

# Stateless Distributed Gradient Descent for Positive Linear Programs

Baruch Awerbuch<sup>\*</sup>  
Johns Hopkins University  
baruch@cs.jhu.edu

Rohit Khandekar  
IBM T.J.Watson Research Center  
rkhandekar@gmail.com

## ABSTRACT

We develop a framework of *distributed* and *stateless* solutions for packing and covering linear programs, which are solved by multiple agents operating in a cooperative but uncoordinated manner. Our model has a separate “agent” controlling each variable and an agent is allowed to read-off the current values only of those constraints in which it has non-zero coefficients. This is a natural model for many distributed applications like *flow control*, *maximum bipartite matching*, and *dominating sets*.

The most appealing feature of our algorithms is their *simplicity* and *polylogarithmic convergence*. For the packing LP  $\max\{c \cdot x \mid Ax \leq b, x \geq 0\}$ , the algorithm associates a dual variable  $y_i = \exp[\frac{1}{\epsilon}(\frac{A_i x}{b_i} - 1)]$  for each constraint  $i$  and each agent  $j$  iteratively increases (resp. decreases)  $x_j$  *multiplicatively* if  $A_j^\top y$  is too small (resp. large) as compared to  $c_j$ . Our algorithm starting from a feasible solution, always maintains feasibility, and computes a  $(1 + \epsilon)$  approximation in  $\text{poly}(\frac{\ln(mn \cdot A_{\max})}{\epsilon})$  rounds. Here  $m$  and  $n$  are number of rows and columns of  $A$  and  $A_{\max}$ , also known as the “width” of the LP, is the ratio of maximum and minimum non-zero entries  $A_{ij}/(b_i c_j)$ . Similar algorithm works for the covering LP  $\min\{b \cdot y \mid A^\top y \geq c, y \geq 0\}$  as well.

While exponential dual variables are used in several packing/covering LP algorithms before [25, 9, 13, 12, 26, 16], this is the first algorithm which is both *stateless* and has *polylogarithmic* convergence. Our algorithms can be thought of as applying *distributed gradient descent/ascent* on a carefully chosen potential. Our analysis differs from those of previous multiplicative update based algorithms and argues that while the current solution is far away from optimality, the potential function decreases/increases by a significant factor.

<sup>\*</sup>Partially supported by NSF grants CCF 0515080, ANIR-0240551, CCR-0311795, and CNS-0617883.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

STOC’08, May 17–20, 2008, Victoria, British Columbia, Canada.  
Copyright 2008 ACM 978-1-60558-047-0/08/05 ...\$5.00.

## Categories and Subject Descriptors

F.2.1 [Analysis of Algorithms and Problem Complexity]: Numerical Algorithms and Problems—*Computations on matrices*; G.1.6 [Numerical Analysis]: Optimization—*Linear programming*

## General Terms

Algorithms, theory

## Keywords

Distributed and stateless algorithms, gradient descent, fast convergence, linear programming

## 1. INTRODUCTION

Achieving *global goal* using only *local interactions* has been a focus of a lot of research in the last two decades [20, 23, 1, 22, 9, 24, 18]. A typical distributed system is composed of loosely connected components taking decisions based on their “local views”. Due to lack of central coordination, it is often non-trivial to achieve any global objective in such a system.

In this paper, we consider such *distributed* and *stateless* solutions for packing and covering linear programming (LP) problems, which are solved by multiple agents operating in a cooperative but uncoordinated manner.

### 1.1 The model

**Distributed Packing.** Imagine trying to solve the following packing LP in a *distributed* and *dynamic* environment:

$$\max\{c \cdot x \mid Ax \leq b, x \geq 0\}. \quad (1)$$

Assume that  $c \in \mathbb{R}_+^n$ ,  $A \in \mathbb{R}_+^{m \times n}$ ,  $b \in \mathbb{R}_+^m$  have non-negative entries.

Analogous to [23, 21, 9, 18], we assume that there is an agent controlling variable  $x_j$ . The agent  $j$  is assumed to know (upper bounds on)  $m$  and  $n$ , the  $j$ th column of  $A$ , the value  $c_j$ , and is provided with the current values of “relative congestion”  $\frac{A_i x}{b_i}$  for the constraints on which agent  $j$  has non-zero *influence*, i.e.,  $i$  for which the  $(i, j)$ th entry  $A_{ij} \neq 0$ . Here  $A_i$  denotes the  $i$ th row of  $A$  and we can safely assume that  $b_i > 0$ . This is the *only* information known to agents. Note that the knowledge of these values does not disclose directly information about other agents. In particular, an agent  $j$  is oblivious to the existence of rows  $i$  for which  $A_{ij} = 0$ , or other agents and their  $x$  values.

Model		Problem [citation]	Convergence time
Stateless	Distributed		
no	no	packing/covering LP [25, 13, 12, 26, 16]	$m \cdot [\ln(m)/\epsilon]^{O(1)}$
no	yes	packing/covering LP [21, 9, 26, 17]	$[\ln(m \cdot A_{\max})/\epsilon]^{O(1)}, [\ln(mn)/\epsilon]^{O(1)}$
no	yes	multi-commodity flow routing [6, 5, 4]	$L \cdot [\ln(m \cdot c_{\max})/\epsilon]^{O(1)}$
yes	yes	multi-commodity flow control [14]	$c_{\max} \cdot [\ln(b_{\max})/\epsilon]^{O(1)}$
yes	yes	bipartite load balancing [2]	$[\ln(m)/\epsilon]^{O(1)}$
yes	yes	multi-commodity flow control [3]	$[\ln(mn \cdot c_{\max})/\epsilon]^{O(1)}$
yes	yes	packing/covering LP [this paper]	$[\ln(mn \cdot A_{\max})/\epsilon]^{O(1)}$

**Figure 1: A comparison of some combinatorial  $(1 + \epsilon)$ -approximation algorithms for multicommodity flows, bipartite load balancing, or packing/covering LPs. Here  $m$  and  $n$  denote the number of constraints and variables resp.,  $A_{\max}$  (resp.  $c_{\max}$  or  $b_{\max}$ ) is the ratio of maximum to minimum non-zero entries in  $A$  (resp.  $c$  or  $b$ ),  $L$  denotes the maximum-path-length.**

One instance of packing LP is distributed optimization of multi-commodity flow [9, 14, 3], or more precisely, flow control over fixed paths. In case of flow control, an agent associated with each flow-path has to decide how much flow is sent along that path, while a capacity constraint is enforced for each edge. Each agent  $j$  observes the relative congestion  $\frac{A_{ij}x}{b_i}$  only on the edges  $i$  on path  $j$ , and makes decision about modifying the flow on its path based on this partial information. Here  $A_{ij}$  denotes how much flow results on edge  $i$  due to unit flow on path  $j$  and  $b_i$  denotes the capacity of edge  $i$ .

**Distributed Covering.** We also consider solving the dual covering LP

$$\min\{b \cdot y \mid A^\top y \geq c, y \geq 0\} \quad (2)$$

in a similar model. An agent, in this case, controls a variable  $y_i$  and knows (upper bounds on)  $m$  and  $n$ , the  $i$ th column of  $A^\top$ , the value  $b_i$ , and is provided with the values of “relative coverage”  $\frac{A_{ij}^\top y}{c_j}$  for all constraints  $j$  for which  $A_{ij} \neq 0$ . Here  $A_j^\top$  denotes the  $j$ th row of  $A^\top$ .

One instance of covering LP is computing minimum-weight fractional *dominating set* [19, 18], which is the first step in obtaining randomized-rounding based distributed  $O(\ln \Delta)$  approximation for the minimum-weight integral dominating set problem (here  $\Delta$  is the maximum degree). In the natural covering LP formulation of this problem, we have a variable  $y_i$  for every vertex  $i$  (indicating the extent to which that vertex is picked in the solution) and a constraint for every vertex (indicating that the total coverage of that vertex is at least the required value). An agent for vertex  $i$  is aware only of its neighbors in the graph and observes the relative coverage  $\frac{A_{ij}^\top y}{c_j}$  of these neighbors  $j$ . The agent  $i$  then decides whether to increase/decrease  $y_i$  based only on this information. Here  $A_{ji}^\top = A_{ij}$  denotes how much coverage vertex  $i$  provides to its neighbor vertex  $j$  and  $c_j$  denotes the total coverage requirement of vertex  $j$ .

It is important to point out that, in our model, solving a primal problem does not necessarily lead to solving the dual problem. So separate solutions are needed for packing and covering LPs above.

**Stateless algorithms.** Similar to [5, 3], we desire that a distributed optimization is performed in a *stateless* manner.

That is, the decisions made by agents are not dependent on the past; they are only dependent on the current local state observable to the agents. This means, for example, that the agents make their decisions in a cooperative but uncoordinated manner, without having access to a global clock and without being able to properly initialize and synchronize their individual executions.

Statelessness is attractive since it implies a number of other important features, which are very desirable in distributed systems with unreliable components:

*Self-stabilization.* It is a classical and a very elegant notion in the theory of robust distributed systems [10, 15, 11, 8, 7]. It means that the solution can withstand adversarial but finite sequence of “hard reset” events, namely, crashes accompanied with loss of all memory contents, except of course the code of the program to be executed. Note that a self-stabilizing solution allows some of the agents to fall asleep for an undetermined period of time, and then to wake up. Also it means that the algorithm does not need to be initialized.

*Incremental and local adjustment.* Our dynamic model allows existing agents to “leave” and new agents to “join” over time. This corresponds to deleting/adding a variable from/to the LP. It also allows adding or deleting constraints. Even if such changes occur in the system, the algorithm does not need to be restarted. Rather, the algorithm adjusts the state (its variables) in a local and incremental manner, without disrupting the variables that are not affected.

*No global clock.* Algorithms are *not* driven by a global clock and can proceed asynchronously.

## 1.2 Existing work

A summary of existing work on designing combinatorial approximation schemes for packing and covering LPs or multi-commodity flow problems is given in Figure 1. The last row refers to the results in this paper.

**Statefull algorithms.** In centralized or distributed setting, efficient “primal-dual” algorithms in the statefull model have been widely studied in the past, both for general packing/covering LPs [25, 13, 21, 9, 17, 26, 16] and for network flows [13, 12, 26, 6, 4]. Most of these algorithms share features like exponential dual variables with our algorithms. However, these algorithms crucially depend on maintaining a state, e.g., proper initialization or some global information

about the current solution, and perform globally optimum updates in each round. Many of these algorithms initialize the variables/flows to zero. Thus they have to be restarted whenever the instance changes due to addition/deletion of variables/constraints; and do not satisfy the *incremental* and *local adjustment* property. These algorithms also increment only the “best” variables w.r.t. the current dual solution; and these variables are identified based on some global information. For example, in many multi-commodity flow algorithms [25, 13, 12, 26], flow is increased only along (approximately) shortest paths w.r.t. the current (dual) length function. The packing/covering LP algorithm of Plotkin-Shmoys-Tardos [25] or the multi-commodity flow algorithm of Awerbuch and Khandekar [5] fall short of being stateless since they have to keep track of the maximum violation in a constraint or the global maximum congestion of the current solution. The algorithm of [25] converges in time *linear* in the width<sup>1</sup>  $A_{\max}$  and that of [5] in time *linear* in the maximum path-length.

**Stateless algorithms.** Garg and Young [14] presented a stateless flow control algorithm. While their algorithm resembles ours in the case of flow control, the convergence time of their algorithm depends *linearly* on the ratio  $c_{\max}/c_{\min}$  of the maximum and minimum benefit of the flows. This linear dependence is inherent to their algorithm due to a severe limit on how much flow of a commodity is allowed to increase in a single round. Furthermore, their algorithm is based on the packet drop-rates at various routers/links and it is not clear how to generalize their techniques to arbitrary packing/covering LPs.

A stateless flow control algorithm with  $\text{polylog}(c_{\max}/c_{\min})$  convergence is given in a recent work of the authors [3]. The current paper greatly simplifies the proofs of [3], generalizes these techniques to arbitrary packing/covering LPs and *improves* the convergence time. Recently, Awerbuch, Azar, and Khandekar [2] presented a stateless algorithm for a special case of load balancing in bipartite graphs. Their algorithm and techniques, which do not use any exponential duals, do not appear to generalize to arbitrary LPs and hence new techniques are required.

## 2. OUR RESULTS AND TECHNIQUES

**Reduction to the normal form:**  $\max\{\mathbb{1} \cdot x \mid Ax \leq \mathbb{1}, x \geq 0\}$  and  $\min\{\mathbb{1} \cdot y \mid A^T y \geq \mathbb{1}, y \geq 0\}$ . Let  $\mathbb{1}_k$  denote the  $k$ -dimensional vector with all entries 1; we often omit the dimension from the notation when it is clear from the context. Consider the dual pair of packing and covering LPs (1) and (2). To simplify the description of our algorithms, we assume that  $b = \mathbb{1}$  and  $c = \mathbb{1}$ . For the packing LP, this can be done, without loss of generality, by replacing entry  $A_{ij}$  by  $\hat{A}_{ij} = A_{ij}/(b_i c_j)$  and working with variables  $\hat{x}_j = c_j x_j$ . Note that we can assume that  $b_i > 0$  for all  $i$  (otherwise all variables  $x_j$  with  $A_{ij} > 0$  must be set to zero) and  $c_j > 0$  for all  $j$  (otherwise  $x_j$  can be set to zero w.l.o.g.). We emphasize that our algorithm can be modified appropriately to work in the described model even without such scaling. This is true since the feedback  $\frac{A_{ij}x}{b_i}$  in the original (packing) LP is same as the feedback  $\hat{A}_{ij}\hat{x}$  in the new LP.

<sup>1</sup>In [25], the *width* of a polytope is defined slightly differently. However in our case, it becomes equal to  $A_{\max}$ .

Start with any solution  $x \in \mathbb{R}_+^n$  such that  $Ax \leq \mathbb{1}$ .

Repeat forever:

1. Let  $y_i = \exp\left[\mu \cdot \left(A_i x - 1\right)\right]$  for all  $i$ .
2. For  $1 \leq j \leq n$  do:
  - (a) If  $A_j^T y \leq 1 - \alpha$ , then  $x_j \leftarrow \max\{x_j(1 + \beta), \delta\}$ .
  - (b) If  $A_j^T y \geq 1 + \alpha$ , then  $x_j \leftarrow x_j(1 - \beta)$ .

**Figure 2: Algorithm for the packing LP:**  $\max\{\mathbb{1} \cdot x \mid Ax \leq \mathbb{1}, x \geq 0\}$ . The parameters  $\mu$ ,  $\alpha$ ,  $\beta$ , and  $\delta$  are given in Section 3.

For technical reasons, we also assume that the smallest non-zero entry in  $A$  is at least 1. Let  $A_{\max}$  denote (an upper bound on) the largest entry in  $A$ ; this is often referred to as the “width”<sup>1</sup> of the LP [25]. We assume that  $A_{\max}$  is known to all agents. A similar reduction holds for the covering LP.

**Main result.** Our algorithms for the packing and covering LPs are given in Figures 2 and 3 and our main results are Theorems 2.1 and 2.2. These algorithms are stateless, fast converging, and very simple to describe and implement. A round refers to one execution of the Repeat loop. The convergence time, measured in terms of number of rounds, is, strictly speaking, polynomial in the input  $\ln(A_{\max})$  and  $1/\epsilon$ . However, if  $A_{\max}$  is polynomially bounded in  $n$  and  $m$  (as is the case in many applications), the convergence time depends poly-logarithmically on  $n$  and  $m$  and polynomially on  $1/\epsilon$ .

Although the algorithms are described in a centralized setting, it is easy to verify that they fit into the model described in the previous section. In case of packing LP, for example, each agent  $j$  simply evaluates the variables  $y_i$  for  $i$  such that  $A_{ij} \neq 0$ , computes  $A_j^T y$ , and updates  $x_j$  as stated. The algorithms also work in asynchronous manner—the convergence time in that case is given by the number of rounds of the slowest agent.

The algorithms can also be made to work in the *dynamic* model as follows. For the packing LP, a new arriving variable is set to 0. If a new constraint arrives, all the variables in that constraint are set to 0. For the covering LP, on the other hand, a new arriving variable is set to 1. If a new constraint arrives, all the variables in that constraint are set to 1. It is easy to see that, this way, we always maintain feasibility on the arrival of variables/constraints. Theorems 2.1 and 2.2 state that once the system stabilizes, i.e., no new arrivals or departures, the near-optimality is achieved quickly.

**THEOREM 2.1.** *With parameters  $\mu, \alpha, \beta, \delta$  given in Section 3, the algorithm in Figure 2 starts from any feasible solution  $Ax_0 \leq \mathbb{1}$ , always maintains a feasible solution for (1), and satisfies that the number of rounds in which the solution  $x$  is not  $(1 + \epsilon)$ -approximate for (1) is*

$$\tilde{O}\left(\frac{\ln^2(m \cdot A_{\max}) \cdot \ln^2(n \cdot A_{\max})}{\epsilon^5}\right).$$

Here  $\tilde{O}$  hides lower order terms like  $\ln \ln(mn \cdot A_{\max})$  and  $\ln(1/\epsilon)$ .

Start with any solution  $y \in \mathbb{R}_+^m$  such that  $A^\top y \geq \mathbb{1}$ .

Repeat forever:

1. Let  $x_j = \exp\left[\mu \cdot \left(1 - A_j^\top y\right)\right]$  for all  $j$ .
2. For  $1 \leq i \leq m$  do:
  - (a) If  $A_i x \geq 1 + \alpha$ , then  $y_i \leftarrow \max\{y_i(1 + \beta), \delta\}$ .
  - (b) If  $A_i x \leq 1 - \alpha$ , then  $y_i \leftarrow y_i(1 - \beta)$ .

**Figure 3: Algorithm for the covering LP:**  $\min\{\mathbb{1} \cdot y \mid A^\top y \geq \mathbb{1}, y \geq 0\}$ . The parameters  $\mu, \alpha, \beta$ , and  $\delta$  are given in Section 4.

**THEOREM 2.2.** *With parameters  $\mu, \alpha, \beta, \delta$  given in Section 4, the algorithm in Figure 3 starts from any feasible solution  $A^\top y_0 \geq \mathbb{1}$ , always maintains a feasible solution for (2), and satisfies that the number of rounds in which the solution  $y$  is not  $(1 + \epsilon)$ -approximate for (2) is*

$$\tilde{O}\left(\frac{\ln^2(n \cdot A_{\max}) \cdot \ln^2(mn \cdot A_{\max} \cdot y_{\max})}{\epsilon^5}\right).$$

Here  $y_{\max} = \max_i(y_0)_i$ .

**Intuition.** Consider, for example, the algorithm for the packing LP given in Figure 2. With each constraint  $A_i x \leq 1$ , we associate a dual variable  $y_i$  that has an exponential dependence on how tightly that constraint is satisfied—tighter the constraint, higher the dual variable. The algorithm can be thought of as working toward satisfying the *dual feasibility*:  $A^\top y \geq \mathbb{1}$  and *complementary slackness* conditions:  $x_j > 0$  implies  $A_j^\top y = 1$ . Thus it increases  $x_j$  if  $A_j^\top y$  is too small as compared to 1 and decreases it if it is too large as compared to 1. If  $A_j^\top y$  is close to 1, then  $x_j$  is left unchanged. For fast convergence, it is critical that this increase and decrease be *multiplicative*. Since the multiplicative increase is ineffective when a variable is very small, we also allow a tiny *additive* increase by  $\delta$ .

Even though we do not prove convergence to a fixed point, it is easy to see that any fixed point  $(x^*, y^*)$  of this algorithm constitutes a pair of near-optimal primal and dual solutions. Since  $(1 - \alpha) \cdot \mathbb{1} < A^\top y^* < (1 + \alpha) \cdot \mathbb{1}$ , the solution  $y^*$  is approximately dual feasible. Moreover since every non-zero entry in  $A$  is at least 1, we get  $y_i^* < 1 + \alpha$  for all  $i$ . Therefore  $A_i x^* - 1 < \frac{\ln(1+\alpha)}{\mu}$  for all  $i$  and hence  $x^*$  is also approximately primal feasible. Furthermore, we have  $\mathbb{1} \cdot x^* > (1 - \alpha) \cdot y^{*\top} A x^* \geq (1 - \alpha)(1 - 4\epsilon) \cdot \mathbb{1} \cdot y^*$ , where the second inequality follows from the fact that  $y_i$  is a fast-growing function of  $A_i x$  (see Lemma 3.2 for the proof). Thus the primal objective value  $\mathbb{1} \cdot x^*$  is not much less than the dual objective value  $\mathbb{1} \cdot y^*$ , and hence  $x^*$  and  $y^*$  form near-optimal primal and dual solutions respectively.

**Distributed gradient descent/ascent.** It is not enough to study the properties of a fixed point without understanding whether/how fast the algorithm converges to the fixed point or any approximation. The main purpose of this paper is to analyze the convergence of the algorithm to a near-optimum solution. To this end, we introduce the following

potential function:

$$\Phi(x) = \mathbb{1} \cdot x - \frac{\mathbb{1} \cdot y}{\mu}.$$

The intuition behind this potential function lies in the observation that the partial derivative of  $\Phi$  w.r.t. a variable  $x_j$  is given as

$$\begin{aligned} \frac{\partial \Phi}{\partial x_j} &= 1 - \frac{1}{\mu} \sum_i \frac{\partial y_i(x)}{\partial x_j} \\ &= 1 - \frac{1}{\mu} \sum_i \mu \cdot A_{ij} \cdot y_i(x) \\ &= 1 - A_j^\top y(x). \end{aligned}$$

Note that the RHS is precisely the term on which the decision to increase/decrease  $x_j$  is based. Ignoring the second (and higher) order terms, we can then approximate<sup>2</sup> the increase  $\Delta \Phi$  in the potential in a single round by

$$\begin{aligned} \Delta \Phi &\approx \sum_j \frac{\partial \Phi}{\partial x_j} \cdot \Delta x_j \\ &= \sum_j (1 - A_j^\top y(x)) \cdot \Delta x_j \\ &\geq \Omega(\alpha) \cdot \sum_j |\Delta x_j| \end{aligned}$$

where  $\Delta x_j$  denotes the (possibly negative) increase in  $x_j$ . The above inequality follows from the fact that  $\Delta x_j > 0$  only if  $1 - A_j^\top y(x) \geq \alpha$  and  $\Delta x_j < 0$  only if  $1 - A_j^\top y(x) \leq -\alpha$ . Thus our algorithm can be thought of as *distributed gradient ascent* on  $\Phi$ .

Any “significant” change in the  $x$ -values thus leads to a “significant” increase in the potential. Since the potential is bounded, this cannot happen consistently. We thus conclude that there must be so-called “stationary” interval of at least *logarithmic* number of rounds in which the  $x$ -values (and hence  $y$ -values) do not change significantly. We then argue that the current solution is near-optimal as follows. If on the contrary the solution is *not* near-optimal, we prove that there must be a variable  $x_j$  (on which optimum outperforms the current solution) which satisfies  $A_j^\top y \leq 1 - \alpha$  consistently. However in this case, the variable  $x_j$  would increase multiplicatively over the logarithmic interval. This in turn contradicts the fact that the solution is always feasible. A similar line of argument has been used in [5, 2] in the context of multi-commodity flows and load balancing in bipartite graphs.

**Organization.** The rest of the paper is organized as follows. In Section 3, we prove Theorem 2.1—in Section 3.2, we prove that the potential increases whenever the “state” of the system changes “significantly” and in Section 3.3, we prove the convergence results. A similar proof for Theorem 2.2 is given in Section 4. Finally, we conclude with some open questions in Section 5.

### 3. PROOF OF THEOREM 2.1

Given  $x \in \mathbb{R}_+^n$ , let  $y(x) = (y_1(x), \dots, y_m(x))^\top \in \mathbb{R}_+^m$  to be the vector defined in step 1 of Figure 2,

$$y_i(x) = \exp\left[\mu \cdot \left(A_i x - 1\right)\right]. \quad (3)$$

<sup>2</sup>A rigorous proof is given in Section 3.

We chose the parameters  $\mu$ ,  $\alpha$ ,  $\beta$ , and  $\delta$  as follows. Let

$$\mu = \frac{1}{\epsilon} \cdot \ln \frac{mA_{\max}}{\epsilon} \quad \text{and} \quad \alpha = \frac{\epsilon}{4}$$

where  $\epsilon$  is the given error parameter. The parameters  $\beta$  and  $\delta$  are chosen so that, assuming that  $x$  satisfies  $Ax \leq \mathbb{1}$  in the beginning of a round, the maximum change (increase or decrease) in that round, in the value of  $A_j^\top y(x)$  for any  $j$  is at most by factor  $\frac{\alpha}{4}$ . Assume that the solution  $x$  in the beginning of a round is feasible:  $Ax \leq \mathbb{1}$ . Note that the maximum increase or decrease in  $A_i x$  in any single round is at most  $\beta + n \cdot A_{\max} \cdot \delta$ . Thus  $y_i(x)$  increases to at most  $y_i(x) \cdot \exp[\mu(\beta + n \cdot A_{\max} \cdot \delta)]$ . We ensure that this is at most  $y_i(x) \cdot (1 + \frac{\alpha}{4})$  by setting

$$\beta = \frac{\alpha}{10\mu} = \Theta\left(\frac{\epsilon}{\mu}\right), \quad \text{and} \quad (4)$$

$$\delta = \frac{\alpha}{10\mu \cdot n \cdot A_{\max}} = \Theta\left(\frac{\epsilon}{\mu \cdot n \cdot A_{\max}}\right). \quad (5)$$

Hence  $A_j^\top y(x)$  increases to at most  $A_j^\top y(x) \cdot (1 + \frac{\alpha}{4})$ . Similarly with the values of  $\beta$  and  $\delta$  as in (4) and (5), we can argue that the maximum decrease in  $A_j^\top y(x)$  is by a factor at most  $\frac{\alpha}{4}$ .

### 3.1 Preliminaries

LEMMA 3.1. *During the course of the algorithm, the solution  $x$  always remains feasible, i.e.,  $x \geq 0$  and  $Ax \leq \mathbb{1}$ .*

PROOF. We prove this by induction on the rounds. Since we start from a feasible solution, it is true in the beginning of the algorithm. Assume now that  $Ax \leq \mathbb{1}$  holds in the beginning of a round  $r$  and assume on the contrary that  $A_i x > 1$  for some  $i$  after round  $r$ . This is possible only due to an increase in some variable  $x_j$  such that  $A_{ij} > 0$ . Since  $x_j$  is increased in round  $r$ , we have  $A_j^\top y(x) \leq 1 - \alpha$  in the beginning of round  $r$ . Since  $A_j^\top y(x)$  increases in any single round by a factor of at most  $\frac{\alpha}{4}$ , we conclude that  $A_j^\top y(x) < 1$  after round  $r$ . Now recall that  $A_{ij} > 0$  in fact implies that  $A_{ij} \geq 1$ . Thus we have  $y_i(x) < 1$  after round  $r$ . This contradicts the fact that  $A_i x > 1$  after round  $r$  since, by definition,  $y_i(x) = \exp[\mu \cdot (A_i x - 1)]$ . Hence the proof is complete.  $\square$

The following lemma states that after poly-logarithmic number of rounds, at least one constraint is approximately tight, which is necessary for near-optimality. Recall that the complementary slackness conditions state that  $y_i > 0$  implies  $A_i x = 1$ , i.e.,  $y_i \cdot A_i x = y_i$ . The following lemma states that after poly-logarithmic number of rounds, these conditions hold in an approximate and aggregate sense.

LEMMA 3.2. *After first  $\tau_0 = O(\frac{1}{\beta} \ln \frac{1}{\delta})$  rounds, it is always true that*

- $\max_i A_i x \geq 1 - 2\epsilon$ , and
- $\mathbb{1} \cdot y \geq y^\top Ax \geq (1 - 4\epsilon) \cdot \mathbb{1} \cdot y$ .

PROOF. While  $\max_i A_i x < 1 - \epsilon$ , we have  $\max_i y_i < \exp[-\mu \cdot \epsilon] \leq \frac{\epsilon}{mA_{\max}}$ . Thus for all  $j$ , it holds that  $A_j^\top y(x) < \epsilon < 1 - \alpha$ . Thus all  $x_j$  increase by a factor of at least  $(1 + \beta)$ . Since the minimum non-zero entry in  $A$  is at least 1 and each  $x_j$  increases to at least  $\delta$  in a single round, in  $\tau_0$  rounds, we

get  $\max_i A_i x \geq 1 - \epsilon$ . For  $\max_i A_i x$  to decrease in any round, some  $x_j$  must decrease. Thus for some  $j$ , it must hold that  $A_j^\top y \geq 1 + \alpha$ , and hence it must hold that  $\max_i A_i x \geq 1 - \epsilon$  before the round. Since the maximum decrease in  $A_i x$  in any single round is at most  $\beta A_i x \leq \beta < \epsilon$ , it implies that  $\max_i A_i x \geq 1 - 2\epsilon$  always holds thereafter.

Since  $A_i x \leq 1$ , we always have  $\mathbb{1} \cdot y \geq y^\top Ax$ . Now fix a round  $t \geq \tau_0$  and let  $S = \{i \mid A_i x < \max_{i'} A_{i'} x - \epsilon\}$  in round  $t$ . For  $i \in S$ , we have  $y_i < \exp[-\mu \cdot \epsilon] \cdot \max_{i'} y_{i'} \leq \frac{\epsilon}{m} \cdot \max_{i'} y_{i'}$ . Therefore  $\sum_{i \in S} y_i < \epsilon \cdot \max_{i'} y_{i'}$  and hence  $\mathbb{1} \cdot y < (1 + \epsilon) \sum_{i \notin S} y_i$ . Thus we have

$$\begin{aligned} (1 - 2\epsilon) \cdot \mathbb{1} \cdot y &< (\max_i A_i x) \cdot (1 + \epsilon) \sum_{i \notin S} y_i \\ &\leq (1 + \epsilon) \sum_{i \notin S} y_i (A_i x + \epsilon) \\ &\leq (1 + \epsilon) \cdot y^\top Ax + \epsilon(1 + \epsilon) \cdot \mathbb{1} \cdot y. \end{aligned}$$

After simplifying this, we get the desired inequality.  $\square$

### 3.2 The potential

We work with the following potential:

$$\Phi(x) = \mathbb{1} \cdot x - \frac{\mathbb{1} \cdot y(x)}{\mu}. \quad (6)$$

The following fact is an elementary property of a differentiable convex function and approximates the potential by the linear term in Taylor series.

FACT 3.3. *For a differentiable convex function  $f : \mathbb{R}^k \rightarrow \mathbb{R}$ , for any  $x^0, x^1 \in \mathbb{R}^k$  we have*

$$f'(x^0) \cdot (x^1 - x^0) \leq f(x^1) - f(x^0) \leq f'(x^1) \cdot (x^1 - x^0)$$

where  $f'(x) = (\frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_k})^\top$  denotes the gradient evaluated at  $x$ .

PROOF. Using convexity, for  $\eta \in [0, 1]$ , we have  $f(x^0 + \eta(x^1 - x^0)) = f((1 - \eta)x^0 + \eta x^1) \leq (1 - \eta)f(x^0) + \eta f(x^1)$ . Now subtracting  $f(x^0)$  from both sides, then dividing by  $\eta$  on both sides, and then taking limit as  $\eta \rightarrow 0$ , we get the inequality on the left. The inequality on the right follows from exchanging the roles of  $x^0$  and  $x^1$ .  $\square$

LEMMA 3.4 (POTENTIAL INCREASE). *The potential  $\Phi$  does not decrease during the course of the algorithm. Moreover, if  $\Phi^0$  (resp.  $\Phi^1$ ),  $x^0$  (resp.  $x^1$ ), and  $y(x^0)$  (resp.  $y(x^1)$ ) denote the values of  $\Phi$ ,  $x$ , and  $y(x)$  before (resp. after) a round, then*

$$\Phi^1 - \Phi^0 \geq \Omega\left(\frac{\alpha}{\mu}\right) \cdot \sum_i |y_i(x^1) - y_i(x^0)|, \quad (7)$$

$$\Phi^1 - \Phi^0 \geq \Omega(\beta) \cdot \left[ (1 - \alpha) \cdot \mathbb{1} \cdot x^0 - y(x^0)^\top Ax^0 \right], \quad (8)$$

$$\Phi^1 - \Phi^0 \geq \Omega(\beta) \cdot \left[ y(x^0)^\top Ax^0 - (1 + \alpha) \cdot \mathbb{1} \cdot x^0 \right]. \quad (9)$$

PROOF. The negative potential  $-\Phi$  is a differentiable convex function of  $x$ . Using Fact 3.3, we conclude that

$$\begin{aligned} \Phi^1 - \Phi^0 &= -\Phi^0 - (-\Phi^1) \\ &\geq -\Phi'(x^1) \cdot (x^0 - x^1) \\ &= \sum_j (x_j^1 - x_j^0) \cdot (1 - A_j^\top y(x^1)). \quad (10) \end{aligned}$$

Let  $S^+ = \{j \mid A_j^\top y(x^0) \leq 1 - \alpha\}$  and  $S^- = \{j \mid A_j^\top y(x^0) \geq 1 + \alpha\}$ . Note that  $x_j^1 = x_j^0$  if  $j \notin S^+ \cup S^-$ .

**Case  $j \in S^+$ :** We have  $x_j^1 - x_j^0 > 0$ . Since  $A_j^\top y(x)$  does not change by a factor of more than  $\frac{\alpha}{4}$ , we conclude that  $A_j^\top y(x^1) \leq A_j^\top y(x^0) \cdot (1 + \frac{\alpha}{4}) \leq (1 - \alpha) \cdot (1 + \frac{\alpha}{4}) < 1 - \frac{3\alpha}{4}$ . Thus  $1 - A_j^\top y(x^1) \geq \frac{3\alpha}{4} = \Omega(\alpha)$ .

**Case  $j \in S^-$ :** We have  $x_j^1 - x_j^0 < 0$  and  $A_j^\top y(x^1) \geq A_j^\top y(x^0) \cdot (1 - \frac{\alpha}{4}) \geq 1 + \frac{\alpha}{2}$ . Thus  $1 - A_j^\top y(x^1) \leq -\frac{\alpha}{2} = -\Omega(\alpha)$ .

Now we prove inequality (7). Denote  $z = A_j^\top y(x^1) > 0$ . If  $j \in S^+$ , we have  $1 - z \geq \frac{3\alpha}{4}$  which implies  $1 - z \geq \frac{3\alpha}{4} \cdot z = \Omega(\alpha) \cdot z$ . If  $j \in S^-$ , we have  $1 - z \leq -\frac{\alpha}{2}$  which implies  $1 \leq z(1 + \frac{\alpha}{2})^{-1}$  which, in turn, implies that  $1 - z \leq z \cdot [(1 + \frac{\alpha}{2})^{-1} - 1] = -\frac{\alpha}{2} \cdot z \cdot (1 + \frac{\alpha}{2})^{-1} = -\Omega(\alpha) \cdot z$ . Thus the expression in (10) is at least

$$\Omega(\alpha) \cdot \sum_j |x_j^1 - x_j^0| \cdot A_j^\top y(x^1). \quad (11)$$

Now using Fact 3.3 for the convex function  $y_i(x)$  we conclude

$$\begin{aligned} y_i(x^0) \cdot \mu \cdot (A_i x^1 - A_i x^0) &\leq y_i(x^1) - y_i(x^0) \\ &\leq y_i(x^1) \cdot \mu \cdot (A_i x^1 - A_i x^0). \end{aligned}$$

Therefore

$$\begin{aligned} &\sum_i |y_i(x^1) - y_i(x^0)| \\ &\leq \sum_i \max\{y_i(x^0), y_i(x^1)\} \cdot \mu \cdot |A_i x^1 - A_i x^0| \\ &\leq \sum_i \max\{y_i(x^0), y_i(x^1)\} \cdot \mu \cdot \sum_j A_{ij} |x_j^1 - x_j^0| \\ &= O(\mu) \cdot \sum_j |x_j^1 - x_j^0| \cdot A_j^\top y(x^1). \end{aligned} \quad (12)$$

The equation (12) follows from the fact that  $y_i(x^0) \leq (1 - \frac{\alpha}{4})^{-1} \cdot y_i(x^1)$  and by interchanging the order of summations over  $i$  and  $j$ . From equations (11) and (12), we conclude that  $\Phi^1 - \Phi^0 \geq \Omega(\frac{\alpha}{\mu}) \cdot \sum_i |y_i(x^1) - y_i(x^0)|$ . Thus the proof of (7) is complete. Now we prove (8). From (10), we know

$$\begin{aligned} \Phi^1 - \Phi^0 &\geq \sum_j (x_j^1 - x_j^0) \cdot (1 - A_j^\top y(x^1)) \\ &\geq \sum_{j \in S^+} (x_j^1 - x_j^0) \cdot (1 - A_j^\top y(x^1)). \end{aligned}$$

Since  $A_j^\top y$  changes by a factor of at most  $\alpha/4$  in a single round, the RHS above is at least

$$\begin{aligned} &\sum_{j \in S^+} (x_j^1 - x_j^0) \cdot (1 - (1 + \frac{\alpha}{4}) \cdot A_j^\top y(x^0)) \\ &\geq \sum_{j \in S^+} \Omega(\beta) \cdot x_j^0 \cdot ((1 - \alpha) - A_j^\top y(x^0)) \end{aligned} \quad (13)$$

$$\begin{aligned} &\geq \sum_j \Omega(\beta) \cdot x_j^0 \cdot ((1 - \alpha) - A_j^\top y(x^0)) \\ &= \Omega(\beta) \left[ (1 - \alpha) \cdot \mathbb{1} \cdot x^0 - y(x^0)^\top A x^0 \right]. \end{aligned} \quad (14)$$

The inequality (13) follows since  $x_j^0$  (for  $j \in S^+$ ) increases by a factor of at least  $\beta$ ; inequality (14) follows from the definition of  $S^+$ . The proof of (9) is similar and is omitted.  $\square$

### 3.3 Convergence of the dynamics

Let  $x(t)$ ,  $y(t)$ , and  $\Phi(t)$  denote the values of  $x$ ,  $y(x)$ , and  $\Phi(x)$  respectively at the beginning of round  $t$ . In light of the potential increase lemma 3.4, we consider the following definition.

**DEFINITION 3.5 (STATIONARY INTERVAL).** Consider  $t_0 \geq \tau_0$ . We call an interval  $\mathcal{T} = [t_0, t_1]$  of rounds stationary if

- For all  $t \in \mathcal{T}$ , we have  $(1 - 2\alpha) \cdot \mathbb{1} \cdot x(t) \leq y(t)^\top A x(t) \leq (1 + 2\alpha) \cdot \mathbb{1} \cdot x(t)$ , and
- For all  $t \in \mathcal{T}$ , we have  $\sum_{t' \in \mathcal{T}} \sum_i |\Delta y_i(t')| \leq \frac{\alpha}{2} \cdot \mathbb{1} \cdot x(t)$ , where  $\Delta y_i(t') = y_i(t' + 1) - y_i(t')$ .

We call an interval unstationary if it is not stationary.

Now we prove two of our main lemmas. Lemma 3.6 states that the potential  $\Phi$  increases significantly in any unstationary interval (and hence there cannot be too many disjoint unstationary intervals) while Lemma 3.7 states that a stationary interval of logarithmic length implies near-optimality throughout the interval.

**LEMMA 3.6.** During any unstationary interval  $\mathcal{T} = [t_0, t_1]$ , the potential increases by at least  $\Omega(\frac{\alpha^2}{\mu}) \cdot \mathbb{1} \cdot x(t)$  for some  $t \in \mathcal{T}$ .

**PROOF.** By definition, the interval  $\mathcal{T}$  is unstationary because either  $y(t)^\top A x(t) < (1 - 2\alpha) \cdot \mathbb{1} \cdot x(t)$ , or  $(1 + 2\alpha) \cdot \mathbb{1} \cdot x(t) < y(t)^\top A x(t)$ , or  $\sum_{t' \in \mathcal{T}} \sum_i |\Delta y_i(t')| > \frac{\alpha}{2} \cdot \mathbb{1} \cdot x(t)$  for some  $t \in \mathcal{T}$ . In any case, from Lemma 3.4, we get that the potential increases by either  $\Omega(\beta) \cdot \alpha \cdot \mathbb{1} \cdot x(t) = \Omega(\frac{\alpha^2}{\mu}) \cdot \mathbb{1} \cdot x(t)$  or  $\Omega(\frac{\alpha}{\mu}) \cdot \frac{\alpha}{2} \cdot \mathbb{1} \cdot x(t) = \Omega(\frac{\alpha^2}{\mu}) \cdot \mathbb{1} \cdot x(t)$ .  $\square$

**LEMMA 3.7.** Suppose there is a stationary interval  $\mathcal{T} = [t_0, t_1]$  such that  $t_0 \geq \tau_0$  and that has length  $t_1 - t_0 \geq \tau_0$  where  $\tau_0 = O(\frac{1}{\beta} \ln \frac{1}{\delta})$ . Then for all rounds  $\hat{t} \in \mathcal{T}$ , the solution  $x(\hat{t})$  is an  $(1 + 6\epsilon)$ -approximate solution to the packing LP (1).

**PROOF.** Fix a round  $\hat{t} \in \mathcal{T}$  and let  $x_0$  and  $y_0$  denote  $x(\hat{t})$  and  $y(\hat{t})$  respectively. Let  $x^*$  be the optimum solution to the packing LP (1). Let  $\gamma = \mathbb{1} \cdot x^* / \mathbb{1} \cdot x_0$  be the approximation factor of the solution  $x_0$ . Assume, on the contrary, that  $\gamma > 1 + 6\epsilon$ . We have

$$\begin{aligned} (1 + 2\alpha) \cdot \mathbb{1} \cdot x^* &= (1 + 2\alpha) \cdot \gamma \cdot \mathbb{1} \cdot x_0 \\ &\geq \gamma \cdot y_0^\top A x_0 \quad (\text{since } \mathcal{T} \text{ is stationary}) \\ &\geq \gamma \cdot (1 - 4\epsilon) \cdot \mathbb{1} \cdot y_0 \quad (\text{from Lemma 3.2}) \\ &\geq \frac{1}{1 - \epsilon} \cdot \mathbb{1} \cdot y_0 \quad (\text{since } \gamma > 1 + 6\epsilon) \\ &\geq \frac{1}{1 - 4\alpha} \cdot y_0^\top A x^* \quad (\text{since } A x^* \leq \mathbb{1}) \end{aligned}$$

Thus

$$\frac{\sum_j (y_0^\top A)_j \cdot x_j^*}{\sum_j x_j^*} \leq 1 - 2\alpha. \quad (15)$$

Furthermore, from the definition of stationary interval, we

have  $\sum_{t \in \mathcal{T}} \sum_i |\Delta y_i(t)| \leq \frac{\alpha}{2} \cdot \mathbb{1} \cdot x_0 < \frac{\alpha}{2} \cdot \mathbb{1} \cdot x^*$ . Thus

$$\begin{aligned} & \sum_j x_j^* \cdot \sum_{t \in \mathcal{T}} |(y(t+1)^\top A)_j - (y(t)^\top A)_j| \\ & \leq \sum_j x_j^* \cdot \sum_{t \in \mathcal{T}} \sum_i A_{ij} \cdot |y_i(t+1) - y_i(t)| \quad (16) \\ & = \sum_{t \in \mathcal{T}} \sum_i |y_i(t+1) - y_i(t)| \cdot A_i x^* \\ & \leq \sum_{t \in \mathcal{T}} \sum_i |y_i(t+1) - y_i(t)| \\ & < \frac{\alpha}{2} \cdot \mathbb{1} \cdot x^*. \quad (17) \end{aligned}$$

The inequality (16) follows from the ‘‘triangle inequality’’  $|\sum_i a_i - \sum_i a'_i| \leq \sum_i |a_i - a'_i|$  for real numbers  $a_i, a'_i \in \mathfrak{R}$ .

From inequalities (15) and (17), we conclude that

$$\begin{aligned} & \frac{1}{\sum_j x_j^*} \sum_j x_j^* \cdot \left[ (y_0^\top A)_j + \sum_{t \in \mathcal{T}} |(y(t+1)^\top A)_j - (y(t)^\top A)_j| \right] \\ & < 1 - 2\alpha + \frac{\alpha}{2} < 1 - \alpha. \end{aligned}$$

Now think of the LHS above as the (weighted) average of the terms in square brackets. Since the average is less than  $1 - \alpha$ , there exists  $j$  such that

$$\left[ (y_0^\top A)_j + \sum_{t \in \mathcal{T}} |(y(t+1)^\top A)_j - (y(t)^\top A)_j| \right] < 1 - \alpha.$$

Since the first term in the square bracket above is the value of  $A_j^\top y$  at round  $\hat{t}$  and the second term is the absolute change in it throughout the interval, we conclude that  $A_j^\top y(t) < 1 - \alpha$  for all  $t \in \mathcal{T}$ . Thus the variable  $x_j$  in the algorithm must be increased by a factor of at least  $1 + \beta$  in each of the  $\tau_0 = \Theta(\frac{1}{\beta} \ln \frac{1}{\delta})$  rounds in  $\mathcal{T}$ . Note that  $x_j$  increases to at least  $\delta$  in a single round and therefore would become larger than 1 in  $\tau_0$  rounds, which when combined with the fact that every non-zero entry in  $A$  is at least 1, contradicts the fact that  $x(t)$  always remains feasible for  $Ax \leq \mathbb{1}$ . Thus our assumption that  $\gamma > 1 + 6\epsilon$  must be incorrect; this completes the proof.  $\square$

We are now ready to prove our main theorem.

**THEOREM 3.8.** *After  $O(\frac{\mu}{\alpha^2} \cdot \frac{1}{\beta} \ln \frac{1}{\delta} \cdot \ln(n \cdot A_{\max})) = \tilde{O}(\frac{1}{\epsilon^5} \ln^2(m \cdot A_{\max}) \cdot \ln^2(n \cdot A_{\max}))$  rounds, there must exist a round  $t$  such that  $x(t)$  forms an  $(1 + 6\epsilon)$ -approximate solution to the packing LP (1). Moreover the total number of rounds in which the solution is not  $(1 + 6\epsilon)$ -approximate is also at most the above bound.*

**PROOF.** The proof is based on Lemmas 3.6 and 3.7 and the fact that the potential is always between  $-m/\mu$  and  $n$ .

After first  $\tau_0$  rounds, the conditions in Lemma 3.2 hold. Consider now the duration in which  $\Phi < 0$ . Thus we have  $y^\top Ax \geq (1 - 4\epsilon) \cdot \mathbb{1} \cdot y > (1 - 4\epsilon) \cdot \mu \cdot \mathbb{1} \cdot x$ . From Lemma 3.4-Eqn. (9), we have that the potential in each round increases by  $\Omega(\beta) \cdot \mathbb{1} \cdot y$ . Thus the negative potential  $-\Phi$  decreases by a factor of  $\Omega(\beta\mu)$ . Note also that Lemma 3.2 states that  $\max_i A_i x \geq 1 - 2\epsilon$ , which implies that  $\mathbb{1} \cdot x \geq (1 - 2\epsilon)/A_{\max}$ . Thus in  $O(\frac{1}{\beta\mu} \ln \frac{m \cdot A_{\max}}{\mu})$  rounds, the potential must become non-negative. A similar argument in fact shows that the potential becomes  $\Omega(1/A_{\max})$  in  $O(\frac{1}{\beta\mu} \ln \frac{m \cdot A_{\max}}{\mu})$  rounds.

Now consider the duration in which  $\Phi = \Omega(1/A_{\max})$ . Partition the rounds into contiguous intervals of length  $\tau_0$  each. If an interval is stationary, from Lemma 3.7, we have near-optimality throughout the interval. If an interval  $\mathcal{T} = [t_0, t_1]$  is unstationary, from Lemma 3.6, for some  $t \in \mathcal{T}$ , the potential  $\Phi$  increases by  $\Omega(\frac{\alpha^2}{\mu} \cdot \mathbb{1} \cdot x(t)) \geq \Omega(\frac{\alpha^2}{\mu} \cdot \Phi(t)) \geq \Omega(\frac{\alpha^2}{\mu} \cdot \Phi(t_0))$  in this interval. Thus after encountering  $O(\frac{\mu}{\alpha^2} \cdot \ln(n \cdot A_{\max}))$  unstationary intervals, the potential becomes more than  $n$ . This contradicts the fact that  $x$  always remains feasible and hence  $\Phi \leq \mathbb{1} \cdot x \leq n$ . Thus the total number of unstationary intervals is at most  $O(\frac{\mu}{\alpha^2} \cdot \ln(n \cdot A_{\max}))$ . Hence the proof is complete.  $\square$

## 4. PROOF OF THEOREM 2.2

Given  $y \in \mathfrak{R}_+^m$ , let  $x(y) = (x_1(y), \dots, x_n(y))^\top \in \mathfrak{R}_+^n$  to be the vector defined in step 1 of Figure 3,

$$x_j = \exp \left[ \mu \cdot \left( 1 - A_j^\top y \right) \right]. \quad (18)$$

Now we let

$$\mu = \frac{1}{\epsilon} \ln \frac{n \cdot A_{\max}}{\epsilon} \quad \text{and} \quad \alpha = \frac{\epsilon}{4}.$$

We set

$$\begin{aligned} \beta &= \frac{\alpha}{20\mu} = \Theta \left( \frac{\epsilon}{\mu} \right), \quad \text{and} \\ \delta &= \frac{\alpha}{20\mu \cdot m \cdot A_{\max}} = \Theta \left( \frac{\epsilon}{\mu \cdot m \cdot A_{\max}} \right). \end{aligned}$$

With these values, we ensure that if  $A_j^\top y \leq 3$  in the beginning of a round, then in this round  $A_j^\top y$  changes by at most  $\frac{\alpha}{5\mu}$  and  $x_j$  changes by a factor of at most  $\exp[\frac{\alpha}{5}] \leq 1 + \frac{\alpha}{4}$ .

### 4.1 Preliminaries

**LEMMA 4.1.** *During the course of the algorithm, the solution  $y$  always remains feasible, i.e.,  $y \geq 0$  and  $A^\top y \geq \mathbb{1}$ .*

**PROOF.** We prove this by induction on the rounds. The solution  $y$  is feasible in the beginning of the algorithm. Assume, on the contrary, that  $A^\top y \geq \mathbb{1}$  holds in the beginning of round  $r$  and  $A_j^\top y < 1$  holds for some  $j$  after round  $r$ . This is possible only due to a decrease in some variable  $y_i$  such that  $A_{ij} > 0$ . Since the variables  $y_i$  decrease by a factor of at most  $1 - \beta$ , we have that  $A_j^\top y < \frac{1}{1 - \beta}$  must hold in the beginning of round  $r$ . Thus  $x_j > \exp[\mu \cdot \frac{-\beta}{1 - \beta}] > 1 - \frac{\mu\beta}{1 - \beta} \geq 1 - \alpha$ . Since  $A_{ij} \geq 1$ , we get that  $A_i x > 1 - \alpha$  holds in the beginning of round  $r$ . This implies, from the algorithm, that the variable  $y_i$  cannot decrease, in contradiction to our assumption. Thus the proof is complete.  $\square$

Analogous to Lemma 3.2, the following lemma states that after poly-logarithmic number of rounds, at least one constraint is approximately tight and the complementary slackness conditions:  $x_j > 0$  implies  $A_j^\top y = 1$  hold in an approximate and *aggregate* sense.

**LEMMA 4.2.** *Let  $y_{\max}$  be the maximum value of any  $y_i$  in round 0. After first  $\tau_1 = O(\frac{1}{\beta} \ln \frac{y_{\max}}{\delta})$  rounds, it is always true that*

$$\bullet \min_j A_j^\top y \leq 1 + \frac{3}{2}\epsilon, \quad \text{and}$$

$$\bullet \mathbb{1} \cdot x \leq x^\top A^\top y \leq (1 + 4\epsilon) \cdot \mathbb{1} \cdot x.$$

PROOF. While  $\min_j A_j^\top y > 1 + \epsilon$ , we have  $\max_j x_j < \exp[-\mu \cdot \epsilon] \leq \frac{\epsilon}{n A_{\max}}$ . Thus for all  $i$ , it holds that  $A_i x(y) < \epsilon < 1 - \alpha$ . Thus all  $y_i$  decrease by a factor of  $(1 - \beta)$ . Since the minimum non-zero entry in  $A$  is at least 1, in  $O(\frac{1}{\beta} \ln(m \cdot A_{\max} \cdot y_{\max})) \leq O(\frac{1}{\beta} \ln \frac{y_{\max}}{\delta})$  rounds, we get  $\min_j A_j^\top y \leq 1 + \epsilon$ .

For  $\min_j A_j^\top y$  to increase in any round, some  $y_i$  must increase. Thus for some  $i$ , it must hold that  $A_i x \geq 1 + \alpha$ , and hence  $\min_j A_j^\top y \leq 1 + \epsilon$  holds before that round. From the values of  $\beta$  and  $\delta$ , the maximum increase in  $A_j^\top y \leq 1 + \epsilon$  in a single round is at most  $\frac{\alpha}{5\mu} < \frac{\epsilon}{2}$ , it implies that  $\min_j A_j^\top y \leq 1 + \frac{3}{2}\epsilon$  always holds thereafter.

Since  $A_j^\top y \geq 1$ , we always have  $\mathbb{1} \cdot x \leq x^\top A^\top y$ . Now fix a round  $t \geq \tau_1$  and let  $j_0 = j$  be the index such that  $A_j^\top y$  is minimum and let  $S = \{j \mid A_j^\top y > A_{j_0}^\top y + \epsilon\}$  in round  $t$ . For  $j \in S$ , we have  $x_j < \exp[-\mu \cdot \epsilon] \cdot x_{j_0} \leq \frac{\epsilon}{n} \cdot x_{j_0}$ . Therefore  $\sum_{j \in S} x_j < \epsilon \cdot x_{j_0}$  and hence  $\mathbb{1} \cdot x < (1 + \epsilon) \sum_{j \notin S} x_j$ . Thus we have

$$\begin{aligned} \left(1 + \frac{3}{2}\epsilon\right) \cdot \mathbb{1} \cdot x &\geq (\min_j A_j^\top y) \cdot \sum_{j \notin S} x_j \\ &\geq \sum_{j \notin S} x_j (A_j^\top y - \epsilon) \\ &\geq \sum_{j \notin S} x_j A_j^\top y - \epsilon \cdot \mathbb{1} \cdot x \\ &\geq (1 - \epsilon(1 + \epsilon)) \cdot x^\top A^\top y - \epsilon \cdot \mathbb{1} \cdot x. \end{aligned}$$

The last inequality follows from the fact that for each  $j \in S$ , the definition of  $x_j$  implies that  $x_j A_j^\top y < \frac{\epsilon(1+\epsilon)}{n} \cdot x_{j_0} A_{j_0}^\top y$ ; hence  $x^\top A^\top y < (1 + \epsilon(1 + \epsilon)) \cdot \sum_{j \notin S} x_j A_j^\top y$ . After simplifying the above expression, we get the desired inequality.  $\square$

## 4.2 The potential

We work with the following potential:

$$\Psi(y) = \mathbb{1} \cdot y + \frac{\mathbb{1} \cdot x(y)}{\mu}. \quad (19)$$

Note that  $\Psi(y)$  is a *convex* function of  $y$  and that  $\frac{\partial \Psi}{\partial y_i} = 1 - A_i x$  holds. We argue that the potential  $\Phi$  does not increase after first  $\tau_1$  rounds.

LEMMA 4.3 (POTENTIAL DECREASE). *The potential  $\Psi$  does not increase after first  $\tau_1$  rounds. Moreover the following holds after first  $\tau_1$  rounds. If  $\Psi^0$  (resp.  $\Psi^1$ ),  $y^0$  (resp.  $y^1$ ), and  $x(y^0)$  (resp.  $x(y^1)$ ) denote the values of  $\Psi$ ,  $y$ , and  $x(y)$  before (resp. after) a round, then*

$$\Psi^0 - \Psi^1 \geq \Omega\left(\frac{\alpha}{\mu}\right) \cdot \sum_j |x_j(y^1) - x_j(y^0)| \quad (20)$$

$$\Psi^0 - \Psi^1 \geq \Omega(\beta) \cdot \left[ (1 - \alpha) \cdot \mathbb{1} \cdot y^0 - x(y^0)^\top A^\top y^0 \right], \quad (21)$$

$$\Psi^0 - \Psi^1 \geq \Omega(\beta) \cdot \left[ x(y^0)^\top A^\top y^0 - (1 + \alpha) \cdot \mathbb{1} \cdot y^0 \right]. \quad (22)$$

PROOF. The potential  $\Psi$  is a differentiable convex function of  $y$ . Using Fact 3.3, we conclude that

$$\begin{aligned} \Psi^0 - \Psi^1 &\geq \Psi'(y^1) \cdot (y^0 - y^1) \\ &= \sum_i (y_i^0 - y_i^1) \cdot (1 - A_i x(y^1)). \quad (23) \end{aligned}$$

Let  $S^+ = \{i \mid A_i x(y^0) \geq 1 + \alpha\}$  and  $S^- = \{i \mid A_i x(y^0) \leq 1 - \alpha\}$ . Note that  $y_i^1 = y_i^0$  if  $i \notin S^+ \cup S^-$ .

**Case  $i \in S^-$ :** We have  $y_i^0 - y_i^1 > 0$ . We know that  $1 - A_i x(y^0) \geq \alpha$ . We now argue that  $1 - A_i x(y^1) \geq \Omega(\alpha)$ . If  $A_j^\top y^0 \leq 3$ , then from the way  $\beta$  and  $\delta$  is defined,  $x_j$  changes by at most factor of  $1 + \frac{\alpha}{4}$ . If on the other hand,  $A_j^\top y^0 > 3$ , then  $A_j^\top y^1 > 3(1 - \beta)$  and hence  $x_j(y^1)$  is smaller than  $\exp[-\mu(3(1 - \beta) - 1)]$ . Recall from Lemma 4.2 that  $\min_{j'} A_{j'}^\top y \leq 1 + \frac{3}{2}\epsilon$  and hence  $\max_{j'} x_{j'} \geq \exp[\mu \cdot \frac{3}{2}\epsilon]$ . Therefore the contribution of  $x_j(y^1)$  for which  $A_j^\top y^0 > 3$  to  $A_i x(y^1)$  is at most  $\frac{\alpha}{10}$  fraction. Thus  $A_i x(y)$  changes by at most  $\frac{\alpha}{2}$  factor in this round and hence  $1 - A_i x(y^1) \geq \Omega(\alpha)$ .

**Case  $i \in S^+$ :** We have  $y_i^0 - y_i^1 < 0$ . We know that  $1 - A_i x(y^0) \leq -\alpha$ . By a similar argument as above, we argue that  $1 - A_i x(y^1) \leq -\Omega(\alpha)$ .

Now we prove inequality (20). Denote  $z = A_i x(y^1) > 0$ . If  $i \in S^-$ , we have  $1 - z \geq \Omega(\alpha)$  which implies  $1 - z \geq \Omega(\alpha) \cdot z$ . If  $i \in S^+$ , we have  $1 - z \leq -\Omega(\alpha)$  which implies  $1 - z \leq -\Omega(\alpha) \cdot z$ . Similar expressions also hold with  $z = A_i x(y^0)$ . Thus the expression in (23) is at least

$$\Omega(\alpha) \cdot \sum_i |y_i^0 - y_i^1| \cdot (A_i x(y^0) + A_i x(y^1)). \quad (24)$$

Now using Fact 3.3 for the convex function  $x_j(y)$  we conclude

$$\begin{aligned} x_j(y^0) \cdot \mu \cdot (A_j^\top y^0 - A_j^\top y^1) &\leq x_j(y^1) - x_j(y^0) \\ &\leq x_j(y^1) \cdot \mu \cdot (A_j^\top y^0 - A_j^\top y^1). \end{aligned}$$

Therefore

$$\begin{aligned} &\sum_j |x_j(y^1) - x_j(y^0)| \\ &\leq \sum_j \max\{x_j(y^0), x_j(y^1)\} \cdot \mu \cdot |A_j^\top y^0 - A_j^\top y^1| \\ &\leq \sum_j (x_j(y^0) + x_j(y^1)) \cdot \mu \cdot \sum_i A_{ij} |y_i^0 - y_i^1| \\ &= O(\mu) \cdot \sum_i |y_i^0 - y_i^1| \cdot (A_i x(y^0) + A_i x(y^1)). \quad (25) \end{aligned}$$

The equation (25) follows by interchanging the order of summations over  $i$  and  $j$ . From equations (24) and (25), we conclude that

$$\Psi^0 - \Psi^1 \geq \Omega\left(\frac{\alpha}{\mu}\right) \cdot \sum_j |x_j(y^1) - x_j(y^0)|.$$

Thus the proof of (20) is complete. Now we prove (21). From (23), we know

$$\begin{aligned} \Psi^0 - \Psi^1 &\geq \sum_i (y_i^0 - y_i^1) \cdot (1 - A_i x(y^1)) \\ &\geq \sum_{i \in S^-} (y_i^0 - y_i^1) \cdot (1 - A_i x(y^1)) \\ &\geq \sum_{i \in S^-} (y_i^0 - y_i^1) \cdot (1 - (1 + \frac{\alpha}{2}) \cdot A_i x(y^0)) \quad (26) \\ &\geq \sum_{i \in S^-} \Omega(\beta) \cdot y_i^0 \cdot ((1 - \alpha) - A_i x(y^0)) \quad (27) \\ &\geq \sum_i \Omega(\beta) \cdot y_i^0 \cdot ((1 - \alpha) - A_i x(y^0)) \quad (28) \\ &= \Omega(\beta) \left[ (1 - \alpha) \cdot \mathbb{1} \cdot y^0 - x(y^0)^\top A^\top y^0 \right]. \end{aligned}$$

The argument for inequality (26) is similar to the one described in case  $i \in S^-$  above and uses the fact that  $A_{i,x}$  changes by a factor of  $\frac{\alpha}{2}$  in a single round; inequality (27) follows since  $y_i^0$  (for  $i \in S^-$ ) decreases by a factor of  $\beta$ ; inequality (28) follows from the definition of  $S^-$ . The proof of (22) is similar and is omitted.  $\square$

### 4.3 Convergence of the dynamics

Let  $y(t)$ ,  $x(t)$ , and  $\Psi(t)$  denote the values of  $y$ ,  $x(y)$ , and  $\Psi(y)$  respectively at the beginning of round  $t$ . In light of the potential decrease lemma 4.3, we consider the following definition.

**DEFINITION 4.4 (STATIONARY INTERVAL).** Consider  $t_0 \geq \tau_1$ . We call an interval  $\mathcal{T} = [t_0, t_1]$  of rounds stationary if

- For all  $t \in \mathcal{T}$ , we have  $(1-2\alpha) \cdot \mathbb{1} \cdot y(t) \leq x(t)^\top A^\top y(t) \leq (1+2\alpha) \cdot \mathbb{1} \cdot y(t)$ , and
- For all  $t \in \mathcal{T}$ , we have  $\sum_{t' \in \mathcal{T}} \sum_j |\Delta x_j(t')| \leq \frac{\alpha}{2} \cdot \mathbb{1} \cdot y(t)$  where  $\Delta x_j(t') = x_j(t'+1) - x_j(t')$ .

We call an interval unstationary if it is not stationary.

Analogous to Lemma 3.6, Lemma 4.5 states that the potential  $\Phi$  decreases significantly in any unstationary interval (and hence there cannot be too many disjoint unstationary intervals) while Lemma 4.6 states that a stationary interval of logarithmic length implies near-optimality throughout the interval.

**LEMMA 4.5.** During any unstationary interval  $\mathcal{T} = [t_0, t_1]$ , the potential decreases by at least  $\Omega(\frac{\alpha^2}{\mu}) \cdot \mathbb{1} \cdot y(t)$  for some  $t \in \mathcal{T}$ .

**PROOF.** By definition, the interval  $\mathcal{T}$  is unstationary because either  $x(t)^\top A^\top y(t) < (1-2\alpha) \cdot \mathbb{1} \cdot y(t)$ , or  $(1+2\alpha) \cdot \mathbb{1} \cdot y(t) < x(t)^\top A^\top y(t)$ , or  $\sum_{t' \in \mathcal{T}} \sum_j |\Delta x_j(t')| > \frac{\alpha}{2} \cdot \mathbb{1} \cdot y(t)$  for some  $t \in \mathcal{T}$ . In any case, from Lemma 4.3, we get that the potential decreases by either  $\Omega(\beta) \cdot \alpha \cdot \mathbb{1} \cdot y(t) = \Omega(\frac{\alpha^2}{\mu}) \cdot \mathbb{1} \cdot y(t)$  or  $\Omega(\frac{\alpha}{\mu}) \cdot \frac{\alpha}{2} \cdot \mathbb{1} \cdot y(t) = \Omega(\frac{\alpha^2}{\mu}) \cdot \mathbb{1} \cdot y(t)$ .  $\square$

**LEMMA 4.6.** Suppose there is a stationary interval  $\mathcal{T} = [t_0, t_1]$  such that  $t_0 \geq \tau_1$  and that has length  $t_1 - t_0 \geq \tau_0$  where  $\tau_0 = O(\frac{1}{\beta} \ln \frac{1}{\delta})$ . Then for all rounds  $\hat{t} \in \mathcal{T}$ , the solution  $y(\hat{t})$  is an  $(1+6\epsilon)$ -approximate solution to the covering LP (2).

**PROOF.** Fix a round  $\hat{t} \in \mathcal{T}$  and let  $y_0$  and  $x_0$  denote  $y(\hat{t})$  and  $x(\hat{t})$  respectively. Let  $y^*$  be the optimum solution to the covering LP (2). Let  $\gamma = \mathbb{1} \cdot y_0 / \mathbb{1} \cdot y^*$  be the approximation factor of the solution  $y_0$ . Assume, on the contrary, that  $\gamma > 1 + 6\epsilon$ . Note that  $A^\top y^* \geq \mathbb{1}$ . We now divide  $j$ th row of  $A^\top$  by  $A_j^\top y^*$  for every  $j$  to obtain a matrix  $\hat{A}^\top$ . Thus we have  $\hat{A}^\top y^* = \mathbb{1}$  and hence  $x_0^\top \hat{A}^\top y^* = \mathbb{1} \cdot x_0$ . Thus

$$\begin{aligned} (1-2\alpha) \cdot \mathbb{1} \cdot y^* &= \frac{1-2\alpha}{\gamma} \cdot \mathbb{1} \cdot y_0 \\ &\leq \frac{1}{\gamma} \cdot x_0^\top A^\top y_0 \quad (\text{since } \mathcal{T} \text{ is stationary}) \\ &\leq \frac{1+4\epsilon}{\gamma} \cdot \mathbb{1} \cdot x_0 \quad (\text{from Lemma 4.2}) \\ &= \frac{1+16\alpha}{\gamma} \cdot x_0^\top \hat{A}^\top y^*. \end{aligned}$$

Thus

$$\frac{\sum_i \hat{A}_i x_0 \cdot y_i^*}{\sum_i y_i^*} = \frac{\sum_i (x_0^\top \hat{A}^\top)_i \cdot y_i^*}{\sum_i y_i^*} > \gamma \cdot (1-18\alpha). \quad (29)$$

Furthermore, from the definition of stationary interval, we have  $\sum_{t \in \mathcal{T}} \sum_j |\Delta x_j(t)| \leq \frac{\alpha}{2} \cdot \mathbb{1} \cdot y_0 = \frac{\alpha}{2} \cdot \gamma \cdot \mathbb{1} \cdot y^*$ . Thus

$$\begin{aligned} &\sum_i y_i^* \cdot \sum_{t \in \mathcal{T}} |\hat{A}_i x(t+1) - \hat{A}_i x(t)| \\ &\leq \sum_i y_i^* \cdot \sum_{t \in \mathcal{T}} \sum_j \hat{A}_{ij} \cdot |x_j(t+1) - x_j(t)| \\ &= \sum_{t \in \mathcal{T}} \sum_j |x_j(t+1) - x_j(t)| \cdot \hat{A}_j^\top y^* \\ &= \sum_{t \in \mathcal{T}} \sum_j |x_j(t+1) - x_j(t)| \\ &\leq \frac{\alpha}{2} \cdot \gamma \cdot \mathbb{1} \cdot y^*. \end{aligned} \quad (30)$$

From inequalities (29) and (30), we conclude that

$$\begin{aligned} \frac{1}{\sum_i y_i^*} \sum_i y_i^* \cdot \left[ \hat{A}_i x_0 - \sum_{t \in \mathcal{T}} |\hat{A}_i x(t+1) - \hat{A}_i x(t)| \right] \\ > \gamma \cdot \left( 1 - 18\alpha - \frac{\alpha}{2} \right) \geq 1 + \alpha. \end{aligned}$$

The last inequality holds since  $\gamma > 1 + 6\epsilon = 1 + 24\alpha$ . Now think of the LHS above as the (weighted) average of the terms in square brackets. Since the average is greater than  $1 + \alpha$ , there exists  $i$  such that

$$\left[ \hat{A}_i x_0 - \sum_{t \in \mathcal{T}} |\hat{A}_i x(t+1) - \hat{A}_i x(t)| \right] > 1 + \alpha,$$

and hence  $\hat{A}_i x(t) \geq 1 + \alpha$  for all  $t \in \mathcal{T}$ . Since the entries in  $A$  are at least the corresponding entries in  $\hat{A}$ , we conclude that  $A_i x(t) > 1 + \alpha$  for all  $t \in \mathcal{T}$ . Thus the variable  $y_i$  in the algorithm must be increased by a factor of at least  $1 + \beta$  in each of the  $\Theta(\frac{1}{\beta} \ln \frac{1}{\delta})$  rounds in  $\mathcal{T}$ . Note that  $y_i$  increases to at least  $\delta$  in a single round and therefore would become larger than 2 in  $\tau_0$  rounds. This when combined with the fact that every non-zero entry in  $A$  is at least 1, implies that  $A_j^\top y$  would become larger than 2 for all  $j$  such that  $A_{ij} > 0$ . Thus  $x_j$  for such  $j$  would become smaller than  $\exp[-\mu]$ . This in turn contradicts the fact that  $A_i x(t) > 1 + \alpha$  in all rounds  $t \in \mathcal{T}$ .  $\square$

**THEOREM 4.7.** After  $O(\frac{\mu}{\alpha^2} \cdot \frac{1}{\beta} \ln \frac{1}{\delta} \cdot \ln(mn \cdot A_{\max} \cdot y_{\max})) = \tilde{O}(\frac{1}{\epsilon^5} \ln^2(n \cdot A_{\max}) \ln^2(mn \cdot A_{\max} \cdot y_{\max}))$  rounds, there must exist a round  $t$  such that  $y(t)$  forms an  $(1+6\epsilon)$ -approximate solution to the covering LP (2). Moreover the total number of rounds in which the solution is not  $(1+6\epsilon)$ -approximate is also at most the above bound.

**PROOF.** The proof is based on Lemmas 4.5 and 4.6 and the fact that the potential is always between  $\frac{1}{A_{\max}}$  and  $m \cdot y_{\max} + \frac{n}{\mu}$ .

Partition the rounds into contiguous intervals of length  $\tau_0$  each. If an interval is stationary, from Lemma 4.6, we have near-optimality throughout the interval. If an interval  $\mathcal{T} = [t_0, t_1]$  is unstationary, from Lemma 4.5, the potential reduces by  $\Omega(\frac{\alpha^2}{\mu} \cdot \mathbb{1} \cdot y(t))$  for some  $t \in \mathcal{T}$ . If  $\frac{1}{\mu} \cdot \mathbb{1} \cdot x(y(t)) \leq \mathbb{1} \cdot y(t)$ , we get that the potential in  $\mathcal{T}$  reduces by  $\Omega(\frac{\alpha^2}{\mu} \cdot \Psi(t)) \geq \Omega(\frac{\alpha^2}{\mu} \cdot \Psi(t_1))$ . On the other hand, if  $\mathbb{1} \cdot x(y(t)) > \mu \cdot \mathbb{1} \cdot y(t)$ , we get  $x(y(t))^\top A^\top y(t) \geq \mathbb{1} \cdot x(y(t)) > \mu \cdot \mathbb{1} \cdot y(t)$ . Thus from Lemma 4.3-Eqn. (22), we conclude that the potential decreases by  $\Omega(\beta) \cdot \mathbb{1} \cdot x(y(t)) = \Omega(\beta\mu) \cdot \Psi(t) \geq \Omega(\frac{1}{\mu}) \cdot \Psi(t_1)$ .

In either case, the potential reduces by a factor of  $\Omega(\frac{\alpha^2}{\mu})$  in  $\mathcal{T}$ .

Since the the potential is always between  $\frac{1}{A_{\max}}$  and  $m \cdot y_{\max} + \frac{n}{\mu}$ , we get that the total number of unstationary intervals is at most  $O(\frac{\mu}{\alpha^2} \cdot \ln(mn \cdot A_{\max} \cdot y_{\max}))$ . Hence the proof is complete.  $\square$

## 5. CONCLUSIONS

We presented *stateless* algorithms for packing/covering LPs in a natural distributed model. Our algorithms are arguably simpler than the previous algorithms and have poly-logarithmic convergence guarantees. We feel that the *distributed gradient descent* technique developed in this paper should generalize to larger class of mathematical programs, e.g., general LPs or convex programs including SDPs. Can we show that similar algorithms work for implicitly given packing/covering linear programs, e.g., multi-commodity flows with unspecified paths, without associating separate agents for separate paths?

## 6. REFERENCES

- [1] B. Awerbuch and Y. Azar. Local optimization of global objectives: competitive distributed deadlock resolution and resource allocation. *FOCS*, 1994.
- [2] B. Awerbuch, Y. Azar, and R. Khandekar. Fast load balancing via bounded best response. *SODA*, 2008.
- [3] B. Awerbuch and R. Khandekar. Stateless near optimal distributed flow control with poly-logarithmic convergence. *LATIN*, 2008.
- [4] B. Awerbuch and R. Khandekar. Distributed network monitoring and multicommodity flows: a primal-dual approach. *PODC*, 2007.
- [5] B. Awerbuch and R. Khandekar. Greedy distributed optimization of multi-commodity flows. *PODC*, 2007. 2007.
- [6] B. Awerbuch, R. Khandekar, and S. Rao. Distributed algorithms for multicommodity flow problems via approximate steepest descent framework. *SODA*, 2007.
- [7] B. Awerbuch, B. Patt-Shamir, and G. Varghese. Self-stabilization by local checking and correction. *FOCS*, 1991.
- [8] B. Awerbuch and G. Varghese. Distributed program checking: a paradigm for building self-stabilizing distributed protocols. *FOCS*, 1991.
- [9] Y. Bartal, J. W. Byers, and D. Raz. Global optimization using local information with applications to flow control. *FOCS*, 1997.
- [10] E. Dijkstra. Self stabilizing systems in spite of distributed control. *CACM*, 17:643–644, Nov 1974.
- [11] S. Dolev, A. Israeli, and S. Moran. Self-stabilization of dynamic systems assuming only read/write atomicity. *PODC*, 1990.
- [12] L. Fleischer. Approximating fractional multicommodity flow independent of the number of commodities. *SIAM Journal on Discrete Mathematics*, 13(4):505–520, 2000.
- [13] N. Garg and J. Könemann. Faster and simpler algorithms for multicommodity flow and other fractional packing problems. *SIAM Journal on Computing*, 37(2):630–652, 2007.
- [14] N. Garg and N. E. Young. On-line end-to-end congestion control. *FOCS*, 2002.
- [15] M. G. Gouda and N. J. Multari. Stabilizing communication protocols. Technical Report TR-90-20, Dept. of Computer Science, University of Texas at Austin, June 1990.
- [16] C. Koufogiannakis and N. E. Young. Beating simplex for fractional packing and covering linear programs. *FOCS*, 2007.
- [17] F. Kuhn. *The price of locality: exploring the complexity of distributed coordination primitives*. PhD Thesis, ETH Zurich, Diss. ETH No. 16213, December 2005.
- [18] F. Kuhn, T. Moscibroda, and R. Wattenhofer. The price of being near-sighted. *SODA*, 2006.
- [19] F. Kuhn and R. Wattenhofer. Constant-time distributed dominating set approximation. *PODC*, 2003.
- [20] N. Linial. Locality in distributed graph algorithms. *SIAM Journal of Computing*, 21(1):193–201, 1992.
- [21] M. Luby and N. Nisan. A parallel approximation algorithm for positive linear programming. *STOC*, 1993.
- [22] M. Naor and L. Stockmeyer. What can be computed locally? *SIAM Journal of Computing*, 24(6):1259–1277, 1995.
- [23] C. Papadimitriou and M. Yannakakis. Linear programming without the matrix. *STOC*, 1993.
- [24] D. Peleg. *Distributed Computing: A Locality-Sensitive Approach*. SIAM, 2000.
- [25] S. Plotkin, D. Shmoys, and E. Tardos. Fast approximation algorithms for fractional packing and covering problems. *Mathematics of Operations Research*, 20(2):257–301, 1995.
- [26] N. E. Young. Sequential and parallel algorithms for mixed packing and covering. *FOCS*, 2001.