

# Entropy Regularization and Faster Decremental Matching in General Graphs<sup>\*</sup>

Jiale Chen<sup>†</sup>

Aaron Sidford<sup>‡</sup>

Ta-Wei Tu<sup>§</sup>

## Abstract

We provide an algorithm that maintains, against an adaptive adversary, a  $(1 - \varepsilon)$ -approximate maximum matching in  $n$ -node  $m$ -edge general (not necessarily bipartite) undirected graph undergoing edge deletions with high probability with (amortized)  $O(\text{poly}(\varepsilon^{-1}, \log n))$  time per update. We also obtain the same update time for maintaining a fractional approximate weighted matching (and hence an approximation to the value of the maximum weight matching) and an integral approximate weighted matching in dense graphs.<sup>†</sup> Our unweighted result improves upon the prior state-of-the-art which includes a  $\text{poly}(\log n) \cdot 2^{O(1/\varepsilon^2)}$  update time [Assadi–Bernstein–Dudeja 2022] and an  $O(\sqrt{m}\varepsilon^{-2})$  update time [Gupta–Peng 2013], and our weighted result improves upon the  $O(\sqrt{m}\varepsilon^{-O(1/\varepsilon)} \log n)$  update time due to [Gupta–Peng 2013].

To obtain our results, we generalize a recent optimization approach to dynamic algorithms from [Jambulapati–Jin–Sidford–Tian 2022]. We show that repeatedly solving entropy-regularized optimization problems yields a lazy updating scheme for fractional decremental problems with a near-optimal number of updates. To apply this framework we develop optimization methods compatible with it and new dynamic rounding algorithms for the matching polytope.

## 1 Introduction

*Dynamic matching* is a fundamental and incredibly well-studied problem in data structure design and dynamic graph algorithms. In this problem, there is a sequence of modifications, e.g., edge insertion and deletions, to an undirected graph  $G = (V, E)$  with vertices  $V$  and edges  $E$ . The goal is to maintain a *matching*  $M \subseteq E$ , i.e., a subset of edges that are pairwise disjoint, of approximately maximum size, i.e.,  $|M| \geq (1 - \varepsilon)M^*(G)$  where  $\varepsilon$  is a specified accuracy parameter and  $M^*(G)$  is the size of the maximum matching in the current graph,  $G$ . In this paper we consider solving this problem with the goal of obtaining improved (amortized) update time per operation.<sup>‡</sup>

The dynamic matching problem is also notoriously challenging to solve efficiently. In the general setting of maintaining such a  $(1 - \varepsilon)$ -approximate maximum matching in a general undirected graph undergoing edge insertions and deletions, the state-of-the-art includes an  $O(\sqrt{m}\varepsilon^{-2})$  update time algorithm of [41] for all  $\varepsilon$  and an  $n/(\log^* n)^{\Omega(1)}$  update time algorithm of [2] for  $\varepsilon = \Omega((\log^* n)^{-c})$  where  $c > 0$  is some small constant. Subsequent to the initial release of our paper, the work of [6], building on top of [15], obtained an  $O(n^{o(1)} \cdot \text{ORS}(n, \Theta_\varepsilon(n)))$  update time algorithm for constant  $\varepsilon$ , where  $\text{ORS}(n, r)$  is the maximum possible density of the so-called *ordered Ruzsa-Szemerédi Graphs*: the current, state-of-the-art bound is  $n^{o(1)} \leq \text{ORS}(n, \Theta_\varepsilon(n)) \leq n^{1-o(1)}$  for constant  $\varepsilon$  [6]. Additional improvements are only known in special cases. For example, when there are vertex (instead of edge) updates to a bipartite graph then [30, 70] provide  $\tilde{O}(\varepsilon^{-1})$  time<sup>§</sup> algorithms. Additionally, recent breakthrough

<sup>\*</sup>The full version of the paper can be accessed at <https://arxiv.org/abs/2312.09077>

<sup>†</sup>Stanford University, [jialec@stanford.edu](mailto:jialec@stanford.edu). Supported by a Lawrence Tang Graduate Fellowship, a Microsoft Research Faculty Fellowship, and NSF CAREER Award CCF-1844855.

<sup>‡</sup>Stanford University, [sidford@stanford.edu](mailto:sidford@stanford.edu). Supported in part by a Microsoft Research Faculty Fellowship, NSF CAREER Award CCF-1844855, NSF Grant CCF-1955039, a PayPal research award, and a Sloan Research Fellowship.

<sup>§</sup>Stanford University, [taweitu@stanford.edu](mailto:taweitu@stanford.edu). Supported by a Stanford School of Engineering Fellowship, a Microsoft Research Faculty Fellowship, and NSF CAREER Award CCF-1844855.

<sup>1</sup>Independently and concurrently, Aditi Dudeja obtained new decremental weighted matching results for general graphs [37].

<sup>2</sup>Our new algorithms periodically rebuild a fractional matching when its value degrades. Consequently, their update times are (naïvely) amortized instead of worst-case. We leave de-amortizing our results as an open problem.

<sup>3</sup>In this paper, we use  $\tilde{O}(\cdot)$  to hide  $\text{poly}(\log(n), \log(\varepsilon^{-1}))$  factors and  $\hat{O}(\cdot)$  to hide subpolynomial,  $(n\varepsilon^{-1})^{o(1)}$  factors in the  $O(\cdot)$  notation.  $\tilde{\Omega}(\cdot)$ ,  $\hat{\Omega}(\cdot)$ ,  $\tilde{\Theta}(\cdot)$ , and  $\hat{\Theta}(\cdot)$  are defined analogously.

results maintain better-than-2 approximations to the size of the maximum matching (rather than the matching itself) in near-optimal update time [13, 28, 3] and better approximations in sub-linear update time [25].

Towards explaining the challenging nature of dynamic matching, recent work established conditional lower bounds on the problem's complexity. To maintain exactly the maximum matching, there are conditional lower bounds on the update times of  $\Omega(\sqrt{m})$  [9, 46, 52] and  $\Omega(n^{1.407})$  [68]. To maintain approximate maximum matchings, very recently, [53] ruled out the possibility of a truly sublinear,  $\varepsilon^{-O(1)}n^{1-O(1)}$  update time algorithm in the fully dynamic setting, assuming a new dynamic approximate OMv conjecture. Additionally, assuming only the standard OMv conjecture, [53] ruled out the possibility of truly sublinear update time algorithms for the closely related problem of maintaining a  $(1 + \varepsilon)$ -approximate vertex cover.

Corresponding to the difficulty of dynamic matching, recent work studied the *decremental setting* where starting from an initial graph only edge deletions are permitted until the graph is deleted. Excitingly, [17] provided an  $\tilde{O}(\varepsilon^{-4})$  time per update algorithm for this problem. [49] then obtained update times of  $\tilde{O}(\varepsilon^{-3})$  and  $\hat{O}(\varepsilon^{-2})$  for the problem of maintaining fractional matchings. [29] then obtained update times of  $\tilde{O}(\varepsilon^{-3})$  and  $\hat{O}(\varepsilon^{-2})$  for maintaining integral matchings through improved dynamic rounding algorithms for fractional matchings. Recently, [1] obtained an  $\tilde{O}(1) \cdot 2^{O(1/\varepsilon^2)}$  update time for general (not necessarily bipartite) graphs.

Strikingly, these results show that decremental matching can be solved near-optimally in *bipartite graphs* for  $\varepsilon = 1/\text{poly}(\log n)$ . However, in *general graphs*, for such  $\varepsilon$ , the state-of-the-art is still  $O(\sqrt{m}\varepsilon^{-2})$  time [41]. A central question motivating this work is whether it is possible to narrow this gap and develop improved decremental matching algorithms for general graphs. One of our main results is an affirmative answer to this question; we provide  $\tilde{O}(\varepsilon^{-41})$  and  $\hat{O}(\varepsilon^{-17})$  update time algorithms for maintaining a  $(1 - \varepsilon)$ -approximate matching in a general graph undergoing edge deletions. Our algorithm succeeds with high probability (w.h.p.<sup>4</sup>) against an adaptive adversary that has access to the internal randomness used by our algorithm and can design its future updates based on that. This is the first decremental dynamic matching algorithm for general graphs that achieves subpolynomial accuracy and subpolynomial update time simultaneously.<sup>5</sup>

We obtain our results through the development of a general framework that also allows us to obtain improved runtimes in the decremental setting of the more challenging dynamic *weighted* matching as well. In this problem we must maintain a  $(1 - \varepsilon)$ -approximate maximum weight matching given polynomially bounded edge weights  $\mathbf{w} \in \mathbb{N}^E$ . Prior to our work, the state-of-the-art was an  $\tilde{O}(\sqrt{m}\varepsilon^{-O(1/\varepsilon)})$  time [41] and an  $\tilde{O}(\varepsilon^{-3})$  time per update algorithm for maintaining fractional matchings in bipartite graphs [26] (which also applied to a broader class of partially-dynamic packing/covering linear programs). In contrast, we provide algorithms which decrementally maintain  $(1 - \varepsilon)$ -approximate fractional matchings in general weighted graphs in  $\tilde{O}(\varepsilon^{-41})$  and  $\hat{O}(\varepsilon^{-17})$  time per update and integral  $(1 - \varepsilon)$ -approximate matchings in dense graphs in the same update times. We also provide an  $\tilde{O}(\varepsilon^{-O(1/\varepsilon)})$  update time integral matching algorithm, which is not limited to dense graphs, using a weight reduction framework of [41]. These are the first near-optimal partially dynamic algorithms for weighted matchings with edge updates.

## 1.1 Approach

**Lazy Updates and Congestion Balancing.** Our algorithms follow a natural, time-tested *lazy* approach to solving dynamic matching problems [41, 31, 32, 17, 49, 1]. Broadly, we compute an approximate fractional matching, delete edges from it as needed, and then, when the updates cause the solution to change, we *rebuild*, computing a new fractional matching. By efficiently computing the fractional matching, computing fractional matchings that limit the number of rebuilds, and efficiently rounding, we obtain our dynamic matching algorithms.

More specifically, our algorithms follow a template common to [17, 49, 1] (see Section 3.1 for a more precise description). In this framework, we first, compute a  $(1 - \delta)$ -approximate fractional matching. Then, when edges are deleted, the corresponding fractional assignment on those edges is removed as well. Once the value of the fractional matching decreases by  $(1 - \delta)$  multiplicatively, a new  $(1 - \delta)$ -approximate fractional matching is computed and the process is repeated; we call each computation of a  $(1 - \delta)$ -approximate fractional matching a *rebuild*. Any algorithm following this *lazy update framework* clearly maintains a  $(1 - \delta)^2$ -approximate fractional

<sup>4</sup>We use *w.h.p.* in this paper as an abbreviation of *with high probability*, meaning the success probability can be made at least  $1 - n^{-c}$  for any constant  $c > 0$  by increasing the runtime by a constant factor.

<sup>5</sup>More precisely, when  $\varepsilon = \Theta(1/\text{poly}(\log(n)))$ , or even  $\varepsilon = \Theta(1/n^{o(1)})$ , previous algorithms require  $\text{poly}(n)$  or even  $\exp(n)$  update times, while our algorithms achieve an  $n^{o(1)}$  update time.

matching and consequently, by picking, e.g.,  $\delta = \varepsilon/2$ , this algorithm maintains a  $(1 - \varepsilon)$ -approximate fractional solution. What is perhaps unclear, is how to make this approach efficient.

The update time of an algorithm following the lazy update framework is governed by:<sup>6</sup>

1. the number of rebuilds, i.e., the number of approximate fractional matchings computed,
2. the cost per rebuilding, i.e., the cost of computing each fractional matching, and
3. the cost of rounding, i.e., the cost of turning these dynamically maintained fractional matchings into dynamically maintained integral matchings.

Note that computing a  $(1 - \varepsilon)$ -approximate matching takes  $\tilde{\Omega}(m)$  time in the worst case [27]. Consequently, if each rebuild is computed from scratch (as they are in our algorithms) then the cost of each rebuild is  $\Omega(m)$  (Item 2) and to obtain an  $O(\text{poly}(\log n, \varepsilon^{-1}))$ -update time it must be that the number of rebuilds (Item 1) is  $O(\text{poly}(\log n, \varepsilon^{-1}))$ . However, it is unclear, just from the approach, whether or why this should be obtainable.

Nevertheless, in a striking result, [17] showed that it was possible to follow this framework and obtain  $\tilde{O}(\varepsilon^{-4})$  time per update for bipartite graphs. The algorithm had  $\tilde{O}(\varepsilon^{-3})$  rebuilds (Item 1) at a cost of  $\tilde{O}(m\varepsilon^{-1})$  time per rebuild<sup>7</sup> (Item 2) for a total runtime of  $\tilde{O}(m\varepsilon^{-4})$  and, therefore, an update time of  $\tilde{O}(\varepsilon^{-4})$  for maintaining fractional matchings. The technique they used to construct the fractional matchings they call *congestion balancing*. Furthermore, [1] generalized this approach to non-bipartite graphs. Their algorithm had  $\tilde{O}(2^{O(1/\varepsilon^2)})$  rebuilds (Item 1) at a cost of  $\tilde{O}(m\varepsilon^{-1})$  per rebuild (Item 2) for a total runtime of  $\tilde{O}(m \cdot 2^{O(1/\varepsilon^2)})$  and, therefore, an update time of  $\tilde{O}(2^{O(1/\varepsilon^2)})$  for maintaining fractional matchings. To obtain integral matchings, both papers applied known dynamic rounding algorithms [69] to solve (Item 3).

**Entropy Regularization and Weighted Matching.** A key question motivating our result is, *how powerful and general is the lazy update framework for decremental problems?* Recent work of [49] opened the door to studying this question. This work showed that, for bipartite graphs, to bound the number of rebuilds (Item 1) it sufficed to set the fractional matchings to be sufficiently accurate solutions to natural, regularized optimization problems. The current state-of-the-art decremental bipartite matching algorithms follow this framework [49, 29].

In this paper, we provide a broad generalization of this result. We show that for any non-negative, non-degenerate, compact, downward closed, and convex  $\mathcal{X} \subseteq \mathbb{R}_{\geq 0}^d$ , to maintain approximate maximizers of  $\mathbf{w}^\top \mathbf{x}$  for  $\mathbf{x} \in \mathcal{X}$  under deletions of coordinates to  $\mathcal{X}$  one can apply the lazy update framework and rebuild only  $\tilde{O}(\varepsilon^{-2})$  times! Furthermore, we show that this rebuild count is obtained so long as the algorithm solves certain entropy-regularized versions of the problem, i.e., finding an  $\mathbf{x} \in \mathcal{X}$  approximately maximizing  $\mathbf{w}^\top \mathbf{x} + \mu \cdot H(\mathbf{x})$  for some trade-off parameter  $\mu$ , where  $H(\mathbf{x})$  is an appropriately weighted and scaled measure of the entropy of  $\mathbf{x}$ .

This result, when combined with [33] and [29], recovers the  $\tilde{O}(\varepsilon^{-2})$  update time of [29, 49] and enables our main results on decremental matching in general graphs. Letting  $\mathcal{X} = \mathcal{M}_G$  be the non-bipartite matching polytope of the input graph  $G$ , this result implies that to obtain a fractional solution we simply need to repeatedly solve an entropy-regularized matching problem. Additionally, our entropy regularization result immediately implies the same for maintaining a weighted fractional matching! This leads to improved algorithms even for bipartite graphs, where the previous best algorithm has an  $\tilde{O}(\varepsilon^{-3})$  update time [28].

However, to obtain efficient decremental fractional algorithms for general graphs, we still need to efficiently solve these entropy-regularized problems over the matching polytope (Item 2). In Section 5 we show how to solve these entropy-regularized problems to  $(1 - \delta)$  accuracy in  $\tilde{O}(m \cdot \delta^{-13})$  and  $\tilde{O}(m \cdot \delta^{-5})$  time for all  $\delta \geq \tilde{\Omega}(n^{-1/2})$ . To maintain a  $(1 - \varepsilon)$ -approximate matching, it suffices to set  $\delta = 1/\text{poly}(\varepsilon^{-1}, \log n)$ . We obtain this by modularizing and generalizing the framework of [5] for capacitated weighted general  $\mathbf{b}$ -matching and provide two different instantiations of the framework. In one approach we apply the recent convex flow algorithm of [33], leading to the running time of  $\tilde{O}(m \cdot \delta^{-5})$ . In the other one we reduce entropy regularization to capacity-constrained weighted general matching, leading to the other running time of  $\tilde{O}(m \cdot \delta^{-13})$ .

<sup>6</sup>Additionally, the algorithm must remove fractional assignments to the deleted edges from the fractional matching, but that is trivial to implement in  $O(1)$  per update.

<sup>7</sup>The  $\tilde{O}(m\varepsilon^{-1})$  runtimes stems from running a push-relabel-style flow algorithm to find an approximate flow. Using recent almost linear time maximum flow algorithms [33, 66], this can be improved to be  $\tilde{O}(m)$ , leading to an  $\tilde{O}(\varepsilon^{-3})$  update time algorithm for maintaining fractional matchings. See [17, Lemma 5.2].

**Dynamic Rounding.** To turn our dynamic fractional matching results into dynamic integral matching results we develop two unweighted rounding algorithms for general graphs, one randomized and the other deterministic. For the randomized algorithm, we analyze the standard sampling procedure for rounding in bipartite graphs and argue that, excitingly, it can also be used in general graphs. On the other hand, our deterministic algorithm is based on a recently dynamized pipage rounding procedure [29]. However, we differ from [29] by stopping their algorithm earlier and running static algorithms on the result periodically. Although the randomized rounding algorithm does not have a strictly better update time than the deterministic one, it has stronger guarantees that we leverage to round weighted matchings. We remark that while [1] also required rounding, the fractional matching they maintained has a special property (in particular it is  $\text{poly}(\varepsilon)$ -restricted [32, Definition 1]) that allows them to argue the bipartite rounding algorithm of [69] directly applies to the fractional matching they found. In contrast, our rounding algorithms generically work for *any* fractional matching in general graphs, and are the first to achieve such a guarantee. (See Section 1.3 for a more detailed comparison of rounding algorithms.)

**Optimality of Entropy Regularization.** Given the utility of entropy regularization and the lazy update framework, we ask, *can we further decrease the number of rebuilds (Item 1)?* Interestingly, we show that this is not the case for decremental matching, even on unweighted bipartite graphs. We show that for any  $n \geq 1$  and  $\varepsilon = \Omega(n^{-1/2})$ , there is an adversarial choice of the initial graph with  $n$  vertices and a sequence of deletions such that, regardless of what fractional matching the algorithm maintains, there are at least  $\Omega(\log^2(\varepsilon^2 n) \cdot \varepsilon^{-2})$  rebuilds. This shows the optimality of entropy regularization in both  $\log n$  and  $\varepsilon^{-1}$  factors for  $\varepsilon = \Omega(n^{-1/2+\delta})$  for constant  $\delta > 0$ .

**Summary.** We obtain improved results for decremental matching in general graphs by building new tools to follow the lazy update framework: we prove a general bound on the number of rebuilds when using entropy regularization (Item 1), we develop efficient algorithms for solving entropy-regularized optimization problems over the matching polytope (Item 2), and we develop new dynamic rounding algorithms for fractional matchings in general graphs (Item 3). We think that each tool could be of independent interest. Additionally, given the generality and optimality (in terms of the number of rebuilds) of our entropy regularization approach to decremental problems, we hope that our results may open the door to new dynamic algorithms in broader settings.

**Paper Organization.** In the remainder of this introduction we present our results in Section 1.2 and compare and survey previous work in Section 1.3. We then cover preliminaries in Section 2 and provide a technical overview of our approach in Section 3. In Section 4 we show the robustness and optimality of entropy regularization for the decremental linear optimization problem. We then turn our attention to the special case of decremental matching in non-bipartite graphs, presenting our algorithms for solving entropy-regularized matchings in Section 5 and our rounding algorithms in Section 6. In Sections A and B we provide additional proofs that are included for completeness.

**1.2 Results** In this paper we consider the *decremental matching problem* formally defined as follows. Note that in this definition, when  $W = 1$ , the problem is the aforementioned *unweighted* matching problem in which the maintained matching is simply a  $(1 - \varepsilon)$ -approximate maximum (cardinality) matching.

**PROBLEM 1.1. (DECREMENTAL MATCHING)** *In the decremental matching problem, we are given an  $n$ -vertex  $m$ -edge graph  $G = (V, E)$ , integer<sup>8</sup> edge weights  $\mathbf{w} \in \mathbb{N}^E$  bounded by  $W = \text{poly}(n)$ , and an accuracy parameter  $\varepsilon \in (0, 1)$ . The goal is to maintain a  $(1 - \varepsilon)$ -approximate maximum weight matching  $M \subseteq E$  at all times under deletions to  $E$  until  $G$  becomes empty.*

We develop a variety of randomized algorithms to solve Problem 1.1. Due to the use of randomization, it is important to distinguish between different kinds of update sequences that they can support. We say a dynamic algorithm is *output-adaptive* (respectively, *fully-adaptive*) if it works for update sequences that are chosen adaptively based on the output (respectively, internal randomness) of the algorithm. Note that a fully-adaptive

<sup>8</sup>Note that the assumption that  $\mathbf{w}$  is integral is without loss of generality as it can be achieved by scaling. For instance, we can first make the minimum weight 1, and then scale each entry to the nearest value of  $\lceil (1 + O(\varepsilon))^i \rceil$  since we are only aiming for an approximate solution.

algorithm is automatically output-adaptive. These are in contrast to the *oblivious* algorithms which only work for updated sequences that are fixed in advance<sup>9</sup>.

We now describe our main results. Recall that in the unweighted case, the previous state-of-the-art algorithms that solve [Problem 1.1](#) are an  $\tilde{O}(1) \cdot 2^{O(1/\varepsilon^2)}$  update time algorithm of [\[1\]](#) and an  $O(\sqrt{m}\varepsilon^{-2})$  update time algorithm of [\[41\]](#).

**THEOREM 1.1. (UNWEIGHTED DECREMENTAL MATCHING)** *There are randomized fully-adaptive  $\tilde{O}(\varepsilon^{-41})$  and  $\hat{O}(\varepsilon^{-17})$  update time algorithms that solve [Problem 1.1](#) in the unweighted case w.h.p.*

For the weighted case, the algorithm of [\[1\]](#) does not apply, and the previous state-of-the-art, even for bipartite graphs, is an  $\tilde{O}(\sqrt{m}\varepsilon^{-O(1/\varepsilon)})$  update time algorithm, also by [\[41\]](#). Our algorithms have near-optimal update time when either  $\varepsilon$  is a constant or the input graph is dense.

**THEOREM 1.2. (WEIGHTED DECREMENTAL MATCHING)** *There is a randomized fully-adaptive  $\tilde{O}(\varepsilon^{-O(1/\varepsilon)})$  update time algorithm that solves [Problem 1.1](#) w.h.p. Additionally, when  $m = \hat{\Theta}(n^2)$ , there are randomized output-adaptive  $\tilde{O}(\varepsilon^{-41})$  and  $\hat{O}(\varepsilon^{-17})$  update time algorithms.*

For dense bipartite graphs we obtain an even better update time.

**THEOREM 1.3. (WEIGHTED BIPARTITE DECREMENTAL MATCHING)** *For bipartite graphs with  $m = \hat{\Theta}(n^2)$  there is an output-adaptive randomized  $\hat{O}(\varepsilon^{-6})$  update time algorithm that solves [Problem 1.1](#) w.h.p.*

[Theorem 1.1](#) to [1.3](#) are all obtained by solving the intermediate problem of decremental fractional matching ([Problem 3.2](#)) and then converting the fractional results to integral via rounding algorithms. Following the entropy regularization approach in the lazy update framework, the fractional matching algorithms we develop achieve an upper bound of  $\tilde{O}(\varepsilon^{-2})$  on the number of rebuilds ([Item 1](#)) in [Theorem 3.1](#). We investigate whether this bound on the number of rebuilds can be improved. Interestingly, we show that this is not possible and that entropy regularization is a near-optimal strategy for the lazy update framework in the decremental setting for any  $n$  and  $\varepsilon = \Omega(n^{-1/2})$ . This lower bound holds even in the simple case of unweighted bipartite matching.

**THEOREM 1.4. (INFORMAL, SEE [THEOREM 3.2](#))** *For any  $n \in \mathbb{N}$ ,  $\varepsilon \geq 2/\sqrt{n}$ , and any output-adaptive algorithm implementing the rebuilding subroutine in the lazy update framework, there exists an unweighted bipartite graph of  $n$  vertices on each side such that an output-adaptively chosen sequence of edge deletions causes  $\Omega(\log^2(\varepsilon^2 n) \cdot \varepsilon^{-2})$  rebuilds.*

**1.3 Related Work** Here we give a more extensive summary of previous work related to our results in this paper.

**General Matching.** Due to the existence of odd cycles and blossoms, matching problems in general graphs are often considerably more challenging than in bipartite graphs. Starting from the blossom algorithm of Edmonds [\[39\]](#), with additional ideas and techniques, several works culminated in general matching algorithms that are equally efficient as classic bipartite algorithms, both in the exact [\[58, 44, 36\]](#) and approximate [\[58, 35, 5\]](#) settings (we make particular use of [\[5\]](#) in our dynamic algorithms). Yet, it is still open whether modern optimization-based algorithms for bipartite matching [\[55, 54, 56, 67, 51, 8, 33\]](#) can lead to runtime improvements for computing matchings in general graphs.

**Incremental Matching.** Another studied partially-dynamic matching problem is the *incremental matching problem*, where instead of edge deletions, there are only edge insertions. In this setting, near-optimal results are known for obtaining  $(1 - \varepsilon)$ -approximate matchings, achieving  $\tilde{O}(\text{poly}(\varepsilon^{-1}))$  and recently  $O(\text{poly}(\varepsilon^{-1}))$  update times in bipartite graphs [\[45, 24\]](#) and  $\varepsilon^{-O(1/\varepsilon)}$  update time in general graphs [\[40\]](#). For bipartite graphs, an  $\tilde{O}(\text{poly}(\varepsilon^{-1}))$  update time is also obtainable by a more general partially dynamic packing/covering LP algorithm [\[28\]](#).

<sup>9</sup>It is also common in the dynamic algorithm literature to model the adaptiveness of an algorithm in the *adversarial* setting in which the algorithm is working *against* an *adversary* designing the update sequence either on-the-fly (in which case it is an *adaptive* adversary) or in advance (an *oblivious* adversary). We also remark that the term *adaptive* is widely used in the literature but has mixed meanings and can refer to either output-adaptive or fully-adaptive based on the context. Consequently, in this paper we make an explicit distinction between the two notions.



**Additional Matching Results.** Aside from  $(1 - \varepsilon)$ -approximations, dynamic matching with other approximation ratios has been studied, particularly in the fully dynamic setting. Notably, for  $1/2$ -approximate, maximal matching, a line of work culminated into optimal, constant-update-time algorithms [60, 16, 18, 64, 20, 10, 21, 11]. The work of [31] introduced the notion of edge-degree constrained subgraphs (EDCS) and with this initiated a line of work on non-trivial algorithms to maintain  $(2/3 - \varepsilon)$  approximate matchings [32, 42] and beyond [23]. Improvements in other directions such as derandomization, de-amortization, and frameworks converting unweighted results to the weighted case have also been studied [19, 65, 34, 21, 14, 12, 22, 50, 62, 29]. Dynamic rounding algorithms which reduce dynamic integral matching algorithms to dynamic fractional matching algorithms are also well-studied in various regimes [4, 69, 22, 50, 29].

**Entropy Regularization.** Even in prior dynamic matching results that do not explicitly use entropy regularization, entropy does play a role implicitly. More precisely, [45, 26, 70] all used the multiplicative weight update (MWU) method, which can be viewed as an iterative method for optimization method which uses an entropy regularizer to determine the steps it makes.

**Comparison to [1].** The  $\tilde{O}(1) \cdot 2^{O(1/\varepsilon^2)}$  update time algorithm of [1] used the congestion balancing approach introduced in [17] to implement rebuilds in the lazy update framework. Informally, in congestion balancing, a capacity constraint  $c \in [0, 1]^E$  is maintained, and in each iteration the goal is to find a matching  $M_c$  respecting this capacity constraint of size comparable to the actual maximum matching or, in that case that such a matching does not exist, find a set of edges whose capacity constraint is “critical” to  $M_c$  being small. In the first case,  $M_c$  is used as the output fractional matching until future deletions decrease its value significantly. In the second case, the capacities of these critical edges are increased to accommodate larger matchings. The exponential dependence on  $\varepsilon^{-1}$  of [1]’s algorithm stems from the difficulty of solving the above subroutine in general graphs. As we will show later in this paper, there is indeed an  $\tilde{O}(m \cdot \text{poly}(\varepsilon^{-1}))$  algorithm for the capacity-constrained matching problem in general graphs (see Lemma 5.5), but [1] additionally needed a dual certificate (obtainable from, e.g., [35], on uncapacitated graphs) of the matching problem to identify critical edges. Interestingly, obtaining such certificates is still left open by our work.<sup>10</sup>

**Comparison to Previous Randomized Rounding Algorithms.** Our randomized rounding algorithm adopts the same strategy central to previous algorithms [4, 69]. Their algorithms build upon the subgraph, hereafter referred to as a *sparsifier*, obtained by sampling each edge, either independently or dependently, with probability proportional to the fractional mass assigned to it. Informally, [4] showed that the independently sampled sparsifier preserves<sup>11</sup> approximately maximal fractional matchings which are  $(1/2 - \varepsilon)$ -approximate. This holds in general graphs as well. [69], on the other hand, studied fully-adaptive rounding, in which case updating the sparsifier only partially as in [4] no longer works. [69] therefore designed a *dependent* sampling scheme that has marginal the same as the independent one, which turned out to be more efficiently sample-able from scratch. Their analysis also showed that the sparsifier preserves  $(1 - \varepsilon)$ -approximate fractional matchings in *bipartite* graphs. [1] then extended the analysis to show that  $O(\varepsilon)$ -restricted matchings in general graphs are preserved in [69]’s sparsifier as well. Finally, our rounding algorithm goes back and considers the independently sampled sparsifier which can now be used in the (output-)adaptive setting due to the dynamic sampler of [29]. Adopting a more direct analysis, we show that this sparsifier can in fact preserve *arbitrary* fractional matchings in general graphs. Given that the sampling distribution of [69] is, in essence, an easier-to-sample-from but correlated version of the distribution we consider in this paper, we suspect their analysis can be extended to work for any fractional matching in general graphs, perhaps at the cost of a larger runtime. After the initial publication of our manuscript, [38] showed that indeed the rounding algorithm of [69] can be extended to work in general graphs, albeit with a slightly larger update time.

## 2 Preliminaries

**Notation.** We let  $[d] \stackrel{\text{def}}{=} \{1, 2, \dots, d\}$  for  $d \in \mathbb{N}$ ,  $\mathbb{R}_\infty \stackrel{\text{def}}{=} \mathbb{R} \cup \{\infty\}$ , and  $\mathbb{R}_{-\infty} \stackrel{\text{def}}{=} \mathbb{R} \cup \{-\infty\}$ . We use  $\log(\cdot)$  to denote logarithm base 2 and  $\ln(\cdot)$  for natural logarithm. We let  $\llbracket \phi \rrbracket$  be evaluated to 1 if the expression  $\phi$  is true

<sup>10</sup>More precisely, [1] ran the algorithm of [35] only on a carefully sampled subgraph of  $G$ , and thus the dual certificate they obtained and used is different from that of the capacity-constrained matching problem in Problem 5.1. Interestingly, one can show that a dual certificate to Problem 5.1 also suffices to identify critical edges, so this leaves extracting the dual from our algorithm the final step toward speeding up [1].

<sup>11</sup>More precisely the sparsifier contains an integral matching whose size is the same as the fractional one, up to  $(1 - \delta)$  multiplicatively for any  $\delta$  (on which the algorithm’s runtime depends).

and 0 otherwise.

Consider a finite set  $U$  and  $S \subseteq U$ . Let  $\mathbf{0}^S$  and  $\mathbf{1}^S$  be the all-zero and all-one vector in  $\mathbb{R}^S$ , respectively.<sup>12</sup> Let  $\Delta^S$  denote the *simplex* in  $\mathbb{R}^S$ , i.e.,  $\Delta^S \stackrel{\text{def}}{=} \{\mathbf{x} \in \mathbb{R}_{\geq 0}^S \mid \|\mathbf{x}\|_1 = 1\}$ . For  $\mathbf{x} \in \mathbb{R}^U$ , let  $\mathbf{x}_S \in \mathbb{R}^S$  be  $\mathbf{x}$  with coordinates restricted to  $S$ , i.e.,  $(\mathbf{x}_S)_i = \mathbf{x}_i$  for all  $i \in S$ . For  $\mathcal{X} \subseteq \mathbb{R}^U$ , let  $\mathcal{X}_S$  be  $\mathcal{X}$  restricted to coordinates  $S$ , i.e.,  $\mathcal{X}_S \stackrel{\text{def}}{=} \{\mathbf{x}_S : \mathbf{x} \in \mathcal{X}\}$ , and  $\mathcal{X}_{S,+} \stackrel{\text{def}}{=} \{\mathbf{x} \in \mathcal{X}_S : \mathbf{x}_i > 0 \forall i \in S\}$  be the subset of  $\mathcal{X}_S$  in  $\mathbb{R}_{>0}^S$ . For  $\mathbf{x} \in \mathbb{R}^S$  and  $i \in U$ , let  $\mathbf{x} \setminus \{i\} \in \mathbb{R}^{S \setminus \{i\}}$  be  $\mathbf{x}$  excluding coordinate  $i$ , i.e.,  $(\mathbf{x} \setminus \{i\})_j = \mathbf{x}_j$  for all  $j \in S \setminus \{i\}$ . For  $S' \subseteq S$  and  $\mathbf{x} \in \mathbb{R}_{S'}^S$ , let  $\mathbf{x}^S \in \mathbb{R}^S$  be  $\mathbf{x}$  extended to  $\mathbb{R}^S$ , i.e.,  $\mathbf{x}_i^S = \mathbf{x}_i$  for  $i \in S'$  and  $\mathbf{x}_i^S = 0$  for  $i \in S \setminus S'$ .

**Runtimes.** In this paper we use the standard word-RAM model where basic arithmetic operations over  $O(\log n)$ -bit words can be performed in constant time. When the input size  $n$  is clear from context, we say an  $x \in \mathbb{R}$  is *polynomially bounded* if  $|x| \in \{0\} \cup [n^{-O(1)}, n^{O(1)}]$ .

**Graphs.** All graphs in this paper are undirected, simple, and not necessarily bipartite, unless stated otherwise. For a graph  $G = (V, E)$ , let  $E_v \stackrel{\text{def}}{=} \{e \in E : v \in e\}$  be the set of edges incident to  $v \in V$ , and  $E[B] \stackrel{\text{def}}{=} \{e \in E : e \subseteq B\}$  be the set of edges whose endpoints are both in  $B \subseteq V$ . For  $F \subseteq E$  let  $F_v \stackrel{\text{def}}{=} F \cap E_v$  and  $F[B] \stackrel{\text{def}}{=} F \cap E[B]$ . For  $\mathbf{x} \in \mathbb{R}_{\geq 0}^E$ , let  $\mathbf{x}(v) \stackrel{\text{def}}{=} \sum_{e \in E_v} \mathbf{x}_e$  for  $v \in V$  and  $\mathbf{x}(B) \stackrel{\text{def}}{=} \sum_{e \in E[B]} \mathbf{x}_e$  for  $B \subseteq V$ . Let  $\mathbf{B}_G \in \{-1, 0, 1\}^{E \times V}$  be the (*edge vertex*) *incidence matrix* of  $G$ , where there are exactly two non-zero entries per row  $e$ , one at entry  $(e, u)$  with value  $-1$  and the other at entry  $(e, v)$  with value  $1$ , for an arbitrary orientation  $(u, v)$  of  $e = \{u, v\}$ . Let  $M^*(G)$  be the size of the maximum matching in  $G$  and  $M_{\mathbf{w}}^*(G)$  for  $\mathbf{w} \in \mathbb{R}_{\geq 0}^E$  be the value of the maximum weight matching in  $G$  for the weights  $\mathbf{w}$ .

**Matching Polytope.** For an undirected graph  $G = (V, E)$ , the *matching polytope* of  $G$  is the convex hull of the indicator vectors of matchings in  $G$ . Let

$$(2.1) \quad \mathcal{P}_G \stackrel{\text{def}}{=} \left\{ \sum_{e \in E_v} \mathbf{x}_e \leq 1, \quad \forall v \in V \right\} \cap \mathbb{R}_{\geq 0}^E.$$

It is a standard fact that when  $G$  is bipartite,  $\mathcal{P}_G$  is the matching polytope of  $G$ . When  $G$  is non-bipartite, we need to further consider odd-set constraints. Formally, we let

$$\mathcal{O}_G \stackrel{\text{def}}{=} \{B \subseteq V : |B| \geq 3 \text{ and } |B| \text{ is odd}\}$$

be the collections of *odd sets* and define

$$(2.2) \quad \mathcal{M}_G \stackrel{\text{def}}{=} \mathcal{P}_G \cap \left\{ \mathbf{x}(B) \leq \left\lfloor \frac{|B|}{2} \right\rfloor, \quad \forall B \in \mathcal{O}_G \right\}.$$

It is known that  $\mathcal{M}_G$  is the matching polytope of  $G$  [63]. We often consider the relaxation of (2.2) to only contain small odd sets  $\mathcal{O}_{G,\varepsilon} \stackrel{\text{def}}{=} \{B \in \mathcal{O}_G : |B| \leq 1/\varepsilon\}$  denoted by

$$(2.3) \quad \mathcal{M}_{G,\varepsilon} \stackrel{\text{def}}{=} \mathcal{P}_G \cap \left\{ \mathbf{x}(B) \leq \left\lfloor \frac{|B|}{2} \right\rfloor, \quad \forall B \in \mathcal{O}_{G,\varepsilon} \right\}$$

when dealing with  $(1 - \varepsilon)$ -approximation algorithms. The following fact about  $\mathcal{M}_G$  versus  $\mathcal{M}_{G,\varepsilon}$  is folklore and key to our algorithm development.

FACT 2.1. (SEE, E.G., [1]) For  $\varepsilon > 0$  and  $\mathbf{x} \in \mathcal{M}_{G,\varepsilon}$  it holds that  $\frac{\mathbf{x}}{1+3\varepsilon} \in \mathcal{M}_G$ .

We may refer to an  $\mathbf{x} \in \mathcal{M}_G$  as a (*fractional*) *matching* in  $G$  and  $\mathbf{x} \in \mathcal{P}_G$  as a *relaxed (fractional) matching*.<sup>13</sup> When  $\mathbf{x} \in \mathcal{M}_G$  or  $\mathbf{x} \in \mathcal{P}_G$  is clear from context, we may refer to  $\mathbf{x}_e$  as the *mass* the matching  $\mathbf{x}$  puts on edge  $e$ .

**Miscellaneous.** The *recourse* of a dynamic algorithm is the total number of changes it makes to its output. When working with  $(1 - \varepsilon)$ -approximation, in the remainder of the paper we may assume without loss of generality that  $\varepsilon$  is upper-bounded by an explicit constant. This only incurs a constant increase in runtimes.

<sup>12</sup>When  $S$  is clear from context, we may drop the superscript and simply use  $\mathbf{0}$  and  $\mathbf{1}$ .

<sup>13</sup>We remark that an  $\mathbf{x} \in \mathcal{P}_G$  is often referred to as a *fractional matching* even in general graphs in the literature (see, e.g., [4, 29]). However, we deviate from this convention so that there is no integrality gap between fractional and integral matchings.

### 3 Technical Overview

In this section we introduce the problems that we consider in this paper, present the main results and technical tools of each section, and illustrate high-level ideas towards proving them. At the end of this section, we utilize these results and tools to prove our main theorems stated in [Section 1.2](#).

**3.1 Lazy Updates for Decremental Linear Optimization** In this paper we consider a unifying framework of congestion balancing for solving *decremental linear optimization problems* formally defined as follows.

**PROBLEM 3.1. (DECREMENTAL LINEAR OPTIMIZATION)** *In the decremental linear optimization problem we are given a positive weight vector  $\mathbf{w} \in \mathbb{R}_{>0}^d$  and a non-negative, non-degenerate<sup>14</sup>, compact<sup>15</sup>, downward closed<sup>16</sup>, convex  $\mathcal{X} \subseteq \mathbb{R}_{\geq 0}^d$  where  $d > 1$ .<sup>17</sup> Starting from the entire coordinate set  $S = [d]$ , under a sequence of deletions of coordinates from  $S$  we must maintain an  $\mathbf{x} \in \mathcal{X}_S$  such that*

$$\mathbf{w}_S^\top \mathbf{x} \geq (1 - \varepsilon) \max_{\mathbf{x}' \in \mathcal{X}_S} \mathbf{w}_S^\top \mathbf{x}'$$

for a given accuracy  $\varepsilon > 0$  until  $S$  becomes empty.

The framework for solving [Problem 3.1](#) that we study in this paper is a generalization of the lazy update scheme that is widely used for dynamic matching problems as we discussed in [Section 1](#). Specifically, we consider algorithms that maintain an approximate solution  $\mathbf{x} \in \mathcal{X}_S$  and use it as the solution until its value drops by an  $O(\varepsilon)$  fraction at which point we perform a rebuild. The following [Algorithm 1](#) is a template for the lazy update approaches, for which we will later specify what approximate solutions  $\mathbf{x}$  will be used in `Rebuild()` at Line [9](#).

---

#### Algorithm 1: LAZYUPDATE( $\mathcal{X}$ )

---

**global:** weight vector  $\mathbf{w} \in \mathbb{R}_{>0}^d$  and accuracy parameter  $\varepsilon \in (0, 1)$ .  
**global:** current coordinates  $S \subseteq [d]$  and solution  $\mathbf{x} \in \mathcal{X}_S$  with “rebuild” value  $\nu$ .  
**global:** number of rebuilds  $t \in \mathbb{Z}_{\geq 0}$ .  
**global:** snapshots  $\mathbf{x}^{(t)}$  and  $S^{(t)}$  for analysis.

- 1 **function** Initialize( $\mathbf{w} \in \mathbb{R}_{>0}^d, \varepsilon \in (0, 1)$ )
  - 2   Save  $\mathbf{w}$  and  $\varepsilon$  as global variables.
  - 3   Initialize  $S \leftarrow [d]$  and  $t \leftarrow 0$ .
  - 4    $(\mathbf{x}, \nu) \leftarrow \text{Rebuild}()$ .
- 5 **function** Delete( $i \in [d]$ )
  - 6   Set  $\mathbf{x} \leftarrow \mathbf{x} \setminus \{i\}$  and  $S \leftarrow S \setminus \{i\}$ .
  - 7   **if**  $\mathbf{w}_S^\top \mathbf{x} < (1 - \frac{\varepsilon}{2}) \nu$  **then**  $(\mathbf{x}, \nu) \leftarrow \text{Rebuild}()$ .
- 8 **function** Rebuild()
  - 9   // What  $\mathbf{x}^{(t)}$  is computed below depends on the specific algorithm
  - 9   Set  $\mathbf{x}^{(t)}$  to an element in  $\mathcal{X}_S$  with  $\mathbf{w}_S^\top \mathbf{x}^{(t)} \geq (1 - \frac{\varepsilon}{2}) \max_{\mathbf{x}' \in \mathcal{X}_S} \mathbf{w}_S^\top \mathbf{x}'$ .
  - 10   Set  $\nu^{(t)} \leftarrow \mathbf{w}_S^\top \mathbf{x}^{(t)}$  and  $S^{(t)} \leftarrow S$ .
  - 11   **return**  $(\mathbf{x}^{(t)}, \nu^{(t)})$  and set  $t \leftarrow t + 1$ .

---

**OBSERVATION 3.1.** [Algorithm 1](#) solves the decremental linear optimization problem [\(Problem 3.1\)](#).

<sup>14</sup>That is, for each coordinate  $i \in [d]$  there exists an  $\mathbf{x} \in \mathcal{X}$  such that  $\mathbf{x}_i > 0$ . This is a natural assumption since we can always ignore the degenerate dimensions.

<sup>15</sup>Note that this is equivalent to being bounded and closed, which implies  $\max_{\mathbf{x} \in \mathcal{X}} \mathbf{w}^\top \mathbf{x}$  is bounded as well.

<sup>16</sup>That is, for each  $\mathbf{x} \in \mathcal{X}$  and  $\mathbf{y} \in \mathbb{R}_{\geq 0}^d$  with  $\mathbf{y} \leq \mathbf{x}$  entry-wise, we have  $\mathbf{y} \in \mathcal{X}$  also.

<sup>17</sup>The assumption of  $d > 1$  is natural as otherwise the problem degenerates into a 1D optimization and becomes not decremental in essence.



*Proof.* Note that  $\mathbf{x} \in \mathcal{X}_S$  at all times since  $\mathcal{X}$  is downward closed. The vector  $\mathbf{x}^{(t)}$ , when constructed in `Rebuild()`, is an  $(1 - \frac{\varepsilon}{2})$ -approximate solution with value  $\nu^{(t)}$ . Since the `Delete` operations only decrease  $\max_{\mathbf{x}' \in \mathcal{X}_S} \mathbf{w}_S^\top \mathbf{x}'$ , as long as  $\mathbf{w}_S^\top \mathbf{x} \geq (1 - \frac{\varepsilon}{2}) \nu^{(t)}$ , we have

$$\mathbf{w}_S^\top \mathbf{x} \geq \left(1 - \frac{\varepsilon}{2}\right) \nu^{(t)} \geq (1 - \varepsilon) \max_{\mathbf{x}' \in \mathcal{X}_S} \mathbf{w}_S^\top \mathbf{x}'.$$

On the other hand, whenever  $\mathbf{w}_S^\top \mathbf{x} < (1 - \frac{\varepsilon}{2}) \nu^{(t)}$ , we call `Rebuild()`. [Algorithm 1](#) thus maintains a  $(1 - \varepsilon)$ -approximate solution at all times.  $\square$

**3.2 Entropy Regularization** To solve the above decremental linear optimization problem using the lazy update scheme, we apply a variant of the entropy regularization strategy previously used decremental dynamic matching in unweighted bipartite graphs [\[49\]](#). Intuitively, to avoid the adversary from deleting large weight from our solution at once, the idea is to find an  $\mathbf{x} \in \mathcal{X}_S$  with uniformly distributed value on each coordinate. As such, the approach is to use the entropy-regularized solution as  $\mathbf{x}$  in `Rebuild()`, prioritizing vectors with higher entropy when they have similar weights.

To formally describe our results, consider a fixed positive weight  $\mathbf{w} \in \mathbb{R}_{>0}^d$ . For  $S \subseteq [d]$ ,  $\mu, \gamma \in \mathbb{R}_{>0}$ , we define our *entropy-regularized objective*  $f_{S,\gamma}^\mu : \mathcal{X}_S \rightarrow \mathbb{R}$  for all  $\mathbf{x} \in \mathcal{X}_S$  by

$$(3.4) \quad f_{S,\gamma}^\mu(\mathbf{x}) \stackrel{\text{def}}{=} \mathbf{w}_S^\top \mathbf{x} + \mu \cdot \sum_{i \in S} \mathbf{w}_i x_i \log \frac{\gamma}{\mathbf{w}_i x_i}$$

and let

$$(3.5) \quad \mathbf{x}_{S,\gamma}^\mu \stackrel{\text{def}}{=} \operatorname{argmax}_{\mathbf{x} \in \mathcal{X}_S} f_{S,\gamma}^\mu(\mathbf{x})$$

be the optimal solution to [\(3.4\)](#). The main result we show later in [Section 4](#) is that solutions to [\(3.4\)](#) with  $\mu = \tilde{\Theta}(\varepsilon)$  lead to a lazy update scheme with bounded rebuilds.

**LEMMA 3.1.** *For any  $\alpha \geq 0$ , accuracy parameter  $\varepsilon > 0$ , and  $0 < \mu \leq \frac{\varepsilon}{8 \log d}$ , if the subroutine `Rebuild()` in [Algorithm 1](#) returns  $\mathbf{x}_{S^{(t)},\gamma}^\mu$  as  $\mathbf{x}^{(t)}$  for  $\gamma \stackrel{\text{def}}{=} \alpha$ , then `Rebuild()` will be called at most  $O(\frac{\log d}{\mu \cdot \varepsilon})$  times before  $\max_{\mathbf{x}' \in \mathcal{X}_S} \mathbf{w}_S^\top \mathbf{x}'$  drops from at most  $\alpha$  to below  $\alpha/d$ .*

Analogous to [\[49\]](#), to prove [Lemma 3.1](#), we use  $f_{S,\gamma}^\mu(\mathbf{x}_{S,\gamma}^\mu)$ , the optimal value of the entropy-regularized objective on  $\mathcal{X}_S$ , as a potential function to capture the progress that [Algorithm 1](#) makes. Applying the optimality conditions for concave optimization to  $f_{S,\gamma}^\mu(\mathbf{x}_{S,\gamma}^\mu)$  allows us to lower bound the decrease from  $f_{S,\gamma}^\mu(\mathbf{x}_{S,\gamma}^\mu)$  to  $f_{S',\gamma}^\mu(\mathbf{x}_{S',\gamma}^\mu)$  using the Bregman divergence of the entropy regularizer, which has a close relationship to the weighted value of the deleted coordinates in  $S \setminus S'$ . The choice of  $\mu$  and  $\gamma$  is to guarantee that the entropy-regularized solution serves as a valid solution for `Rebuild()`.

In case the exact solution  $\mathbf{x}_{S,\gamma}^\mu$  is computationally expensive to compute, we also show that an accurate enough approximation to it admits the same robustness property, by combining the proof of [Lemma 3.1](#) and the strong concavity of the entropy-regularized objective. In the following, we say an  $\mathbf{x} \in \mathcal{X}_S$  is a  $(1 - \delta)$ -approximate solution to  $f_{S,\gamma}^\mu$  if  $f_{S,\gamma}^\mu(\mathbf{x}) \geq (1 - \delta) f_{S,\gamma}^\mu(\mathbf{x}_{S,\gamma}^\mu)$ .

**LEMMA 3.2.** *For any  $\alpha \geq 0$ , accuracy parameter  $\varepsilon > 0$ , and  $0 < \mu \leq \frac{\varepsilon}{128 \log d}$ , if the subroutine `Rebuild()` in [Algorithm 1](#) returns any  $(1 - \frac{\mu \varepsilon^2}{512})$ -approximate solution to  $f_{S^{(t)},\gamma}^\mu$  as  $\mathbf{x}^{(t)}$  for  $\gamma \stackrel{\text{def}}{=} \alpha$ , then `Rebuild()` will be called at most  $O(\frac{\log d}{\mu \cdot \varepsilon})$  times before  $\max_{\mathbf{x}' \in \mathcal{X}_S} \mathbf{w}_S^\top \mathbf{x}'$  drops from at most  $\alpha$  to below  $\alpha/d$ .*

In the general case that  $\max_{\mathbf{x}' \in \mathcal{X}_S} \mathbf{w}_S^\top \mathbf{x}'$  drops by more than a factor of  $d$ , we can simply re-run the algorithm with different values of  $\gamma$ . [Algorithm 2](#) below implements this strategy and [Theorem 3.1](#) bounds its performance when used as the `Rebuild()` subroutine in [Algorithm 1](#).

**THEOREM 3.1.** *For parameters  $\alpha, k, \varepsilon > 0$ , and  $0 < \mu \leq \frac{\varepsilon}{128 \log d}$ , using [Algorithm 2](#) as the `Rebuild()` subroutine in [Algorithm 1](#), before  $\max_{\mathbf{x}' \in \mathcal{X}_S} \mathbf{w}_S^\top \mathbf{x}'$  drops from at most  $\alpha$  to below  $\alpha/k$  there will be at most  $O(\frac{\log k}{\mu \cdot \varepsilon})$  calls to `Rebuild()`. Moreover, the value  $\gamma$  in `Rebuild()` satisfies  $\max_{\mathbf{x}' \in \mathcal{X}_S} \mathbf{w}_S^\top \mathbf{x}' \leq \gamma \leq \alpha$ , and the while-loop in [Line 3](#) will be run at most  $O(\log_d k)$  times in total.*

---

**Algorithm 2:** Implementation of `Rebuild()` for Theorem 3.1


---

**global:** weight vector  $\mathbf{w} \in \mathbb{R}_{>0}^d$ , accuracy parameter  $\varepsilon \in (0, 1)$ , and  $\mu \in (0, 1)$ .  
**global:** current coordinates  $S \subseteq [d]$ .  
**global:** number of rebuilds  $t \in \mathbb{Z}_{\geq 0}$  and snapshots  $\mathbf{x}^{(t)}$  and  $S^{(t)}$  for analysis.  
**global:** an estimate  $\tilde{\nu}$  for the current phase, initially set to  $\alpha$

```

1 function Rebuild()
2   Let  $\mathbf{x}^{(t)} \in \mathcal{X}_S$  be an arbitrary  $(1 - \frac{\mu\varepsilon^2}{512})$ -approximate solution to  $f_{S,\gamma}^\mu$  for  $\gamma \stackrel{\text{def}}{=} \tilde{\nu}$ .
3   while  $\mathbf{w}^\top \mathbf{x}^{(t)} < \tilde{\nu}/d$  do
4      $\tilde{\nu} \leftarrow \frac{1}{1-\varepsilon} \tilde{\nu}/d$ .
5     Recompute  $\mathbf{x}^{(t)} \in \mathcal{X}_S$  as a  $(1 - \frac{\mu\varepsilon^2}{512})$ -approximate solution to  $f_{S,\gamma}^\mu$  for  $\gamma \stackrel{\text{def}}{=} \tilde{\nu}$ .
6   Set  $\nu^{(t)} \leftarrow \mathbf{w}_S^\top \mathbf{x}^{(t)}$  and  $S^{(t)} \leftarrow S$ .
7   return  $(\mathbf{x}^{(t)}, \nu^{(t)})$  and set  $t \leftarrow t + 1$ .

```

---

*Proof.* Let us assume  $\varepsilon \leq 1/3$ . Consider dividing Algorithm 1 into phases, where each phase ends when  $\tilde{\nu}$  is decreased in Line 4. At the start of each phase, we maintain the invariant that  $\tilde{\nu} \geq \max_{\mathbf{x}' \in \mathcal{X}_S} \mathbf{w}_S^\top \mathbf{x}'$ . The invariant implies that the number of phases is  $O(\log_d k)$ .

Within each phase, as long as  $\mathbf{w}_S^\top \mathbf{x}^{(t)} \geq \tilde{\nu}/d$ , we know that  $\max_{\mathbf{x}' \in \mathcal{X}_S} \mathbf{w}_S^\top \mathbf{x}' \geq \tilde{\nu}/d$ . Lemma 3.2 therefore shows that the vector  $\mathbf{x}$  we maintain throughout this phase is a valid approximation. On the other hand, when  $\mathbf{w}^\top \mathbf{x}^{(t)} < \tilde{\nu}/d$ , we know that  $\max_{\mathbf{x}' \in \mathcal{X}_S} \mathbf{w}_S^\top \mathbf{x}'$  must fall below  $\frac{\tilde{\nu}/d}{1-\varepsilon} \leq \frac{3\tilde{\nu}}{4}$ . Hence, the new  $\tilde{\nu}$  remains an upper bound on it and the invariant holds. Lemma 3.2 also shows that within each phase the subroutine `Rebuild()` will be called at most  $O\left(\frac{\log d}{\mu \cdot \varepsilon}\right)$  times. As such throughout the  $O(\log_d k)$  phases the number of rebuilds is at most  $O\left(\frac{\log k}{\mu \cdot \varepsilon}\right)$ .  $\square$

To complement our algorithmic results, we investigate the limit of the lazy update framework. It turns out that entropy regularization is a nearly optimal strategy for the lazy update framework in its dependence on both  $\log n$  and  $\varepsilon^{-1}$  for the decremental matching problem for any given  $n$  and  $\varepsilon = \Omega(n^{-1/2})$ , even in unweighted bipartite graphs.

**THEOREM 3.2.** *For any  $n \in \mathbb{N}$ ,  $2/\sqrt{n} \leq \varepsilon \leq 1/4$ , and implementation of `Rebuild()` in Algorithm 1, there exists a bipartite graph  $G$  with  $n$  vertices on each side and an output-adaptively chosen sequence of deletions that when  $\mathcal{X} \stackrel{\text{def}}{=} \mathcal{P}_G$  Algorithm 1 calls `Rebuild()`  $\Omega(\log^2(\varepsilon^2 n) \cdot \varepsilon^{-2})$  times before  $G$  is empty.*

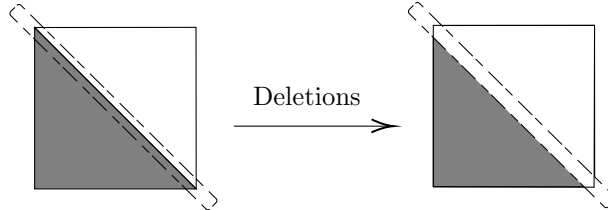


Figure 1: Illustration of the adversarial strategy for proving the lower bound.

To give an intuition of the construction, Figure 1 shows the choice of the initial graph and the output-adaptive deletion strategy. We use the square to represent the adjacency matrix of the chosen bipartite graph, where the gray area represents edges between row vertices and column vertices. In this graph, the only maximum matching is the one on the diagonal of the square. We generalize this observation to any  $(1 - O(\varepsilon))$ -approximate fractional maximum matching and show that a constant fraction of the mass will be concentrated around the diagonal. Thus edge deletion near the diagonal suffices to cause rebuilds of the algorithm. We show that within each phase,

which is defined by deleting everything in the dashed rectangle, an output-adaptive deletion sequence can cause  $\Omega(\log(\varepsilon^2 n)/\varepsilon)$  calls to `Rebuild()`. Additionally, there are  $\Omega(\log(\varepsilon^2 n)/\varepsilon)$  such phases in total, which together establish the lower bound.

**3.3 Dynamic Fractional Matching** We obtain our algorithms for [Problem 1.1](#) by first solving the following intermediate fractional version of the problem.

**PROBLEM 3.2. (DECREMENTAL FRACTIONAL MATCHING)** *In the decremental fractional matching problem, we are given the same set of inputs as [Problem 1.1](#) does, and the goal is to maintain a  $(1 - \varepsilon)$  approximate maximum weight fractional matching, i.e., an  $\mathbf{x} \in \mathcal{M}_G$  such that  $\sum_{e \in E} \mathbf{w}_e \mathbf{x}_e \geq (1 - \varepsilon) M_{\mathbf{w}}^*(G)$ , at all times under deletions to  $E$  until  $G$  becomes empty.*

Our dynamic algorithms use the framework established in the preceding sections, and in particular they apply [Theorem 3.1](#) with  $\mathcal{X} = \mathcal{M}_G$ . This suffices to solve [Problem 3.2](#) provided we can solve the following entropy-regularized problem efficiently.

**PROBLEM 3.3. (ENTROPY-REGULARIZED MATCHING)** *Given an  $n$ -vertex  $m$ -edge graph  $G = (V, E)$  with edge weights  $\mathbf{w} \in \mathbb{N}^E$  bounded by  $W = \text{poly}(n)$ , a trade-off parameter  $\mu \leq 1$ , a  $\gamma \in \mathbb{R}_{\geq 0}$  such that  $M_{\mathbf{w}}^*(G) \leq \gamma \leq n^3 W$ , and an accuracy parameter  $\varepsilon \in (0, 1)$ , the entropy-regularized matching problem is to compute a  $(1 - \varepsilon)$ -approximate solution  $\mathbf{x} \in \mathcal{M}_G$  to*

$$(3.6) \quad \max_{\mathbf{x} \in \mathcal{M}_G} \left\{ \mathbf{w}^\top \mathbf{x} + \mu \cdot \sum_{e \in E} \mathbf{w}_e \mathbf{x}_e \log \frac{\gamma}{\mathbf{w}_e \mathbf{x}_e} \right\}.$$

We develop two algorithms for solving the entropy-regularized matching problem as specified in the following theorem.

**THEOREM 3.3.** *For any  $\varepsilon \geq \tilde{\Omega}(n^{-1/2})$ , there are randomized  $\tilde{O}(m\varepsilon^{-6} + n\varepsilon^{-13})$  and  $\hat{O}(m\varepsilon^{-5})$  time algorithms that solve [Problem 3.3](#) w.h.p.*

This when combined with [Theorem 3.1](#) immediately implies algorithms for solving the decremental fractional matching problem.

**THEOREM 3.4.** *There are fully-adaptive randomized algorithms that, for  $\varepsilon \geq \tilde{\Omega}(n^{-1/6})$ , solve [Problem 3.2](#) w.h.p. with amortized update times  $\tilde{O}(\varepsilon^{-41})$  and  $\hat{O}(\varepsilon^{-17})$ . Additionally, if  $G$  is bipartite, then there is such an algorithm that works for any  $\varepsilon \geq 1/\text{poly}(n)$  with amortized update time  $\hat{O}(\varepsilon^{-2})$ . The recourse of the algorithms is  $\hat{O}(m\varepsilon^{-2})$ .*

*Proof.* Letting  $\mathcal{X} \stackrel{\text{def}}{=} \mathcal{M}_G$  for the decremental matching problem, [Theorem 3.1](#) shows that by using [Problem 3.3](#) with  $\mu \stackrel{\text{def}}{=} \frac{\varepsilon}{128 \log m}$  and accuracy parameter  $\varepsilon' \stackrel{\text{def}}{=} \frac{\mu \varepsilon^2}{512} = \tilde{\Theta}(\varepsilon^3)$  inside [Algorithm 2](#) as the subroutine `Rebuild()`, there will be at most  $\tilde{O}(\varepsilon^{-2})$  calls to `Rebuild()` throughout the algorithm, as the weight of any non-empty matching is at most  $nW$  and at least one. For general graphs, we have  $\varepsilon' \geq \tilde{\Omega}(\varepsilon^{-1/2})$ , and the update times of  $\tilde{O}(\varepsilon^{-41})$  and  $\hat{O}(\varepsilon^{-17})$  can be obtained by running [Theorem 3.3](#) as the subroutine `Rebuild()` and amortizing over the  $m$  updates. For bipartite graphs, we can use [\[33, Theorem 10.16\]](#) to solve [Problem 3.3](#) to high accuracy in  $\hat{O}(m)$  time, resulting in the amortized update time of  $\hat{O}(\varepsilon^{-2})$ . The recourse is  $\hat{O}(m\varepsilon^{-2})$  since there are at most  $\tilde{O}(\varepsilon^{-2})$  different matchings.  $\square$

Our algorithms for [Theorem 3.3](#) build upon the MWU-based algorithm for weighted non-bipartite  $\mathbf{b}$ -matching by [\[5\]](#). Informally, [\[5\]](#) showed that the weighted  $\mathbf{b}$ -matching problem in non-bipartite graphs reduces to solving a sequence of the same problem in bipartite graphs, possibly with different weights. We observe that the analysis of [\[5\]](#) seamlessly extends to general concave objective optimization over the non-bipartite matching polytope. We then leverage the recent almost-linear time convex flow algorithm of [\[33\]](#) for our almost-linear time algorithm for entropy-regularized matching. Alternatively, by approximating the concave weight with piecewise linear functions

and splitting each edge into multiple copies, we reduce [Problem 3.3](#) to a capacity-constrained maximum weight matching problem, which is then solved by similarly applying the generalized approach of [5](#). The runtime of this algorithm does not have the subpolynomial factor incurred by the use of [33](#) but suffers from a larger dependence on  $\varepsilon^{-1}$ .

**3.4 Dynamic Rounding of Fractional Matchings** The results in previous sections show that we can solve the decremental fractional matching problem adaptively. To turn the fractional matching into an integral one, we further design dynamic rounding algorithms for general graphs. To present our algorithms in a unified way, we consider the following weighted definition of dynamic rounding algorithms.

**DEFINITION 3.1. (DYNAMIC ROUNDING ALGORITHM)** *A dynamic rounding algorithm, for a given  $n$ -vertex graph  $G = (V, E)$ , edge weights  $\mathbf{w} \in \mathbb{N}^E$  bounded by  $W = \text{poly}(n)$ , and accuracy parameter  $\varepsilon > 0$ , initializes with an  $\mathbf{x} \in \mathcal{M}_G$  and must maintain an integral matching  $M \subseteq \text{supp}(\mathbf{x})$  with  $\mathbf{w}(M) \geq (1 - \varepsilon)\mathbf{w}^\top \mathbf{x}$  under entry updates to  $\mathbf{x}$  that guarantee  $\mathbf{x} \in \mathcal{M}_G$  after each operation.*

We prove the following deterministic rounding algorithm which has near-optimal overhead in the unweighted case.

**THEOREM 3.5.** *There is a deterministic dynamic rounding algorithm for general graphs with amortized update time  $\tilde{O}(W\varepsilon^{-4})$ .*

Our algorithm for [Theorem 3.5](#) builds on top of the pipage-rounding algorithm recently dynamized by [29](#) for bipartite graphs. Their algorithm circumvents the inherent barrier of the periodic-recomputation approaches by directly rounding to integral matchings without creating an intermediate sparsifier. Although this approach does not generalize to non-bipartite graphs due to odd-set constraints, we observe that terminating their algorithm early in fact creates a good sparsifier for general graphs. The main intuition is that the first few rounds of their algorithm only have small additive effects on the value  $\mathbf{x}_e$ , and perturbing each edge slightly indeed does not have a huge impact on odd-set constraints. While [Definition 3.1](#) is weighted, we remark that the algorithm of [Theorem 3.5](#) is essentially unweighted in its design, hence the linear dependence on  $W$ . We thus apply it in the unweighted case or through reductions to the case when  $W$  is small. See [Section 3.5](#) for more details.

We further consider rounding with better dependence on  $W$ . For this we directly adopt the standard sampling approach of creating a matching sparsifier. Similar approaches were studied before in, e.g., [4](#), [69](#), [29](#), for rounding matchings in bipartite graphs or certain structured matchings in general graphs. Our analysis of the sparsifier, however, differs from the previous ones in that we directly analyze the violation of each odd-set constraint while previous work used various proxies (e.g., kernels or  $\varepsilon$ -restrictness) when arguing the integrality gap. In particular, by standard Chernoff bounds, we show that the sparsifier maintains (i) the degree of the vertices, (ii) the total edge mass in odd sets, and (iii) the unweighted matching size. Coupled with a dynamic set sampler from, e.g., [29](#), we obtain from the sampling approach an unweighted rounding algorithm for general graphs. Though the sampling scheme itself does not lead to a runtime improvement over [Theorem 3.5](#), we further show that in the decremental setting, surprisingly, properties (i) and (ii) suffice to round the entropy-regularized fractional matching maintained by [Theorem 3.4](#) in weighted graphs with large  $W$ . The resulting rounding algorithm in [Theorem 3.6](#) below has a near-optimal overhead in dense graphs.

**THEOREM 3.6.** *There are randomized output-adaptive algorithms that solve [Problem 1.1](#) w.h.p. with amortized update times  $\tilde{O}(\varepsilon^{-41} + (n^2/m) \cdot \varepsilon^{-6})$  and  $\hat{O}(\varepsilon^{-17} + (n^2/m) \cdot \varepsilon^{-6})$ . Additionally, if  $G$  is bipartite then there is such an algorithm with amortized update time  $\hat{O}((n^2/m) \cdot \varepsilon^{-6})$ .*

**3.5 Putting Everything Together** We conclude this overview by using the previously stated results to prove our main theorems.

**THEOREM 3.7. (UNWEIGHTED DECREMENTAL MATCHING)** *There are randomized fully-adaptive  $\tilde{O}(\varepsilon^{-41})$  and  $\hat{O}(\varepsilon^{-17})$  update time algorithms that solve [Problem 1.1](#) in the unweighted case w.h.p.*

*Proof.* For  $\varepsilon < n^{-1/6}$ , the update times of  $\tilde{O}(\varepsilon^{-41})$  and  $\hat{O}(\varepsilon^{-17})$  can be obtained by re-running the static algorithm of [35](#) after each update. As a result we assume  $\varepsilon \geq n^{-1/6}$  in the remainder of the proof. By [Theorem 3.4](#) with

accuracy parameter  $\varepsilon/2$ , we can maintain a  $(1 - \frac{\varepsilon}{2})$ -approximate fractional matching in amortized update times  $\tilde{O}(\varepsilon^{-41})$  and  $\hat{O}(\varepsilon^{-17})$ . We then apply [Theorem 3.5](#) with accuracy parameter  $\varepsilon/2$  to round the fractional matching we maintain to a  $(1 - \varepsilon)$ -approximate integral matching. Since the recourse of [Theorem 3.4](#) is  $\tilde{O}(m\varepsilon^{-2})$ , there will be  $\tilde{O}(m\varepsilon^{-2})$  updates to [Theorem 3.5](#) in total, incurring an additional  $\tilde{O}(\varepsilon^{-6})$  amortized time per update that is subsumed by the update time of [Theorem 3.4](#). Since [Theorem 3.5](#) is deterministic, our final algorithm works against a fully-dynamic adversary like [Theorem 3.4](#) does.  $\square$

**THEOREM 3.8. (WEIGHTED DECREMENTAL MATCHING)** *There is a randomized fully-adaptive  $\tilde{O}(\varepsilon^{-O(1/\varepsilon)})$  update time algorithm that solves [Problem 1.1](#) w.h.p. Additionally, when  $m = \tilde{\Theta}(n^2)$ , there are randomized output-adaptive  $\tilde{O}(\varepsilon^{-41})$  and  $\hat{O}(\varepsilon^{-17})$  update time algorithms.*

We make use of the following weight reduction framework from [\[41\]](#) which allows us to assume that the maximum weight is bounded by  $\varepsilon^{-O(1/\varepsilon)}$ .

**PROPOSITION 3.1. ([\[41\]](#))** *Given a fully-dynamic/incremental/decremental algorithm for  $(1 - \varepsilon)$ -approximate maximum weighted matching on  $n$ -vertex  $m$ -edge graphs of maximum weight  $W$  with worst-case/amortized update time  $T(n, m, \varepsilon, W)$ , there is a fully-dynamic/incremental/decremental algorithm for the same task with worst-case/amortized update time  $\tilde{O}(T(n, m, \Theta(\varepsilon), \varepsilon^{-O(1/\varepsilon)}))$ .*

*Proof of [Theorem 1.2](#).* As in the proof of [Theorem 1.1](#) we assume  $\varepsilon \geq n^{-1/6}$ . For the first algorithm, we apply the weight reduction framework of [\[41\]](#) in [Proposition 3.1](#) to make  $W \leq \varepsilon^{-O(1/\varepsilon)}$ . Again, running [Theorem 3.4](#) with accuracy parameter  $\varepsilon/2$  we maintain a  $(1 - \frac{\varepsilon}{2})$ -approximate fractional matching in amortized update time  $\tilde{O}(\varepsilon^{-41})$ . The rounding algorithm [Theorem 3.5](#) now has amortized update time  $\tilde{O}(\varepsilon^{-O(1/\varepsilon)})$ , incurring an additional  $\tilde{O}(\varepsilon^{-O(1/\varepsilon)})$  amortized time per update which subsumes the update time of the fractional matching. The algorithms for dense graphs follow from [Theorem 3.6](#).  $\square$

Finally, the bipartite result [Theorem 1.3](#) also follows from [Theorem 3.6](#).

**THEOREM 3.9. (WEIGHTED BIPARTITE DECREMENTAL MATCHING)** *For bipartite graphs with  $m = \hat{\Theta}(n^2)$  there is an output-adaptive randomized  $\hat{O}(\varepsilon^{-6})$  update time algorithm that solves [Problem 1.1](#) w.h.p.*

## 4 Entropy Regularization for Decremental Linear Optimization

In this section we analyze our entropy regularization strategy for the lazy update framework that solves the decremental linear optimization problem. In [Sections 4.2](#) and [4.3](#) we show the robustness of the entropy regularization strategy, and in [Section 4.4](#) we prove its optimality. We consider a fixed instance of the decremental linear optimization problem, including the  $d$ -dimensional convex set  $\mathcal{X}$  and the input weight  $\mathbf{w} \in \mathbb{R}_{\geq 0}^d$ .

**4.1 Notation and General Setup** Before showing the robustness of our framework, we first set up the notation and various optimization constructs that will be used throughout the section. For  $S \subseteq [d]$  and  $\gamma \geq 0$ , we consider the entropy regularizer  $r_{S,\gamma} : \mathcal{X}_S \rightarrow \mathbb{R}$  defined by

$$(4.7) \quad r_{S,\gamma}(\mathbf{x}) \stackrel{\text{def}}{=} \sum_{i \in S} \mathbf{w}_i \mathbf{x}_i \log \frac{\gamma}{\mathbf{w}_i \mathbf{x}_i},$$

which is a weighted and scaled version of the original entropy function  $H_S(\mathbf{y}) \stackrel{\text{def}}{=} \sum_{i \in S} \mathbf{y}_i \log \frac{1}{\mathbf{y}_i}$  for  $\mathbf{y} \in \Delta^S$  that is usually applied on the simplex. For intuition, observe that for an  $\mathbf{x} \in \mathcal{X}_S$  with  $\sum_{i \in S} \mathbf{w}_i \mathbf{x}_i = \gamma$ , we have  $r_{S,\gamma}(\mathbf{x}) = \gamma \cdot H_S(\Pi(\mathbf{x}))$ , where  $\Pi(\mathbf{x}) \in \Delta^S$  with  $(\Pi(\mathbf{x}))_i \stackrel{\text{def}}{=} \mathbf{w}_i \mathbf{x}_i / \gamma$ . Below we give upper and lower bounds on the value of  $r_{S,\gamma}(\mathbf{x})$  which generalize known properties of entropy on the simplex.

**LEMMA 4.1.** *For  $\mathbf{x} \in \mathcal{X}_S$  with  $\mathbf{w}_S^\top \mathbf{x} = \nu$ , we have  $\nu \log(\gamma/\nu) \leq r_{S,\gamma}(\mathbf{x}) \leq \nu \log(d\gamma/\nu)$ .*

*Proof.* For the upper bound, consider the relaxation of the problem

$$(4.8) \quad \max_{\mathbf{x}' \in \mathbb{R}^S : \|\mathbf{x}'\|_1 = \nu/\gamma} \gamma \cdot \sum_{i \in S} \mathbf{x}'_i \log \frac{1}{\mathbf{x}'_i}.$$



For any fixed  $\mathbf{x} \in \mathcal{X}_S$ , there is a feasible point  $\mathbf{y}'$  of (4.8) with  $\mathbf{y}'_i \stackrel{\text{def}}{=} \mathbf{w}_i \mathbf{x}_i / \gamma$ , and  $\gamma \cdot \sum_{i \in S} \mathbf{y}' \log \frac{1}{\mathbf{y}'} = r_{S,\gamma}(\mathbf{x})$ . Thus the optimal value of (4.8) is an upper bound on  $r_{S,\gamma}(\mathbf{x})$ . Let  $\mathbf{g}_{\mathbf{x}'} \in \mathbb{R}^S$  with  $(\mathbf{g}_{\mathbf{x}'})_i = -(1 + \log \mathbf{x}'_i)$  be the gradient of the objective of (4.8) at  $\mathbf{x}'$ . Note that the optimality conditions of the problem are that  $\mathbf{g}_{\mathbf{x}'} \perp \ker(\mathbf{1}^{S^\top})$  or equivalently that  $\mathbf{g}_{\mathbf{x}'} = \alpha \cdot \mathbf{1}^S$  for some  $\alpha \in \mathbb{R}$ . Thus, it holds that  $\mathbf{x}' = \beta \cdot \mathbf{1}^S$  for some  $\beta \in \mathbb{R}$ . Combining with  $\|\mathbf{x}'\|_1 = \nu/\gamma$ , we have  $\mathbf{x}'_i = \nu/(d\gamma)$ , implying that the maximizing value of (4.8) is  $\nu \log(d\gamma/\nu)$ .

For the lower bound, since  $\mathbf{w}_S^\top \mathbf{x} = \nu$ , we have  $\max_{i \in S} \mathbf{w}_i \mathbf{x}_i \leq \nu$ , and thus

$$r_{S,\gamma}(\mathbf{x}) \geq \sum_{i \in S} \mathbf{w}_i \mathbf{x}_i \log(\gamma/\nu) = \nu \log(\gamma/\nu).$$

□

Using  $r_{S,\gamma}$  we may rewrite the entropy-regularized objective defined in (3.4) as  $f_{S,\gamma}^\mu(\mathbf{x}) \stackrel{\text{def}}{=} \mathbf{w}_S^\top \mathbf{x} + \mu \cdot r_{S,\gamma}(\mathbf{x})$ . Note that  $\mu > 0$  in the definition. Let  $Z_{S,\gamma}^\mu \stackrel{\text{def}}{=} \max_{\mathbf{x} \in \mathcal{X}_S} f_{S,\gamma}^\mu(\mathbf{x})$  be the maximum entropy-regularized objective value and  $\nu_S^* \stackrel{\text{def}}{=} \max_{\mathbf{x} \in \mathcal{X}_S} \mathbf{w}_S^\top \mathbf{x}$  be the maximum of the actual linear objective. Below are properties of  $f_{S,\gamma}^\mu$  that we will use in this section.

LEMMA 4.2. *The entropy-regularized objective function  $f_{S,\gamma}^\mu$  admits the following properties:*

- (i)  $f_{S,\gamma}^\mu$  has a unique maximizer  $\mathbf{x}_{S,\gamma}^\mu$  on  $\mathcal{X}_S$ .
- (ii)  $\mathbf{x}_{S,\gamma}^\mu$  has positive coordinates, i.e.,  $\mathbf{x}_{S,\gamma}^\mu \in \mathcal{X}_{S,+}$ , and thus  $f_{S,\gamma}^\mu$  and  $r_{S,\gamma}$  are differentiable at  $\mathbf{x}_{S,\gamma}^\mu$ .
- (iii)  $f_{S,\gamma}^\mu(\mathbf{x}) \leq f_{S,\gamma}^\mu(\mathbf{x}_{S,\gamma}^\mu) - \frac{\mu}{2\nu_S^*} \left\| \mathbf{x} - \mathbf{x}_{S,\gamma}^\mu \right\|_{\mathbf{w},S}^2$  for all  $\mathbf{x} \in \mathcal{X}_S$ , where  $\|\mathbf{x}\|_{\mathbf{w},S} \stackrel{\text{def}}{=} \sum_{i \in S} \mathbf{w}_i |\mathbf{x}_i|$  is the weighted  $\ell_1$ -norm.<sup>18</sup>

We provide a fairly standard proof of the properties from first principle.

*Proof.* By the compactness of  $\mathcal{X}_S$ ,  $f_{S,\gamma}^\mu$  has a maximizer on  $\mathcal{X}_S$ . We prove the following claims.

CLAIM 4.1. *Any maximizer of  $f_{S,\gamma}^\mu$  on  $\mathcal{X}_S$  must be in  $\mathcal{X}_{S,+}$ .*

*Proof.* Consider the first-order partial derivative of  $f_{S,\gamma}^\mu$  with respect to an  $\mathbf{x} \in \mathcal{X}_{S,+}$  and coordinate  $i \in S$ , which by calculation is

$$\partial_i f_{S,\gamma}^\mu(\mathbf{x}) \stackrel{\text{def}}{=} \frac{\partial f_{S,\gamma}^\mu(\mathbf{x})}{\partial \mathbf{x}_i} = (1 - \mu) \mathbf{w}_i + \mu \mathbf{w}_i \log \left( \frac{\gamma}{\mathbf{w}_i \mathbf{x}_i} \right).$$

Since  $\mathcal{X}$  contains no degenerate dimension and is convex, there is a point  $\mathbf{y} \in \mathcal{X}_{S,+}$ . Consider any point  $\mathbf{x} \in \mathcal{X}_S \setminus \mathcal{X}_{S,+}$ , and let  $\mathbf{x}_\alpha \stackrel{\text{def}}{=} \mathbf{x} + \alpha(\mathbf{y} - \mathbf{x})$  for  $\alpha \in [0, 1]$ . Note that  $\mathbf{x}_\alpha \in \mathcal{X}_{S,+}$  for  $\alpha > 0$ . By the mean value theorem, for any  $\alpha > 0$  there is a  $0 < \beta < \alpha$  such that

$$f_{S,\gamma}^\mu(\mathbf{x}_\alpha) - f_{S,\gamma}^\mu(\mathbf{x}) = \nabla f_{S,\gamma}^\mu(\mathbf{x}_\beta)^\top (\mathbf{x}_\alpha - \mathbf{x}) = \sum_{i \in S} \partial_i f_{S,\gamma}^\mu(\mathbf{x}_\beta) \cdot ((\mathbf{x}_\alpha)_i - \mathbf{x}_i).$$

For  $i \in S$  with  $\mathbf{x}_i > 0$ ,  $|\partial_i f_{S,\gamma}^\mu(\mathbf{x}_\beta)|$  is bounded since  $(\mathbf{x}_\beta)_i$  is between  $\mathbf{x}_i$  and  $\mathbf{y}_i$ . For  $i \in S$  with  $\mathbf{x}_i = 0$ , however, when  $(\mathbf{x}_\beta)_i$  approaches 0,  $\partial_i f_{S,\gamma}^\mu(\mathbf{x}_\beta)$  goes to infinity. This shows that we can pick an  $\alpha$  close enough to 0 so that  $f_{S,\gamma}^\mu(\mathbf{x}_\alpha) > f_{S,\gamma}^\mu(\mathbf{x})$ , proving that  $\mathbf{x} \in \mathcal{X}_S \setminus \mathcal{X}_{S,+}$  is not a maximizer. □

CLAIM 4.2. *For any  $\mathbf{x}, \mathbf{z} \in \mathcal{X}_{S,+}$  it holds that  $\mathbf{z}^\top \nabla^2 f_{S,\gamma}^\mu(\mathbf{x}) \mathbf{z} \leq -\frac{\mu}{\nu_S^*} \|\mathbf{z}\|_{\mathbf{w},S}^2$ .*

Note that  $f_{S,\gamma}^\mu$  is twice-differentiable on  $\mathcal{X}_{S,+}$  and thus Claim 4.2 is well-defined.

<sup>18</sup>  $\|\cdot\|_{\mathbf{w},S}$  is indeed a norm since  $\mathbf{w}_i > 0$  for all  $i \in [d]$ .

*Proof.* The second-order partial derivatives of  $f_{S,\gamma}^\mu$  satisfy

$$\frac{\partial^2 f_{S,\gamma}^\mu(\mathbf{x})}{\partial x_i^2} = -\frac{\mu \mathbf{w}_i}{x_i} \quad \text{and} \quad \frac{\partial^2 f_{S,\gamma}^\mu(\mathbf{x})}{\partial x_i \partial x_j} = 0 \text{ for } j \neq i.$$

Using Cauchy-Schwarz inequality, for any  $\mathbf{x}, \mathbf{z} \in \mathcal{X}_{S,+}$  we have

$$\mathbf{z}^\top \nabla^2 f_{S,\gamma}^\mu(\mathbf{x}) \mathbf{z} = \sum_{i \in S} -\frac{\mu \mathbf{w}_i}{x_i} z_i^2 \leq -\mu \frac{\|\mathbf{z}\|_{\mathbf{w},S}^2}{\|\mathbf{x}\|_{\mathbf{w},S}} \leq -\frac{\mu}{\nu_S^*} \|\mathbf{z}\|_{\mathbf{w},S}^2,$$

as claimed.  $\square$

Now consider a maximizer  $\mathbf{y}$  of  $f_{S,\gamma}^\mu$  and an  $\mathbf{x} \in \mathcal{X}_{S,+}$ . Since  $\mathbf{y} \in \mathcal{X}_{S,+}$ ,  $f_{S,\gamma}^\mu$  is differentiable at  $\mathbf{y}$ . Letting  $\mathbf{y}_\alpha \stackrel{\text{def}}{=} \mathbf{y} + \alpha(\mathbf{x} - \mathbf{y})$  for  $\alpha \in [0, 1]$ , we have

$$(4.9) \quad f_{S,\gamma}^\mu(\mathbf{x}) = f_{S,\gamma}^\mu(\mathbf{y}) + \left( \nabla f_{S,\gamma}^\mu(\mathbf{y}) \right)^\top (\mathbf{x} - \mathbf{y}) + \int_0^1 \int_0^t (\mathbf{x} - \mathbf{y})^\top \nabla^2 f_{S,\gamma}^\mu(\mathbf{y}_\alpha) (\mathbf{x} - \mathbf{y}) d\alpha dt$$

$$(4.10) \quad \stackrel{(i)}{\leq} f_{S,\gamma}^\mu(\mathbf{y}) - \frac{\mu}{\nu_S^*} \|\mathbf{x} - \mathbf{y}\|_{\mathbf{w},S}^2 \cdot \int_0^1 \int_0^t d\alpha dt = f_{S,\gamma}^\mu(\mathbf{y}) - \frac{\mu}{2\nu_S^*} \|\mathbf{x} - \mathbf{y}\|_{\mathbf{w},S}^2,$$

where (i) uses the optimality conditions of  $f_{S,\gamma}^\mu$  at  $\mathbf{y}$  and [Claim 4.2](#). On the other hand, for  $\mathbf{x} \in \mathcal{X}_S \setminus \mathcal{X}_{S,+}$ , there is a sequence  $\{\mathbf{x}_n\} \subseteq \mathcal{X}_{S,+}$  approaching  $\mathbf{x}$ . By continuity of  $f_{S,\gamma}^\mu$  we have

$$(4.11) \quad f_{S,\gamma}^\mu(\mathbf{x}) = \lim_{n \rightarrow \infty} f_{S,\gamma}^\mu(\mathbf{x}_n) \leq \lim_{n \rightarrow \infty} f_{S,\gamma}^\mu(\mathbf{y}) - \frac{\mu}{2\nu_S^*} \|\mathbf{x}_n - \mathbf{y}\|_{\mathbf{w},S}^2 = f_{S,\gamma}^\mu(\mathbf{y}) - \frac{\mu}{2\nu_S^*} \|\mathbf{x} - \mathbf{y}\|_{\mathbf{w},S}^2.$$

This implies that there is a unique maximizer  $\mathbf{x}_{S,\gamma}^\mu$  of  $f_{S,\gamma}^\mu$  on  $\mathcal{X}_S$ . The rest of the lemma follows from [Claim 4.1](#) and [Equations \(4.10\)](#) and [\(4.11\)](#).  $\square$

Finally, our analysis of the framework uses the Bregman divergence of  $r_{S,\gamma}$  as a proxy to bound the decrease of a certain potential.

**DEFINITION 4.1. (BREGMAN DIVERGENCE)** For differentiable  $r : \mathcal{X} \rightarrow \mathbb{R}$  and  $\mathbf{x}, \mathbf{y} \in \mathcal{X}$  for a domain  $\mathcal{X}$ , the Bregman divergence of  $r$  from  $\mathbf{y}$  to  $\mathbf{x}$  is

$$V_{\mathbf{y}}^r(\mathbf{x}) \stackrel{\text{def}}{=} r(\mathbf{x}) - \left( r(\mathbf{y}) + (\nabla r(\mathbf{y}))^\top (\mathbf{x} - \mathbf{y}) \right).$$

Overloading notation, let  $V_{\mathbf{y}}^{S,\gamma}(\mathbf{x})$  for  $S \subseteq [d]$ ,  $\mathbf{x} \in \mathcal{X}_S$  and  $\mathbf{y} \in \mathcal{X}_{S,+}$  be the Bregman divergence induced by  $r_{S,\gamma}$ , which is non-positive since  $r_{S,\gamma}$ , as a generalization of the entropy function, is concave. We have by direct calculation from the definition that

$$(4.12) \quad V_{\mathbf{y}}^{S,\gamma}(\mathbf{x}) \stackrel{\text{def}}{=} V_{\mathbf{y}}^{r_{S,\gamma}}(\mathbf{x}) = r_{S,\gamma}(\mathbf{x}) - \left( r_{S,\gamma}(\mathbf{y}) + (\nabla r_{S,\gamma}(\mathbf{y}))^\top (\mathbf{x} - \mathbf{y}) \right)$$

$$(4.13) \quad = \left( \sum_{i \in S} \mathbf{w}_i x_i \log \frac{\gamma}{\mathbf{w}_i x_i} \right) - \left[ \left( \sum_{i \in S} \mathbf{w}_i y_i \log \frac{\gamma}{\mathbf{w}_i y_i} \right) - \left( \sum_{i \in S} \mathbf{w}_i \left( 1 + \log \frac{\mathbf{w}_i y_i}{\gamma} \right) (x_i - y_i) \right) \right]$$

$$(4.14) \quad = \sum_{i \in S} \mathbf{w}_i x_i \log \frac{y_i}{x_i} + \sum_{i \in S} \mathbf{w}_i (x_i - y_i).$$

Note that we need  $\mathbf{y} \in \mathcal{X}_{S,+}$  because the gradient does not exist on  $\mathcal{X}_S \setminus \mathcal{X}_{S,+}$ . Indeed, in the remainder of the section we will only use  $V_{\mathbf{y}}^{S,\gamma}(\mathbf{x})$  for  $\mathbf{y} = \mathbf{x}_{S,\gamma}^\mu$ , which lies in  $\mathcal{X}_{S,+}$  by [Lemma 4.2\(ii\)](#). The following lemma bounds the entropy-regularized objective by the Bregman divergence from  $\mathbf{x}_{S,\gamma}^\mu$ .

**LEMMA 4.3.** For any  $\mathbf{x} \in \mathcal{X}_S$ , we have  $f_{S,\gamma}^\mu(\mathbf{x}) \leq f_{S,\gamma}^\mu(\mathbf{x}_{S,\gamma}^\mu) + \mu V_{\mathbf{x}_{S,\gamma}^\mu}^{S,\gamma}(\mathbf{x})$ .

*Proof.* For clarity let us write  $\mathbf{x}^* \stackrel{\text{def}}{=} \mathbf{x}_{S,\gamma}^\mu$ . Optimality conditions for concave optimization applied to  $f_{S,\gamma}^\mu(\mathbf{x}^*)$  imply that

$$(4.15) \quad (\mathbf{w}_S + \mu \nabla r_{S,\gamma}(\mathbf{x}^*))^\top (\mathbf{x} - \mathbf{x}^*) = \nabla f_{S,\gamma}^\mu(\mathbf{x}^*)^\top (\mathbf{x} - \mathbf{x}^*) \leq 0.$$

As such, we have

$$\begin{aligned} f_{S,\gamma}^\mu(\mathbf{x}^*) - f_{S,\gamma}^\mu(\mathbf{x}) &\stackrel{(i)}{=} \mathbf{w}_S^\top (\mathbf{x}^* - \mathbf{x}) + \mu(r_{S,\gamma}(\mathbf{x}^*) - r_{S,\gamma}(\mathbf{x})) \\ &\stackrel{(ii)}{=} (\mathbf{w}_S + \mu \nabla r_{S,\gamma}(\mathbf{x}^*))^\top (\mathbf{x}^* - \mathbf{x}) - \mu V_{\mathbf{x}^*}^{S,\gamma}(\mathbf{x}) \stackrel{(iii)}{\geq} -\mu V_{\mathbf{x}^*}^{S,\gamma}(\mathbf{x}), \end{aligned}$$

where (i) is by definition of  $f_{S,\gamma}^\mu$ , (ii) is by definition of the Bregman divergence, and (iii) follows from (4.15).  $\square$

The following Table 1 summarizes the notation introduced here for future reference.

Notation	Definition	Description
$r_{S,\gamma}(\mathbf{x})$	$\sum_{i \in S} \mathbf{w}_i \mathbf{x}_i \log \frac{\gamma}{\mathbf{w}_i \mathbf{x}_i}$	entropy regularizer
$V_{\mathbf{y}}^{S,\gamma}(\mathbf{x})$	$\sum_{i \in S} \mathbf{w}_i \mathbf{x}_i \log(\mathbf{y}_i / \mathbf{x}_i) + \sum_{i \in S} \mathbf{w}_i (\mathbf{x}_i - \mathbf{y}_i)$	Bregman divergence of entropy regularizer
$f_{S,\gamma}^\mu(\mathbf{x})$	$\mathbf{w}_S^\top \mathbf{x} + \mu \cdot r_{S,\gamma}(\mathbf{x})$	entropy-regularized objective
$\mathbf{x}_{S,\gamma}^\mu$	$\operatorname{argmax}_{\mathbf{x} \in \mathcal{X}_S} f_{S,\gamma}^\mu(\mathbf{x})$	optimal entropy-regularized solution
$\nu_S^*$	$\max_{\mathbf{x} \in \mathcal{X}_S} \mathbf{w}_S^\top \mathbf{x}$	optimal linear objective value
$Z_{S,\gamma}^\mu$	$\max_{\mathbf{x} \in \mathcal{X}_S} f_{S,\gamma}^\mu(\mathbf{x})$	optimal entropy-regularized objective value
$\ \mathbf{x}\ _{\mathbf{w},S}$	$\sum_{i \in S} \mathbf{w}_i  \mathbf{x}_i $	norm of $\mathbf{x} \in \mathcal{X}_S$ induced by $\mathbf{w}$

Table 1: Frequently used notation

**4.2 Robustness of Entropy Regularization** In this section we show that if the subroutine `Rebuild()` in Algorithm 1 returns the exact maximizer of the entropy-regularized objective, then `Rebuild()` will be called at most  $O(\varepsilon^{-2})$  times. The first step is to derive a proper choice of  $\gamma$  and  $\mu$  so that  $\mathbf{x}_{S,\gamma}^\mu$  becomes an approximate maximum weight solution that can be used in Algorithm 1. We call a tuple containing an accuracy parameter  $\varepsilon > 0$ , coordinate subset  $S \subseteq [d]$ , and estimate

$$(4.16) \quad \nu_S^* \leq \gamma \leq d \cdot \nu_S^*$$

a *valid iterate* which will repeatedly appear as input to lemmas in the remainder of this section.

LEMMA 4.4. *For any valid iterate  $(\varepsilon, S, \gamma)$  and  $0 < \mu \leq \frac{\varepsilon}{8 \log d}$ , it holds that  $0 \leq \mu \cdot r_{S,\gamma}(\mathbf{x}) \leq \frac{\varepsilon}{2} \cdot \nu_S^*$  for all  $\mathbf{x} \in \mathcal{X}_S$ .*

*Proof.* The lower bound follows from Lemma 4.1 and the assumption of  $\gamma$  in (4.16). For the upper bound, let  $\mathbf{x} \in \mathcal{X}_S$  be arbitrary and define  $\nu \stackrel{\text{def}}{=} \mathbf{w}_S^\top \mathbf{x}$ . Lemma 4.1 implies that

$$\begin{aligned} \mu \cdot r_{S,\gamma}(\mathbf{x}) &\leq \mu \nu \log \frac{d\gamma}{\nu} = \mu \nu \log \frac{d\gamma}{\nu_S^*} + \mu \nu_S^* \cdot \frac{\nu}{\nu_S^*} \log \frac{\nu_S^*}{\nu} \\ &\stackrel{(i)}{\leq} \mu \nu \cdot (2 \log d) + \mu \nu_S^* \stackrel{(ii)}{\leq} \frac{\varepsilon}{2} \cdot \nu_S^*, \end{aligned}$$

where (i) comes from  $x \log(1/x) \leq 1$  for all  $x \in \mathbb{R}_{\geq 0}$  and the assumption of  $\gamma$  in (4.16) and (ii) is by the assumption of  $\mu$  and  $\nu \leq \nu_S^*$ .  $\square$

LEMMA 4.5. For any valid iterate  $(\varepsilon, S, \gamma)$  and  $0 < \mu \leq \frac{\varepsilon}{8 \log d}$ , it holds that

$$\mathbf{w}_S^\top \mathbf{x}_{S,\gamma}^\mu \geq \left(1 - \frac{\varepsilon}{2}\right) \cdot \nu_S^*.$$

*Proof.* For any  $\mathbf{x} \in \mathcal{X}_S$  such that  $\mathbf{w}_S^\top \mathbf{x} < (1 - \varepsilon/2) \cdot \nu_S^*$ , we have

$$f_{S,\gamma}^\mu(\mathbf{x}_{S,\gamma}^\mu) - f_{S,\gamma}^\mu(\mathbf{x}) \stackrel{(i)}{\geq} \nu_S^* - \left(\mathbf{w}_S^\top \mathbf{x} + \frac{\varepsilon}{2} \cdot \nu_S^*\right) \stackrel{(ii)}{>} 0,$$

where (i) comes from [Lemma 4.4](#) (note that the non-negativity of the entropy-regularizer implies  $f_{S,\gamma}^\mu(\mathbf{x}_{S,\gamma}^\mu) \geq \nu_S^*$ ) and (ii) comes from the assumption that  $\mathbf{w}_S^\top \mathbf{x} < (1 - \varepsilon/2) \cdot \nu_S^*$ . The lemma follows since  $\mathbf{x}_{S,\gamma}^\mu$  is a maximizer of  $f_{S,\gamma}^\mu$ .  $\square$

[Lemma 4.5](#) shows that for  $\gamma$  being a  $d$ -approximate upper bound of  $\nu_S^*$ , if we choose  $\mu$  to be roughly proportional to  $\varepsilon$ , then the corresponding value of  $\mathbf{w}_S^\top \mathbf{x}_{S,\gamma}^\mu$  is a  $(1 - \varepsilon/2)$ -approximation to  $\nu_S^*$ . Thus we can use the entropy-regularized solution for the `Rebuild()` procedure in [Algorithm 1](#). Moreover, as we will show below, a decrease in the linear objective value of  $\mathbf{x}_{S,\gamma}^\mu$  implies a decrease in the entropy-regularized objective globally. This enables the use of the latter quantity as a potential to bound the number of calls to `Rebuild()` [Algorithm 1](#) needs.

LEMMA 4.6. For any valid iterate  $(\varepsilon, S, \gamma)$ ,  $0 < \mu \leq \frac{\varepsilon}{8 \log d}$ , and  $S' \subseteq S$ , if

$$\mathbf{w}_{S'}^\top (\mathbf{x}_{S,\gamma}^\mu)_{S'} < \left(1 - \frac{\varepsilon}{2}\right) \cdot \mathbf{w}_S^\top \mathbf{x}_{S,\gamma}^\mu,$$

then  $Z_{S',\gamma}^\mu \leq \left(1 - \frac{\mu\varepsilon}{3}\right) Z_{S,\gamma}^\mu$ .

*Proof.* For clarity let  $\mathbf{x}_S \stackrel{\text{def}}{=} \mathbf{x}_{S,\gamma}^\mu$ , and  $\mathbf{x}_{S'} \stackrel{\text{def}}{=} (\mathbf{x}_{S',\gamma}^\mu)^S$  be  $\mathbf{x}_{S',\gamma}^\mu$  extended to have coordinates  $S$ . [Lemma 4.3](#) shows that

$$f_{S,\gamma}^\mu(\mathbf{x}_{S'}) \leq f_{S,\gamma}^\mu(\mathbf{x}_S) + \mu V_{\mathbf{x}_S}^{S,\gamma}(\mathbf{x}_{S'}).$$

Since  $S' \subseteq S$ , we have

$$(4.17) \quad Z_{S',\gamma}^\mu = f_{S',\gamma}^\mu(\mathbf{x}_{S',\gamma}^\mu) = f_{S,\gamma}^\mu(\mathbf{x}_{S'}) \leq f_{S,\gamma}^\mu(\mathbf{x}_S) + \mu V_{\mathbf{x}_S}^{S,\gamma}(\mathbf{x}_{S'}) = Z_{S,\gamma}^\mu + \mu V_{\mathbf{x}_S}^{S,\gamma}(\mathbf{x}_{S'}).$$

Further, by above letting  $D \stackrel{\text{def}}{=} S \setminus S'$  and therefore  $\mathbf{x}_{S',i} = 0$  for  $i \in D$ , we have from [\(4.14\)](#) and the non-positiveness of the Bregman divergence for concave functions that

$$\begin{aligned} V_{\mathbf{x}_S}^{S,\gamma}(\mathbf{x}_{S'}) &= \sum_{i \in S} \mathbf{w}_i(\mathbf{x}_{S'})_i \log \frac{(\mathbf{x}_S)_i}{(\mathbf{x}_{S'})_i} + \sum_{i \in S} \mathbf{w}_i((\mathbf{x}_{S'})_i - (\mathbf{x}_S)_i) \\ &\leq \sum_{i \in D} \mathbf{w}_i(\mathbf{x}_{S'})_i \log \frac{(\mathbf{x}_S)_i}{(\mathbf{x}_{S'})_i} + \sum_{i \in D} \mathbf{w}_i((\mathbf{x}_{S'})_i - (\mathbf{x}_S)_i) \\ &= - \sum_{i \in D} \mathbf{w}_i(\mathbf{x}_S)_i \leq -\frac{\varepsilon}{2} \cdot \mathbf{w}_S^\top \mathbf{x}_S, \end{aligned}$$

which when combined with [\(4.17\)](#) and [Lemma 4.4](#) shows that

$$Z_{S',\gamma}^\mu \leq Z_{S,\gamma}^\mu - \frac{\mu\varepsilon}{2} \cdot \mathbf{w}_S^\top \mathbf{x}_S \leq \left(1 - \frac{\mu\varepsilon}{(1 + \varepsilon/2) \cdot 2}\right) \cdot Z_{S,\gamma}^\mu \leq \left(1 - \frac{\mu\varepsilon}{3}\right) \cdot Z_{S,\gamma}^\mu.$$

$\square$

We now conclude the number of rebuilds before the actual optimal value  $\nu_S^*$  drops to a certain number, if we use an entropy-regularized solution  $\mathbf{x}_{S,\gamma}^\mu$  for `Rebuild()`.

LEMMA 3.1. For any  $\alpha \geq 0$ , accuracy parameter  $\varepsilon > 0$ , and  $0 < \mu \leq \frac{\varepsilon}{8 \log d}$ , if the subroutine `Rebuild()` in [Algorithm 1](#) returns  $\mathbf{x}_{S^{(t)}, \gamma}^\mu$  as  $\mathbf{x}^{(t)}$  for  $\gamma \stackrel{\text{def}}{=} \alpha$ , then `Rebuild()` will be called at most  $O(\frac{\log d}{\mu \cdot \varepsilon})$  times before  $\max_{\mathbf{x}' \in \mathcal{X}_S} \mathbf{w}_S^\top \mathbf{x}'$  drops from at most  $\alpha$  to below  $\alpha/d$ .

*Proof.* Note that while  $\nu_S^*$  is between  $\alpha/d$  and  $\alpha$ ,  $\gamma \stackrel{\text{def}}{=} \alpha$  is a  $d$ -approximate upper bound of  $\nu_S^*$  and thus the preceding lemmas apply. By [Lemma 4.5](#),  $\mathbf{x}^{(t)}$  is a  $(1 - \frac{\varepsilon}{2})$ -approximate solution to the linear objective  $\mathbf{w}_{S^{(t)}}^\top \mathbf{x}$  before  $\nu_S^*$  drops below  $\alpha$ . Fix a round  $t$  and consider  $S^{(t+1)}$  which is obtained by deleting some coordinates from  $S^{(t)}$  so that

$$\mathbf{w}_{S^{(t+1)}}^\top \mathbf{x}_{S^{(t+1)}}^{(t)} < \left(1 - \frac{\varepsilon}{2}\right) \mathbf{w}_{S^{(t)}}^\top \mathbf{x}^{(t)}.$$

In other words, the quality of the current solution  $\mathbf{x}^{(t)}$  decreases by a multiplicative factor of  $1 - \frac{\varepsilon}{2}$  when we go from  $S^{(t)}$  to  $S^{(t+1)}$ . [Lemma 4.6](#) then implies

$$Z_{S^{(t+1)}, \gamma}^\mu \leq \left(1 - \frac{\mu \varepsilon}{3}\right) Z_{S^{(t)}, \gamma}^\mu$$

for each round  $t$ . This allows us to bound the number of calls to `Rebuild()` as follows. Initially when  $\max_{\mathbf{x} \in \mathcal{X}_S} \mathbf{w}_S^\top \mathbf{x} \leq \alpha$ , the optimal entropy-regularized objective is no more than  $(1 + \varepsilon/2) \cdot \alpha$  by [Lemma 4.4](#). Likewise, at the end before  $\nu_S^*$  drops below  $\alpha/d$ , the objective is at least  $\alpha/d$ . Thus there will be at most

$$\log_{1 - \frac{\mu \varepsilon}{3}}((1 + \varepsilon/2)d) = O\left(\frac{\log d}{\mu \varepsilon}\right)$$

calls to `Rebuild()`.  $\square$

**4.3 Sufficiency of Approximate Solutions** From [Section 4.2](#) we have seen that the maximizer of the entropy-regularized objective solves the decremental linear optimization problem. However, exact maximizers are not always easy to obtain, and therefore in this section, we show that any accurate enough approximate maximizer of  $f_{S, \gamma}^\mu$  suffices for the lazy update framework to work efficiently. This is by the following lemma which states that such a solution is also close in  $\|\cdot\|_{\mathbf{w}, S}$ -distance to the actual maximizer  $\mathbf{x}_{S, \gamma}^\mu$ .

LEMMA 4.7. For any valid iterate  $(\varepsilon, S, \gamma)$ ,  $0 < \mu \leq \frac{\varepsilon}{8 \log d}$ , and  $\mathbf{x} \in \mathcal{X}_S$  with  $f_{S, \gamma}^\mu(\mathbf{x}) \geq \left(1 - \frac{\mu \varepsilon^2}{2}\right) Z_{S, \gamma}^\mu$ , it holds that  $\|\mathbf{x} - \mathbf{x}_{S, \gamma}^\mu\|_{\mathbf{w}, S} \leq \varepsilon \cdot Z_{S, \gamma}^\mu$ .

*Proof.* [Lemma 4.2\(iii\)](#) states that

$$f_{S, \gamma}^\mu(\mathbf{x}) \leq f_{S, \gamma}^\mu(\mathbf{x}_{S, \gamma}^\mu) - \frac{\mu}{2\nu_S^*} \|\mathbf{x} - \mathbf{x}_{S, \gamma}^\mu\|_{\mathbf{w}, S}^2$$

for all  $\mathbf{x} \in \mathcal{X}_S$ . Thus for every  $\mathbf{x} \in \mathcal{X}_S$  such that  $f_{S, \gamma}^\mu(\mathbf{x}) \geq \left(1 - \frac{\mu \varepsilon^2}{2}\right) Z_{S, \gamma}^\mu$ , we have

$$\|\mathbf{x} - \mathbf{x}_{S, \gamma}^\mu\|_{\mathbf{w}, S} \leq \sqrt{\frac{2\nu_S^*}{\mu} \left(f_{S, \gamma}^\mu(\mathbf{x}_{S, \gamma}^\mu) - f_{S, \gamma}^\mu(\mathbf{x})\right)} \leq \sqrt{\varepsilon^2 \cdot \nu_S^* \cdot Z_{S, \gamma}^\mu} \leq \varepsilon \cdot Z_{S, \gamma}^\mu,$$

where the last inequality follows from [Lemma 4.4](#) and therefore  $\nu_S^* \leq Z_{S, \gamma}^\mu$ .  $\square$

The closeness of an approximate maximizer  $\mathbf{x}$  to the actual one  $\mathbf{x}_{S, \gamma}^\mu$  allows us to bound the decrease in the objective value of  $\mathbf{x}_{S, \gamma}^\mu$  that is hidden to us. This establishes the number of rebuilds needed if we only have an accurate enough approximation to  $f_{S, \gamma}^\mu$ .

LEMMA 3.2. For any  $\alpha \geq 0$ , accuracy parameter  $\varepsilon > 0$ , and  $0 < \mu \leq \frac{\varepsilon}{128 \log d}$ , if the subroutine `Rebuild()` in [Algorithm 1](#) returns any  $(1 - \frac{\mu \varepsilon^2}{512})$ -approximate solution to  $f_{S^{(t)}, \gamma}^\mu$  as  $\mathbf{x}^{(t)}$  for  $\gamma \stackrel{\text{def}}{=} \alpha$ , then `Rebuild()` will be called at most  $O(\frac{\log d}{\mu \cdot \varepsilon})$  times before  $\max_{\mathbf{x}' \in \mathcal{X}_S} \mathbf{w}_S^\top \mathbf{x}'$  drops from at most  $\alpha$  to below  $\alpha/d$ .



*Proof.* Setting  $\varepsilon' \stackrel{\text{def}}{=} \varepsilon/4$  and  $\varepsilon'' \stackrel{\text{def}}{=} \varepsilon/16$ , we have  $\mu \leq \frac{\varepsilon''}{8 \log d} \leq \frac{\varepsilon'}{8 \log d}$  and  $\mathbf{x}^{(t)}$  being a  $(1 - \frac{\mu \varepsilon''}{2})$ -approximate solution to  $f_{S^{(t)}, \gamma}^\mu$ . By Lemma 4.5 with accuracy parameter  $\varepsilon'$  we know that  $\mathbf{x}_{S^{(t)}, \gamma}^\mu$  is a  $(1 - \frac{\varepsilon}{8})$ -approximate solution to the linear objective  $\mathbf{w}_{S^{(t)}}^\top \mathbf{x}$  before  $\nu_S^*$  drops below  $\alpha/d$ . By triangle inequality of the norm  $\|\cdot\|_{\mathbf{w}, S^{(t)}}$  and Lemma 4.7 with accuracy parameter  $\varepsilon''$ , the value of  $\mathbf{w}_{S^{(t)}}^\top \mathbf{x}^{(t)}$  satisfies

$$(4.18) \quad \mathbf{w}_{S^{(t)}}^\top \mathbf{x}^{(t)} = \|\mathbf{x}^{(t)}\|_{\mathbf{w}, S^{(t)}} \geq \|\mathbf{x}_{S^{(t)}, \gamma}^\mu\|_{\mathbf{w}, S^{(t)}} - \|\mathbf{x}_{S^{(t)}, \gamma}^\mu - \mathbf{x}^{(t)}\|_{\mathbf{w}, S^{(t)}}$$

$$(4.19) \quad \geq \left(1 - \frac{\varepsilon}{8}\right) \nu_{S^{(t)}}^* - \frac{\varepsilon}{16} Z_{S^{(t)}, \gamma}^\mu \geq \left(1 - \frac{\varepsilon}{4}\right) \nu_{S^{(t)}}^*,$$

where the last inequality uses that  $Z_{S^{(t)}, \gamma}^\mu \leq (1 + \frac{\varepsilon}{8}) \nu_{S^{(t)}}^* \leq 2\nu_{S^{(t)}}^*$  by Lemma 4.4 with accuracy parameter  $\varepsilon'$ . This shows that  $\mathbf{x}^{(t)}$  is indeed a  $(1 - \frac{\varepsilon}{2})$ -approximate solution to the linear objective, as required by Algorithm 1. Fix a round  $t$  and consider  $S^{(t+1)}$  which is obtained by deleting some coordinates from  $S^{(t)}$  so that

$$\mathbf{w}_{S^{(t+1)}}^\top \mathbf{x}_{S^{(t+1)}}^{(t)} < \left(1 - \frac{\varepsilon}{2}\right) \mathbf{w}_{S^{(t)}}^\top \mathbf{x}^{(t)}.$$

This implies

$$\begin{aligned} \mathbf{w}_{S^{(t+1)}}^\top (\mathbf{x}_{S^{(t)}, \gamma}^\mu)_{S^{(t+1)}} &= \|(\mathbf{x}_{S^{(t)}, \gamma}^\mu)_{S^{(t+1)}}\|_{\mathbf{w}, S^{(t+1)}} \\ &\leq \|(\mathbf{x}^{(t)})_{S^{(t+1)}}\|_{\mathbf{w}, S^{(t+1)}} + \|(\mathbf{x}_{S^{(t)}, \gamma}^\mu - \mathbf{x}^{(t)})_{S^{(t+1)}}\|_{\mathbf{w}, S^{(t+1)}} \\ &< \left(1 - \frac{\varepsilon}{2}\right) \mathbf{w}_{S^{(t)}}^\top \mathbf{x}^{(t)} + \|\mathbf{x}_{S^{(t)}, \gamma}^\mu - \mathbf{x}^{(t)}\|_{\mathbf{w}, S^{(t)}} \\ &\stackrel{(i)}{\leq} \left(\frac{1 - \varepsilon/2}{1 - \varepsilon/4}\right) \mathbf{w}_{S^{(t)}}^\top \mathbf{x}_{S^{(t)}, \gamma}^\mu + \frac{\varepsilon}{16} \cdot Z_{S^{(t)}, \gamma}^\mu \\ &\stackrel{(ii)}{\leq} \left(\frac{1 - \varepsilon/2}{1 - \varepsilon/4}\right) \mathbf{w}_{S^{(t)}}^\top \mathbf{x}_{S^{(t)}, \gamma}^\mu + \frac{\varepsilon}{16(1 - \varepsilon/8)} \mathbf{w}_{S^{(t)}}^\top \mathbf{x}_{S^{(t)}, \gamma}^\mu \\ &\stackrel{(iii)}{\leq} \left(1 - \frac{\varepsilon}{8}\right) \mathbf{w}_{S^{(t)}}^\top \mathbf{x}_{S^{(t)}, \gamma}^\mu, \end{aligned}$$

where (i) is by (4.19), (ii) is by Lemma 4.4 with accuracy parameter  $\varepsilon'$ , and (iii) uses  $\frac{1 - \varepsilon/2}{1 - \varepsilon/4} \leq 1 - \varepsilon/4$  for  $\varepsilon \in (0, 1)$ . The theorem then follows from Lemma 4.6 with accuracy parameter  $\varepsilon'$  and the same reasoning that proves Lemma 3.1.  $\square$

**4.4 Near-Optimality of Entropy Regularization** We have shown earlier in this section that Algorithm 1 with the entropy regularization strategy Algorithm 2 solves the decremental linear optimization problem with at most  $O(\log^2 d/\varepsilon^2)$  calls to **Rebuild()** before the optimal linear objective drops from  $d \cdot \alpha$  to  $\alpha$ . Complementing this result, in this section we show that this bound is near optimal for a certain range of  $\varepsilon$ .

More specifically, we focus on the special case of unweighted bipartite matching, i.e., when  $\mathcal{X} = \mathcal{P}_G = \mathcal{M}_G$  for some bipartite graph  $G$ , and consider any algorithm that implements the **Rebuild()** subroutine. In this decremental unweighted bipartite matching problem, Theorem 3.1 gives an upper bound of  $O(\log^2 n/\varepsilon^2)$  on the number of calls to **Rebuild()**, where  $n$  is the number of vertices. The following theorem establishes an  $\Omega(\log^2(\varepsilon^2 n)/\varepsilon^2)$  lower bound in the regime of  $\varepsilon \geq \Omega(1/\sqrt{n})$  against an output-adaptive adversary.

**THEOREM 4.1.** *For any  $n \in \mathbb{N}$ ,  $2/\sqrt{n} \leq \varepsilon \leq 1/4$ , and implementation of **Rebuild()** in Algorithm 1, there exists a bipartite graph  $G$  with  $n$  vertices on each side and an output-adaptively chosen sequence of deletions that when  $\mathcal{X} \stackrel{\text{def}}{=} \mathcal{P}_G$  Algorithm 1 calls **Rebuild()**  $\Omega(\log^2(\varepsilon^2 n) \cdot \varepsilon^{-2})$  times before  $G$  is empty.*

We first introduce a graph structure  $G_k$  that will repeatedly appear during the deletion process.

**DEFINITION 4.2.** *For  $k \in \mathbb{N}$ ,  $G_k$  is a bipartite graph with  $k$  vertices on each side, and the edge set  $E(G_k)$  of  $G_k$  is  $\{\{i^\ell, j^r\} \mid 1 \leq j \leq i \leq k\}$ .*

It is straightforward to check that  $G_k$  has a unique maximum matching  $M = \{\{i^\ell, i^r\} \mid 1 \leq i \leq k\}$ . The following lemma shows a generalization of the observation, that any large enough fractional matching has a mass concentration on edges with small differences in their endpoint labels.

LEMMA 4.8. *For  $k \in \mathbb{N}$  and  $\eta, \delta > 0$ , any fractional matching  $\mathbf{x}$  of  $G_k$  with matching size  $\|\mathbf{x}\|_1 \geq (1 - \eta)k$  satisfies that*

$$\sum_{0 \leq i-j \leq \delta k} \mathbf{x}_{\{i,j\}} \geq (1 - (1 + \delta^{-1})\eta) \cdot k.$$

*Proof.* By Definition 4.2, for any edge  $\{i^\ell, j^r\}$  in  $G_k$ , we have  $i - j \geq 0$ . Also, we can upper-bound the weighted sum of difference by

$$\sum_{\{i^\ell, j^r\} \in E(G_k)} (i - j) \mathbf{x}_{\{i^\ell, j^r\}} = \sum_{i \in [k]} i \cdot \mathbf{x}(i^\ell) - \sum_{j \in [k]} j \cdot \mathbf{x}(j^r) \leq \sum_{j \in [k]} j \cdot (1 - \mathbf{x}(j^r)) \leq k(k - \|\mathbf{x}\|_1) \leq \eta k^2.$$

By Markov's inequality,  $\sum_{i-j > \delta k} \mathbf{x}_{\{i^\ell, j^r\}} \leq \eta \delta^{-1} k$ , and thus  $\sum_{0 \leq i-j \leq \delta k} \mathbf{x}_{\{i^\ell, j^r\}} \geq (1 - (1 + \delta^{-1})\eta) \cdot k$ .  $\square$

Now we are ready to construct an output-adaptive adversary to achieve the previously claimed lower bound.

*Proof.* [Proof of Theorem 3.2] Let  $\mathcal{A}$  be an instance of Algorithm 1 with the given implementation of `Rebuild()`. The adversarial input graph to the algorithm is  $G_n$ , and the adversary works in phases. At the beginning of the  $t$ -th phase, let  $k_t$  be the largest number such that  $G_{k_t}$  is a subgraph of the current graph. The adversary will guarantee that  $k_t = (1 - 4\varepsilon)^t \cdot n$ . In the  $t$ -th phase, the adversary will cause  $\Omega(\log(\varepsilon^2 k_t)/\varepsilon)$  rebuilds in this phase. Before  $k_t$  reaches  $4/\varepsilon^2$ , there will be at least  $\log_{1-4\varepsilon}(4/(\varepsilon^2 n)) = \Omega(\log(\varepsilon^2 n)/\varepsilon)$  phases, achieving a

$$\sum_{t=0}^{\Omega(\log(\varepsilon^2 n)/\varepsilon)} \Omega(\log(\varepsilon^2 k_t)/\varepsilon) = \sum_{t=0}^{\Omega(\log(\varepsilon^2 n)/\varepsilon)} \Omega(\log(\varepsilon^2 (1 - 4\varepsilon)^t n)/\varepsilon) = \Omega(\log^2(\varepsilon^2 n)/\varepsilon^2)$$

lower bound on the total number of rebuilds. In the remainder of the proof, we focus on a single phase  $t$ , and show how the adversary can cause  $\Omega(\log(\varepsilon^2 k_t)/\varepsilon)$  rebuilds in this phase.

---

**Algorithm 3:** The adversarially chosen sequence of deletions

---

```

1 Let  $G \leftarrow G_n$  be the initial graph and feed it to  $\mathcal{A}$ .
2 Let  $t \leftarrow 0$ .
3 while  $k_t \stackrel{\text{def}}{=} (1 - 4\varepsilon)^t$  satisfies  $k_t > 4/\varepsilon^2$  do                                     // phase  $t$ 
4   Identify a subgraph  $G_{k_t} \subseteq G$ , delete  $G \setminus G_{k_t}$ , and re-label vertices so that
      $E(G) = \{\{i^\ell, j^r\} : 1 \leq j \leq i \leq k_t\}$ .                                     // preprocessing
5   Let  $\mathcal{M} \stackrel{\text{def}}{=} \{M_p : 0 \leq p \leq \varepsilon k_t/2\}$ , where  $M_p \stackrel{\text{def}}{=} \{\{i^\ell, j^r\} : i = j + p\}$ .
6   while  $|\mathcal{M}| > 1/\varepsilon$  do                                                         // regular deletion
7     Let  $\mathbf{x}$  be the current matching output by  $\mathcal{A}$ .
8     Choose  $r \stackrel{\text{def}}{=} 2\varepsilon|\mathcal{M}|$  matchings  $M_{p_1}, \dots, M_{p_r}$  such that  $\|\mathbf{x}_{M_{p_1}}\|_1 + \dots + \|\mathbf{x}_{M_{p_r}}\|_1 \geq \varepsilon/2 \cdot k_t$ .
       // guaranteed by Lemma 4.8
9     Delete  $M_{p_1} \cup \dots \cup M_{p_r}$  and set  $\mathcal{M} = \mathcal{M} \setminus \{M_{p_j} : 1 \leq j \leq r\}$ .       // cause a rebuild
10   $t \leftarrow t + 1$ .
```

---

**Preprocessing.** At the beginning of phase  $t$ , by definition of  $k_t$ ,  $G_{k_t}$  is a subgraph of the current graph  $G$ . The adversary will first delete edges outside this subgraph, then relabel vertices in this subgraph on each side from 1 to  $k_t$  in a way that the current set of edges is  $\{\{i^\ell, j^r\} \mid 1 \leq j \leq i \leq k_t\}$ . The adversary will then delete edges  $\{\{i^\ell, j^r\} \mid j + \varepsilon k_t/2 < i < j + 4\varepsilon k_t\}$ . After that, regular deletion starts.

**Regular Deletion.** A regular deletion starts after a preprocessing finishes. During the regular deletion of phase  $t$ , the adversary will only delete edges in the subgraph  $G_{k_t}^\varepsilon \stackrel{\text{def}}{=} \{\{i^\ell, j^r\} \mid j \leq i \leq j + \varepsilon k_t/2\}$  of  $G_{k_t}$ , and the deletion continues until the maximum matching size  $M^*(G)$  becomes less than  $(1 - \varepsilon/2)k_t$ . After that, the current phase ends. Since no edges in the subgraph  $\{\{i^\ell, j^r\} \mid i \geq j + 4\varepsilon k_t\}$  are deleted,  $G_{(1-4\varepsilon)k_t} = G_{k_{t+1}}$  is a subgraph of the current graph at the beginning of phase  $t + 1$ . It remains to show that  $\Omega(\log(\varepsilon^2 k_t)/\varepsilon)$  rebuilds could be caused during the deletion described above in  $G_{k_t}^\varepsilon$ .

Note that  $G_{k_t}^\varepsilon$  is the union of  $\Theta(\varepsilon k_t)$  matchings where the  $p$ -th matching is  $M_p \stackrel{\text{def}}{=} \{\{i^\ell, j^r\} \mid i = j + p\}$ , each of size at least  $(1 - \varepsilon/2)k_t$ . Therefore, as long as one of the  $M_p$ 's remains intact, we have  $(1 - \varepsilon/2)k_t \leq M^*(G) \leq k_t$ . This implies that the fractional matching  $\mathbf{x}$  that  $\mathcal{A}$  maintains must have size at least

$$\|\mathbf{x}\|_1 \geq \left(1 - \frac{\varepsilon}{2}\right)^2 k_t - \frac{\varepsilon}{2} \cdot k_t \geq \left(1 - \frac{3\varepsilon}{2}\right) k_t$$

throughout this phase, since by definition of [Algorithm 1](#)  $\mathbf{x}$  was a  $(1 - \varepsilon/2)$ -approximate matching since the last rebuild, and after  $\frac{\varepsilon}{2} \cdot k_t$  units of mass get deleted, a rebuild must be caused.

By the above argument it suffices to delete  $\varepsilon/2 \cdot k_t$  units of mass to cause a rebuild from  $\mathcal{A}$ . Applying [Lemma 4.8](#) with  $\eta = 3\varepsilon/2$  and  $\delta = 4\varepsilon$  on the fractional matching  $\mathbf{x}$ , we know that at least  $k_t/4$  units of mass are on  $G_{k_t}^\varepsilon$ , i.e.,  $\|\mathbf{x}_{E(G_{k_t}^\varepsilon)}\|_1 \geq k_t/4$ . Thus via an averaging argument the adversary can delete  $2\varepsilon$  fraction of the matchings  $M_p$ 's to cause a rebuild. After the rebuild, we repeat the same argument again, choosing  $2\varepsilon$  fraction of the remaining  $M_p$ 's to cause another rebuild. Note that after the first batch of deletions, the current graph is no longer  $G_{k_t}$ . Nevertheless, we can still apply [Lemma 4.8](#) by interpreting the future  $\mathbf{x}$ 's as a fractional matching on  $G_{k_t}$  by assigning a mass of zero on deleted edges. This shows that the mass of  $\mathbf{x}$  is still concentrated on the remaining  $M_p$ 's. Before the number of intact matchings in  $M_p$ 's reaches  $1/(2\varepsilon)$ , there will be at least  $\log_{1-2\varepsilon}(4/(\varepsilon^2 k_t)) = \Omega(\log(\varepsilon^2 k_t)/\varepsilon)$  reconstructions, if  $\varepsilon k_t/2 \geq 1/(2\varepsilon)$  or equivalently  $k_t \leq 4/\varepsilon^2$ . This completes the proof.  $\square$

## 5 Decremental Algorithms for Fractional Matching

In [Section 4](#) we showed that the decremental linear optimization problem reduces to computing an approximate solution to the entropy-regularized problem to moderate accuracy. In the particular, in the special case of decremental bipartite matching, i.e., when  $\mathcal{X}$  is the matching polytope  $\mathcal{M}_G$ , it suffices to solve the entropy-regularized matching problem recalled below.

**PROBLEM 3.3. (ENTROPY-REGULARIZED MATCHING)** *Given an  $n$ -vertex  $m$ -edge graph  $G = (V, E)$  with edge weights  $\mathbf{w} \in \mathbb{N}^E$  bounded by  $W = \text{poly}(n)$ , a trade-off parameter  $\mu \leq 1$ , a  $\gamma \in \mathbb{R}_{\geq 0}$  such that  $M_{\mathbf{w}}^*(G) \leq \gamma \leq n^3 W$ , and an accuracy parameter  $\varepsilon \in (0, 1)$ , the entropy-regularized matching problem is to compute a  $(1 - \varepsilon)$ -approximate solution  $\mathbf{x} \in \mathcal{M}_G$  to*

$$(3.6) \quad \max_{\mathbf{x} \in \mathcal{M}_G} \left\{ \mathbf{w}^\top \mathbf{x} + \mu \cdot \sum_{e \in E} \mathbf{w}_e \mathbf{x}_e \log \frac{\gamma}{\mathbf{w}_e \mathbf{x}_e} \right\}.$$

Consequently, in the remainder of the section we focus on solving [Problem 3.3](#). We prove the following theorem.

**THEOREM 5.1.** *For any  $\varepsilon \geq \tilde{\Omega}(n^{-1/2})$ , there are randomized  $\tilde{O}(m\varepsilon^{-6} + n\varepsilon^{-13})$  and  $\hat{O}(m\varepsilon^{-5})$  time algorithms that solve [Problem 3.3](#) w.h.p.*

Our approaches for [Theorem 3.3](#) follow from a generalization of the MWU-style algorithm of [5](#) that solves the uncapacitated and capacitated versions of weighted matching. Given a graph  $G = (V, E)$  and consider a downward closed convex set  $\mathcal{P} \subseteq \mathbb{R}_{\geq 0}^E$  of interest. By modularizing and generalizing the framework of [5](#), we derive the following [Lemma 5.1](#) on concave function optimization over the matching polytope. Informally speaking, [Lemma 5.1](#) gives a reduction from approximately maximizing a concave function  $\mathcal{P}$  together with odd-set constraints to the same task just in  $\mathcal{P}$  by increasing the dependence on  $\varepsilon^{-1}$ . We then, in [Sections 5.1](#) and [5.2](#) respectively, instantiate the framework of [Lemma 5.1](#) in two different ways with the entropy-regularized function to obtain the two runtimes stated in [Theorem 3.3](#).

Let  $\kappa_{\mathcal{P}}$  be the minimum number such that  $\mathcal{P} \subseteq \kappa_{\mathcal{P}} \cdot \mathcal{M}_G$ , which for both  $\mathcal{P}_G$  and  $\mathcal{P}_G^c$  that we will consider later in this section are  $O(1)$ . A function  $f : \mathbb{R}_{\geq 0}^E \rightarrow \mathbb{R}_{-\infty}$  is *coordinate-separable concave/linear* if  $f$  is of the form  $f(\mathbf{x}) = \sum_{e \in E} f_e(\mathbf{x}_e)$  where each  $f_e : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{-\infty}$  is concave/linear. Specifically, if  $\ell : \mathbb{R}_{\geq 0}^E \rightarrow \mathbb{R}$  is coordinate-separable linear, we use  $\ell \in \mathbb{R}^E$  to denote the linear coefficients, i.e.,  $\ell(\mathbf{x}) \stackrel{\text{def}}{=} \ell^\top \mathbf{x}$ .

DEFINITION 5.1. For  $\beta \geq 1$  and  $0 \leq \zeta \leq 1$ , an algorithm  $\mathcal{A}$  is a  $(\beta, T_{\mathcal{A}}, \zeta)$ -oracle for  $(\mathcal{P}, f)$ , where  $\mathcal{P} \subseteq \mathbb{R}_{\geq 0}^E$  is convex and downward closed and  $f$  is concave, if given any concave function  $g$  of the form  $g = f - \ell$  for some coordinate-separable linear  $\ell$  such that each  $\ell_e$  is polynomially bounded, with an additional guarantee that

$$(5.20) \quad \max_{\mathbf{x} \in \mathcal{P}} g(\mathbf{x}) \geq \frac{1}{\text{poly}(n)},$$

it outputs an  $\mathbf{x}_g \in \beta \cdot \mathcal{P}$  with  $g(\mathbf{x}_g) \geq \zeta \cdot \max_{\mathbf{x} \in \mathcal{P}} g(\mathbf{x})$  in  $T_{\mathcal{A}}(n, m) = \Omega(m)$  time.

We defer the proof and discussion on modularization and explicit dependence on  $\varepsilon^{-1}$  of the following Lemma 5.1 to Section A for completeness. It is worth noting that the two stated runtimes below are obtained by leveraging recently developed algorithms in certain components (specifically the computation of Gomory-Hu trees) of [5].

LEMMA 5.1. (ADAPTED FROM [5]) For any  $\varepsilon \geq \tilde{\Omega}(n^{-1/2})$ , downward closed and convex  $\mathcal{P} \subseteq \mathbb{R}_{\geq 0}^E$ , and concave function  $f : \mathbb{R}_{\geq 0}^E \rightarrow \mathbb{R}_{-\infty}$  such that  $\max_{\mathbf{x} \in \mathcal{P} \cap \mathcal{M}_G} f(\mathbf{x}) \geq 1$ , given an  $(\beta, T_{\mathcal{A}}, \zeta)$ -oracle  $\mathcal{A}$  for  $(\mathcal{P}, f)$  where  $\kappa_{\mathcal{P}} \cdot \beta < n$ , there are randomized  $\tilde{O}((T_{\mathcal{A}}(n, m) + m\varepsilon^{-1} + n\varepsilon^{-9}) \cdot \kappa_{\mathcal{P}}\beta\varepsilon^{-4})$  and  $\hat{O}((T_{\mathcal{A}}(n, m) + m\varepsilon^{-1}) \cdot \kappa_{\mathcal{P}}\beta\varepsilon^{-4})$  time algorithms that w.h.p. compute an  $\mathbf{x}_f \in (\beta \cdot \mathcal{P}) \cap \mathcal{M}_G$  such that  $f(\mathbf{x}_f) \geq \zeta(1 - \varepsilon) \max_{\mathbf{x} \in \mathcal{P} \cap \mathcal{M}_G} f(\mathbf{x})$ .

For instance, if we set  $\mathcal{P}$  to be  $\mathcal{P}_G$ , then  $\kappa_{\mathcal{P}} = O(1)$  since for any  $\mathbf{x} \in \mathcal{P}_G$  we have

$$\sum_{e \in E[B]} x_e \leq \frac{\sum_{v \in B} \sum_{e \in E_v} x_e}{2} \leq \frac{|B|}{2} \leq 2 \cdot \left\lfloor \frac{|B|}{2} \right\rfloor$$

for all odd sets  $B \in \mathcal{O}_G$ . As such, optimizing concave functions (specifically the entropy-regularized objective) over the general matching polytope reduces to optimizing over  $\mathcal{P}_G$ , which may be substantially simpler to solve. In Sections 5.1 and 5.2 we will develop two incomparable algorithms that both follow Lemma 5.1. These ultimately assemble into the two runtimes of Theorem 3.3 in Section 5.3.

**5.1 Almost-Linear Time Oracle via Convex Flow Algorithms** Using the convex flow algorithm of [33], we can indeed optimize any “efficiently-computable” concave function over  $\mathcal{P}_G$ . We then use that result to obtain an algorithm optimizing concave functions over the general matching polytope in Lemma 5.3, with the help of Lemma 5.1. The algorithm of [33] requires as input self-concordant barriers on the domain  $\{(x, y) : y > h_e(x)\}$ , where  $h_e$  is the convex edge weight, that satisfy assumptions detailed in [33, Assumption 10.2] which ensures the numbers encountered during the algorithm are quasi-polynomially bounded.

LEMMA 5.2. ([33, THEOREM 10.13] DERANDOMIZED BY [66]) Given  $m$ -edge graph  $G = (V, E)$ , demands  $\mathbf{d} \in \mathbb{R}^V$ , convex  $h_e : \mathbb{R} \rightarrow \mathbb{R}_{\infty}$ , and  $\nu$ -self-concordant barriers  $\psi_e(x, y)$  on the domain  $\{(x, y) : y > h_e(x)\}$  that satisfy [33, Assumption 10.2], there is a deterministic  $m^{1+o(1)}$  time algorithm that computes a flow  $\mathbf{f} \in \mathbb{R}^E$  with  $B_G^\top \mathbf{f} = \mathbf{d}$  such that

$$h(\mathbf{f}) \leq \min_{B_G^\top \mathbf{f}^* = \mathbf{d}} h(\mathbf{f}^*) + \exp(-\log^C m)$$

for any fixed constant  $C > 0$ , where  $h(\mathbf{f}) \stackrel{\text{def}}{=} \sum_{e \in E} h_e(\mathbf{f}_e)$ .

The following is our main lemma in the section which turns the flow algorithm above into an algorithm for optimizing concave functions over  $\mathcal{M}_G$ . For simplicity of the construction of barriers, we allow the functions  $f_e$  to be decomposed into  $\tilde{O}(1)$  portions, each with its own barrier.

LEMMA 5.3. Suppose we are given a graph  $G = (V, E)$ ,  $\varepsilon \geq \tilde{\Omega}(n^{-1/2})$ , and a coordinate-separable concave function  $f : \mathbb{R}_{\geq 0}^E \rightarrow \mathbb{R}_{-\infty}$  such that  $\max_{\mathbf{x} \in \mathcal{M}_G} f(\mathbf{x}) \geq 1$ , where the function on each edge is given as  $f_e(x) \stackrel{\text{def}}{=} f_e^{(1)}(x) + \dots + f_e^{(k_e)}(x)$  for some  $k_e = \tilde{O}(1)$ , each equipped with a  $\nu$ -self-concordant barrier  $\psi_e^{(i)}(x, y)$  on the domain  $\{(x, y) : y > -f_e^{(i)}(x)\}$  that satisfy [33, Assumption 10.2]. Then, there is a randomized  $\tilde{O}(m\varepsilon^{-5})$  time algorithm that w.h.p. computes an  $\mathbf{x} \in \mathcal{M}_G$  such that

$$f(\mathbf{x}) \geq (1 - \varepsilon) \max_{\mathbf{x}^* \in \mathcal{M}_G} f(\mathbf{x}^*).$$

*Proof.* By Lemma 5.1, to prove Lemma 5.3 it suffices to provide a  $(1, m^{1+o(1)}, 1 - \varepsilon/2)$ -oracle for  $(\mathcal{P}_G, f)$ . We achieve this by reducing the optimization problem over the relaxed matching polytope  $\mathcal{P}_G$  to a minimum cost circulation<sup>19</sup> problem as follows. Consider the following directed graph  $\tilde{G} = (\tilde{V}, \tilde{E})$  with  $\tilde{V} \stackrel{\text{def}}{=} \{v_{\text{in}} : v \in V\} \cup \{v_{\text{out}} : v \in V\}$  and

$$\tilde{E} \stackrel{\text{def}}{=} \{(v_{\text{in}}, v_{\text{out}}) : v \in V\} \cup \{(u_{\text{out}}, v_{\text{in}}) : \{u, v\} \in E\} \cup \{(v_{\text{out}}, u_{\text{in}}) : \{u, v\} \in E\}.$$

In other words, each vertex  $v$  is split into two copies  $v_{\text{in}}$  and  $v_{\text{out}}$ , and each edge  $\{u, v\}$  is directed from  $u_{\text{out}}$  to  $v_{\text{in}}$  and from  $v_{\text{out}}$  to  $u_{\text{in}}$ . For each  $e = \{u, v\}$  in  $E$ , let  $e^{(1)}$  and  $e^{(2)}$  denote  $(u_{\text{out}}, v_{\text{in}}) \in \tilde{E}$  and  $(v_{\text{out}}, u_{\text{in}}) \in \tilde{E}$  respectively. Let  $e^{(v)} \in \tilde{E}$  for  $v \in V$  be  $(v_{\text{in}}, v_{\text{out}})$ .

Let  $g \stackrel{\text{def}}{=} f - \ell$  be the coordinate-separable concave function that the oracle needs to optimize which we write as  $g(\mathbf{x}) \stackrel{\text{def}}{=} \sum_{e \in E} g_e(\mathbf{x}_e)$ . We translate the concave edge weights  $g_e$ 's on  $E$  to convex  $h_e$ 's on  $\tilde{E}$  as follows. For each  $e \in E$ , let  $h_{e(1)}, h_{e(2)} : \mathbb{R} \rightarrow \mathbb{R}_{\infty}$  be defined as

$$h_{e(1)}(x) = h_{e(2)}(x) \stackrel{\text{def}}{=} \frac{-g_e(x)}{2} + \phi(x) = \frac{-f_e(x)}{2} + \frac{\ell_e(x)}{2} + \phi(x),$$

where

$$\phi(x) = \begin{cases} 0, & \text{if } 0 \leq x \leq 1 \\ \infty, & \text{otherwise} \end{cases}$$

is a convex regularizer which enforces each edge having flow at most one. For each  $v \in V$ , let  $h_{e(v)} : \mathbb{R} \rightarrow \mathbb{R}_{\infty}$  be defined as

$$h_{e(v)}(x) \stackrel{\text{def}}{=} \phi(x).$$

Observe that we can translate between circulations in  $\tilde{G}$  and relaxed fractional matchings in  $G$  as follows. For a circulation  $\mathbf{f} \in \mathbb{R}^{\tilde{E}}$  with  $h(\mathbf{f}) < \infty$ , by definition of the edge weights we may assume  $0 \leq \mathbf{f}_e \leq 1$  for every  $e \in \tilde{E}$ . Setting  $\mathbf{x}_e \stackrel{\text{def}}{=} \frac{\mathbf{f}_{e(1)} + \mathbf{f}_{e(2)}}{2}$  for each edge  $e \in E$ , we see that by concavity of  $g_e$  that  $g_e(\mathbf{x}_e) \geq -h_{e(1)}(\mathbf{f}_{e(1)}) - h_{e(2)}(\mathbf{f}_{e(2)})$ . Conversely, for any relaxed fractional matching  $\mathbf{x} \in \mathcal{P}_G$ , let  $\mathbf{f}_{e(1)} = \mathbf{f}_{e(2)} \stackrel{\text{def}}{=} \mathbf{x}_e$  and  $\mathbf{f}_{e(v)} \stackrel{\text{def}}{=} \mathbf{x}(v)$ . It can be easily checked that  $\mathbf{f}$  is a circulation and has weight  $h(\mathbf{f}) = -f(\mathbf{x})$ . This shows that the optimal values of these two problems are the same (up to negation). Consequently, if we get a  $\delta$ -additive approximate minimizing circulation  $\mathbf{f}$ , then the corresponding relaxed fractional matching  $\mathbf{x}$  would satisfy

$$g(\mathbf{x}) \geq -h(\mathbf{f}) \geq -\min_{\mathbf{B}_G^\top \mathbf{f}^* = \mathbf{0}} h(\mathbf{f}^*) - \delta = \max_{\mathbf{x}^* \in \mathcal{P}_G} g(\mathbf{x}^*) - \delta.$$

We will then choose  $\delta$  to be sufficiently small to make  $\mathbf{x}$  a  $(1 - \varepsilon/2)$ -approximate maximizer of  $g$  in  $\mathcal{P}_G$ .

To apply the convex flow algorithm of Lemma 5.2 to minimize  $h(\mathbf{f})$ , we need to provide self-concordant barriers for the edge weights. Consider edges  $e^{(1)}, e^{(2)} \in \tilde{E}$  for some  $e \in E$ . The edge weights  $h_{e(1)}$  and  $h_{e(2)}$  can be divided into  $k_e + 2$  parts which we handle by splitting  $e^{(1)}$  and  $e^{(2)}$  further into paths of length  $k_e + 2$  (recall that  $k_e$  the function  $f_e$  is given to Lemma 5.3 as  $k_e$  portions, each with its own barrier). Among the first  $k_e$  parts, the  $i$ -th of which has weight  $\frac{-f_e^{(i)}(x)}{2}$  which we use  $\tilde{\psi}(x, y) \stackrel{\text{def}}{=} \psi_e^{(i)}(x, 2y)$  as the barrier, where recall that  $\psi_e^{(i)}$  is the given barrier to the  $i$ -th portion of edge  $e$ ; the second last part has weight  $\frac{\ell_e(x)}{2}$  which we use  $\tilde{\psi}(x, y) \stackrel{\text{def}}{=} -\log\left(y - \frac{\ell_e(x)}{2}\right)$

<sup>19</sup>A circulation is a flow  $\mathbf{f}$  such that routes the demand  $\mathbf{0}$ , i.e.,  $\mathbf{B}_G^\top \mathbf{f} = \mathbf{0}$ .



as the barrier; and the last part has weight  $\phi(x)$  which we use  $\tilde{\psi}(x, y) \stackrel{\text{def}}{=} x^{-\alpha} + (1-x)^{-\alpha}$  for  $\alpha \stackrel{\text{def}}{=} \frac{1}{1000 \log mU}$  as the barrier. The barrier  $-\log\left(y - \frac{\ell_e(x)}{2}\right)$  is the same barrier that [33, Theorem 10.16] used for linear functions (note that  $\ell_e$  is polynomially bounded); the barrier  $x^{-\alpha} + (1-x)^{-\alpha}$  is the same barrier that [33] used in their min-cost flow algorithm to enforce capacity constraints. Both of the barriers were shown to satisfy the assumption in [33]. The barrier for  $\phi(x)$  is also used for edges  $e^{(v)} \in \tilde{E}$ .

This gives us an exact reduction from maximizing concave weights over  $\mathcal{P}_G$  to finding a circulation minimizing convex weights in a graph with  $\tilde{O}(m)$  edges. Applying Lemma 5.2 with the constant  $C$  chosen in such a way that  $\exp(-\log^C m) \leq \varepsilon/2 \cdot \max_{\mathbf{x} \in \mathcal{P}_G} g(\mathbf{x}) = \Theta(1/\text{poly}(n))$  thus results in a  $(1, m^{1+o(1)}, 1-\varepsilon/2)$ -oracle for  $(\mathcal{P}_G, f)$ . The theorem then follows from Lemma 5.1 with accuracy parameter  $\varepsilon/2$ .  $\square$

**5.2 Near-Linear Time Reduction to Linear Optimization** To remove the  $n^{o(1)}$  factors incurred by the almost-linear time flow algorithm in Lemma 5.3, we can instead reduce concave function maximization over the matching polytope directly to a capacity-constrained weighted matching problem at the cost of a larger  $\varepsilon^{-1}$  factor in the running time. For  $\mathbf{c} \in [0, 1]^E$ , let  $\Gamma_{\mathbf{c}} \stackrel{\text{def}}{=} \{0 \leq \mathbf{x}_e \leq \mathbf{c}_e, \forall e \in E\} \subseteq \mathbb{R}_{\geq 0}^E$  be the capacity-constrained polytope. Let  $\mathcal{M}_G^{\mathbf{c}} \stackrel{\text{def}}{=} \mathcal{M}_G \cap \Gamma_{\mathbf{c}}$  and  $\mathcal{P}_G^{\mathbf{c}} \stackrel{\text{def}}{=} \mathcal{P}_G \cap \Gamma_{\mathbf{c}}$ .

**PROBLEM 5.1.** *In the capacity-constrained weighted matching problem, we are given a graph  $G = (V, E)$ , an accuracy  $\varepsilon > 0$ , edge weights  $\mathbf{w} \in \mathbb{R}_{\geq 0}^E$ , capacities  $\mathbf{c} \in [0, 1]^E$ , all polynomially bounded. The goal is find an  $\mathbf{x} \in \mathcal{M}_G^{\mathbf{c}}$  such that*

$$\mathbf{w}^\top \mathbf{x} \geq (1 - \varepsilon) \max_{\mathbf{x}^* \in \mathcal{M}_G^{\mathbf{c}}} \mathbf{w}^\top \mathbf{x}^*.$$

To solve Problem 5.1, we use the following constant-approximate algorithm to the “relaxed” capacitated  $\mathbf{b}$ -matching problem as an oracle and apply Lemma 5.1. Their algorithm works for multigraphs with integral demands and capacities.

**LEMMA 5.4.** ([5, THEOREM 13]) *Given an  $m$ -edge multigraph  $G = (V, E)$ , edge weights  $\mathbf{w} \in \mathbb{R}_{\geq 0}^E$ , demands  $\mathbf{b} \in \mathbb{Z}_{\geq 0}^V$ , capacities  $\mathbf{c} \in \mathbb{Z}_{\geq 0}^E$ , all polynomially bounded entrywise, there is a deterministic  $\tilde{O}(m)$  time algorithm that obtains a  $1/8$ -approximate maximizer to the following “relaxed” capacitated  $\mathbf{b}$ -matching problem:*

$$\begin{aligned} & \text{maximize} && \mathbf{w}^\top \mathbf{x} \\ (5.21) \quad & \text{subject to} && \mathbf{x}(v) \leq \mathbf{b}_v, \quad \forall v \in V, \\ & && 0 \leq \mathbf{x}_e \leq \mathbf{c}_e, \quad \forall e \in E. \end{aligned}$$

The criterion of  $\mathbf{b}$  and  $\mathbf{c}$  being integral in Lemma 5.4 can be relaxed via scaling.

**COROLLARY 5.1.** *For polynomially bounded  $\mathbf{b} \in \mathbb{R}_{\geq 0}^V$  and  $\mathbf{c} \in \mathbb{R}_{\geq 0}^E$  there is an  $\tilde{O}(m)$  time algorithm that obtains a  $1/16$ -approximate maximizer to (5.21).*

*Proof.* Since  $\mathbf{b}$  and  $\mathbf{c}$  are polynomially bounded, we can scale them to integers by replacing each  $\mathbf{b}_v$  and  $\mathbf{c}_e$  with  $\lfloor \mathbf{b}_v / \mathbf{b}_{\min} \rfloor$  and  $\lfloor \mathbf{c}_e / \mathbf{c}_{\min} \rfloor$ , respectively. Observe that  $\lfloor \mathbf{b}_v / \mathbf{b}_{\min} \rfloor \geq \mathbf{b}_v / (2\mathbf{b}_{\min})$  and  $\lfloor \mathbf{c}_e / \mathbf{c}_{\min} \rfloor \geq \mathbf{c}_e / (2\mathbf{c}_{\min})$ , and thus we only lose a factor of 2 in the approximation ratio from Lemma 5.4.  $\square$

Using Corollary 5.1 as an oracle, Lemma 5.1 now implies the following algorithm for Problem 5.1

**LEMMA 5.5.** *There is a randomized  $\tilde{O}(m\varepsilon^{-5} + n\varepsilon^{-13})$  time algorithm for  $\varepsilon \geq \tilde{\Omega}(n^{-1/2})$  that solves Problem 5.1 w.h.p.*

*Proof.* Consider the function  $f_{\mathbf{w}}(\mathbf{x}) \stackrel{\text{def}}{=} \mathbf{w}^\top \mathbf{x}$ . By scaling  $\mathbf{w}$  we may assume  $\max_{e \in E} \mathbf{w}_e \mathbf{c}_e \geq 1$  and therefore  $\max_{\mathbf{x} \in \mathcal{P}_G^{\mathbf{c}} \cap \mathcal{M}_G} f_{\mathbf{w}}(\mathbf{x}) \geq 1$ . By running Corollary 5.1 and returning the vector  $\mathbf{x}$  it outputs by 16, we get a  $(16, \tilde{O}(m), 1)$ -oracle for  $(\mathcal{P}_G^{\mathbf{c}}, f_{\mathbf{w}})$ . Since  $\mathcal{M}_G^{\mathbf{c}} = \mathcal{P}_G^{\mathbf{c}} \cap \mathcal{M}_G$ , the lemma follows from Lemma 5.1.  $\square$

Note that an immediate corollary of the above [Lemma 5.5](#) is that we can solve the subproblem in [\[1\]](#) of finding an approximate maximum matching obeying capacity and odd-set constraints in  $\tilde{O}(m \cdot \text{poly}(\varepsilon^{-1}))$  time. However, we remark again this does not suffice to make their framework run in  $\tilde{O}(m \cdot \text{poly}(\varepsilon^{-1}))$  completely, as a dual variable to [Problem 5.1](#) is still required to identify the set of critical edges along which the capacity is increased.

We now present the reduction from maximizing convex objective to [Problem 5.1](#) by approximating the objective with piecewise linear functions, thereby effectively splitting each edge into  $O(\varepsilon^{-1})$  copies of different capacities and weights. Similar approaches were used before, e.g., in [\[57\]](#).

**LEMMA 5.6.** *Given an  $n$ -vertex  $m$ -edge graph  $G = (V, E)$  and a coordinate-separable concave function  $f : \mathbb{R}_{\geq 0}^E \rightarrow \mathbb{R}$  satisfying*

- (1)  $f_e(x)$  is polynomially bounded for  $x \geq 1/\text{poly}(n)$ ,
- (2) each  $f_e$  can be evaluated in  $O(1)$  time, and
- (3)  $z_e^* \stackrel{\text{def}}{=} \arg\max_{x \in [0,1]} f_e(x)$  is given and satisfies  $z_e^* \geq 1/\text{poly}(n)$ ,

for any  $\varepsilon \geq \tilde{\Omega}(n^{-1/2})$  there is a randomized  $\tilde{O}(m\varepsilon^{-6} + n\varepsilon^{-13})$  time algorithm that computes an  $\mathbf{x}_f \in \mathcal{M}_G$  such that w.h.p.,

$$f(\mathbf{x}_f) \geq (1 - \varepsilon) \max_{\mathbf{x} \in \mathcal{M}_G} f(\mathbf{x}).$$

*Proof.* Let us consider a fixed  $\varepsilon' = O(\varepsilon)$  that we will set later. For each edge  $e \in E$ , let  $r_e^{(0)} \stackrel{\text{def}}{=} z_e^*$  and  $r_e^{(i)} \stackrel{\text{def}}{=} r_e^{(i-1)}/(1 + \varepsilon')$  for  $i \in \{1, \dots, k\}$ , where  $k \stackrel{\text{def}}{=} \left\lceil \log_{1+\varepsilon'} \frac{n^2}{\varepsilon'} \right\rceil = \tilde{O}(\varepsilon^{-1})$ . Let  $r_e^{(k+1)} \stackrel{\text{def}}{=} 0$ . Splitting each edge  $e$  into  $k+1$  copies  $e^{(0)}, e^{(1)}, \dots, e^{(k)}$ , we get a graph  $G' = (V, E')$  with  $m' = \tilde{O}(m\varepsilon^{-1})$  edges. Define  $\mathbf{c}' \in \mathbb{R}^{E'}$  and  $\mathbf{w}' \in \mathbb{R}^{E'}$  with  $\mathbf{c}'_{e^{(i)}} \stackrel{\text{def}}{=} r_e^{(i)} - r_e^{(i+1)}$  and  $\mathbf{w}'_{e^{(i)}} \stackrel{\text{def}}{=} \frac{f_e(r_e^{(i)}) - f_e(r_e^{(i+1)})}{\mathbf{c}'_{e^{(i)}}}$  for each  $e \in E$  and  $i \in \{0, \dots, k\}$ . Note that  $\mathbf{c}_{e^{(i)}}$  and  $\mathbf{w}_{e^{(i)}}$  are both polynomially bounded by [\(1\)](#) and [\(3\)](#). Recall that  $\mathcal{M}_{G'}^{\mathbf{c}'}$  is the capacity-constrained matching polytope of  $G'$ . We show the following two claims.

**CLAIM 5.1.** *For any  $\mathbf{x}' \in \mathcal{M}_{G'}^{\mathbf{c}'}$ , the vector  $\mathbf{x} \in \mathbb{R}_{\geq 0}^E$  given by  $\mathbf{x}_e \stackrel{\text{def}}{=} \mathbf{x}'_{e^{(0)}} + \dots + \mathbf{x}'_{e^{(k)}}$  satisfies  $\mathbf{x} \in \mathcal{M}_G$  and  $f(\mathbf{x}) \geq \mathbf{w}'^\top \mathbf{x}'$ .*

*Proof.* That  $\mathbf{x} \in \mathcal{M}_G$  is immediate from  $\mathbf{x}' \in \mathcal{M}_{G'}^{\mathbf{c}'}$ . Let

$$\tilde{f}_e(x) \stackrel{\text{def}}{=} \int_0^x \tilde{\mathbf{w}}_e(y) dy, \quad \text{where} \quad \tilde{\mathbf{w}}_e(y) \stackrel{\text{def}}{=} \mathbf{w}'_{e^{(i)}} \text{ for } r_e^{(i+1)} \leq y < r_e^{(i)}$$

be the piecewise linearized version of  $f_e$ . By concavity of  $f_e$ , it holds that  $\tilde{f}_e(x) \leq f_e(x)$  for all  $x \in [0, z_e^*]$ . We have

$$f(\mathbf{x}) \geq \sum_{e \in E} \tilde{f}_e(\mathbf{x}_e) = \sum_{e \in E} \sum_{0 \leq i \leq k} \mathbf{w}'_{e^{(i)}} \cdot \min \left\{ \mathbf{c}'_{e^{(i)}}, \max \left\{ 0, \mathbf{x}_e - r_e^{(i+1)} \right\} \right\} \geq \mathbf{w}'^\top \mathbf{x}',$$

where the last inequality uses the fact that  $\mathbf{w}'_{e^{(k)}} \geq \mathbf{w}'_{e^{(k-1)}} \geq \dots \geq \mathbf{w}'_{e^{(0)}}$  by concavity of  $f_e$  so it is always better to saturate  $e^{(i)}$  before putting mass on  $e^{(i-1)}$ .  $\square$

**CLAIM 5.2.** *For any  $\mathbf{x} \in \mathcal{M}_G$ , there exists an  $\mathbf{x}' \in \mathcal{M}_{G'}^{\mathbf{c}'}$  with  $\mathbf{w}'^\top \mathbf{x}' \geq (1 - 2\varepsilon')f(\mathbf{x})$ .*

*Proof.* Let  $\tilde{\mathbf{x}} \in \mathbb{R}_{\geq 0}^{E'}$  by defined as

$$\tilde{\mathbf{x}}_{e^{(i)}} \stackrel{\text{def}}{=} \begin{cases} \min \left\{ \mathbf{c}'_{e^{(i)}}, \max \left\{ 0, \mathbf{x}_e - r_e^{(i+1)} \right\} \right\}, & \text{if } i < k, \\ \mathbf{c}'_{e^{(k)}}, & \text{if } i = k. \end{cases}$$

Observe that  $\tilde{\mathbf{x}}_{e^{(k)}} \leq \varepsilon'/n^2$  for every  $e \in E$  by definition. As such we have  $\tilde{\mathbf{x}}_{e^{(0)}} + \dots + \tilde{\mathbf{x}}_{e^{(k)}} \leq \mathbf{x}_e + \varepsilon'/n^2$ , which implies  $\tilde{\mathbf{x}} \in (1+\varepsilon')\mathcal{M}_{G'}^{\mathbf{c}'}$ , given  $\mathbf{x} \in \mathcal{M}_G$ . Letting  $0 \leq t_e \leq k$  be the smallest integer such that  $r_e^{(t_e)} \leq \max\{\mathbf{x}_e, \mathbf{c}'_{e^{(k)}}\}$ , we also have

$$\mathbf{w}'^\top \tilde{\mathbf{x}} \geq \sum_{e \in E} \sum_{i=t_e}^k \mathbf{w}'_{e^{(i)}} \mathbf{c}'_{e^{(i)}} \stackrel{(i)}{=} \sum_{e \in E \setminus F} f_e(r_e^{(t_e)}) \stackrel{(ii)}{\geq} \frac{1}{1+\varepsilon'} \sum_{e \in E} f_e(\mathbf{x}_e) \geq (1-\varepsilon')f(\mathbf{x}),$$

where (i) is by definition of  $\mathbf{w}'$  and  $\mathbf{c}'$ , and (ii) uses the fact that  $r_e^{(t_e)} \geq \mathbf{x}_e/(1+\varepsilon')$  by definition and concavity of  $f_e$ . Therefore, the vector  $\mathbf{x}' \stackrel{\text{def}}{=} \tilde{\mathbf{x}}/(1+\varepsilon')$  satisfies  $\mathbf{x}' \in \mathcal{M}_{G'}^{\mathbf{c}'}$  and  $\mathbf{w}'^\top \mathbf{x}' \geq (1-2\varepsilon')f(\mathbf{x})$ .  $\square$

Going back to the proof of [Lemma 5.6](#) with  $\varepsilon' \stackrel{\text{def}}{=} \varepsilon/4$ , we run [Lemma 5.5](#) on the split graph  $G'$  with edge weights  $\mathbf{w}'$  and capacities  $\mathbf{c}'$  to accuracy  $1-\varepsilon'$  in time  $\tilde{O}(m'\varepsilon^{-5} + n\varepsilon^{-13}) = \tilde{O}(m\varepsilon^{-6} + n\varepsilon^{-13})$ , obtaining an  $\mathbf{x}' \in \mathcal{M}_{G'}^{\mathbf{c}'}$ . Letting  $\mathbf{x} \in \mathcal{M}_G$  be derived from  $\mathbf{x}'$  as in [Claim 5.1](#), it follows that

$$f(\mathbf{x}) \geq \mathbf{w}'^\top \mathbf{x}' \geq (1-\varepsilon') \max_{\mathbf{x}'' \in \mathcal{M}_{G'}^{\mathbf{c}'}} \mathbf{w}'^\top \mathbf{x}'' \geq (1-\varepsilon')(1-2\varepsilon') \max_{\mathbf{x}^* \in \mathcal{M}_G} f(\mathbf{x}^*) \geq (1-\varepsilon) \max_{\mathbf{x}^* \in \mathcal{M}_G} f(\mathbf{x}^*),$$

where we used [Claim 5.2](#). This concludes the proof.  $\square$

**5.3 Putting Everything Together** We can now prove [Theorem 3.3](#) by combining [Lemmas 5.3](#) and [5.6](#).

*Proof of [Theorem 3.3](#).* Let  $f_e(x) \stackrel{\text{def}}{=} \mathbf{w}_e x + \mu \cdot \mathbf{w}_e x \log \frac{\gamma}{\mathbf{w}_e x}$  so that the entropy-regularized matching objective is a coordinate-separable concave function  $f(\mathbf{x}) \stackrel{\text{def}}{=} \sum_{e \in E} f_e(\mathbf{x}_e)$ .

The runtime of  $\tilde{O}(m\varepsilon^{-6} + n\varepsilon^{-13})$  follows from [Lemma 5.6](#) since the function  $f$  satisfies (1)  $\mathbf{w}_e x \leq f_e(x) \leq \mathbf{w}_e + \mu \mathbf{w}_e \log(\gamma/\mathbf{w}_e)$  is polynomially bounded for  $x \geq 1/\text{poly}(n)$ , (2) each  $f_e$  can be evaluated in  $O(1)$  time, and (3)  $z_e^* = 1$  since  $\mu \leq 1$ .

For the runtime of  $\hat{O}(m\varepsilon^{-5})$ , we use [Lemma 5.3](#) in which each  $f_e(x)$  is decomposed into  $f_e^{(1)}(x) + f_e^{(2)}(x)$ , where  $f_e^{(1)}(x) \stackrel{\text{def}}{=} (\mathbf{w}_e + \mu \mathbf{w}_e \log \gamma)x$  is linear and  $f_e^{(2)}(x) \stackrel{\text{def}}{=} -\mu \cdot \mathbf{w}_e x \log \mathbf{w}_e x$ . We use the barriers  $\psi_e^{(1)}(x, y) \stackrel{\text{def}}{=} -\log(y + (\mathbf{w}_e + \mu \mathbf{w}_e \log \gamma)x)$  for  $\{(x, y) : y \geq -f_e^{(1)}(x)\}$  and  $\psi_e^{(2)}(x, y) \stackrel{\text{def}}{=} -\log(\mathbf{w}_e x) - \log(y/\mu - \mathbf{w}_e x \log(\mathbf{w}_e x))$  for  $\{(x, y) : y \geq -f_e^{(2)}(x)\}$ . The barrier  $\psi_e^{(1)}$  is the same as the one used in [\[33, Theorem 10.16\]](#) for linear functions, and  $\psi_e^{(2)}$  is the same barrier that [\[33, Theorem 10.16\]](#) used for the entropy term. Both of the barriers were shown to satisfy [\[33, Assumption 10.2\]](#) as long as the coefficients are polynomially bounded (note that we apply an affine substitution for  $\psi_e^{(2)}$  which preserves self-concordance [\[59, Proposition 3.1.1\]](#)).  $\square$

## 6 Dynamic Rounding Algorithms

In [Sections 4](#) and [5](#) we have shown how to solve the decremental fractional matching problem with  $\tilde{O}(\text{poly}(\varepsilon^{-1}))$  amortized update time and  $\tilde{O}(m \cdot \text{poly}(\varepsilon^{-1}))$  recourse. Here we further show how to obtain an integral matching from the fractional one with dynamic rounding algorithms whose definition is recalled below.

**DEFINITION 3.1. (DYNAMIC ROUNDING ALGORITHM)** A dynamic rounding algorithm, for a given  $n$ -vertex graph  $G = (V, E)$ , edge weights  $\mathbf{w} \in \mathbb{N}^E$  bounded by  $W = \text{poly}(n)$ , and accuracy parameter  $\varepsilon > 0$ , initializes with an  $\mathbf{x} \in \mathcal{M}_G$  and must maintain an integral matching  $M \subseteq \text{supp}(\mathbf{x})$  with  $\mathbf{w}(M) \geq (1-\varepsilon)\mathbf{w}^\top \mathbf{x}$  under entry updates to  $\mathbf{x}$  that guarantee  $\mathbf{x} \in \mathcal{M}_G$  after each operation.

Note that although the previous sections of our paper focus on the decremental setting, the rounding algorithm as defined in [Definition 3.1](#) and given below in [Theorem 3.5](#) is fully dynamic. In other words, it works under arbitrary updates to the fractional matching  $\mathbf{x}$ , irrespective of how the underlying fractional algorithm maintains  $\mathbf{x}$ . Our rounding algorithms are obtained by maintaining a sparse subgraph in which the maximum weight matching is approximately preserved.

**DEFINITION 6.1.** For a fractional matching  $\mathbf{x} \in \mathcal{M}_G$ , a subgraph  $H \subseteq \text{supp}(\mathbf{x})$  is an  $s$ -sparse  $\varepsilon$ -sparsifier for  $s \stackrel{\text{def}}{=} s(n, m, \varepsilon)$  of  $G$  if  $|H| \leq s \cdot \|\mathbf{x}\|_1$  and  $M_{\mathbf{w}}^*(H) \geq (1-\varepsilon)\mathbf{w}^\top \mathbf{x}$ . We call a fractional matching  $\mathbf{x}^{(H)} \in \mathcal{M}_H$  a certificate of  $H$  if  $\mathbf{w}^\top \mathbf{x}^{(H)} \geq (1-\varepsilon)\mathbf{w}^\top \mathbf{x}$ .

Following standard techniques of periodic recomputation, if we can maintain an  $\tilde{O}(\text{poly}(\varepsilon^{-1}))$ -sparse  $O(\varepsilon)$ -sparsifier under updates to  $\mathbf{x}$ , then this gives the desired fully-dynamic rounding algorithm using the below static algorithm of [35].

PROPOSITION 6.1. ([35]) *There is an  $\tilde{O}(m\varepsilon^{-1})$  time algorithm that given an  $m$ -edge graph  $G = (V, E)$  weighted by  $\mathbf{w} \in \mathbb{R}_{\geq 0}^E$  computes a matching  $M \subseteq E$  such that*

$$\mathbf{w}(M) \geq (1 - \varepsilon) \cdot \max_{\text{matching } M' \subseteq E} \mathbf{w}(M').$$

The notion of sparsifier maintenance is formalized as follows.

DEFINITION 6.2. *An algorithm  $\mathcal{S}$  is an  $(s, T_{\text{init}}, T_{\text{update}}, T_{\text{output}})$ -sparsifier-maintainer if given a fractional matching  $\mathbf{x}$  and parameter  $\varepsilon > 0$ , it initializes in  $T_{\text{init}}(n, m, \varepsilon)$  time, processes each entry update to  $\mathbf{x}$  in  $T_{\text{update}}(n, m, \varepsilon)$  amortized time, and outputs an  $s$ -sparse  $\varepsilon$ -sparsifier  $H$  of the current  $\mathbf{x}$  in  $T_{\text{output}}(n, m, \varepsilon) \cdot |H|$  time.*

LEMMA 6.1. (ROUNDING VIA SPARSIFICATION) *Given an  $(s, T_{\text{init}}, T_{\text{update}}, T_{\text{output}})$ -sparsifier-maintainer  $\mathcal{S}$ , there is a dynamic rounding algorithm that initializes in  $\tilde{O}(m \cdot \varepsilon^{-1} + T_{\text{init}}(n, m, \varepsilon/4))$  time and maintains an integral matching  $M \subseteq \text{supp}(\mathbf{x})$  with  $\mathbf{w}(M) \geq (1 - \varepsilon) \mathbf{w}^\top \mathbf{x}$  in amortized*

$$\tilde{O} \left( T_{\text{update}}(n, m, \varepsilon/4) + s(n, m, \varepsilon/4) \cdot W \cdot \left( \frac{T_{\text{output}}(n, m, \varepsilon/4)}{\varepsilon} + \frac{1}{\varepsilon^2} \right) \right)$$

*time per update to  $\mathbf{x}$ . The dynamic rounding algorithm has the same properties as  $\mathcal{S}$  does in terms of being deterministic/randomized and being fully/output-adaptive.*

*Proof.* Let  $M$  be an initial  $(1 - \frac{\varepsilon}{2})$ -approximate maximum weight matching over  $\text{supp}(\mathbf{x})$  obtained by the static algorithm of [35]. We initialize  $\mathcal{S}$  and feed every update to  $\mathbf{x}$  to it to maintain an  $\varepsilon/4$ -sparsifier  $H$  of  $\mathbf{x}$  in  $T_{\text{update}}(n, m, \varepsilon/4)$  time per update. If  $\mathbf{x}_e$  is set to zero for some  $e \in M$ , we remove  $e$  from  $M$ . Every time  $\mathbf{w}(M) < (1 - \varepsilon) \mathbf{w}^\top \mathbf{x}$ , we query  $\mathcal{S}$  to get an  $s(n, m, \varepsilon/4)$ -sparse  $\varepsilon/4$ -sparsifier  $H$  and re-compute  $M$  as a  $(1 - \frac{\varepsilon}{4})$ -approximate matching of  $H$ , again using [35]. This step takes time

$$\tilde{O} \left( T_{\text{output}}(n, m, \varepsilon/4) \cdot s(n, m, \varepsilon/4) \cdot \|\mathbf{x}\|_1 + s(n, m, \varepsilon/4) \cdot \|\mathbf{x}\|_1 \cdot \varepsilon^{-1} \right).$$

By definition of an  $\varepsilon/4$ -sparsifier, we have  $\mathbf{w}(M) \geq (1 - \frac{\varepsilon}{4}) (1 - \frac{\varepsilon}{4}) \mathbf{w}^\top \mathbf{x} \geq (1 - \frac{\varepsilon}{2}) \mathbf{w}^\top \mathbf{x}$ . The re-computation happens at most once every  $\frac{\varepsilon}{2W} \cdot \mathbf{w}^\top \mathbf{x} \geq \frac{\varepsilon}{2W} \cdot \|\mathbf{x}\|_1$  updates, as we need that many updates to either decrease  $\mathbf{w}(M)$  by  $\frac{\varepsilon}{2} \cdot \mathbf{w}^\top \mathbf{x}$  or increase the value of the maximum weight matching size by  $\frac{\varepsilon}{2} \cdot \mathbf{w}^\top \mathbf{x}$ , so the time of re-computation amortizes to  $\tilde{O} \left( s(n, m, \varepsilon/4) \cdot W \cdot \left( \frac{T_{\text{output}}(n, m, \varepsilon/4)}{\varepsilon} + \frac{1}{\varepsilon^2} \right) \right)$  time per update. Combined with the update time of  $\mathcal{S}$ , the lemma follows.  $\square$

In [Section 6.1] we design a deterministic  $(\tilde{O}(\varepsilon^{-2}), \tilde{O}(m), \tilde{O}(W\varepsilon^{-1}), O(1))$ -sparsifier-maintainer in [Lemma 6.7]. This together with [Lemma 6.1] proves the following theorem.

THEOREM 6.1. *There is a deterministic dynamic rounding algorithm for general graphs with amortized update time  $\tilde{O}(W\varepsilon^{-4})$ .*

In [Section 6.3] we further obtain a rounding algorithm specifically for [Theorem 3.4], proving the following.

THEOREM 6.2. (INFORMAL, SEE [THEOREM 3.6]) *We can round the entropy-regularized matching maintained in [Theorem 3.4] in  $\tilde{O}((n^2/m) \cdot \varepsilon^{-6})$  additional amortized time per update.*

We emphasize that while [Theorem 6.2] has a near-optimal dependence on  $W$ , it is not a generic dynamic rounding algorithm as defined in [Definition 3.1]. See [Section 3.4] for a more detailed exposition on this. We leave obtaining a generic, fully-dynamic weighted rounding algorithm that has polylogarithmic or even better dependence on  $W$  as an important open question. No such algorithms are known even for bipartite graphs.

**6.1 Deterministic Sparsifier** To obtain a deterministic sparsifier, we employ the “bit-by-bit” rounding approach of [29] that iteratively sparsifies the support of  $\mathbf{x}$  while maintaining the degree value of each vertex. For bipartite graphs, [29] showed that using this approach, we can directly round to an integral matching without resorting to periodic re-computation, e.g., as in [Lemma 6.1]. However, it is unclear how to extend this approach to general graphs due to odd-set constraints. Our key observation here is that, nevertheless, if we stop the bit-by-bit rounding process earlier, we still get a sparsifier on which we can then perform periodic re-computation.

LEMMA 6.2. *For any  $\varepsilon > 0$ , given a fractional matching  $\mathbf{x} \in \mathcal{M}_G$  and a vector  $\tilde{\mathbf{x}} \in \mathbb{R}_{\geq 0}^E$  such that (1)  $\text{supp}(\tilde{\mathbf{x}}) \subseteq \text{supp}(\mathbf{x})$ , (2)  $\mathbf{w}^\top \tilde{\mathbf{x}} \geq (1 - \varepsilon/2)\mathbf{w}^\top \mathbf{x}$ , (3)  $\tilde{\mathbf{x}}(v) \leq \mathbf{x}(v) + \varepsilon/4$  for all  $v \in V$ , and (4)  $|\mathbf{x}_e - \tilde{\mathbf{x}}_e| \leq \frac{\varepsilon^2}{144}$  for all  $e \in E(G)$ , we can conclude that  $H \stackrel{\text{def}}{=} \text{supp}(\tilde{\mathbf{x}})$  is an  $\varepsilon$ -sparsifier of  $\mathbf{x}$  and it has a certificate  $\mathbf{x}^{(H)}$  satisfying  $\mathbf{x}_e^{(H)} = \Theta(\tilde{\mathbf{x}}_e)$  for all  $e \in E$ .*

*Proof.* Let  $\varepsilon' \stackrel{\text{def}}{=} \varepsilon/4$  and let  $\mathbf{x}' \stackrel{\text{def}}{=} \frac{\tilde{\mathbf{x}}}{1+\varepsilon'}$  whose support is  $\text{supp}(\mathbf{x}') = \text{supp}(\tilde{\mathbf{x}}) = H \subseteq \text{supp}(\mathbf{x})$ . We have  $\mathbf{x}'(v) = \frac{\tilde{\mathbf{x}}(v)}{1+\varepsilon'} \leq 1$  for each  $v \in V$ . For each odd set  $B \in \mathcal{O}_G$  of size  $|B| \leq 3/\varepsilon'$ , it follows from  $|\mathbf{x}_e - \tilde{\mathbf{x}}_e| \leq \frac{\varepsilon'^2}{9}$  that  $|\mathbf{x}(B) - \tilde{\mathbf{x}}(B)| \leq \frac{\varepsilon'}{3}|B|$ . As such, we have  $\mathbf{x}'(B) = \frac{\tilde{\mathbf{x}}(B)}{1+\varepsilon'} \leq \frac{\mathbf{x}(B) + \varepsilon'|B|/3}{1+\varepsilon'} \leq \lfloor |B|/2 \rfloor$ . Therefore by [Fact 2.1] we have  $\mathbf{x}^{(H)} \stackrel{\text{def}}{=} \frac{\mathbf{x}'}{1+\varepsilon'}$  satisfies  $\mathbf{x}^{(H)} \in \mathcal{M}_G$  and this fractional matching has weight  $\mathbf{w}^\top \mathbf{x}^{(H)} \geq (1 - \varepsilon/2)(1 - \varepsilon')^2 \mathbf{w}^\top \mathbf{x} \geq (1 - \varepsilon)\mathbf{w}^\top \mathbf{x}$ . This shows that  $H$  is indeed an  $\varepsilon$ -sparsifier of  $\mathbf{x}$ . That  $\mathbf{x}_e^{(H)} = \Theta(\tilde{\mathbf{x}}_e)$  is obvious from definition of  $\mathbf{x}^{(H)}$ .  $\square$

The pipage-rounding algorithm of [29] is based on the following **degree-split** algorithm for dividing a graph into a collection of cycles and paths.

PROPOSITION 6.2. ([29, PROPOSITION 2.4]) *There exists an algorithm **degree-split**, which on multigraph  $G = (V, E)$  with maximum edge multiplicity at most two (i.e., no edge has more than two copies) computes in  $O(|E|)$  time two (simple) edge-sets  $E^{(1)}$  and  $E^{(2)}$  of two disjoint sub-graphs of  $G$ , such that  $E^{(1)}$ ,  $E^{(2)}$  and the degrees  $d_G(v)$  and  $d^{(i)}(v)$  of  $v$  in  $G$  and  $H^{(i)} \stackrel{\text{def}}{=} (V, E^{(i)})$  satisfy:*

- (1)  $|E^{(1)}| = \left\lceil \frac{|E|}{2} \right\rceil$  and  $|E^{(2)}| = \left\lfloor \frac{|E|}{2} \right\rfloor$  and
- (2)  $d^{(i)}(v) \in \left[ \frac{d_G(v)}{2} - 1, \frac{d_G(v)}{2} + 1 \right]$  for all  $v \in V$ .

The rounding algorithm of [29] works in a bit-by-bit fashion, and thus for  $\mathbf{x} \in [0, 1]^E$  we encode each entry  $\mathbf{x}_e$  as a binary string  $\mathbf{x}_e = \sum_i (\mathbf{x}_e)_i \cdot 2^{-i}$ . The following observation suggests that it is without loss of optimality to focus only on the first  $\tilde{O}(\varepsilon^{-1})$  bits of this encoding.

OBSERVATION 6.1. (SIMILAR TO [29, OBSERVATION 2.2]) *For an  $\mathbf{x} \in \mathcal{M}_G$  with  $\mathbf{w}^\top \mathbf{x} \geq 1$  and accuracy parameter  $\varepsilon > 0$ , the matching  $\mathbf{x}'$  obtained from  $\mathbf{x}$  by zeroing out edges  $e$  with  $\mathbf{x}_e < \frac{\varepsilon}{2n^2W}$  and setting  $(\mathbf{x}'_e)_i = 0$  for all  $e \in E$  and  $i > L$  for  $L \stackrel{\text{def}}{=} 1 + \left\lceil \log \frac{2n^2W}{\varepsilon} \right\rceil$  satisfies  $\mathbf{w}^\top \mathbf{x}' \geq (1 - \varepsilon)\mathbf{w}^\top \mathbf{x}$ .*

*Proof.* Direct calculation shows

$$\mathbf{w}^\top \mathbf{x}' \geq \mathbf{w}^\top \mathbf{x} - W \cdot \left( \binom{n}{2} \cdot \frac{\varepsilon}{2n^2W} + \sum_{e \in E} \sum_{i > L} 2^{-i} \right) \geq \mathbf{w}^\top \mathbf{x} - \varepsilon \geq (1 - \varepsilon)\mathbf{w}^\top \mathbf{x}.$$

$\square$

Our algorithm for deterministically maintaining a dynamic sparsifier is [Algorithm 4], which is almost identical to [29, Algorithm 2] except that we only run it until level  $L_{\min}$  instead of 0 (modulo rather straightforward changes needed for generalization to the weighted case). Stopping the algorithm early disallows us from having the same “straight-to-integral” property as [29, Algorithm 2] does but enables rounding in general graphs.

For simplicity of analysis, we depart from [29] in the implementation of **Rebuild**( $i$ ) in that we are using a tail-recursion style implementation while they use a loop from  $j = i$  to  $L_{\min} + 1$ . These two implementations are equivalent, and we choose the current one to emphasize that a call to **Rebuild**( $i$ ) causes a **Rebuild**( $i - 1$ ).



---

**Algorithm 4: DETSPARSIFIER**

---

**global:** edge weights  $\mathbf{w} \in \mathbb{N}^E$  and accuracy parameter  $\varepsilon \in (0, 1)$ .  
**global:** fractional matching  $\mathbf{x} \in \mathcal{M}_G$ .  
**global:** maximum and minimum level  $L_{\max}, L_{\min} \in \mathbb{N}$ .  
**global:** edge sets  $E_i \subseteq \text{supp}_i(\mathbf{x})$  and  $F_i \subseteq E_i$  for  $i \in \{L_{\min} + 1, \dots, L_{\max}\}$ .  
**global:** counters  $c_i \in \mathbb{N}$  for  $i \in \{L_{\min} + 1, \dots, L_{\max}\}$ .

```
1 function Initialize( $G = (V, E, \mathbf{w}), \mathbf{x} \in \mathcal{M}_G$  such that  $(\mathbf{x}_e)_i = 0$  for all  $i > L, \varepsilon \in (0, 1)$ )
2   Save  $\mathbf{w}, \varepsilon$ , and  $\mathbf{x}$  as global variables.
3   Set  $L_{\min} \leftarrow \lceil \log \frac{2000L}{\varepsilon^2} \rceil$  and  $L_{\max} \leftarrow L$ .
4   Call Rebuild( $L_{\max}$ ).

5 function Rebuild( $i$ )
6   if  $i \leq L_{\min}$  then return.
7    $E_i \leftarrow \text{supp}_i(\mathbf{x})$  and  $c_i \leftarrow 0$ .
8   Let  $(E^{(1)}, E^{(2)}) \leftarrow \text{degree-split}(G[E_i \cup F_i])$ .
9   Set  $F_{i-1}$  to be the  $E^{(b)}$  for  $b \in \{1, 2\}$  with larger weight, i.e., the one that satisfies
    $\mathbf{w}(E^{(b)}) \geq \mathbf{w}(E^{(3-b)})$ .
10  Call Rebuild( $i - 1$ ).

11 function Update( $e, \nu$  such that  $(\nu)_i = 0$  for all  $i > L_{\max}$ )
12   $\mathbf{x}_e \leftarrow \nu$ .
13  for  $i = L_{\max}, L_{\max} - 1, \dots, L_{\min} + 1$  do
14    if  $e \in E_i$  then remove  $e$  from  $E_i$ .
15    if  $e \in F_{i-1}$  then remove  $e$  from  $F_{i-1}$ .
16    else remove one edge adjacent to each endpoint of  $e$  from  $F_{i-1}$  (if there is one).
17     $c_i \leftarrow c_i + 1$ .
18    if  $c_i > 2^{i-2} \cdot \frac{\varepsilon \cdot \mathbf{w}^\top \mathbf{x}}{L_{\max} \cdot W}$  then call Rebuild( $i$ ) and return.

19 function Output()
20  return  $H \stackrel{\text{def}}{=} \text{supp}_0(\mathbf{x}) \cup \dots \cup \text{supp}_{L_{\min}}(\mathbf{x}) \cup F_{L_{\min}}$  as the sparsifier.
```

---

**Notation from [29].** For  $\mathbf{x} \in \mathcal{M}_G$ , let  $\text{supp}_i(\mathbf{x})$  be the set of edges  $e$  whose  $\mathbf{x}_e$  has its  $i$ -th bit set to one, i.e.,  $\text{supp}_i(\mathbf{x}) \stackrel{\text{def}}{=} \{e \in E : (\mathbf{x}_e)_i = 1\}$ . Let  $F_i(v)$  and  $S_i(v)$  be the number of edges in  $F_i$  and  $\text{supp}_i(\mathbf{x})$  incident to  $v$  respectively. Let  $F_i(e) \stackrel{\text{def}}{=} \llbracket e \in F_i \rrbracket$  and  $E_i(e) \stackrel{\text{def}}{=} \llbracket e \in E_i \rrbracket$ . Let  $\mathbf{x}^{(i)} \in \mathbb{R}_{\geq 0}^E$  be given by

$$(6.22) \quad \mathbf{x}_e^{(i)} \stackrel{\text{def}}{=} F_i(e) \cdot 2^{-i} + \sum_{j=0}^{L_{\min}} S_j(e) \cdot 2^{-j} + \sum_{j=L_{\min}+1}^i E_j(e) \cdot 2^{-j}.$$

for  $L_{\min} \leq i \leq L_{\max}$ .

LEMMA 6.3. (SIMILAR TO [29, LEMMA 3.8]) *If  $\mathbf{x} \in \mathcal{M}_G$  holds for each update, then  $\mathbf{x}^{(i)}(v) \leq \mathbf{x}(v) + \varepsilon/2$  for all  $L_{\min} \leq i \leq L_{\max}$  holds after each update.*

*Proof.* We adopt a similar proof strategy except that the **degree-split** subroutine can only give us a weaker guarantee in non-bipartite graphs. By backward induction from  $i = L_{\max}$  to  $L_{\min} - 1$  we show that

$$(6.23) \quad F_i(v) \leq \left( 2^i \cdot \sum_{j=i+1}^{L_{\max}} S_j(v) \cdot 2^{-j} \right) + L_{\max} - i.$$

The base case when  $i = L_{\max}$  is trivial as the left-hand side is 0 and the right-hand side is non-negative. Now consider an  $i < L_{\max}$  and suppose first that a **Rebuild**( $i + 1$ ) happened right before the current moment. Then

$E_{i+1}$  is set to  $S_{i+1}$  and we have

$$\begin{aligned} F_i(v) &\stackrel{(i)}{\leq} \frac{S_{i+1}(v) + F_{i+1}(v)}{2} + 1 \stackrel{(ii)}{\leq} \frac{1}{2} \left( S_{i+1}(v) + \left( 2^{i+1} \cdot \sum_{j=i+2}^{L_{\max}} S_j(v) \cdot 2^{-j} \right) \right) + L_{\max} - (i+1) + 1 \\ &\leq \left( 2^i \cdot \sum_{j=i+1}^{L_{\max}} S_j(v) \cdot 2^{-j} \right) + L_{\max} - i, \end{aligned}$$

where (i) is by Property (2) and (ii) is by the inductive hypothesis. On the other hand, if the **Update**( $e, \nu$ ) operation did not cause a **Rebuild**( $i+1$ ), then the right-hand side might decrease by at most one as an edge incident to  $v$  is removed. If  $e \in F_i$ , then the left-hand side was also decreased by one so the inequality holds. If  $e \notin F_i$ , then we still decrease the left-hand side by one via removing an edge in  $F_i$  adjacent to  $v$  unless it is empty, for which the left-hand side was already 0 so the inequality holds again. Observe that by definition

$$\mathbf{x}(v) = \sum_{j=0}^{L_{\max}} S_j(v) \cdot 2^{-j},$$

and therefore

$$\begin{aligned} \mathbf{x}^{(i)}(v) - \mathbf{x}(v) &= F_i(v) \cdot 2^{-i} - \left( \sum_{j=L_{\min}+1}^{L_{\max}} S_j(v) \cdot 2^{-j} - \sum_{j=L_{\min}+1}^i E_j(v) \cdot 2^{-j} \right) \\ &\stackrel{(i)}{\leq} L_{\max} \cdot 2^{-i} - \left( \sum_{j=L_{\min}+1}^i S_j(v) \cdot 2^{-j} - \sum_{j=L_{\min}+1}^i E_j(v) \cdot 2^{-j} \right) \\ &\leq \varepsilon/2, \end{aligned}$$

where (i) uses (6.23) and that  $S_j(v) \geq E_j(v)$  for all  $j$  since  $E_j \subseteq \text{supp}_j(\mathbf{x})$  and (ii) is by our choice of  $L_{\min}$  and  $L_{\max}$ .  $\square$

Let  $\tilde{\mathbf{x}} \stackrel{\text{def}}{=} \mathbf{x}^{(L_{\min})}$ . In particular Lemma 6.3 shows that  $\tilde{\mathbf{x}}(v) \leq \mathbf{x}(v) + \varepsilon/2$  for all  $v \in V$ . Also observe that  $\tilde{\mathbf{x}}$  is a vector supported on  $H$ , the sparsifier outputted by Algorithm 4. The following helper lemmas analogous to ones in [29] can be shown, and to avoid repeating their arguments we defer the formal proofs to Section B.

LEMMA 6.4. (ANALOGOUS TO [29, LEMMA 3.10])  $\text{supp}(\mathbf{x}^{(i)}) \subseteq \text{supp}(\mathbf{x})$  holds for each  $L_{\min} \leq i \leq L_{\max}$  after each operation.

LEMMA 6.5. (ANALOGOUS TO [29, LEMMA 3.11])  $\mathbf{w}^\top \tilde{\mathbf{x}} \geq (1 - \varepsilon) \mathbf{w}^\top \mathbf{x}$  holds after each operation.<sup>20</sup>

LEMMA 6.6. (ANALOGOUS TO [29, LEMMA 3.13]) The amortized time per **Update**( $e, \nu$ ) of Algorithm 4 is  $O(W \cdot \varepsilon^{-1} \cdot L_{\max}^2)$ .

Using these helper lemmas we show that Algorithm 4 maintains a deterministic sparsifier.

LEMMA 6.7. There is a deterministic  $(\tilde{O}(\varepsilon^{-2}), \tilde{O}(m), \tilde{O}(W\varepsilon^{-1}), O(1))$ -sparsifier-maintainer.

*Proof.* By preprocessing the input fractional matching  $\mathbf{x}$  to a  $\mathbf{x}'$  with large-enough entries and low bit-complexity as in Observation 6.1 with accuracy parameter  $\varepsilon/2$ , we may run Algorithm 4 on  $\mathbf{x}'$  with accuracy parameter  $\varepsilon/2$  and  $L \stackrel{\text{def}}{=} 1 + \lceil \log \frac{4n^2 W}{\varepsilon} \rceil$ . The initialization, update, and output time of Algorithm 4 follow from definition and Lemma 6.6. Observe that since the rounding algorithm is only run until level  $L_{\min}$ , we have  $|\mathbf{x}'_e - \tilde{\mathbf{x}}_e| \leq 2^{-L_{\min}+1} \leq \frac{(\varepsilon/2)^2}{144}$  for all  $e \in E$ . By Lemma 6.4 we also have  $\text{supp}(\tilde{\mathbf{x}}) \subseteq \text{supp}(\mathbf{x}') \subseteq \text{supp}(\mathbf{x})$ . Therefore, together with Lemmas 6.3 and 6.5, Lemma 6.2 applies and the  $H = \text{supp}(\tilde{\mathbf{x}})$  returned by Algorithm 4 is an  $\varepsilon/2$ -sparsifier of  $\mathbf{x}'$  with certificate  $\mathbf{x}^{(H)}$ , hence an  $\varepsilon$ -sparsifier of  $\mathbf{x}$ . Finally, as  $\mathbf{x}_e^{(H)} = \Theta(\tilde{\mathbf{x}}_e) \geq \Omega(2^{-L_{\min}}) = \tilde{\Omega}(\varepsilon^2)$ ,  $H$  is  $\tilde{O}(\varepsilon^{-2})$ -sparse.  $\square$

<sup>20</sup>Unlike [29], we assume the input matching  $\mathbf{x}$  is already preprocessed by Observation 6.1 and thus the bound stated here is  $(1 - \varepsilon)$  instead of  $(1 - 2\varepsilon)$ .

**6.2 Randomized Degree Sparsifier** Toward getting the near-optimal weighted rounding algorithm for decremental dense graphs, in this section we consider the following notion of degree sparsifiers. We remark that Property (1) below in fact is not used later in Section 6.3. Nevertheless, we choose to include it to make degree-sparsifier a stronger notion than the sparsifier in previous sections on unweighted graphs.<sup>21</sup> This, as we will demonstrate later, also shows that the simple and standard sampling approach previously studied for bipartite graphs can also be used for rounding in general graphs.

**DEFINITION 6.3. (DEGREE SPARSIFIERS)** *For a fractional matching  $\mathbf{x} \in \mathcal{M}_G$ , a subgraph  $H \subseteq \text{supp}(\mathbf{x})$  is an  $s$ -sparse  $\varepsilon$ -degree-sparsifier for  $s \stackrel{\text{def}}{=} s(n, m, \varepsilon)$  of  $G$  if  $|H| \leq s \cdot \|\mathbf{x}\|_1$  and there exists a fractional matching  $\mathbf{x}^{(H)} \in \mathcal{M}_G$  supported on  $H$ , which is similarly called a certificate of  $H$ , such that*

- (1)  $\|\mathbf{x}^{(H)}\|_1 \geq (1 - \varepsilon)\|\mathbf{x}\|_1$ .
- (2)  $\mathbf{x}^{(H)}(v) \geq \mathbf{x}(v) - \varepsilon$  for all  $v \in V$ .
- (3)  $\mathbf{x}^{(H)}(B) \geq \mathbf{x}(B) - \frac{\varepsilon}{3} \cdot |B| \geq \mathbf{x}(B) - \varepsilon \cdot \lfloor |B|/2 \rfloor$  for all odd sets  $B \in \mathcal{O}_G$ .

Given  $\mathbf{x} \in \mathcal{M}_G$ , our randomized degree sparsifier  $H$  is obtained from a simple sampling scheme that includes each edge  $e$  into  $H$  independently with probability

$$(6.24) \quad \mathbf{p}_e \stackrel{\text{def}}{=} \min \left( 1, \frac{\mathbf{x}_e}{\gamma} \right) \quad \text{where } \gamma \stackrel{\text{def}}{=} \frac{\varepsilon^2}{320 \cdot d \cdot \ln n},$$

where  $d \stackrel{\text{def}}{=} \max\{c, 1\}$  for  $c > 0$  a given constant to Lemma 6.8 below.

**LEMMA 6.8. (DEGREE SPARSIFICATION)** *For any fractional matching  $\mathbf{x} \in \mathcal{M}_G$  with  $\|\mathbf{x}\|_1 \geq 1$  and constant  $c > 0$ , the random subgraph  $H$  obtained by sampling each edge  $e$  independently with probability  $\mathbf{p}_e$  defined in (6.24), is an  $\tilde{O}(\varepsilon^{-2})$ -sparse  $\varepsilon$ -degree-sparsifier with probability at least  $1 - n^{-c}$ .*

To prove Lemma 6.8, we will use the following version of the Chernoff bound to show concentration on all the constraints.

**PROPOSITION 6.3. (CHERNOFF BOUND)** *Given  $m$  independent random variables  $X_1, \dots, X_m$  in  $[0, b]$ , for  $X = \sum_{i=1}^m X_i$  and any  $\delta > 0$  it holds that*

$$\Pr(|X - \mathbb{E}[X]| > \delta \cdot \mathbb{E}[X]) \leq 2 \exp \left( \frac{-\delta^2 \cdot \mathbb{E}[X]}{(2 + \delta)b} \right).$$

Let  $\mathbf{x}' \in \mathbb{R}^E$  be a random vector conditioned and supported on  $H$  defined by

$$\mathbf{x}'_e \stackrel{\text{def}}{=} \begin{cases} \mathbf{x}_e, & \text{if } \mathbf{x}_e \geq \gamma, \\ \gamma, & \text{if } \mathbf{x}_e < \gamma \text{ and } e \in H, \\ 0, & \text{if } \mathbf{x}_e < \gamma \text{ and } e \in E \setminus H. \end{cases}$$

Equivalently, for  $\mathbf{x}_e < \gamma$ ,  $\mathbf{x}'_e$  is a random variable that takes value  $\gamma$  with probability  $\mathbf{p}_e = \mathbf{x}_e/\gamma$  and 0 with probability  $1 - \mathbf{p}_e$ . Therefore,  $\mathbf{x}'$  is an unbiased estimator of  $\mathbf{x}$ . Let  $F \stackrel{\text{def}}{=} \{e \in E : \mathbf{x}_e < \gamma\}$ . We prove the following claims about  $\mathbf{x}'$ .

**CLAIM 6.1.** *With probability at least  $1 - 2 \cdot n^{-20d}$ , we have  $|\mathbf{x}'(v) - \mathbf{x}(v)| \leq \frac{\varepsilon}{2}$ .*

<sup>21</sup>One can fairly directly extend Property (1) to be weighted, i.e.,  $\mathbf{w}^\top \mathbf{x}^{(H)} \geq (1 - \varepsilon)\mathbf{w}^\top \mathbf{x}$ , at the cost of an  $O(W)$  blow-up in the sparsity.

*Proof.* By definition,  $|\mathbf{x}'(v) - \mathbf{x}(v)| = \left| \|\mathbf{x}'_{F_v}\|_1 - \|\mathbf{x}_{F_v}\|_1 \right|$ , and thus it suffices to bound the right-hand side. Using the Chernoff bound with  $\delta \stackrel{\text{def}}{=} \frac{\varepsilon}{2\|\mathbf{x}_{F_v}\|_1}$ , we have

$$\Pr \left( \left| \|\mathbf{x}'_{F_v}\|_1 - \|\mathbf{x}_{F_v}\|_1 \right| > \frac{\varepsilon}{2} \right) \leq 2 \exp \left( \frac{-\left(\frac{\varepsilon}{2\|\mathbf{x}_{F_v}\|_1}\right)^2 \cdot \|\mathbf{x}_{F_v}\|_1}{\left(2 + \frac{\varepsilon}{2\|\mathbf{x}_{F_v}\|_1}\right) \frac{\varepsilon^2}{320 \cdot d \cdot \ln n}} \right) \leq 2 \exp(-20d \ln n) \leq 2 \cdot n^{-20d}$$

using that  $\mathbf{x}_{F_v} \leq 1$ .  $\square$

CLAIM 6.2. *With probability at least  $1 - 2 \cdot n^{-2d|B|}$ , we have  $|\mathbf{x}'(B) - \mathbf{x}(B)| \leq \varepsilon \cdot \frac{|B|}{12}$ .*

*Proof.* Again  $|\mathbf{x}'(B) - \mathbf{x}(B)| = \left| \|\mathbf{x}'_{F[B]}\|_1 - \|\mathbf{x}_{F[B]}\|_1 \right|$  and therefore we will bound the right-hand side. We set  $\delta \stackrel{\text{def}}{=} \frac{|B|\varepsilon}{12\|\mathbf{x}_{F[B]}\|_1}$  and apply the Chernoff bound, giving

$$\begin{aligned} \Pr \left( \left| \|\mathbf{x}'_{F[B]}\|_1 - \|\mathbf{x}_{F[B]}\|_1 \right| > \frac{|B|}{12} \cdot \varepsilon \right) &\leq 2 \exp \left( \frac{-\left(\frac{|B|\varepsilon}{12\|\mathbf{x}_{F[B]}\|_1}\right)^2 \cdot \|\mathbf{x}_{F[B]}\|_1}{\left(2 + \frac{|B|\varepsilon}{12\|\mathbf{x}_{F[B]}\|_1}\right) \frac{\varepsilon^2}{320 \cdot d \cdot \ln n}} \right) \\ &\leq 2 \exp \left( \frac{-26d \ln n |B|^2}{24\|\mathbf{x}_{F[B]}\|_1 + |B|\varepsilon} \right). \end{aligned}$$

Since  $\|\mathbf{x}_{F[B]}\|_1 \leq \frac{|B|}{2}$ , we get

$$\Pr \left( \left| \|\mathbf{x}'_{F[B]}\|_1 - \|\mathbf{x}_{F[B]}\|_1 \right| > \frac{|B|}{12} \cdot \varepsilon \right) \leq 2n^{-2d|B|}.$$

$\square$

CLAIM 6.3. *With probability at least  $1 - 2 \cdot n^{-32d}$ , we have  $(1 - \frac{\varepsilon}{2}) \|\mathbf{x}\|_1 \leq \|\mathbf{x}'\|_1 \leq (1 + \frac{\varepsilon}{2}) \|\mathbf{x}\|_1$ .*

*Proof.* An application of the Chernoff bound shows that

$$\begin{aligned} \Pr \left( \left| \|\mathbf{x}'\|_1 - \|\mathbf{x}\|_1 \right| > \frac{\varepsilon}{2} \|\mathbf{x}\|_1 \right) &= \Pr \left( \left| \|\mathbf{x}'_F\|_1 - \|\mathbf{x}_F\|_1 \right| > \frac{\varepsilon}{2} \|\mathbf{x}\|_1 \right) \\ &\leq 2 \exp \left( \frac{-\left(\frac{\varepsilon \cdot \|\mathbf{x}\|_1}{2\|\mathbf{x}_F\|_1}\right)^2 \cdot \|\mathbf{x}_F\|_1}{\left(2 + \frac{\varepsilon \cdot \|\mathbf{x}\|_1}{2\|\mathbf{x}_F\|_1}\right) \frac{\varepsilon^2}{320 \cdot d \cdot \ln n}} \right) \\ &\leq 2 \exp(-32\|\mathbf{x}\|_1 \ln n) \leq 2n^{-32} \end{aligned}$$

using that  $\|\mathbf{x}_F\|_1 \leq \|\mathbf{x}\|_1$  and  $\|\mathbf{x}\|_1 \geq 1$ .  $\square$

*Proof.* [Proof of Lemma 6.8] Take  $\mathbf{x}^{(H)} \stackrel{\text{def}}{=} \frac{\mathbf{x}'}{1+\varepsilon/2}$  as the fractional matching over  $H$ . By union bounds over Claims 6.1 to 6.3 (note that there are at most  $n^k$  odd sets of size  $k$ ), with probability  $1 - n^{-d} \geq 1 - n^{-c}$  we have  $\mathbf{x}^{(H)}(v) = \frac{\mathbf{x}'(v)}{1+\varepsilon/2} \leq 1$  for all  $v \in V$ ,  $\mathbf{x}^{(H)}(B) = \frac{\mathbf{x}'(B)}{1+\varepsilon/2} \leq \lfloor |B|/2 \rfloor$  for all  $B \in \mathcal{O}_G$ , and  $(1 - \varepsilon) \cdot \|\mathbf{x}\|_1 \leq \|\mathbf{x}^{(H)}\|_1 \leq (1 + \varepsilon) \cdot \|\mathbf{x}\|_1$ . This implies  $\mathbf{x}^{(H)} \in \mathcal{M}_G$  and shows Property (1). Claim 6.1 also implies that  $\mathbf{x}^{(H)}(v) = \frac{\mathbf{x}'(v)}{1+\varepsilon/2} \geq \mathbf{x}(v) - \varepsilon$  which establishes Property (2). Similarly, Claim 6.2 implies that  $\mathbf{x}^{(H)}(B) = \frac{\mathbf{x}'(B)}{1+\varepsilon/2} \geq \mathbf{x}(B) - \frac{\varepsilon}{3} \cdot |B|$  which establishes Property (3). For the sparsity of  $H$ , note that  $\mathbf{x}_e^{(H)} \geq \Omega(\varepsilon^2)$  and by  $\|\mathbf{x}^{(H)}\|_1 \leq O(\|\mathbf{x}\|_1)$  we have  $|H| \leq \tilde{O}(\varepsilon^{-2}) \cdot \|\mathbf{x}\|_1$ . Finally,  $H \subseteq \text{supp}(\mathbf{x})$  is apparent by definition. This concludes the proof.  $\square$

Combining the above sampling scheme with the dynamic set sampler of [29], we get the following algorithm for maintaining a degree sparsifier.

LEMMA 6.9. ([29, THEOREM 5.2]) *There is an output-adaptive data structure that, given  $\mathbf{p} \in [0, 1]^d$  with the guarantee that  $\min_{i \in [d]: \mathbf{p}_i \neq 0} \mathbf{p}_i \geq 1/\text{poly}(n)$  for each update, initializes in  $O(d)$  time and supports (1) updating an entry  $\mathbf{p}_i$  in  $O(1)$  time and (2) sampling a set  $T \subseteq [d]$  in  $O(1 + |T|)$  time such that each  $i \in [d]$  is included independently with probability  $\mathbf{p}_i$ .*

LEMMA 6.10. *There is a randomized output-adaptive  $(\tilde{O}(\varepsilon^{-2}), O(m), O(1), O(1))$ -degree-sparsifier-maintainer for any  $\varepsilon \geq 1/\text{poly}(n)$  that succeeds w.h.p.*

*Proof.* For a given  $\varepsilon > 0$ , let  $\varepsilon' \stackrel{\text{def}}{=} \varepsilon/2$ . We can preprocess the matching  $\mathbf{x}$  to be  $\mathbf{x}'$  by zeroing out edges with  $\mathbf{x}_e < \frac{\varepsilon'}{2n^2}$ . By [Observation 6.1], we have  $\|\mathbf{x}'\|_1 \geq (1 - \varepsilon')\|\mathbf{x}\|$ . Moreover,  $|\mathbf{x}'(v) - \mathbf{x}(v)| \leq \varepsilon'$  and  $|\mathbf{x}'(B) - \mathbf{x}(B)| \leq \varepsilon'$  hold for all  $v \in V$  and  $B \in \mathcal{O}_G$ . Now we can simply run the dynamic sampler of [Lemma 6.9] with  $\mathbf{p}_e$  being the probability defined in (6.24) with accuracy parameter  $\varepsilon'$  and constant  $c > 0$  chosen to make [Lemma 6.10] succeed w.h.p. Note that as  $\mathbf{x}'_e \geq \frac{\varepsilon'}{2n^2} \geq 1/\text{poly}(n)$  for all  $e$  with non-zero  $\mathbf{x}'_e$ ,  $\mathbf{p}_e$ 's are polynomially bounded and thus the runtime of [Lemma 6.9] holds. Every time there is an update to  $\mathbf{x}_e$ , we interpret that as an update to  $\mathbf{x}'_e$  (by again zeroing out the coordinate if  $\mathbf{x}_e < \frac{\varepsilon'}{2n^2}$ ), and hence to  $\mathbf{p}_e$ , and feed it to the dynamic sampler. Every time we are asked to output a sparsifier, we use the dynamic sampler to sample each edge independently with probability  $\mathbf{p}_e$ , obtaining a subgraph  $H$ , in time  $O(1 + |H|)$ . By [Lemma 6.8],  $H$  is an  $\tilde{O}(\varepsilon^{-2})$ -sparse  $\varepsilon'$ -degree-sparsifier for  $\mathbf{x}'$  w.h.p. This implies that  $H$  is an  $\varepsilon$ -degree-sparsifier for  $\mathbf{x}$  as well. Thus, we have  $T_{\text{init}}(n, m, \varepsilon) = O(m)$ ,  $T_{\text{update}}(n, m, \varepsilon) = O(1)$ , and  $T_{\text{output}}(n, m, \varepsilon) = O(1)$ . The algorithm works against an output-adaptive adversary as [Lemma 6.9] does.  $\square$

As alluded to at the beginning of the section, by including Property (1) into the definition, we get the following immediate corollary. This can straightforwardly be generalized to the weighted setting by imposing a linear dependence on  $W$ , but we omit this part since it is irrelevant to this paper.

COROLLARY 6.1. *For  $\varepsilon \geq 1/\text{poly}(n)$  there is a randomized output-adaptive  $(\tilde{O}(\varepsilon^{-2}), O(m), O(1), O(1))$ -sparsifier-maintainer for unweighted graphs that succeeds w.h.p.*

**6.3 Weighted Rounding for Entropy-Regularized Matching** While the rounding algorithms presented previously work with weights, their dependence on  $W$  is polynomial, which in general might incur a  $\text{poly}(n)$  or  $\varepsilon^{-O(1/\varepsilon)}$  runtime overhead. As such, here we further leverage the primal-dual properties of entropy-regularized matchings and provide a weighted rounding algorithm in the decremental setting that has near-optimal update time in dense graphs, proving Part (II) of [Theorem 3.4]. Before diving into the technical calculations that form the rest of the sections, we first give a short and intuitive explanation of this weighted rounding algorithm. It leverages the following properties.

- (1) For any fractional matching  $\mathbf{x}_e$ , deleting edges  $e$  with  $\mathbf{x}_e = O(\varepsilon/n)$  does not affect the weight of  $\mathbf{x}$  by more than an  $O(\varepsilon)$  factor.
- (2) For the entropy-regularized fractional matching, a degree-sparsifier obtained from the unweighted rounding algorithm keeps most of its weight.

As such, (1) ensures that we can only consider edges with  $\mathbf{x}_e \geq \Omega(\varepsilon/n)$ . This implies that any deletion to the integral matching that we maintain will also delete a comparable portion from the underlying fractional matching, which bounds the number of rebuilds of the integral matchings. Rounding with nearly optimal dependence on  $W$  is then made possible by (2).

With this intuition we now proceed to the proofs. Let  $G = (V, E)$  be the input graph to the decremental matching problem with edge weights  $\mathbf{w} \in \mathbb{N}^E$ . In the remainder of this section we consider a fixed set of input  $G^{(t)} = (V, E^{(t)})$ ,  $\mu$ , and  $\gamma$  to [Problem 3.3], where  $E^{(t)} \subseteq E$  is the edge set in [Algorithm 1] for the  $t$ -th rebuild. Fixing  $t$ , we let  $\mathbf{x}^* \stackrel{\text{def}}{=} \mathbf{x}_{E^{(t)}, \gamma}^\mu \in \mathcal{M}_{G^{(t)}}$  be the unique optimal entropy-regularized fractional matching. Note that we adopt the notations from [Section 4] for the special case of  $\mathcal{X} \stackrel{\text{def}}{=} \mathcal{M}_G$ , and thus the subscripts of certain notations are changed to  $E$  from  $S$ . For instance,  $Z_{E^{(t)}, \gamma}^\mu$  denotes the optimal value of the entropy-regularized matching problem on  $G^{(t)}$  with parameters  $\mu$  and  $\gamma$ .

LEMMA 6.11. *There exists a pair of dual solutions  $(\mathbf{y}, \mathbf{z}) \in \mathbb{R}^V \times \mathbb{R}^{\mathcal{O}_G}$  such that for every edge  $e \in E^{(t)}$  it holds that*

$$\mathbf{x}_e^* = 2^{\frac{1}{\mu} - 1 - \frac{s_e}{\mu \cdot \mathbf{w}_e} + \log \frac{\gamma}{\mathbf{w}_e}},$$

where  $\mathbf{s}_e \stackrel{\text{def}}{=} \mathbf{y}_u + \mathbf{y}_v + \sum_{B \supseteq \{u,v\}} \mathbf{z}_B$  for  $e = \{u, v\}$ . Further, the optimal objective value  $Z_{E^{(t)}, \gamma}^\mu$  satisfies

$$Z_{E^{(t)}, \gamma}^\mu = \mu \cdot \sum_{e \in E^{(t)}} \mathbf{w}_e \mathbf{x}_e^* + \sum_{v \in V} \mathbf{y}_v + \sum_{B \in \mathcal{O}_G} \mathbf{z}_B \cdot \left\lfloor \frac{|B|}{2} \right\rfloor.$$

*Proof.* Denote

$$L(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{r}) \stackrel{\text{def}}{=} f_{E^{(t)}, \gamma}^\mu(\mathbf{x}) + \sum_{v \in V} \mathbf{y}_v \left(1 - \sum_{e \in E_v^{(t)}} \mathbf{x}_e\right) + \sum_{B \in \mathcal{O}_G} \mathbf{z}_B \left(\left\lfloor \frac{|B|}{2} \right\rfloor - \sum_{e \in E^{(t)}[B]} \mathbf{x}_e\right) + \sum_{e \in E^{(t)}} \mathbf{r}_e \mathbf{x}_e$$

as the Lagrangian of [Problem 3.3](#) (see [\(2.2\)](#) for the definition of  $\mathcal{M}_G$ ). Strong duality of [Problem 3.3](#) (by Slater's condition) implies that

$$Z_{E^{(t)}, \gamma}^\mu = f_{E^{(t)}, \gamma}^\mu(\mathbf{x}^*) = \max_{\mathbf{x}} \min_{\mathbf{y}, \mathbf{z}, \mathbf{r} \geq \mathbf{0}} L(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{r}) = \min_{\mathbf{y}, \mathbf{z}, \mathbf{r} \geq \mathbf{0}} \max_{\mathbf{x}} L(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{r}).$$

Suppose  $(\mathbf{y}, \mathbf{z}, \mathbf{r})$  is the minimizer of the dual problem. From the stationarity of the KKT condition, the corresponding optimal primal solution  $\mathbf{x}^*$  satisfies that

$$(6.25) \quad \mathbf{x}_e^* = 2^{\frac{1}{\mu} - 1 - \frac{s_e}{\mu \cdot \mathbf{w}_e} + \frac{\mathbf{r}_e}{\mu \cdot \mathbf{w}_e} + \log \frac{\gamma}{\mathbf{w}_e}}$$

for every edge  $e \in E^{(t)}$ . From the complementary slackness of the KKT condition, we know  $\mathbf{r}_e \mathbf{x}_e^* = 0$  for every  $e \in E^{(t)}$ . Since  $\mathbf{x}^* > \mathbf{0}$  from [\(6.25\)](#), we have  $\mathbf{r} = \mathbf{0}$ , and therefore

$$\mathbf{x}_e^* = 2^{\frac{1}{\mu} - 1 - \frac{s_e}{\mu \cdot \mathbf{w}_e} + \log \frac{\gamma}{\mathbf{w}_e}}.$$

Further, by plugging in  $(\mathbf{x}^*, \mathbf{y}, \mathbf{z}, \mathbf{r})$  to  $L$  and expanding the formula, we can see that the optimal objective value can be written as

$$Z_{E^{(t)}, \gamma}^\mu = L(\mathbf{x}^*, \mathbf{y}, \mathbf{z}, \mathbf{r}) = \mu \cdot \sum_{e \in E^{(t)}} \mathbf{w}_e \mathbf{x}_e^* + \sum_{v \in V} \mathbf{y}_v + \sum_{B \in \mathcal{O}_G} \mathbf{z}_B \cdot \left\lfloor \frac{|B|}{2} \right\rfloor.$$

□

An immediate corollary of [Lemma 6.11](#) that will be used throughout is the following.

COROLLARY 6.2. *It holds that*

$$Z_{E^{(t)}, \gamma}^\mu \geq \sum_{v \in V} \mathbf{y}_v + \sum_{B \in \mathcal{O}_G} \mathbf{z}_B \cdot \left\lfloor \frac{|B|}{2} \right\rfloor.$$

For  $\varepsilon > 0$  and  $\mathbf{x} \in \mathbb{R}_{\geq 0}^{E^{(t)}}$ , let  $E_\varepsilon^{(t)}(\mathbf{x}) \stackrel{\text{def}}{=} \{e \in E^{(t)} \mid \mathbf{x}_e \geq \frac{\varepsilon}{3n}\}$ . Recall from [Table 1](#) that  $\nu_{E^{(t)}}^* \stackrel{\text{def}}{=} \max_{\mathbf{x} \in \mathcal{M}_{G^{(t)}}} \mathbf{w}_{E^{(t)}}^\top \mathbf{x}$  which satisfies  $\nu_{E^{(t)}}^* \leq nW$ .

LEMMA 6.12. *For  $\varepsilon > 0$ ,  $\nu_{E^{(t)}}^* \leq \gamma \leq m \cdot \nu_{E^{(t)}}^*$ , and  $\mu \leq \frac{\varepsilon}{8 \log(n^4 W / \varepsilon)}$ , the optimal dual solution  $(\mathbf{y}, \mathbf{z})$  approximately covers all edge weights, i.e., for every  $e \in E^{(t)}$  it holds that  $s_e \geq (1 - \varepsilon/8) \mathbf{w}_e$ . Moreover, the covering is almost tight on the subgraph  $E_\varepsilon^{(t)}(\mathbf{x})$ , i.e.,  $s_e \leq (1 + \varepsilon/8) \mathbf{w}_e$  holds for every  $e \in E_\varepsilon^{(t)}(\mathbf{x})$ .*

*Proof.* From [Problem 3.3](#) and the choice of  $\gamma$ , we have  $\max_{e \in E^{(t)}} \mathbf{w}_e \leq \nu_{E^{(t)}}^* \leq \gamma \leq m \cdot \nu_{E^{(t)}}^* \leq n^3 W$ . [Lemma 6.11](#) shows that

$$\mathbf{x}_e^* = 2^{\frac{1}{\mu} - 1 - \frac{s_e}{\mu \cdot \mathbf{w}_e} + \log \frac{\gamma}{\mathbf{w}_e}}.$$



Since the optimal primal solution is feasible, we have  $\mathbf{x}_e^* \leq 1$ , implying

$$\mathbf{s}_e \geq (1 - \mu)\mathbf{w}_e + \mu\mathbf{w}_e \log \frac{\gamma}{\mathbf{w}_e} \geq \left(1 - \frac{\varepsilon}{8}\right) \mathbf{w}_e$$

for all  $e \in E^{(t)}$  because  $\gamma \geq \mathbf{w}_e$  and  $\mu \leq \varepsilon/8$ . Similarly, for all  $e \in E_\varepsilon^{(t)}(\mathbf{x})$ , we have

$$\mathbf{s}_e \leq \left(1 - \mu + \mu \log \left(\frac{n}{3\varepsilon}\right)\right) \mathbf{w}_e + \mu\mathbf{w}_e \log \frac{\gamma}{\mathbf{w}_e} \leq \left(1 + \frac{\varepsilon}{8}\right) \mathbf{w}_e,$$

since  $\gamma \leq n^3W$  and  $\mu \leq \frac{\varepsilon}{8 \log(n^4W/\varepsilon)}$ .  $\square$

We argue that  $E^{(t)}(\varepsilon, \mathbf{x})$  for any fractional matching  $\mathbf{x} \in \mathcal{M}_{G^{(t)}}$  keeps most of the weight of  $\mathbf{x}$ .

LEMMA 6.13. *For any fractional matching  $\mathbf{x} \in \mathcal{M}_{G^{(t)}}$ , accuracy parameter  $\varepsilon > 0$ ,  $\nu_{E^{(t)}}^* \leq \gamma \leq m \cdot \nu_{E^{(t)}}^*$ , and  $\mu \leq \frac{\varepsilon}{8 \log(n^4W/\varepsilon)}$ , we have*

$$\sum_{e \in E_\varepsilon^{(t)}(\mathbf{x})} \mathbf{w}_e \mathbf{x}_e \geq \sum_{e \in E^{(t)}} \mathbf{w}_e \mathbf{x}_e - \frac{4\varepsilon}{7} \cdot Z_{E^{(t)}, \gamma}^\mu.$$

*Proof.* For those edges in  $E_\varepsilon^{(t)}(\mathbf{x})$ , we have

$$\begin{aligned} \sum_{e \in E^{(t)} \setminus E_\varepsilon^{(t)}(\mathbf{x})} \mathbf{w}_e \mathbf{x}_e &\stackrel{(i)}{\leq} \frac{1}{1 - \varepsilon/8} \sum_{e \in E^{(t)} \setminus E_\varepsilon^{(t)}(\mathbf{x})} \mathbf{s}_e \cdot \mathbf{x}_e \\ &= \frac{1}{1 - \varepsilon/8} \sum_{v \in V} \mathbf{y}_v \left\| \mathbf{x}_{E_v^{(t)} \setminus E_\varepsilon^{(t)}(\mathbf{x})} \right\|_1 + \frac{1}{1 - \varepsilon/8} \sum_{B \in \mathcal{O}_G} \mathbf{z}_B \left\| \mathbf{x}_{E^{(t)}[B] \setminus E_\varepsilon^{(t)}(\mathbf{x})} \right\|_1 \\ &\stackrel{(ii)}{\leq} \frac{\varepsilon}{3(1 - \varepsilon/8)} \sum_{v \in V} \mathbf{y}_v + \frac{\varepsilon}{3(1 - \varepsilon/8)} \sum_{B \in \mathcal{O}_G} \mathbf{z}_B \frac{|B|}{2} \\ &\stackrel{(iii)}{\leq} \frac{\varepsilon}{3(1 - \varepsilon/8)} \|\mathbf{y}\|_1 + \frac{\varepsilon}{2(1 - \varepsilon/8)} \sum_{B \in \mathcal{O}_G} \mathbf{z}_B \left\lceil \frac{|B|}{2} \right\rceil \\ &\stackrel{(iv)}{\leq} \frac{4\varepsilon}{7} \cdot Z_{E^{(t)}, \gamma}^\mu, \end{aligned}$$

where (i) is by Lemma 6.12 which shows that  $\mathbf{s}_e \geq (1 - \varepsilon/8)\mathbf{w}_e$  for all  $e \in E^{(t)}$ , (ii) is because

$$\sum_{e \in E_v^{(t)} \setminus E_\varepsilon^{(t)}(\mathbf{x})} \mathbf{x}_e \leq \frac{\varepsilon}{3} \quad \text{and} \quad \sum_{e \in E^{(t)}[B] \setminus E_\varepsilon^{(t)}(\mathbf{x})} \mathbf{x}_e \leq \frac{1}{2} \cdot \sum_{v \in B} \sum_{e \in E_v^{(t)} \setminus E_\varepsilon^{(t)}(\mathbf{x})} \mathbf{x}_e \leq \frac{\varepsilon}{3} \cdot \frac{|B|}{2},$$

(iii) is by  $|B| \geq 3$ , and (iv) is by Corollary 6.2  $\square$

Now, we consider a fractional matching  $\mathbf{x} \in \mathcal{M}_{G^{(t)}}$  that is close to  $\mathbf{x}^*$ . We show that a degree-sparsifier of  $\mathbf{x}$  preserves most of its weight. This is perhaps surprising as the definition of degree-sparsifier (Definition 6.3) is purely unweighted, while we can use it to sparsify this particular weighted matching.

LEMMA 6.14. *For any accuracy parameter  $\varepsilon > 0$ ,  $\nu_E^* \leq \gamma \leq m \cdot \nu_E^*$ , and  $\mu \leq \frac{\varepsilon}{64 \log(8n^4W/\varepsilon)}$ , given an  $\mathbf{x} \in \mathcal{M}_G$  with*

$$(6.26) \quad \sum_{e \in E^{(t)}} \mathbf{w}_e |\mathbf{x}_e - \mathbf{x}_e^*| \leq \frac{\varepsilon}{8} \cdot Z_{E^{(t)}, \gamma}^\mu,$$

for any edge subset  $E' \subseteq E^{(t)}$  such that

$$\sum_{e \in E'} \mathbf{w}_e \mathbf{x}_e \geq (1 - \varepsilon/8) \sum_{e \in E^{(t)}} \mathbf{w}_e \mathbf{x}_e,$$

we have that any  $\varepsilon/8$ -degree-sparsifier  $H$  of  $\mathbf{x}_{E' \cap E_{\varepsilon/8}^{(t)}(\mathbf{x})}$  satisfies

$$M_{\mathbf{w}}^*(H) \geq \left(1 - \frac{\varepsilon}{2}\right) \cdot \sum_{e \in E'} \mathbf{w}_e \mathbf{x}_e.$$

*Proof.* Letting  $\varepsilon' \stackrel{\text{def}}{=} \varepsilon/8$ , we have  $\mu \leq \frac{\varepsilon'}{8 \log(n^4 W / \varepsilon')}$ . Let  $\mathbf{x}^{(H)}$  be the certificate of  $H$  being an  $\varepsilon/8$ -degree-sparsifier. It suffices to prove that  $\sum_{e \in E'} \mathbf{w}_e \mathbf{x}_e^{(H)} \geq (1 - \varepsilon/2) \cdot \sum_{e \in E'} \mathbf{w}_e \mathbf{x}_e$ . From [Lemma 6.13](#) with accuracy parameter  $\varepsilon'$ , we have

$$(6.27) \quad \sum_{e \in E' \setminus E_{\varepsilon/8}^{(t)}(\mathbf{x})} \mathbf{w}_e \mathbf{x}_e \leq \frac{\varepsilon}{14} \cdot Z_{E', \gamma}^\mu \leq \frac{\varepsilon}{14} \cdot Z_{E^{(t)}, \gamma}^\mu \quad \text{and} \quad \sum_{e \in E' \setminus E_{\varepsilon/8}^{(t)}(\mathbf{x}^*)} \mathbf{w}_e \mathbf{x}_e^* \leq \frac{\varepsilon}{14} \cdot Z_{E^{(t)}, \gamma}^\mu.$$

By [\(6.26\)](#), [\(6.27\)](#), and triangle inequality we have

$$\sum_{e \in E' \setminus E_{\varepsilon/8}^{(t)}(\mathbf{x}^*)} \mathbf{w}_e \mathbf{x}_e \leq \sum_{e \in E' \setminus E_{\varepsilon/8}^{(t)}(\mathbf{x}^*)} \mathbf{w}_e \mathbf{x}_e^* + \sum_{e \in E' \setminus E_{\varepsilon/8}^{(t)}(\mathbf{x}^*)} \mathbf{w}_e |\mathbf{x}_e - \mathbf{x}_e^*| \leq \frac{3\varepsilon}{14} \cdot Z_{E^{(t)}, \gamma}^\mu$$

and thus

$$(6.28) \quad \sum_{e \in E' \setminus (E_{\varepsilon/8}^{(t)}(\mathbf{x}) \cup E_{\varepsilon/8}^{(t)}(\mathbf{x}^*))} \mathbf{w}_e \mathbf{x}_e \leq \frac{2\varepsilon}{7} \cdot Z_{E^{(t)}, \gamma}^\mu.$$

Applying [Lemma 6.12](#) with accuracy parameter  $\varepsilon'$ , for every edge  $e \in E^{(t)} \cap E_{\varepsilon/8}^{(t)}(\mathbf{x}^*)$ , we have  $\mathbf{s}_e \leq (1 + \frac{\varepsilon}{64}) \mathbf{w}_e$ . Therefore, letting  $\tilde{E} \stackrel{\text{def}}{=} E' \cap E_{\varepsilon/8}^{(t)}(\mathbf{x})$  and  $\hat{E} \stackrel{\text{def}}{=} \tilde{E} \cap E_{\varepsilon/8}^{(t)}(\mathbf{x}^*)$  for clarity, we have

$$\begin{aligned} \sum_{e \in E'} \mathbf{w}_e \mathbf{x}_e^{(H)} &\geq \sum_{e \in \tilde{E}} \mathbf{w}_e \mathbf{x}_e^{(H)} \geq \left(1 - \frac{\varepsilon}{64}\right) \sum_{e \in \tilde{E}} \mathbf{s}_e \mathbf{x}_e^{(H)} \\ &= \left(1 - \frac{\varepsilon}{64}\right) \left( \sum_{v \in V} \mathbf{y}_v \left\| \mathbf{x}_{\tilde{E}_v}^{(H)} \right\|_1 + \sum_{B \in \mathcal{O}_G} \mathbf{z}_B \left\| \mathbf{x}_{\tilde{E}[B]}^{(H)} \right\|_1 \right) \\ &\stackrel{(i)}{\geq} \left(1 - \frac{\varepsilon}{64}\right) \left( \sum_{v \in V} \mathbf{y}_v \left\| \mathbf{x}_{\tilde{E}_v} \right\|_1 + \sum_{B \in \mathcal{O}_G} \mathbf{z}_B \left\| \mathbf{x}_{\tilde{E}[B]} \right\|_1 \right) - \frac{\varepsilon}{8} \left( \left\| \mathbf{y} \right\|_1 + \sum_{B \in \mathcal{O}_G} \mathbf{z}_B \left\lfloor \frac{|B|}{2} \right\rfloor \right) \\ &\geq \left(1 - \frac{\varepsilon}{64}\right) \left( \sum_{v \in V} \mathbf{y}_v \left\| \mathbf{x}_{\hat{E}_v} \right\|_1 + \sum_{B \in \mathcal{O}_G} \mathbf{z}_B \left\| \mathbf{x}_{\hat{E}[B]} \right\|_1 \right) - \frac{\varepsilon}{8} \left( \left\| \mathbf{y} \right\|_1 + \sum_{B \in \mathcal{O}_G} \mathbf{z}_B \left\lfloor \frac{|B|}{2} \right\rfloor \right) \\ &\stackrel{(ii)}{\geq} \left(1 - \frac{\varepsilon}{64}\right) \left(1 - \frac{\varepsilon}{8}\right) \left( \sum_{e \in \hat{E}} \mathbf{w}_e \mathbf{x}_e \right) - \frac{\varepsilon}{8} Z_{E, \gamma}^\mu \\ &\stackrel{(iii)}{\geq} \left(1 - \frac{9\varepsilon}{64}\right) \left( \sum_{e \in E'} \mathbf{w}_e \mathbf{x}_e \right) - \frac{\varepsilon}{6} Z_{E^{(t)}, \gamma}^\mu \stackrel{(iv)}{\geq} \left(1 - \frac{\varepsilon}{2}\right) \left( \sum_{e \in E'} \mathbf{w}_e \mathbf{x}_e \right), \end{aligned}$$

where (i) uses the assumption of  $\mathbf{x}^{(H)}$  being an  $\varepsilon/8$ -degree-sparsifier of  $\mathbf{x}_{E' \cap E_{\varepsilon/8}^{(t)}(\mathbf{x})}$ , (ii) uses [Lemma 6.12](#) and [Corollary 6.2](#), and (iii) uses [\(6.28\)](#). Finally, (iv) follows from

$$\begin{aligned} \sum_{e \in E'} \mathbf{w}_e \mathbf{x}_e &\stackrel{(a)}{\geq} \left(1 - \frac{\varepsilon}{8}\right) \sum_{e \in E^{(t)}} \mathbf{w}_e \mathbf{x}_e \stackrel{(b)}{\geq} \left(1 - \frac{\varepsilon}{8}\right) \left( \sum_{e \in E^{(t)}} \mathbf{w}_e \mathbf{x}_e^* - \frac{\varepsilon}{8} Z_{E^{(t)}, \gamma}^\mu \right) \\ &\stackrel{(c)}{\geq} \left(1 - \frac{\varepsilon}{8}\right) \left( \left(1 - \frac{\varepsilon}{16}\right) Z_{E^{(t)}, \gamma}^\mu - \frac{\varepsilon}{8} Z_{E^{(t)}, \gamma}^\mu \right) \geq \left(1 - \frac{5}{16}\varepsilon\right) Z_{E^{(t)}, \gamma}^\mu, \end{aligned}$$

where (a) and (b) follow from the input assumption, and (c) is from [Lemma 4.4](#) with accuracy parameter  $\varepsilon'$ . This concludes the proof.  $\square$

We are now ready to finish the proof of [Theorem 3.6](#). For completeness we repeat the proof of [Theorem 3.4](#) with a different set of parameters which allows us to perform rounding afterward.

**THEOREM 6.3.** *There are randomized output-adaptive algorithms that solve [Problem 1.1](#) w.h.p. with amortized update times  $\tilde{O}(\varepsilon^{-41} + (n^2/m) \cdot \varepsilon^{-6})$  and  $\hat{O}(\varepsilon^{-17} + (n^2/m) \cdot \varepsilon^{-6})$ . Additionally, if  $G$  is bipartite then there is such an algorithm with amortized update time  $\hat{O}((n^2/m) \cdot \varepsilon^{-6})$ .*

*Proof.* Let us assume  $\varepsilon \geq n^{-1/6}$  for general  $G$  and  $\varepsilon \geq n^{-1}$  for bipartite  $G$ , as the update time of  $\tilde{O}(\varepsilon^{-41})$ ,  $\hat{O}(\varepsilon^{-17})$ , and  $\hat{O}(\varepsilon^{-6})$  can be obtained by re-running the static algorithm of [\[35\]](#) after every update. Letting  $\mathcal{X} \stackrel{\text{def}}{=} \mathcal{M}_G$ , [Theorem 3.1](#) with accuracy parameter  $\varepsilon' \stackrel{\text{def}}{=} \varepsilon/8$  and  $\mu \stackrel{\text{def}}{=} \frac{\varepsilon'}{128 \log((n^4 W)/\varepsilon')}$  shows that by using [Theorem 3.3](#) with accuracy parameter  $\frac{\mu \varepsilon'^2}{512}$  inside [Algorithm 2](#) as the subroutine **Rebuild()**, there will be at most  $\tilde{O}(\varepsilon^{-2})$  calls to **Rebuild()** until the graph becomes empty. As such, we can maintain a  $(1 - \varepsilon/8)$ -approximate fractional maximum weight matching in amortized update times  $\tilde{O}(\varepsilon^{-41})$  and  $\hat{O}(\varepsilon^{-17})$  for general  $G$ . As in the proof of [Theorem 3.4](#) for bipartite  $G$ , the update time improves to  $\hat{O}(\varepsilon^{-2})$  by using [\[33, Theorem 10.16\]](#) instead of [Theorem 3.3](#).

Consider the  $t$ -th rebuild, where the current graph is  $G^{(t)} = (V, E^{(t)})$ . Until the next rebuild, let  $E' \subseteq E^{(t)}$  be the current edge set, i.e., the adversary has deleted  $E^{(t)} \setminus E'$  from the graph. To round the fractional matching into an integral matching, right after the  $t$ -th rebuild we run [Lemma 6.10](#) to maintain an  $\varepsilon/8$ -degree-sparsifier  $H$  of  $\mathbf{x}_{E' \cap E_{\varepsilon/8}^{(t)}}(\mathbf{x})$ , where  $\mathbf{x} \stackrel{\text{def}}{=} \mathbf{x}^{(t)}$  is the new fractional matching that we just got from **Rebuild()**. We then run the static algorithm of [\[35\]](#) to get a  $(1 - \varepsilon/4)$ -approximate matching  $M$  in  $H$  in time  $\tilde{O}(n\varepsilon^{-3})$ , since  $|H| \leq \tilde{O}(n\varepsilon^{-2})$  by the sparsity guarantee of [Lemma 6.10](#). We use  $M$  as the output integral matching, and let  $\nu$  be the value of  $\mathbf{w}(M)$  right after running [\[35\]](#). Until the next rebuild happens, for every deletion, we feed the update of  $\mathbf{x}_{E' \cap E_{\varepsilon/8}^{(t)}}(\mathbf{x})$  to [Lemma 6.10](#) to maintain a degree sparsifier. Whenever the deletions make  $\mathbf{w}(M) < (1 - \varepsilon/8)\nu$ , we query the degree-sparsifier-maintainer of [Lemma 6.10](#) to get a new  $H$  over which we reconstruct an integral  $(1 - \varepsilon/4)$ -approximate matching  $M$  and its value  $\nu$  using [\[35\]](#).

We first analyze the quality of  $M$ . By [Lemma 4.7](#) the  $\mathbf{x}^{(t)}$  returned by **Rebuild()** satisfies<sup>22</sup>

$$\sum_{e \in E^{(t)}} \mathbf{w}_e \left| \mathbf{x}_e^{(t)} - \left( \mathbf{x}_{E^{(t)}, \gamma}^\mu \right)_e \right| \leq \frac{\varepsilon}{8} Z_{E^{(t)}, \gamma}^\mu.$$

By definition of [Algorithm 1](#), until the next rebuild happens, we have

$$\sum_{e \in E'} \mathbf{w}_e \mathbf{x}_e^{(t)} \geq \left(1 - \frac{\varepsilon'}{2}\right) \sum_{e \in E^{(t)}} \mathbf{w}_e \mathbf{x}_e^{(t)} = \left(1 - \frac{\varepsilon}{16}\right) \sum_{e \in E^{(t)}} \mathbf{w}_e \mathbf{x}_e^{(t)}.$$

Since  $\mu \leq \frac{\varepsilon}{64 \log(8n^4 W/\varepsilon)}$ , the conditions of [Lemma 6.14](#) are satisfied, showing that the  $\varepsilon/8$ -degree-sparsifier  $H$  of  $\mathbf{x}_{E' \cap E_{\varepsilon/8}^{(t)}}(\mathbf{x})$  we maintain indeed has  $M_{\mathbf{w}}^*(H) \geq (1 - \frac{\varepsilon}{2}) \left( \sum_{e \in E'} \mathbf{w}_e \mathbf{x}_e^{(t)} \right)$ . This implies that

$$(6.29) \quad \mathbf{w}(M) \geq \left(1 - \frac{\varepsilon}{4}\right) \left(1 - \frac{\varepsilon}{2}\right) \left( \sum_{e \in E'} \mathbf{w}_e \mathbf{x}_e^{(t)} \right) \geq \left(1 - \frac{3\varepsilon}{4}\right) \left(1 - \frac{\varepsilon}{16}\right) \sum_{e \in E^{(t)}} \mathbf{w}_e \mathbf{x}_e^{(t)} \geq \left(1 - \frac{7\varepsilon}{8}\right) M_{\mathbf{w}}^*(G)$$

holds right after we run [\[35\]](#) since  $\mathbf{x}^{(t)}$  was a  $(1 - \varepsilon'/2)$ -approximate fractional matching after the rebuild. Because we re-run [\[35\]](#) whenever  $\mathbf{w}(M) < (1 - \frac{\varepsilon}{8})\nu$ , we have that  $M$  is always a  $(1 - \varepsilon)$ -approximate maximum weight matching.

We now analyze the additional update time spent in rounding. By [\(6.29\)](#), whenever we have to re-run [\[35\]](#), we must have deleted a set of edges  $D \subseteq E^{(t)} \setminus E'$  from  $M$  whose weights sum to at least  $\sum_{e \in D} \mathbf{w}_e \geq \frac{\varepsilon}{8}\nu$ . Because  $M \subseteq E_{\varepsilon/8}^{(t)}(\mathbf{x}^{(t)})$ , we have  $\mathbf{x}_e^{(t)} \geq \Omega(\varepsilon/n)$  for all  $e \in M$ , and thus the adversary also deletes  $\Omega(\nu \cdot \varepsilon^2/n)$  units of weight from  $\mathbf{x}^{(t)}$ , i.e.,  $\sum_{e \in D} \mathbf{w}_e \mathbf{x}_e^{(t)} \geq \Omega(\nu \cdot \varepsilon^2/n)$ . As a result, until the next rebuild of  $\mathbf{x}^{(t)}$  happens (i.e., when

<sup>22</sup>Note that there is a change of notation here, so  $\sum_{e \in E^{(t)}} \mathbf{w}_e \left| \mathbf{x}_e^{(t)} - \left( \mathbf{x}_{E^{(t)}, \gamma}^\mu \right)_e \right|$  is the same as  $\left\| \mathbf{x}^{(t)} - \mathbf{x}_{S^{(t)}, \gamma}^\mu \right\|_{\mathbf{w}, S^{(t)}}$  from [Section 4](#). See also [Table 1](#)

the weight of  $\mathbf{x}^{(t)}$  drops by an  $\Theta(\varepsilon)$  fraction), we will run [35] at most  $\tilde{O}(n/\varepsilon)$  times. Since there are  $\tilde{O}(\varepsilon^{-2})$  rebuilds of  $\mathbf{x}^{(t)}$  by Theorem 3.1, rounding incurs an  $\tilde{O}(n^2\varepsilon^{-6})$  additional running time, which is amortized to  $\tilde{O}((n^2/m) \cdot \varepsilon^{-6})$  time per update. This proves the update times of the algorithms. Finally, the algorithms are output-adaptive as Lemma 6.10 is.  $\square$

## Acknowledgements

Thank you to Aditi Dudeja for coordinating the posting to arXiv. Thank you to Aaron Bernstein, Sayan Bhattacharya, Arun Jambulapati, Peter Kiss, Yujia Jin, Thatchaphol Saranurak, Kevin Tian, and David Wajc for helpful conversations at various stages of the project that ultimately led to this paper. Part of the work for this paper was conducted while the authors were visiting the Simons Institute for the Theory of Computing.

## References

- [1] Sepehr Assadi, Aaron Bernstein, and Aditi Dudeja. Incremental matching in general graphs. In *49th International Colloquium on Automata, Languages, and Programming, ICALP 2022*, volume 229 of *LIPIcs*, pages 11:1–11:19. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2022. Available at <https://arxiv.org/abs/2207.00927>
- [2] Sepehr Assadi, Soheil Behnezhad, Sanjeev Khanna, and Huan Li. On regularity lemma and barriers in streaming and dynamic matching. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing, STOC 2023*, pages 131–144. ACM, 2023. Available at <https://arxiv.org/abs/2207.09354>.
- [3] Amir Azarmehr, Soheil Behnezhad, and Mohammad Roghani. Fully dynamic matching:  $(2 - \sqrt{2})$ -approximation in polylog update time. In *Proceedings of the 35th ACM-SIAM Symposium on Discrete Algorithms, SODA 2024*, 2024. Available at <https://arxiv.org/abs/2307.08772>.
- [4] Moab Arar, Shiri Chechik, Sarel Cohen, Cliff Stein, and David Wajc. Dynamic matching: Reducing integral algorithms to approximately-maximal fractional algorithms. In *45th International Colloquium on Automata, Languages, and Programming, ICALP 2018*, volume 107 of *LIPIcs*, pages 7:1–7:16. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2018. Available at <https://arxiv.org/abs/1711.06625>.
- [5] Kook Jin Ahn and Sudipto Guha. Near linear time approximation schemes for uncapacitated and capacitated b-matching problems in nonbipartite graphs. In *Proceedings of the 25th Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2014*, pages 239–258. SIAM, 2014. Available at <https://arxiv.org/abs/1307.4355>.
- [6] Sepehr Assadi and Sanjeev Khanna. Improved bounds for fully dynamic matching via ordered ruzsa-szemerédi graphs. In *arXiv Preprint*, 2024. Available at <https://arxiv.org/abs/2406.13573>.
- [7] Amir Abboud, Jason Li, Debmalya Panigrahi, and Thatchaphol Saranurak. All-pairs max-flow is no harder than single-pair max-flow: Gomory-hu trees in almost-linear time. In *64th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2023*, pages 2204–2212. IEEE Computer Society, 2023.
- [8] Kyriakos Axiotis, Aleksander Madry, and Adrian Vladu. Faster sparse minimum cost flow by electrical flow localization. In *62nd IEEE Annual Symposium on Foundations of Computer Science, FOCS 2021*, pages 528–539. IEEE Computer Society, 2021. Available at <https://arxiv.org/abs/2111.10368>.
- [9] Amir Abboud and Virginia Vassilevska Williams. Popular conjectures imply strong lower bounds for dynamic problems. In *55th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2014*, pages 434–443. IEEE Computer Society, 2014. Available at <https://arxiv.org/abs/1402.0054>.
- [10] Sayan Bhattacharya, Deeparnab Chakrabarty, and Monika Henzinger. Deterministic fully dynamic approximate vertex cover and fractional matching in  $O(1)$  amortized update time. In *Integer Programming and Combinatorial Optimization - 19th International Conference, IPCO 2017*, volume 10328 of *Lecture Notes in Computer Science*, pages 86–98. Springer, 2017. Available at <https://arxiv.org/abs/1611.00198>.
- [11] Soheil Behnezhad, Mahsa Derakhshan, MohammadTaghi Hajiaghayi, Cliff Stein, and Madhu Sudan. Fully dynamic maximal independent set with polylogarithmic update time. In *60th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2019*, pages 382–405. IEEE Computer Society, 2019. Available at <https://arxiv.org/abs/1909.03478>.
- [12] Aaron Bernstein, Aditi Dudeja, and Zachary Langley. A framework for dynamic matching in weighted graphs. In *Proceedings of the 53rd Annual ACM Symposium on Theory of Computing, STOC 2021*, pages 668–681. ACM, 2021.
- [13] Soheil Behnezhad. Dynamic algorithms for maximum matching size. In *Proceedings of the 34th ACM-SIAM Symposium on Discrete Algorithms, SODA 2023*, pages 129–162. SIAM, 2023. Available at <https://arxiv.org/abs/2207.07607>.
- [14] Aaron Bernstein, Sebastian Forster, and Monika Henzinger. A deamortization approach for dynamic spanner and dynamic maximal matching. In *Proceedings of the 30th Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2019*, pages 1899–1918. SIAM, 2019. Available at <https://arxiv.org/abs/1810.10932>.

- [15] Soheil Behnezhad and Alma Ghafari. Fully dynamic matching and ordered ruzsa-szemerédi graphs. In *65th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2024*. IEEE Computer Society, 2024. Available at <https://arxiv.org/abs/2404.06069>
- [16] Surender Baswana, Manoj Gupta, and Sandeep Sen. Fully dynamic maximal matching in  $O(\log n)$  update time. In *52nd IEEE Annual Symposium on Foundations of Computer Science, FOCS 2011*, pages 383–392. IEEE Computer Society, 2011. Available at <https://arxiv.org/abs/1103.1109>
- [17] Aaron Bernstein, Maximilian Probst Gutenberg, and Thatchaphol Saranurak. Deterministic decremental reachability, scc, and shortest paths via directed expanders and congestion balancing. In *61st IEEE Annual Symposium on Foundations of Computer Science, FOCS 2020*, pages 1123–1134. IEEE Computer Society, 2020. Available at <https://arxiv.org/abs/2009.02584>
- [18] Sayan Bhattacharya, Monika Henzinger, and Giuseppe F. Italiano. Deterministic fully dynamic data structures for vertex cover and matching. In *Proceedings of the 26th Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2015*, pages 785–804. SIAM, 2015. Available at <https://arxiv.org/abs/1412.1318>
- [19] Sayan Bhattacharya, Monika Henzinger, and Danupon Nanongkai. New deterministic approximation algorithms for fully dynamic matching. In *Proceedings of the 48th Annual ACM Symposium on Theory of Computing, STOC 2016*, pages 398–411. ACM, 2016. Available at <https://arxiv.org/abs/1604.05765>
- [20] Sayan Bhattacharya, Monika Henzinger, and Danupon Nanongkai. Fully dynamic approximate maximum matching and minimum vertex cover in  $O(\log^3 n)$  worst case update time. In *Proceedings of the 28th Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2017*, pages 470–489. SIAM, 2017. Available at <https://arxiv.org/abs/1704.02844>
- [21] Sayan Bhattacharya and Janardhan Kulkarni. Deterministically maintaining a  $(2 + \epsilon)$ -approximate minimum vertex cover in  $O(1/\epsilon^2)$  amortized update time. In *Proceedings of the 30th Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2019*, pages 1872–1885. SIAM, 2019. Available at <https://arxiv.org/abs/1805.03498>
- [22] Sayan Bhattacharya and Peter Kiss. Deterministic rounding of dynamic fractional matchings. In *48th International Colloquium on Automata, Languages, and Programming, ICALP 2021*, volume 198 of *LIPIcs*, pages 27:1–27:14. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2021. Available at <https://arxiv.org/abs/2105.01615>
- [23] Soheil Behnezhad and Sanjeev Khanna. New trade-offs for fully dynamic matching via hierarchical EDCS. In *Proceedings of the 33rd ACM-SIAM Symposium on Discrete Algorithms, SODA 2022*, pages 3529–3566. SIAM, 2022. Available at <https://arxiv.org/abs/2201.02905>
- [24] Joakim Blikstad and Peter Kiss. Incremental  $(1 - \epsilon)$ -approximate dynamic matching in  $O(\text{poly}(1/\epsilon))$  update time. In *31st Annual European Symposium on Algorithms, ESA 2023*, volume 274 of *LIPIcs*, pages 22:1–22:19. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2023. Available at <https://arxiv.org/abs/2302.08432>
- [25] Sayan Bhattacharya, Peter Kiss, and Thatchaphol Saranurak. Dynamic  $(1 + \epsilon)$ -approximate matching size in truly sublinear update time. In *64th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2023*, 2023. Available at <https://arxiv.org/abs/2302.05030>
- [26] Sayan Bhattacharya, Peter Kiss, and Thatchaphol Saranurak. Dynamic algorithms for packing-covering lps via multiplicative weight updates. In *Proceedings of the 34th ACM-SIAM Symposium on Discrete Algorithms, SODA 2023*, pages 1–47. SIAM, 2023. Available at <https://arxiv.org/abs/2207.07519>
- [27] Sayan Bhattacharya, Peter Kiss, and Thatchaphol Saranurak. Sublinear algorithms for  $(1.5 + \epsilon)$ -approximate matching. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing, STOC 2023*, pages 254–266. ACM, 2023.
- [28] Sayan Bhattacharya, Peter Kiss, Thatchaphol Saranurak, and David Wajc. Dynamic matching with better-than-2 approximation in polylogarithmic update time. In *Proceedings of the 34th ACM-SIAM Symposium on Discrete Algorithms, SODA 2023*, pages 100–128. SIAM, 2023. Available at <https://arxiv.org/abs/2207.07438>
- [29] Sayan Bhattacharya, Peter Kiss, Aaron Sidford, and David Wajc. Near-optimal dynamic rounding of fractional matchings in bipartite graphs. In *arXiv Preprint*, 2023. Available at <https://arxiv.org/abs/2306.11828>
- [30] Bartłomiej Bosek, Dariusz Leniowski, Piotr Sankowski, and Anna Zych. Online bipartite matching in offline time. In *55th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2014*, pages 384–393. IEEE Computer Society, 2014.
- [31] Aaron Bernstein and Cliff Stein. Fully dynamic matching in bipartite graphs. In *42nd International Colloquium on Automata, Languages, and Programming, ICALP 2015*, volume 9134 of *Lecture Notes in Computer Science*, pages 167–179. Springer, 2015. Available at <https://arxiv.org/abs/1506.07076>
- [32] Aaron Bernstein and Cliff Stein. Faster fully dynamic matchings with small approximation ratios. In *Proceedings of the 27th Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2016*, pages 692–711. SIAM, 2016.
- [33] Li Chen, Rasmus Kyng, Yang P. Liu, Richard Peng, Maximilian Probst Gutenberg, and Sushant Sachdeva. Maximum flow and minimum-cost flow in almost-linear time. In *63rd IEEE Annual Symposium on Foundations of Computer Science, FOCS 2022*, pages 612–623. IEEE Computer Society, 2022. Available at <https://arxiv.org/abs/2203.00671>



- [34] Moses Charikar and Shay Solomon. Fully dynamic almost-maximal matching: Breaking the polynomial worst-case time barrier. In *45th International Colloquium on Automata, Languages, and Programming, ICALP 2018*, volume 107 of *LIPIcs*, pages 33:1–33:14. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2018. Available at <https://arxiv.org/abs/1711.06883>
- [35] Ran Duan and Seth Pettie. Linear-time approximation for maximum weight matching. *J. ACM*, 61(1):1:1–1:23, 2014.
- [36] Ran Duan, Seth Pettie, and Hsin-Hao Su. Scaling algorithms for weighted matching in general graphs. In *Proceedings of the 28th Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2017*, pages 781–800. SIAM, 2017. Available at <https://arxiv.org/abs/1411.1919>.
- [37] Aditi Dudeja. Decremental matching in general weighted graphs. In *51st International Colloquium on Automata, Languages, and Programming, ICALP 2024*, volume 297 of *LIPIcs*, pages 59:1–59:20. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2024. Available at <https://arxiv.org/abs/2402.03068>.
- [38] Aditi Dudeja. A note on rounding matchings in general graphs. In *arXiv Preprint*, 2024. Available at <https://arxiv.org/abs/2402.03068>.
- [39] Jack Edmonds. Paths, trees, and flowers. *Canadian Journal of mathematics*, 17:449–467, 1965.
- [40] Fabrizio Grandoni, Stefano Leonardi, Piotr Sankowski, Chris Schwiegelshohn, and Shay Solomon.  $(1 + \epsilon)$ -approximate incremental matching in constant deterministic amortized time. In *Proceedings of the 30th Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2019*, pages 1886–1898. SIAM, 2019.
- [41] Manoj Gupta and Richard Peng. Fully dynamic  $(1 + \epsilon)$ -approximate matchings. In *54th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2013*, pages 548–557. IEEE Computer Society, 2013. Available at <https://arxiv.org/abs/1304.0378>.
- [42] Fabrizio Grandoni, Chris Schwiegelshohn, Shay Solomon, and Amitai Uzdav. Maintaining an EDCS in general graphs: Simpler, density-sensitive and with worst-case time bounds. In *5th Symposium on Simplicity in Algorithms, SOSA 2022*, pages 12–23. SIAM, 2022. Available at <https://arxiv.org/abs/2108.08825>.
- [43] Harold N. Gabow and Robert Endre Tarjan. A linear-time algorithm for a special case of disjoint set union. In *Proceedings of the 15th Annual ACM Symposium on Theory of Computing, STOC 1983*, pages 246–251. ACM, 1983.
- [44] Harold N. Gabow and Robert Endre Tarjan. Faster scaling algorithms for general graph-matching problems. *J. ACM*, 38(4):815–853, 1991.
- [45] Manoj Gupta. Maintaining approximate maximum matching in an incremental bipartite graph in polylogarithmic update time. In *34th International Conference on Foundation of Software Technology and Theoretical Computer Science, FSTTCS 2014*, volume 29 of *LIPIcs*, pages 227–239. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2014.
- [46] Monika Henzinger, Sebastian Krinninger, Danupon Nanongkai, and Thatchaphol Saranurak. Unifying and strengthening hardness for dynamic problems via the online matrix-vector multiplication conjecture. In *Proceedings of the 47th Annual ACM Symposium on Theory of Computing, STOC 2015*, pages 21–30. ACM, 2015. Available at <https://arxiv.org/abs/1511.06773>.
- [47] Ramesh Hariharan, Telikepalli Kavitha, and Debmalaya Panigrahi. Efficient algorithms for computing all low  $s$ - $t$  edge connectivities and related problems. In *Proceedings of the 18th Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2007*, pages 127–136. SIAM, 2007.
- [48] Ramesh Hariharan, Telikepalli Kavitha, Debmalaya Panigrahi, and Anand Bhalgat. An  $\tilde{O}(mn)$  gomory-hu tree construction algorithm for unweighted graphs. In *Proceedings of the 39th Annual ACM Symposium on Theory of Computing, STOC 2007*, pages 605–614. ACM, 2007.
- [49] Arun Jambulapati, Yujia Jin, Aaron Sidford, and Kevin Tian. Regularized box-simplex games and dynamic decremental bipartite matching. In *49th International Colloquium on Automata, Languages, and Programming, ICALP 2022*, volume 229 of *LIPIcs*, pages 77:1–77:20. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2022. Available at <https://arxiv.org/abs/2204.12721>.
- [50] Peter Kiss. Deterministic dynamic matching in worst-case update time. In *13th Innovations in Theoretical Computer Science Conference, ITCS 2022*, volume 215 of *LIPIcs*, pages 94:1–94:21. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2022. Available at <https://arxiv.org/abs/2108.10461>.
- [51] Tarun Kathuria, Yang P. Liu, and Aaron Sidford. Unit capacity maxflow in almost  $O(m^{4/3})$  time. In *61st IEEE Annual Symposium on Foundations of Computer Science, FOCS 2020*, pages 119–130. IEEE Computer Society, 2020.
- [52] Tsvi Kopelowitz, Seth Pettie, and Ely Porat. Higher lower bounds from the 3sum conjecture. In *Proceedings of the 27th Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2016*, pages 1272–1287. SIAM, 2016. Available at <https://arxiv.org/abs/1407.6756>.
- [53] Yang P. Liu. On approximate fully-dynamic matching and online matrix-vector multiplication. In *65th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2024*. IEEE Computer Society, 2024. Available at <https://arxiv.org/abs/2403.02582>.
- [54] Yin Tat Lee and Aaron Sidford. Path finding methods for linear programming: Solving linear programs in  $\tilde{O}(\sqrt{\text{rank}})$  iterations and faster algorithms for maximum flow. In *55th IEEE Annual Symposium on Foundations of Computer*



- Science*, *FOCS 2014*, pages 424–433. IEEE Computer Society, 2014. Available at <https://arxiv.org/abs/1910.08033>.
- [55] Aleksander Madry. Navigating central path with electrical flows: From flows to matchings, and back. In *54th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2013*, pages 253–262. IEEE Computer Society, 2013. Available at <https://arxiv.org/abs/1307.2205>.
  - [56] Aleksander Madry. Computing maximum flow with augmenting electrical flows. In *57th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2016*, pages 593–602. IEEE Computer Society, 2016. Available at <https://arxiv.org/abs/1608.06016>.
  - [57] Tung Mai, Richard Peng, Anup B. Rao, and Vijay V. Vazirani. Concave flow on small depth directed networks. In *arXiv Preprint*, 2017. Available at <http://arxiv.org/abs/1704.07791>.
  - [58] Silvio Micali and Vijay V. Vazirani. An  $O(\sqrt{|V||E|})$  algorithm for finding maximum matching in general graphs. In *21st IEEE Annual Symposium on Foundations of Computer Science, FOCS 1980*, pages 17–27. IEEE Computer Society, 1980.
  - [59] Yurii Nesterov. *Introductory lectures on convex optimization: A basic course*, volume 87. Springer Science & Business Media, 2003. Available at <https://wwwfr.uni.lu/content/download/92121/1121193/file/NesB.pdf>.
  - [60] Krzysztof Onak and Ronitt Rubinfeld. Maintaining a large matching and a small vertex cover. In *Proceedings of the 42nd ACM Symposium on Theory of Computing, STOC 2010*, pages 457–464. ACM, 2010.
  - [61] Manfred W. Padberg and M. R. Rao. Odd minimum cut-sets and  $b$ -matchings. *Math. Oper. Res.*, 7(1):67–80, 1982.
  - [62] Mohammad Roghani, Amin Saberi, and David Wajc. Beating the folklore algorithm for dynamic matching. In *13th Innovations in Theoretical Computer Science Conference, ITCS 2022*, volume 215 of *LIPIcs*, pages 111:1–111:23. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2022. Available at <https://arxiv.org/abs/2106.10321>.
  - [63] Alexander Schrijver. *Combinatorial optimization: polyhedra and efficiency*, volume 24. Springer, 2003.
  - [64] Shay Solomon. Fully dynamic maximal matching in constant update time. In *57th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2016*, pages 325–334. IEEE Computer Society, 2016. Available at <https://arxiv.org/abs/1604.08491>.
  - [65] Daniel Stubbs and Virginia Vassilevska Williams. Metatheorems for dynamic weighted matching. In *8th Innovations in Theoretical Computer Science Conference, ITCS 2017*, volume 67 of *LIPIcs*, pages 58:1–58:14. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2017.
  - [66] Jan van den Brand, Li Chen, Richard Peng, Rasmus Kyng, Yang P. Liu, Maximilian Probst Gutenberg, Sushant Sachdeva, and Aaron Sidford. A deterministic almost-linear time algorithm for minimum-cost flow. In *64th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2023*, pages 503–514. IEEE Computer Society, 2023. Available at <https://arxiv.org/abs/2309.16629>.
  - [67] Jan van den Brand, Yin Tat Lee, Danupon Nanongkai, Richard Peng, Thatchaphol Saranurak, Aaron Sidford, Zhao Song, and Di Wang. Bipartite matching in nearly-linear time on moderately dense graphs. In *61st IEEE Annual Symposium on Foundations of Computer Science, FOCS 2020*, pages 919–930. IEEE Computer Society, 2020. Available at <https://arxiv.org/abs/2009.01802>.
  - [68] Jan van den Brand, Danupon Nanongkai, and Thatchaphol Saranurak. Dynamic matrix inverse: Improved algorithms and matching conditional lower bounds. In *60th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2019*, pages 456–480. IEEE Computer Society, 2019. Available at <https://arxiv.org/abs/1905.05067>.
  - [69] David Wajc. Rounding dynamic matchings against an adaptive adversary. In *Proceedings of the 52nd Annual ACM Symposium on Theory of Computing, STOC 2020*, pages 194–207. ACM, 2020. Available at <https://arxiv.org/abs/1911.05545>.
  - [70] Da Wei Zheng and Monika Henzinger. Multiplicative auction algorithm for approximate maximum weight bipartite matching. In *Integer Programming and Combinatorial Optimization - 24th International Conference, IPCO 2023*, volume 13904 of *Lecture Notes in Computer Science*, pages 453–465. Springer, 2023. Available at <https://arxiv.org/abs/2301.09217>.

## A Generalizing [5]

As mentioned in Section 5, [5] can be generalized to prove Lemma 5.1 in a fairly straightforward manner. Their results, however, only claimed a  $\text{poly}(\varepsilon^{-1})$  dependence in the runtime instead of an explicit one. As we will swap out certain components in their algorithms with recently developed counterparts, for completeness we give a proof of Lemma 5.1 with explicit dependence on  $\varepsilon^{-1}$ . Most of the proofs in this section are slight variants of analogous proofs in [5], and we do not claim novelty of them.

**Overview.** Here we outline the overall approach of [5] to make it easier to understand what the subsequent lemmas and proofs are about. To begin, note that by Fact 2.1 and concavity of the objective we can focus only on  $\mathcal{M}_{G,\varepsilon}$ . Since the constraints of the matching polytope  $\mathcal{M}_{G,\varepsilon}$  are linear and each  $\mathbf{x} \in \beta \cdot \mathcal{P}$  returned by the oracle only violates these constraints by a factor of  $\kappa_{\mathcal{P}} \cdot \beta$ , the main idea of [5] is to use the multiplicative weights

update (MWU) framework to produce a sequence of vectors in  $\beta \cdot \mathcal{P}$  whose average only violates the constraints by a factor of  $(1 + O(\varepsilon))$ . Here, the width of the MWU algorithms is  $\kappa_{\mathcal{P}} \cdot \beta$ . As long as each vector in the sequence is an approximate maximizer of the objective, the average of them will be one as well (again we use the concavity of the objective). However, classic MWU algorithms need to spend  $\Omega(\text{nnz}(\mathbf{A}))$  time in each iteration where  $\mathbf{A}$  is the constraint matrix, which is exponential for the matching polytope even only considering small odd sets. Thus, to circumvent to issue, [5] showed that (1) a variation of the classic MWU algorithms where only constraints that are violated significantly are evaluated still works and (2) for (a slightly perturbed)  $\mathcal{M}_{G,\varepsilon}$ , these maximally-violated constraints can be found and evaluated efficiently. In what follows we adopt several lemmas from [5], reproving them analogously in the form we need if necessary.

LEMMA A.1. ([5, THEOREM 8(1)]) *Suppose we are given an  $(\beta, T_{\mathcal{A}}, \zeta)$ -oracle  $\mathcal{A}$  for  $(\mathcal{P}, f)$ ,  $\varepsilon \geq \frac{1}{\text{poly}(n)}$ , concave function  $f$ , non-negative linear function  $\ell$  where all  $\ell_e$ 's are polynomially bounded,  $\gamma_1 \geq \Omega(1)$  and  $\gamma_2 \geq 0$ . If*

$$(A.1) \quad \{\mathbf{x} \in \mathcal{P} : f(\mathbf{x}) \geq \gamma_1 \text{ and } \ell(\mathbf{x}) \leq \gamma_2\} \neq \emptyset,$$

*then in  $\tilde{O}(T_{\mathcal{A}}(n, m))$  time we can find an  $\mathbf{x} \in \beta \cdot \mathcal{P}$  such that  $f(\mathbf{x}) \geq \zeta(1 - \varepsilon)\gamma_1$  and  $\ell(\mathbf{x}) \leq \gamma_2$ .*

*Proof.* The proof strategy is the same as that of [5, Lemma 6.1]. Let  $g_{\rho}(\mathbf{x}) \stackrel{\text{def}}{=} f(\mathbf{x}) - \rho \cdot \ell(\mathbf{x})$ . We can use a binary search to find  $0 \leq \rho^- < \rho^+ \leq \frac{\gamma_1}{\gamma_2}$  and  $\mathbf{x}_{\rho^-}, \mathbf{x}_{\rho^+} \in \beta \cdot \mathcal{P}$  such that  $\rho^+ - \rho^- \leq \frac{\varepsilon\gamma_1}{\gamma_2}$  with  $\ell(\mathbf{x}_{\rho^-}) > \gamma_2$  and  $\ell(\mathbf{x}_{\rho^+}) \leq \gamma_2$  in  $\tilde{O}(T_{\mathcal{A}}(n, m))$  time as follows. For  $\rho = 0$ , the  $\mathbf{x}_0$  returned by  $\mathcal{A}$  on input  $g_0$  satisfies  $f(\mathbf{x}_0) \geq \zeta\gamma_1$ . If  $\ell(\mathbf{x}_0) \leq \gamma_2$  then we can simply return  $\mathbf{x}_0$  as the solution. Similarly, for  $\rho = \frac{\gamma_1}{\gamma_2}$ , the vector  $\mathbf{x}_{\gamma_1/\gamma_2} \stackrel{\text{def}}{=} \mathbf{0}$  satisfies  $\ell(\mathbf{x}_{\gamma_1/\gamma_2}) \leq \gamma_2$ . Starting from the left endpoint  $\rho^- = 0$  and the right endpoint  $\rho^+ = \frac{\gamma_1}{\gamma_2}$ , we do a binary search on  $\rho \in [\rho^-, \rho^+]$  using  $\mathcal{A}$  to optimize  $g_{\rho}$ . Let  $\mathbf{x}_{\rho}$  be the returned vector on  $g_{\rho}$ . We maintain that  $\ell(\mathbf{x}_{\rho^-}) > \gamma_2$  and  $\ell(\mathbf{x}_{\rho^+}) \leq \gamma_2$  until  $\rho^+ - \rho^- \leq \frac{\varepsilon\gamma_1}{\gamma_2}$ , i.e., if  $\ell(\mathbf{x}_{\rho}) > \gamma_2$ , then we set  $\rho^- \leftarrow \rho$ ; otherwise, we set  $\rho^+ \leftarrow \rho$ . This takes  $\tilde{O}(T_{\mathcal{A}}(n, m))$  time. To see that each  $g_{\rho}$  is indeed a valid function for  $\mathcal{A}$  to optimize, observe that by the guarantee of (A.1), we have  $\max_{\mathbf{x} \in \mathcal{P}} g_{\rho}(\mathbf{x}) \geq \gamma_1 - \rho\gamma_2$ . Since the binary search is terminated when  $\rho^+ - \rho^- \leq \frac{\varepsilon\gamma_1}{\gamma_2}$ , we have  $\gamma_1 - \rho\gamma_2 \geq \frac{\varepsilon}{2}$  and therefore  $\max_{\mathbf{x} \in \mathcal{P}} g_{\rho}(\mathbf{x}) \geq \frac{1}{\text{poly}(n)}$  for each  $\rho$  that we invoke the oracle. This verifies (5.20). In the end, we take a linear combination of  $\mathbf{x}_{\rho^-}$  and  $\mathbf{x}_{\rho^+}$  to get an  $\mathbf{x} \stackrel{\text{def}}{=} (1 - \alpha)\mathbf{x}_{\rho^-} + \alpha\mathbf{x}_{\rho^+}$  for some  $\alpha \in [0, 1]$ , which is in  $\beta \cdot \mathcal{P}$  by convexity of  $\mathcal{P}$ , with  $\ell(\mathbf{x}) = \gamma_2$ . This  $\mathbf{x}$  is then returned as the solution.

To analyze the quality of  $\mathbf{x}$ , we use by the guarantee of (A.1) that

$$(A.2) \quad g_{\rho^-}(\mathbf{x}_{\rho^-}) \geq \zeta(\gamma_1 - \rho^- \gamma_2) \quad \text{and} \quad g_{\rho^+}(\mathbf{x}_{\rho^+}) \geq \zeta(\gamma_1 - \rho^+ \gamma_2).$$

It then follows that

$$\begin{aligned} f(\mathbf{x}) &\stackrel{(i)}{\geq} (1 - \alpha)f(\mathbf{x}_{\rho^-}) + \alpha f(\mathbf{x}_{\rho^+}) \stackrel{(ii)}{=} (1 - \alpha)(g_{\rho^-}(\mathbf{x}_{\rho^-}) + \rho^- \ell(\mathbf{x}_{\rho^-})) + \alpha(g_{\rho^+}(\mathbf{x}_{\rho^+}) + \rho^+ \ell(\mathbf{x}_{\rho^+})) \\ &\stackrel{(iii)}{\geq} \zeta((1 - \alpha)(\gamma_1 - \rho^- \gamma_2) + \alpha(\gamma_1 - \rho^+ \gamma_2)) + (1 - \alpha)\rho^- \ell(\mathbf{x}_{\rho^-}) + \alpha\rho^+ \ell(\mathbf{x}_{\rho^+}) \\ &\stackrel{(iv)}{\geq} \zeta(\gamma_1 - \alpha \cdot \varepsilon \cdot \gamma_1 - \rho^- \gamma_2) + \ell((1 - \alpha)\rho^- \mathbf{x}_{\rho^-} + \alpha\rho^+ \mathbf{x}_{\rho^+}) \\ &\stackrel{(v)}{\geq} \zeta(1 - \varepsilon)\gamma_1 - \zeta\rho^- \gamma_2 + \rho^- \ell(\mathbf{x}) + \alpha(\rho^+ - \rho^-)\ell(\mathbf{x}_{\rho^+}) \geq \zeta(1 - \varepsilon)\gamma_1, \end{aligned}$$

where (i) is by concavity of  $f$ , (ii) is by definition of  $g_{\rho}$ , (iii) is by (A.2), (iv) is by  $\rho^+ - \rho^- \leq \frac{\varepsilon\gamma_1}{\gamma_2}$  and linearity of  $\ell$ , and (v) is by definition of  $\mathbf{x}$  and again linearity of  $\ell$ . This proves the lemma.  $\square$

In addition to the oracle optimizing the objective, as described in the overview we also need to find constraints that are violated by the current solution significantly. Let

$$(A.3) \quad \widetilde{\mathcal{M}}_{G,\varepsilon} \stackrel{\text{def}}{=} \left\{ \begin{array}{ll} \mathbf{x}(v) \leq \tilde{b}_v, & \forall v \in V \\ \mathbf{x}(B) \leq \tilde{b}_B, & \forall B \in \mathcal{O}_{G,\varepsilon} \end{array} \right\} \cap \mathbb{R}_{\geq 0}^E, \quad \text{where } \tilde{b}_v \stackrel{\text{def}}{=} 1 - 4\varepsilon \text{ and } \tilde{b}_B \stackrel{\text{def}}{=} \left\lfloor \frac{|B|}{2} \right\rfloor - \frac{\varepsilon^2 |B|^2}{4}$$

be a slightly perturbed version of  $\mathcal{M}_{G,\varepsilon}$ . Formally, for  $\mathbf{x} \in \mathbb{R}_{\geq 0}^E$ , let

$$(A.4) \quad \lambda_{\mathbf{x}} \stackrel{\text{def}}{=} \max \left\{ \max_{v \in V} \lambda_{\mathbf{x},v}, \max_{B \in \mathcal{O}_{G,\varepsilon}} \lambda_{\mathbf{x},B} \right\}, \quad \text{where } \lambda_{\mathbf{x},v} \stackrel{\text{def}}{=} \frac{\mathbf{x}(v)}{\tilde{b}_v} \text{ and } \lambda_{\mathbf{x},B} \stackrel{\text{def}}{=} \frac{\mathbf{x}(B)}{\tilde{b}_B}.$$

An algorithm  $\mathcal{B}$  is a  $T_{\mathcal{B}}$ -evaluator for  $\widetilde{\mathcal{M}}_{G,\varepsilon}$  if given any vector  $\mathbf{x} \in \mathbb{R}_{\geq 0}^E$ , when  $\lambda_{\mathbf{x}} > 1 + 8\varepsilon$ , it computes  $\lambda_{\mathbf{x}}$  and the subset of small odd sets  $\mathcal{L}_{\mathbf{x}} \stackrel{\text{def}}{=} \{B \in \mathcal{O}_{G,\varepsilon} : \lambda_{\mathbf{x},B} \geq \lambda_{\mathbf{x}} - \varepsilon^3/10\}$  in  $T_{\mathcal{B}}(n, m)$  time. On the other hand, it correctly identifies when  $\lambda_{\mathbf{x}} \geq 1 + 8\varepsilon$  in the same runtime. Let  $\lambda_{\mathbf{x}, \mathcal{O}_{G,\varepsilon}} \stackrel{\text{def}}{=} \max_{B \in \mathcal{O}_{G,\varepsilon}} \{\lambda_{\mathbf{x},B}\}$ . The following theorem from [5] characterizes the structure of  $\mathcal{L}_{\mathbf{x}}$ .

PROPOSITION A.1. ([5, THEOREM 5]) *If  $\lambda_{\mathbf{x}, \mathcal{O}_{G,\varepsilon}} \geq 1 + 3\varepsilon$ , then  $\mathcal{L}_{\mathbf{x}}$  forms a laminar family.*

We also give the following simple algorithm which helps us compute the multiplicative weights induced by a laminar family without any dependence on  $\varepsilon^{-1}$ .

LEMMA A.2. *Given any  $\mathbf{x} \in \mathbb{R}_{\geq 0}^E$ , a laminar family  $\mathcal{L} \subseteq \mathcal{O}_{G,\varepsilon}$ , and value  $p_B$  associate with each  $B \in \mathcal{L}$ , there is an  $\tilde{O}(m + \sum_{B \in \mathcal{L}} |B|)$  time algorithm that computes*

$$\ell_e \stackrel{\text{def}}{=} \sum_{B \in \mathcal{L}: u, v \in B} p_B$$

for each edge  $e = \{u, v\}$  in  $E$ .

*Proof.* Let  $\mathcal{T}$  be a tree on  $\mathcal{L} \cup \{V\} \cup \{\{v\} : v \in V\}$  such that  $U$  is an ancestor of  $W$  if and only if  $W \subseteq U$ . By laminarity of  $\mathcal{L}$  such a tree exists and can be constructed in  $\tilde{O}(m + \sum_{B \in \mathcal{L}} |B|)$  time. Observe that for each  $e = \{u, v\} \in E$ , the sets of  $B \in \mathcal{L}$  containing both  $u$  and  $v$  correspond to a path from the root of  $\mathcal{T}$  to the lowest common ancestor (LCA) of  $\{u\}$  in  $\{v\}$  in the tree. The LCAs of all edges can be computed in  $\tilde{O}(m)$  time [43], so can the sum of  $p_B$ 's of all root-to-vertex paths be populated in  $O(n)$  time.  $\square$

The following claim is proven in [5].

CLAIM A.1. *It holds that  $\widetilde{\mathcal{M}}_{G,\varepsilon} \subseteq \mathcal{M}_G \subseteq \frac{1}{1-4\varepsilon} \cdot \widetilde{\mathcal{M}}_{G,\varepsilon}$ .*

*Proof.* Consider an  $\mathbf{x} \in \widetilde{\mathcal{M}}_{G,\varepsilon}$ , which clearly satisfies the degree constraints and odd set constraints of size at most  $1/\varepsilon$ . For an odd set  $B \in \mathcal{O}_G$  with  $|B| > 1/\varepsilon$ , it follows that

$$\mathbf{x}(B) \leq \frac{1}{2} \sum_{v \in B} \mathbf{x}(v) \leq \frac{(1-4\varepsilon)|B|}{2} \leq \left\lfloor \frac{|B|}{2} \right\rfloor.$$

That  $\mathcal{M}_G \subseteq \frac{1}{1-4\varepsilon} \cdot \widetilde{\mathcal{M}}_{G,\varepsilon}$  follows from

$$(1-4\varepsilon) \cdot \left\lfloor \frac{|B|}{2} \right\rfloor \leq \left\lfloor \frac{|B|}{2} \right\rfloor - \frac{\varepsilon^2 |B|^2}{4}.$$

$\square$

We now present the algorithm for concave optimization over the matching polytope. [Algorithm 5] is a specialization of [5, Algorithm 2] to the case that the constraint matrices  $\mathbf{A}$  and  $\mathbf{b}$  correspond to that of the perturbed matching polytope  $\widetilde{\mathcal{M}}_{G,\varepsilon}$ , and then modified to work for concave function optimization. The algorithm is divided into *superphases*, where each superphase contains several *phases*, and each phase contains several *iterations*. The only real difference between [Algorithm 5] and [5, Algorithm 1] is that we substitute the subroutine optimizing linear objective (e.g., the solver for LP8 in [5]) to an oracle capable of optimizing general concave objective. Other changes to the algorithm are for consistency with the notations and terminologies that we are using throughout the paper and to highlight when each superphase, phase, and iteration begins and ends. What may appear confusing at first is that [5] used  $\delta$  for the input accuracy parameter and  $\varepsilon$  as the variable used in their algorithm that gradually decreases from  $1/8$  to  $\delta$ . Instead, we use  $\varepsilon$  for the accuracy parameter, and thus the roles of these two symbols are interchanged. Also, [5, Algorithm 2] has other degrees of freedom where they can choose a function  $f(\varepsilon) < \varepsilon$  (which is not to be confused with the concave function  $f$  in our context) and  $\alpha \leq \frac{1}{f(\varepsilon)} \ln \frac{M\lambda_0}{\varepsilon}$ . Nevertheless, for simplicity, we fix  $f(\varepsilon) \stackrel{\text{def}}{=} \varepsilon^3 - 10$  and  $\alpha \stackrel{\text{def}}{=} 50\varepsilon^{-3} \ln n$  as in the uncapacitated  $\mathbf{b}$ -matching algorithm in [5].

---

**Algorithm 5:** Concave function optimization over matching polytope

---

**Input:**  $n$ -vertex  $m$ -edge graph  $G = (V, E)$  and  $\varepsilon \in (3/\sqrt{n}, 1/16)$ .

**Input:** convex, downward closed  $\mathcal{P} \subseteq \mathbb{R}_{\geq 0}^E$  and concave  $f : \mathbb{R}_{\geq 0}^E \rightarrow \mathbb{R}_{-\infty}$ .

**Input:**  $(\beta, T_A, \xi)$ -oracle  $\mathcal{A}$  for  $(\mathcal{P}, f)$  and  $T_B$ -evaluator  $\mathcal{B}$  for  $\mathcal{M}_{G, \varepsilon}$ .

- 1 Let  $\delta \leftarrow \frac{1}{8}$ ,  $\alpha \leftarrow 50\varepsilon^{-3} \ln n$ , and  $\lambda_0 \stackrel{\text{def}}{=} \kappa_{\mathcal{P}} \cdot \beta$ .
  - 2 Use  $\mathcal{A}$  on  $f$  to find an  $\mathbf{x} \in \beta \cdot \mathcal{P}$  with  $f(\mathbf{x}) \geq \max_{\mathbf{x}' \in \mathcal{P}} f(\mathbf{x}')$  and set  $\text{OPT} \stackrel{\text{def}}{=} f(\mathbf{x})$ . Note that  $\lambda_{\mathbf{x}} \leq \lambda_0$  by definition of  $\kappa_{\mathcal{P}}$  and  $\beta$ .
  - 3 **while** *true* **do**
  - 4   Compute  $\lambda_{\mathbf{x}}$  and  $\mathcal{L}_{\mathbf{x}}$  using the evaluator  $\mathcal{B}$  in  $T_B(n, m)$  time.<sup>23</sup>
  - 5   **if**  $\lambda_{\mathbf{x}} \leq 1 + 8\varepsilon$  **then return**  $\mathbf{x}_f \stackrel{\text{def}}{=} \frac{\mathbf{x}}{1+8\varepsilon}$ .
  - 6   Repeatedly set  $\delta \leftarrow \max\{2\delta/3, \varepsilon\}$  until  $\lambda_{\mathbf{x}} > 1 + 8\delta$ .
  - 7   Let  $\mathbf{p}_v \stackrel{\text{def}}{=} \exp(\alpha\lambda_{\mathbf{x},v} - \alpha\lambda_{\mathbf{x}}) / \tilde{b}_v$  for  $v \in V$  with  $\lambda_{\mathbf{x},v} \geq \lambda_{\mathbf{x}} - \varepsilon^3/10$ , and  $\mathbf{p}_v \stackrel{\text{def}}{=} 0$  otherwise.
  - 8   Let  $\mathbf{p}_B \stackrel{\text{def}}{=} \exp(\alpha\lambda_{\mathbf{x},B} - \alpha\lambda_{\mathbf{x}}) / \tilde{b}_B$  for  $B \in \mathcal{L}$ , and  $\mathbf{p}_B \stackrel{\text{def}}{=} 0$  for  $B \notin \mathcal{L}_{\mathbf{x}}$ .
  - 9   Let  $\gamma_{\mathbf{x}} \stackrel{\text{def}}{=} \sum_{v \in V} \mathbf{p}_v \tilde{b}_v + \sum_{B \in \mathcal{O}_{G, \varepsilon}} \mathbf{p}_B \tilde{b}_B$ .
  - 10   Compute  $\ell_e \stackrel{\text{def}}{=} p_u + p_v + \sum_{B \in \mathcal{L}_{\mathbf{x}}: u, v \in B} p_B$  for each  $e = \{u, v\}$  in  $\tilde{O}(m)$  time by [Lemma A.2](#), using that  $\mathcal{L}_{\mathbf{x}}$  forms a laminar family.
  - 11   Use [Lemma A.1](#) with  $f$ ,  $\gamma_1 \stackrel{\text{def}}{=} \text{OPT}$ ,  $\ell(\mathbf{x}') \stackrel{\text{def}}{=} \ell^\top \mathbf{x}'$ , and  $\gamma_2 \stackrel{\text{def}}{=} \gamma_{\mathbf{x}}$  to find an  $\tilde{\mathbf{x}} \in \beta \cdot \mathcal{P}$  such that  $f(\tilde{\mathbf{x}}) \geq \zeta(1 - \varepsilon)\text{OPT}$  and  $\ell(\tilde{\mathbf{x}}) \leq \gamma_{\mathbf{x}}$  in  $\tilde{O}(T_A(n, m))$  time.
  - 12   **while** [Lemma A.1](#) failed to find such an  $\tilde{\mathbf{x}}$  **do**
  - 13    |  $\text{OPT} \leftarrow (1 - \varepsilon)\text{OPT}$  and re-run [Lemma A.1](#) with the new OPT.
  - 14   Update  $\mathbf{x}$  by  $\mathbf{x}_e \leftarrow (1 - \sigma)\mathbf{x}_e + \sigma\tilde{\mathbf{x}}_e$  where  $\sigma \stackrel{\text{def}}{=} \frac{\delta}{4\alpha\lambda_0}$ .
- 

LEMMA A.3. ([\[5\]](#) ALGORITHM 2, THEOREM 16, AND LEMMA 18]) For any  $\varepsilon \geq \tilde{\Omega}(n^{-1/2})$ , downward closed, convex  $\mathcal{P} \subseteq \mathbb{R}_{\geq 0}^E$ , concave function  $f : \mathbb{R}_{\geq 0}^E \rightarrow \mathbb{R}$  such that  $\max_{\mathbf{x} \in \mathcal{P} \cap \mathcal{M}_G} f(\mathbf{x}) \geq 1$ , given an  $(\beta, T_A, \zeta)$ -oracle  $\mathcal{A}$  for  $(\mathcal{P}, f)$  and a  $T_B$ -evaluator  $\mathcal{B}$  for  $\mathcal{M}_{G, \varepsilon}$ , if  $\kappa_{\mathcal{P}} \cdot \beta < n$ , then [Algorithm 5](#) takes

$$\tilde{O}((T_A(n, m) + T_B(n, m)) \cdot \beta \kappa_{\mathcal{P}} \varepsilon^{-4})$$

time and computes an  $\mathbf{x}_f \in (\beta \cdot \mathcal{P}) \cap \mathcal{M}_G$  such that  $f(\mathbf{x}_f) \geq \zeta(1 - 13\varepsilon) \max_{\mathbf{x} \in \mathcal{P} \cap \mathcal{M}_G} f(\mathbf{x})$ .

*Proof.* Let us assume  $\varepsilon \leq 1/16$ . Let  $f^* \stackrel{\text{def}}{=} \max_{\mathbf{x} \in \mathcal{P} \cap \mathcal{M}_G} f(\mathbf{x})$  and  $\mathbf{x}^* \stackrel{\text{def}}{=} \arg\max_{\mathbf{x} \in \mathcal{P} \cap \mathcal{M}_G} f(\mathbf{x})$ . The vector  $\mathbf{x}$  in [Algorithm 5](#) is in  $\beta \cdot \mathcal{P}$  at all times, and thus  $\mathbf{x}_f \in \beta \cdot \mathcal{P}$  as well. [Lemma A.1](#) never fails after OPT decreases to  $(1 - 4\varepsilon)f^*$ : we can take  $\mathbf{x}^* \cdot (1 - 4\varepsilon)$  with  $f(\mathbf{x}^* \cdot (1 - 4\varepsilon)) \geq (1 - 4\varepsilon)f^*$  and by [Claim A.1](#),  $\mathbf{x}^* \cdot (1 - 4\varepsilon) \in \mathcal{P} \cap \mathcal{M}_{G, \varepsilon}$  and in particular  $\ell(\mathbf{x}^* \cdot (1 - 4\varepsilon)) \leq \gamma$  is guaranteed. This shows that indeed the value of  $\gamma_1$  in [Lemma A.1](#) is  $\Omega(1)$  as  $f(\mathbf{x}^*) \geq \Omega(1)$ . Also, all  $\tilde{\mathbf{x}}$ 's returned by [Lemma A.1](#) have  $f(\tilde{\mathbf{x}}) \geq \zeta(1 - \varepsilon)(1 - 4\varepsilon)f^*$ , proving  $f(\mathbf{x}) \geq \zeta(1 - \varepsilon)(1 - 4\varepsilon)f^*$  and therefore  $f(\mathbf{x}_f) \geq \frac{\zeta(1 - \varepsilon)(1 - 4\varepsilon)}{1 + 8\varepsilon} f^* \geq \zeta(1 - 13\varepsilon)f^*$  by concavity of  $f$ . At the end of the algorithm, we have that  $\lambda_{\mathbf{x}} \leq 1 + 8\varepsilon$  and thus  $\mathbf{x}_f \in \mathcal{M}_{G, \varepsilon} \subseteq \mathcal{M}_G$  by [Claim A.1](#). The numbers  $\mathbf{p}_v$ 's are polynomially bounded since  $\lambda_{\mathbf{x},v} \geq \lambda_{\mathbf{x}} - \varepsilon^3/10$  and thus  $n^{-5} \leq \exp(\alpha\lambda_{\mathbf{x},v} - \alpha\lambda_{\mathbf{x}}) \leq 1$ . Similarly  $\mathbf{p}_B$ 's are polynomially bounded. This ensures that  $\ell(\mathbf{x}) \stackrel{\text{def}}{=} \ell^\top \mathbf{x}$  is a valid input to [Lemma A.1](#).

It thus remains to bound the number of iterations in [Algorithm 5](#) until it terminates since each iteration takes  $O(T_A(n, m) + T_B(n, m))$  time. Let  $\Phi_{\mathbf{x}} \stackrel{\text{def}}{=} \sum_{v \in V} \exp(\alpha\lambda_{\mathbf{x},v}) + \sum_{B \in \mathcal{O}_{G, \varepsilon}} \exp(\alpha\lambda_{\mathbf{x},B})$  as defined in [\[5\]](#) Definition 4]. [\[5\]](#) Theorem 16] showed that [\[5\]](#) Algorithm 2] converges in  $\tilde{O}(\lambda_0 \cdot (\varepsilon^{-2} + \alpha\varepsilon^{-1}))$  iterations if we start with  $\Phi_{\mathbf{x}} \leq \gamma_{\mathbf{x}} + \frac{\varepsilon\gamma_{\mathbf{x}}}{\lambda_0}$ . This can be verbatim carried to [Algorithm 5](#) since the analysis is oblivious to the function  $f$  we are optimizing: it applies as long as the vector  $\tilde{\mathbf{x}}$  we found in each iteration satisfies  $\ell^\top \tilde{\mathbf{x}} \leq \gamma_{\mathbf{x}}$ , which is guaranteed by [Lemma A.1](#). Roughly speaking, in [\[5\]](#) Lemma 15] they showed that  $\Phi_{\mathbf{x}}$  decreases after every iteration if the initial condition  $\Phi_{\mathbf{x}} \leq \gamma_{\mathbf{x}} + \frac{\varepsilon\gamma_{\mathbf{x}}}{\lambda_0}$  is satisfied. The proof only used how  $\Phi_{\mathbf{x}}$  can change for the new  $\mathbf{x}$  given by

$\mathbf{x}_e \leftarrow (1 - \sigma)\mathbf{x}_e + \sigma\tilde{\mathbf{x}}_e$  via the fact that  $\ell^\top \tilde{\mathbf{x}} \leq \gamma_{\mathbf{x}}$ . [5, Theorem 16] then used [5, Lemma 15] to argue the total decrease of  $\Phi_{\mathbf{x}}$  after  $\tilde{O}(\lambda_0 \cdot (\varepsilon^{-2} + \alpha\varepsilon^{-1}))$ . Finally, we use [5, Lemma 18] which showed that indeed for the specific value of  $\alpha$  that we choose, when  $\lambda_0 < n$  the initial condition  $\Phi_{\mathbf{x}} \leq \gamma_{\mathbf{x}} + \frac{\varepsilon\gamma_{\mathbf{x}}}{\lambda_0}$  is satisfied. This shows that the number of iterations [Algorithm 5] has is  $\tilde{O}(\kappa_{\mathcal{P}}\beta\varepsilon^{-4})$ .  $\square$

**A.1 Finding Maximally-Violated Constraints** It remains to give an evaluator for  $\widetilde{\mathcal{M}}_{G,\varepsilon}$  that finds all the maximally-violated odd sets. [5, Lemma 17] reduces this to the case where  $\lambda_{\mathbf{x},\mathcal{O}_{G,\varepsilon}} \geq 1 + 3\varepsilon$ . In this regime, [5, Theorem 5] applies (recall that it says  $\mathcal{L}_{\mathbf{x}}$  forms a laminar family), and [5, Theorem 6] presents an  $\tilde{O}(m + n \cdot \text{poly}(\varepsilon^{-1}))$  algorithm. We briefly sketch the algorithm and analyze its dependence on  $\varepsilon^{-1}$  in the remainder of the section.

The algorithm uses a binary search to find an estimate  $\tilde{\lambda}$  such that  $\tilde{\lambda} - \frac{\varepsilon^3}{100} \leq \lambda_{\mathbf{x},\mathcal{O}_{G,\varepsilon}} \leq \tilde{\lambda}$ , where  $\lambda_{\mathbf{x},\mathcal{O}_{G,\varepsilon}} \stackrel{\text{def}}{=} \max_{B \in \mathcal{O}_{G,\varepsilon}} \lambda_{\mathbf{x},B}$ . Fix a current value of  $\tilde{\lambda}$ . For each odd  $3 \leq \ell \leq 1/\varepsilon$ , [5] constructs an integral weighted graph  $G_\varphi(\ell, \tilde{\lambda})$  with  $V(G_\varphi(\ell, \tilde{\lambda})) = V(G) \cup \{s\}$  with  $O(\min\{m, n\varepsilon^{-5}\})$  edges whose weights sum up to  $O(n\varepsilon^{-5})$  in  $O(m)$  time. Let  $\text{cut}(B)$  for  $B \subseteq V$  be the sum of weights of edges between  $B$  and  $(V(G) \cup \{s\}) \setminus B$  in  $G_\varphi(\ell, \tilde{\lambda})$ . If  $\tilde{\lambda}$  is an accurate estimate, i.e.,  $\tilde{\lambda} - \frac{\varepsilon^3}{100} \leq \lambda_{\mathbf{x},\mathcal{O}_{G,\varepsilon}} \leq \tilde{\lambda}$ , then [5, Property 1] showed that

- (1) every  $\ell$ -sized odd set  $B \in \mathcal{L}_{\mathbf{x}}$  satisfies  $\text{cut}(B, \bar{B}) < \kappa(\ell)$ , where  $\kappa(\ell) \stackrel{\text{def}}{=} \left\lceil \varphi \tilde{\lambda} \left(1 - \frac{\varepsilon^2 \ell^2}{2}\right) \right\rceil + \frac{12\ell}{\varepsilon} + 1 < 2\varphi$  for  $\varphi \stackrel{\text{def}}{=} 50\varepsilon^{-4}$ , and
- (2) every  $\ell$ -sized odd set  $B \in \mathcal{O}_{G,\varepsilon}$  that satisfies  $\text{cut}(B, \bar{B}) < \kappa(\ell)$  belongs to the collection  $\mathcal{L}'_{\mathbf{x}} \stackrel{\text{def}}{=} \{B \in \mathcal{O}_{G,\varepsilon} : \lambda_{\mathbf{x},B} \geq \lambda_{\mathbf{x},\mathcal{O}_{G,\varepsilon}} - \varepsilon^3\}$ .

They then applied the following algorithm of [Lemma A.4] below with  $\kappa \stackrel{\text{def}}{=} \kappa(\ell) = O(\varepsilon^{-4})$  to obtain a collection of vertex sets  $\mathcal{L}$ , for which they showed that  $\mathcal{L}_{\mathbf{x}} \subseteq \mathcal{L}$  and therefore we can extract all size- $\ell$  sets in  $\mathcal{L}_{\mathbf{x}}$  in  $\tilde{O}(m)$  time by simply checking their value of  $\lambda_{\mathbf{x},B}$ .<sup>24</sup> The value of  $\tilde{\lambda}$  can then be adjusted based on whether any size- $\ell$  set in  $\mathcal{L}_{\mathbf{x}}$  is found (recall that we are doing a binary search to determine an accurate estimate  $\tilde{\lambda}$ ).

To compute the collection  $\mathcal{L}$ , [5] used the minimum odd-cut approach of [61] using the construct of partial Gomory-Hu trees. A  $\kappa$ -partial Gomory-Hu tree of a (possibly weighted) graph  $G$  is a partition  $\mathcal{U} = \{U_1, \dots, U_k\}$  of  $V(G)$  and a weighted tree  $\mathcal{T}$  on  $\mathcal{U}$  such that each  $x, y \in U_i$  belonging to the same set has minimum cut value greater than  $\kappa$  in  $G$ , and each  $x \in U_i$  and  $y \in U_j$  in different sets has minimum cut (both the value and the actual cut) equal to that induced by  $\mathcal{T}$ .

LEMMA A.4. ([5, LEMMA 19 AND ALGORITHM 3]) *Given a  $T(n, m, \kappa)$ -time algorithm  $\mathcal{G}$  that constructs a  $\kappa$ -partial Gomory-Hu tree of an  $n$ -vertex  $m$ -edge graph, there is an algorithm that computes an  $\tilde{O}(n)$ -sized collection  $\mathcal{L}$  of odd sets in  $G$  where  $s \in V(G)$  in  $\tilde{O}(T(n, m, \kappa))$  time such that (i)  $s \notin B$  for every  $B \in \mathcal{L}$ , (ii)  $E_G(B, \bar{B}) \leq \kappa$  for every  $B \in \mathcal{L}$ , and (iii) every odd set  $B'$  not containing  $s$  with  $E_G(B', \bar{B}') \leq \kappa$  intersects with some  $B \in \mathcal{L}$ . Moreover,  $\mathcal{L}$  is of the form  $\mathcal{L}_1 \cup \dots \cup \mathcal{L}_{O(\log n)}$ , where each  $\mathcal{L}_i$  contains disjoint vertex subsets. The algorithm is deterministic if  $\mathcal{G}$  is.*

*Proof.* The lemma is the same as [5, Lemma 19] modulo the additional guarantee that  $\mathcal{L}$  is the union  $O(\log n)$  collections of disjoint sets. This is manifest from the implementation of [5, Algorithm 3], using that it has  $O(\log n)$  iterations, and in each of them the sets it found are disjoint.  $\square$

We deliberately make the statement of [Lemma A.4] flexible to the choice of partial Gomory-Hu tree algorithm, given that faster algorithms were developed recently. [5] used the following algorithm in the regime where  $\kappa$  is small.

LEMMA A.5. ([48, 47]) *There is a randomized  $\tilde{O}(m + n\kappa^2)$  time algorithm for constructing a  $\kappa$ -partial Gomory-Hu tree on an  $m$ -edge  $n$ -vertex unweighted graph.*

<sup>24</sup>Since  $\mathcal{L} = \mathcal{L}_1 \cup \dots \cup \mathcal{L}_{O(\log n)}$  where sets in  $\mathcal{L}_i$  are disjoint by [Lemma A.4] for each  $1 \leq i \leq O(\log n)$  we can compute  $\lambda_{\mathbf{x},B}$  for  $B \in \mathcal{L}_i$  in  $O(m)$  time.



Alternatively, we can use the recent almost-linear time Gomory-Hu tree algorithm which does not depend on  $\kappa$  at the cost of having a subpolynomial factor  $m^{o(1)}$ . Note that contracting edges greater than  $\kappa$  in a Gomory-Hu tree trivially gives a  $\kappa$ -partial Gomory-Hu tree.

LEMMA A.6. ([7]) *There is a randomized  $m^{1+o(1)}$  time algorithm that constructs a Gomory-Hu tree on an  $m$ -edge weighted graph w.h.p.*

LEMMA A.7. ([5, THEOREM 6 AND LEMMA 17]) *There are randomized  $\tilde{O}(m\varepsilon^{-1} + n\varepsilon^{-9})$ -evaluator and  $\tilde{O}(m\varepsilon^{-1})$ -evaluator for  $\widetilde{\mathcal{M}}_{G,\varepsilon}$ .*

*Proof.* The runtime of the algorithms can be analyzed as follows. The binary search takes  $\tilde{O}(1)$  iterations. In each iteration,  $O(\varepsilon^{-1})$  values of  $\ell$  are enumerated, and for each of them we either use Lemma A.5 for constructing partial Gomory-Hu tree in Lemma A.4, resulting in an  $\tilde{O}(m + n\varepsilon^{-8})$  time algorithm for computing size- $\ell$  sets in  $\mathcal{L}_x$ , or we use Lemma A.6 and get an  $\tilde{O}(m)$  time algorithm.

By [5, Lemma 17], if  $\lambda_x > 1 + 8\varepsilon$  but  $\lambda_{x, O_{G,\varepsilon}} < 1 + 3\varepsilon$ , then  $\mathcal{L}_x = \emptyset$ . Since the above algorithms based on [5, Theorem 6] correctly identify  $\mathcal{L}_x$  when  $\lambda_{x, O_{G,\varepsilon}} \geq 1 + 3\varepsilon$ , by checking whether there is indeed an odd set  $B$  returned with  $\lambda_{x,B} \geq 1 + 3\varepsilon$  we can deduce whether we should return an empty set or not. Similarly, we can also deduce whether  $\lambda_x \leq 1 + 8\varepsilon$  by inspecting if there is a vertex or returned odd set which is violated by a  $1 + 8\varepsilon$  factor by  $x$ .  $\square$

With Lemmas A.3 and A.7, Lemma 5.1 follow by appropriately adjusting the parameter  $\varepsilon$ .

REMARK A.1. *From the discussion above in combination with the deterministic rounding algorithm in Section 6 we can also see that to obtain a deterministic  $\tilde{O}_\varepsilon(1)$  update time decremental matching algorithm, it suffices to get a deterministic,  $\tilde{O}_\kappa(m)$  time algorithm for constructing  $\kappa$ -partial Gomory-Hu tree. We leave this as an interesting open question and future direction.*

## B Omitted Proofs in Section 6

In this section we provide proofs of the following lemmas.

LEMMA 6.4. (ANALOGOUS TO [29, LEMMA 3.10])  *$\text{supp}(x^{(i)}) \subseteq \text{supp}(x)$  holds for each  $L_{\min} \leq i \leq L_{\max}$  after each operation.*

*Proof.* By reverse induction on  $i$  we show that  $\text{supp}(x^{(i)}) \subseteq \text{supp}(x^{(i+1)}) \subseteq \text{supp}(x)$ . From (3.5) we have  $\text{supp}(x^{(i)}) = F_i \cup \bigcup_{0 \leq j \leq L_{\min}} \text{supp}_j(x) \cup \bigcup_{L_{\min}+1 \leq j \leq i} E_j$  and thus it suffices to show that  $F_i \subseteq \text{supp}(x)$  and  $E_j \subseteq \text{supp}(x)$  for all  $L_{\min} + 1 \leq j \leq i$  at all times. That  $E_j \subseteq \text{supp}(x)$  is immediate as  $E_j$  is set to  $\text{supp}_j(x)$  in `Rebuild()`, and in each subsequent update we will remove  $e$  from  $E_j$  if  $x_e$  is changed. By the property of `degree-split`, we have  $F_i \subseteq E_{i+1} \cup F_{i+1}$ , which by the inductive hypothesis satisfies  $E_{i+1} \cup F_{i+1} \subseteq \text{supp}(x^{(i+1)}) \subseteq \text{supp}(x)$ , and thus  $F_i \subseteq \text{supp}(x^{(i+1)}) \subseteq \text{supp}(x)$  after a call to `Rebuild()`. Similarly, in each update after the rebuild, we will remove  $e$  from  $F_i$  whose  $x_e$  is changed, and thus the containment is maintained. This proves the lemma.  $\square$

LEMMA 6.5. (ANALOGOUS TO [29, LEMMA 3.11])  *$w^\top \tilde{x} \geq (1 - \varepsilon)w^\top x$  holds after each operation.*<sup>25</sup>

*Proof.* Choosing the subgraph returned by `degree-split` with large weight ensures that, right after a call to `Rebuild(i)`, we have  $w(F_{i-1}) \geq \frac{w(E_i) + w(F_i)}{2}$  and thus  $w^\top x^{(i-1)} \geq w^\top x^{(i)}$  by that

$$(B.5) \quad x_e^{(i-1)} - x_e^{(i)} = F_{i-1}(e) \cdot 2^{-(i-1)} - (F_i(e) + E_i(e)) \cdot 2^{-i}.$$

On the other hand, there are at most  $2^{i-2} \cdot \frac{\varepsilon \cdot w^\top x}{L_{\max} \cdot W}$  updates between two calls to `Rebuild(i)` and thus at most  $2^{i-1} \cdot \frac{\varepsilon \cdot w^\top x}{L_{\max} \cdot W}$  edges are deleted from  $F_{i-1}$ , decreasing  $w^\top x^{(i-1)}$  by at most  $\frac{\varepsilon}{L_{\max}} \cdot w^\top x$ . By (B.5), any change

<sup>25</sup>Unlike [29], we assume the input matching  $x$  is already preprocessed by Observation 6.1 and thus the bound stated here is  $(1 - \varepsilon)$  instead of  $(1 - 2\varepsilon)$ .



to  $E_j$  for  $j \neq i$  does not affect the difference between  $\mathbf{w}^\top \mathbf{x}^{(i-1)}$  and  $\mathbf{w}^\top \mathbf{x}^{(i)}$ . Furthermore,  $E_i$  and  $F_i$  can only decrease until **Rebuild**( $i$ ) is called (if **Rebuild**( $i+1$ ) is called and  $F_i$  increases consequently, then **Rebuild**( $i$ ) will be called as well). This shows except for operations that remove edges from  $F_{i-1}$ ,  $\mathbf{w}^\top \mathbf{x}^{(i-1)} - \mathbf{w}^\top \mathbf{x}^{(i)}$  cannot decrease, and as such

$$(B.6) \quad \mathbf{w}^\top \mathbf{x}^{(i-1)} \geq \mathbf{w}^\top \mathbf{x}^{(i)} - \frac{\varepsilon}{L_{\max}} \cdot \mathbf{w}^\top \mathbf{x}$$

holds at all times for each  $i \in \{L_{\min} + 1, \dots, L_{\max}\}$ . Chaining (B.6) for all  $i$  and noticing that  $\mathbf{x} = \mathbf{x}^{(L_{\max})}$  and  $\tilde{\mathbf{x}} \stackrel{\text{def}}{=} \mathbf{x}^{(L_{\min})}$  conclude the proof.  $\square$

LEMMA 6.6. (ANALOGOUS TO [29, LEMMA 3.13]) *The amortized time per **Update**( $e, \nu$ ) of Algorithm 4 is  $O(W \cdot \varepsilon^{-1} \cdot L_{\max}^2)$ .*

*Proof.* We first show using a similar argument as the proof of Lemma 6.3 that  $|F_i| \leq O(2^i \cdot \|\mathbf{x}\|_1)$ . By backward induction on  $i$  we prove

$$(B.7) \quad |F_i| \leq \left\lceil 2^i \cdot \sum_{j=i+1}^{L_{\max}} |S_j| \cdot 2^{-j} \right\rceil.$$

For  $i = L_{\max}$  this is trivially true. For  $i < L_{\max}$ , right after a call to **Rebuild**( $i+1$ ) when  $E_{i+1} \leftarrow S_{i+1}$ , by Property (1) of the subroutine **degree-split** we have

$$|F_i| \leq \left\lceil \frac{|S_{i+1}| + |F_{i+1}|}{2} \right\rceil \leq \left\lceil \frac{|S_{i+1}| + \left\lceil 2^{i+1} \cdot \sum_{j=i+2}^{L_{\max}} |S_j| \cdot 2^{-j} \right\rceil}{2} \right\rceil \leq \left\lceil 2^i \cdot \sum_{j=i+1}^{L_{\max}} |S_j| \cdot 2^{-j} \right\rceil.$$

On the other hand, for an update after which **Rebuild**( $i+1$ ) is not called, as in the proof of Lemma 6.3, if the right-hand side decreases by one, then we will remove an extra edge from  $F_i$  and thus the inequality still holds. This proves the (B.7) which in turn implies  $|F_i| \leq O(2^i \cdot \|\mathbf{x}\|_1)$ .

This shows that the call to **degree-split** in **Rebuild**( $i$ ) takes  $O(2^i \cdot \|\mathbf{x}\|_1)$  time. Since **Rebuild**( $i$ ) causes **Rebuild**( $i-1$ ), the total running time of **Rebuild**( $i$ ) is  $\sum_{j=0}^i O(2^j \cdot \|\mathbf{x}\|_1) = O(2^i \cdot \|\mathbf{x}\|_1)$ . Because **Rebuild**( $i$ ) is called once every  $2^{i-2} \cdot \frac{\varepsilon \cdot \mathbf{w}^\top \mathbf{x}}{L_{\max} \cdot W} \geq 2^{i-2} \cdot \frac{\varepsilon \cdot \|\mathbf{x}\|_1}{L_{\max} \cdot W}$  updates, the amortized update time of Algorithm 4 is  $\tilde{O}(W \cdot \varepsilon^{-2} \cdot L_{\max})$ .  $\square$