A Fast Distributed Algorithm for $\alpha-$Fair Packing Problems

Jelena Marašević$^1$, Cliff Stein$^2$, Gil Zussman$^1$

$^1$Department of Electrical Engineering, $^2$Department of Industrial Engineering and Operations Research
Columbia University
{jelena@ee, cliff@ieor, gil@ee}.columbia.edu

Abstract
Over the past two decades, fair resource allocation problems received considerable attention in a variety of application areas. While polynomial time distributed algorithms have been designed for max-min fair resource allocation, the design of distributed algorithms with convergence guarantees for the more general $\alpha-$fair allocations received little attention. In this paper, we study weighted $\alpha$-fair packing problems, that is, the problems of maximizing the objective functions $\sum_j w_j x_j^{\frac{1-\alpha}{(1-\alpha)}}$ when $\alpha \neq 1$ and $\sum_j w_j \ln x_j$ when $\alpha = 1$ over linear constraints $Ax \leq b$, $x \geq 0$, where $w_j$ are positive weights and $A$ and $b$ are non-negative. We consider the distributed computation model that was used for packing linear programs and network utility maximization problems. Under this model, we provide a distributed algorithm for general $\alpha$. The algorithm uses simple local update rules and is stateless (namely, it allows asynchronous updates, is self-stabilizing, and allows incremental and local adjustments). It converges to approximate solutions in running times that have an inverse polynomial dependence on the approximation parameter $\varepsilon$. The convergence time has polylogarithmic dependence on the problem size for $\alpha \neq 1$, and a nearly-linear dependence on the number of variables for $\alpha = 1$. These are the best convergence times known for these problems.
1 Introduction

Over the past two decades, fair resource allocation problems have received considerable attention in a variety of application areas. Perhaps the most familiar are networking applications such as Internet congestion control, rate control in software defined networks, and scheduling in wireless networks [26, 29, 35]. However, fair resource allocation has many other applications, such as those that arise in traditional operations research settings (see [9] for an overview), and in economics and game theory, as in the work by Jain and Vazirani on Eisenberg-Gale markets [16].

While polynomial time distributed algorithms have been designed for max-min fair resource allocation problems (see e.g., [11] and the follow-up work), the design of distributed algorithms with convergence guarantees for the more general case of $\alpha$-fair allocations has received very little attention. Classical control-theoretic approaches for network congestion control problems yield algorithms that are guaranteed to converge after a "finite" time [18, 20, 26, 32, 35, 38]. Nonetheless, their convergence time as a function of the input size is poorly understood.

On the other hand, significant recent progress has been made in the design of efficient distributed algorithms for linear programs with packing constraints [1, 5, 6, 23, 27, 39]. Linear programming problems can be interpreted as a special case of $\alpha$-fair resource allocation problems with $\alpha = 0$ (see below). Therefore, a natural question arises whether distributed algorithms for the more general class of $\alpha$-fair packing problems can be solved as efficiently by combining ideas from packing linear programming algorithms [1, 5, 6, 23, 27, 39] and network congestion control [18, 20, 26, 32, 35, 38]. The answer to this question is the main focus of this paper.

In general, $\alpha$-fairness provides a trade-off between efficiency (sum of allocated resources) and fairness (minimum allocated resource) as a function of $\alpha$: the higher the $\alpha$, the better the fairness guarantees and the lower the efficiency [20, 25]. According to the definition in [32], for a vector of positive weights $w$ and $\alpha > 0$, an allocation vector $x^*$ of size $n$ is weighted $\alpha$-fair (also referred to as $(w, \alpha)$-proportionally fair or $(w, \alpha)$-fair), if for any alternative feasible vector $x$: $\sum_j w_j x_j \frac{x_j-x_j^*}{(x_j^*)^\alpha} \leq 0$. For a compact and convex feasible region, $x^*$ can be equivalently defined as a vector that solves the problem of maximizing the $w$-weighted sum of $\alpha$-fair utilities $f_\alpha(x_j)$, where $f_\alpha(x_j) = x_j^{1-\alpha}/(1-\alpha)$ for $\alpha \neq 1$ and $f_\alpha(x_j) = \ln(x_j)$ for $\alpha = 1$ [32]. $\alpha$-fairness subsumes three special cases: (i) $\alpha = 1$ – known as proportional fairness [18], (ii) $\alpha = 0$ – a utilitarian resource allocation, i.e., $f_\alpha(x_j)$ is linear in $x_j$, and (iii) $\alpha \to \infty$ that converges to the max-min fair solution [32], the most egalitarian resource allocation. In the bargaining theory, proportional fairness can be interpreted as a Nash solution [34] and max-min fairness corresponds to the Kalai-Smorodinsky solution [17] (see [9] for discussion).

We consider the problem of efficient distributed weighted $\alpha$-fair packing, namely, the problem of efficiently solving $\max \{ \sum_j w_j f_\alpha(x_j) : Ax \leq b, x \geq 0 \}$ in a distributed manner. We adopt the model of distributed computation that was used for packing linear programs [1, 5, 6, 23, 27, 36]. Under this model, an agent $j$ controls the variable $x_j$ and has information about: (i) the $j$th column of the constraint matrix $A$, (ii) weight $w_j$, (iii) upper bounds on global problem parameters $m, n, w_{\text{max}}$, and $A_{\text{max}}$, and (iv) relative slack of each constraint in which $x_j$ appears with a non-zero coefficient, in each round. We remark that this model of distributed computation is a generalization of the model considered in network congestion control problems [20] (see Section 2.3).

Our Results. We provide a unified algorithm for the general class of distributed weighted $\alpha$-fair resource allocation problems subject to positive linear (packing) constraints. Similar to [15], we state the algorithm and prove the convergence results for the normalized version of the problem in the following form: $\max \{ \sum_j w_j f_\alpha(x_j) : Ax \leq 1, x \geq 0 \}$, where all non-zero elements $A_{ij}$ of the matrix $A$
satisfy \( A_{ij} \geq 1 \). Adopting such a scaled norm is without loss of generality (see Appendices A and B).

To the best of our knowledge, our algorithm is the fastest distributed algorithm for weighted \( \alpha \)-fair packing problems developed to date. Our main results are summarized in the following theorem, where for technical reasons we assume that \( \alpha \) is bounded away from 0 and 1 whenever \( \alpha \neq 1 \). A more detailed statement of the results appears in Theorems 4.6, 4.7, and 4.8.

**Theorem 1.1.** (Main Result) For a given weighted \( \alpha \)-fair packing problem \( \max \{ \sum_j w_j f_\alpha(x_j) : Ax \leq 1, x \geq 0 \} \), there exists a stateless and distributed algorithm (Algorithm 1) that computes: (i) a \((1 + \varepsilon)\)-approximate solution in \( O(\frac{\ln(n m A_{\max})}{\varepsilon^2}) \) rounds for \( \alpha \neq 1 \), and (ii) an additive \( \varepsilon \sum_j w_j \)-approximate solution in \( O(\frac{n \ln(n m A_{\max})}{\varepsilon^2}) \) rounds for \( \alpha = 1 \).

Our algorithm is stateless according to the definition by Awerbuch and Khandekar [3–5]: it starts from any initial state, the agents update the variables \( x_j \) in a cooperative but uncoordinated manner, reacting only to the current state of the constraints that they can observe, and without an access to a global clock. Statelessness implies a number of desirable properties of a distributed algorithm, such as: asynchronous updates, self-stabilization, and incremental and local adjustments [3–5]. For the purpose of clear analysis, Algorithm 1 is presented in Section 3 in a non-stateless form. However, minor modifications that do not affect the analysis make this algorithm stateless (see Appendix E).

Algorithm 1 associates a dual variable \( y_i \) with each constraint \( i \), so that \( y_i \) is an exponential function of the \( i \)th constraint’s relative slack: \( y_i = C \cdot \exp(\kappa(\sum_j A_{ij} x_j - 1)) \). A similar form of dual variables has been used for packing linear programming algorithms [1, 5, 6, 13, 14, 22, 37, 39]. However, due to the different objective functions, linear programs adopt \( C = 1 \), while in our case \( C \) is a function of the global input parameters \( \alpha, w_{\max}, n, m, A_{\max} \).

Algorithm 1 leverages non-obvious connections between a control-theoretic algorithm for network congestion control in the \( \alpha = 1 \) case (by Kelly et al. [18]) and a packing linear programming algorithm for the \( \alpha = 0 \) case (by Awerbuch and Khandekar [5]). The primal algorithm in [18] is described as a set of differential equations that guide primal updates of each variable \( x_j \) proportional to the difference \( w_j - x_j^\alpha \sum_i y_i A_{ij} \). The dual variables \( y_i \) are not fully specified, but only required to be non-negative continuous functions of \( \sum_j A_{ij} x_j \). For the choice of dual variables \( y_i = C \cdot \exp(\kappa(\sum_j A_{ij} x_j - 1)) \) as in our work, the Lyapunov function used for showing finite convergence time in [18] is equivalent (up to a constant additive term) to the potential function used for showing polylogarithmic convergence time of the packing linear program [5]. At the same time, the primal updates in the algorithm of [5] are similar to those from [18]: each primal variable \( x_j \) is updated by a constant multiplicative factor whenever \( w_j = x_j^\alpha \sum_i y_i A_{ij} \) (\( \Leftrightarrow 1 = \sum_i y_i A_{ij} \) in [5]) is not approximately satisfied.

Our algorithm is similar in spirit to the algorithm for packing linear programs of [5]: it maintains a primal and dual feasible solutions and updates each primal variable \( x_j \) whenever a Karush-Kuhn-Tucker (KKT) condition \( x_j^\alpha \sum_i y_i A_{ij} = w_j \) is not approximately satisfied. Our update rule is slightly different: we generally employ a multiplicative update, but use an additional threshold value \( \delta_j \) to make sure that \( x_j \) does not become too small. There are significant challenges that needed to be overcome in order to obtain results for the non-linear case. One major issue is due to different KKT conditions that guide the updates of the primal variables: as already mentioned, for \( \alpha \)-fair objectives the condition is determined by \( x_j^\alpha \cdot \sum_{i=1}^m y_i A_{ij} \), while for the linear objectives [5] the equivalent term is \( \sum_{i=1}^m y_i A_{ij} \). While this may seem as a minor difference, it is actually the main reason why Algorithm 1 (unlike the algorithm of [5]) disallows any variable \( x_j \) to be decreased below a fixed threshold \( \delta_j \). If there were no thresholds \( \delta_j \), then it would be difficult to claim that the algorithm always maintains a feasible solution; in fact, the argument used for proving Lemma 4.1 would not be valid. On the other hand, the existence of the

---

\(^2\)For the purpose of presentation, convergence time bounds provided here are looser than the actual bounds that we obtain. For tighter bounds, see Section 4.4.

\(^3\)The primal algorithm in [18] uses slightly different notation and is provided for \( \alpha = 1 \) and \( A_{ij} \in \{0, 1\} \). We adapt the notation and the results from [18] to our model.

2
lower threshold $\delta_j$ implies that we cannot lower-bound the change in every variable by a multiplicative factor $(1 \pm \beta)$ – the change in the variables that decrease to $\delta_j$ in some round may be very small. Non-multiplicative change in $x_j$ complicates the argument for a multiplicative increase in the potential function, and requires ideas that were not needed in [5]. Moreover, as will be discussed later, $\alpha$–fair objectives for $\alpha \geq 1$ are significantly different than the linear objectives, requiring novel arguments in the proof of the algorithm’s convergence.

The convergence analysis of Algorithm 1 relies on the appropriately chosen potential function that is bounded from below and from above for $x_j \in [\delta_j, 1] \forall j$, and that increases with every primal update. A similar potential function was used for packing linear programs in [1–5], and (as already mentioned) is equivalent to the Lyapunov function of [18]. The main idea in the analysis is to show that whenever a solution $x$ does not approximate well the optimal one, the potential function increases substantially.

In particular, the convergence proofs of the $\alpha < 1$ and $\alpha > 1$ use suitably chosen definitions of stationary intervals and stationary rounds, respectively. The stationary interval (respectively the stationary round), is roughly a time when the $x_j$ variables do not change much and is chosen so that it guarantees near-optimality of the current solution. Polylogarithmic convergence time is then obtained by showing that the potential function increases multiplicatively within any non-stationary interval (non-stationary round).

In the $\alpha = 1$ case, the arguments used in the proofs of $\alpha < 1$ and $\alpha > 1$ cases do not apply. We are not able to obtain a polylogarithmic convergence time in the input size, mainly due to the fact that we bound the convergence time by lower-bounding the increase in the potential due to a single variable update, instead of aggregate arguments used for $\alpha < 1$ and $\alpha > 1$ cases.

The convergence proof of the $\alpha < 1$ case follows the same line of argument as the linear programming algorithm [5], but is not a straightforward extension of [5]. Our setting requires additional results to obtain a multiplicative increase in the potential function, and the appropriate choice of the algorithm parameters is essential to obtaining all the intermediate results. The same set of arguments does not lead to the convergence proof for the cases $\alpha = 1$ and $\alpha > 1$, mainly because in these two cases the function $f_\alpha(x_j)$ is negative throughout the feasible region. More details are provided in Section 4.

**Related Work.** Traditionally, the literature in the area of algorithm design has mostly focused on the two limit cases of $\alpha$–fair objectives: linear programming and max-min fairness. Efficient algorithms for packing linear programs have been widely studied both in centralized [14, 22, 37] and distributed [1–5, 6, 23, 27, 39] settings. Max-min fair resource allocations have been studied in terms of various network flow problems both in centralized [5, 15, 21, 24, 28, 30] and distributed [11] settings.

Very little attention has been devoted to the design of efficient algorithms for the general class of $\alpha$–fair objectives. Classical work on distributed rate control algorithms in the networking literature uses control-theoretic approach to optimize $\alpha$–fair objectives. This approach has been extensively studied and applied to various network settings [18, 20, 26, 32, 35, 38]. However, it does not lead to guaranteed convergence time: the convergence results state that the algorithm converges after some “finite time” where it is unclear whether this “finite time” is polynomial as a function of the input size.

Since $\alpha$–fair objectives are concave, their optimization over a region determined by linear constraints is solvable in polynomial time in centralized setting through convex programming (see, e.g., [10]). Distributed gradient methods for network utility maximization problems such as [7, 33] can be employed to address the problem of $\alpha$–fair packing. However, the convergence time of these algorithms depends on the dual gradient’s Lipschitz constant to produce good approximations. This, in general, leads to a polynomial convergence time as a function of $n$ and $A_{\max}$. Moreover, the algorithms in [7, 33] lack desirable properties of distributed algorithms such as asynchronous updates and self-stabilization.
2 Preliminaries

2.1 Weighted $\alpha$-Fairness

Consider the following optimization problem with positive linear (packing) constraints:

\[
(Q_\alpha) \quad \max \left\{ \sum_{j=1}^{n} w_j f_\alpha(x_j) : Ax \leq b, x \geq 0 \right\}, \quad \text{where } f_\alpha(x_j) = \begin{cases} \ln(x_j), & \text{if } \alpha = 1 \\ x_j^{-\alpha}, & \text{if } \alpha \neq 1 \end{cases}
\]

where \(x = (x_1, ..., x_n)\) is the vector of variables, \(A\) is the \(m \times n\) constraint matrix with non-negative elements, and \(b = (b_1, ..., b_m)\) is a vector with strictly positive elements. We will refer to the problem \((Q_\alpha)\) as the weighted \(\alpha\)-fair packing.

The following definition and lemma introduced by Mo and Walrand [32] characterize weighted \(\alpha\)-fair allocations. In the rest of the paper we will use the terms weighted \(\alpha\)-fair and \(\alpha\)-fair interchangeably.

**Definition 2.1.** [32] Let \(w = (w_1, ..., w_n)\) and \(\alpha\) be positive numbers. A vector \(x^* = (x_1^*, ..., x_n^*)\) is weighted \(\alpha\)-fair if it is feasible and for any other feasible vector \(x: \sum_{j=1}^{n} w_j \frac{x_j - x_j^*}{x_j^*} \leq 0\).

**Lemma 2.2.** [32] A vector \(x^*\) solves \((Q_\alpha)\) for functions \(f_\alpha(x_j^*)\) if and only if it is weighted \(\alpha\)-fair.

Notice in \((Q_\alpha)\) that since \(b_i\)'s are strictly positive, and the partial derivative of the objective with respect to any of the variables \(x_j\) goes to \(-\infty\) as \(x_j \to 0\), the optimal solution will always lie in the positive orthant. Moreover, since the objective is a strictly concave function maximized over a convex region (in this case, a polytope), the optimal solution is unique and \((Q_\alpha)\) satisfies strong duality (see, e.g., [10]). The same observations are true for the scaled version of the problem denoted by \((P_\alpha)\) and introduced in the following subsection.

2.2 Our Problem

We consider the weighted \(\alpha\)-fair packing problem in the following normalized form:

\[
(P_\alpha) \quad \max \left\{ \sum_{j=1}^{n} w_j f_\alpha(x_j) : Ax \leq 1, x \geq 0 \right\},
\]

where \(f_\alpha\) is defined by \([1]\), \(w = (w_1, ..., w_n)\) is a vector of positive weights, \(x = (x_1, ..., x_n)\) is the vector of variables, \(A\) is an \(m \times n\) matrix of non-negative weights, and \(1\) is a size-\(m\) vector of 1's. We let \(A_{\text{max}}\) denote the maximum element of the constraint matrix \(A\), and assume that every element \(A_{ij}\) of \(A\) is non-negative, and moreover, that \(A_{ij} \geq 1\) whenever \(A_{ij} \neq 0\). The maximum weight is denoted by \(w_{\text{max}}\).

We remark that considering problem \((Q_\alpha)\) in the normalized form \((P_\alpha)\) is without loss of generality: any problem \((Q_\alpha)\) can be scaled to the form of \((P_\alpha)\) by (i) dividing both sides of each inequality \(i\) by \(b_i\), and (ii) working with scaled variables \(c \cdot x_j\), where \(c = \min\{1, \min_{i,j: A_{ij} \neq 0} \frac{A_{ij}}{b_i}\}\). Observe that if \(c \cdot x\) solves \((P_\alpha)\), then \(x\) solves \((Q_\alpha)\). Moreover, such scaling preserves the approximation (see Appendix A).

For technical reasons, we assume that the ratio between any two weights is bounded by a constant: \(\frac{w_j}{w_{\text{max}}} = \Omega(1)\) and that the sum of all weights is at least a constant: \(W \equiv \sum_{j=1}^{n} w_j = \Omega(1)\). This is a reasonable assumption for most settings of practical interest.

2.3 Model of Distributed Computation

We adopt the same model of distributed computation as [1, 5, 6, 23, 27, 36], described as follows. We assume that for each \(j \in \{1, ..., n\}\), there is an agent controlling the variable \(x_j\). The agent \(j\) is assumed
to have information about the following problem parameters: (i) the $j^{th}$ column of $A$, (ii) weight $w_j$, and (iii) (an upper bound on) $m, n, w_{\text{max}},$ and $A_{\text{max}}$. In each round, agent $j$ collects the relative slack $e_j$ of all constraints $i$ for which $A_{ij} \neq 0$.

We note that this model of distributed computation is a generalization of the model considered in network congestion control problems [20] where a variable $x_j$ corresponds to the rate of node $j$, $A$ is a 0-1 routing matrix, such that $A_{ij} = 1$, if and only if a node $j$ sends flow over link $i$, and $b$ is the vector of capacities. Under this model, the knowledge about the price of each constraint corresponds to each node collecting (a function of) congestion on each link that it utilizes. Such a model was used in network utility maximization problems with $\alpha-$fair objective [18] and general strongly-concave objectives [7].

2.4 Solution Lower Bound and Duality Gap

Recall (from Section 2.1) that the optimal solution $x^*$ that solves $(P_\alpha)$ must lie in the positive orthant. We show in Lemma 2.3 that not only does $x^*$ lie in the positive orthant, but the minimum element of $x^*$ can be bounded from below as a function of the problem parameters. This lemma motivates the choice of parameters $\delta_j$ in Algorithm 1 (presented in Section 3). The proof is provided in Appendix [B].

**Lemma 2.3.** Let $x^* = (x_1^*, ..., x_n^*)$ be the optimal solution to $(P_\alpha)$. Then $\forall j \in \{1, ..., n\}$:

- $x_j^* \geq \left(\frac{w_j}{w_{\text{max}}^{1/\alpha}} \min_{i: A_{ij} \neq 0} n_i A_{ij}\right)^{1/\alpha}$, if $\alpha \leq 1$,

- $x_j^* \geq A_{\text{max}}^{(1-\alpha)/\alpha} \left(\frac{w_j}{w_{\text{max}}^{1/\alpha}} \min_{i: A_{ij} \neq 0} n_i A_{ij}\right)^{1/\alpha}$, if $\alpha > 1$,

where $n_i = \sum_{j=1}^n I\{A_{ij} \neq 0\}$ is the number of non-zero elements in the $i^{th}$ row of the constraint matrix $A$, and $M = \min\{m, n\}$.

Apart from the lower bound on $\min_j x_j^*$, another useful piece of information for understanding the intuition behind Algorithm 1 and its analysis are the KKT conditions for $(P_\alpha)$ and the duality gap. We will denote the Lagrange multipliers for the problem $(P_\alpha)$ as $y = (y_1, ..., y_m)$ and refer to them as “dual variables”. The KKT conditions for $(P_\alpha)$ are (see Appendix [C]):

- $\sum_{j=1}^n A_{ij} x_j \leq 1 \ \forall i \in \{1, ..., m\}, \quad x_j \geq 0 \ \forall j \in \{1, ..., n\}$ (primal feasibility) (K1)

- $y_i \geq 0 \ \forall i \in \{1, ..., m\}$ (dual feasibility) (K2)

- $y_i \left(\sum_{j=1}^m A_{ij} x_j - 1\right) = 0 \ \forall i \in \{1, ..., m\}$ (complementary slackness) (K3)

- $x_j^\alpha \sum_{i=1}^m y_i A_{ij} = w_j \ \forall j \in \{1, ..., m\}$ (zero-gradient of the Lagrangian) (K4)

The duality gap for $\alpha \neq 1$ is (see Appendix [C]):

$$G_\alpha(x, y) = \sum_{j=1}^n w_j x_j^{1-\alpha} \left(\frac{1}{\alpha} - 1\right) + \sum_{i=1}^m y_i - \sum_{j=1}^m w_j x_j^{1-\alpha} \cdot \frac{\sum_{i=1}^m A_{ij} y_i}{w_j}^{\frac{\alpha}{\alpha-1}},$$

while for $\alpha = 1$:

$$G_1(x, y) = \sum_{j=1}^n w_j \ln(w_j) - \sum_{j=1}^n w_j \ln \left( x_j \sum_{i=1}^m y_i A_{ij} \right) + \sum_{i=1}^m y_i - W.$$
3 Algorithm

The pseudocode of the algorithm that is being run at each node $j$ is provided in Algorithm 1. The basic intuition is that the algorithm keeps KKT conditions $[K1]$ and $[K2]$ satisfied and works towards (approximately) satisfying the remaining two KKT conditions $[K3]$ and $[K4]$ to minimize the duality gap. It is clear that the algorithm can run in distributed setting described in Section 2.3. In each round, every agent $j$ makes updates to their value $x_j$ based on the relative slack of all the constraints in which $j$ takes part, as long as the KKT condition $[\text{K4}]$ of agent $j$ is not approximately satisfied. We remark that it is not necessary that agents update their values synchronously – if the updates are asynchronous, then the convergence time is measured by the number of rounds of the slowest agent.

![Algorithm 1 α-Fair Resource Allocation](image)

Although Algorithm 1 is not stateless, it requires only minor modifications that are provided in Appendix E to become stateless. These modifications do not affect the convergence time analysis. We keep the algorithm in the current form for simplicity of the exposition and for the purpose of the analysis. To allow for self-stabilization and dynamic changes, the algorithm runs forever at all the agents. This is a standard requirement for self-stabilizing algorithms (see, e.g., [12]). The convergence of the algorithm is, therefore, measured as the number of rounds between the round in which the algorithm starts from some feasible solution and the round in which it reaches an approximate solution, assuming that there are no hard reset events or node/constraint insertions/deletions in between. While Algorithm 1 starts from a feasible solution, we remark that (after minor modifications provided in Appendix E) even if the algorithm starts from an arbitrary (not necessarily feasible) solution, it reaches a feasible solution in time that is upper-bounded by its convergence time.

Without loss of generality, we assume that the input parameter $\varepsilon$ that determines the approximation quality satisfies $\varepsilon \leq \min\{\frac{\delta_j}{\epsilon_m}, \frac{\epsilon}{m\alpha}\}$. The parameters $\delta_j, C, \kappa, \gamma,$ and $\beta$ are set as follows. For technical reasons (mainly due to reinforcing dominant multiplicative updates of the variables $x_j$), we set the values of lower thresholds $\delta_j$ below the actual lower bound of the optimal solution given by Lemma 2.3

$$\delta_j = \frac{1}{21/\alpha} \cdot \left\{ \begin{array}{ll} \left(\frac{w_j}{w_{\max}}\frac{1}{n-1}A_{\max}\right)^{1/\alpha}, & \text{if } 0 < \alpha \leq \frac{1}{2} \\ \left(\frac{w_j}{w_{\max}}\frac{1}{n-1}A_{\max}\right)^{1/\alpha} \cdot \frac{1}{n}, & \text{if } \frac{1}{2} < \alpha < 1 \\ \frac{w_j}{w_{\max}}\frac{1}{n-1}A_{\max}, & \text{if } \alpha = 1 \\ \frac{w_j}{w_{\max}}\frac{1}{n-1}A_{\max}^{2-1/\alpha}, & \text{if } \alpha > 1 \end{array} \right.$$  

The constant $C$ that multiplies the exponent in dual variables $y_i$ is chosen as $C = \frac{W}{\sum_{j=1}^{n} \delta_j}$. Because $\delta_j$ only depends on $w_j$ and on gloabal paramters, we also have $C = \frac{w_j}{\delta_j}, \forall j$. The parameter $\kappa$ that appears in the exponent of $y_i$’s is chosen to be $\kappa = \frac{1}{\varepsilon} \ln \left(\frac{C/\alpha A_{\max}}{\min_j w_j}\right)$. The “absolute error” of $[\text{K4}]$ $\gamma$ is set to $\varepsilon/4$.
Finally, similar to \[5\], we choose the value of $\beta$ so that in any round the value of each $\frac{x_j \sum_{i=1}^{m} y_i(x) A_{ij}}{w_j}$ changes by a multiplicative factor of at most $(1 \pm \gamma/4)$. Since the maximum increase over any $x_j$ in each iteration is by a factor $1 + \beta$, and the solution $x$ is feasible in each iteration (see Lemma 4.1), we have that $\sum_{j=1}^{n} A_{ij} x_j \leq 1$, and therefore, the maximum increase in each $y_i$ is by a factor of $e^{\kappa \beta}$. A similar argument holds for the maximum decrease. Hence, the following should be satisfied:

$$(1 + \beta)^\alpha e^{\kappa \beta} \leq 1 + \gamma/4 \quad \text{and} \quad (1 - \beta)^\alpha e^{-\kappa \beta} \geq 1 - \gamma/4.$$  

Observing that $1 + \beta \leq \frac{1}{1 - \gamma}$, to satisfy the first inequality it is enough to satisfy $\frac{1}{(1 - \beta)^\alpha} e^{\kappa \beta} \leq 1 + \frac{\gamma}{4}$, which is equivalent to $(1 - \beta)^\alpha e^{-\kappa \beta} \geq (1 + \frac{\gamma}{4})^{-1} \geq 1 - \frac{\gamma}{4}$. There are two cases: $0 < \alpha \leq 1$ and $\alpha > 1$. If $\alpha \leq 1$, then:

$$(1 - \beta)^\alpha e^{-\kappa \beta} \geq (1 - \beta)(1 - \kappa \beta) = 1 - (\kappa + 1)\beta + \kappa \beta^2,$$

and setting $\beta = \frac{\gamma}{4(\kappa + 1)}$ is sufficient to satisfy both inequalities. If $\alpha > 1$, then:

$$(1 - \beta)^\alpha e^{-\kappa \beta} \geq (1 - \alpha \beta)(1 - \kappa \beta) = 1 - (\kappa + \alpha)\beta + \kappa \alpha \beta^2,$$

and choosing $\beta = \frac{\gamma}{4(\kappa + \alpha)}$ satisfies the desired inequalities. Therefore:

$$\beta = \begin{cases} \frac{\gamma}{4(\kappa + 1)}, & \text{if } \alpha \leq 1 \\ \frac{\gamma}{4(\kappa + \alpha)}, & \text{if } \alpha > 1. \end{cases}$$

### 4 Convergence Analysis

In this section, we analyze the convergence time of Algorithm 1. We will first provide some general results that hold for all $\alpha > 0$, and then we will analyze the algorithm separately for three cases, depending on the value of $\alpha$. For $\alpha < 1$, we follow the general line of argument used in \[5\] to analyze packing linear programs, although several additional insights will be needed. Intuitively, this case is the most similar to the linear case because, for $\alpha < 1$, $f_\alpha(x_j)$ is “similar enough” to the linear function; e.g., in both cases the objective is equal to 0 for $x_j = 0$, positive for $x_j > 0$, and bounded on the interval $[0, 1)$. We also require that $\alpha$ be bounded away from 0 and 1. When $\alpha$ is close to 0, the initial values of $\delta_i$ are also close to 0, and hence, the multiplicative increase will lead to a very small progress in $x_j$. When $\alpha$ approaches 1, the term $1/(1 - \alpha)$ blows up, and our analysis fails. Of course, when $\alpha$ is 0 we can just use the algorithm of \[5\], and as $\alpha$ approaches 1 we are close to the $\alpha = 1$ case of our algorithm.

The $\alpha \geq 1$ cases are more challenging, and do not follow \[5\] as closely. The biggest difference is that when $\alpha \geq 1$, then $f_\alpha(x_j)$ is negative throughout the feasible region, which invalidates most of the convergence proof used in the $\alpha = 0$ case and $\alpha \in (0, 1)$ cases. Furthermore, the argument used for showing near-optimality within stationary intervals is not valid for $\alpha \geq 1$.

For $\alpha > 1$, we introduce a definition of a stationary round, and show that the negative potential function decreases multiplicatively in any non-stationary round, while the solution $x$ in any stationary round is near-optimal. We remark that the arguments used for proving convergence in this case rely on bounding the duality gap and are significantly different from those used for proving convergence of $\alpha = 0$ \[5\] and $\alpha \in (0, 1)$ cases. Similarly as for $\alpha < 1$, we assume that $\alpha$ is bounded away from 1.

For $\alpha = 1$, we bound the convergence time using an additive increase in the potential due to a single variable update, which leads us to a quasi-linear, rather than polylogarithmic convergence time. This case differs significantly from the previous two largely because of a different relationship between the primal objective, sum of dual variables, and $\frac{x_j^\gamma \sum_{i=1}^{m} y_i(x) A_{ij}}{w_j}$. In contrast to the other two cases, the duality gap contains a term $-\sum_j \ln(x_j) \sum_{i=1}^{m} y_i(x) A_{ij}$ (see Eq. \[B\]), and it is sufficient for one term $x_j \sum_{i=1}^{m} y_i(x) A_{ij}$ to be small for the duality gap to become very large. This complicates the use of an
aggregate argument over a sum of terms \( x_j \sum_i y_i(x) A_{ij} \) as in \( \alpha \neq 1 \) cases, since it does not necessarily lead to a small duality gap. Similarly, the term \( x_j^{1-\alpha} \) that appears in the analysis, and that in other cases is related closely to the primal objective, in this case is just 1, which is clearly not as useful.

**Feasibility and Approximate Complementary Slackness.** The following two lemmas are preliminary for the convergence time analysis, and were given in a similar form in [5]. Lemma 4.1 shows that the solution \( x \) is always feasible, while Lemma 4.2 shows that after a polylogarithmic number of rounds approximate complementary slackness (KKT condition [K3]) holds in aggregate sense: 
\[
\sum_{i=1}^{m} y_i(x) \left( \sum_{j=1}^{n} A_{ij} x_j - 1 \right) \approx 0.
\]
While the proofs of these two lemmas follow the same line of argument as [5], we remark that the appropriate choice of algorithm parameters, existence of lower thresholds \( \delta_j \), and the choice of sufficiently large constant C in dual variables \( y_i \) are essential for validity of the arguments. Proofs of Lemmas 4.1 and 4.2 are provided in Appendix D.1.

**Lemma 4.1.** In any round of the algorithm, the solution \( x \) is always feasible: 
\[
\sum_{j=1}^{n} A_{ij} x_j \leq 1, \forall i.
\]

**Lemma 4.2.** After at most \( \tau_0 = \min_{\varepsilon \in \{1, \ldots, n\}} \frac{1}{\beta} \ln \left( \frac{1}{\varepsilon} \right) \) it is always true that:
1. There exists at least one approximately tight constraint: \( \max \{ \sum_{j=1}^{n} A_{ij} x_j \} \geq 1 - (1 + 1/\kappa) \varepsilon, \)
2. \( \sum_{i=1}^{m} y_i \leq (1 + 6 \varepsilon) \sum_{j=1}^{n} x_j \sum_{i=1}^{m} y_i A_{ij} \), and
3. \( (1 - 4 \varepsilon) \sum_{i=1}^{m} y_i \leq \sum_{j=1}^{n} x_j \sum_{i=1}^{m} y_i A_{ij} \leq \sum_{i=1}^{m} y_i. \)

**Decrease of Small Variables.** The following lemma is another preliminary for the convergence analysis. It shows that if there is a variable \( x_j \) that decreases by less than a multiplicative factor \( (1 - \beta) \), i.e., \( x_j < \frac{\delta_j}{1 - \beta} \) and \( x_j \) decreases, then \( x_j \) must be part of at least one approximately tight constraint. This lemma will be used later to show that in any round the increase in the potential due to decrease of the “small-value” variables is dominated by the decrease of “large-value” variables, that is, the variables that decrease by a multiplicative factor \( (1 - \beta) \). The proof of Lemma 4.3 is provided in Appendix D.2.

**Lemma 4.3.** If in some round there is a variable \( x_j < \frac{\delta_j}{1 - \beta} \) that decreases, then in the same round for some \( i \) with \( A_{ij} \neq 0 \) it holds that: 
\[
y_i \geq \sum_{i=1}^{m} A_{ij} y_i(x) \quad \text{and} \quad \sum_{k=1}^{n} A_{ik} x_k > 1 - \frac{\varepsilon}{\kappa}.
\]

**Potential.** We use the following potential function for analyzing the convergence time of the algorithm:
\[
\Phi(x) = \sum_{j=1}^{n} w_j f_\alpha(x_j) - \frac{1}{\kappa} \sum_{i=1}^{m} y_i(x),
\]
where \( f_\alpha \) is defined by (1). The potential function is strictly concave, and its partial derivative with respect to any variable \( x_j \) is:
\[
\frac{\partial \Phi(x)}{\partial x_j} = \frac{w_j}{x_j^\alpha} - \sum_{i=1}^{m} y_i(x) A_{ij} = \frac{w_j}{x_j^\alpha} \left( 1 - \frac{x_j^\alpha \sum_{i=1}^{m} y_i(x) A_{ij}}{w_j} \right). \tag{4}
\]

The following fact (given in a similar form in [5]), which follows directly from the Taylor series representation of concave functions, will be useful for the potential increase analysis:

**Fact 4.4.** For a differentiable concave function \( f : \mathbb{R}^n \to \mathbb{R} \) and any two points \( x^0, x^1 \in \mathbb{R}^n \):
\[
\sum_{j=1}^{n} \frac{\partial f(x^0)}{\partial x_j} (x_j^1 - x_j^0) \geq f(x^1) - f(x^0) \geq \sum_{j=1}^{n} \frac{\partial f(x^1)}{\partial x_j} (x_j^1 - x_j^0). \tag{3}
\]

\(^7\)Note that in [5] \( C = 1. \)
Using Fact 4.4 and 4.5, it is relatively simple to show the following result, proof of which is provided in Appendix D.3.

**Lemma 4.5.** Throughout the course of the algorithm, the potential function $\Phi(x)$ never decreases. Letting $\Phi^0$, $x^0$, $y(x^0)$ and $\Phi^1$, $x^1$, $y(x^1)$ denote the values of $\Phi$, $x$, and $y$ before and after a round, respectively, the potential function increase is lower-bounded as:

$$\Phi^1 - \Phi^0 \geq \sum_{j=1}^{n} w_j \left| \frac{x^1_j - x^0_j}{(x^0_j)^\alpha} \right| 1 - \frac{(x^1_j)\sum_{i=1}^{m} y_i(x^1)A_{ij}}{w_j}.$$  

### 4.1 Main Results

Our main results are summarized in the following three theorems. Notation $\tilde{O}(.)$ hides $\text{polylog}(\frac{1}{\epsilon})$ terms.

The objective function $\sum_j w_j f_\alpha(x_j)$ is denoted by $p_\alpha(x)$, $x^t$ denotes the solution at the beginning of round $t$, $x^*$ denotes the optimal solution to $(P_\alpha)$.

**Theorem 4.6.** *(Convergence for $\alpha < 1$.)* Algorithm 1 solves $(P_\alpha)$ approximately for $\alpha < 1$ in time that is polynomial in $\frac{\ln(nmA_{\max})}{\alpha(1-\alpha)\epsilon}$. In particular, after at most

$$\tilde{O} \left( \frac{1}{\alpha(1-\alpha)^3} \ln(nmA_{\max}) \ln(mA_{\max}) \ln^2 \left( \frac{mA_{\max}}{\epsilon} \right) \right) = \tilde{O} \left( \ln^4(nmA_{\max}) \right) \quad (5)$$

rounds there exists at least one round $t$ such that $p_\alpha(x^*) \leq (1 + 6\epsilon)p_\alpha(x^t)$. Moreover, the total number of rounds $t$ in which $p_\alpha(x^*) > (1 + 6\epsilon)p_\alpha(x^t)$ is also upper-bounded by $5$.

**Theorem 4.7.** *(Convergence for $\alpha = 1$.)* Algorithm 1 solves $(P_1)$ approximately in time that is linear in $n$ and polynomial in $\frac{\ln(nmA_{\max})}{\epsilon}$. In particular, after at most

$$\tilde{O} \left( \frac{W \ln(nmA_{\max}) \ln(mA_{\max})}{\epsilon^3} \right) = \tilde{O} \left( \frac{n \ln^2(nmA_{\max})}{\epsilon^3} \right) \quad (6)$$

rounds of the algorithm, the algorithm converges to a solution $x$ that provides an additive $7W\epsilon$-approximation to $(P_\alpha)$, that is $p_1(x^*) - p_1(x) \leq 7W\epsilon$.

**Theorem 4.8.** *(Convergence for $\alpha > 1$.)* Algorithm 1 solves $(P_\alpha)$ approximately for $\alpha > 1$ in time that is polynomial in $\frac{\ln(nmA_{\max})}{\epsilon}$. In particular, after at most:

$$\tilde{O} \left( \frac{\ln(nmA_{\max})(\ln(nmA_{\max}/\epsilon) + \alpha\epsilon)}{\epsilon^4} \right) = \tilde{O} \left( \frac{\ln^2(nmA_{\max})}{\epsilon^5} \right) \quad (7)$$

rounds there exists at least one round $t$ such that the vector of variables $x^t$ satisfies $p_\alpha(x^*) - p_\alpha(x^t) \leq \epsilon(8\alpha - 7)(-p_\alpha(x^t))$. Moreover, the total number of rounds in which $p_\alpha(x^*) - p_\alpha(x^t) > \epsilon(8\alpha - 7)(-p_\alpha(x^t))$ is upper-bounded by $7$.

Proofs of Theorem 4.6 and Theorem 4.7 are provided in Appendix D.4 and Appendix D.5 respectively.

### 4.2 Proof Sketch of Theorem 4.8

In this section, we outline the main ideas of the proof of Theorem 4.8, while the technical details are omitted and are instead provided in Appendix D.6. First, we show that in any round of the algorithm the variables that decrease by a multiplicative factor $(1 - \beta)$ dominate the potential increase due to all the variables that decrease (see Lemma D.7 in Appendix D.6). This result is then used in Lemma 4.9 to show the following lower bound on the potential increase:
Lemma 4.9. Let $\Phi^0, x^0, y(x^0)$ and $\Phi^1, x^1, y(x^1)$ denote the values of $\Phi$, $x$, and $y$ before and after any fixed round, respectively, and let $S^+ = \{ j : x_j^1 > x_j^0 \}$, $S^- = \{ j : x_j^1 < x_j^0 \}$. The potential increase in the round is lower-bounded as:

1. $\Phi^1 - \Phi^0 \geq \Omega(\beta \gamma) \sum_{j \in (S^+ \cup S^-)} x_j^0 \sum_{i=1}^m y_i(x^0) A_{ij}$;
2. $\Phi^1 - \Phi^0 \geq \Omega \left( \frac{\beta}{(1-\beta)^\alpha} \right) \left( \sum_{j=1}^n x_j^0 \sum_{i=1}^m y_i(x^0) - (1+\gamma) \sum_{j=1}^n w_j(x_j^0)^{1-\alpha} \right)$;
3. $\Phi^1 - \Phi^0 \geq \Omega \left( \frac{\beta}{(1+\beta)^\alpha} \right) \left( (1-\gamma) \sum_{j=1}^n w_j(x_j^0)^{1-\alpha} - \sum_{j=1}^n x_j^0 \sum_{i=1}^m y_i(x^0) \right)$.

Observe that for $\alpha > 1$ the objective function $p_\alpha(x)$, and consequently the potential function $\Phi$ are negative for any feasible $x$. To yield a poly-logarithmic convergence time in $n, m, A_{max}$, the idea is to show that the negative potential $-\Phi$ decreases by some multiplicative factor whenever $x$ is not a “good” approximation to $x^*$ – the optimal solution to $(P_\alpha)$. This idea, combined with the fact that the potential never decreases (and therefore $-\Phi$ never increases) and with upper and lower bounds on the potential then leads to the desired convergence time.

Consider the following definition of a stationary round:

Definition 4.10. (Stationary round.) A round is stationary, if both of the following conditions hold:

1. $\sum_{j \in (S^+ \cup S^-)} x_j \sum_{i=1}^m y_i(x) A_{ij} \leq \gamma \sum_{j=1}^n w_j x_j^{1-\alpha}$;
2. $(1-\gamma) \sum_{j=1}^n w_j x_j^{1-\alpha} \leq \sum_{j=1}^n x_j \sum_{i=1}^m y_i(x) A_{ij}$.

Otherwise, the round is non-stationary.

Recall the expression for the negative potential: $-\Phi = \frac{1}{\alpha-1} \sum_j w_j x_j^{1-\alpha} + \frac{1}{\kappa} \sum_i y_i(x)$. Then using Lemma 4.9 it suffices to show that in a non-stationary round the decrease in the negative potential $-\Phi$ is a multiplicative factor of the larger of the two terms $\frac{1}{\alpha-1} \sum_j w_j x_j^{1-\alpha}$ and $\frac{1}{\kappa} \sum_i y_i(x)$.

The last part of the proof is showing that the solution $x$ that corresponds to any stationary round is close to the optimal solution. This part is done by appropriately upper-bounding the duality gap. Denoting by $S^+ \cup S^-$ the set of coordinates $j$ for which $x_j$ either increases or decreases in the observed stationary round, using Definition D.2 we show that the terms $j \in \{ S^+ \cup S^- \}$ contribute to the duality gap by no more than $O(\varepsilon \alpha \cdot (-p_\alpha(x))$. The terms corresponding to $j \notin \{ S^+ \cup S^- \}$ are bounded recalling (from Algorithm 1) that for such terms $\frac{x_j^0 \sum_{i=1}^m y_i(x^0) A_{ij}}{w_j} \in (1-\gamma, 1+\gamma)$.

5 Conclusion

We presented an efficient stateless distributed algorithm for a class of $\alpha$-fair packing problems. To the best of our knowledge, this is the most efficient distributed algorithm for this problem. We obtained polylogarithmic convergence time in the input size for all the cases of positive fairness parameter $\alpha$ that are bounded away from 0 and 1. For the $\alpha = 1$ case, the convergence time is quasi-linear. We conjecture that the actual convergence time for the $\alpha = 1$ case should be polylogarithmic as well. However, it is an open problem to find an argument that leads to this result. Another research direction is to study if ideas developed in this paper can be used to address the open question from [19] regarding polynomial algorithms for market equilibria, where the considered optimization problem is very similar to the $\alpha$-fair packing problem for $\alpha = 1$. Finally, we believe that the techniques introduced in this paper can be used for addressing a broader class of distributed convex programming problems.
References

[1] Z. Allen-Zhu and L. Orecchia. Using optimization to break the epsilon barrier: A faster and simpler width-independent algorithm for solving positive linear programs in parallel. In Proc. ACM-SIAM SODA'15, 2015.

[2] A. B. Atkinson. On the measurement of inequality. Journal of Economic Theory, 2(3):244–263, 1970.

[3] B. Awerbuch and R. Khandekar. Greedy distributed optimization of multi-commodity flows. In Proc. ACM PODC’07, 2007.

[4] B. Awerbuch and R. Khandekar. Stateless near optimal flow control with poly-logarithmic convergence. In E. Laber, C. Bornstein, L. Nogueira, and L. Faria, editors, LATIN 2008: Theoretical Informatics, volume 4957 of Lecture Notes in Computer Science, pages 580–592. Springer Berlin Heidelberg, 2008.

[5] B. Awerbuch and R. Khandekar. Stateless distributed gradient descent for positive linear programs. SIAM Journal on Computing, 38(6):2468–2486, 2009.

[6] Y. Bartal, J. Byers, and D. Raz. Global optimization using local information with applications to flow control. In Proc. IEEE FOCS’97, 1997.

[7] A. Beck, A. Nedic, A. Ozdaglar, and M. Teboulle. An O(1/k) gradient method for network resource allocation problems. IEEE Transactions on Control of Network Systems, 1(1):64–73, 2014.

[8] D. Bertsekas and R. Gallager. Data Networks. Prentice Hall, 1992.

[9] D. Bertsimas, V. F. Farias, and N. Trichakis. On the efficiency-fairness trade-off. Management Science, 58(12):2234–2250, 2012.

[10] S. Boyd and L. Vandenberghe. Convex optimization. Cambridge university press, 2009.

[11] A. Charny, D. D. Clark, and R. Jain. Congestion control with explicit rate indication. In Proc. IEEE ICC’95, 1995.

[12] S. Dolev. Self-stabilization. MIT press, 2000.

[13] L. Fleischer. Approximating fractional multicommodity flow independent of the number of commodities. SIAM Journal on Discrete Mathematics, 13(4):505–520, 2000.

[14] N. Garg and J. Koenemann. Faster and simpler algorithms for multicommodity flow and other fractional packing problems. SIAM Journal on Computing, 37(2):630–652, 2007.

[15] J. Jaffe. Bottleneck flow control. IEEE Transactions on Communications, 29(7):954–962, 1981.

[16] K. Jain and V. Vazirani. Eisenberg-gale markets: Algorithms and structural properties. In Proc. ACM STOC’07, 2007.

[17] E. Kalai and M. Smorodinsky. Other solutions to Nash’s bargaining problem. Econometrica: Journal of the Econometric Society, 43(3):513–518, 1975.

[18] F. Kelly, A. Maulloo, and D. Tan. Rate control for communication networks: shadow prices, proportional fairness and stability. Journal of the Operational Research Society, 49(3):237–252, 1998.

[19] F. Kelly and V. Vazirani. Rate control as a market equilibrium. Unpublished manuscript, 2002.

[20] F. Kelly and E. Yudovina. Stochastic networks, volume 2. Cambridge University Press, 2014.

[21] J. Kleinberg, Y. Rabani, and É. Tardos. Fairness in routing and load balancing. In Proc. IEEE FOCS’99, 1999.

[22] C. Koufogiannakis and N. Young. Beating simplex for fractional packing and covering linear programs. In Proc. IEEE FOCS’07, 2007.

[23] F. Kuhn. The price of locality: Exploring the Complexity of Distributed Coordination Primitives. Diss., Eidgenössische Technische Hochschule ETH Zürich, Nr. 16213, 2005. PhD Thesis.

[24] A. Kumar and J. Kleinberg. Fairness measures for resource allocation. In Proc. IEEE FOCS’00, 2000.
[25] T. Lan, D. Kao, M. Chiang, and A. Sabharwal. An axiomatic theory of fairness in network resource allocation. In Proc. IEEE INFOCOM’10, 2010.

[26] S. Low, F. Paganini, and J. Doyle. Internet congestion control. IEEE Control Systems, 22(1):28–43, 2002.

[27] M. Luby and N. Nisan. A parallel approximation algorithm for positive linear programming. In Proc. ACM STOC’93, 1993.

[28] J. Marašević, C. Stein, and G. Zussman. Max-min fair rate allocation and routing in energy harvesting networks: Algorithmic analysis. In Proc. ACM MobiHoc’14, 2014.

[29] B. McCormick, F. Kelly, P. Plante, P. Gunning, and P. Ashwood-Smith. Real time alpha-fairness based traffic engineering. In Proc. ACM HotSDN’14, 2014.

[30] N. Megiddo. Optimal flows in networks with multiple sources and sinks. Mathematical Programming, 7(1):97–107, 1974.

[31] D. Mitroinovic and P. Vasic. Analytic inequalities, volume 1. Springer, 1970.

[32] J. Mo and J. Walrand. Fair end-to-end window-based congestion control. IEEE/ACM Transactions on Networking, 8(5):556–567, Oct. 2000.

[33] D. Mosk-Aoyama, T. Roughgarden, and D. Shah. Fully distributed algorithms for convex optimization problems. In Proc. DISC’07, 2007.

[34] J. F. Nash Jr. The bargaining problem. Econometrica: Journal of the Econometric Society, 18(2):155–162, 1950.

[35] F. Paganini, Z. Wang, J. Doyle, and S. Low. Congestion control for high performance, stability, and fairness in general networks. IEEE/ACM Transactions on Networking, 13(1):43–56, 2005.

[36] C. Papadimitriou and M. Yannakakis. Linear programming without the matrix. In Proc. ACM STOC’93, 1993.

[37] S. Plotkin, D. Shmoys, and É. Tardos. Fast approximation algorithms for fractional packing and covering problems. Mathematics of Operations Research, 20(2):257–301, 1995.

[38] Y. Yi and M. Chiang. Stochastic network utility maximisation - a tribute to Kelly’s paper published in this journal a decade ago. European Transactions on Telecommunications, 19(4):421–442, 2008.

[39] N. Young. Sequential and parallel algorithms for mixed packing and covering. In Proc. IEEE FOCS’01, 2001.
A Scaling Preserves Approximation

Let the $\alpha$–fair allocation problem be given in the form:

$$
(P_\alpha) \quad \text{max} \quad \sum_{j=1}^{n} w_j f_\alpha(x_j)
$$

where

$$f_\alpha(x_j) = \begin{cases} 
\ln(x_j), & \text{if } \alpha = 1 \\
\frac{x_j^{1-\alpha}}{1-\alpha}, & \text{if } \alpha \neq 1
\end{cases},$$

s.t. $Ax \leq b$

$x \geq 0$

$w$ is an $n$–length vector of positive weights, $x$ is the vector of variables, $A$ is an $n \times m$ constraint matrix, and $b$ is an $m$–length vector with positive entries.

It is not hard to see that the assumption $b_i = 1 \forall i$ is without loss of generality, since for $b_i \neq 1$ we can always divide both sides of the inequality by $b_i$ and obtain 1 on the right-hand side, since for (non-trivial) packing problems $b_i > 0$. Therefore, we can assume that the input problem has constraints of the form $A \cdot x \leq 1$, although it may not necessarily be the case that $A_{ij} \geq 1 \forall A_{ij} \neq 0$.

The remaining transformation that is performed on the input problem is:

$$
\hat{x}_j = c \cdot x_j, \quad \hat{A}_{ij} = A_{ij}/c.
$$

where

$$c = \begin{cases} 
\min_{i,j:A_{ij}\neq 0} A_{ij}, & \text{if } \min_{i,j:A_{ij}\neq 0} A_{ij} < 1 \\
1, & \text{otherwise}
\end{cases}.$$

The problem $(P_\alpha)$ after the scaling becomes:

$$
\hat{P}_\alpha \quad \text{max} \quad \sum_{j=1}^{n} w_j f_\alpha(\hat{x}_j) \cdot c^{1-\alpha}
$$

s.t. $\hat{A} \hat{x} \leq 1$

$x \geq 0,$

which is equivalent to:

$$
\hat{P}_\alpha \quad \text{max} \quad \sum_{j=1}^{n} w_j f_\alpha(\hat{x}_j)
$$

s.t. $\hat{A} \hat{x} \leq 1$

$x \geq 0,$

as $c^{1-\alpha}$ is a positive constant. Recall that Algorithm 1 returns an approximate solution to $(\hat{P}_\alpha)$.

Choose the dual variables (Lagrange multipliers) in the original solution as:

$$y_i = c^{\alpha-1} C \cdot e^{\kappa(\sum_{i=1}^{n} A_{ij} x_j - 1)} = c^{\alpha-1} C \cdot e^{\kappa(\sum_{i=1}^{n} A_{ij} \hat{x}_j - 1)} = c^{\alpha-1} \hat{y}_i, \quad (8)$$

and notice that

$$x_j^\alpha \sum_{i=1}^{m} y_i A_{ij} = \hat{x}_j^\alpha \cdot c^{-\alpha} \sum_{i=1}^{m} (c^{\alpha-1} \cdot \hat{y}_i \cdot c \cdot \hat{A}_{ij}) = \hat{x}_j^\alpha \sum_{i=1}^{m} \hat{y}_i \hat{A}_{ij}. \quad (9)$$

It is clear that $y_i$’s are feasible dual solutions, since the only requirement for the duals is non-negativity.
A.1 Approximation for Proportional Fairness

Recall (from (2)) that the duality gap for a given primal- and dual-feasible $x$ and $y$ is given as:

$$G(x, y) = \sum_{j=1}^{n} w_{j} \ln(w_{j}) - \sum_{j=1}^{n} w_{j} \ln \left( x_{j} \sum_{i=1}^{m} y_{i} A_{ij} \right) + \sum_{i=1}^{m} y_{i} - 1.$$ 

Since $\alpha = 1$, we have that $\hat{y}_{i} = y_{i}$ for all $i$, and using (3), it follows that

$$G(\hat{x}, \hat{y}) = G(x, y).$$

Since we demonstrate an additive approximation for the proportional fairness via the duality gap: $p(\hat{x}^{*}) - p(\hat{x}) \leq G(\hat{x}, \hat{y})$, the same additive approximation follows for the original (non-scaled) problem.

A.2 Approximation for $\alpha$-Fairness and $\alpha \neq 1$

For $\alpha \neq 1$, we show that the algorithm achieves a multiplicative approximation for the scaled problem. In particular, we show that after the algorithm converges we have that: $p_{\alpha}(\hat{x}^{*}) - p_{\alpha}(\hat{x}) \leq r_{\alpha} p_{\alpha}(\hat{x})$, where $\hat{x}^{*}$ is the optimal solution, $\hat{x}$ is the solution returned by the algorithm, and $r_{\alpha}$ is a constant.

Observe that since $\hat{x} = c \cdot x$, we have that $p_{\alpha}(\hat{x}^{*}) = c^{1-\alpha} p(x^{*})$ and $p_{\alpha}(\hat{x}) = c^{1-\alpha} p_{\alpha}(x)$. Therefore:

$$p_{\alpha}(x^{*}) - p_{\alpha}(x) = c^{\alpha-1} (p_{\alpha}(\hat{x}^{*}) - p_{\alpha}(\hat{x}))$$

$$\leq c^{\alpha-1} r_{\alpha} p_{\alpha}(\hat{x})$$

$$= r_{\alpha} p_{\alpha}(x).$$

B Solution Lower Bound

Proof of Lemma 2.3. Fix $\alpha$. Let:

$$\mu_{j}(\alpha) = \begin{cases} \left( \frac{w_{j}}{w_{\max} M} \min_{i: A_{ij} \neq 0} \frac{1}{n_{i} A_{ij}} \right)^{1/\alpha}, & \text{if } \alpha \leq 1 \\ A_{\max}(1-\alpha)/\alpha \left( \frac{w_{j}}{w_{\max} M} \right)^{1/\alpha} \min_{i: A_{ij} \neq 0} \frac{1}{n_{i} A_{ij}}, & \text{if } \alpha > 1 \end{cases}.$$ 

For the purpose of contradiction, suppose that $x^{*} = (x_{1}^{*}, ..., x_{n}^{*})$ is the optimal solution to $(P_{\alpha})$, and $x_{j}^{*} < \mu_{j}(\alpha)$ for some fixed $j \in \{1, ..., n\}$.

To establish the desired result, we will need to introduce additional notation. We first break the set of (the indices of) constraints of the form $Ax \leq 1$ in which variable $x_{j}$ appears with a non-zero coefficient into two sets, $U$ and $T$:

- Let $U$ denote the set of the constraints from $(P_{\alpha})$ that are not tight at the given optimal solution $x^{*}$, and are such that $A_{u,j} \neq 0$ for $u \in U$. Let $s_{u} = 1 - \sum_{k=1}^{n} A_{uk} x_{k}$ denote the slack of the constraint $u \in U$.

- Let $T$ denote the set of tight constraints from $(P_{\alpha})$ that are such that $A_{t,j} \neq 0$ for $t \in T$. Note that since $x^{*}$ is assumed to be optimal, it must be $T \neq \emptyset$.

Let $\varepsilon_{j} = \min \left\{ \mu_{j}(\alpha) - x_{j}^{*}, \min_{u \in U} s_{u}/A_{uj} \right\}$. Notice that by increasing $x_{j}$ to $x_{j}^{*} + \varepsilon_{j}$ none of the constraints from $U$ can be violated (although all the constraints in $T$ will; we deal with this in what follows).

14
In each constraint \( t \in T \), there must exist at least one variable \( x_k \) such that \( x_k^* > \frac{1}{n_t A_{tk}} \), because \( \sum_{l=1}^n A_{tl} x_l^* = 1 \), as each \( t \in T \) is tight, and \( x_j^* < \mu_j(\alpha) \leq \min_{i:A_{ij} \neq 0} \frac{1}{n_t A_{ij}} \leq \frac{1}{n_t A_{ij}} \). Select one such \( x_k \) per each constraint \( t \in T \), and denote by \( K \) the set of indices of selected variables. Observe that \( |K| \leq |T| (\leq M) \), since one \( x_k \) can appear in more than one constraint.

For each \( k \in K \), let \( T_k \) denote the constraints in which \( x_k \) is selected, and let

\[
\varepsilon_k = \max_{t \in T_k : A_{tk} \neq 0} \frac{A_{tj} \varepsilon_j}{A_{tk}},
\]

(10)

If we increase \( x_j \) by \( \varepsilon_j \) and decrease \( x_k \) by \( \varepsilon_k \) \( \forall k \in K \), each of the constraints \( t \in T \) will be satisfied since, from (10) and from the fact that only one \( x_k \) gets selected per constraint \( t \in T \), \( \varepsilon_j A_{tj} - \sum_{k \in K} \varepsilon_k A_{tk} \leq 0 \). Therefore, to construct an alternative feasible solution \( x' \), we set \( x_j' = x_j^* + \varepsilon_j, x_k' = x_k^* - \varepsilon_k \) for \( k \in K \), and \( x_l' = x_l^* \) for all the remaining coordinates \( l \in \{1, ..., n\} \setminus (K \cup \{j\}) \).

Since \( j \) is the only coordinate over which \( x \) gets increased in \( x' \), all the constraints \( A x' \leq 1 \) are satisfied. For \( x' \) to be feasible, we must have in addition that \( x_k' \geq 0 \) for \( k \in K \). This is shown by observing that for \( k \in K \):

\[
\varepsilon_k = \varepsilon_j \cdot \max_{t \in T_k : A_{tk} \neq 0} \frac{A_{tj}}{A_{tk}} \leq \mu_j(\alpha) \cdot \max_{t \in T_k : A_{tk} \neq 0} \frac{A_{tj}}{A_{tk}} \leq \min_{i:A_{ij} \neq 0} \frac{1}{n_t A_{ij}} \cdot \max_{t \in T_k : A_{tk} \neq 0} \frac{A_{tj}}{A_{tk}} \leq \max_{t \in T_k : A_{tk} \neq 0} \frac{1}{n_t A_{tk}} \cdot \max_{t \in T_k : A_{tk} \neq 0} \frac{A_{tj}}{A_{tk}} < x_k^*,
\]

where the second line follows from \( \varepsilon_j \leq \mu_j(\alpha) - x_j^* \leq \mu_j(\alpha) \), and the last line follows from the choice of \( x_k \).

The last part of the proof is to show that \( \sum_{l=1}^n w_l x_l' - \sum_{i=1}^n x_i^* = 0 \), which contradicts the initial assumption that \( x^* \) is optimal, by the definition of \( \alpha \)-fairness from Section 2.1. We have that:

\[
\sum_{l=1}^n w_l x_l' - \sum_{i=1}^n x_i^* \leq \sum_{k \in K} w_k \frac{\varepsilon_k}{x_k^*},
\]

(11)

Observe one fixed term from the summation (11). From the choice of \( \varepsilon_k \)'s, we know that for each \( \varepsilon_k \) there exist \( t \in T \) such that \( \varepsilon_k = \frac{\varepsilon_j A_{tj}}{A_{tk}} \), and at the same time (by the choice of \( x_k \)) we have \( x_k^* > \frac{1}{n_t A_{tk}} \), so that

\[
w_j \varepsilon_j x_j^* > w_j \frac{\varepsilon_j A_{tk}}{A_{tk}} = w_j \frac{\varepsilon_j A_{tj}}{A_{tk}} \left( \frac{1}{A_{tk} n_t} \right)^{\alpha} \geq \frac{w_k w_j \varepsilon_k A_{tk}}{w_{\max} A_{tj}} \left( \frac{1}{A_{tk} n_t} \right)^{\alpha}.
\]

(12)
Case 1. Suppose first that \( \alpha \leq 1 \). Then \( x^*_k > \left( \frac{1}{A_{tk} n_t} \right)^\alpha \geq \frac{1}{A_{tk} n_t} \), as \( A_{tk} \neq 0 \Rightarrow A_{tk} \geq 1 \). Plugging into (12) we have:

\[
\sum_j w_j x^*_k \alpha > w_k \frac{w_j}{w_{\text{max}}} \frac{1}{n_t A_{tj}},
\]

By the initial assumption, \( x^*_j < \mu_j(\alpha) = \left( \frac{w_j}{w_{\text{max}}} A_{\text{max}} \frac{1}{\min_i : A_{ij} \neq 0} \right)^{1/\alpha} \), and therefore

\[
w_k x^*_j = \frac{w_k x^*_j}{w_{\text{max}}} |K| > \frac{w_k w_j}{w_{\text{max}}} \frac{|K|}{M} \frac{1}{n_t A_{tj}} \leq \frac{w_k w_j}{w_{\text{max}}} \frac{1}{n_t A_{tj}},
\]

since it must be \( |K| \leq M (= \min\{m, n\}) \). From (13) and (14), we get that every term in the summation (11) is strictly positive, which implies:

\[
\sum_{l=1}^n w_l (x'_l - x^*_l) > 0,
\]

and therefore \( x^* \) is not optimal.

Case 2. Now suppose that \( \alpha > 1 \). Then

\[
x^*_j < \mu_j(\alpha) = A_{\text{max}}^{(1-\alpha)/\alpha} \left( \frac{w_j}{w_{\text{max}}} M \right)^{1/\alpha} \frac{1}{n_t A_{tj}} \leq A_{\text{max}}^{(1-\alpha)/\alpha} \left( \frac{w_j}{w_{\text{max}}} M \right)^{1/\alpha} \frac{1}{n_t A_{tj}}.
\]

Therefore:

\[
w_k x^*_j = \frac{w_k}{w_{\text{max}} M} A_{\text{max}}^{1-\alpha} \left( \frac{1}{n_t A_{tj}} \right)^\alpha |K|
\leq w_k \frac{w_j}{w_{\text{max}}} A_{\text{max}}^{1-\alpha} \frac{1}{n_t A_{tj}} \frac{A_{tk}^\alpha}{A_{tj}^\alpha}
= w_k \frac{w_j}{w_{\text{max}}} \frac{\varepsilon_k A_{tk}}{A_{tj}} \frac{(A_{tk}/A_{tj})^{\alpha-1}}{A_{\text{max}}^{\alpha-1}} \frac{1}{A_{tk} n_t} ^\alpha
\leq w_k \frac{w_j}{w_{\text{max}}} \frac{\varepsilon_k A_{tk}}{A_{tj}} \left( \frac{1}{A_{tk} n_t} \right) ^\alpha,
\]

as \( |K| \leq M \), and \( \frac{A_{tk}}{A_{tj}} \leq A_{\text{max}} \) (since for any \( i, j \): \( 1 \leq A_{ij} \leq A_{\text{max}} \)).

Finally, from (12) and (15) we get that every term in the summation (11) is positive, which yields a contradiction. \( \square \)

C Primal, Dual, and the Duality Gap

C.1 Proportionally Fair Resource Allocation

In this section we consider \((w, 1)\)-proportional resource allocation, often referred to as the weighted proportionally fair resource allocation. Recall that the primal is of the form:

\[
(P_1) \quad \max \sum_{j=1}^n w_j \ln(x_j)
\]

s.t. \( Ax \leq 1 \),
\( x \geq 0 \).
The Lagrangian for this problem can be written as:

\[ L_1(x; y, z) = \sum_{j=1}^{n} w_j \ln(x_j) + \sum_{i=1}^{m} y_i \left( \frac{1}{\chi} - \sum_{j=1}^{n} A_{ij} x_j - z_i \right) \]

where \( y_1, ..., y_m \) are Lagrange multipliers, and \( z_1, ..., z_m \) are slack variables. The dual to this problem is:

\[
(D_1) \quad \min \ g(y) \\
\text{s.t.} \quad y \geq 0,
\]

where \( g(y) = \max_{x, z \geq 0} L(x; y, z) \). To maximize \( L_1(x; y, z) \), we first differentiate with respect to \( x_j, j \in \{1, ..., n\} \):

\[
\frac{\partial L_1(x; y, z)}{\partial x_j} = \frac{w_j}{x_j} - \sum_{i=1}^{m} y_i A_{ij} = 0,
\]

which gives:

\[
x_j \cdot \sum_{i=1}^{m} y_i A_{ij} = w_j, \quad \forall j \in \{1, ..., n\}. \tag{16}
\]

Plugging this back into the expression for \( L_1(x; y, z) \), and noticing that, since \( y_i, z_i \geq 0 \ \forall i \in \{1, ..., m\} \), \( L_1(x; y, z) \) is maximized for \( z_i = 0 \), we get that:

\[
g_1(y) = \sum_{j=1}^{n} w_j \ln \left( \frac{w_j}{\sum_{i=1}^{m} y_i A_{ij}} \right) + \sum_{i=1}^{m} y_i - \sum_{i=1}^{m} y_i \sum_{j=1}^{n} A_{ij} \frac{w_j}{\sum_{k=1}^{m} y_k A_{kj}}
\]

\[
= \sum_{j=1}^{n} w_j \ln(w_j) - \sum_{j=1}^{n} w_j \ln \left( \sum_{i=1}^{m} y_i A_{ij} \right) + \sum_{i=1}^{m} y_i - \sum_{j=1}^{n} w_j \sum_{i=1}^{m} y_i \frac{A_{ij}}{\sum_{k=1}^{m} y_k A_{kj}}
\]

\[
= \sum_{j=1}^{n} w_j \ln(w_j) - \sum_{j=1}^{n} w_j \ln \left( \sum_{i=1}^{m} y_i A_{ij} \right) + \sum_{i=1}^{m} y_i - W,
\]

since \( \sum_{i=1}^{m} \frac{y_i A_{ij}}{\sum_{k=1}^{m} y_k A_{kj}} = 1 \ \forall j \in \{1, ..., n\} \), and \( \sum_{j=1}^{n} w_j = W \).

Let \( p_1(x) = \sum_{j=1}^{n} w_j \ln(x_j) \) denote the primal objective. The duality gap for any pair of primal-feasible \( x \) and dual-feasible (nonnegative) \( y \) is given by:

\[
G_1(x, y) = g_1(y) - p_1(x)
\]

\[
= \sum_{j=1}^{n} w_j \ln(w_j) - \sum_{j=1}^{n} w_j \ln \left( x_j \sum_{i=1}^{m} y_i A_{ij} \right) + \sum_{i=1}^{m} y_i - W.
\]

Since the primal problem maximizes a concave function over a polytope, the strong duality holds \[10\], and therefore \( G_1(x, y) \geq 0 \) for any pair of primal-feasible \( x \) and dual-feasible \( y \), with equality if and only if \( x \) and \( y \) are primal- and dual- optimal, respectively.

### C.2 \( \alpha \)-Fair Resource Allocation for \( \alpha \neq 1 \)

Recall that for \( \alpha \neq 1 \) the primal problem is:

\[
(P_\alpha) \quad \max \ \sum_{j=1}^{n} w_j x_j^{1-\alpha} = p_\alpha(x)
\]

\[
\text{s.t.} \quad Ax \leq 1,
\]

\[
x \geq 0.
\]
The Lagrangian for this problem can be written as:

\[ L_\alpha(x; y, z) = \sum_{j=1}^{n} w_j \frac{x_j^{1-\alpha}}{1-\alpha} + \sum_{i=1}^{m} y_i \left( 1 - \sum_{j=1}^{n} A_{ij} x_j - z_i \right) , \]

where \( y_i \) and \( z_i \), for \( i \in \{1, \ldots, m\} \), are Lagrangian multipliers and slack variables, respectively.

The dual to \((P_\alpha)\) can be written as:

\[
(D_\alpha) \quad \min \quad g(y) \\
\text{s.t.} \quad y \geq 0,
\]

where \( g_\alpha(y) = \max_{x,z \geq 0} L_\alpha(x; y, z) \).

Since \( L_\alpha(x; y, z) \) is differentiable with respect to \( x_j \) for \( j \in \{1, \ldots, n\} \), it is maximized for:

\[
\frac{\partial L_\alpha(x; y, z)}{\partial x_j} = \frac{w_j}{x_j^{\alpha}} - \sum_{i=1}^{m} y_i A_{ij} = 0 \\
\Rightarrow w_j = x_j^{\alpha} \sum_{i=1}^{m} y_i A_{ij}.
\] (17)

As \( z_i \cdot y_i \geq 0 \ \forall i \in \{1, \ldots, m\} \), we get that:

\[
g_\alpha(y) = \sum_{j=1}^{n} \frac{w_j}{1-\alpha} \left( \frac{w_j}{\sum_{i=1}^{m} y_i A_{ij}} \right)^{1-\alpha} + \sum_{i=1}^{m} y_i - \sum_{i=1}^{m} y_i \sum_{j=1}^{n} A_{ij} \left( \frac{w_j}{\sum_{k=1}^{m} y_k A_{kj}} \right)^{1/\alpha} \\
= \sum_{j=1}^{n} \frac{w_j}{1-\alpha} \left( \frac{w_j}{\sum_{i=1}^{m} y_i A_{ij}} \right)^{1-\alpha} + \sum_{i=1}^{m} y_i - \sum_{j=1}^{m} w_j^{1/\alpha} \left( \sum_{k=1}^{m} y_k A_{kj} \right)^{-1/\alpha} \sum_{i=1}^{m} A_{ij} y_i \\
= \sum_{j=1}^{n} \frac{w_j}{1-\alpha} \left( \frac{w_j}{x_j^{\alpha} \sum_{i=1}^{m} y_i A_{ij}} \right)^{1-\alpha} - 1 \right) + \sum_{i=1}^{m} y_i - \sum_{j=1}^{m} w_j^{1/\alpha} \left( \sum_{i=1}^{m} A_{ij} y_i \right)^{\alpha-1/\alpha}.
\]

Similarly as before, for primal-feasible \( x \) and dual-feasible \( y \), the duality gap is given as:

\[
G_\alpha(x, y) = g_\alpha(y) - p_\alpha(x) \\
= \sum_{j=1}^{n} \frac{w_j}{1-\alpha} \left( \frac{w_j}{\sum_{i=1}^{m} y_i A_{ij}} \right)^{1-\alpha} + \sum_{i=1}^{m} y_i - \sum_{j=1}^{m} w_j^{1/\alpha} \left( \sum_{k=1}^{m} y_k A_{kj} \right)^{-1/\alpha} - \sum_{i=1}^{m} w_j^{1-\alpha} \\
= \sum_{j=1}^{n} \frac{w_j}{1-\alpha} \left( \frac{w_j}{x_j^{\alpha} \sum_{i=1}^{m} y_i A_{ij}} \right)^{1-\alpha} X_j^{1-\alpha} (\frac{\sum_{i=1}^{m} A_{ij} y_i}{w_j})^{\alpha-1/\alpha}.
\]

Observing that:

\[
w_j^{1/\alpha} \left( \frac{\sum_{i=1}^{m} A_{ij} y_i}{w_j} \right)^{\alpha-1/\alpha} = w_j \cdot w_j^{-\alpha-1/\alpha} \cdot x_j^{1-\alpha} \cdot x_j^{\alpha-1/\alpha} \cdot \left( \frac{\sum_{i=1}^{m} A_{ij} y_i}{w_j} \right)^{\alpha-1/\alpha} \\
= w_j x_j^{1-\alpha} \cdot \left( \frac{x_j^{\alpha} \sum_{i=1}^{m} A_{ij} y_i}{w_j} \right)^{\alpha-1/\alpha},
\]

we finally get:

\[
G_\alpha(x, y) = \sum_{j=1}^{n} \frac{w_j}{1-\alpha} \left( \frac{w_j}{x_j^{\alpha} \sum_{i=1}^{m} y_i A_{ij}} \right)^{1-\alpha} - 1 \right) + \sum_{i=1}^{m} y_i - \sum_{j=1}^{m} w_j^{1-\alpha} \cdot \left( \frac{x_j^{\alpha} \sum_{i=1}^{m} A_{ij} y_i}{w_j} \right)^{\alpha-1/\alpha}.
\]

18
D Omitted Proofs from Section 4

D.1 Feasibility and Approximate Complementary Slackness

Proof of Lemma 4.1. The claim holds initially, since for any constraint $i$ at the beginning of the first round we have $\sum_{\ell=1}^{n} A_{i\ell}x_{\ell}^{0} \leq A_{\text{max}} \sum_{\ell=1}^{n} \delta_{\ell} \leq (\frac{1}{2})^{1/\alpha} < 1$.

Now assume that $x$ becomes infeasible in some round, and let $x^{0}$ denote the (feasible) solution before that round, $x^{1}$ denote the (infeasible) solution after the round. We have:

$$\sum_{\ell=1}^{n} A_{i\ell}x_{\ell}^{0} \leq 1, \quad \forall i \in \{1, \ldots, m\}, \quad \text{and}$$

$$\sum_{\ell=1}^{n} A_{k\ell}x_{\ell}^{1} > 1, \quad \text{for some } k \in \{1, \ldots, m\}.$$

For this to be true, $x$ must have increased over at least one coordinate $j$ such that $A_{kj} \neq 0$. For such change to be triggered by the algorithm, it must also be true that:

$$(x_{j}^{0})^{\alpha} \sum_{i=1}^{m} y_{i}(x^{0})A_{ij} \leq w_{j}(1 - \gamma).$$

Since, by the choice of $\beta$, this term can increase by a factor of at most $1 + \gamma/4$, it follows that:

$$(x_{j}^{1})^{\alpha} \sum_{i=1}^{m} y_{i}(x^{1})A_{ij} \leq w_{j}(1 - \gamma) \left(1 + \frac{\gamma}{4}\right) < w_{j}.$$

This further implies:

$$(x_{j}^{1})^{\alpha} y_{k}(x^{1})A_{kj} < w_{j},$$

and since whenever $A_{kj} \neq 0$ we also have $A_{kj} \geq 1$, we get:

$$(x_{j}^{1})^{\alpha} y_{k}(x^{1}) < w_{j}.$$

On the other hand, since $x_{j}^{1} \geq \delta_{j}$, $\delta_{j}^{\alpha} = \frac{w_{j}}{C}$, and $\sum_{j=1}^{n} A_{kj}x_{j}^{1} > 1$:

$$(x_{j}^{1})^{\alpha} y_{k}(x^{1}) \geq \frac{w_{j}}{C} \cdot C \cdot e^{\kappa(\sum_{j=1}^{n} A_{kj}x_{j}^{1} - 1)} > w_{j},$$

which is a contradiction. \(\square\)

Proof of Lemma 4.2. Suppose that $\max_{i} \sum_{j=1}^{n} A_{ij}x_{j} < 1 - \varepsilon$. Then for each $y_{i}$ we have:

$$y_{i} \leq C \cdot e^{-\kappa \varepsilon} = C \cdot \frac{\varepsilon \min_{j} w_{j}}{CmA_{\text{max}}} = \frac{\varepsilon \min_{j} w_{j}}{mA_{\text{max}}}.$$

Due to Lemma 4.1, we have that $x$ is feasible in every round, which implies that $x_{j} \leq 1 \forall j$. This further gives:

$$x_{j}^{\alpha} \sum_{i=1}^{m} y_{i}A_{ij} \leq w_{j}\varepsilon \leq w_{j}(1 - \gamma),$$

and, therefore, all variables $x_{j}$ increase by a factor $1 + \beta$. From Lemma 4.1 since the solution always remains feasible, none of the variables can increase to a value larger than 1. Therefore, after at most $\tau_{0} = \min_{j} \log_{1+\beta} \left(\frac{1}{\delta_{j}}\right) \leq \min_{j} \frac{1}{\beta} \ln \left(\frac{1}{\delta_{j}}\right)$ rounds there must exist at least one $i$ such that $\sum_{j=1}^{n} A_{ij}x_{j} \geq$
1 - \varepsilon. If in any round \( \max_i \sum_{j=1}^n A_{ij}x_j \) decreases, it can decrease by at most \( \beta \sum_{j=1}^n A_{ij}x_j \leq \beta < \frac{\varepsilon}{\kappa} \). Therefore, in every subsequent round

\[
\max_i \sum_{j=1}^n A_{ij}x_j > (1 - (1 + 1/\kappa)\varepsilon).
\]

For the second part of the lemma, let \( S = \{ i : \sum_{j=1}^n A_{ij}x_j < \max_{k \in \{1, \ldots, m\}} A_{kj}x_j - \frac{\varepsilon}{\kappa} \} \) be the set of constraints that are at least \( \frac{\kappa-1}{\kappa} \varepsilon \)-looser than the tightest constraint. Then for \( i \in S \) we have

\[
y_i \leq e^{-(\kappa-1)\varepsilon} \max_{k \in \{1, \ldots, m\}} y_k \leq \frac{\varepsilon}{m} e^{\varepsilon} \max_{k \in \{1, \ldots, m\}} y_k \leq 2 \frac{\varepsilon}{m} \max_{k \in \{1, \ldots, m\}} y_k.
\]

This further gives:

\[
\sum_{i=1}^m y_i = \sum_{i \in S} y_i + \sum_{k \notin S} y_k < (1 + 2\varepsilon) \sum_{i \in S} y_i.
\]

Moreover, for each \( i \notin S \) we have (since \( \max_{k \in \{1, \ldots, m\}} A_{kj}x_j \geq 1 - (1+1/\kappa)\varepsilon \)):

\[
y_i \sum_{j=1}^n A_{ij}x_j \geq (1 - 2\varepsilon) y_i,
\]

and, therefore:

\[
\sum_{i=1}^m y_i < \frac{1 + 2\varepsilon}{1 - 2\varepsilon} \sum_{i \notin S} y_i \sum_{j=1}^n A_{ij}x_j
\]

\[
\leq (1 + 2\varepsilon)(1 + 3\varepsilon) \sum_{i \notin S} y_i \sum_{j=1}^n A_{ij}x_j
\]

\[
= (1 + 5\varepsilon + 6\varepsilon^2) \sum_{i \notin S} y_i \sum_{j=1}^n A_{ij}x_j
\]

\[
\leq (1 + 6\varepsilon) \sum_{i \notin S} y_i \sum_{j=1}^n A_{ij}x_j \quad \text{(from } \varepsilon \leq 1/6)\]

\[
\leq (1 + 6\varepsilon) \sum_{i=1}^m y_i \sum_{j=1}^n A_{ij}x_j.
\]

Interchanging the order of summation in the last line, we reach the desired inequality.

The proof of the last part is equivalent to the proof of the second part of Lemma 3.2 in [5] and is omitted.

**D.2 Decrease of Small Variables**

**Proof of Lemma 4.3.** Suppose that some \( x_j < \frac{\delta_j}{1-\beta} \) triggers a decrease over the \( j^{th} \) coordinate. The first part of the Lemma is easy to show, simply by using the argument that at least one term of a summation must be higher than the average:

\[
y_i A_{ij} \geq \frac{\sum_{l=1}^m A_{lj}y_l(x)}{m} \quad \Rightarrow \quad y_i \geq \frac{\sum_{l=1}^m A_{lj}y_l(x)}{mA_{\text{max}}}.
\]

For the second part, as \( x_j < \frac{\delta_j}{1-\beta} \), we have that:

\[
x_j^\alpha y_i \geq \frac{x_j^\alpha \sum_{l=1}^m A_{lj}y_l(x)}{mA_{\text{max}}} \quad \Rightarrow \quad y_i > \frac{(1 - \beta)^\alpha}{\delta_j^\alpha} \frac{x_j^\alpha \sum_{l=1}^m A_{lj}y_l(x)}{mA_{\text{max}}}.
\]
Since \( x_j \) decreases, we have that \( x_j^\alpha \sum_{i=1}^m y_i(x)A_{ij} \geq w_j(1+\gamma) \), and therefore \( y_i(x) > \frac{w_j(1+\gamma)(1-\beta)^\alpha}{A_{\max}m} \).

Moreover, as \( y_i(x) = C \cdot e^{\kappa(\sum_{i=1}^n A_{ik}x_k^{-1})} \), and \( C = \frac{w_j}{\delta_j} \), it follows that:

\[
e^{\kappa(\sum_{i=1}^n A_{ik}x_k^{-1})} > \frac{(1 + \gamma)(1 - \beta)^\alpha}{A_{\max}m}.
\]

(18)

Observe that for \( \alpha \leq 1 \):

\[
(1 + \gamma)(1 - \beta)^\alpha \geq (1 + \gamma)(1 - \beta) = \left(1 + \frac{\epsilon}{4}\right) \left(1 - \frac{\epsilon}{16(\kappa + 1)}\right) > 1 + \frac{\epsilon}{8} > \sqrt{\epsilon},
\]

(19)

while for \( \alpha > 1 \), since \( \epsilon \alpha \leq \frac{9}{10} \):

\[
(1 + \gamma)(1 - \beta)^\alpha \geq (1 + \gamma)(1 - \alpha \beta) \geq (1 + \gamma) \left(1 - \frac{\gamma \epsilon \alpha}{4}\right) \geq 1 > \sqrt{\epsilon},
\]

(20)

where we have used the generalized Bernoulli’s inequality for \( (1 - \beta)^\alpha \geq (1 - \alpha \beta) \), and then \( \beta = \frac{\gamma}{4(\kappa + \alpha)} \leq \frac{\sqrt{\epsilon}}{4} \). Recalling that \( \kappa = \frac{1}{\epsilon} \ln \left(\frac{CmA_{\max}}{\epsilon \min_j w_j}\right) \), and combining (18) with (19) and (20):

\[
\left(\frac{\epsilon \min_j w_j}{CmA_{\max}}\right)^{1 - \sum_{i=1}^n A_{ik}x_k^0} > \frac{\sqrt{\epsilon}}{mA_{\max}}.
\]

Finally, as \( C \geq 2w_{\max}nmA_{\max} \), it follows that \( \min_j w_j \leq \frac{\epsilon \min_j w_j}{2w_{\max}nmA_{\max}} < \left(\frac{\sqrt{\epsilon}}{mA_{\max}}\right)^2 < 1 \), which gives:

\[
1 - \frac{1}{\epsilon} \sum_{k=1}^n A_{ik}x_k^0 < 1 - \frac{\epsilon}{2} \iff \sum_{k=1}^n A_{ik}x_k^0 > 1 - \frac{\epsilon}{2}.
\]

\(\Box\)

D.3 Potential

Proof of Lemma 4.5. Since \( \Phi \) is concave, using Fact 4.4 and (1) it follows that:

\[
\Phi^1 - \Phi^0 \geq \sum_{j=1}^n w_j \frac{x_j^1 - x_j^0}{(x_j^1)^\alpha} \left(1 - \frac{(x_j^1)^\alpha \sum_{i=1}^m y_i(x^1)A_{ij}}{w_j}\right).
\]

(21)

If \( x_j^1 = x_j^0 \), then the term in the summation [21] corresponding to the change in \( x_j \) is equal to zero, and \( x_j \) has no contribution to the sum in [21].

If \( x_j^1 - x_j^0 > 0 \), then, as \( x_j \) increases over the observed round, it must be \( \frac{(x_j^1)^\alpha \sum_{i=1}^m y_i(x^1)A_{ij}}{w_j} \leq 1 - \gamma \).

By the choice of the parameters, \( \frac{(x_j^1)^\alpha \sum_{i=1}^m y_i(x^1)A_{ij}}{w_j} \leq \left(1 + \frac{\gamma}{4}\right) \left(1 - \frac{\gamma}{2}\right) = 1 - \frac{3}{4}\gamma \leq 1 - \frac{3\gamma}{4} \), and therefore

\[
\frac{(x_j^1)^\alpha \sum_{i=1}^m y_i(x^1)A_{ij}}{w_j} \leq \left(1 + \frac{\gamma}{4}\right) (1 - \gamma) = 1 - \frac{3\gamma}{4} - \frac{\gamma^2}{4} < 1 - \frac{3\gamma}{4}.
\]

(22)

It follows that \( 1 - \frac{(x_j^1)^\alpha \sum_{i=1}^m y_i(x^1)A_{ij}}{w_j} > \frac{3\gamma}{4} > 0 \), and therefore

\[
w_j \frac{x_j^1 - x_j^0}{(x_j^1)^\alpha} \left(1 - \frac{(x_j^1)^\alpha \sum_{i=1}^m y_i(x^1)A_{ij}}{w_j}\right) = w_j \frac{x_j^1 - x_j^0}{(x_j^1)^\alpha} \left|1 - \frac{(x_j^1)^\alpha \sum_{i=1}^m y_i(x^1)A_{ij}}{w_j}\right|.
\]
Finally, if \( x_j^1 - x_j^0 < 0 \), then it must be 
\[
\frac{(x_j^0)^\alpha \sum_{i=1}^m y_i(x^0) A_{ij}}{w_j} \geq \left( 1 - \frac{\gamma}{4} \right) \left( \frac{(x_j^0)^\alpha \sum_{i=1}^m y_i(x^0) A_{ij}}{w_j} \right),
\]
implying
\[
\frac{(x_j^1)^\alpha \sum_{i=1}^m y_i(x^1) A_{ij}}{w_j} \geq \left( 1 - \frac{\gamma}{4} \right) (1 + \gamma) = 1 + \frac{3}{4} \gamma - \frac{\gamma^2}{4} > 1 + \frac{1}{2} \gamma.
\]
(23)

We get that
\[
1 - \frac{(x_j^1)^\alpha \sum_{i=1}^m y_i(x^1) A_{ij}}{w_j} < -\frac{1}{2} \gamma < 0,
\]
and therefore
\[
w_j \frac{x_j^1 - x_j^0}{(x_j^1)^\alpha} \left( 1 - \frac{(x_j^1)^\alpha \sum_{i=1}^m y_i(x^1) A_{ij}}{w_j} \right) = w_j \frac{\left| x_j^1 - x_j^0 \right|}{(x_j^1)^\alpha} \left| 1 - \frac{(x_j^1)^\alpha \sum_{i=1}^m y_i(x^1) A_{ij}}{w_j} \right|,
\]
completing the proof.

\[ \square \]

D.4 Convergence Proof for \( \alpha < 1 \)

The following lemma appears in a similar form in [5] and is essential for obtaining multiplicative updates of the potential function. We remark that while the proof of the lemma uses similar arguments as [5], the third part requires an additional result for bounding the potential change due to decrease of small \( x \)'s, that is, \( x \)'s that are smaller than \( \frac{\delta_j}{1 - \beta} \).

**Lemma D.1.** If \( \alpha < 1 \) and \( \Phi^0, x^0, y(x^0) \) and \( \Phi^1, x^1, y(x^1) \) denote the values of \( \Phi, x, \) and \( y \) before and after a round, respectively, then:

1. \( \Phi^1 - \Phi^0 \geq \Omega \left( \frac{n}{\alpha} \right) \cdot \sum_{i=1}^m \left| y_i(x^1) - y_i(x^0) \right| \);
2. \( \Phi^1 - \Phi^0 \geq \Omega(\beta) \left( (1 - \gamma) \sum_{j=1}^n w_j (x_j^0)^{1-\alpha} - \sum_{i=1}^m y_i(x^0) \sum_{j=1}^n A_{ij} x_j^0 \right) \);
3. \( \Phi^1 - \Phi^0 \geq \Omega(\beta) \left( \sum_{i=1}^m y_i(x^0) \sum_{j=1}^n A_{ij} x_j^0 - (1 + \gamma) \sum_{j=1}^n w_j (x_j^0)^{1-\alpha} \right) \).

**Proof.**

**Proof of 1.** Let \( z_j = \frac{(x_j^1)^\alpha \sum_{i=1}^m y_i(x^1) A_{ij}}{w_j} \). From the proof of Lemma 4.5 if \( x_j^1 - x_j^0 > 0 \), then \( 1 - z_j \geq \frac{3}{4} \gamma \geq \frac{3}{4} \gamma j, \) as \( 0 < z_j \leq 1 - \frac{3}{4} \gamma j. \) If \( x_j^1 - x_j^0 < 0, \) then \( 1 - z_j \leq \frac{\gamma}{2}, \) which implies \( 1 \leq z_j (1 + \frac{\gamma}{2})^{-1}, \) and thus \( 1 - z_j \leq z_j ((1 + \gamma/2)^{-1} - 1) = z_j - z_j (1 + \frac{\gamma}{2})^{-1} < -z_j \frac{\gamma}{3} = -z_j. \) It follows immediately from Lemma 4.5 that:

\[
\Phi^1 - \Phi^0 \geq \frac{2}{3} \sum_{j=1}^n w_j \frac{|x_j^1 - x_j^0|}{(x_j^1)^\alpha} \cdot z_j = \frac{2}{3} \sum_{j=1}^n w_j \frac{|x_j^1 - x_j^0|}{(x_j^1)^\alpha} \cdot \frac{(x_j^1)^\alpha \sum_{i=1}^m y_i(x^1) A_{ij}}{w_j}
\]
\[
= \frac{2}{3} \sum_{j=1}^n |x_j^1 - x_j^0| \sum_{i=1}^m y_i(x^1) A_{ij} = \frac{2}{3} \sum_{i=1}^m y_i(x^1) \sum_{j=1}^n A_{ij} |x_j^1 - x_j^0|.
\]
(24)

Since each \( y_i \) is convex, \( -y_i \) is concave, and applying Fact 4.4 it follows that \( \forall i : \)
\[
y_i(x^0) k \sum_{j=1}^n A_{ij} (x_j^1 - x_j^0) \leq y_i(x^1) - y_i(x^0) \leq y_i(x^1) k \sum_{j=1}^n A_{ij} (x_j^1 - x_j^0).
\]
(25)

From the choice of parameters, since every \( x_j \) can change by at most a factor \( (1 + \beta) \) or \( (1 - \beta) \) over a round, \( e^{-\kappa \beta} y_i(x^0) \leq y_i(x^1) \leq e^{\kappa \beta} y_i(x^0), \) and therefore, recalling that \( \beta = \frac{\gamma}{4(\kappa + 1)} \), \( \frac{y_i(x^1)}{y_i(x^0)} = \Theta(1). \)
Therefore, (25) implies that:

\[ |y_i(x^1) - y_i(x^0)| = \Theta(\kappa) y_i(x^1) \sum_{j=1}^{n} A_{ij}(x_j^1 - x_j^0) = O(\kappa) y_i(x^1) \sum_{j=1}^{n} A_{ij} |x_j^1 - x_j^0|. \]  

(26)

Combining (24) and (26):

\[ \Phi^1 - \Phi^0 \geq \Omega \left( \frac{\gamma}{\kappa} \right) \cdot \sum_{i=1}^{m} |y_i(x^1) - y_i(x^0)|. \]

Proof of 2. Let \( S^+ \) denote the set of \( j \)'s such that \( x_j \) increases in the current round. Then, recalling that for \( j \in S^+ \) \( \frac{(x_j^0)^{\alpha} \sum_{i=1}^{m} y_i(x^0) A_{ij}}{w_j} \leq 1 - \gamma \) and that from the choice of parameters \( \frac{(x_j^1)^{\alpha} \sum_{i=1}^{m} y_i(x^1) A_{ij}}{w_j} \leq (1 + \gamma/4) \frac{(x_j^0)^{\alpha} \sum_{i=1}^{m} y_i(x^0) A_{ij}}{w_j} \):

\[
\Phi^1 - \Phi^0 \geq \sum_{j=1}^{n} w_j x_j^1 - x_j^0 \left( 1 - \frac{(x_j^1)^{\alpha} \sum_{i=1}^{m} y_i(x^1) A_{ij}}{w_j} \right) \\
\geq \sum_{j \in S^+} w_j x_j^1 - x_j^0 \left( 1 - \frac{(x_j^1)^{\alpha} \sum_{i=1}^{m} y_i(x^1) A_{ij}}{w_j} \right) \\
\geq \sum_{j \in S^+} w_j x_j^1 - x_j^0 \left( 1 - (1 + \gamma/4) \frac{(x_j^0)^{\alpha} \sum_{i=1}^{m} y_i(x^0) A_{ij}}{w_j} \right) \\
\geq \sum_{j \in S^+} w_j x_j^1 - x_j^0 \left( 1 - (1 - \gamma) - \frac{(x_j^0)^{\alpha} \sum_{i=1}^{m} y_i(x^0) A_{ij}}{w_j} \right). \]  

(27)

Since \( j \in S^+ \), \( x_j^1 = (1+\beta)x_j^0 \), it follows that \( \Phi^1 - \Phi^0 \geq \frac{\beta}{(1+\beta)^{\alpha}} \sum_{j \in S^+} w_j (x_j^0)^{-\alpha} (1 - \gamma) - \frac{(x_j^0)^{\alpha} \sum_{i=1}^{m} y_i(x^0) A_{ij}}{w_j} \).

Observing that for any \( x_j \notin S^+ \) we have that \( (1 - \gamma) - \frac{(x_j^0)^{\alpha} \sum_{i=1}^{m} y_i(x^0) A_{ij}}{w_j} < 0 \), we get:

\[
\Phi^1 - \Phi^0 \geq \frac{\beta}{(1+\beta)^{\alpha}} \sum_{j=1}^{n} w_j (x_j^0)^{1-\alpha} \left( (1 - \gamma) - \frac{(x_j^0)^{\alpha} \sum_{i=1}^{m} y_i(x^0) A_{ij}}{w_j} \right) \\
= \Omega(\beta) \left( (1 - \gamma) \sum_{j=1}^{n} w_j (x_j^0)^{1-\alpha} - \sum_{j=1}^{n} x_j^0 \sum_{i=1}^{m} y_i(x^0) A_{ij} \right). \]

Proof of 3. Let \( S^- \) denote the set of \( j \)'s such that \( x_j \) decreases in the current round. Note that in this case not all the \( x_j \)'s with \( j \in S^- \) decrease by a multiplicative factor \( (1 - \beta) \), since for \( j \in S^- \): \( x_j^1 = \max\{(1 - \beta)x_j^0, \delta_j\} \). We will first lower-bound the potential increase over \( x_j \)'s that decrease multiplicatively: \( \{j : j \in S^- \land x_j^0(1 - \beta) \geq \delta_j\} \), so that \( x_j^1 = x_j^0(1 - \beta) \). Recall that for \( j \in S^- \)
\[
\frac{(x_j^0)^\alpha \sum_{i=1}^{m} y_i(x^0) A_{ij}}{w_j} \geq 1 + \gamma \text{ and } \frac{(x_j^0)^\alpha \sum_{i=1}^{m} y_i(x^1) A_{ij}}{w_j} \geq (1 - \gamma/4) \frac{(x_j^0)^\alpha \sum_{i=1}^{m} y_i(x^0) A_{ij}}{w_j}. \]
It follows that:

\[
\Phi^1 - \Phi^0 \geq \frac{\beta}{(1 - \beta)\alpha} \sum_{\{i,j \in S^- \land x_j^0(1 - \beta) \geq \delta_j\}} \sum_{i = 1}^{m} y_i(x^0) A_{ij} = \left(1 - \gamma/4\right) \frac{(x_j^0)^\alpha \sum_{i=1}^{m} y_i(x^0) A_{ij}}{w_j} - 1 \right)
\]

\[
\geq \frac{\beta}{(1 - \beta)\alpha} \sum_{\{i,j \in S^- \land x_j^0(1 - \beta) \geq \delta_j\}} \sum_{i = 1}^{m} y_i(x^0) A_{ij} = \left(1 - \gamma/4\right) \frac{(x_j^0)^\alpha \sum_{i=1}^{m} y_i(x^0) A_{ij}}{w_j} - 1 \right)
\]

\[= \Omega(\beta) \sum_{i \in S^- \land x_j^0(1 - \beta) \geq \delta_j} \sum_{i = 1}^{m} y_i(x^0) A_{ij} = \left(1 + \gamma \right). \quad (28)\]

Next, we prove that the potential increase due to decrease of \(x_j\) such that \(\{ j : j \in S^- \land x_j^0(1 - \beta) < \delta_j \}\) is dominated by the potential increase due to \(x_k\)‘s that decrease multiplicatively.

Choose any \(x_j\) such that \(\{ j : j \in S^- \land x_j^0(1 - \beta) < \delta_j \}\), and let \(\xi_j = \frac{(x_j^0)^\alpha \sum_{i=1}^{m} A_{ij} y_i(x^0)}{w_j}\). From Lemma 4.34 there exists at least one \(i\) with \(A_{ij} \neq 0\), such that:

\[
y_i \geq \frac{w_j(x_j^0)^\alpha}{w_j(x_j^0)^\alpha} \sum_{i=1}^{m} y_i(x^0) A_{ij} \geq \frac{1}{m A_{\text{max}}} \sum_{i=1}^{m} y_i(x^0) A_{ij} \geq 1 - \beta \sum_{i=1}^{m} y_i(x^0) A_{ij} \geq 1 - \beta \sum_{i=1}^{m} y_i(x^0) A_{ij} \geq \frac{1 - \beta}{m A_{\text{max}}} \xi_j, \quad \text{and},
\]

\[
\sum_{i=1}^{m} A_{ij} x_j^0 > 1 - \frac{\xi_j}{2}. \quad (30)
\]

From (30), there exists at least one \(p\) such that \(A_{ip} \neq 0\) and

\[
A_{ip} x_p^0 > \frac{1 - \frac{\xi_j}{2}}{n}. \quad (31)
\]

Since \(A_{ip} \leq A_{\text{max}}\) and for \(\alpha < 1\):

\[
x_p^0 > \left(1 - \frac{\xi_j}{2}\right) \sum_{k=1}^{n} \delta_k > \left(1 - \frac{\xi_j}{2}\right) (1 - \beta) \sum_{\{k : k \in S^- \land x_k^0(1 - \beta) < \delta_k\}} x_k^0 \geq \Theta(1) \sum_{\{k : k \in S^- \land x_k^0(1 - \beta) < \delta_k\}} x_k^0. \quad (32)
\]

Since \(x_p^0 \in (0, 1)\) and \(\alpha \in (0, 1)\), we have that \(A_{ip}(x_p^0)^\alpha \geq A_{ip}(x_p^0) \geq \frac{1 - \frac{\xi_j}{2}}{n} \). Using (29) and (32):

\[
(x_p^0)^\alpha \sum_{l=1}^{m} A_{ip} y_l(x^0) \geq (x_p^0)^\alpha A_{ip} y_i(x^0) \geq \frac{1 - \frac{\xi_j}{2}}{n} \frac{1 - \beta}{m A_{\text{max}}} \xi_j.
\]

Recalling that \(\frac{w_j}{\delta_j} = C \geq 2w_{\text{max}} n m A_{\text{max}}\), it further follows that:

\[
(x_p^0)^\alpha \sum_{l=1}^{m} A_{ip} y_l(x^0) \geq 2 \left(1 - \frac{\xi_j}{2}\right) (1 - \beta) w_{\text{max}} \xi_j.
\]

Since \(\varepsilon \leq \frac{1}{6}\) and \(\beta = \frac{\gamma}{4(\kappa + 1)} = \frac{\gamma}{16(\kappa + 1)} < \frac{\varepsilon}{16}\), it follows that \(2 \left(1 - \frac{\xi_j}{2}\right) (1 - \beta) > 1\). Therefore:

\[
\frac{(x_p^0)^\alpha \sum_{i=1}^{m} A_{ip} y_l(x^0)}{w_p} \geq \frac{(x_p^0)^\alpha \sum_{i=1}^{m} A_{ip} y_l(x^0)}{w_{\text{max}}} \geq \frac{(x_j^0)^\alpha \sum_{i=1}^{m} A_{ij} y_i(x^0)}{w_j}. \quad (33)
\]

24
Since \( \xi_j \geq (1 + \gamma) \) and \( x_p^0 > 1 - \beta \), it immediately follows that \( x_p \) decreases by a factor \( (1 - \beta) \). Combining this result with (39), (32), and (28):

\[
\Phi^1 - \Phi^0 \geq \Omega(\beta) \sum_{j \in S^-} w_j(x_j^0)^{1-\alpha} \left( \frac{(x_j^0)^\alpha \sum_{i=1}^n y_i(x^0) A_{ij}}{w_j} - (1 + \gamma) \right).
\]

Finally, since for \( j \notin S^- \):

\[
\Phi^1 - \Phi^0 \geq \Omega(\beta) \sum_{j=1}^n w_j(x_j^0)^{1-\alpha} \left( \frac{(x_j^0)^\alpha \sum_{i=1}^m y_i(x^0) A_{ij}}{w_j} - (1 + \gamma) \right)
= \Omega(\beta) \left( \sum_{j=1}^n x_j^0 \sum_{i=1}^m y_i(x^0) A_{ij} - (1 + \gamma) \sum_{j=1}^n w_j(x_j^0)^{1-\alpha} \right),
\]

completed the proof. \( \square \)

The rest of the convergence results follow by appropriately adapting the results from Section 3.3 in [5]. They are provided here for completeness.

Let \( x^t, y(x^t), \Phi^t \) denote the values of \( x, y(x), \Phi \) at the beginning of round \( t \).

**Definition D.2.** (Stationary interval.) For \( t_0 \geq \tau = \min \frac{1}{\beta} \ln \left( \frac{1}{\delta} \right) \), an interval \( T = [t_0, t_1] \) of rounds is said to be stationary if for all \( t \in T \):

- \((1 - 2\gamma) \sum_{j=1}^n w_j(x_j^t)^{1-\alpha} \leq \sum_{j=1}^n x_j^t \sum_{i=1}^m y_i(x^t) A_{ij} \leq (1 + 2\gamma) \sum_{j=1}^n w_j(x_j^t)^{1-\alpha} \), and
- \( \sum_{t' \in T} \sum_{i=1}^{\infty} |y_i(x^{t'+1}) - y_i(x^{t'})| \leq \frac{2}{\beta} \sum_{j=1}^n w_j(x_j^t)^{1-\alpha} \).

**Lemma D.3.** During any non-stationary interval \( T = [t_0, t_1] \) the potential increases by at least \( \Omega \left( \frac{\kappa^2}{\kappa} \right) \).

\((1 - \alpha) p_\alpha(x^t) = \Omega \left( \frac{\kappa^2}{\kappa} \right) \cdot \sum_{j=1}^n w_j(x_j^t)^{1-\alpha} \) for some \( t \in T \).

**Proof.** Follows directly from Definition [D.2] and Lemma [D.1]. \( \square \)

**Lemma D.4.** In any stationary interval \( T = [t_0, t_1] \), where \( t_0 \geq \tau_0 = \min \frac{1}{\beta} \ln \left( \frac{1}{\delta} \right) \) and \( t_1 - t_0 \geq \tau_1 = \max_{j} \frac{1}{\beta} \ln \left( \frac{1}{\delta_j} \right) \), \( x^t \) is a \((1 + 6\varepsilon)\)-approximate solution to the \( \alpha \)-fair packing problem for all \( t \in T \).

**Proof.** The proof is by contradiction and follows a similar line of argument as the proof of Lemma 3.7 in [5]. Let \( x^* \) denote the optimal solution. Recall that \( p_\alpha(x) = \frac{1}{1 - \alpha} \sum_j w_j x_j^{1-\alpha} \). Suppose that for some round \( t' \in T \) the approximation ratio \( r = \frac{p_\alpha(x^t)}{p_\alpha(x^{t'})} > 1 + 6\varepsilon \). Using that \( T \) is stationary:

\[
(1 + 2\gamma)(1 - \alpha)p_\alpha(x^t) = r(1 + 2\gamma)(1 - \alpha)p_\alpha(x^{t'})
\geq r \sum_{j=1}^n x_j^{t'} \sum_{i=1}^m y_i(x^{t'}) A_{ij}
\geq (1 + 6\varepsilon) \sum_{j=1}^n x_j^{t'} \sum_{i=1}^m y_i(x^{t'}) A_{ij}.
\]

From Lemma [4.2] \( \sum_{j=1}^n x_j^{t'} \sum_{i=1}^m y_i(x^{t'}) A_{ij} \geq (1 - 4\varepsilon) \sum_{i=1}^m y_i(x^{t'}) \), and therefore:

\[
(1 + 2\gamma)(1 - \alpha)p_\alpha(x^t) > (1 + 6\varepsilon)(1 - 4\varepsilon) \sum_{i=1}^m y_i(x^{t'}).
\]

25
Since $x^*$ is feasible, we have that $\sum_{j=1}^n A_{ij} x^*_j \leq 1$ for all $i$. Therefore:

$$(1 + 2\gamma)(1 - \alpha)p_\alpha(x^*) > (1 + 6\varepsilon)(1 - 4\varepsilon) \sum_{i=1}^m y_i(x') \sum_{j=1}^n A_{ij} x^*_j.$$ Changing the order of summation, and using $\gamma = \frac{2}{3}$ we get:

$$\frac{\sum_{j=1}^n x^*_j \sum_{i=1}^m y_i(x') A_{ij}}{(1 - \alpha)p_\alpha(x^*)} < 1 - 2\gamma.$$ (34)

On the other hand:

$$\begin{align*}
\sum_{j=1}^n x^*_j \sum_{t \in T} \sum_{i=1}^m y_i(x^{t+1}) A_{ij} - \sum_{i=1}^m y_i(x') A_{ij} &\leq \sum_{j=1}^n x^*_j \sum_{t \in T} \sum_{i=1}^m A_{ij} |y_i(x^{t+1}) - y_i(x')| \\
&= \sum_{i=1}^m \sum_{t \in T} |y_i(x^{t+1}) - y_i(x')| \sum_{j=1}^n A_{ij} x^*_j \\
&\leq \sum_{i=1}^m \sum_{t \in T} |y_i(x^{t+1}) - y_i(x')| \\
&< \frac{\gamma}{2}(1 - \alpha)p_\alpha(x^*)
\end{align*}$$

The last two inequalities follow from feasibility ($\sum_{j=1}^n A_{ij} x^*_j \leq 1$), the initial assumption that the interval $T$ is stationary ($\sum_{t \in T} \sum_{i=1}^m |y_i(x^{t+1}) - y_i(x')| \leq \frac{\gamma}{2}(1 - \alpha)p_\alpha(x^*)$), and optimality of $x^*$ ($p(x') = \frac{1}{\gamma}p(x^*) < p(x^*)$ by the initial assumption).

Adding (33) and (35):

$$\begin{align*}
\sum_{j=1}^n x^*_j \left( \sum_{i=1}^m y_i(x') A_{ij} + \sum_{t \in T} \sum_{i=1}^m y_i(x^{t+1}) A_{ij} - \sum_{i=1}^m y_i(x') A_{ij} \right) \\
&\leq \sum_{j=1}^n x^*_j \left( \frac{\sum_{i=1}^m y_i(x') A_{ij}}{w_j} + \sum_{t \in T} \left( \frac{\sum_{i=1}^m y_i(x^{t+1}) A_{ij}}{w_j} - \frac{\sum_{i=1}^m y_i(x') A_{ij}}{w_j} \right) \right) \\
&\leq \sum_{k=1}^n w_k (x^*)^{1-\alpha} < 1 - \gamma.
\end{align*}$$ (36)

Since $x^*_j \in (0, 1) \forall j$ and $\alpha \in (0, 1)$, it follows that $(x^*_j)^{1-\alpha} \geq x^*_j \forall j$, and therefore $\sum_{k=1}^n w_k (x^*)^{1-\alpha} \leq 1$. Observe that (36) is a weighted sum of positive terms $\sum_{k=1}^n w_k (x^*)^{1-\alpha}$ with positive weights $\sum_{k=1}^n w_k (x^*)^{1-\alpha}$, such that all the weights sum up to a value that is $\leq 1$. Therefore, there must exist at least one $j$ such that:

$$\begin{align*}
\sum_{i=1}^m y_i(x') A_{ij} + \sum_{t \in T} \left( \frac{\sum_{i=1}^m y_i(x^{t+1}) A_{ij}}{w_j} - \frac{\sum_{i=1}^m y_i(x') A_{ij}}{w_j} \right) &< 1 - \gamma.
\end{align*}$$ (37)

The first summand in (37) is the value of $\sum_{i=1}^m y_i(x') A_{ij}$ in one specific round $t' \in T$, while the the second is the sum of the absolute increments over the interval $T$. It follows that $\forall t \in T$: $\sum_{i=1}^m y_i(x') A_{ij} < 1 - \gamma$,
and, since \( x_j^t \in (0, 1] \) (from feasibility) and \( \alpha > 0 \), we have in fact that \( \forall t \in \mathcal{T} \):

\[
\frac{(x_j^t)^\alpha \sum_{i=1}^m y_i(x^t) A_{ij}}{w_j} < 1 - \gamma.
\]

From the algorithm description, \( x_j \) increases multiplicatively by a factor \((1 + \beta)\) in each round \( t \in \mathcal{T} \). But then after \( \tau_1 = \max_j \frac{1}{\beta} \ln \left( \frac{1}{\delta_j} \right) \geq \frac{1}{\beta} \ln \left( \frac{1}{\delta_j} \right) \) rounds, it must be \( x_j \geq 1 \). Since \( |\mathcal{T}| = t_1 - t_0 > \tau_1 \), after interval \( \mathcal{T} \): \( x_j > 1 \). But then the solution \( x \) becomes infeasible, which is a contradiction (recall Lemma 4.1), and therefore the initial assumption \( \tau > 1 + 6\epsilon \) is not true.

**Proof of Theorem 4.6** First, consider the minimum (initial) and the maximum (final) values of the potential \( \Phi \).

At initialization, \( x_j^0 = \delta_j \leq 2n^2_{A_{\text{max}}} \forall j \). Therefore, \( p_\alpha(x^0) = \sum_j w_j \delta_j^{1-\alpha} > 0 \) and \( \sum_j A_{ij}x_j^0 = \sum_j A_{ij} \delta_j \leq n A_{\text{max}} \frac{1}{nA_{\text{max}}} = \frac{1}{2n} \leq \frac{1}{2} \forall i \). Recalling that \( \epsilon \leq \frac{1}{8} \):

\[
y_i(x^0) = C \cdot e^{\kappa(\sum_j A_{ij}x_j^0 - 1)} \leq C \cdot \left( \frac{\min_j w_j}{w_{\text{max}} C m A_{\text{max}}} \right)^{1-1/\epsilon} \leq C \cdot \left( \frac{1}{C m A_{\text{max}}} \right)^3 \leq \frac{1}{4n^2 m^3 (A_{\text{max}})^5}.
\]

It follows that \( \sum_{i=1}^m y_i(x^0) \leq m 4n^2 m^3 (A_{\text{max}})^5 \leq \frac{1}{4n^2 (A_{\text{max}})^5} \), and therefore:

\[
\Phi_{\text{min}} = \Phi(x^0) = p_\alpha(x^0) - \frac{1}{\kappa} \sum_{i=1}^m y_i(x^0) \geq -\frac{1}{4n^2 (A_{\text{max}})^5}.
\]

On the other hand, since \( y_i(x) > 0 \) and \( x_j \in (0, 1] \) for all feasible \( x \):

\[
\Phi_{\text{max}} \leq \sum_{j=1}^n w_j \frac{1}{1 - \alpha} = \frac{W}{1 - \alpha}.
\] (39)

From Lemma 4.2 after at most \( \tau_0 = O \left( \frac{1}{\beta} \log \left( \frac{1}{\max_i \delta_j} \right) \right) \) rounds, there exists at least one \( i \) such that \( \sum_j A_{ij}x_j \geq 1 - (1 + 1/\kappa) \epsilon \). Since \( A_{ij} \leq A_{\text{max}} \forall i, j \), it is also true that \( \sum_j x_j \geq 1 - (1 + 1/\kappa) \epsilon \), and as \( x_j^{1-\alpha} \geq x_j \) and \( \frac{w_j}{w_{\text{max}}} = \Theta(1) \forall j, k \), it follows that \( \sum_j w_j x_j^{\alpha} \geq \Omega \left( \frac{W}{(n A_{\text{max}})} \right) \).

Consider the value of the potential after at most \( \tau_0 \) rounds. Recall from Lemma 4.5 that the potential never decreases. If \( \Phi = o \left( \frac{W}{(1 - \alpha) n A_{\text{max}}} \right) \), then \( \sum_j y_i(x) \geq \kappa \frac{1}{\alpha} \sum_j w_j x_j^{\alpha} \geq \Omega \left( \frac{W}{(1 - \alpha) n A_{\text{max}}} \right) \), and from Lemma 4.2 \( \sum_i y_i(x) \sum_j A_{ij} x_j \geq (1 - 4\epsilon) \sum_i y_i(x) = \Omega \left( \frac{W}{(1 - \alpha) n A_{\text{max}}} \right) \). From the third part of Lemma 4.7, the potential increases additively by at least \( \Omega \left( \beta \kappa \frac{W}{(1 - \alpha) n A_{\text{max}}} \right) \), and after at most \( O \left( \frac{1}{\gamma} \right) \) rounds:

\[
\Phi = \Omega \left( \frac{W}{(1 - \alpha) n A_{\text{max}}} \right).
\]

Finally, consider the rounds in which \( \Omega \left( \frac{W}{(1 - \alpha) n A_{\text{max}}} \right) \leq \Phi \leq \Phi_{\text{max}} \leq \frac{W}{1 - \alpha} \). From Lemmas 4.3 and 4.4 in any stationary interval, \( p_\alpha(x^*) \leq (1 + 6\epsilon) p_\alpha(x) \) throughout the interval, while in any non-stationary interval (of length \( \leq \tau_1 \)), there exists at least one round \( t \) in which the potential increases by at least \( \Omega \left( \frac{2\epsilon}{\kappa} \right) \cdot (1 - \alpha) \) \( p_\alpha(x^*) \geq \Omega \left( \frac{2\epsilon}{\kappa} \right) \cdot (1 - \alpha) \Phi(x^*) \). Since the potential never decreases, and in each non-stationary interval it increases by a multiplicative factor \( \Omega \left( 1 + \frac{2\epsilon}{\kappa} \cdot (1 - \alpha) \right) \), there can be at most \( O \left( \frac{\ln(n A_{\text{max}})}{\ln(1 + \frac{2\epsilon}{\kappa} - (1 - \alpha))} \right) = O \left( \frac{\kappa}{\gamma^2 (1 - \alpha)} \ln \left( n A_{\text{max}} \right) \right) \) non-stationary intervals, each of length at most \( \tau_1 \).

Therefore, the total convergence time is at most:

\[
O \left( \frac{\kappa \tau_1}{\gamma^2 (1 - \alpha)} \ln \left( n A_{\text{max}} \right) \right) = O \left( \frac{1}{\alpha (1 - \alpha) \epsilon^5} \ln \left( n A_{\text{max}} \right) \ln \left( mn A_{\text{max}} \right) \ln^2 \left( \frac{mn A_{\text{max}}}{\epsilon} \right) \right).
\]
D.5 Convergence Proof for $\alpha = 1$

The proof outline for the convergence of Algorithm 1 in the $\alpha = 1$ case is as follows. First, we show that in any round it cannot be the case that only “small-value” $x_j$’s (i.e., $x_j$’s that are smaller than $\frac{\delta_j}{1-\beta}$) decrease (Lemma D.5). This result implies that in any round in which at least one $x_j$ updates (i.e., increases or decreases), at least one update must be due to multiplicative increase or decrease, which provides a sufficient progress in the potential increase (Lemma D.6). Finally, since the total increase in the potential is bounded, we yield a bound on the convergence time, that is, on the number of rounds until $x_j \sum_{m=1}^{x_j} \frac{y_i(x)}{w_j} \in (1 - \gamma, 1 + \gamma)$ for all $j$. Using the expression for the duality gap (see Eq. 3), we show that after convergence the duality gap is upper-bounded by $7W\varepsilon$, which further implies additive $7W\varepsilon$ approximation of the solution to which the algorithm converges.

**Lemma D.5.** Whenever some variable $x_j < \frac{\delta_j}{1-\beta}$ decreases, there exists another variable $x_l > \frac{\delta_l}{1-\beta}$ that also decreases.

**Proof.** Fix any round, and let $x^0$, $y(x^0)$ and $x^1$, $y(x^1)$ denote the values of $x$, $y$ at the beginning and at the end of the round, respectively. Suppose that some $x_j^0 < \frac{\delta_j}{1-\beta}$ decreases. Then from Lemma 4.3 there exists at least one $i \in \{1, ..., m\}$ such that $A_{ij} \neq 0$, and:

- $\sum_{k=1}^{n} A_{ik} x_k^0 > 1 - \frac{\varepsilon}{2}$, and
- $y_i(x) \geq \frac{\sum_{k=1}^{n} A_{ik} x_k^0}{m_{\text{max}}} > (1 - \beta) \frac{w_i}{\delta_j} \frac{1}{m_{\text{max}}} \frac{x_j^0 \sum_{m=1}^{x_j} y_i(x) A_{ij}}{w_j}$.

Since $\sum_{k=1}^{n} A_{ik} x_k^0 > 1 - \frac{\varepsilon}{2}$, there exists at least one $p$ such that $A_{ip} x_p^0 > \frac{1 - \frac{\varepsilon}{n}}{n}$. As $x_j$ decreases, we must have $\frac{x_j^0 \sum_{m=1}^{x_j} y_i(x) A_{ij}}{w_j} \geq (1 + \gamma)$, and therefore $y_i \geq (1 - \beta) \frac{w_i}{\delta_j} \frac{1}{m_{\text{max}}} (1 + \gamma)$. Recalling that $C = \frac{w_i}{\delta_j} \geq 2w_{\text{max}}nmA_{\text{max}}$:

$$(x_p^0) A_{ip} y_i(x^0) \geq C \cdot \frac{(1 + \gamma)(1 - \beta)}{m_{\text{max}}} \cdot \frac{1 - \frac{\varepsilon}{n}}{n} \geq 2w_{\text{max}}nmA_{\text{max}} \cdot \frac{(1 + \gamma)(1 - \beta)}{m_{\text{max}}} \cdot \frac{1 - \frac{\varepsilon}{n}}{n} \geq 2w_{\text{max}}(1 + \gamma)(1 - \beta) \left(1 - \frac{\varepsilon}{2}\right) \geq 2w_{\text{max}} \left(1 + \gamma \beta \frac{\varepsilon}{4}\right) \geq 2w_{\text{max}} \left(1 + \gamma \frac{\varepsilon}{4}\right) \geq 2w_{\text{max}} \cdot \frac{3}{4} \cdot \left(1 + \frac{\gamma}{2}\right) \geq w_{\text{max}}(1 + \gamma).$$

Finally, we have that:

$$(x_p^0) \sum_{i=1}^{m} y_i(x^0) A_{ip} \geq (x_p^0) A_{ip} y_i(x^0) > w_{\text{max}}(1 + \gamma) \geq w_p(1 + \gamma),$$

and, therefore $x_p^0$ must decrease as well. Since $x_p^0 > \frac{1 - \frac{\varepsilon}{n} A_{\text{max}}}{10nA_{\text{max}}} > \frac{11}{12nA_{\text{max}}} > \frac{\delta_p}{1-\beta}$, the decrease is by a factor of $1 - \beta$, i.e., $x_p^0 = (1 - \beta)x_p^0$. □
Lemma D.6. In any round of the algorithm run if there exists at least one $j$ such that $\frac{x_j \sum_{i=1}^m y_i(x)A_{ij}}{w_j} \notin (1-\gamma, 1+\gamma)$, then the potential increases by at least $\min_k w_k \Omega(\beta \gamma)$.

Proof. Suppose that in the observed round there exists at least one $j$ such that $\frac{x_j \sum_{i=1}^m y_i(x)A_{ij}}{w_j} \notin (1-\gamma, 1+\gamma)$. Then $x_j$ is either increased or decreased by the algorithm. Let $x^0, y(x^0), \Phi^0$ and $x^1, y(x^1), \Phi^1$ denote the values of $x, y(x), \Phi$ before and after the round, respectively. From Lemma 4.5:

$$\Phi^1 - \Phi^0 \geq \sum_{k=1}^n w_k \left| x_k^1 - x_k^0 \right| \left( 1 - \frac{x_k^1 \sum_{i=1}^m y_i(x^1)A_{ik}}{w_k} \right).$$

If $\frac{x_j^0 \sum_{i=1}^m y_i(x^0)A_{ij}}{w_j} \leq 1-\gamma$, then $x_j^1 = (1+\beta)x_j^0$ and $\left| 1 - \frac{x_j^1 \sum_{i=1}^m y_i(x^1)A_{ij}}{w_j} \right| > \frac{\gamma}{2}$ (see Proof of Lemma 4.5), and therefore $\Phi^1 - \Phi^0 \geq w_j \left| x_j^1 - x_j^0 \right| \left( 1 - \frac{x_j^1 \sum_{i=1}^m y_i(x^1)A_{ij}}{w_j} \right) > w_j \frac{\beta}{1+\beta} \frac{\gamma}{2} = w_j \Omega(\beta \gamma)$.

If $\frac{x_j^0 \sum_{i=1}^m y_i(x^0)A_{ij}}{w_j} \geq 1+\gamma$, then, from Lemma D.5, either $x_j^1 = (1-\beta)x_j^0$, or there exists some other $x_p$ with $x_p^1 = (1-\beta)x_p^0$. In the former case $\left| 1 - \frac{x_j^1 \sum_{i=1}^m y_i(x^1)A_{ij}}{w_j} \right| > \frac{\gamma}{2}$, while in the latter $\left| 1 - \frac{x_p^1 \sum_{i=1}^m y_i(x^1)A_{ip}}{w_p} \right| > \frac{\gamma}{2}$ (see Proof of Lemma 4.5). Therefore:

$$\Phi^1 - \Phi^0 \geq \min \left\{ w_j \frac{|x_j^1 - x_j^0|}{x_j^1} \left( 1 - \frac{x_j^1 \sum_{i=1}^m y_i(x^1)A_{ij}}{w_j} \right), w_p \frac{|x_p^1 - x_p^0|}{x_p^1} \left( 1 - \frac{x_p^1 \sum_{i=1}^m y_i(x^1)A_{ip}}{w_p} \right) \right\} \geq \min \{ w_j, w_p \} \frac{\beta}{1-\beta} \frac{\gamma}{2},$$

which completes the proof. \qed

Proof of Theorem 4.7. At initialization, we have that $\forall j \in \{1, ..., n\}$ $x_j = \delta_j = \frac{w_j}{\min_i w_i n A_{max} \max_i}$, and therefore $\forall i: \sum_{j=1}^m A_{ij} x_j = \sum_{j=1}^m A_{ij} \delta_j \leq \frac{1}{2}$. Recalling that $\varepsilon \leq \frac{1}{m}$, we can upper-bound each $y_i$ as:

$$y_i(x) \leq Ce^{\kappa(A_{max} \sum_{j=1}^m \delta_j - 1)} = C \left( \frac{\varepsilon}{C m A_{max}} \right)^{(1-1/2)/\varepsilon} \leq C \left( \frac{\varepsilon}{C m A_{max}} \right)^{3} < \frac{1}{m^3}.$$

It follows that the initial potential $\Phi_0$ can be lower-bounded as:

$$\Phi_0 \geq \sum_{j=1}^n w_j \ln(\delta_j) - \frac{1}{\kappa} \sum_{i=1}^m y_i \geq \sum_{j=1}^n w_j \ln \left( \frac{2A_{max} w_j}{w_i \min_i w_i n A_{max}} \right) - \frac{1}{\kappa m^2} = -\Theta(W \ln(n m A_{max})). \quad (40)$$

As $x$ remains feasible throughout the course of the algorithm, we have that $x_j \leq 1 \forall j \in \{n\}$. Noticing that $y_i(x) > 0 \forall x \in \mathbb{R}^n$, we get that $\Phi(x) \leq 0$ at any point throughout the algorithm execution. Therefore, the maximum total increase in $\Phi$ is: $\Delta \Phi = \Theta(W \ln(n m A_{max}))$.

From Lemma D.6 there exists at least one $x_j$ with:

$$\frac{x_j \sum_{i=1}^m y_i(x)A_{ij}}{w_j} \notin (1-\gamma, 1+\gamma),$$

29
potential increases by at least $\Delta \phi = \min_k w_k \Omega(\beta \gamma)$. Therefore, after at most
\[
O\left(\frac{\Delta \phi}{\Delta \phi}\right) = O\left(\frac{W}{\ln(nm A_{\max})} \ln\left(\frac{nm A_{\max}}{\varepsilon \min_k w_k}\right)\right)
\]
rounds, we have:
\[
\frac{x_j \sum_{i=1}^{m} y_i(x) A_{ij}}{w_j} \in (1 - \gamma, 1 + \gamma), \quad \forall j \in \{1, ..., n\}.
\]
Recall the expression (3) for the duality gap:
\[
G_1(x, y) = -\sum_{j=1}^{n} w_j \ln\left(\frac{x_j \sum_{i=1}^{m} y_i A_{ij}}{w_j}\right) + \sum_{i=1}^{m} y_i - W.
\]
From the second part of Lemma 4.2, after $O\left(\frac{1}{\beta} \ln\left(\frac{1}{\min_j \alpha_j}\right)\right)$ rounds:
\[
\sum_{i=1}^{m} y_i \leq (1 + 6\varepsilon) \sum_{j=1}^{n} x_j \sum_{i=1}^{m} y_i A_{ij}.
\]
Therefore, after $O\left(\frac{\Delta \phi}{\Delta \phi}\right)$ rounds:
\[
G_1(x, y) \leq -\sum_{j=1}^{n} w_j \ln\left(\frac{x_j \sum_{i=1}^{m} y_i A_{ij}}{w_j}\right) + (1 + 6\varepsilon) \sum_{j=1}^{n} w_j \frac{x_j \sum_{i=1}^{m} y_i A_{ij}}{w_j} - W,
\]
and
\[
\frac{x_j \sum_{i=1}^{m} y_i(x) A_{ij}}{w_j} \in (1 - \gamma, 1 + \gamma), \quad \forall j.
\]
Let $\xi_j = \frac{x_j \sum_{i=1}^{m} y_i(x) A_{ij}}{w_j}$. Then:
\[
G_1(x, y) \leq -\sum_{j=1}^{n} w_j \ln(\xi_j) + (1 + 6\varepsilon) \sum_{j=1}^{n} w_j \xi_j - W = \sum_{j=1}^{n} w_j \Gamma_j(\xi_j) - W,
\]
(41)
where $\Gamma_j(\xi_j) = -\ln(\xi_j) + (1 + 6\varepsilon)\xi_j$. Now each function $\Gamma_j(\xi_j)$ is strictly convex in $\xi_j$, and since $\xi_j \in (1 - \gamma, 1 + \gamma)$, we get:
\[
\Gamma_j(\xi_j) < \max\{\Gamma_j(1 - \gamma), \Gamma_j(1 + \gamma)\} = \max\{-\ln(1 - \gamma) + (1 + 6\varepsilon)(1 - \gamma), -\ln(1 + \gamma) + (1 + 6\varepsilon)(1 + \gamma)\}.
\]
The inequality $\ln(1 \pm \gamma) \geq \pm \gamma - \gamma^2$ holds since $\gamma = \frac{\varepsilon}{4} \leq \frac{\varepsilon}{4}$ (the inequality is in fact satisfied for any $\gamma \leq 0.65$). Therefore:
\[
\Gamma_j(\xi_j) < (1 + 6\varepsilon) + \max\{\gamma - (1 + 6\varepsilon)\gamma + \gamma^2, -\gamma + (1 + 6\varepsilon)\gamma + \gamma^2\} = (1 + 6\varepsilon) + 6\varepsilon\gamma + \gamma^2.
\]
(42)
Plugging (42) back into (41), and recalling that $\sum_{j=1}^{n} w_j = W$:
\[
\sum_{j=1}^{n} w_j \Gamma_j(\xi_j) - W < W(6\varepsilon + \varepsilon^2) + \left(\frac{\varepsilon}{4}\right)^2 + \frac{\varepsilon^2}{4} < 7W\varepsilon,
\]
which completes the proof.
\[\square\]
D.6 Omitted Proofs from Section 4.2

The following technical Lemma is used in the proof of Lemma 4.2:

**Lemma D.7.** In any round of the algorithm:

\[
\sum_{\{i,j\in S^- \land x_j^0 \geq \frac{\delta_j}{1-\beta}\}} x_j^0 \sum_{i=1}^{m} y_i(x^0)A_{ij} \geq \frac{1}{2} \sum_{j \in S^-} x_j^0 \sum_{i=1}^{m} y_i(x^0)A_{ij};
\]

and

\[
\sum_{\{i,j\in S^- \land x_j^0 \geq \frac{\delta_j}{1-\beta}\}} \left( x_j^0 \sum_{i=1}^{m} y_i(x^0)A_{ij} - (1 + \gamma)w_j(x_j^0)^{1-\alpha} \right) \geq \frac{1}{2} \sum_{j \in S^-} \left( x_j^0 \sum_{i=1}^{m} y_i(x^0)A_{ij} - (1 + \gamma)w_j(x_j^0)^{1-\alpha} \right).
\]

**Proof.** If \( x_j^0 \geq \frac{\delta_j}{1-\beta} \) \( \forall j \) there is nothing to prove, so assume that there exists at least one \( j \) with \( x_j^0 < \frac{\delta_j}{1-\beta} \). The proof proceeds as follows. First, we show that for each \( j \) for which \( x_j \) decreases by a factor less than \( (1 - \beta) \) there exists at least one \( x_p \) that appears in at least one constraint \( i \) in which \( x_j \) appears and decreases by a factor \( (1 - \beta) \). We then proceed to show that \( x_p \) is in fact such that \( x_p^0 \sum_{i=1}^{m} y_i(x^0)A_{ip} = \Omega(n)x_p^0 \sum_{i=1}^{m} y_i(x^0)A_{ij} \) and \( x_p^0 \sum_{i=1}^{m} y_i(x^0)A_{ip} - (1 + \gamma)w_p(x_p^0)^{1-\alpha} = \Omega(n)(x_p^0 \sum_{i=1}^{m} y_i(x^0)A_{ij} - (1 + \gamma)w_j(x_j^0)^{1-\alpha}) \). This further implies that the terms \( x_p^0 \sum_{i=1}^{m} y_i(x^0)A_{ip} \) and \( x_p^0 \sum_{i=1}^{m} y_i(x^0)A_{ip} - (1 + \gamma)w_p(x_p^0)^{1-\alpha} \) dominate the sum of all the terms corresponding to \( x_j \)'s with \( A_{ij} \neq 0 \) and \( x_j < \frac{\delta_j}{1-\beta} \), thus completing the proof.

From Lemma 4.3 for each \( j \in S^- \) with \( x_j < \frac{\delta_j}{1-\beta} \) there exists at least one constraint \( i \) such that:

- \( \sum_{k=1}^{n} A_{ik}x_k^0 > 1 - \frac{\epsilon}{2} \), and
- \( y_i(x^0) \geq \frac{\sum_{i=1}^{m} y_i(x^0)A_{ij}}{mA_{\max}} \Rightarrow y_i(x^0) > (1 - \beta)^{\alpha} \frac{1}{mA_{\max}} \frac{w_j(x_j^0)^{\alpha}}{\delta_j} \sum_{i=1}^{m} y_i(x^0)A_{ij} \).

Therefore, there exists at least one \( x_p \) with \( A_{ip} \neq 0 \) such that \( A_{ip}x_p^0 > \frac{1 - \frac{\epsilon}{2}}{n} \), which further gives \( A_{ip}(x_p^0)^{\alpha} > \frac{1 - \frac{\epsilon}{2}}{n^{\alpha}} A_{ip}^{1-\alpha} \geq \frac{1 - \frac{\epsilon}{2}}{n^{\alpha}} \cdot A_{\max}^{1-\alpha} \), where the last inequality follows from \( 1 \leq A_{ip} \leq A_{\max} \) and \( \alpha > 1 \). Combining the inequality for \( A_{ip}(x_p^0)^{\alpha} \) with the inequality for \( y_i(x^0) \) above:

\[
\sum_{l=1}^{m} y_l(x^0)A_{il} \geq (x_p^0)^{\alpha} A_{ip}y_i(x^0)
\]

\[
\geq \frac{(1 - \frac{\epsilon}{2})^\alpha}{n^{\alpha}} \cdot A_{\max}^{1-\alpha} (1 - \beta)^{\alpha} \frac{1}{mA_{\max}} \frac{w_j(x_j^0)^{\alpha}}{\delta_j} \sum_{i=1}^{m} y_i(x^0)A_{ij}
\]

\[
= C \cdot \frac{(1 - \frac{\epsilon}{2})^\alpha}{n^{\alpha}mA_{\max}^{\alpha}} (1 - \beta)^{\alpha} \frac{w_j(x_j^0)^{\alpha}}{\delta_j} \frac{\sum_{i=1}^{m} y_i(x^0)A_{ij}}{w_j} \quad \text{(from } C = \frac{w_j}{\delta_j})
\]

\[
\geq 2nw_{\max}(1 - \beta)^{\alpha} \left( 1 - \frac{\epsilon}{2} \right) \frac{(x_p^0)^{\alpha}}{2} \sum_{i=1}^{m} y_i(x^0)A_{ij} \quad \text{(from } C \geq 2w_{\max}n^{\alpha+1}mA_{\max}^{2\alpha-1}).
\]

Using the generalized Bernoulli’s inequality: \( (1 - \frac{\epsilon}{2})^\alpha > 1 - \frac{\epsilon^\alpha}{2} \) and \( (1 - \beta)^{\alpha} > (1 - \beta) \) \cite{31}, and
recalling that $\varepsilon \alpha \leq \frac{9}{10}$, $\beta \leq \frac{16}{9}$, we further get:

$$\left( x_p^0 \right)^\alpha \sum_{l=1}^m y_l(x_0^0) A_{lp} \geq 2n w_{\text{max}} \left( 1 - \frac{9}{10 \cdot 96} \right) \left( 1 - \frac{9}{20} \right) \frac{\left( x_j^0 \right)^\alpha \sum_{i=1}^m y_i(x_0^0) A_{ij}}{w_j}$$

$$\geq n w_{\text{max}} \frac{\left( x_j^0 \right)^\alpha \sum_{i=1}^m y_i(x_0^0) A_{ij}}{w_j},$$

which further implies:

$$\frac{\left( x_p^0 \right)^\alpha \sum_{l=1}^m y_l(x_0^0) A_{lp}}{w_p} \geq n \cdot \frac{\left( x_j^0 \right)^\alpha \sum_{i=1}^m y_i(x_0^0) A_{ij}}{w_j}, \quad (43)$$

as $w_p \leq w_{\text{max}}$. Since $x_j$ decreases, $\frac{\left( x_j^0 \right)^\alpha \sum_{i=1}^m y_i(x_0^0) A_{ij}}{w_j} \geq 1 + \gamma$, and therefore $x_p$ decreases as well.

Using similar arguments, as $A_{ip} x_p^0 > \frac{1 - \frac{8}{n}}{\sum l=1^m A_{ij}}$ and recalling that $y_i(x_0^0) \geq \frac{1}{m \lambda_{\text{max}}} \sum l=1^m A_{ij} y_i(x_0^0) > \frac{1}{m \lambda_{\text{max}}} \frac{1 - \beta}{\delta_j} \cdot x_j^0 \sum_{l=1}^m A_{ij} y_i(x_0^0)$:

$$\sum_{l=1}^m y_l(x_0^0) A_{lp} \geq x_p^0 A_{ip} y_i(x_0^0) \geq \frac{1}{n} \frac{1}{\lambda_{\text{max}}} \frac{1 - \beta}{\delta_j} \cdot x_j^0 \sum_{l=1}^m A_{ij} y_i(x_0^0)$$

$$\geq n x_j^0 \sum_{l=1}^m A_{ij} y_i(x_0^0), \quad (44)$$

as $\delta_j \leq \frac{1}{2^{1/\alpha} m \lambda_{\text{max}}}$ and $2^{1/\alpha} (1 - \frac{8}{n}) (1 - \beta) \geq 2^{1/\alpha} (1 - \frac{8}{9}) (1 - \frac{2}{9}) \geq 1$ (since $\varepsilon \in (0, 1/6]$).

From (44), it follows that

$$x_p^0 \sum_{l=1}^m y_l(x_0^0) A_{lp} \geq \sum_{\{k \in S^- : x_k < \frac{1}{3} \alpha \land A_{ik} \neq 0\}} x_k^0 \sum_{l=1}^m y_l(x_0^0) A_{lk},$$

which further implies the first part of the lemma.

For the second part, consider the following two cases:

**Case 1:** $w_p(x_p^0)^{1-\alpha} \geq w_j(x_j^0)^{1-\alpha}$. Then:

$$x_p^0 \sum_{l=1}^m y_l(x_0^0) A_{lp} - (1 + \gamma) w_p(x_p^0)^{1-\alpha} = w_p(x_p^0)^{1-\alpha} \left( \frac{\left( x_p^0 \right)^\alpha \sum_{l=1}^m y_l(x_0^0) A_{lp}}{w_p} - (1 + \gamma) \right)$$

$$\geq w_j(x_j^0)^{1-\alpha} \left( \frac{\left( x_j^0 \right)^\alpha \sum_{l=1}^m y_l(x_0^0) A_{lp}}{w_p} - (1 + \gamma) \right)$$

$$\geq w_j(x_j^0)^{1-\alpha} \left( \frac{\left( x_j^0 \right)^\alpha \sum_{l=1}^m y_l(x_0^0) A_{ij}}{w_j} - (1 + \gamma) \right) \quad \text{(from (43))}$$

$$\geq n w_j(x_j^0)^{1-\alpha} \left( \frac{\left( x_j^0 \right)^\alpha \sum_{l=1}^m y_l(x_0^0) A_{ij}}{w_j} - (1 + \gamma) \right)$$

$$= n \left( x_j^0 \sum_{l=1}^m y_l(x_0^0) A_{ij} - (1 + \gamma) w_j(x_j^0)^{1-\alpha} \right),$$

implying the second part of the lemma.
Case 2: \(w_p(x_p^0)^{1-\alpha} < w_j(x_j^0)^{1-\alpha}\). Then:
\[
x_p^0 \sum_{l=1}^{m} y_l(x^0) A_{lp} - (1 + \gamma)w_p(x_p^0)^{1-\alpha} > x_p^0 \sum_{l=1}^{m} y_l(x^0) A_{lp} - (1 + \gamma)w_j(x_j^0)^{1-\alpha}
\geq n x_j^0 \sum_{l=1}^{m} y_l(x^0) A_{lj} - (1 + \gamma)w_j(x_j^0)^{1-\alpha} \quad \text{(from (44))}
\geq n \left( x_j^0 \sum_{l=1}^{m} y_l(x^0) A_{lj} - (1 + \gamma)w_j(x_j^0)^{1-\alpha} \right),
\]
thus implying the second part of the lemma and completing the proof. \(\square\)

**Proof of Lemma 4.9**

**Proof of 1.** From Lemma 4.5:
\[
\Phi^1 - \Phi^0 \geq \sum_{j=1}^{n} w_j \left| \frac{x_j^1 - x_j^0}{(x_j^1)^{\alpha}} \right| 1 - \frac{(x_j^1)^{\alpha} \sum_{i=1}^{m} y_i(x^1) A_{ij}}{w_j}.
\]
Using the same arguments as in the proof of part 1 of Lemma D.1, we have in fact that 
\[
\left| 1 - \frac{(x_j^1)^{\alpha} \sum_{i=1}^{m} y_i(x^1) A_{ij}}{w_j} \right| \geq \frac{\gamma (x_j^1)^{\alpha} \sum_{i=1}^{m} y_i(x^1) A_{ij}}{w_j},
\]
which further gives:
\[
\Phi^1 - \Phi^0 \geq \frac{\gamma}{2} \left( x_j^1 - x_j^0 \right) \sum_{i=1}^{m} y_i(x^1) A_{ij} = \gamma \sum_{j=1}^{n} \left| x_j^1 - x_j^0 \right| \sum_{i=1}^{m} y_i(x^1) A_{ij}.
\]
If \(j \in S^+\), then \(x_j^1 = (1 + \beta)x_j^0\), and therefore 
\[
\sum_{i=1}^{m} y_i(x^1) A_{ij} \geq \left( 1 - \frac{\gamma}{4(1 + \beta)} \right) x_j^1 \sum_{i=1}^{m} y_i(x^1) A_{ij} \geq \left( 1 - \frac{\gamma}{4} \right) x_j^1 \sum_{i=1}^{m} y_i(x^0) A_{ij}.
\]
Similarly, if \(j \in S^-\) and \(x_j^0 \geq \frac{\delta_j}{1 - \beta}\), then \(x_j^1 = (1 - \beta)x_j^0\), and therefore 
\[
\sum_{i=1}^{m} y_i(x^1) A_{ij} \geq \left( 1 - \frac{\gamma}{4} \right) x_j^1 \sum_{i=1}^{m} y_i(x^0) A_{ij} \geq \left( 1 - \frac{\gamma}{4} \right) x_j^0 \sum_{i=1}^{m} y_i(x^0) A_{ij}.
\]
Using part 1 of Lemma D.7:
\[
\Phi^1 - \Phi^0 \geq \frac{\gamma}{4} \frac{\beta}{1 + \beta} \sum_{j \in \{S^+ \cup S^-\}} x_j^0 \sum_{i=1}^{m} y_i(x^0) A_{ij}.
\]

**Proof of 2:** Consider \(j \in S^-\) such that \(x_j^0 \geq \frac{\delta_j}{1 - \beta}\). Then \(x_j^1 = (1 - \beta)x_j^0\), and 
\[
\frac{(x_j^1)^{\alpha} \sum_{i=1}^{m} y_i(x^1) A_{ij}}{w_j} \geq (1 + \gamma),
\]
and using Lemma 4.5,

\[
\Phi^1 - \Phi^0 \geq \sum_{\{j \in S^{-} \mid x_j^0 \leq \delta_j^{-} \}} w_j \frac{|x_j^1 - x_j^0|}{(x_j^1)^\alpha} \left(1 - \frac{(x_j^1)^\alpha \sum_{i=1}^{m} y_i(x^1)A_{ij}}{w_j}ight)
\]

\[
\sum_{\{j \in S^{-} \mid x_j^0 \leq \delta_j^{-} \}} w_j \frac{\beta}{(1 - \beta)^\alpha} (x_j^0)^{1-\alpha} \left(1 - \frac{(x_j^0)^\alpha \sum_{i=1}^{m} y_i(x^0)A_{ij}}{w_j}ight) - (1 + \gamma)
\]

\[
(1 - \gamma/4) \frac{\beta}{(1 - \beta)^\alpha} \sum_{\{j \in S^{-} \mid x_j^0 \leq \delta_j^{-} \}} w_j (x_j^0)^{1-\alpha} (\gamma/4) \sum_{i=1}^{m} y_i(x^0)A_{ij} - (1 + \gamma)
\]

Using the second part of Lemma D.7 and the fact that for \(k \notin S^{-} \): \(\frac{(x_k^0)^\alpha \sum_{i=1}^{m} y_i(x^0)A_{ik}}{w_k} < (1 + \gamma)\), we get the desired result:

\[
\Phi^1 - \Phi^0 \geq \frac{1}{2} (1 - \gamma/4) \frac{\beta}{(1 - \beta)^\alpha} \left(\sum_{j=1}^{n} x_j^0 \sum_{i=1}^{m} y_i(x^0) - (1 + \gamma) \sum_{j=1}^{n} w_j (x_j^0)^{1-\alpha}\right).
\]

**Proof of 3:** The proof is equivalent to the proof of Lemma D.1 part 2, and is omitted for brevity. \(\square\)

The following lemma shows that in any non-stationary round (defined in Section 4.2) that happens after initial \(\tau_0\) rounds the negative potential decreases by a multiplicative factor.

**Lemma D.8.** Let \(\Phi^1\) and \(\Phi^0\) denote the potential values before and after any non-stationary round that happens after first \(\tau_0 = \frac{1}{\beta} \min_j \ln \left(\frac{1}{\delta_j}\right)\) rounds. Then:

\[
(-\Phi^0) - (-\Phi^1) = \Phi^1 - \Phi^0 \geq \Omega(\beta\gamma^2) \frac{\alpha - 1}{\alpha} (-\Phi^0).
\]

**Proof.** Consider the changes in the potential in any non-stationary round that happens after initial \(\frac{1}{\beta} \min_j \ln \left(\frac{1}{\delta_j}\right)\) rounds. By definition of a stationary round, we have either of the following two cases:

**Case 1:** \(\sum_{j \in S^{-} \cup S^{-}} x_j \sum_{i=1}^{m} y_i(x)A_{ij} \geq \gamma \sum_{j=1}^{n} w_j x_j^{1-\alpha}\). From the first part of Lemma 4.9, the increase in the potential is:

\[
\Phi^1 - \Phi^0 \geq \Omega(\beta\gamma^2) \sum_{j=1}^{n} w_j x_j^{1-\alpha}.
\]

If \(\sum_{j=1}^{n} w_j x_j^{1-\alpha} \geq \frac{1}{\alpha} \sum_{i=1}^{m} y_i(x)\), then

\[
-\Phi^0 \leq \sum_{j=1}^{n} w_j x_j^{1-\alpha} \left(\frac{1}{\alpha - 1} + 1\right),
\]

and the increase in the potential is at least:

\[
\Phi^1 - \Phi^0 \geq \Omega(\beta\gamma^2) \left(\frac{1}{\alpha - 1} + 1\right)^{-1} (-\Phi^0)
\]

\[
= \Omega(\beta\gamma^2) \frac{\alpha - 1}{\alpha} (-\Phi^0).
\]

Recall from Lemma 4.2 that after at most \(\tau_0 = \frac{1}{\beta} \min_j \ln \left(\frac{1}{\delta_j}\right)\) rounds it is always true that:

\[
\sum_{i=1}^{m} y_i(x) \leq (1 + 6\varepsilon) \sum_{j=1}^{n} x_j \sum_{i=1}^{m} y_i(x)A_{ij} \text{ and } \sum_{j=1}^{n} x_j \sum_{i=1}^{m} y_i(x)A_{ij} \geq (1 - 4\varepsilon) \sum_{i=1}^{m} y_i(x).
\]

If
\[ \sum_{j=1}^{n} \frac{w_j}{x_j^{1-\alpha}} < \frac{1}{\kappa} \sum_{i=1}^{m} y_i(x), \text{ then} \]

\[ -\Phi^0 < \frac{1}{\kappa} \sum_{i=1}^{m} y_i(x) \left( 1 + \frac{1}{\alpha - 1} \right) \leq \frac{1 + 6\varepsilon}{\kappa} \sum_{j=1}^{n} x_j \sum_{i=1}^{m} y_i(x) A_{ij}, \quad (46) \]

and

\[ \sum_{j=1}^{n} x_j \sum_{i=1}^{m} y_i(x) A_{ij} \geq (1 - 4\varepsilon) \sum_{i=1}^{m} y_i(x) > \kappa(1 - 4\varepsilon) \sum_{j=1}^{n} w_j x_j^{1-\alpha}. \quad (47) \]

Since \( \kappa = \frac{1}{\varepsilon} \ln \left( Cm A_{\text{max}}/ (\min_j w_j \varepsilon) \right) \geq \frac{1}{\varepsilon} \) and \( \varepsilon \leq \frac{1}{n} \), it follows that \( \kappa(1 - 4\varepsilon) \geq 2 \), and therefore (using (47)):

\[ \sum_{j=1}^{n} x_j \sum_{i=1}^{m} y_i(x) A_{ij} - (1 + \gamma) \sum_{j=1}^{n} w_j x_j^{1-\alpha} \geq \frac{1 - \gamma}{2} \sum_{j=1}^{n} x_j \sum_{i=1}^{m} y_i(x) A_{ij}. \]

From the second part of Lemma 4.9 we have that: \( \Phi^1 - \Phi^0 \geq \Omega \left( \frac{\beta}{(1+\beta)^{\alpha}} \right) \frac{1-\gamma}{\alpha} \sum_{j=1}^{n} x_j \sum_{i=1}^{m} y_i(x) A_{ij} \), which combined with (46) gives:

\[ \Phi^1 - \Phi^0 \geq \Omega (\beta \kappa) (-\Phi^0) > \Omega (\beta \gamma^2) \frac{\alpha - 1}{\alpha} (-\Phi^0). \quad (48) \]

**Case 2:** \( (1 - 2\gamma) \sum_{j=1}^{n} w_j x_j^{1-\alpha} > \sum_{j=1}^{n} x_j \sum_{i=1}^{m} y_i(x) A_{ij} \). Using part 3 of Lemma 4.9, the increase in the potential is then \( \Phi^1 - \Phi^0 \geq \Omega \left( \frac{\beta}{(1+\beta)^{\alpha}} \right) \sum_{j=1}^{n} w_j x_j^{1-\alpha} \). On the other hand, as \( \sum_{j=1}^{n} x_j \sum_{i=1}^{m} y_i(x) A_{ij} \geq (1 - 4\varepsilon) \sum_{i=1}^{m} y_i(x) \), we have that \( \sum_{j=1}^{n} w_j x_j^{1-\alpha} > \frac{1 - 4\varepsilon}{1 - 2\gamma} \sum_{i=1}^{m} y_i(x) \), and therefore

\[ -\Phi^0 < \sum_{j=1}^{n} w_j x_j^{1-\alpha} \left( \frac{1}{\alpha - 1} + \frac{1 - 4\varepsilon}{\kappa (1 - 2\gamma)} \right) < \sum_{j=1}^{n} w_j x_j^{1-\alpha} \left( \frac{1}{\alpha - 1} + 1 \right). \]

It follows that:

\[ \Phi^1 - \Phi^0 \geq \Omega \left( \frac{\beta}{(1+\beta)^{\alpha}} \right) \frac{\alpha - 1}{\alpha} (-\Phi^0). \]

Recalling that \( \beta = \frac{\gamma}{\alpha (\alpha + \alpha)} < \frac{\gamma}{4\alpha} \), we have that \( (1 + \beta)^{\alpha} \leq e^{\beta \alpha} < e^{\gamma/4} < 1.04 \) (as \( \gamma = \varepsilon / 4 \leq 1/24 \)). Therefore:

\[ \Phi^1 - \Phi^0 \geq \Omega (\beta \gamma) \frac{\alpha - 1}{\alpha} (-\Phi^0) > \Omega (\beta \gamma^2) \frac{\alpha - 1}{\alpha} (-\Phi^0). \quad (49) \]

Finally, (15), (48), and (49) yield the proof. \( \square \)

The following two technical propositions are used in Lemma 4.11 for bounding the duality gap in non-stationary rounds.

**Proposition D.9.** After at most \( n_{0} = \frac{1}{\beta} \min_{j} \ln \left( \frac{1}{\Delta_j} \right) \) rounds it is always true that \( G_{\alpha}(x, y(x)) \leq \sum_{j=1}^{n} \frac{w_j x_j^{1-\alpha}}{\alpha - 1} \left( 1 + (1 + 6\varepsilon)(\alpha - 1) \xi_j - \alpha \xi_j^{\alpha - 1} \right), \) where \( \xi_j = \frac{x_j^\alpha \sum_i y_i(x) A_{ij}}{w_j}. \)

**Proof.** Recall from (2) that the duality gap for \( x, y \) in (P_\alpha) is given as:

\[ G_{\alpha}(x, y) = \sum_{j=1}^{n} \frac{w_j x_j^{1-\alpha}}{\alpha - 1} \left( \frac{\sum_i y_i(x) A_{ij}}{x_j^\alpha} \right)^{\frac{\alpha - 1}{\alpha}} - 1 \right) + \sum_{i=1}^{m} y_i - \sum_{j=1}^{n} \frac{w_j x_j^{1-\alpha}}{\alpha - 1} \cdot \left( \frac{\sum_{j=1}^{m} A_{ij} y_i}{w_j} \right)^{\frac{\alpha - 1}{\alpha}}. \]
From Lemma 4.2, after at most \( \tau_0 = \frac{1}{\beta} \ln \left( \frac{1}{\max_j \delta_j} \right) \) rounds:

\[
\sum_{i=1}^{m} y_i \leq (1 + 6\varepsilon) \sum_{j=1}^{n} x_j \sum_{i=1}^{m} y_i A_{ij}
\]

\[
= (1 + 6\varepsilon) \sum_{j=1}^{n} w_j x_j^{1-\alpha} \left( \frac{x_j^{\alpha} \sum_{i=1}^{m} y_i A_{ij}}{w_j} \right),
\]

and letting \( \xi_j = \frac{x_j^{\alpha} \sum_{i=1}^{m} y_i A_{ij}}{w_j} \), we get:

\[
G_{\alpha}(x, y) \leq \sum_{j=1}^{n} w_j \frac{x_j^{1-\alpha}}{1-\alpha} \left( \xi_j^{1-\alpha} - 1 + (1 + 6\varepsilon)(1 - \alpha)\xi_j - (1 - \alpha)\xi_j^{1-\alpha} \right)
\]

\[
= \sum_{j=1}^{n} w_j \frac{x_j^{1-\alpha}}{1-\alpha} \left( \alpha \xi_j^{\alpha-1} + (1 + 6\varepsilon)(1 - \alpha)\xi_j - 1 \right)
\]

\[
= \sum_{j=1}^{n} w_j \frac{x_j^{1-\alpha}}{1-\alpha} \left( 1 + (1 + 6\varepsilon)(\alpha - 1)\xi_j - \alpha \xi_j^{1-\alpha} \right).
\]

\[\square\]

**Proposition D.10.** Let \( r_{\alpha}(\xi_j) = \left( 1 + (1 + 6\varepsilon)(\alpha - 1)\xi_j - \alpha \xi_j^{\alpha-1} \right) \), where \( \xi_j = \frac{x_j^{\alpha} \sum_{i=1}^{m} y_i A_{ij}}{w_j} \). If \( \alpha > 1 \) and \( \xi_j \in (1 - \gamma, 1 + \gamma) \) \( \forall j \in \{1, ..., n\} \), then \( r_{\alpha}(\xi_j) \leq \varepsilon(7\alpha - 6) \).

**Proof.** Observe the first and the second derivative of \( r_{\alpha}(\xi_j) \):

\[
\frac{dr_{\alpha}(\xi_j)}{d\xi_j} = (\alpha - 1)(1 + 6\varepsilon - \xi_j^{-1/\alpha});
\]

\[
\frac{d^2r_{\alpha}(\xi_j)}{d\xi_j^2} = \frac{1}{\alpha}(\alpha - 1)\xi_j^{-1/\alpha-1}.
\]

As \( \xi_j > 0 \), \( r(\xi_j) \) is convex for \( \alpha > 1 \), and therefore: \( r(\xi_j) \leq \max\{r(1 - \gamma), r(1 + \gamma)\} \). We have that:

\[
r(1 - \gamma) = r(1 - \varepsilon/4) = 1 - \left(1 - \frac{\varepsilon}{4} \right) \left( (1 - \alpha)(1 + 6\varepsilon) + \alpha(1 - \varepsilon/4)^{-1/\alpha} \right)
\]

\[
\leq 1 - \left(1 - \frac{\varepsilon}{4} \right) \left( 1 - \alpha + 6\varepsilon - 6\varepsilon \alpha + \alpha(1 + \varepsilon/4)^{1/\alpha} \right)
\]

\[
\leq 1 - \left(1 - \frac{\varepsilon}{4} \right) \left( 1 + 6\varepsilon(1 - \alpha) \right)
\]

\[
= 1 - 1 + 6\varepsilon(\alpha - 1) + \frac{\varepsilon}{4}(1 - 6\varepsilon(\alpha - 1))
\]

\[
= \frac{\varepsilon}{4} + 6\varepsilon(\alpha - 1) \left(1 - \frac{\varepsilon}{4} \right)
\]

\[
\leq 6\varepsilon(\alpha - 1).
\]
On the other hand:

\[
    r(1 + \gamma) = r(1 + \varepsilon/4) = 1 - \left(1 + \frac{\varepsilon}{4}\right) ((1 - \alpha)(1 + 6\varepsilon) + \alpha(1 + \varepsilon/4)^{-1/\alpha}) \\
    \leq 1 - \left(1 + \frac{\varepsilon}{4}\right) (1 - \alpha + 6\varepsilon - 6\varepsilon\alpha + \alpha(1 - \varepsilon/4)^{1/\alpha}) \\
    \leq 1 - \left(1 + \frac{\varepsilon}{4}\right) (1 + 6\varepsilon - \frac{25}{4}\varepsilon\alpha) \\
    \leq \frac{25}{4}\varepsilon\alpha - 6\varepsilon - \frac{\varepsilon}{4}(1 + 6\varepsilon - \frac{25}{4}\varepsilon\alpha) \\
    \leq \varepsilon(7\alpha - 6),
\]

completing the proof. \(\square\)

The following lemma states that in any stationary round current solution is an \((1 + \varepsilon(8\alpha - 7))\)-approximate solution.

**Lemma D.11.** In any stationary round that happens after first \(\tau_0 = \frac{1}{\beta} \min_j \ln \left(\frac{1}{\delta_j}\right)\) rounds, \(p_\alpha(x^*) - p_\alpha(x) \leq \varepsilon(8\alpha - 7)(-p_\alpha(x)),\) where \(x^*\) is the optimal solution to \((P_\alpha),\) \(x\) is the solution at the beginning of the round.

**Proof.** Observe that for any \(k \notin \{S^+ \cup S^-\}\) (by the definition of \(S^+\) and \(S^-\)) we have that \(1 - \gamma < x_k \sum_{i=1}^{m} y_i(x) A_{ik} < 1 + \gamma,\) which is equivalent to:

\[
    (1 - \gamma)w_kx_k^{1-\alpha} < x_k \sum_{i=1}^{m} y_i(x) A_{ik} < (1 + \gamma)w_kx_k^{1-\alpha} \quad \forall k \notin \{S^+ \cup S^-\}. \tag{50}
\]

Using stationarity and (50):

\[
    (1 - 2\gamma) \sum_{j=1}^{n} w_j x_j^{1-\alpha} \leq \sum_{j=1}^{n} x_j \sum_{i=1}^{n} y_i(x) A_{ij} \\
    = \sum_{l \in \{S^+ \cup S^-\}} x_l \sum_{i=1}^{n} y_i(x) A_{il} + \sum_{k \notin \{S^+ \cup S^-\}} x_k \sum_{i=1}^{n} y_i(x) A_{ik} \\
    < \gamma \sum_{j=1}^{n} w_j x_j^{1-\alpha} + (1 + \gamma) \sum_{k \notin \{S^+ \cup S^-\}} w_kx_k^{1-\alpha}. \tag{51}
\]

Since \(\sum_{l \in \{S^+ \cup S^-\}} w_l x_l^{1-\alpha} = \sum_{j=1}^{n} w_jx_j^{1-\alpha} - \sum_{k \notin \{S^+ \cup S^-\}} w_kx_k^{1-\alpha},\) using (51):

\[
    (1 - 2\gamma) \sum_{l \in \{S^+ \cup S^-\}} w_l x_l^{1-\alpha} < \sum_{j=1}^{n} w_j x_j^{1-\alpha} + (1 + \gamma) \sum_{k \notin \{S^+ \cup S^-\}} w_k x_k^{1-\alpha} - (1 - 2\gamma) \sum_{k \notin \{S^+ \cup S^-\}} w_k x_k^{1-\alpha} \\
    = \gamma \sum_{j=1}^{n} w_j x_j^{1-\alpha} + 3\gamma \sum_{k \notin \{S^+ \cup S^-\}} w_k x_k^{1-\alpha} \\
    \leq 4\gamma \sum_{j=1}^{n} w_j x_j^{1-\alpha},
\]

and therefore:

\[
    \sum_{l \in \{S^+ \cup S^-\}} w_l x_l^{1-\alpha} < \frac{4\gamma}{1 - 2\gamma} \sum_{j=1}^{n} w_j x_j^{1-\alpha} < 5\gamma \sum_{j=1}^{n} w_j x_j^{1-\alpha}, \tag{52}
\]
as \( \gamma = \frac{\varepsilon}{4} \) and \( \varepsilon \leq \frac{1}{6} \).

As \( p_\alpha(x^*) - p_\alpha(x) \leq G(x, y(x)) \), from Proposition D.9

\[
p_\alpha(x^*) - p_\alpha(x) \leq \sum_{j=1}^{n} w_j \frac{x_j^{1-\alpha}}{\alpha-1} \left( 1 + (1 + 6\varepsilon)(\alpha - 1)\xi_j - \alpha^{\frac{\alpha-1}{\alpha}} \right)
\]

\[
= \sum_{k \notin \{S^+ \cup S^-\}} w_k \frac{x_k^{1-\alpha}}{\alpha-1} \left( 1 + (1 + 6\varepsilon)(\alpha - 1)\xi_k - \alpha^{\frac{\alpha-1}{\alpha}} \right)
\]

\[
+ \sum_{l \notin \{S^+ \cup S^-\}} w_l \frac{x_l^{1-\alpha}}{\alpha-1} \left( 1 + (1 + 6\varepsilon)(\alpha - 1)\xi_l - \alpha^{\frac{\alpha-1}{\alpha}} \right).
\]

From Proposition D.10

\[
\sum_{k \notin \{S^+ \cup S^-\}} w_k \frac{x_k^{1-\alpha}}{\alpha-1} \left( 1 + (1 + 6\varepsilon)(\alpha - 1)\xi_k - \alpha^{\frac{\alpha-1}{\alpha}} \right) \leq \varepsilon (7\alpha - 6) \sum_{k \notin \{S^+ \cup S^-\}} w_k \frac{x_k^{1-\alpha}}{\alpha-1}
\]

\[
\leq \varepsilon (7\alpha - 6) \sum_{j=1}^{n} w_j \frac{x_j^{1-\alpha}}{\alpha-1}
\]

\[
= \varepsilon (7\alpha - 6) ( - p_\alpha(x) ). \quad (53)
\]

Observe \( \sum_{l \notin \{S^+ \cup S^-\}} w_l \frac{x_l^{1-\alpha}}{\alpha-1} \left( 1 + (1 + 6\varepsilon)(\alpha - 1)\xi_l - \alpha^{\frac{\alpha-1}{\alpha}} \right) \). Since \( \alpha > 1 \), each \( w_l \frac{x_l^{1-\alpha}}{\alpha-1} > 0 \), and therefore:

\[
\sum_{l \notin \{S^+ \cup S^-\}} w_l \frac{x_l^{1-\alpha}}{\alpha-1} \left( 1 + (1 + 6\varepsilon)(\alpha - 1)\xi_l - \alpha^{\frac{\alpha-1}{\alpha}} \right)
\]

\[
\leq \sum_{l \notin \{S^+ \cup S^-\}} w_l \frac{x_l^{1-\alpha}}{\alpha-1} \left( (1 + 6\varepsilon)(\alpha - 1)\xi_l + 1 \right)
\]

\[
= \sum_{l \notin \{S^+ \cup S^-\}} w_l \frac{x_l^{1-\alpha}}{\alpha-1} \left( (1 + 6\varepsilon)(\alpha - 1)\sum_{i=1}^{m} y_i(x)A_{il} + 1 \right)
\]

\[
= (1 + 6\varepsilon) \sum_{l \notin \{S^+ \cup S^-\}} x_l \sum_{i=1}^{m} y_i(x)A_{il} + \sum_{l \notin \{S^+ \cup S^-\}} w_l \frac{x_l^{1-\alpha}}{\alpha-1}.
\]

Now, from stationarity \( \sum_{l \notin \{S^+ \cup S^-\}} x_l \sum_{i=1}^{m} y_i(x)A_{il} < \gamma \sum_{j=1}^{n} w_j x_j^{1-\alpha} \) and using (52) we get:

\[
\sum_{l \notin \{S^+ \cup S^-\}} w_l \frac{x_l^{1-\alpha}}{\alpha-1} \left( 1 + (1 + 6\varepsilon)(\alpha - 1)\xi_l - \alpha^{\frac{\alpha-1}{\alpha}} \right) < \sum_{j=1}^{n} w_j \frac{x_j^{1-\alpha}}{\alpha-1} (\gamma (1 + 6\varepsilon)(\alpha - 1) + 5\gamma)
\]

\[
\leq -p_\alpha(x) (2\gamma (\alpha - 1) + 5\gamma)
\]

\[
= -p_\alpha(x) \left( \frac{\varepsilon}{2} - \frac{3\varepsilon}{4} \right). \quad (54)
\]

Finally, combining (53) and (54): 

\[
p_\alpha(x^*) - p_\alpha(x) < \varepsilon (8\alpha - 7)( - p_\alpha(x)) \quad \square
\]

**Proof of Theorem 4.8** First, we will bound the total change in the potential. Observe that since \( \alpha < 1 \) and \( x_j, y_i > 0 \), the potential is always negative. Recall that initially \( x_j = \delta_j \). Using similar
arguments as in the proof of convergence for $\alpha < 1$ and $\alpha = 1$, it is simple to show that $\sum_{i=1}^{m} y_i = o(1)$.

On the other hand:

$$
\sum_{j=1}^{n} w_j \frac{\delta_j^{1-\alpha}}{1-\alpha} = -\frac{1}{\alpha-1} \sum_{j=1}^{n} w_j \frac{1}{\delta_j^{1-\alpha}} \\
\geq -\frac{1}{\alpha-1} \sum_{j=1}^{n} w_j n^{2(\alpha-1)m^{\alpha-1}A_{\text{max}}^{2\alpha-1}} \\
\geq -\frac{W n^{2(\alpha-1)m^{\alpha-1}A_{\text{max}}^{2\alpha-1}}}{\alpha-1}.
$$

Therefore, the initial and minimum $\Phi$ is bounded as:

$$
\Phi_{\text{min}} \geq -O \left( \frac{W n^{2(\alpha-1)m^{\alpha-1}A_{\text{max}}^{2\alpha-1}}}{\alpha-1} \right). 
\tag{55}
$$

From Lemma 4.1, $x$ is always feasible, and therefore $x_j \leq 1 \forall j$, which implies:

$$
\Phi_{\text{max}} \leq -\frac{W}{\alpha - 1}. 
\tag{56}
$$

From Lemma D.8, after the first $\tau_0$ rounds in any non-stationary round the negative potential always decreases by a multiplicative factor: $-\Phi^1 \leq (1 - \Omega \left( \beta^2 \frac{\alpha-1}{\alpha} \right))(-\Phi^0)$. Using the bounds on the potential given in (55) and (56), we get that the total number of non-stationary rounds after the initial $\tau_0$ rounds is bounded by

$$
O \left( \log_{1-\Omega(\beta^2)\frac{\alpha-1}{\alpha}} \left( \frac{W}{W n^{2(\alpha-1)m^{\alpha-1}A_{\text{max}}^{2\alpha-1}}} \right) \right) = O \left( \frac{\alpha}{\alpha - 1} \beta^2 \ln \left( n^{2(\alpha-1)m^{\alpha-1}A_{\text{max}}^{2\alpha-1}} \right) \right) \\
= O \left( \frac{\ln(nmA_{\text{max}}) \ln(nmA_{\text{max}}/\varepsilon) + \varepsilon \alpha}{\varepsilon^4} \right),
$$

where we have used that $\beta = \frac{\gamma}{4(\kappa+\alpha)}$, $\kappa = \frac{1}{\varepsilon} \ln \left( \frac{CmA_{\text{max}}}{\min_j w_j \varepsilon} \right) = O \left( \frac{1}{\varepsilon} \ln \left( \frac{nmA_{\text{max}}}{\varepsilon} \right) \right)$.

Finally, since $\tau_0 = O \left( \frac{\ln(nmA_{\text{max}})(\ln(nmA_{\text{max}}/\varepsilon) + \alpha)}{\varepsilon^2} \right)$, the total number of non-stationary rounds is at most:

$$
O \left( \frac{\ln(nmA_{\text{max}})(\ln(nmA_{\text{max}}/\varepsilon) + \varepsilon \alpha)}{\varepsilon^4} \right) = \tilde{O} \left( \frac{\ln^2(nmA_{\text{max}})}{\varepsilon^5} \right),
$$

as $\varepsilon \alpha \leq \frac{9}{10}$. \qed

### E Stateless Implementation

Algorithm 1 is not stateless because it requires a proper initialization of each variable $x_j$ to the lower threshold $\delta_j$. We show that if we replace the initialization by a line that places $x_j$ in the interval $[\delta_j, 1]$, the algorithm can converge starting from any initial state. The pseudocode is provided in Algorithm 2.

It is easy to see that if the initial state is feasible, then all the results that were obtained for Algorithm 1 also apply to Algorithm 2. What is left to show is that if at initial state any constraint $i$ is infeasible, then $i$ becomes feasible after a finite number of rounds. We show this in the following lemma.
Algorithm 2 Stateless $\alpha$-Fair Resource Allocation

In each round of the algorithm:
1: if $x_j \in [\delta_j, 1]$ then
2: Update prices: $y_i = C \cdot e^{\alpha(\sum_{j=1}^{n} A_{ij} x_j - 1)} \forall i \in \{1, ..., m\}$
3: if $x_j \cdot \sum_{j=1}^{m} y_i A_{ij} \leq (1 - \gamma)$ then
4: $x_j = x_j \cdot (1 + \beta)$
5: else
6: if $x_j \cdot \sum_{j=1}^{m} y_i A_{ij} \geq (1 + \gamma)$ then
7: $x_j = \max\{x_j \cdot (1 - \beta), \delta_j\}$
8: else $x_j = \max\{x_j, \delta_j\}$, $x_j = \min\{x_j, 1\}$

Lemma E.1. If for any $i \sum_{j=1}^{n} A_{ij} x_j > 1$, then after at most $O(\frac{1}{\gamma} \ln(nA_{\text{max}}))$ rounds we have that it is always true that $\sum_{j=1}^{n} A_{ij} x_j \leq 1$.

Proof. Suppose that $\sum_{j=1}^{n} A_{ij} x_j > 1$ for some $i$. Then $y_i > C$, and for every $x_j$ with $A_{ij} \neq 0$:

$$x_j \sum_{i}^{m} y_i(x) A_{ij} \geq x_j^\alpha y_i(x) A_{ij} \geq A_{ij}^\alpha C \geq w_j (1 - \gamma),$$

and therefore none of the variables that appear in $i$ decreases.

Since $\sum_{j=1}^{n} A_{ij} x_j > 1$, there exists at least one $x_k$ with $A_{ik} \neq 0$ such that $x_k = \frac{\sum_{j=1}^{n} A_{ij} x_j}{A_{ik}} > \frac{1}{nA_{\text{max}}}$. For each such $x_k$, since $C \geq 2nA_{\text{max}}$:

$$x_k \sum_{i=1}^{m} y_i(x) A_{ij} \geq C \frac{1}{nA_{\text{max}}} \geq 2w_{\text{max}} > w_k (1 + \gamma),$$

and therefore $x_k$ decreases (by a factor $(1 - \beta)$). As $x_k \leq 1$, after at most $O(\frac{1}{\gamma} \ln(nA_{\text{max}}))$ rounds, we must have $x_k \leq \frac{1}{nA_{\text{max}}}$, and therefore $\sum_{j=1}^{n} A_{ij} x_j \leq 1$.

Another result that follows as a corollary is that the algorithm reaches a feasible solution after node or constraint insertion/deletion, as long as all the nodes can maintain a proper estimate of the upper bound on $n, m, A_{\text{max}}, w_{\text{max}}$. We remark that under minor modifications, the algorithm can run on any non-scaled input, provided that in addition all the nodes have available information about: $c = \min\{1, \min_{k,j:A_{ij} \neq 0} \frac{A_{ij}}{w_j}\}$. The pseudocode for the equivalent algorithm (see Appendix A) that can be run on the non-scaled input is given in Algorithm 3.

Algorithm 3 Non-scaled Stateless $\alpha$-Fair Resource Allocation

In each round of the algorithm:
1: if $x_j \in [\delta_j/c, 1/c]$ then
2: Update prices: $y_i = e^{\alpha - 1} \cdot C \cdot e^{\alpha(\sum_{j=1}^{n} A_{ij} x_j - 1)/b_i} \forall i \in \{1, ..., m\}$
3: if $x_j \cdot \sum_{j=1}^{m} y_i A_{ij} \leq (1 - \gamma)$ then
4: $x_j = x_j \cdot (1 + \beta)$
5: else
6: if $x_j \cdot \sum_{j=1}^{m} y_i A_{ij} \geq (1 + \gamma)$ then
7: $x_j = \max\{x_j \cdot (1 - \beta), \delta_j\}$
8: else $x_j = \max\{x_j, \delta_j/c\}$, $x_j = \min\{x_j, 1/c\}$