Multi-budgeted Matchings and Matroid Intersection via Dependent Rounding

Chandra Chekuri* Jan Vondrák† Rico Zenklusen‡

Abstract

Motivated by multi-budgeted optimization and other applications, we consider the problem of randomly rounding a fractional solution $\mathbf{x}$ in the (non-bipartite graph) matching and matroid intersection polytopes. We show that for any fixed $\delta > 0$, a given point $\mathbf{x}$ can be rounded to a random solution R such that $\mathbf{E}[\mathbf{1}_R] = (1-\delta)\mathbf{x}$ and *any* linear function of $\mathbf{x}$ satisfies dimension-free Chernoff-Hoeffding concentration bounds (the bounds depend on δ and the expectation μ). We build on and adapt the swap rounding scheme in our recent work [9] to achieve this result. Our main contribution is a non-trivial martingale based analysis framework to prove the desired concentration bounds. In this paper we describe two applications. We give a randomized PTAS for matroid intersection and matchings with any fixed number of budget constraints. We also give a deterministic PTAS for the case of matchings. The concentration bounds also yield related results when the number of budget constraints is not fixed. As a second application we obtain an algorithm to compute in polynomial time an ε-approximate Pareto-optimal set for the multi-objective variants of these problems, when the number of objectives is a fixed constant. We rely on a result of Papadimitriou and Yannakakis [26].

1 Introduction

Randomized rounding of a fractional solution into an integral solution is a powerful and ubiquitous technique in approximation algorithm design. Following the influential work of Raghavan and Thompson [27] for routing, packing and covering problems, several different randomized rounding methods have been developed over the years. *Dependent* randomized rounding methods have recently found many applications [1, 31, 14, 11, 19, 3, 2, 29, 6]. The term dependent refers to the property of the rounding scheme that ensures that the rounded solution satisfies some additional constraints in a deterministic fashion — this implies that the coordinates of the fractional solution cannot be independently rounded. An abstract view of such a scheme is the following: given a fractional point $\mathbf{x}$ in a polytope $P \subset \mathbb{R}^n$, randomly round $\mathbf{x}$ to a solution R corresponding to a vertex of P. Here P captures the deterministic constraints that we wish the rounding to satisfy, and it is natural to assume that P is an integer polytope (typically a $\{0,1\}$ polytope). Of course the important issue is what properties we need R to satisfy, and this is dictated by the application at hand. A property that is useful in several applications is that R satisfies *concentration* properties for linear functions of $\mathbf{x}$: that is, for any vector $a \in [0,1]^n$, we want the linear function $a(R) = \sum_{i \in R} a_i$ to be concentrated around its expectation[1]. Ideally, we would like to have $\mathbf{E}[\mathbf{1}_R] = \mathbf{x}$, which would mean $\mathbf{E}[a(R)] = \sum_i a_i x_i$. If this is not feasible, we would like $\mathbf{E}[\mathbf{1}_R]$ to be approximately equal to $\mathbf{x}$.

A natural question is the following. For which polytopes P can we implement such a rounding procedure? A strong property that implies concentration for linear functions is that of *negative correlation*[2], as shown by Panconesi and Srinivasan [25]. In our recent work [9] we described a rounding scheme called randomized swap rounding, to unify and generalize some existing results. In particular, we showed that if P is a matroid polytope (the convex hull of independent sets of the matroid), then one can round $\mathbf{x} \in P$ to a random independent set R such that $\mathbf{E}[\mathbf{1}_R] = \mathbf{x}$ and the coordinates of $\mathbf{1}_R$ satisfy negative correlation. We also showed negative correlation properties for certain restricted subsets of variables when P is the intersection of two matroids; this restricted setting nevertheless captures the important and useful special case of the assignment polytope where one obtains negative correlation for the edge variables incident to each vertex of the underlying bipartite graph [14]. There are, however, applications where one requires concentration properties for arbitrary linear functions of the variables. If one wants to use the property of negative correlation, then matroids are essentially the only structures where this can be done: We have recently shown that a $\{0,1\}$ polytope P has the property that any $\mathbf{x} \in P$ can be rounded to a vertex R of P such that $\mathbf{E}[\mathbf{1}_R] = \mathbf{x}$ and the coordinates of $\mathbf{1}_R$ are negatively correlated, if and only if P is an axis-parallel projection of a matroid base polytope [10].

*Department of Computer Science, University of Illinois, Urbana, IL 61801. Supported in part by NSF grants CCF-0728782 and CCF-1016684. E-mail: chekuri@cs.illinois.edu.

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

‡Dept. of Mathematics, MIT, Cambridge, MA 02139. E-mail: ricoz@math.mit.edu. Supported by the Swiss National Science Foundation, grant number: PBEZP2-129524.

[1] In this paper we focus on non-negative vectors $a \in [0,1]^n$ although there are applications for vectors $a \in [-1,1]^n$ (see [1]).

[2] We call $\{0,1\}$-random variables $X_1, \ldots, X_r$ negatively correlated, if $\mathbf{E}[\prod_{i \in T} X_i] \le \prod_{i \in T} \mathbf{E}[X_i]$ and $\mathbf{E}[\prod_{i \in T}(1 - X_i)] \le \prod_{i \in T}(1 - \mathbf{E}[X_i])$ for all subsets $T \subseteq [r]$.

Given that negative correlation is too strong a property to hold beyond matroid polytopes, it is natural to consider whether the property of concentration for linear functions can still be obtained via different means for other polytopes. An important special case to consider is the assignment polytope (equivalently the bipartite simple b-matching polytope). Arora, Frieze and Kaplan [1], motivated by applications to approximation schemes for dense-graph optimization problems, considered this problem in a certain restricted setting and guaranteed concentration bounds in an additive sense (with standard deviation on the order of $\sqrt{n} \cdot poly(\log n)$ and $poly(\log n)$, respectively, depending on whether the linear function under consideration has arbitrary coefficients or non-negative coefficients). We discuss more details of their result and its relationship to our work later in the paper. It is easy to show that even in the setting of the assignment polytope, one cannot obtain concentration bounds while preserving the expectation exactly.

In this paper we give a rounding scheme for two (classes of) polytopes, namely the polytope corresponding to the intersection of two matroids, and the non-bipartite graph matching polytope. The rounding procedure for non-bipartite graph matching polytopes can easily be applied to simple b-matchings by a standard reduction from b-matchings to matchings [30]. Note that the assignment polytope is a special case of both the intersection of two matroids and simple b-matchings. Our main result is the following.

THEOREM 1.1. *Let P be either a matroid intersection polytope or a (non-bipartite graph) matching polytope. For any fixed $0 < \gamma \leq \frac{1}{2}$, there is an efficient randomized rounding procedure, such that given a point $\mathbf{x} \in P$, it outputs a random feasible solution R corresponding to a (integer) vertex of P such that $\mathbf{E}[\mathbf{1}_R] = (1-\gamma)\mathbf{x}$. In addition, for any linear function $a(R) = \sum_{i \in R} a_i$ with $a_i \in [0,1]$, and for any $\varepsilon \in [0,1]$, we have:*

- *If $\mu \leq \mathbf{E}[a(R)]$,* $\Pr[a(R) \leq (1-\varepsilon)\mu] \leq e^{-\mu\gamma\varepsilon^2/20}$.
- *If $\mu \geq \mathbf{E}[a(R)]$,* $\Pr[a(R) \geq (1+\varepsilon)\mu] \leq e^{-\mu\gamma\varepsilon^2/20}$.

For any $t \geq 2$,

- *If $\mu \geq \mathbf{E}[a(R)]$,* $\Pr[a(R) \geq t\mu] \leq e^{-\mu\gamma(2t-3)/20}$.

We emphasize that the concentration bounds above depend only on μ, γ, ε and are independent of n, the *dimension* of the polytope. Moreover, there is no restriction on the linear functions $a(R)$ for which we prove concentration bounds, other than the normalization condition $a_i \in [0,1]$. Roughly speaking, $a(R)$ for the rounded solution is concentrated in a window of size $O(\sqrt{\mu})$ around $\mu = \mathbf{E}[a(R)]$, even if $\mu \ll n$.

The rounding scheme and concentration bounds in the above theorem also hold for non-bipartite graph b-matchings. This can be seen from a well-known pseudo-polynomial time reduction of b-matchings to regular matchings (see [30]), and the fact that our bounds are dimension free. Obtaining an efficient (that is, polynomial instead of pseudo-polynomial) rounding procedure requires more work, and we defer the details.

Our methods also yield related bounds that depend on n, and that work for negative coefficients. Our main applications in this paper, however, require dimension-free bounds and we defer further discussion to a later version of the paper. Our framework extends to polytopes/combinatorial structures that have certain exchange properties; Section 4 outlines this framework.

1.1 Applications

In this paper, we present two applications, each of which applies to non-bipartite matchings and matroid intersection.

Multi-budgeted Optimization: There has been substantial recent interest in the problem of optimizing linear (and also submodular functions) over a combinatorial structure with additional linear/budget/packing constraints [5, 16, 22, 18, 2, 9]. Given a polynomial-time solvable problem, such as maximum matching, can we optimize (in particular maximize) over feasible solutions with an additional budget constraint of the form $\mathbf{a}^T\mathbf{x} \leq b$ where $\mathbf{a}$ is a non-negative weight vector? It is easy to see that the problem becomes NP-hard via a reduction from knapsack. Randomized and deterministic polynomial-time approximation schemes (PTASes) have been developed for important combinatorial optimization problems subject to a single budget constraint: these include spanning trees [28], matchings and independent sets in two matroids [5]. Two approaches have been used.

Algebraic methods give a randomized pseudo-polynomial time algorithm for the *exact* version of some of these problems [8]; that is, given an integer B and an integer weight vector w, is there a solution S such that $w(S) = B$? Such algorithms can be used to obtain a randomized PTAS for any fixed number of budgets [16]. In the case of matroids and matroid intersection, this approach is limited to representable (i.e. linear) matroids [8].

A different, deterministic approach is via the use of Lagrangian relaxation combined with various technical properties of the underlying combinatorial structure. This approach was first used to develop a PTAS for spanning trees [28], and recently, matchings and matroid intersection with one budget [5]. These results have been also extended to matchings subject to two budget constraints and matroid independent sets with multiple budgets [16]. We note that a randomized PTAS for multi-budgeted matroid independent sets is also given in [9] via the randomized swap rounding scheme.

In this paper we obtain the following theorem as a relatively straightforward consequence of Theorem 1.1.

THEOREM 1.2. *For any fixed number of budget constraints, there is a randomized PTAS for multi-budgeted matroid intersection and multi-budgeted non-bipartite graph matching.*

Our techniques offer several advantages over previous methods. First, we obtain a PTAS for matroid intersection without any restriction on representability; this was not known prior to our work. Second, the concentration bounds naturally allow for a good approximation even when the number of budget constraints is not fixed as long as the constraints are sufficiently "loose". More precisely, if the size of each element is $O(\frac{1}{\log m})B$ where B is minimum budget size, then we can handle m budget constraints simultaneously. We can also handle the case when all except a small fixed number of constraints are loose. Alternatively, if we can tolerate an overflow of the budget constraints by an $O(\log m)$ factor, we can handle m budget constraints without any assumption on element sizes. These results arise naturally from our concentration results.

Finally, our approach can potentially be derandomized with the method of pessimistic estimators to obtain deterministic PTASes. (Unlike techniques relying on randomized exact-weight algorithms, whose derandomization is a major open problem.) In the case of matroid intersection, derandomization of our approach seems non-trivial but possible in principle. In this work, we give a deterministic PTAS for the case of multi-budgeted matchings; a deterministic PTAS was previously known only for two budgets [16]. Our algorithm can be viewed (indirectly) as a derandomization of our randomized algorithm.

THEOREM 1.3. *For any fixed number of budgets, there is a deterministic PTAS for multi-budgeted matching.*

Multi-objective Optimization: In multi-objective optimization we have several different objective functions $f_1, \dots, f_k$ and the goal is to simultaneously optimize them over a given polytope P. We restrict our attention to linear objective functions given by weight vectors $w_1, \dots, w_k$. Here, one is interested in the so called Pareto-optimal solutions; a solution S is Pareto-optimal if no other solution S' dominates S for each objective. Papadimitriou and Yannakakis [26] defined the notion of a succinct ε-approximate Pareto-optimal set which consists of a polynomial number of solutions $\mathcal{S}$ for given objectives $w_1, \dots, w_k$ such that every solution S' is approximately dominated by some point $S \in \mathcal{S}$; that is $w_i(S) \geq (1-\varepsilon)w_i(S')$ for all $1 \leq i \leq k$.

We use our rounding scheme combined with a result in [26] to obtain the following.

THEOREM 1.4. *There is a randomized polynomial-time algorithm that for any fixed $\varepsilon > 0$ computes a ε-approximate Pareto-optimal set for matroid intersection with a fixed number of objectives.*

The above was not previously known for general matroids. We remark that if the matroids are representable, then the results in [26] combined with the known exact algorithm for linear matroid intersection [8] imply a randomized fully polynomial-time approximation scheme (FPTAS). The same can be done for multi-objective matchings as well, using the randomized exact algorithm presented in [24]. We note that our techniques can be used to obtain a deterministic algorithm for multi-objective matchings, similar to Theorem 1.3. We omit the details here.

Further extensions. Our framework allows for a common generalization of the multi-budgeted and multi-objective setting, with any fixed number of budgets and a fixed number of objectives. More precisely, we can find an ε-approximate Pareto-optimal set in the presence of a constant number of budget constraints. This is dicussed further in Section 5. Other applications may arise in settings that involve a matching or matroid intersection problem with additional constraints.

1.2 Technical contribution and relation to prior work

Our rounding builds on and modifies the swap-rounding scheme from our recent work [9] that, given a point $\mathbf{x}$ in a n-dimensional polytope P, produces a random vertex corresponding to a solution R as follows:

- Express $\mathbf{x}$ as a convex combination of vertices of P, that is, $\mathbf{x} = \sum_{i=1}^{n} \alpha_i \mathbf{v}_i$ where each $\mathbf{v}_i$ is a vertex of P and $\sum_i \alpha_i = 1$.
- Let $\mathbf{w}_1 = \mathbf{v}_1$, $\beta_1 = \alpha_1$, and in each stage, merge two vertices $\mathbf{w}_i, \mathbf{v}_{i+1}$ into a new vertex $\mathbf{w}_{i+1}$ with coefficient $\beta_{i+1} = \beta_i + \alpha_{i+1}$ in such a way that $\mathbf{E}[\beta_{i+1}\mathbf{w}_{i+1}] = \beta_i \mathbf{w}_i + \alpha_{i+1}\mathbf{v}_{i+1}$ (**the merge operation**).
- After $n-1$ stages, obtain a vertex $\mathbf{w}_n = \mathbf{1}_R$ such that $\mathbf{E}[\mathbf{1}_R] = \sum_{i=1}^{n} \alpha_i \mathbf{v}_i = \mathbf{x}$.

Clearly, the crucial step which we have not described here is the merge operation. For matroids, the strong base-exchange property is useful in merging two bases by swapping two elements at a time and this leads to negative correlation. However, more complicated structures such as matchings and independent sets in the intersection of two matroids cannot be merged with a bounded number of elements being swapped at a time. For example, if M_1 and M_2 are two matchings, their symmetric difference $M_1 \Delta M_2$ consists of a collection of disjoint cycles and paths. If one of these cycles/paths C is long, merging the matchings to preserve expectation requires one to pick all edges in $C \cap M_1$ or $C \cap M_2$, leading to positive correlation among many variables.

In order to avoid the above, we modify the merge process so that the cycles and paths used in the merge operation are short. We do this by breaking long cycles and paths, effectively deleting some edges. (The idea of breaking long cycles/paths is natural and not new; it has been used in related contexts, for instance in [1] and more recently in [15].) Therefore, instead of $\mathbf{E}[\mathbf{1}_R] = \mathbf{x}$, we can only ensure that $\mathbf{E}[\mathbf{1}_R]$ is close to $\mathbf{x}$. The main technical difficulty is to analyze the above process. Although the above is intuitive for matchings, the process is considerably more complicated for matroid intersection. Despite this, it was shown in [9] that one can define a merge process that is based on covering the symmetric difference by a distribution over cycles; this decomposition was originally developed in [23] for the analysis of a local-search algorithm. One can adapt the idea of breaking cycles even to this more general setting.

The first difficulty in proving concentration bounds for our rounding scheme is that negative correlation is no longer available. Therefore, we resort to a martingale-based analysis. Our goal is to prove dimension-free bounds, and therefore direct applications of Azuma's inequality are insufficient. Our approach is to bound the total variance of the rounding process as function of the expectation $\mu = \sum a_i x_i$, in a way inspired by the lower-tail bound for submodular functions in [9]. However, further difficulties arise because of the fact that swap operations are randomly distributed over large sets of elements. Hence it is not obvious how to estimate the variance of a sequence of swap operations. This leads to a fairly technical analysis which we present in Section 4. The swap-rounding process is crucial for the analysis since an element that participates in a swap step effectively disappears after the step. We set up a somewhat generic analysis framework that is applicable to both of our problems. It could also be useful for related problems.

Apart from our recent work [9], the work that is most related to this paper is that of Arora, Frieze and Kaplan [1]. They considered the problem of rounding a fractional perfect matching $\mathbf{x}$ in a bipartite graph into a near-perfect matching such that concentration properties hold for any linear function of the edge variables. They developed an algorithm that is similar in spirit to swap-rounding with the idea of breaking long cycles when merging two cycles. However, there are crucial differences in both the results and the techniques. First, they obtain concentration bounds that allow an additive deviation depending polynomially in n (essentially as $\tilde{O}(\sqrt{n})$), where n is the size of the graph. This is necessary for them since they consider negative coefficients, and these bounds were sufficient for their applications. Further, these large additive terms imply that the rounding scheme and the analysis could be coarse — after some preliminary processing to reduce the number of merge steps to be poly-logarithmic in n, the analysis uses standard Chernoff-Hoeffding bounds for each merge, and a union bound over the merge steps. However, the bounds we seek are relative to the expectation $\mu = \sum a_i x_i$ which can be quite small when compared to n. Moreover, unlike the simpler setting of bipartite matchings, the merge process for matroid intersection is complex and hard to analyze in a direct fashion as in [1]. We remark that our martingale setup can be used to derive results similar to those in [1] also for the more general setting of matroid intersection.

We also mention the work in [15] on multi-budgeted matroid bases and bipartite matchings, when the number of budgets is a fixed constant. Unlike our work, they allow the budget constraints to be violated by a $(1+\varepsilon)$-factor; for matroid bases the weight of the solution obtained is optimal, while for bipartite matchings the weight of the solution is at least $(1-\varepsilon)$ times the optimum. The main technique of [15] is iterated rounding, with additional ideas for bipartite matchings.

Organization: In Section 2, we present our rounding procedure for matchings. In Section 3, we present our rounding procedure for matroid intersection. In Section 4, we prove the main concentration results. Finally, we give more details on our applications in Section 5. We describe our deterministic PTAS for multibudgeted matchings in Appendix A.

All our algorithms can be implemented in polynomial time via well-known and standard assumptions/algorithms for matchings and matroids. Due to space constraints, we do not discuss the details in this version of the paper.

2 Rounding in the matching polytope

In this section, we describe how our rounding approach can be applied to a point in the (non-bipartite) matching polytope. Let $G = (V, E)$ be an undirected graph. We denote by $\mathcal{M} \subseteq 2^E$ the set of matchings in G, and by $\mathcal{P} = conv(\{\mathbf{1}_M \mid M \in \mathcal{M}\})$ the matching polytope for G. Let $\mathbf{x} \in \mathcal{P}$ be a point in the matching polytope that is an input to our rounding procedure.

Our rounding procedure builds on the swap-rounding framework from [9]. First, a convex decomposition $\mathbf{x} = \sum_{\ell=1}^{n} \alpha_\ell \mathbf{1}_{I_\ell}$ with $I_\ell \in \mathcal{M}$ is obtained for $\mathbf{x}$. Assume that the coefficients are ordered so that $\alpha_1 \geq \cdots \geq \alpha_n$. (Even though the ordering of the coefficients is not crucial, it simplifies the analysis.) Let $J_1 := I_1$. We proceed in stages, where in the k-th stage, a new matching $J_{k+1} \in \mathcal{M}$ is produced by "merging" J_k and I_{k+1}. The formal rounding procedure **SwapRound** is described in the figure below.

Algorithm **SwapRound**$(\mathbf{x} = \sum_{\ell=1}^{n} \alpha_\ell \mathbf{1}_{I_\ell})$:
  Reorder so that $\alpha_1 \geq \alpha_2 \geq \ldots \geq \alpha_n$;
  $J_1 = I_1$;
  For $(k = 1$ to $n-1)$ do
    J_{k+1} =**MergeMatchings**$(\alpha_{k+1}, I_{k+1}, \sum_{\ell=1}^{k} \alpha_\ell, J_k)$;
  EndFor
  Output J_n.

It remains to specify the **MergeMatchings** procedure. Given two matchings $I, J \in \mathcal{M}$ with corresponding coefficients α, β, we want to choose **MergeMatchings**(α, I, β, J) such that a (random) matching M is returned that is in expectation close to the given linear combination of I and J, i.e., $\mathbf{E}[(\alpha+\beta)\mathbf{1}_M] \approx \alpha\mathbf{1}_I + \beta\mathbf{1}_J$, and such that the correlation between different edges of M is weak enough for concentration bounds to hold. The approach in [9] is to consider the symmetric difference $I\Delta J$ which consists of alternating paths and cycles $S_1, \ldots, S_m$; a merged matching M is obtained by including all edges $I \cap J$ in M, and for each set S_i, with probability $\frac{\alpha}{\alpha+\beta}$ all edges of $S_i \cap I$ are included in M, otherwise all edges of $S_i \cap J$ are included in M. This scheme satisfies the property that $\mathbf{E}[(\alpha+\beta)\mathbf{1}_M] = \alpha\mathbf{1}_I + \beta\mathbf{1}_J$, however, it creates positive correlations between many variables inside the sets S_i. As mentioned earlier, such strong positive correlations are unavoidable if $\mathbf{x}$ is to be preserved in expectation. Our approach returns a merged matching M whose expectation is slightly below the target value $\alpha\mathbf{1}_I + \beta\mathbf{1}_J$, in return we obtain concentration results for any linear function $f(M)$. Intuitively, we would like to break the S_i into smaller segments by removing some edges. The lemma below describes the properties we desire for the segments. The choice of the parameter p will be discussed later.

LEMMA 2.1. *Let $I, J \in \mathcal{M}$, $p \in \mathbb{Z}_+$, and $\delta = 1/(p-1)$. Then we can find in polynomial time a collection of sets $P_1, \ldots, P_m \subseteq I\Delta J$ with coefficients $\rho_i = 1/m$, such that $m \le p|I\Delta J|$, and for $\rho = p-1$:*

(i) For $i \in \{1, \ldots, m\}$, $I\Delta P_i \in \mathcal{M}$.

(ii) For $1 \le i \le m$, $|I \cap P_i| \le p$, $|J \cap P_i| \le p$ and hence $|P_i| \le 2p$.

(iii) $\sum_{i=1}^{m} \rho_i \mathbf{1}_{P_i} = (1+\delta)\rho\mathbf{1}_{I\setminus J} + \rho\mathbf{1}_{J\setminus I}$.

We defer the proof of the lemma to Section 2.1, and describe our merging procedure **MergeMatchings**. It works in iterations, where in each iteration either I or J is modified such that $|I\Delta J|$ strictly decreases, until $I = J$. If we (randomly) decide to alter I, we construct a collection of short alternating paths/cycles $\{P_i\}$, according to Lemma 2.1. Then a path P_i is chosen out of $\{P_i\}$ with probability ρ_i and I is replaced by $I\Delta P_i$. Similarly, when we (randomly) decide to change J, a collection of paths/cycles $\{P'_i\}$ of length at most $2p$ is constructed with probabilities $\{\sigma_i\}$ such that $\sum \sigma_i \mathbf{1}_{P'_1} = \frac{p-1}{p}\sigma\mathbf{1}_{I\setminus J} + \sigma\mathbf{1}_{J\setminus I}$, for some constant $\sigma > 0$. Again, such a family can be obtained through Lemma 2.1, by exchanging the roles of I and J, and by scaling ρ. Because of their use in modifying I and J, we call the alternating paths/cycles $\{P_i\}$ and $\{P'_i\}$ also *swap sets*. The **MergeMatchings** algorithm is summarized in the box where $\delta = 1/(p-1)$.

The random step to choose which set I or J to alter is based on $\alpha, \beta, \sigma, \rho$ in a natural fashion to ensure two properties: (i) the merged matching proportionally represents I, J with respect to α, β, and (ii) assuming $\delta = 0$, the step ensures that each element of $I\Delta J$ is equally likely to be removed as added. However, since $\delta > 0$, the important point here (in contrast to the work in [9]) is that the process creates a slight bias towards the elements we are removing. This bias allows the use of short swap sets which in turn helps prove concentration bounds. The bias also means that the process is no longer a martingale. Nevertheless, we show in Section 4 that the process can be analyzed by defining a related martingale.

Algorithm **MergeMatchings**(α, I, β, J):
  While $(I \ne J)$ do
    Generate a collection of alternating paths/cycles $\{P_i\}$ in $I\Delta J$ of length $\le 2p$ such that
      $\sum \rho_i \mathbf{1}_{P_i} = (1+\delta)\rho\mathbf{1}_{I\setminus J} + \rho\mathbf{1}_{J\setminus I}$, $\sum \rho_i = 1$, $\rho_i \ge 0$,
    and a collection of alternating paths/cycles $\{P'_i\}$ in $I\Delta J$ of length $\le 2p$ such that
      $\sum \sigma_i \mathbf{1}_{P'_i} = \frac{\sigma}{1+\delta}\mathbf{1}_{I\setminus J} + \sigma\mathbf{1}_{J\setminus I}$, $\sum \sigma_i = 1$, $\sigma_i \ge 0$.
    With probability $\frac{\beta\sigma\rho_i}{\alpha\rho+\beta\sigma}$ for each i, let $I := I\Delta P_i$.
    else with probability $\frac{\alpha\rho\sigma_i}{\alpha\rho+\beta\sigma}$ for each i, let $J := J\Delta P'_i$.
  EndWhile
  Output I.

The choice of the parameter p will depend on the parameter γ of Theorem 1.1 which controls the loss in expectation. There is a trade-off in choosing p: the larger p is chosen, the smaller is the bias, i.e., the closer is the matching M returned by **SwapRound** to the point $\mathbf{x}$ in expectation. However, a larger value of p leads also to longer swap sets in the collections $\{P_i\}$ and $\{P'_i\}$ used in **MergeMatchings**, which results in weaker concentration. The relation between p and γ together with the concentration results will be discussed in Section 4.

2.1 Creating collections of swap sets of small size In this section we prove Lemma 2.1, the basis of which is the following lemma.

LEMMA 2.2. *Let $I, J \in \mathcal{M}$, $p \in \mathbb{Z}_+$, and let $S \subseteq I\Delta J$ be a connected component of $I\Delta J$. In polynomial time, a collection of sets $P_1, \ldots, P_m \subseteq S$ can be obtained such that $m \le \max\{p, |S|\}$ and:*

- *For $1 \le i \le m$, $I\Delta P_i \in \mathcal{M}$.*
- *For $1 \le i \le m$, $|I \cap P_i| \le p$, $|J \cap P_i| \le p$ and hence $|P_i| \le 2p$.*
- *Each edge of $S \cap I$ appears in exactly p of the P_i and each edge of $S \cap J$ in exactly $p-1$ of the P_i.*

Proof. The set S is either an alternating path or cycle. Assume first that S is an alternating path $(e_1, \ldots, e_r)$ that is not a cycle. The case of alternating cycles is very similar and we discuss it briefly at the end of the proof. If $|S| \le 2p$, then we choose $m = p$ paths $P_1, \dots, P_m$, where each path P_i consists of all edges in S. To make sure that each edge of $S \cap J$ is contained in exactly $p-1$ sets, for each edge in J we select a unique path from the set of $m = p$ paths and remove it from that path. Since $|S| \leq 2p$ and S is an alternating path, $|S \cap J| \leq p$, and hence this is possible to do. One can easily check that the resulting sets fulfill the conditions of the lemma.

Now consider the case $|S| > 2p$. Assume $e_1 \in I$ (the case $e_1 \in J$ is analogous). For a fixed $j \in \{1, \dots, p\}$, we can break the path $(e_1, \dots, e_r)$ into short subpaths by removing the edges $e_{2j}, e_{2j+2p}, e_{2j+4p}, \dots$. Let $P_1, P_2, \dots, P_m \subseteq S$ be the collection of all subpaths obtained in this way for all choices of $j \in \{1, \dots, p\}$. Each subpath has length at most $2p-1$. It is easy to verify that each edge $e \in I \cap S$ is contained in exactly p of the subpaths $P_1, \dots, P_m$, each edge $e \in J \cap S$ is contained in exactly $p-1$ subpaths, and $m \leq |S|$. It remains to observe that $I \Delta P_i \in \mathcal{M}$ for $i \in \{1, \dots, m\}$, i.e., that $d_{I\Delta P_i}(v) \leq 1$ for $v \in V$. We have $d_{I\Delta P_i}(v) = d_I(v) = 1$ for all vertices $v \in V$ that are not endpoints of P_i. If v is an endpoint of P_i, then by the way how the P_i are constructed, the edge $e \in P_i$ that is adjacent to v is either an edge in I, in which case we have $d_{I\Delta P_i}(v) = 0$, or if $e \in J$ then v is also an endpoint of the path S, and hence $d_{I\Delta P_i}(v) \leq d_S(v) = 1$.

The argument above can be easily extended to S being an alternating cycle. If $|S| \leq 2p$ then again $m = p$ paths $P_1, \dots, P_m$ can be chosen where each path consists of all edges in S except up to one edge of J, to fulfill the condition that each edge in $S \cap J$ is contained in exactly $p-1$ of the paths. If $|S| > 2p$, then a collection of paths $P_1, \dots, P_m$ fulfilling the claims of the lemma is obtained by taking all paths of length $2p-1$ starting with an edge in I and going along the cycle in a fixed direction. One can check that this collection of paths satisfies the conditions of the lemma. ■

Finally, applying Lemma 2.2 to each component of $I\Delta J$ separately, a collection of short paths/cycles $P_1, \dots, P_m$ is obtained that covers all elements of $I\Delta J$, and implies Lemma 2.1.

Lemma 2.2 shows that in fact all probabilities $\{\rho_i\}$ in Lemma 2.1 can be chosen to be equal, and the same holds for the probabilities $\{\sigma_i\}$. Still, we presented the algorithm in the slightly more general setting to present a unified framework for matchings and matroid intersection. For matroid intersection the decomposition into swap sets requires non-uniform probabilities $\{\rho_i\}$.

3 Rounding for matroid intersection

In this section, we describe our new procedure to round a fractional solution in the matroid intersection polytope, i.e. the intersection of two matroid polytopes $P(\mathcal{M}_1) \cap P(\mathcal{M}_2)$ where both $\mathcal{M}_1 = (N, \mathcal{I}_1)$ and $\mathcal{M}_2 = (N, \mathcal{I}_2)$ are defined over the same ground set N. It was proved by Edmonds [12] that this polytope is integral and hence any fractional solution can be expressed as a convex combination of integer solutions. We assume that the matroids are given via membership oracles: that is, given a set $S \subseteq N$ the oracle returns whether S is independent in the matroid. Let $n = |N|$ denote the cardinality of the ground set.

3.1 Decomposition into irreducible paths and cycles

The basis of our rounding procedure is a decomposition of the symmetric difference between two independent sets into feasible exchanges. This is similar to the decomposition developed in [23, 9]; however, with some modifications due to the fact that we want to break long exchange cycles into shorter exchange paths. We use the construction of [9] as a black box and describe the important differences here.

Again, we use the standard constructs of matroid intersection; see [30]. For $I \in \mathcal{I}_1 \cap \mathcal{I}_2$, we define two digraphs $D_{\mathcal{M}_1}(I)$ and $D_{\mathcal{M}_2}(I)$ as follows.

- For each $i \in I, j \in N \setminus I$ with $I + j - i \in \mathcal{I}_1$, we have an arc $(i, j) \in D_{\mathcal{M}_1}(I)$;
- For each $i \in I, j \in N \setminus I$ with $I + j - i \in \mathcal{I}_2$, we have an arc $(j, i) \in D_{\mathcal{M}_2}(I)$.

When we refer to a matching in $D_{\mathcal{M}_l}(I)$ for $l = 1, 2$ we mean a matching in an undirected graph where the arcs of $D_{\mathcal{M}_l}(I)$ are treated as undirected edges. We define a digraph $D_{\mathcal{M}_1,\mathcal{M}_2}(I)$ as the union of $D_{\mathcal{M}_1}(I)$ and $D_{\mathcal{M}_2}(I)$. A directed cycle in $D_{\mathcal{M}_1,\mathcal{M}_2}(I)$ corresponds to a chain of feasible swaps. It is not necessarily the case that the entire cycle gives a valid exchange in both matroids. Nonetheless, it is known that if a cycle decomposes into two matchings which are unique on their set of vertices, respectively in $D_{\mathcal{M}_1}(I)$ and $D_{\mathcal{M}_2}(I)$, then the cycle corresponds to a feasible swap. This motivates the following definition.

DEFINITION 3.1. *We call a directed cycle C in $D_{\mathcal{M}_1,\mathcal{M}_2}(I)$* irreducible *if $C \cap D_{\mathcal{M}_1}(I)$ is the unique matching in $D_{\mathcal{M}_1}(I)$ and $C \cap D_{\mathcal{M}_2}(I)$ is the unique matching in $D_{\mathcal{M}_2}(I)$ covering exactly the vertex set $V(C)$. Otherwise, we call C* reducible.

Let us assume for simplicity that we have two sets $I, J \in \mathcal{I}_1 \cap \mathcal{I}_2$ and $|I| = |J|$. We also assume that $I \cap J = \emptyset$; if not, we can formally replace elements in $I \cap J$ by parallel copies which appear in I, J respectively that can always be exchanged without affecting independence. The following lemma, building on previous work [23], was proved in [9].

LEMMA 3.1. *Let $\mathcal{M}_\ell = (N, \mathcal{I}_\ell)$, $\ell = 1, 2$, be matroids on ground set N. Suppose that $I, J \in \mathcal{I}_1 \cap \mathcal{I}_2$ and $|I| = |J|$. Then we can find in polynomial time a collection of irreducible cycles $\{C_1, \dots, C_m\}, m \leq |I\Delta J|$, in $D_{\mathcal{M}_1,\mathcal{M}_2}(I)$, with coefficients $\gamma_i \geq 0$, $\sum_{i=1}^m \gamma_i = 1$, such that for some $\gamma > 0$, $\sum_{i=1}^m \gamma_i \mathbf{1}_{V(C_i)} = \gamma \mathbf{1}_{I\Delta J}$.*

In this work, we need a modification of this decomposition lemma which uses irreducible paths in addition to cycles. We define an irreducible path as follows.

DEFINITION 3.2. *We call a directed path* P *in* $D_{\mathcal{M}_1,\mathcal{M}_2}(I)$ irreducible *if both its endpoints are in* I, $P \cap D_{\mathcal{M}_1}(I)$ *is the unique matching on its vertices in* $D_{\mathcal{M}_1}(I)$, *and* $P \cap D_{\mathcal{M}_2}(I)$ *is the unique matching on its vertices in* $D_{\mathcal{M}_2}(I)$.

LEMMA 3.2. *For any irreducible path* P *in* $D_{\mathcal{M}_1,\mathcal{M}_2}(I)$, *the set* $I \Delta V(P)$ *is independent in* $\mathcal{I}_1 \cap \mathcal{I}_2$.

Proof. Observe that the edges of P alternate between $D_{\mathcal{M}_1}(I)$ and $D_{\mathcal{M}_2}(I)$, the vertices of P alternate between I and $N \setminus I$, and since both endpoints are in I, the length of P is even. Consider the matching $P \cap D_{\mathcal{M}_1}(I)$ and its vertex set $W_1 = V(P) \setminus \{w_2\}$, where w_2 is the endpoint of P incident to an edge of $D_{\mathcal{M}_2}(I)$. By assumption, this is the unique matching in $D_{\mathcal{M}_1}(I)$ covering W_1, and so $I \Delta W_1 \in \mathcal{I}_1$. By removing another element, we cannot violate independence, and hence we also have $I \Delta V(P) = (I \Delta W_1) \setminus \{w_2\} \in \mathcal{I}_1$. Similarly, we prove that $I \Delta V(P) \in \mathcal{I}_2$. ∎

The next lemma is an adaptation of Lemma 3.1 for our purposes.

LEMMA 3.3. *Let* $\mathcal{M}_\ell = (N, \mathcal{I}_1)$, $\ell = 1, 2$, *be matroids on ground set* N, *and let* $\delta = 1/(p-1)$, $p \in \mathbb{Z}$. *Suppose that* $I, J \in \mathcal{I}_1 \cap \mathcal{I}_2$ *and* $|I| = |J|$. *Then we can find in polynomial time a collection of irreducible cycles/paths* $\{P_1, \ldots, P_m\}$ *of length at most* $2p-1$, $m \le |I \Delta J|^2$, *in* $D_{\mathcal{M}_1,\mathcal{M}_2}(I)$, *with coefficients* $\rho_i \ge 0$, $\sum_{i=1}^m \rho_i = 1$, *such that for some* $\rho > 0$, $\sum_{i=1}^m \rho_i \mathbf{1}_{V(P_i)} = (1+\delta)\rho \mathbf{1}_{I \setminus J} + \rho \mathbf{1}_{J \setminus I}$.

Proof. Let $\{C_1, \ldots, C_{m'}\}$ be the collection of irreducible cycles on $I \Delta J$ provided by Lemma 3.1. Consider a cycle C_i with coefficient γ_i, which has length $2\ell_i$. Assume for now that $\ell_i > p$. For each vertex $v \in V(C_i) \cap I$, let P_{iv} be a directed path of length $2p-1$ starting at v, going along the cycle. The number of such paths is $\ell_i \le |I \Delta J|$. We assign a coefficient $\rho_i = \gamma_i/(p-1) = \delta\gamma_i$ to each of these paths. Since every vertex in $V(C) \cap I$ is contained in p such paths, and every vertex in $V(C) \setminus I$ is contained in $p-1$ such paths, these paths contribute $(1+\delta)\gamma_i$ to each vertex of $V(C) \cap I$ and γ_i to each vertex of $V(C) \setminus I$.

If $\ell_i \le p$, we can keep the cycle C_i in our collection with $\rho_i = (1+\delta)\gamma_i$. This cycle would contribute $(1+\delta)\gamma_i$ to all its vertices. To make the contributions consistent with the case above, we replace C_i with suitable coefficients by paths where some vertex of $V(C_i) \cap J$ is removed. It is easy to see that in this way we can decrease the contribution to each vertex of $V(C_i) \cap J$ to γ_i.

We repeat this for every cycle C_i, to produce a collection of at most $|I \Delta J|^2$ paths/cycles. These cover each vertex of $I \setminus J$ with a coefficient of exactly $(1+\delta)\gamma$ and each vertex of $J \setminus I$ with a coefficient of γ. Finally, we normalize the coefficients so that $\sum \rho_i = 1$, which gives the statement of the lemma for some value $\rho > 0$. ∎

Observe that switching the roles of I and J, we can work with $D_{\mathcal{M}_1,\mathcal{M}_2}(J)$ and obtain a collection of paths/cycles P'_i with coefficients σ_i such that $\sum_{i=1}^m \sigma_i \mathbf{1}_{V(P'_i)} = \sigma \mathbf{1}_{I \setminus J} + (1+\delta)\sigma \mathbf{1}_{J \setminus I}$. Equivalently, we can scale σ so that $\sum_{i=1}^m \sigma_i \mathbf{1}_{V(P'_i)} = \frac{\sigma}{1+\delta} \mathbf{1}_{I \setminus J} + \sigma \mathbf{1}_{J \setminus I}$.

3.2 The rounding procedure The rounding procedure follows the same framework as the rounding algorithm for matchings (Section 2). Again, we start with a convex combination $\mathbf{x} = \sum_{i=1}^n \alpha_i \mathbf{1}_{I_i}$ where $I_i \in \mathcal{I}_1 \cap \mathcal{I}_2$. Any fractional solution $\mathbf{x} \in P(\mathcal{M}_1) \cap P(\mathcal{M}_2)$ can be efficiently decomposed in this manner, see [30]. Assume also that the coefficients are ordered so that $\alpha_1 \ge \ldots \ge \alpha_n$. Let $J_1 := I_1$. As in the matchings case we proceed in stages, where in the k-th stage, a new independent set $J_{k+1} \in \mathcal{I}_1 \cap \mathcal{I}_2$ is produced from J_k and I_{k+1}.

Algorithm **SwapRound**($\mathbf{x} = \sum_{\ell=1}^n \alpha_\ell \mathbf{1}_{I_\ell}$):
  Reorder so that $\alpha_1 \ge \alpha_2 \ge \ldots \ge \alpha_n$
  $J_1 = I_1$;
  For ($k = 1$ to $n-1$) do
    $J_{k+1} =$ **MergeIntersectionSets**($\alpha_{k+1}, I_{k+1}, \sum_{\ell=1}^k \alpha_\ell, J_k$);
  EndFor
  Output J_n.

The merge operation is performed in much the same way as in [9], the difference being that we use the swap path/cycle structure provided by Lemma 3.3. Suppose we are merging two sets I, J which are disjoint. If not, we formally add copies of the shared elements, and we consider trivial exchanges between copies of the same element. This does not affect the actual rounding procedure, but helps in the analysis in the sense that each element of $I \cup J$ will be processed exactly once in the merging stage.

We apply Lemma 3.3 twice, to obtain (1) a convex combination of irreducible paths/cycles $\sum \rho_i \mathbf{1}_{V(P_i)}$ in $D_{\mathcal{M}_1,\mathcal{M}_2}(I)$, representing feasible swaps from I to J, and (2) a convex combination of irreducible paths/cycles $\sum \sigma_i \mathbf{1}_{V(P'_i)}$ in $D_{\mathcal{M}_1,\mathcal{M}_2}(J)$, representing feasible swaps from J to I. These swap sets cover the symmetric difference $I \Delta J$ almost uniformly, but as in the rounding algorithm for matchings, there is a slight bias towards the elements we are removing. For example, we have $\sum \rho_i \mathbf{1}_{V(P_i)} = (1+\delta)\rho \mathbf{1}_{I \setminus J} + \rho \mathbf{1}_{J \setminus I}$; i.e., elements of I are covered more often than the elements of J, when we work with I. This means elements are removed from I more often than added to I.

Algorithm **MergeIntersectionSets**(α, I, β, J):
While ($I \neq J$) do
Generate a collection of irreducible paths/cycles $\{P_i\}$ in $D_{\mathcal{M}_1,\mathcal{M}_2}(I)$ each containing at most $2p$ vertices such that $\sum \rho_i \mathbf{1}_{V(P_i)} = (1+\delta)\rho \mathbf{1}_{I\setminus J} + \rho \mathbf{1}_{J\setminus I}$, $\sum \rho_i = 1$, $\rho_i \geq 0$,
and a collection of irreducible paths/cycles $\{P'_i\}$ in $D_{\mathcal{M}_1,\mathcal{M}_2}(J)$ each containing at most $2p$ vertices such that $\sum \sigma_i \mathbf{1}_{V(P'_i)} = \frac{\sigma}{1+\delta}\mathbf{1}_{I\setminus J} + \sigma \mathbf{1}_{J\setminus I}$, $\sum \sigma_i = 1$, $\sigma_i \geq 0$.
With probability $\frac{\beta\sigma\rho_i}{\alpha\rho+\beta\sigma}$ for each i, let $I := I \Delta V(P_i)$.
else with probability $\frac{\alpha\rho\sigma_i}{\alpha\rho+\beta\sigma}$ for each i, let $J := J \Delta V(P'_i)$.
EndWhile
Output I.

We proceed in a sequence of random swaps, where each swap is chosen randomly from a distribution corresponding to this path/cycle structure, as can be seen from a concise description of the merge operation in the box. Every time we perform a swap, the size of $I\Delta J$ shrinks. We remark that it is necessary to recompute the entire path/cycle structure after each swap. We repeat this procedure, until I and J become identical.

4 Dimension-free concentration for swap-based processes

In this section, we prove concentration bounds for a general random process which encapsulates both our rounding procedures for matroid intersection and matching. The rounding procedures from the previous two sections can be described at a high level as follows.

4.1 The swap-based random process Let $P \subseteq [0,1]^N$ be an integer polytope and let $\mathcal{F} \subseteq 2^N$ be the family of the subsets of N that correspond to vertices of P. Given is a convex combination $\mathbf{x} = \sum_{i=1}^n \alpha_i \mathbf{1}_{I_i} \in P$, where $\alpha_i \geq 0$, $\sum_{i=1}^n \alpha_i = 1$ and $I_i \in \mathcal{F}$. We further assume for technical reasons that the indices are ordered so that $\alpha_1 \geq \ldots \geq \alpha_n$. Assuming this ordering slightly simplifies our analysis. However, with some modifications, the proof technique we employ could also be used without the assumption of the α_i being ordered, leading to slightly weaker constants in the exponents of the concentration bounds presented in Theorem 1.1.

We let $J_1 := I_1$. In the k-th merge operation, we process J_k and I_{k+1} to produce a new set J_{k+1}. The merge operation proceeds in a sequence of swap operations between J_k and I_{k+1}, where each swap set D is chosen from a certain distribution so that D contains at most p elements from each of J_k, I_{k+1}. The exact structure of the random swaps is given by Lemma 2.1 for matchings and by Lemma 3.3 for matroid intersection. In summary, if the current sets are I, J with coefficients $\alpha, \beta > 0$, there are coefficients $\rho, \sigma > 0$ such that

- with probability $\frac{\beta\sigma}{\alpha\rho+\beta\sigma}$, we replace I by $I\Delta D$ where $\mathbf{E}[\mathbf{1}_D] = (1+\delta)\rho\mathbf{1}_{I\setminus J} + \rho\mathbf{1}_{J\setminus I}$,
- with probability $\frac{\alpha\rho}{\alpha\rho+\beta\sigma}$, we replace J by $J\Delta D'$ where $\mathbf{E}[\mathbf{1}_{D'}] = \frac{\sigma}{1+\delta}\mathbf{1}_{I\setminus J} + \sigma\mathbf{1}_{J\setminus I}$.

Our goal is to prove the following.

THEOREM 4.1. *Let $\mathbf{a} \in [0,1]^n$ and let $a(S) = \sum_{i\in S} a_i$ be the associated linear function. Let $\mathbf{x} = \sum_{i=1}^n \alpha_i \mathbf{1}_{I_i} \in P$ and let $\mu = \mathbf{E}[a(R)]$, where $R = J_n \in \mathcal{F}$ is a random set obtained by the randomized swap rounding procedure using swaps satisfying $|D\cap I_{k+1}|, |D\cap J_k| \leq p$ as explained above. Then $\mathbf{E}[\mathbf{1}_R] \geq (1-\frac{1}{p})^2\mathbf{x}$ and*

i) For $\varepsilon \in [0,1]$, $\Pr[a(R) \leq (1-\varepsilon)\mu] \leq e^{-\mu\varepsilon^2/8p}$,
ii) For $\varepsilon \in [0,1]$, $\Pr[a(R) \geq (1+\varepsilon)\mu] \leq e^{-\mu\varepsilon^2/8p}$.
iii) For $t \geq 2$, $\Pr[a(R) \geq t\mu] \leq e^{-\mu(2t-3)/8p}$.

Given this theorem, we can derive Theorem 1.1 as follows. Given $\gamma \in (0, \frac{1}{2}]$ in Theorem 1.1, we pick the smallest integer p such that $\frac{2}{p} \leq \gamma$. (Note that $p \geq 4$ and $\delta = \frac{1}{p-1} \leq \frac{1}{3}$.) Then we have $\mathbf{E}[\mathbf{1}_R] \geq (1-\frac{1}{p})^2\mathbf{x} \geq (1-\frac{2}{p})\mathbf{x} \geq (1-\gamma)\mathbf{x}$. As we show below, we can in fact scale down the initial coefficients in such a way that $\mathbf{E}[\mathbf{1}_R] = (1-\gamma)\mathbf{x}$, as required by Theorem 1.1. Since we picked the smallest integer p satisfying $\frac{2}{p} \leq \gamma$, we have $p \leq \frac{2}{\gamma} + 1 \leq \frac{2.5}{\gamma}$. This is why $1/8p$ in the exponent becomes $\gamma/20$ in Theorem 1.1.

Let us point out some differences between this work and [9]. In [9], the random process in terms of the fractional solution throughout the algorithm forms a martingale; i.e., the expectation is preserved in each step. This is not true here, since the structure of exchange cycles and paths provided by Lemmas 2.1 and 3.3 is biased, in the sense that elements are more often removed than added. In order to facilitate the analysis, we first define a *modified random process* which is related to the evolution of the fractional solution and in fact *forms a martingale*. This allows us to apply our martingale analysis and finally prove the concentration results that we claimed.

4.2 A modified martingale process We define $\beta_\ell = \sum_{i=1}^{\ell} \alpha_i$, and consider a related linear combination $\tilde{\mathbf{x}} = \sum_{i=1}^{n} \tilde{\alpha}_i \mathbf{1}_{I_i}$, with

$$\tilde{\alpha}_i = \frac{\alpha_i + \beta_{i-1}}{\alpha_i + (1+\delta)^2 \beta_{i-1}} \alpha_i,$$

where $\delta = 1/(p-1)$ is the parameter used in the rounding procedure. The modified process starts from the point $\tilde{\mathbf{x}} = \sum_{i=1}^{n} \tilde{\alpha}_i \mathbf{1}_{I_i}$. As we merge the first k sets, we produce a new set J_k, and the coefficient assigned to this set will be $\beta_k = \sum_{i=1}^{k} \alpha_i$. I.e., at this point the linear combination becomes $\beta_k \mathbf{1}_{J_k} + \sum_{\ell=k+1}^{n} \tilde{\alpha}_\ell \mathbf{1}_{I_\ell}$. Eventually, we obtain a set J_n with coefficient $\beta_n = \sum_{i=1}^{n} \alpha_i = 1$.

The new random process defines a martingale, as we state more precisely in Lemma 4.1. Thus if $R = J_n$ is the final outcome of the rounding procedure, we will show that $\mathbf{E}[\mathbf{1}_R] = \tilde{\mathbf{x}} = \sum_{i=1}^{n} \tilde{\alpha}_i \mathbf{1}_{I_i}$. Note that each coefficient $\tilde{\alpha}_i$ is at least $\alpha_i/(1+\delta)^2 = \alpha_i/(1+\frac{1}{p-1})^2 = (1-\frac{1}{p})^2 \alpha_i$. Hence $\mathbf{E}[\mathbf{1}_R] \geq (1-\frac{1}{p})^2 \mathbf{x}$, as required by Theorem 4.1.

We remark that in order to achieve the condition $\mathbf{E}[\mathbf{1}_R] = (1-\gamma)\mathbf{x}$ for some $\gamma \geq \frac{2}{p}$, as required by Theorem 1.1, we can proceed as follows. Assuming that the initial linear combination is $\mathbf{x}' = \sum_{i=1}^{n} \alpha'_i \mathbf{1}_{I_i}$, we can scale the coefficients down in a suitable way to obtain $\mathbf{x} = \sum_{i=1}^{n} \alpha_i \mathbf{1}_{I_i}$ (formally adding an empty set with coefficient $1 - \sum_{i=1}^{n} \alpha_i$), so that $\tilde{\alpha}_i = (1-\gamma)\alpha'_i$. We omit the details, as this would further encumber the notation.

Next, we refine the definition of the modified random process and describe what we mean by $\tilde{\mathbf{x}}$ in the middle of the merge operation. The following lemma describes the process and proves that it is a martingale.

LEMMA 4.1. *Let $\mathcal{F}$ denote the family of feasible solutions (either matchings or sets in matroid intersection). Let $\mathbf{x} = \sum_{i=1}^{n} \alpha_i \mathbf{1}_{I_i}$, where $I_i \in \mathcal{F}$. Let $J_1 = I_1$, and for $k \in \{1, \ldots, n-1\}$ let J_{k+1} be the set produced by merging J_k, I_{k+1} as in the procedure* **MergeMatchings** *or* **MergeIntersectionSets**. *Define $\beta_k = \sum_{i=1}^{k} \alpha_i$ and $\tilde{\mathbf{x}} = \sum_{i=1}^{n} \tilde{\alpha}_i \mathbf{1}_{I_i}$, where*

$$\tilde{\alpha}_i = \frac{\alpha_i + \beta_{i-1}}{\alpha_i + (1+\delta)^2 \beta_{i-1}} \alpha_i.$$

After t steps of merging J_k with I_{k+1}, assuming the elements already processed are in $J_k \cap I_{k+1}$, let us define

$$\tilde{\mathbf{x}}_{k,t} = (\beta_k + \alpha_{k+1}) \mathbf{1}_{J_k \cap I_{k+1}} + \beta_k \mathbf{1}_{J_k \setminus I_{k+1}} + \tilde{\alpha}_{k+1} \mathbf{1}_{I_{k+1} \setminus J_k} + \sum_{\ell=k+2}^{n} \tilde{\alpha}_\ell \mathbf{1}_{I_\ell}.$$

Then we have $\tilde{\mathbf{x}} = \sum_{i=1}^{n} \tilde{\alpha}_i \mathbf{1}_{I_i} = \tilde{\mathbf{x}}_{1,0}$ and the process $(\tilde{\mathbf{x}}_{1,0}, \tilde{\mathbf{x}}_{1,1}, \ldots, \tilde{\mathbf{x}}_{2,0}, \tilde{\mathbf{x}}_{2,1}, \ldots)$ forms a martingale. In particular, if R is the final rounded solution, $\mathbf{E}[\mathbf{1}_R] = \tilde{\mathbf{x}} = \tilde{\mathbf{x}}_{1,0}$.

Proof. Note that the elements that have not been processed yet are counted in $\tilde{\mathbf{x}}_{k,t}$ with coefficients $\tilde{\alpha}_i$ which are smaller than α_i - this accounts for the fact that they will be kicked out more often than added. At the beginning of the process, we have $J_1 = I_1$ and formally $J_1 \cap I_1 = \emptyset$; therefore,

$$\tilde{\mathbf{x}}_{1,0} = \beta_1 \mathbf{1}_{J_1} + \tilde{\alpha}_2 \mathbf{1}_{I_2} + \sum_{\ell=3}^{n} \tilde{\alpha}_\ell \mathbf{1}_{I_\ell} = \sum_{\ell=1}^{n} \tilde{\alpha}_\ell \mathbf{1}_{I_\ell}$$

because $\beta_1 = \alpha_1 = \tilde{\alpha}_1$.

To prove the martingale property, consider a fractional solution $\tilde{\mathbf{x}}_{k,t}$. The procedure operates on the sets J_k and I_{k+1}. If it decides to modify J_k, some coordinates on J_k lose β_k (when they are removed from J_k) and some coordinates on I_{k+1} gain $\beta_k + \alpha_{k+1} - \tilde{\alpha}_{k+1}$ (when they move from $I_{k+1} \setminus J_k$ to $I_{k+1} \cap J_k$). If the procedure decides to modify I_{k+1}, some coordinates on J_k gain α_{k+1} (when they move to $J_k \cap I_{k+1}$), and some coordinates on I_{k+1} lose $\tilde{\alpha}_{k+1}$ (when they are removed from I_{k+1}).

Now let us compute the expected change for a fixed element $j \in J_k$. To simplify notation, let us use $\alpha = \alpha_{k+1}$, $\beta = \beta_k$ and $\tilde{\alpha} = \tilde{\alpha}_{k+1}$. The probability that we remove j from J_k is the probability that we modify J_k and j participates in the chosen swap path, i.e.

$$\Pr[j \text{ is removed from } J_k] = \sum_{i:j \in V(P_i)} \frac{\alpha\rho\sigma_i}{\alpha\rho + \beta\sigma} = \frac{\alpha\rho\sigma}{\alpha\rho + \beta\sigma}.$$

Similarly, the probability that an element $j \in J_k$ gains by being added to I_{k+1} is

$$\Pr[j \text{ is added to } I_{k+1}] = \sum_{i:i \in V(P'_i)} \frac{\beta\sigma\rho_i}{\alpha\rho + \beta\sigma} = \frac{\beta\sigma\rho}{\alpha\rho + \beta\sigma}.$$

In the first case, the coordinate X_j loses β, while in the second case, it gains α. Therefore, if X'_j denotes the coordinate value after this step,

$$\mathbf{E}[X'_j \mid X_j] = X_j - \frac{\alpha\rho\sigma}{\alpha\rho + \beta\sigma}\beta + \frac{\beta\sigma\rho}{\alpha\rho + \beta\sigma}\alpha = X_j.$$

It is similar but slightly more involved to analyze the change in coordinates on I_{k+1}. The probability that we remove j from I_{k+1} is the probability that I_{k+1} is modified and j happens to be on the chosen swap path:

$$\Pr[j \text{ is removed from } I_{k+1}] = \sum_{i:j \in V(P'_i)} \frac{\beta\sigma\rho_i}{\alpha\rho + \beta\sigma} = (1+\delta)\frac{\beta\sigma\rho}{\alpha\rho + \beta\sigma}.$$

The probability that j is added to J_k is the following:

$$\Pr[j \text{ is added to } J_k] = \sum_{i:i \in V(P_i)} \frac{\alpha\rho\sigma_i}{\alpha\rho + \beta\sigma} = \frac{1}{1+\delta}\frac{\alpha\rho\sigma}{\alpha\rho + \beta\sigma}.$$

Note the asymmetry here, due to the factors $(1+\delta)$. Recall that in the first case, X_j loses $\tilde{\alpha}$, while in the second case, X_j gains $\alpha+\beta-\tilde{\alpha}$. A computation using $\tilde{\alpha} = \frac{(\alpha+\beta)\alpha}{\alpha+(1+\delta)^2\beta}$ reveals that

$$\begin{aligned}
\mathbf{E}[X'_j \mid X_j] &= X_j - (1+\delta)\frac{\beta\sigma\rho}{\alpha\rho+\beta\sigma}\tilde{\alpha} \\
&\quad + \frac{1}{1+\delta}\frac{\alpha\rho\sigma}{\alpha\rho+\beta\sigma}(\alpha+\beta-\tilde{\alpha}) \\
&= X_j - (1+\delta)\frac{\beta\sigma\rho}{\alpha\rho+\beta\sigma}\frac{\alpha+\beta}{\alpha+(1+\delta)^2\beta}\alpha \\
&\quad + \frac{1}{1+\delta}\frac{\alpha\rho\sigma}{\alpha\rho+\beta\sigma}\frac{(\alpha+\beta)(1+\delta)^2\beta}{\alpha+(1+\delta)^2\beta} \\
&= X_j.
\end{aligned}$$

■

4.3 Martingale analysis In the following, we work towards the proof of Theorem 4.1. By scaling $\mathbf{a}$ by $1/p$, let us assume that $a_i \in [0, 1/p]$, so that $a(D\cap I), a(D\cap J) \le 1$ for any swap set D. We prove Theorem 4.1 by proving $\Pr[a(R) \le (1-\varepsilon)\mu] \le e^{-\mu\varepsilon^2/8}$, $\Pr[a(R) \ge (1+\varepsilon)\mu] \le e^{-\mu\varepsilon^2/8}$ and $\Pr[a(R) \ge t\mu] \le e^{-\mu(2t-3)/8}$.

The core of the proof is to estimate the exponential moment $\mathbf{E}[e^{\lambda(a(R)-\mu)}]$ and bound it by a factor depending only on the expectation, of the form $e^{O(\lambda^2\mu)}$. The difficulty in analyzing this process is that the swap set in each step comes from a probability distribution and hence it is not easy to charge the variance of the current step to the value of elements that have been processed. We develop an inductive approach to this problem relying on two probabilistic lemmas (Lemma 4.2 and 4.3) which appear to be new and somewhat different from traditional martingale analysis.

We focus on the analysis of the modified random process in terms of $\tilde{\mathbf{x}}$. In the k-th stage, we are merging sets J_k and I_{k+1}, with coefficients $\beta_k = \sum_{i=1}^{k}\alpha_k$ and $\tilde{\alpha}_{k+1}$. Define also $\tilde{\beta}_k$ so that $\tilde{\beta}_k + \tilde{\alpha}_{k+1} = \beta_k + \alpha_{k+1}$. Recall that the coefficients are ordered in descending order of magnitude, so we always have $\alpha_{k+1} \le \beta_k$. For simplicity, let us drop the indices and denote the sets by I, J, and the coefficients by $\alpha, \beta, \tilde{\alpha}, \tilde{\beta}$. Since $\tilde{\alpha} \le \alpha$ and $\tilde{\beta}+\tilde{\alpha} = \beta+\alpha$, we have $\tilde{\alpha} \le \alpha \le \beta \le \tilde{\beta}$. While $I\Delta J$ is nonempty, we perform the following rounding step. Assume that the current (modified) fractional solution is $\tilde{\mathbf{x}}$. One swap operation can be summarized as follows.

Swap Operation.

i) With some probability, choose a random $D \subseteq I\Delta J$ and define $I' := I\Delta D$, $\tilde{\mathbf{x}}' := \tilde{\mathbf{x}} - \tilde{\alpha}\mathbf{1}_{D\cap I} + \alpha\mathbf{1}_{D\cap J}$.

ii) Otherwise, choose a random $D \subseteq I\Delta J$ (according to a possibly different distribution), and let $J := J\Delta D$, $\tilde{\mathbf{x}}' := \tilde{\mathbf{x}} + \tilde{\beta}\mathbf{1}_{D\cap I} - \beta\mathbf{1}_{D\cap J}$.

iii) The distributions are chosen so that $\mathbf{E}[\tilde{\mathbf{x}}' \mid \tilde{\mathbf{x}}] = \tilde{\mathbf{x}}$.

iv) We have $a(D\cap I), a(D\cap J) \le 1$ with probability 1.

Let us now analyze how a linear function $\mathbf{a}\cdot\tilde{\mathbf{x}}$ behaves under such swap operations. Let Z be the change of $\mathbf{a}\cdot\tilde{\mathbf{x}}$ due to the first swap, i.e., $Z = \mathbf{a}\cdot(\tilde{\mathbf{x}}' - \tilde{\mathbf{x}})$. Furthermore, let W be the random variable given by $a(D)$, where D is the swap set used in the first swap iteration. We have $|Z| \le W$ and $W = a(D\cap I) + a(D\cap J) \le 2$. First, the following elementary inequality.

LEMMA 4.2. *Let $Y = \alpha AX - \beta(1-A)X$, where $\alpha, \beta \in [0,1]$, A, X are (possibly correlated) random variables with $A \in \{0,1\}, X \in [0,1]$, and $\mathbf{E}[Y] = 0$. Then*

$$\mathbf{E}[Y^2] \le \alpha\beta\mathbf{E}[X].$$

Proof. Since $A \in \{0,1\}$, $Y^2 = \alpha^2AX^2 + \beta^2(1-A)X^2$. Therefore,

$$\begin{aligned}
\mathbf{E}[Y^2] &= \alpha^2\mathbf{E}[AX^2] + \beta^2\mathbf{E}[(1-A)X^2] \\
&\le \alpha^2\mathbf{E}[AX] + \beta^2\mathbf{E}[(1-A)X]
\end{aligned}$$

using the fact that $X \in [0,1]$. Since we assume $\mathbf{E}[Y] = \alpha\mathbf{E}[AX] - \beta\mathbf{E}[(1-A)X] = 0$, we have

$$\begin{aligned}
0 &= (\beta-\alpha)\mathbf{E}[Y] \\
&= (\alpha\beta-\alpha^2)\mathbf{E}[AX] + (\alpha\beta-\beta^2)\mathbf{E}[(1-A)X].
\end{aligned}$$

Adding this to the inequality above, we get

$$\mathbf{E}[Y^2] \le \alpha\beta\mathbf{E}[AX] + \alpha\beta\mathbf{E}[(1-A)X] = \alpha\beta\mathbf{E}[X].$$

■

Next, we have the following inequality for a single swap.

LEMMA 4.3. *Let $\tilde{\mathbf{x}}'$ be obtained from $\tilde{\mathbf{x}}$ by a single swap step, using a random swap set D as above, where $\alpha \le \beta$, and let $Z = \mathbf{a}\cdot(\tilde{\mathbf{x}}' - \tilde{\mathbf{x}})$ and $W = a(D)$. Then for any $\lambda \in [-\frac{1}{4}, \frac{1}{4}]$,*

$$\mathbf{E}[e^{\lambda Z - \lambda^2\alpha\beta W}] \le 1.$$

Proof. Since $\lambda Z - \lambda^2\alpha\beta W \le \lambda Z \le \frac{1}{4}$, we use the following elementary bound: $e^x \le 1 + x + \frac{5}{9}x^2$ for $x \le \frac{1}{4}$. We obtain

$$\begin{aligned}
\mathbf{E}[e^{\lambda Z - \lambda^2\alpha\beta W}] \le& 1 + \mathbf{E}[\lambda Z - \lambda^2\alpha\beta W] \\
&+ \frac{5}{9}\mathbf{E}[(\lambda Z - \lambda^2\alpha\beta W)^2].
\end{aligned}$$

We have $\mathbf{E}[Z] = \mathbf{a}\cdot\mathbf{E}[\tilde{\mathbf{x}}' - \tilde{\mathbf{x}}] = 0$ due to condition (iii). Therefore,

$$\mathbf{E}[e^{\lambda Z - \lambda^2\alpha\beta W}] \le 1 - \lambda^2\alpha\beta\mathbf{E}[W] + \frac{5}{9}\lambda^2\mathbf{E}[(Z - \lambda\alpha\beta W)^2].$$

The key is to estimate the last expectation. We claim that

$$\mathbf{E}[(Z - \lambda\alpha\beta W)^2] \le \alpha\beta(1+|\lambda|)^2\mathbf{E}[W]. \tag{4.1}$$

If we prove this, we are done, because $\frac{5}{9}(1+|\lambda|)^2 \leq \frac{5}{9}(1+\frac{1}{4})^2 < 1$, and therefore

$$\begin{aligned}&\mathbf{E}[e^{\lambda Z-\lambda^2\alpha\beta W}]\\&\qquad\leq 1-\lambda^2\alpha\beta\mathbf{E}[W]+\frac{5}{9}\lambda^2\alpha\beta(1+|\lambda|)^2\mathbf{E}[W]\leq 1.\end{aligned}$$

Proof of (4.1). We introduce the following random variables:

- $W_I = a(D\cap I), W_J = a(D\cap J)$.
- $Z_I = \mathbf{a}_I\cdot(\tilde{\mathbf{x}}'-\tilde{\mathbf{x}}), Z_J = \mathbf{a}_J\cdot(\tilde{\mathbf{x}}'-\tilde{\mathbf{x}})$, where $\mathbf{a}_I, \mathbf{a}_J$ are the weight functions restricted to I, J respectively.
- A is an indicator variable such that $A = 1$ if we are modifying I (step i)), and $A = 0$ if we are modifying J (step ii)).
- I.e., $W = W_I + W_J$ and $Z = Z_I + Z_J$, where

$$Z_I = -A\tilde{\alpha}W_I + (1-A)\tilde{\beta}W_I \tag{4.2}$$

and

$$Z_J = A\alpha W_J - (1-A)\beta W_J. \tag{4.3}$$

The assumption that $\mathbf{E}[\tilde{\mathbf{x}}' \mid \tilde{\mathbf{x}}] = 0$ implies that $\mathbf{E}[Z_I] = \mathbf{E}[Z_J] = 0$. Let us write $\mathbf{E}[(Z-\lambda\alpha\beta W)^2]$ as follows:

$$\mathbf{E}[(Z-\lambda\alpha\beta W)^2] = \mathbf{E}[Z^2]-2\lambda\alpha\beta\mathbf{E}[WZ]+\lambda^2\alpha^2\beta^2\mathbf{E}[W^2].$$

First, let us estimate $\mathbf{E}[Z^2]$. We have

$$\begin{aligned}\mathbf{E}[Z^2] &= \mathbf{E}[(Z_I+Z_J)^2]\\&= \mathbf{E}[Z_I^2]+2\mathbf{E}[Z_IZ_J]+\mathbf{E}[Z_J^2]\\&\leq \mathbf{E}[Z_I^2]+\mathbf{E}[Z_J^2]\end{aligned}$$

because Z_I and Z_J always have opposite signs. The variable Z_J when decomposed as given by (4.3) satisfies the assumptions of Lemma 4.2 and hence $\mathbf{E}[Z_J^2] \leq \alpha\beta\mathbf{E}[W_J]$. Similarly, Z_I as given by (4.2) satisfies the same assumption with $\tilde{\alpha}, \tilde{\beta}$ instead of α, β, and so $\mathbf{E}[Z_I^2] \leq \tilde{\alpha}\tilde{\beta}\mathbf{E}[W_I]$. Recall that we ordered the coefficients, so that $\alpha \leq \beta$; this also means that $\tilde{\alpha} \leq \alpha \leq \beta \leq \tilde{\beta}$, since $\alpha+\beta = \tilde{\alpha}+\tilde{\beta}$. Hence, $\tilde{\alpha}\tilde{\beta} \leq \alpha\beta$.

We can conclude that

$$\mathbf{E}[Z^2] \leq \mathbf{E}[Z_I^2]+\mathbf{E}[Z_J^2] \leq \alpha\beta\mathbf{E}[W_I]+\alpha\beta\mathbf{E}[W_J] = \alpha\beta\mathbf{E}[W].$$

To estimate $\mathbf{E}[ZW]$, we use the Cauchy-Schwartz inequality:

$$\begin{aligned}|\mathbf{E}[ZW]| &\leq (\mathbf{E}[Z^2]\mathbf{E}[W^2])^{1/2} \leq (\alpha\beta\mathbf{E}[W]\mathbf{E}[W^2])^{1/2}\\&\leq \sqrt{2\alpha\beta}\,\mathbf{E}[W],\end{aligned}$$

using the fact that $\mathbf{E}[W^2] \leq 2E[W]$ since $0 \leq W \leq 2$. Hence,

$$\begin{aligned}&\mathbf{E}[(Z-\lambda\alpha\beta W)^2]\\&\qquad= \mathbf{E}[Z^2]-2\lambda\alpha\beta\mathbf{E}[ZW]+\lambda^2\alpha^2\beta^2\mathbf{E}[W^2]\\&\qquad\leq \alpha\beta\mathbf{E}[W]+2|\lambda|\alpha\beta\sqrt{2\alpha\beta}\mathbf{E}[W]\\&\qquad\qquad+2\lambda^2\alpha^2\beta^2\mathbf{E}[W]\\&\qquad= \alpha\beta(1+|\lambda|\sqrt{2\alpha\beta})^2\mathbf{E}[W]\\&\qquad\leq \alpha\beta(1+|\lambda|)^2\mathbf{E}[W]\end{aligned}$$

using $\alpha+\beta \leq 1$ which implies that $\alpha\beta \leq 1/4$. ■

Remark. The proof above is the only place where we exploit the ordering of the exponents $\alpha_1 \leq \cdots \leq \alpha_n$. One can show that even without assuming any particular ordering of the coefficients, $\tilde{\alpha}\tilde{\beta} \leq (1+\delta)^2\alpha\beta$ holds. By slightly modifying the statement of this lemma and the remaining part of the proof, this additional factor of $(1+\delta)^2$ can be propagated in the exponent, still leading to a concentration bound, however with a slightly weaker constant in the exponent.

Now we can add up the contributions to the exponential moment over an entire merge operation. Let us define $Z_1, Z_2, Z_3, \ldots$ to be the changes of $\mathbf{a}\cdot\tilde{\mathbf{x}}$ during successive swaps in merging I and J. Similarly, let us define $W_1 = a(D_1)$ to be the weight of the first swap, etc. Notice that the number of swaps needed to merge I and J is a random variable since the swap sets are not necessarily of the same size. However, this does not impose any further difficulty in the analysis below. In particular, by adding dummy swaps at the end if necessary we can assume that the number of swaps needed to merge I and J is always the same. Notice that $Z_j \in [-1,1]$ and $\mathbf{E}[Z_j \mid \mathcal{H}_{j-1}] = 0$, where $\mathcal{H}_{j-1}$ is the history of the merging procedure up to swap iteration $j-1$.

LEMMA 4.4. *Let $\lambda \in [-\frac{1}{4}, \frac{1}{4}]$ and $k \in \mathbb{N}$ be some fixed number of swap iterations, then*

$$\mathbf{E}[e^{\lambda\sum_{j=1}^k Z_j}] \leq e^{\lambda^2\alpha\beta a(I\Delta J)}.$$

Proof. We prove the result by induction on k. For $k = 1$ we have

$$\begin{aligned}\mathbf{E}[e^{\lambda Z_1}] &= \mathbf{E}[e^{\lambda Z_1-\lambda^2\alpha\beta W_1}\cdot e^{\lambda^2\alpha\beta W_1}]\\&\leq \mathbf{E}[e^{\lambda Z_1-\lambda^2\alpha\beta W_1}]e^{\lambda^2\alpha\beta a(I\Delta J)}\\&\leq e^{\lambda^2\alpha\beta a(I\Delta J)},\end{aligned}$$

where the first inequality follows from $W_1 \leq a(I\Delta J)$ with probability 1, and the second inequality follows from Lemma 4.3.

Consider now the case $k > 1$. We have

$$\mathbf{E}[e^{\lambda\sum_{j=1}^k Z_j}] = \mathbf{E}[e^{\lambda Z_1}\mathbf{E}[e^{\lambda\sum_{j=2}^k Z_j} \mid \mathcal{H}_1]],$$

where $\mathcal{H}_1$ encodes a particular outcome of the first swap operation. Let $D_1 \subseteq I\Delta J$ be the swap set used in the first iteration. Let I' and J' be the updated (random) sets obtained from I and J after the first swap using the swap set D_1. Since $a(I'\Delta J') = a(I\Delta J) - W_1$, we can apply the inductive hypothesis to obtain

$$\mathbf{E}[e^{\lambda \sum_{j=2}^{k} Z_j} \mid \mathcal{H}_1] \leq e^{\lambda^2 \alpha\beta(a(I\Delta J) - W_1)},$$

and hence

$$\begin{aligned}\mathbf{E}[e^{\lambda \sum_{j=1}^{k} Z_j}] &\leq e^{\lambda^2\alpha\beta a(I\Delta J)}\mathbf{E}[e^{\lambda Z_1 - \lambda^2\alpha\beta W_1}] \\ &\leq e^{\lambda^2\alpha\beta a(I\Delta J)},\end{aligned}$$

where the last inequality follows from Lemma 4.3. ∎

So far, we only considered one merge operation. In the following we show how to bound the exponential moment for the whole swap rounding procedure. Given is a fractional solution $\tilde{\mathbf{x}} = \sum_{i=1}^{n} \tilde{\alpha}_i \mathbf{1}_{I_i}$ with $I_i \in \mathcal{F}$ for $i \in [n]$. Let $J_1 = I_1$, and let J_k for $k \in \{1, \ldots, n\}$ be the set obtained by merging J_{k-1} and I_k. After k merge operations, the current fractional solution is given by $\beta_{k+1}\mathbf{1}_{J_{k+1}} + \sum_{i=k+2}^{n} \tilde{\alpha}_i \mathbf{1}_{I_i}$ where $\beta_{k+1} = \sum_{i=1}^{k+1} \alpha_i$. The set returned by the algorithm is $R = J_n$.

LEMMA 4.5. *Let $\lambda \in [-\frac{1}{4}, \frac{1}{4}]$, let J_n be obtained from $\tilde{\mathbf{x}}$ by randomized swap rounding as described above, and let $\mu = \mathbf{a} \cdot \tilde{\mathbf{x}} = \mathbf{E}[a(J_n)]$. Then*

$$\mathbf{E}[e^{\lambda(a(J_n) - \mu)}] \leq e^{2\lambda^2\mu}.$$

Proof. We proceed by induction on the number of terms n in the convex decomposition of $\tilde{\mathbf{x}} = \sum_{i=1}^{n} \tilde{\alpha}_i \mathbf{1}_{I_i}$. Notice that the base step of the inductive proof is trivial, since if $n = 1$, then $a(J_1) = a(I_1) = \mu$ and the left-hand side is 1. So assume $n \geq 2$. We have

$$\begin{aligned}&\mathbf{E}[e^{\lambda(a(J_n)-\mu)}] \\ &\quad= \mathbf{E}_{\mathcal{H}_{n-1}}[e^{\lambda[\beta_{n-1}a(J_{n-1}) - (\mu - \tilde{\alpha}_n a(I_n))]} \\ &\qquad\times \mathbf{E}_{\mathcal{H}_n}[e^{\lambda[a(J_n) - (\beta_{n-1}a(J_{n-1}) + \tilde{\alpha}_n a(I_n))]} \mid \mathcal{H}_{n-1}]] \\ &\quad= \mathbf{E}_{\mathcal{H}_{n-1}}[e^{\lambda[\beta_{n-1}a(J_{n-1}) - \sum_{i=1}^{n-1}\tilde{\alpha}_i a(I_i)]} \\ &\qquad\times \mathbf{E}_{\mathcal{H}_n}[e^{\lambda[a(J_n) - (\beta_{n-1}a(J_{n-1}) + \tilde{\alpha}_n a(I_n))]} \mid \mathcal{H}_{n-1}]],\end{aligned}$$

where $\mathcal{H}_k$ denotes the history of the first k merge operations in the rounding process. The inner expectation corresponds to the change in $e^{\lambda(\mathbf{a}\cdot\tilde{\mathbf{x}})}$ due to the last merge operation, merging J_{n-1} and I_n. By Lemma 4.4 we have

$$\begin{aligned}\mathbf{E}_{\mathcal{H}_n}[e^{\lambda[a(J_n) - (\beta_{n-1}a(J_{n-1}) + \tilde{\alpha}_n a(I_n))]} &\mid \mathcal{H}_{n-1}] \\ &\leq e^{\lambda^2\alpha_n\beta_{n-1}a(J_{n-1}\Delta I_n)} \\ &\leq e^{\lambda^2\alpha_n\beta_{n-1}(a(J_{n-1}) + a(I_n))},\end{aligned}$$

and hence

(4.4)
$$\begin{aligned}\mathbf{E}[e^{\lambda(a(J_n)-\mu)}] &\leq \mathbf{E}_{\mathcal{H}_{n-1}}[e^{\lambda(\beta_{n-1}a(J_{n-1}) - \sum_{i=1}^{n-1}\tilde{\alpha}_i a(I_i))} \\ &\quad\times e^{\lambda^2\alpha_n\beta_{n-1}(a(J_{n-1}) + a(I_n))}] \\ &= \mathbf{E}_{\mathcal{H}_{n-1}}\left[e^{(\lambda + \lambda^2\alpha_n)\left(\beta_{n-1}a(J_{n-1}) - \sum_{i=1}^{n-1}\tilde{\alpha}_i a(I_i)\right)}\right] \\ &\quad\times e^{\lambda^2\alpha_n\left(\beta_{n-1}a(I_n) + \sum_{i=1}^{n-1}\tilde{\alpha}_i a(I_i)\right)} \\ &= \mathbf{E}_{\mathcal{H}_{n-1}}\left[e^{\lambda'\left(a(J_{n-1}) - \sum_{i=1}^{n-1}\frac{\tilde{\alpha}_i}{\beta_{n-1}}a(I_i)\right)}\right] \\ &\quad\times e^{\lambda^2\alpha_n\left(\beta_{n-1}a(I_n) + \sum_{i=1}^{n-1}\tilde{\alpha}_i a(I_i)\right)}\end{aligned}$$

where we set $\lambda' = \lambda(1 + \lambda\alpha_n)\beta_{n-1}$. Notice that $|\lambda'| \leq |\lambda|(1 + |\lambda|\alpha_n)\beta_{n-1} \leq |\lambda|(1 + \alpha_n)(1 - \alpha_n) \leq |\lambda| \leq \frac{1}{4}$. Hence $\lambda' \in [-\frac{1}{4}, \frac{1}{4}]$.

Next, we will apply the inductive hypothesis to the expectation over $\mathcal{H}_{n-1}$. The idea is to observe that J_{n-1} can be seen as a random set that corresponds to applying the swap rounding procedure to the point $\mathbf{z} = \sum_{i=1}^{n-1} \frac{\alpha_i}{\beta_{n-1}} \mathbf{1}_{I_i}$, which is a convex combination of $n-1$ sets, and thus allows us to apply the inductive hypothesis. The only difference, when $I_1, \ldots, I_{n-1}$ are merged into J_{n-1} while applying the rounding procedure to $\mathbf{x}$ compared to rounding $\mathbf{z}$, is that the coefficients of the terms in the convex decomposition of $\mathbf{z}$ are scaled by a factor of $\frac{1}{\beta_{n-1}}$ compared to those of $\mathbf{x}$. The same is true for all coefficients $\tilde{\alpha}_i$ and $\tilde{\beta}_i$ used in the merging procedure. However, this does not make any difference in the rounding procedure, since the rounding probabilities depend only on relative ratios of the coefficients. Thus, we obtain by induction

$$\begin{aligned}\mathbf{E}_{\mathcal{H}_{n-1}}&\left[e^{\lambda'\left(a(J_{n-1}) - \sum_{i=1}^{n-1}\frac{\tilde{\alpha}_i}{\beta_{n-1}}a(I_i)\right)}\right] \\ &\leq e^{2\lambda'^2\sum_{i=1}^{n-1}\frac{\tilde{\alpha}_i}{\beta_{n-1}}a(I_i)} \\ &= e^{2\lambda^2(1+\lambda\alpha_n)^2\beta_{n-1}\sum_{i=1}^{n-1}\tilde{\alpha}_i a(I_i)}.\end{aligned}$$

Combining the above equation with (4.4), we get

$$\begin{aligned}\mathbf{E}[e^{\lambda(a(J_n)-\mu)}] \leq\ & e^{\lambda^2\alpha_n\left(\beta_{n-1}a(I_n) + \sum_{i=1}^{n-1}\tilde{\alpha}_i a(I_i)\right)} \\ & \times e^{2\lambda^2(1+\lambda\alpha_n)^2\beta_{n-1}\sum_{i=1}^{n-1}\tilde{\alpha}_i a(I_i)} \\ = e^{\lambda^2\alpha_n\beta_{n-1}a(I_n)}\ & \times e^{\lambda^2(\alpha_n + 2(1+\lambda\alpha_n)^2\beta_{n-1})\sum_{i=1}^{n-1}\tilde{\alpha}_i a(I_i)}.\end{aligned}$$

Taking logs to simplify the writing, we obtain

(4.5)

$$\log \mathbf{E}\left[e^{\lambda(a(J_n)-\mu)}\right] \le \lambda^2 \underbrace{\alpha_n\beta_{n-1}a(I_n)}_{\substack{\le 2\tilde{\alpha}_n a(I_n) \\ =2\mu-2\sum_{i=1}^{n-1}\tilde{\alpha}_i a(I_i)}}$$

$$+\lambda^2(\underbrace{\alpha_n+2\beta_{n-1}}_{\substack{=\alpha_n+2(1-\alpha_n)\\ =2-\alpha_n}}+(4\lambda\alpha_n+2\lambda^2\alpha_n^2)\beta_{n-1})\sum_{i=1}^{n-1}\tilde{\alpha}_i a(I_i)$$

$$\le \lambda^2\left(2\mu+\alpha_n(-1+(4\lambda+2\lambda^2\alpha_n)\beta_{n-1})\sum_{i=1}^{n-1}\tilde{\alpha}_i a(I_i)\right)$$

where we used $\alpha_n \le (1+\delta)^2\tilde{\alpha}_n \le 2\tilde{\alpha}_n$ (recall that $\delta = \frac{1}{p-1} \le \frac{1}{3}$). The lemma will finally be proven by showing that $(4\lambda+2\lambda^2\alpha_n)\beta_{n-1} \le 1$ for $\lambda \in [-\frac{1}{4}, \frac{1}{4}]$. Since the left-hand side is a convex function of λ, it suffices to check the two endpoints. For $\lambda = -\frac{1}{4}$, the LHS is $(-1+\frac{1}{8}\alpha_n)\beta_{n-1} \le 0$. For $\lambda = \frac{1}{4}$, we get $(1+\frac{1}{8}\alpha_n)\beta_{n-1} \le (1+\alpha_n)(1-\alpha_n) \le 1$. By (4.5), $\log \mathbf{E}\left[e^{\lambda(a(J_n)-\mu)}\right] \le 2\lambda^2\mu$. ■

Now we can finish the proof of Theorem 4.1.

Proof. [Proof of Theorem 4.1]

i) For $\lambda \in [-\frac{1}{4}, 0]$, we have

$$\begin{aligned}\Pr[a(R) \le (1-\varepsilon)\mu] &\le \Pr[e^{\lambda(a(R)-\mu)} \ge e^{-\lambda\varepsilon\mu}] \\ &\le \frac{\mathbf{E}[e^{\lambda(a(R)-\mu)}]}{e^{-\lambda\varepsilon\mu}} \\ &\le e^{(2\lambda^2+\lambda\varepsilon)\mu},\end{aligned}$$

where the second inequality follows from Markov's inequality and the third one follows from Lemma 4.5. Notice that J_n as defined in Lemma 4.5 equals R. Choosing $\lambda = -\varepsilon/4$ proves the claim.

ii) For $\lambda \in [0, \frac{1}{4}]$ we have,

$$\begin{aligned}\Pr[a(R) \ge (1+\varepsilon)\mu] &\le \Pr[e^{\lambda(a(R)-\mu)} \ge e^{\lambda\varepsilon\mu}] \\ &\le \frac{\mathbf{E}[e^{\lambda(a(R)-\mu)}]}{e^{\lambda\varepsilon\mu}} \\ &\le e^{(2\lambda^2-\lambda\varepsilon)\mu}\end{aligned}$$

where again the second inequality follows Markov's inequality and the third one from Lemma 4.5. Choosing $\lambda = \varepsilon/4$ proves the second part of Theorem 4.1.

iii) For the third part of Theorem 4.1, we consider $\varepsilon = t-1 \ge 1$ and we fix $\lambda = \frac{1}{4}$. As above, we obtain

$$\Pr[a(R) \ge t\mu] \le e^{(2\lambda^2-\lambda(t-1))\mu} = e^{(3-2t)\mu/8}.$$

Recall that we scaled the coefficients a_i and hence μ by $1/p$, so the theorem follows. ■

5 Applications

In this section, we show how our concentration bounds imply the results on multi-budgeted and multi-objective optimization. In fact, we prove a more general result, which implies both Theorem 1.2 and Theorem 1.4. First, let us define the following.

Multi-objective/multi-budget matching. *Given a graph $G = (V, E)$, k linear functions ("demands") $f_1, \dots, f_k : 2^E \to \mathbb{R}_+$, and ℓ linear functions ("budgets") $g_1, \dots, g_\ell : 2^E \to \mathbb{R}_+$, is there a matching M satisfying $f_i(M) \ge V_i$ for all $i \in [k]$ and $g_i(M) \le B_i$ for all $i \in [\ell]$?*

Multi-objective/multi-budget matroid intersection. *Given two matroids $\mathcal{M}_1 = (N, \mathcal{I}_1)$ and $\mathcal{M}_2 = (N, \mathcal{I}_2)$, k linear functions ("demands") $f_1, \dots, f_k : 2^N \to \mathbb{R}_+$, and ℓ linear functions ("budgets") $g_1, \dots, g_\ell : 2^N \to \mathbb{R}_+$, is there an independent set $I \in \mathcal{I}_1 \cap \mathcal{I}_2$ satisfying $f_i(I) \ge D_i$ for all $i \in [k]$ and $g_i(I) \le B_i$ for all $i \in [\ell]$?*

We show that these problems can be solved up to a small relative error in the objective functions.

THEOREM 5.1. *For any $\varepsilon > 0$ and any constant number of demands and budgets $k + \ell$, there is a polynomial-time randomized algorithm which for any feasible instance of multi-objective/multi-budget matching or matroid intersection finds with high probability a feasible solution S such that*

- *Each linear budget constraint is satisfied: $g_i(S) \le B_i$.*
- *Each linear demand is nearly satisfied: $f_i(S) \ge (1-\varepsilon)D_i$.*

If such a solution is not found, the algorithm returns a certificate that the instance is not feasible with demands D_i and budgets B_i.

Implications. First, let us consider instances with one objective $k = 1$ and a constant number of budgets ℓ. Via the algorithm in Theorem 5.1, we can perform binary search on the objective function and estimate within a factor of $1-\varepsilon$ the maximum value D_1 such that there is a solution of value $f_1(S) \ge (1-\varepsilon)D_1$ and satisfying $g_i(S) \le B_i$ for all budgets. This gives a (randomized) PTAS for the multi-budgeted versions of matching and matroid intersection, hereby proving Theorem 1.2.

Now, let us consider the case of k objective functions and no budget constraints ($\ell = 0$). Here, we can solve the following promise problem in polynomial time: is there a feasible solution S satisfying $f_i(S) \ge (1-\varepsilon)D_i$ for all i, or no solution satisfies $f_i(S) \ge D_i$ for all i. Using the multi-objective optimization framework of Papadimitriou and Yannakakis (see Theorem 2 in [26]), this is sufficient to find in polynomial time a ε-approximate Pareto set with respect to the ℓ objectives. This proves Theorem 1.4.

Proof. [Sketch of proof of Theorem 5.1.] Fix $\varepsilon > 0$ and let $\gamma = \varepsilon/2$. We guess a constant (depending on $k, \ell, 1/\varepsilon$) number of elements so that for each remaining element j, the value in each f_i is at most $\varepsilon^4 D_i$ and also its size with respect to each budget is at most $\varepsilon^4 B_i$. In the following, we just assume that $f_i(j) \leq \varepsilon^4 D_i$ and $g_i(j) \leq \varepsilon^4 B_i$ for all i and $j \in N$.

Let P be either the matching polytope or the matroid intersection polytope, corresponding to the instance. We add linear constraints to P, to obtain

$$P' = \{\mathbf{x} \in P : \forall i \in [k]; f_i(\mathbf{x}) \geq D_i, \forall j \in [\ell]; g_j(\mathbf{x}) \leq B_i\}.$$

Since P has a polynomial-time separation oracle, we also have a polynomial-time separation oracle for P'; therefore, via the ellipsoid method, we can determine in polynomial time whether P' is empty or not. If P' is empty, we have a certificate that the problem infeasible. Otherwise, we can find a (fractional) solution $\mathbf{y} \in P'$.

We apply randomized swap rounding to $\mathbf{y}$, to obtain a random solution R. This random solution satisfies $\mathbf{E}[\mathbf{1}_R] = (1-\gamma)\mathbf{y} = \mathbf{y}'$ and the concentration bounds of Theorem 1.1 hold. For each budget constraint g_i, the coefficients are in $[0, \varepsilon^4 D_i]$, and $g_i(\mathbf{y}') \leq (1-\gamma)B_i$. Let $\mu = (1-\gamma)B_i$ and apply Theorem 1.1 to the function $g_i(R)/(\varepsilon^4 B_i)$. We obtain

$$\begin{aligned}
\Pr[g_i(R) > B_i] &\leq \Pr[g_i(R) > (1+\gamma)\mu] \\
&< e^{-\mu\gamma^3/(20\varepsilon^4 B_j)} \\
&= e^{-(1-\gamma)\gamma^3/(20\varepsilon^4)}.
\end{aligned}$$

We chose $\gamma = \varepsilon/2$ and so the probability is at most $e^{-(2-\varepsilon)/(320\varepsilon)}$.

Similarly, for each objective function f_i, we have $f_i(\mathbf{y}') \geq (1-\gamma)D_i = \mu$ and the coefficients of f_i are in $[0, \varepsilon^4 D_i]$. Theorem 1.1 implies

$$\begin{aligned}
\Pr[f_i(R) < (1-2\gamma)D_i] &< e^{-\mu\gamma^3/(20\varepsilon^4 D_i)} \\
&= e^{-(1-\gamma)\gamma^3/(20\varepsilon^4)} \\
&= e^{-(2-\varepsilon)/(320\varepsilon)}.
\end{aligned}$$

For fixed k and ℓ one can choose a sufficiently small $\varepsilon(k,\ell) > 0$, such that, by the union bound, the probability that $f_i(R) < (1-2\gamma)D_i$ or $g_i(R) > B_i$ for some i is at most $1/2$. Therefore, we find a solution as required with probability at least $1/2$, and this probability can be boosted by standard techniques. ∎

A bicriteria approximation for many budgets: So far we considered multi-budgeted optimization with a fixed number of budgets. Now we examine the case where the number of budgets can be large. One can formulate this problem as $\max\{\mathbf{w}^T\mathbf{x} \mid A\mathbf{x} \leq \mathbf{1}, \mathbf{x} \in P, \mathbf{x} \in \mathbb{Z}_+^n\}$ where $A\mathbf{x} \leq \mathbf{1}$ is a set of m packing/budget constraints (entries of A are non-negative and lie in $[0,1]$ without loss of generality) and P is an integer polytope of interest such as matroid intersection or matching in our case. If m is part of the input, then the maximum independent set problem is a special case (even with only the packing constraints); unless $P = NP$ there is no $O(n^{1-\varepsilon})$-factor approximation for any fixed $\varepsilon > 0$. However, if the packing constraints are loose or if they can be violated, our rounding procedure yields a good approximation. We can obtain the following results easily via our concentration bounds, and hence omit a formal proof.

- A $(1-\varepsilon)$-approximation if constraints are "loose", that is, $\max_{i,j} A_{ij} \leq c \cdot \varepsilon / \log m$ for some sufficiently small but fixed constant c.
- A $(1-\varepsilon, O(\frac{1}{\varepsilon}\log m))$ bicriteria approximation where the constraints are violated by an $O(\frac{1}{\varepsilon}\log m)$ additive/multiplicative factor.

References

[1] S. Arora, A. Frieze and H. Kaplan. A new rounding procedure for the assignment problem with applications to dense graph assignment problems. *Math. Programming Ser A*, 92: 1–36, 2002. Preliminary version in *Proc. of IEEE FOCS*, 1996.

[2] A. Asadpour, M. Goemans, A. Madry, S. Oveis Gharan, and A. Saberi. An $O(\log n/\log\log n)$-approximation algorithm for the assymetric traveling salesman problem. *Proc. of ACM-SIAM SODA*, 379–289, 2010.

[3] A. Asadpour and A. Saberi. An approximation algorithm for max-min fair allocation of indivisible goods. *SIAM J. Comput.*, 39(7): 2970–2989, 2010. Preliminary version in *Proc. of ACM STOC*, 2007.

[4] 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.

[5] A. Berger, V. Bonifaci, F. Grandoni and G. Schäfer. Budgeted matching and budgeted matroid intersection via the gasoline puzzle. *Math. Programming A*, 2009. Preliminary version in *Proc. of IPCO*, 2008.

[6] J. Byrka, A. Srinivasan, and C. Swamy. Fault-tolerant facility location: a randomized dependent LP-rounding algorithm. *Proc. of IPCO*, 244–257, 2010.

[7] 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*.

[8] P. M. Camerini, G. Galbiati, and F. Maffioli. Random pseudo-polynomial algorithms for exact matroid problems. *J. of algorithms* 13, 258–273, 1992.

[9] C. Chekuri, J. Vondrák and R. Zenklusen. Dependent randomized rounding via exchange properties of combinatorial structures. *Proc. of IEEE FOCS*, 2010.

[10] C. Chekuri, J. Vondrák and R. Zenklusen. Matroid polytopes and negative correlation. Manuscript, 2010.

[11] B. Doerr. Randomly rounding rationals with cardinality constraints and derandomizations. *Proc. of STACS*, 441–452, 2007.

[12] J. Edmonds. Matroids, submodular functions and certain polyhedra. *Combinatorial Structures and Their Applications*, Gordon and Breach, New York, 1970, 69–87.

[13] J. Edmonds. Matroid intersection. *Discrete optimization I*, North Holland, Amsterdam, 1979, 39–49.

[14] R. Gandhi, S. Khuller, S. Parthasarathy and A. Srinivasan. Dependent rounding and its applications to approximation algorithms. *JACM*, 53:324–360, 2006.

[15] F. Grandoni, R. Ravi and M. Singh. Iterative rounding for multi-objective optimization problems. *Proc. of ESA*, 95–106, 2009.

[16] F. Grandoni and R. Zenklusen. Approximation Schemes for Multi-Budgeted Independence Systems. *Proc. of ESA*, 536–548, 2010.

[17] T. Hagerup. A guided tour of Chernoff Bounds *Information Processing Letters* 33(6):305-308, 1990.

[18] A. Kulik, H. Shachnai and T. Tamir. Maximizing submodular set functions subject to multiple linear constraints. *Proc. of ACM-SIAM SODA*, 545–554, 2009.

[19] V. S. A. Kumar, M. V. Marathe, S. Parthasarathy and A. Srinivasan. A unified approach to scheduling on unrelated parallel machines. *JACM*, Vol. 56(5), 2009.

[20] E. L. Lawler. Optimal matroid intersections. *Combinatorial structures and their applications*, Gordon and Breach, New York, 1970, 233–234.

[21] E. L. Lawler. Matroid intersection algorithms. *Mathematical programming* 9, 31–56, 1975.

[22] J. Lee, V. Mirrokni, V. Nagarajan and M. Sviridenko. Maximizing non-monotone submodular functions under matroid and knapsack constraints. *SIAM J. Discrete Math.*, 23(4): 2053–2078, 2010. Preliminary version in *Proc. of ACM STOC*, 2009.

[23] J. Lee, M. Sviridenko, and J. Vondrák. Submodular maximization over multiple matroids via generalized exchange properties. *Math. of OR*, to appear. Preliminary version in *Proc. of APPROX*, 244–257, 2009.

[24] K. Mulmuley, U. V. Vazirani, and V. V. Vazirani. Matching is as easy as matrix inversion. *Combinatorica*, 7(1):101–104, 1987.

[25] 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.

[26] C. H. Papadimitriou and M. Yannakakis. On the approximability of trade-offs and optimal access of web sources. *Proc. of IEEE FOCS*, 86–92, 2000.

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

[28] R. Ravi and M. Goemans. The constrained minimum spanning tree problem. *Proc. of SWAT*, 66–75, 1996.

[29] B. Saha and A. Srinivasan. A new approximation technique for resource-allocation problems. *Proc. of ICS*, 342–357, 2010.

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

[31] A. Srinivasan. Distributions on level-sets with applications to approximation algorithms, *Proc. of IEEE FOCS*, 588–597, 2001.

A A deterministic PTAS for multi-budgeted matching

In this section, we consider the multi-budgeted matching problem, and show how an adapted version of our rounding scheme can be derandomized in this setting to obtain a deterministic PTAS.

The main idea is to use slightly different swap sets when merging a linear combination $\alpha_1\mathbf{1}_{I_1}+\alpha_2\mathbf{1}_{I_2}$ of two matchings I_1, I_2. Assume $\alpha_1 \geq \alpha_2$. In our randomized procedure, $I_1 \Delta I_2$ was cut into swap sets whose size was bounded by a constant. This was done to obtain concentration for any linear function with coefficients in $[0,1]$. However, when concentration is needed only for a constant number of linear functions–in this case the objective function and a constant number of length functions–a much coarser partition of $I_1 \Delta I_2$ into swap sets suffices. In particular we will show that it is possible to partition $I_1 \Delta I_2$ into swap paths, whose lengths may vary and is not necessarily bounded by a constant, such that only a constant number of swaps have to be performed to I_1 to obtain a good merged solution, i.e., a matching J such that the objective value as well as the lengths of $(\alpha_1+\alpha_2)\mathbf{1}_J$ and $\alpha_1\mathbf{1}_{I_1}+\alpha_2\mathbf{1}_{I_2}$ are very close to each other in a well-defined sense. A set of good swaps can then be guessed to derandomized the approach. The swap paths are chosen depending on the objective function and the given length functions. Due to this dependent choice of swap sets, concentration for other linear functions does not necessarily hold anymore. This is a major difference compared to the randomized approach.

More formally, we are given an edge-weighted graph $G=(V,E)$ with a weight function $\mathrm{w}: E \rightarrow \mathbb{Z}_+$. Additionally, a constant number of ℓ non-negative length functions $g_i : E \rightarrow \mathbb{Z}_+, i \in [\ell] := \{1,\ldots,\ell\}$ are given. With every function g_i, a budget $B_i \in \mathbb{Z}_+$ is associated. For convenience, we often consider the weights w and lengths g_i as vectors in $\mathbb{Z}_+^E$, in which case we use the boldface notation $\mathbf{w}$ and $\mathbf{g}_i$. Let $\mathcal{M} \subseteq 2^E$ be the set of all matchings in G, and let $P_{\mathcal{M}}$ be the matching polytope, i.e., the convex hull of $\mathcal{M}$. We are interested in the following multi-budgeted maximum matching problem.

$$\text{(A.1)} \qquad \max\{\mathrm{w}(I) \mid I \in \mathcal{M},\ g_i(I) \leq B_i\ \forall i \in [\ell]\}$$

In the following, we present a PTAS for problem (A.1). More precisely, we present a multi-criteria PTAS for problem (A.1), i.e., an algorithm running in polynomial time for a constant $\varepsilon \in (0,1]$ that returns a matching of weight at least $(1-\varepsilon)OPT$ and violates each budget by a factor of at most $1+\varepsilon$. Such an algorithm can then be transformed into a PTAS for problem (A.1) using standard techniques (see for example [16]).

In a first step we formulate an adapted version of our randomized procedure which is easier to derandomize. For the derandomization step we use the following standard Chernoff bound for Poisson trials. A proof can be found in [17][3].

THEOREM A.1. *Let $X_1,\dots,X_n$ be independent random variables taking values in $[0,1]$, let $X=\sum_{i=1}^n X_i$ and let $\delta\in[0,1]$. Then*
i) for $\mu\geq \mathbf{E}[X]$, $\Pr[X\geq(1+\delta)\mu]\leq e^{-\mu\delta^2/3}$,
ii) for $\mu\leq \mathbf{E}[X]$, $\Pr[X\leq(1-\delta)\mu]\leq e^{-\mu\delta^2/2}$.

A.1 Merging two matchings In this section we show how a convex combination $x=\alpha_1\mathbf{1}_{I_1}+\alpha_2\mathbf{1}_{I_2}$ of two matchings $I_1,I_2\in\mathcal{M}$ can be transformed into a matching $J\in\mathcal{M}$ with a similar objective value and such that a constant number of linear functions do not increase too much. We use the following notion of an *r-almost* matching, which was introduced in a slightly more general form in [16]. A set $I\subseteq E$ is called an *r-almost* matching for some $r\in\mathbb{Z}_+$, if I can be transformed into a matching by removing at most r elements of I.

Let $g_i^{\max}=\max_{e\in I_1\Delta I_2} g_i(e)$ for $i\in[\ell]$, $\mathrm{w}^{\max}=\max_{e\in I_1\Delta I_2}\mathrm{w}(e)$, and $N=\frac{6}{\varepsilon^2}\ln(2\ell)$. The following theorem is the backbone of the PTAS to be presented. Its proof shows how to define swap paths such that at the same time we have concentration for the objective function and budgets, and efficient derandomization through exhaustive search is possible.

THEOREM A.2. *Let $\mathbf{x}=\alpha_1\mathbf{1}_{I_1}+\alpha_2\mathbf{1}_{I_2}$ be a convex combination of two matchings. If $\mathbf{g}_i^T\mathbf{x}\geq Ng_i^{\max}$ for $i\in[\ell]$ and $\mathbf{w}^T\mathbf{x}\geq N\mathrm{w}^{\max}$, then a matching J can be obtained in time $O(n^{3(1+\varepsilon)(\ell+1)N})$ such that $I_1\cap I_2\subseteq J\subseteq I_1\cup I_2$, and*
(i) $g_i(J)\leq(1+\varepsilon)\mathbf{g}_i^T x \quad \forall i\in[\ell]$,
(ii) $\mathrm{w}(J)\geq(1-\varepsilon)\mathbf{w}^T x-6(1+\varepsilon)(\ell+1)N\mathrm{w}_i^{\max}$.

Proof. The theorem will be proven by presenting first a randomized algorithm returning with high probability a $6(1+\varepsilon)(\ell+1)N$-almost matching $\widetilde{J}$ that can be transformed into a matching J satisfying the claim of the theorem. The algorithm will then be derandomized to get the result. Without loss of generality we assume $\alpha_1\geq\alpha_2$, and $I_1\cap I_2=\emptyset$. Let $D=I_1\Delta I_2=I_1\cup I_2$. We number the edges in $D=\{e_1,\dots,e_{|D|}\}$ such that edges belonging to the same cycle/path in D are numbered consecutively, and any two edges e_j,e_{j+1} for $j\in[|D|-1]$ are either adjacent in D or belong to different paths/cycles of D. This can easily be achieved by cutting each cycle of D at an arbitrary vertex and appending the resulting set of paths one to the other by gluing together the endpoints of the paths.

Consider a fixed $i\in[\ell]$, and let $T_i=\mathbf{g}_i^T\mathbf{x}/N$. We partition the edges of D into a collection $\mathcal{P}_i=\{P_i^1,\dots,P_i^{k_i}\}$ such that $g_i(P)\leq T_i\ \forall P\in\mathcal{P}$ as follows. The sets $P_i^1,\dots,P_i^{k_i}$ are constructed iteratively: $P_i^1=\{e_1,\dots,e_{\gamma_1}\}$ consists of the largest sequence of edges $e_1,\dots,e_{\gamma_1}$, such that $\ell_i(P_i^1)\leq T_i$, i.e., $\gamma_1=\max\{h\in[|D|]\mid\sum_{j=1}^h g_i(e_j)\leq T_i\}$. Analogously, P_i^2 consists of the largest sequence of edges $e_{\gamma_1+1},\dots e_{\gamma_2}$ such that $\ell_i(P_i^2)\leq T_i$ and so on.

Next, we show that the number k_i of sets in $\mathcal{P}_i$ satisfies $k_i\leq 3N/\alpha_2$. Notice that by construction of $\mathcal{P}_i$, we have $g_i(P_i^j\cup P_i^{j+1})>T_i$ for $j\in[k_i-1]$. Using this fact, one can easily observe that the average length of the sets in $\mathcal{P}_i$ is at least $T_i/3$. Hence,

$$g_i(D)=\sum_{j=1}^{k_i} g_i(P_i^j)\geq\frac{k_iT_i}{3}. \tag{A.2}$$

Furthermore, we have $\mathbf{g}_i^T\mathbf{x}=\mathbf{g}_i^T(\alpha_1\mathbf{1}_{I_1}+\alpha_2\mathbf{1}_{I_2})$ and $g_i(D)=\mathbf{g}_i^T(\mathbf{1}_{I_1}+\mathbf{1}_{I_2})$, and hence by non-negativity of g_i and since $\alpha_2\leq\alpha_1$, we get $g_i(D)\leq\frac{\mathbf{g}_i^T\mathbf{x}}{\alpha_2}=\frac{NT_i}{\alpha_2}$. Combining this result with (A.2) we obtain $k_i\leq 3N/\alpha_2$ as claimed.

For every $i\in[\ell]$, let $\mathcal{P}_i$ be a partition of D as described above. Similarly, let $\mathcal{P}_{\mathrm{w}}$ be a partition of D constructed analogously to the other ones with respect to the weight function w instead of a length function. Each partition $\mathcal{P}_1,\dots,\mathcal{P}_\ell,\mathcal{P}_{\mathrm{w}}$ represents a way of cutting the sequence of edges $(e_1,\dots,e_{|D|})$ into subsequences. Let $\mathcal{P}$ be the partition corresponding to cutting the sequence $(e_1,\dots,e_{|D|})$ at all places where at least one of the partitions $\mathcal{P}_1,\dots,\mathcal{P}_\ell,\mathcal{P}_{\mathrm{w}}$ cuts the sequence. Hence, $\mathcal{P}$ can be described as follows

$$\mathcal{P}=\{P_{\mathrm{w}}\cap P_1\cap\cdots\cap P_\ell\mid P_i\in\mathcal{P}_i \text{ for } i\in[\ell], P_{\mathrm{w}}\in\mathcal{P}_{\mathrm{w}}\}.$$

Since every partition of $\mathcal{P}_1,\dots,\mathcal{P}_\ell,\mathcal{P}_{\mathrm{w}}$ contains at most $3N/q$ subsequences of $(e_1,\dots,e_{|D|})$, the partition $\mathcal{P}$ contains at most $3N(\ell+1)/q$ elements.

We consider the following random process to create a merge $\widetilde{J}$ of I_1 and I_2. We start with I_1 and, independently for every set $P\in\mathcal{P}$, we perform an edge-flip on the set P with probability α_2. In the following we show that with high probability, the random set $\widetilde{J}$ satisfies i) $g_i(\widetilde{J})\leq\mathbf{g}_i^T\mathbf{x}$ for $i\in[\ell]$, and ii) $\mathrm{w}(\widetilde{J})\geq(1-\varepsilon)\mathbf{w}^T\mathbf{x}$.

Let $i\in[\ell]$. The length $g_i(\widetilde{J})$ can be written as a sum of independent random variables

$$g_i(\widetilde{J})=\sum_{P\in\mathcal{P}} X_P, \tag{A.3}$$

with $\Pr[X_P=g_i(I_1\cap P)]=\alpha_1$ and $\Pr[X_P=g_i(I_2\cap P)]=\alpha_2$. Notice that by construction of $\mathcal{P}$, we have $X_P\in[0,\mathbf{g}_i^T\mathbf{x}/N]$. Furthermore, $\mathbf{E}[g_i(\widetilde{J})]=\alpha_1 g_i(I_1)+$

[3]In [17], a proof is presented for Theorem A.1 for the case $\mu=\mathbf{E}[X]$. However, the same proof also shows the slightly more general statement given by Theorem A.1.

$\alpha_2 g_i(I_2) = \mathbf{g}_i^T\mathbf{x}$. Hence,

(A.4)
$$\Pr[g_i(\widetilde{J}) \geq (1+\varepsilon)\mathbf{g}_i^T\mathbf{x}] = \Pr\left[\frac{g_i(\widetilde{J})N}{\mathbf{g}_i^T\mathbf{x}} \geq (1+\varepsilon)N\right] \leq e^{-N\varepsilon^2/3} = \frac{1}{4\ell^2},$$

where the inequality follows by the Chernoff bound of Theorem (A.1) point (i) with $\mu = N$. Analogously, we can obtain the following bound for the weight function using Theorem A.1 point (ii), i.e.,

$$\text{(A.5)} \qquad \Pr[\mathrm{w}_i(\widetilde{J}) \leq (1-\varepsilon)\mathbf{w}^T\mathbf{x}] \leq e^{-N\varepsilon^2/2} \leq \frac{1}{4\ell^2}.$$

We now show that with high probability the set $\widetilde{J}$ is an $6(1+\varepsilon)(\ell+1)N$-almost matching, i.e., it can be transformed into a matching J by removing from $\widetilde{J}$ at most $6(1+\varepsilon)(\ell+1)N$ edges. The number of edges that have to be removed from $\widetilde{J}$ to obtain a matching can easily be bounded as follows. Let $\mathcal{P}' \subseteq \mathcal{P}$ be the paths that were used to obtain $\widetilde{J}$ from I_1 by flips, i.e., $\widetilde{J} = I_1\Delta(\cup_{P\in\mathcal{P}'}P)$. Clearly, $\widetilde{J}$ can be transformed into a matching by removing all edges at the two ends of all paths in $\mathcal{P}'$. Hence, $\widetilde{J}$ is a $2|\mathcal{P}'|$-matching. To show that $\widetilde{J}$ is with high probability a $6(1+\varepsilon)(\ell+1)$ matching as claimed, we will show that with high probability $|\mathcal{P}'| \leq 3(1+\varepsilon)(\ell+1)$.

Notice that $|\mathcal{P}'|$ is the sum of independent 0-1 random variables Y_P for $P \in \mathcal{P}$, with $P[Y_P = 1] = \alpha_2$. Furthermore, since the number of sets in $\mathcal{P}$ is bounded by $3(\ell+1)N/\alpha_2$, and each set in $\mathcal{P}$ is used to perform an edge-flip with probability α_2, we have that the expected number of flips is bounded by $3(\ell+1)N$. Again, using the Chernoff bound presented in Theorem A.1 point (i), we obtain

(A.6)
$$\Pr[\sum_{P\in\mathcal{P}} Y_P \geq (1+\varepsilon)3N(\ell+1)] \leq e^{-3N(\ell+1)\varepsilon^2/3} \leq e^{N\varepsilon^2/3} = \frac{1}{4\ell^2}.$$

Hence with probability at least $1 - \frac{1}{4\ell^2}$, we have $|\mathcal{P}'| \leq 3N(1+\varepsilon)(\ell+1)$.

Using (A.4), (A.5) and (A.6) we can apply a union bound to obtain that the probability of Y satisfying simultaneously

- (i) $g_i(\widetilde{J}) \leq (1+\varepsilon)\mathbf{g}_i^T\mathbf{x} \quad \forall i \in [\ell]$,
- (ii) $\mathrm{w}(\widetilde{J}) \geq (1-\varepsilon)\mathbf{w}^T\mathbf{x}$ and
- (iii) $\widetilde{J}$ is obtained from I_1 by at most $3N(1+\varepsilon)(\ell+1)$ edge-flips,

is lower-bounded by $1 - (\ell\frac{1}{4\ell^2} + 2\frac{1}{4\ell^2}) > 0$. The algorithm to find a set $\widetilde{J}$ satisfying the above conditions can easily be derandomized as follows. Since we know that at most $3N(1+\varepsilon)(\ell+1)$ sets of the partition $\mathcal{P}$ have to be used for edge-flips to obtain $\widetilde{J}$, we can guess these sets in $O(|\mathcal{P}|^{3N(1+\varepsilon)(\ell+1)})$ time, and since trivially $|\mathcal{P}| \leq n$, where $n = |V|$, this computational complexity is bounded by $O(n^{3N(1+\varepsilon)(\ell+1)})$.

By removing from $\widetilde{J}$ all pending edges $e_{\gamma_j}, e_{\gamma_{j+1}}$ of the sets $P = \{e_{\gamma_j}, \ldots, e_{\gamma_{j+1}}\} \in \mathcal{P}$ that were used to perform edge-flips to obtain $\widetilde{J}$, a matching J is obtained. Notice that $\mathrm{w}(J) \geq (1-\varepsilon)\mathbf{w}^T\mathbf{x} - 6(1+\varepsilon)(\ell+1)N\mathrm{w}_i^{\max}$ since $\mathrm{w}(\widetilde{J}) \geq (1-\varepsilon)\mathbf{w}^T\mathbf{x}$ and at most $6(1+\varepsilon)(\ell+1)N$ edges of $D\cap\widetilde{J}$ have to be removed from $\widetilde{J}$ to obtain J. Furthermore, because all lengths are positive we have $g_i(J) \leq g_i(\widetilde{J}) \leq (1+\varepsilon)\mathbf{g}_i^T\mathbf{x}$. ■

A.2 A PTAS for multi-criteria matching Let $\varepsilon \in (0,1]$ be the desired accuracy, and we set $\varepsilon' = \ln(2)\varepsilon/2\ell$ and let $N' = \frac{6}{\varepsilon'^2}\ln(2\ell)$. In a first step of our algorithm, for each $i \in [\ell]$, we guess the N' longest edges $E_i \subseteq E$ with respect to g_i in an optimal solution. Furthermore, we guess the $12(1+\frac{1}{\varepsilon'})\ell(\ell+1)N'$ heaviest edges $E_{\mathrm{w}} \subseteq E$ with respect to w. For a fixed ε, there is only a constant number of edges that have to be guessed, which can be done in polynomial time.

Let $E_i^{\uparrow}$ be the set of all edges $e \in E \setminus E_i$ such that $g_i(e) > \min\{g_i(f) \mid f \in E_i\}$. Since E_i should contain the longest edges of an optimal solution with respect to g_i, no edges in $E_i^{\uparrow}$ will be added to the solution at a later stage. Similarly we define $E_{\mathrm{w}}^{\uparrow} = \{e \in E \setminus E_{\mathrm{w}} \mid \mathrm{w}(e) > \min\{\mathrm{w}(f) \mid f \in E_{\mathrm{w}}\}\}$. Let $F = E_{\mathrm{w}} \cup \bigcup_{i=1}^{\ell} E_i$ be the set of all guessed edges, and let $F^{\uparrow} = E_{\mathrm{w}}^{\uparrow} \cup \bigcup_{i=1}^{\ell} E_i^{\uparrow}$. Notice that the case $F \cap F^{\uparrow} \neq \emptyset$ corresponds to an inconsistent guess. Hence, such a guess can be discarded. Problem (A.1) can then be reduced accordingly, leading to the following problem.

$$\text{(A.7)} \qquad \max\{\mathrm{w}(I) \mid I \in \mathcal{M}, F \subseteq I, I \cap F^{\uparrow} = \emptyset, g_i(I) \leq B_i \ \forall i \in [\ell]\}$$

Let $\mathbf{x}^*$ be an optimal vertex-solution of the following LP relaxation of problem (A.7).

(A.8)
$$\max\{\mathbf{w}^T\mathbf{x} \mid \mathbf{x} \in P_{\mathcal{M}}, \mathbf{x}(e) = 1 \ \forall e \in F, \mathbf{x}(e) = 0 \ \forall e \in F^{\uparrow}, \mathbf{g}_i^T\mathbf{x} \leq B_i \ \forall i \in [\ell]\}$$

Notice that problem (A.8) corresponds to optimizing a linear function on the face of the matching polytope $P_{\mathcal{M}}$ defined by $\mathbf{x}(e) = 1 \ \forall e \in F$ and $\mathbf{x}(e) = 0 \ \forall e \in F^{\uparrow}$, with the ℓ additional linear constraints $\mathbf{g}_i^T\mathbf{x} \leq B_i$ for $i \in [\ell]$. Hence, $\mathbf{x}^*$ lies on a face of $P_{\mathcal{M}}$ of dimension at most ℓ. Thus, $\mathbf{x}^*$ can be written as a convex combination of at most $\ell+1$ matchings. Furthermore, such a decomposition can be found in polynomial time using a constructive version of Carathéodory's theorem. Hence, let $\mathbf{x}^* = \sum_{j=1}^{\ell+1}\alpha_j\mathbf{1}_{I_j}$, where $I_j \in \mathcal{M}, \alpha_j \geq 0 \ \forall j \in [\ell+1]$ and $\sum_{j=1}^{\ell+1}\alpha_j = 1$. The algorithm then iteratively proceeds as our randomized scheme, where for merging $\alpha_1\mathbf{1}_{I_1}+\alpha_2\mathbf{1}_{I_2}$ we apply Theorem A.2 with $\varepsilon=\varepsilon'$ to the convex combination $\frac{1}{\alpha_1+\alpha_2}(\alpha_1\mathbf{1}_{I_1}+\alpha_2\mathbf{1}_{I_2})$. As in our randomized rounding scheme, let $\beta_j=\sum_{s=1}^{j}\alpha_s$ for $j\in[\ell+1]$. Furthermore, let $J_1=I_1$, and for $j\in\{2,\dots,\ell+1\}$ we denote by J_j the set obtained by merging $\beta_{j-1}\mathbf{1}_{J_{j-1}}+\alpha_j\mathbf{1}_{I_j}$. After ℓ steps, a matching $J=J_{\ell+1}$ is obtained and returned by the algorithm. Let $\mathbf{z}_j=\beta_j\mathbf{1}_{J_j}+\sum_{s=j+1}^{\ell+1}\alpha_s\mathbf{1}_{I_j}$ for $j\in[\ell+1]$ be the fractional matching after $j-1$ merging steps.

By Theorem A.2, the algorithm has polynomial running time for a fixed ε. Notice, that the conditions of Theorem A.2 are always satisfied because of the following. Every matching used throughout the algorithm contains the guessed edges F. Furthermore, the guessed edges satisfy $g_i(F)\geq N'g_i^{max}$ for $i\in[\ell]$ and $\mathrm{w}(F)\geq N'\mathrm{w}^{\max}$, where g_i^{max} is the largest ith length among the non-fixed edges $E\setminus(F\cup F^{\uparrow})$, and $\mathrm{w}_i^{\max}$ is the largest weight among the non-fixed edges $E\setminus(F\cup F^{\uparrow})$.

It remains to show that J is a $(1\pm\varepsilon)$ multi-criteria solution to problem (A.1). We first consider a length function g_i for some fixed $i\in[\ell]$, and show by induction on j that $\mathbf{g}_i^T\mathbf{z}_j\leq(1+\varepsilon')^{j-1}\mathbf{g}_i^T\mathbf{x}^*$ for $j\in[\ell+1]$. For $j=1$ the result trivially holds. For $j\in\{2,\dots,\ell+1\}$ we have

$$\begin{aligned}\mathbf{g}_i^T\mathbf{z}_j&=\beta_j g_i(J_j)+\sum_{s=j+1}^{\ell+1}\alpha_s g_i(I_s)\\&\leq(1+\varepsilon')\left(\beta_{j-1}g_i(J_{j-1})+\alpha_j g_i(I_j)\right)\\&\qquad+\sum_{s=j+1}^{\ell+1}\alpha_s g_i(I_s)\\&\leq(1+\varepsilon')\left(\beta_{j-1}g_i(J_{j-1})+\sum_{s=j}^{\ell+1}\alpha_s g_i(I_s))\right)\\&\leq(1+\varepsilon')^j\mathbf{g}_i^T\mathbf{x}^*,\end{aligned}$$

where the first inequality follows by Theorem A.2 point (i) when applied to the merge operation that merged J_{j-1} and I_j to obtain J_j, and the last inequality follows by the inductive hypothesis. Thus we obtain

$$\begin{aligned}g_i(J)=\mathbf{g}_i^T\mathbf{z}_{\ell+1}&\leq(1+\varepsilon')^\ell\mathbf{g}_i^T\mathbf{x}^*=\left(1+\frac{\ln(2)\varepsilon}{2\ell}\right)^\ell\mathbf{g}_i^T\mathbf{x}^*\\&\leq e^{\ln(2)\frac{\varepsilon}{2}}\mathbf{g}_i^T\mathbf{x}^*=2^{\frac{\varepsilon}{2}}\mathbf{g}_i^T\mathbf{x}^*\leq\left(1+\frac{\varepsilon}{2}\right)\mathbf{g}_i^T\mathbf{x}^*\\&\leq\left(1+\frac{\varepsilon}{2}\right)B_i.\end{aligned}$$

Analogously, one can prove

$$\mathrm{w}(J)\geq\left(1-\frac{\varepsilon}{2}\right)\mathbf{w}^T\mathbf{x}^*-6(1+\varepsilon')\ell(\ell+1)N'\mathrm{w}^{\max}. \tag{A.9}$$

Since

$$\mathbf{w}^T\mathbf{x}^*\geq\mathrm{w}(F)\geq|F|\mathrm{w}^{\max}=12\left(1+\frac{1}{\varepsilon'}\right)\ell(\ell+1)N'\mathrm{w}^{\max},$$

we have $6(1+\varepsilon')\ell(\ell+1)N'\mathrm{w}^{\max}\leq\frac{\varepsilon'}{2}\mathbf{w}^T\mathbf{x}^*\leq\frac{\varepsilon}{2}\mathbf{w}^T\mathbf{x}^*$, and hence by (A.9) we obtain $\mathrm{w}(J)\geq(1-\varepsilon)\mathbf{w}^T\mathbf{x}^*$ as desired.