much more flexible. For example, consider the spanning tree problem with packing constraints: each edge has a weight w_e and we want to minimize the maximum load on any vertex, $\max_{v \in V} \sum_{e \in \delta(v)} w_e$. This problem also falls within our framework.

Theorem 5.1. *There is an $O(\log m/\log\log m)$-approximation for the problem*

$$\min\left\{\lambda : \exists x \in \{0,1\}^N, x \in B(\mathcal{M}), Ax \leq \lambda \mathbf{1}\right\},$$

where m is the number of packing constraints, i.e. $A \in [0,1]^{m\times N}$.

Proof. Fix a value of λ. Let $Z(\lambda) = \{j \mid \exists i; A_{ij} > \lambda\}$. We can force $x_j = 0$ for all $j \in Z(\lambda)$, because no element $j \in Z(\lambda)$ can be in a feasible solution for λ. In polynomial time, we can check the feasibility of the following LP:

$$P_\lambda = \left\{x \in B(\mathcal{M}) : Ax \leq \lambda\mathbf{1}, x|_{Z(\lambda)} = 0\right\}$$

(because we can separate over $B(\mathcal{M})$ and the additional packing constraints efficiently). By binary search, we can find (within $1+\epsilon$) the minimum value of λ such that $P_\lambda \neq \emptyset$. This is a lower bound on the actual optimum λ_{OPT}. We also obtain the corresponding fractional solution x^*.

We apply randomized pipage rounding to x^*, obtaining a random set R. R satisfies the matroid base constraint by definition. Consider a fixed packing constraint (the i-th row of A). We have

$$\sum A_{ij}x_j^* \leq \lambda$$

and all entries A_{ij} such that $x_j^* > 0$ are bounded by λ. We set $\tilde{A}_{ij} = A_{ij}/\lambda$, so that we can use Corollary 1.3. We get

$$\Pr[\sum_{j\in R} A_{ij} > (1+\delta)\lambda] = \Pr[\sum_{j\in R}\tilde{A}_{ij} > 1+\delta] < \left(\frac{e^\delta}{(1+\delta)^{1+\delta}}\right)^\mu.$$

For $\mu = 1$ and $1+\delta = \frac{4\log m}{\log\log m}$, this probability is bounded by

$$\Pr[\sum_{j\in R} A_{ij} > (1+\delta)\lambda] \leq \left(\frac{e\log\log m}{4\log m}\right)^{\frac{4\log m}{\log\log m}} < \left(\frac{1}{\sqrt{\log m}}\right)^{\frac{4\log m}{\log\log m}} = \frac{1}{m^2}$$

for sufficiently large m. Therefore, all m constraints are satisfied within a factor of $1+\delta = \frac{4\log m}{\log\log m}$ with high probability. □

We remark that the approximation guarantee can be made an "almost additive" $O(\log m)$, in the following sense: Assuming that the optimum value is λ^*, for any fixed $\epsilon > 0$ we can find a solution of value $\lambda \leq (1+\epsilon)\lambda^* + O(\frac{1}{\epsilon}\log m)$. Scaling is important here: recall that we assumed $A \in [0,1]^{N\times m}$. We omit the proof, which follows by a similar application of the Chernoff bound as above, with $\mu = \lambda^*$ and $\delta = \epsilon + O(\frac{1}{\epsilon\lambda^*}\log m)$.

5.2 Min-cost matroid bases with packing constraints

We can similarly handle the case where in addition we want to minimize a linear objective function. An example of such a problem would be a multi-path routing problem minimizing the total cost in addition to congestion. Another example is the minimum-cost spanning tree with packing constraints for the edges incident with each vertex. We remark that in case the packing constraints are simply degree bounds, strong results are known - namely, there is an algorithm that finds a spanning tree of optimal cost and violating the degree bounds by at most one [21]. In the general case of finding a matroid base satisfying certain "degree constraints", there is an algorithm [10] that finds a base of optimal cost and violating the degree constraints by an additive error of at most $\Delta - 1$, where each element participates in at most Δ constraints (e.g. $\Delta = 2$ for degree-bounded spanning trees). The algorithm of [10] also works for upper and lower bounds, violating each constraint by at most $2\Delta - 1$. See [10] for more details.

We consider a variant of this problem where the packing constraints can involve arbitrary weights and capacities. We show that we can find a matroid base of near-optimal cost which violates the packing constraints by a multiplicative factor of $O(\log m / \log \log m)$, where m is the total number of packing constraints.

Theorem 5.2. *There is a $(1+\epsilon, O(\log m/\log\log m))$-bicriteria approximation for the problem*

$$\min \left\{ c^T x : x \in \{0,1\}^N, x \in B(\mathcal{M}), Ax \leq b \right\},$$

where $A \in [0,1]^{m \times N}$ and $b \in \mathbb{R}^N$; the first guarantee is w.r.t. the cost of the solution and the second guarantee w.r.t. to overflow on the packing constraints.

Proof. We give a sketch of the proof. First, we throw away all elements that on their own violate some packing constraint. Then, we solve the following LP:

$$\min \left\{ c^T x : x \in B(\mathcal{M}), Ax \leq b \right\}.$$

Let the optimum solution be x^*. We apply pipage rounding to x^*, yielding a random solution R. Since each of the m constraints is satisfied in expectation, and each element alone satisfies each packing constraint, we get by the same analysis as above that with high probability, R violates every constraint by a factor of $O(\log m / \log \log m)$.

Finally, the expected cost of our solution is $c^T x^* \leq OPT$. By Markov's inequality, the probability that $c(R) > (1+\epsilon)OPT$ is at most $1/(1+\epsilon) \leq 1 - \epsilon/2$. With probability at least $\epsilon/2 - o(1)$, $c(R) \leq (1+\epsilon)OPT$ and all packing constraints are satisfied within $O(\log m / \log \log m)$. □

Let us rephrase this result in the more familiar setting of spanning trees. Given packing constraints on the edges incident with each vertex, using arbitrary weights and capacities, we can find a spanning tree of near-optimal cost, violating each packing constraint by a multiplicative factor of $O(\log m / \log \log m)$. As in the previous section, if we assume that the weights are in $[0,1]$, this can be replaced by an additive factor of $O(\frac{1}{\epsilon} \log m)$ while making the multiplicative factor $1 + \epsilon$ (see the end of Section 5.1).

In the general case of matroid bases, our result is incomparable to that of [10], which provides an additive guarantee of $\Delta - 1$. (The assumption here is that each element participates in at most Δ degree constraints; in our framework, this corresponds to $A \in \{0,1\}^{m \times N}$ with Δ-sparse columns.) When elements participate in many degree constraints ($\Delta \gg \log m$) and the degree bounds are $b_i = O(\log m)$, our result is actually stronger in terms of the packing constraint guarantee.

Acknowledgments: We are grateful to Anupam Gupta for asking about the approximability of maximizing a monotone submodular set function subject to a matroid constraint and a constant number of knapsack constraints that led to this work.

References

[1] A. Ageev and M. Sviridenko. Pipage rounding: a new method of constructing algorithms with proven performance guarantee. *J. of Combinatorial Optimization*, 8:307–328, 2004.

[2] G. Calinescu, C. Chekuri, M. Pál and J. Vondrák. Maximizing a submodular set function subject to a matroid constraint. *Proc. of* 12^{th} *IPCO*, 182–196, 2007.

[3] G. Calinescu, C. Chekuri, M. Pál and J. Vondrák. Maximizing a submodular set function subject to a matroid constraint. To appear in *SIAM Journal on Computing, special issue for STOC 2008*.

[4] C. Chekuri, A. Ene and N. Korula. UFP in paths and trees and column-restricted packing integer programs. *Proc. of APPROX*, 2009.

[5] J. Edmonds. Matroids, submodular functions and certain polyhedra. Combinatorial Structures and Their Applications, 69–87, 1970.

[6] U. Feige. A threshold of $\ln n$ for approximating set cover. *Journal of the ACM*, 45(4):634–652, 1998.

[7] M. L. Fisher, G. L. Nemhauser and L. A. Wolsey. An analysis of approximations for maximizing submodular set functions - II. *Math. Prog. Study*, 8:73–87, 1978.

[8] R. Gandhi, S. Khuller, S. Parthasarathy and A. Srinivasan. Dependent rounding and its applications to approximation algorithms. *Journal of the ACM* 53:324-360, 2006.

[9] A. Gupta, V. Nagarajan and R. Ravi. Personal communication, 2009.

[10] T. Király, L. C. Lau, and M. Singh. Degree bounded matroids and submodular flows. *In* 13^{th} *IPCO* (2008).

[11] A. Kulik, H. Shachnai and T. Tamir. Maximizing submodular set functions subject to multiple linear constraints. *Proc. of* 20^{th} *ACM-SIAM SODA* (2009), 545–554.

[12] J. Lee, V. Mirrokni, V. Nagarajan and M. Sviridenko. Maximizing non-monotone submodular functions under matroid and knapsack constraints. *Proc. of* 41^{st} *ACM STOC 2009*, 323–332.

[13] J. Lee, M. Sviridenko, and J. Vondrák. Submodular maximization over multiple matroids via generalized exchange properties. *Proc. of APPROX*, Springer LNCS, 244–257, 2009.

[14] T. Leighton, C.-J.Lu, S. Rao, and A. Srinivasan. New algorithmic aspects of the local lemma with applications to routing and partitioning. *SIAM J. on Computing*, Vol. 31, 626–641, 2001.

[15] G. L. Nemhauser, L. A. Wolsey and M. L. Fisher. An analysis of approximations for maximizing submodular set functions - I. *Math. Prog.*, 14:265–294, 1978.

[16] G. L. Nemhauser and L. A. Wolsey. Best algorithms for approximating the maximum of a submodular set function. *Math. Oper. Research*, 3(3):177–188, 1978.

[17] A. Panconesi and A. Srinivasan. Randomized distributed edge coloring via an extension of the Chernoff-Hoeffding bounds. *SIAM Journal on Computing* 26:350-368, 1997.

[18] D. Pritchard. Approximability of sparse integer programs. *Proc. of ESA*, 2009.

[19] P. Raghavan and C. D. Thompson. Randomized rounding: a technique for provably good algorithms and algorithmic proofs. *Combinatorica* 7(4):365–374, 1987.

[20] A. Schrijver. Combinatorial optimization - polyhedra and efficiency. Springer, 2003.

[21] M. Singh and L.C. Lau. Approximating minimum bounded degree spanning tress to within one of optimal, *Proc. of* 39^{th} *ACM STOC* (2007).

[22] A. Srinivasan. An extension of the Lovász Local Lemma, and its applications to integer programming. *SIAM J. on Computing*, Vol 36, 609–634, 2006. Preliminary version in *Proc. of ACM-SIAM SODA*, 1996.

[23] A. Srinivasan. Distributions on level-sets with applications to approximation algorithms, *Proc. IEEE Symposium on Foundations of Computer Science (FOCS)*, 588–597, 2001.

[24] J. Vondrák. Optimal approximation for the submodular welfare problem in the value oracle model. *Proc. of ACM STOC*, 67–74, 2008.

[25] J. Vondrák. Symmetry and approximability of submodular maximization problems. To appear in *Proc. of IEEE FOCS*, 2009.

A Randomized pipage rounding

Let us summarize the pipage rounding technique in the context of matroid polytopes [2, 3]. The basic version of the technique assumes that we start with a point in the matroid base polytope, and we want to round it to a vertex of $B(\mathcal{M})$. In each step, we have a fractional solution $y \in B(\mathcal{M})$ and a *tight set* T (satisfying $y(T) = r(T)$) containing at least two fractional variables. We modify the two fractional variables in a such a way that their sum remains constant, until some variable becomes integral or a new constraint becomes tight. If a new constraint becomes tight, we continue with a new tight set, which can be shown to be a proper subset of the previous tight set [2, 3]. Hence, after n steps we produce a new integral variable, and the process terminates after n^2 steps.

In the randomized version of the technique, each step is randomized in such a way that the expectation of each variable is preserved. Here is the randomized version of pipage rounding [3]. The subroutine **HitConstraint**(y, i, j) starts from y and tries to increase y_i and decrease y_j at the same rate, as long as the the solution is inside $B(\mathcal{M})$. It returns a new point y and a tight set A, which would be violated if we go any further. This is used in the main algorithm **PipageRound**$(\mathcal{M}, y)$, which repeats the process until an integral solution in $B(\mathcal{M})$ is found.

Subroutine **HitConstraint**(y, i, j):
  Denote $\mathcal{A} = \{A \subseteq X : i \in A, j \notin A\}$;
  Find $\delta = \min_{A \in \mathcal{A}}(r_{\mathcal{M}}(A) - y(A))$
    and an optimal $A \in \mathcal{A}$;
  If $y_j < \delta$ then $\{\delta \leftarrow y_j,\ A \leftarrow \{j\}\}$;
  $y_i \leftarrow y_i + \delta,\ y_j \leftarrow y_j - \delta$;
  Return (y, A).

Algorithm **PipageRound**$((\mathcal{M}, y))$:
While (y is not integral) do
  $T \leftarrow X$;
  While (T contains fractional variables) do
    Pick $i, j \in T$ fractional;
    $(y^+, A^+) \leftarrow$ **HitConstraint**(y, i, j);
    $(y^-, A^-) \leftarrow$ **HitConstraint**(y, j, i);
    $p \leftarrow ||y^+ - y|| / ||y^+ - y^-||$;
    With probability p, $\{y \leftarrow y^-, T \leftarrow T \cap A^-\}$;
      Else $\{y \leftarrow y^+, T \leftarrow T \cap A^+\}$;
  EndWhile
EndWhile
Output y.

Subsequently [25], pipage rounding was extended to the case when the starting point is in the matroid polytope $P(\mathcal{M})$, rather than $B(\mathcal{M})$. This is not an issue in [3], but it is necessary for applications with non-monotone submodular functions [25] or with additional constraints, such as in this paper.

The following procedure takes care of the case when we start with a fractional solution $x \in P(\mathcal{M})$. It adjusts the solution in a randomized way so that the expectation of each variable is preserved, and the new fractional solution is in the base polytope of a (possibly reduced) matroid.

Algorithm **Adjust**$((\mathcal{M}, x))$:
While (x is not in $B(\mathcal{M})$) do
  If (there is i and $\delta > 0$ such that $x + \delta \mathbf{e}_i \in P(\mathcal{M})$) do
    Let $x_{max} = x_i + \max\{\delta : x + \delta \mathbf{e}_i \in P(\mathcal{M})\}$;
    Let $p = x_i / x_{max}$;
    With probability p, $\{x_i \leftarrow x_{max}\}$;
      Else $\{x_i \leftarrow 0\}$;
  EndIf
  If (there is i such that $x_i = 0$) do
    Delete i from $\mathcal{M}$ and remove the i-coordinate from x.
EndWhile
Output $(\mathcal{M}, x)$.

To summarize, the complete procedure works as follows. For a given $x \in P(\mathcal{M})$, we run $(\mathcal{M}', y) :=$**Adjust**$(\mathcal{M}, x)$, followed by **PipageRound**$((\mathcal{M}', y))$. The outcome is a base in the restricted matroid where some elements have been deleted, i.e. an independent set in the original matroid.

B Chernoff bounds for submodular functions

Here we prove Theorem 1.4, in two parts.

Lemma B.1. *Let $f : \{0,1\}^n \rightarrow \mathbb{R}_+$ be a monotone submodular function, with marginal values always between $[0,1]$. Let $X_1, \ldots, X_n$ be independent random variables in $\{0,1\}$. Let $\mu =$*

$\mathbf{E}[f(X_1, X_2, \ldots, X_n)]$. *Then for any* $\delta > 0$,

$$\Pr[f(X_1, \ldots, X_n) \geq (1+\delta)\mu] \leq \left(\frac{e^\delta}{(1+\delta)^{1+\delta}}\right)^\mu.$$

Proof. Assume WLOG that $f(0, 0, \ldots, 0) = 0$. We decompose the value of $f(X_1, \ldots, X_n)$ into a sum of random variables,

$$f(X_1, \ldots, X_n) = \sum_{k=1}^{n} (f(X_1, \ldots, X_k, 0, \ldots, 0) - f(X_1, \ldots, X_{k-1}, 0, \ldots, 0)) = \sum_{i=1}^{n} Y_i,$$

where $Y_i = f(X_1, \ldots, X_k, 0, \ldots, 0) - f(X_1, \ldots, X_{k-1}, 0, \ldots, 0)$. We would like to mimic Chernoff's proof for the variables $Y_1, \ldots, Y_n$. Note that $Y_1, \ldots, Y_n$ are not independent. There could be negative and even positive correlations between Y_i, Y_j. What is important for us, however, is that we can show the correlation between $\sum_{i=1}^{k-1} Y_i$ and Y_k can be only negative.

As in Chernoff's proof, we fix $\lambda > 0$ and analyze the quantity $\mathbf{E}[e^{\lambda \sum_{i=1}^n Y_i}]$. Denote $p_i = \Pr[X_i = 1]$. For any k, we have

$$\begin{aligned}
& \mathbf{E}[e^{\lambda \sum_{i=1}^k Y_i}] = \mathbf{E}[e^{\lambda f(X_1, \ldots, X_k, 0, \ldots, 0)}] \\
= \;& p_k \; \mathbf{E}[e^{\lambda f(X_1, \ldots, X_{k-1}, 1, \ldots, 0)}] + (1-p_k)\mathbf{E}[e^{\lambda f(X_1, \ldots, X_{k-1}, 0, \ldots, 0)}] \\
= \;& p_k \; \mathbf{E}[e^{\lambda f(X_1, \ldots, X_{k-1}, 0, \ldots, 0)} e^{\lambda f_{(X_1, \ldots, X_{k-1}, 0, \ldots, 0)}(k)}] + (1-p_k)\mathbf{E}[e^{\lambda f(X_1, \ldots, X_{k-1}, 0, \ldots, 0)}]
\end{aligned}$$

where $f_{(X_1, \ldots, X_{k-1}, 0, \ldots, 0)}(k) = f(X_1, \ldots, X_{k-1}, 1, \ldots, 0) - f(X_1, \ldots, X_{k-1}, 0, \ldots, 0)$ denotes the marginal value of X_k being set to 1, given the preceding variables. This can be also seen as Y_k, conditioned on $X_k = 1$. By submodularity, this is a decreasing function of $X_1, \ldots, X_{k-1}$. On the other hand, $f(X_1, \ldots, X_{k-1}, 0, \ldots, 0) = \sum_{i=1}^{k-1} Y_i$ is an increasing function of $X_1, \ldots, X_{k-1}$. We get the same monotonicity properties for the exponential functions $e^{\lambda f(\ldots)}$. By the FKG inequality, $e^{\lambda f(X_1, \ldots, X_{k-1}, 0, \ldots, 0)}$ and $e^{\lambda f_{(X_1, \ldots, X_{k-1}, 0, \ldots, 0)}(k)}$ are negatively correlated, and we get

$$\begin{aligned}
& \mathbf{E}[e^{\lambda f(X_1, \ldots, X_{k-1}, 0, \ldots, 0)} e^{\lambda f_{(X_1, \ldots, X_{k-1}, 0, \ldots, 0)}(k)}] \\
\leq \;& \mathbf{E}[e^{\lambda f(X_1, \ldots, X_{k-1}, 0, \ldots, 0)}]\mathbf{E}[e^{\lambda f_{(X_1, \ldots, X_{k-1}, 0, \ldots, 0)}(k)}] \\
= \;& \mathbf{E}[e^{\lambda \sum_{i=1}^{k-1} Y_i}]\mathbf{E}[e^{\lambda Y_k} \mid X_k = 1].
\end{aligned}$$

Hence, we have

$$\begin{aligned}
\mathbf{E}[e^{\lambda \sum_{i=1}^k Y_i}] \;&\leq\; p_k \; \mathbf{E}[e^{\lambda \sum_{i=1}^{k-1} Y_i}]\mathbf{E}[e^{\lambda Y_k} \mid X_k = 1] + (1-p_k)\mathbf{E}[e^{\lambda \sum_{i=1}^{k-1} Y_i}] \\
&=\; \mathbf{E}[e^{\lambda \sum_{i=1}^{k-1} Y_i}] \cdot (p_k \; \mathbf{E}[e^{\lambda Y_k} \mid X_k = 1] + (1-p_k) \cdot 1) \\
&=\; \mathbf{E}[e^{\lambda \sum_{i=1}^{k-1} Y_i}] \cdot (p_k \; \mathbf{E}[e^{\lambda Y_k} \mid X_k = 1] + (1-p_k) \; \mathbf{E}[e^{\lambda Y_k} \mid X_k = 0]) \\
&=\; \mathbf{E}[e^{\lambda \sum_{i=1}^{k-1} Y_i}] \cdot \mathbf{E}[e^{\lambda Y_k}].
\end{aligned}$$

By induction, we obtain

$$\mathbf{E}[e^{\lambda \sum_{i=1}^n Y_i}] \leq \prod_{i=1}^{n} \mathbf{E}[e^{\lambda Y_i}].$$

Henceforth, the proof proceeds exactly like Chernoff's proof. Applying Markov's bound to the random variable $e^{\lambda \sum_{i=1}^n Y_i}$, we get

$$\Pr[\sum_{i=1}^{n} Y_i \geq (1+\delta)\mu] = \Pr[e^{\lambda \sum_{i=1}^n Y_i} \geq e^{\lambda(1+\delta)\mu}] \leq \frac{\mathbf{E}[e^{\lambda \sum_{i=1}^n Y_i}]}{e^{\lambda(1+\delta)\mu}} \leq \frac{\prod_{i=1}^n \mathbf{E}[e^{\lambda Y_i}]}{e^{\lambda(1+\delta)\mu}}.$$

Here, $\mu = \mathbf{E}[f(X_1, \ldots, X_n)] = \sum_{i=1}^n \mathbf{E}[Y_i]$. Let us denote $\mathbf{E}[Y_i] = \omega_i$. By the convexity of the exponential and the fact that $Y_i \in [0,1]$,

$$\mathbf{E}[e^{\lambda Y_i}] \leq \omega_i e^{\lambda} + (1 - \omega_i) = 1 + (e^{\lambda} - 1)\omega_i \leq e^{(e^{\lambda}-1)\omega_i}.$$

Therefore, the bound becomes

$$\Pr[\sum_{i=1}^n Y_i \geq (1+\delta)\mu] \leq \frac{\prod_{i=1}^n e^{(e^{\lambda}-1)\omega_i}}{e^{\lambda(1+\delta)\mu}} = \frac{e^{(e^{\lambda}-1)\mu}}{e^{\lambda(1+\delta)\mu}}.$$

We choose $e^{\lambda} = 1 + \delta$ which yields

$$\Pr[\sum_{i=1}^n Y_i \geq (1+\delta)\mu] \leq \frac{e^{\delta\mu}}{(1+\delta)^{(1+\delta)\mu}}.$$

□

Lemma B.2. *Let $f : \{0,1\}^n \to \mathbb{R}_+$ be a monotone submodular function, with marginal values always between $[0,1]$. Let $X_1, \ldots, X_n$ be independent random variables in $\{0,1\}$. Let $\mu = \mathbf{E}[f(X_1, X_2, \ldots, X_n)]$. Then for any $\delta \in (0,1]$,*

$$\Pr[f(X_1, \ldots, X_n) \leq (1-\delta)\mu] \leq e^{-\mu\delta^2/2}.$$

Proof. The proof is very similar to the previous one. Assume that $f(0,0,\ldots,0) = 0$. We decompose the value of $f(X_1, \ldots, X_n)$ into a sum of random variables,

$$f(X_1, \ldots, X_n) = \sum_{k=1}^n (f(X_1, \ldots, X_k, 0, \ldots, 0) - f(X_1, \ldots, X_{k-1}, 0, \ldots, 0)) = \sum_{i=1}^n Y_i,$$

where $Y_i = f(X_1, \ldots, X_k, 0, \ldots, 0) - f(X_1, \ldots, X_{k-1}, 0, \ldots, 0)$. We fix $\lambda > 0$ and analyze the quantity $\mathbf{E}[e^{-\lambda \sum_{i=1}^n Y_i}]$. Denote $p_i = \Pr[X_i = 1]$. For any k, we have

$$\begin{aligned}
& \mathbf{E}[e^{-\lambda \sum_{i=1}^k Y_i}] = \mathbf{E}[e^{-\lambda f(X_1,\ldots,X_k,0,\ldots,0)}] \\
= \; & p_k \; \mathbf{E}[e^{-\lambda f(X_1,\ldots,X_{k-1},1,\ldots,0)}] + (1-p_k)\mathbf{E}[e^{-\lambda f(X_1,\ldots,X_{k-1},0,\ldots,0)}] \\
= \; & p_k \; \mathbf{E}[e^{-\lambda f(X_1,\ldots,X_{k-1},0,\ldots,0)} e^{-\lambda f_{(X_1,\ldots,X_{k-1},0,\ldots,0)}(k)}] + (1-p_k)\mathbf{E}[e^{-\lambda f(X_1,\ldots,X_{k-1},0,\ldots,0)}]
\end{aligned}$$

where $f_{(X_1,\ldots,X_{k-1},0,\ldots,0)}(k) = f(X_1, \ldots, X_{k-1}, 1, \ldots, 0) - f(X_1, \ldots, X_{k-1}, 0, \ldots, 0)$ denotes the marginal value of X_k being set to 1, given the preceding variables. By submodularity, this is a decreasing function of $X_1, \ldots, X_{k-1}$. On the other hand, $f(X_1, \ldots, X_{k-1}, 0, \ldots, 0)$ is an increasing function of $X_1, \ldots, X_{k-1}$. The monotonicity is inverted for the exponential functions $e^{-\lambda f(\ldots)}$; still, one quantity is increasing and the other one is decreasing. By the FKG inequality, the two random quantities are negatively correlated, and we get

$$\begin{aligned}
& \mathbf{E}[e^{-\lambda f(X_1,\ldots,X_{k-1},0,\ldots,0)} e^{-\lambda f_{(X_1,\ldots,X_{k-1},0,\ldots,0)}(k)}] \\
\leq \; & \mathbf{E}[e^{-\lambda f(X_1,\ldots,X_{k-1},0,\ldots,0)}]\mathbf{E}[e^{-\lambda f_{(X_1,\ldots,X_{k-1},0,\ldots,0)}(k)}] \\
= \; & \mathbf{E}[e^{-\lambda \sum_{i=1}^{k-1} Y_i}]\mathbf{E}[e^{-\lambda Y_k} \mid X_k = 1].
\end{aligned}$$

Hence, we have

$$\begin{aligned}
\mathbf{E}[e^{-\lambda\sum_{i=1}^{k} Y_i}] &\le p_k\ \mathbf{E}[e^{-\lambda\sum_{i=1}^{k-1} Y_i}]\mathbf{E}[e^{-\lambda Y_k} \mid X_k = 1] + (1-p_k)\mathbf{E}[e^{-\lambda\sum_{i=1}^{k-1} Y_i}] \\
&= \mathbf{E}[e^{-\lambda\sum_{i=1}^{k-1} Y_i}] \cdot (p_k\ \mathbf{E}[e^{-\lambda Y_k} \mid X_k = 1] + (1-p_k)\cdot 1) \\
&= \mathbf{E}[e^{-\lambda\sum_{i=1}^{k-1} Y_i}] \cdot (p_k\ \mathbf{E}[e^{-\lambda Y_k} \mid X_k = 1] + (1-p_k)\ \mathbf{E}[e^{-\lambda Y_k} \mid X_k = 0]) \\
&= \mathbf{E}[e^{-\lambda\sum_{i=1}^{k-1} Y_i}] \cdot \mathbf{E}[e^{-\lambda Y_k}].
\end{aligned}$$

By induction, we obtain

$$\mathbf{E}[e^{-\lambda\sum_{i=1}^{n} Y_i}] \le \prod_{i=1}^{n} \mathbf{E}[e^{-\lambda Y_i}].$$

Applying Markov's bound to the random variable $e^{-\lambda\sum_{i=1}^{n} Y_i}$, we get

$$\Pr[\sum_{i=1}^{n} Y_i \le (1-\delta)\mu] = \Pr[e^{-\lambda\sum_{i=1}^{n} Y_i} \ge e^{-\lambda(1-\delta)\mu}] \le \frac{\mathbf{E}[e^{-\lambda\sum_{i=1}^{n} Y_i}]}{e^{-\lambda(1-\delta)\mu}} \le \frac{\prod_{i=1}^{n} \mathbf{E}[e^{-\lambda Y_i}]}{e^{-\lambda(1-\delta)\mu}}.$$

Here, $\mu = \mathbf{E}[f(X_1,\ldots,X_n)] = \sum_{i=1}^{n} \mathbf{E}[Y_i]$. Let us denote $\mathbf{E}[Y_i] = \omega_i$. By the convexity of the exponential and the fact that $Y_i \in [0,1]$,

$$\mathbf{E}[e^{-\lambda Y_i}] \le \omega_i e^{-\lambda} + (1-\omega_i) = 1 + (e^{-\lambda}-1)\omega_i \le e^{(e^{-\lambda}-1)\omega_i}.$$

Therefore, the bound becomes

$$\Pr[\sum_{i=1}^{n} Y_i \le (1-\delta)\mu] \le \frac{\prod_{i=1}^{n} e^{(e^{-\lambda}-1)\omega_i}}{e^{-\lambda(1-\delta)\mu}} = \frac{e^{(e^{-\lambda}-1)\mu}}{e^{-\lambda(1-\delta)\mu}}.$$

Here, we choose $e^{-\lambda} = 1-\delta$ which yields

$$\Pr[\sum_{i=1}^{n} Y_i \le (1-\delta)\mu] \le \frac{e^{-\delta\mu}}{(1-\delta)^{(1-\delta)\mu}} \le e^{-\mu\delta^2/2}$$

where we used $(1-\delta)^{1-\delta} \ge e^{-\delta+\delta^2/2}$ for $\delta \in (0,1]$. $\square$