

Nearly Linear-Work Algorithms for Mixed Packing/Covering and Facility-Location Linear Programs

Neal E. Young *

University of California, Riverside, neal.young@ucr.edu

Abstract

We describe the first nearly linear-time approximation algorithms for explicitly given mixed packing/covering linear programs, and for (non-metric) fractional facility location. We also describe the first parallel algorithms requiring only near-linear total work and finishing in polylog time. The algorithms compute $(1 + \epsilon)$ -approximate solutions in time (and work) $\tilde{O}(N/\epsilon^2)$, where N is the number of non-zeros in the constraint matrix. For facility location, N is the number of eligible client/facility pairs.

*Research partially supported by NSF grant 1117954.

1 Introduction

Mixed packing/covering LP's are fundamental to combinatorial optimization in computer science and operations research, with numerous applications, including many that are not pure packing or covering — solving linear systems, computer tomography, machine scheduling, routing problems, multicommodity flow with demands, etc. In many approximation algorithms for NP-hard problems, solving such an LP in order to round the solution is the main bottleneck.

Algorithms with linear worst-case run time (the time it takes just to read the input) have always been something of a holy grail. Their importance is increasing with the abundance of big data, and, with the growing reliance on multiprocessors and server farms, linear-work algorithms that can be highly parallelized are of particular interest.

Results. We give the first nearly linear-time $(1 + \epsilon)$ -approximation algorithms and the first nearly linear-work parallel algorithms for mixed packing/covering linear programs and for fractional facility location. Let N be the input size, that is, the number of non-zeroes in the linear program. Let m and n be the numbers of constraints and variables, respectively. Generally $\max(m, n) \leq N \leq mn$.

For mixed packing/covering, Thm. 1 gives a $(1 + \epsilon)$ -approximation algorithm taking time $O(N \log(m)/\epsilon^2)$ and a parallel algorithm doing work $O(N \log m \log(n \log(m)/\epsilon)/\epsilon^2)$ in polylog time, $O(\log N \log^2 m \log(n \log(m)/\epsilon)/\epsilon^4)$. For fractional facility location, Thm. 2 gives a $(1 + \epsilon)$ -approximation algorithm running in time $O(N \log(m)/\epsilon^2)$ and a parallel algorithm doing work $O(N \log N \log(m)/\epsilon^2)$, in polylog time, $O(\log N \log^2(m)/\epsilon^4)$. For facility location, the input size N is the number of eligible (client, facility) pairs; there are m clients and n facilities.

Definitions. A *mixed packing/covering linear program (LP)* is of the form “find $x \in \mathbb{R}_+^n$ such that $Cx \geq c$ and $Px \leq p$ ”¹ where C and P are non-negative. A $(1 + \epsilon)$ -approximate solution is an x such that $Cx \geq c$ and $Px \leq (1 + \epsilon)p$. Here are some special cases: a *(non-negative) linear system* is of the form “find $x \in \mathbb{R}_+^n$ such that $Ax = b$ ”; *pure packing/covering LP's* are primal and dual LPs of the form $\max\{c \cdot x : x \in \mathbb{R}_+^n, Ax \leq b\}$ and $\min\{b \cdot y : y \in \mathbb{R}_+^m, A^\top y \geq c\}$; *covering with box constraints* is of the form $\min\{c \cdot x : x \in \mathbb{R}_+^n, Ax \geq b, x \leq u\}$. Above, A must be non-negative.

For facility location (Section 4), a $(1 + \epsilon)$ -approximate solution is one of cost at most $1 + \epsilon$ times minimum. (The LP is not a mixed packing/covering LP. It has two standard reformulations as a set-cover LP, but both increase LP size super-linearly, so don't yield nearly linear-time algorithms by direct reduction to covering. See appendix Section 5.3 and [15], [17, §6.2].)

Techniques. The algorithms are Lagrangian-relaxation algorithms. Roughly, one starts with an all-zero (or small) vector x , then repeatedly increments x by an increment vector δ . (The direction of δ is guided by the gradient of a scalar-valued penalty function $\phi(x)$; the size ensures $x + \delta$ is within a trust region around x , so that $\delta \cdot \nabla \phi(x) = (1 + O(\epsilon))(\phi(x + \delta) - \phi(x))$.) The penalty function ϕ combines the constraints into a smooth scalar-valued function of the current solution x . Ours for mixed packing/covering is roughly $\phi(x) \approx \log(\sum_i (1 + \epsilon)^{P_i x} \times \sum_i (1 - \epsilon)^{C_i x})$.

For mixed packing/covering, [28] gives an algorithm with time $\tilde{O}(md/\epsilon^2)$ and a parallel algorithm with work $\tilde{O}(md/\epsilon^2)$, where $d \leq m$ is the maximum number of constraints that any variable appears in. (Note md is not generally close to linear.) [28] uses ideas from works on *pure* packing/covering: round-robin consideration of variables [9, 10], non-uniform increments [11, 12], and incrementing multiple variables at once [21]. For the special case of pure packing/covering, [18] is the first to achieve nearly linear time, $\tilde{O}((n + m)/\epsilon^2 + N)$. That algorithm randomly couples primal and dual algorithms (an idea from [14]), and uses a random sampling trick to reduce the

¹[28, Lemma 8], reduces the more general $\min\{\lambda : \exists x \in \mathbb{R}_+^n : Cx \geq x; Px \leq \lambda p\}$ to a small number of these.

intermediate calculations. The algorithms in this paper incorporate and adapt all the above ideas except coupling, improving (for the first time since 2001) the bounds from [28].

Other related work. Lagrangian-relaxation algorithms have a large literature [2, 5, 25]. Generally, they maintain sparsity easily (similar to iterative solvers) and are numerically stable.

By allowing $1/\epsilon^2$ dependence on ϵ , the algorithms here achieve near-optimal dependence on the other parameters. Recent sequential algorithms building on Nesterov ([22]) reduce the dependence to $1/\epsilon$: for mixed packing/covering, [6, Thm. 12] achieves $\tilde{O}(n^{2.5}K_p^{1.5}\max(K_p, K_c)^{0.5}/\epsilon)$ time, where K_p and K_c are, respectively, the maximum number of non-zeros in any packing or covering constraint; [8, Thm’s 3,4,6] achieves time $\tilde{O}(m^{1.5}n^{1.5}/\epsilon)$ for facility location, and similar results for pure packing and set cover. So far, the reduced dependence on ϵ always comes at the expense of polynomial (super-linear) dependence on other parameters, so is asymptotically slower unless $1/\epsilon$ is growing polynomially with N .

For the special case of pure packing/covering, a recent parallel algorithm [1] achieves near-linear work $O(N \log^2(N)/\epsilon^3)$, and time $O(\log^2(N)/\epsilon^3)$, breaking the $1/\epsilon^4$ time barrier for polylog-time parallel algorithms. The algorithm does not apply to *mixed* packing/covering.

Solving linear systems (“*find x such that $Ax = b$* ”) is a fundamental algorithmic problem. The algorithms here find an approximate solution in nearly linear time (or in parallel doing near-linear-work) for the special case when A and the solution x are non-negative (e.g., for computer tomography [4, 28]). Another important special case is when A is a *graph Laplacian* [13, 16, 23, 24]. (Note that graph Laplacians are not non-negative, so the form of approximation differs.) Those solvers are a basic building block in many settings, one celebrated recent example being the nearly linear-time approximation algorithm for maximum flow ([7] and surveys [20, 26]).

Online mixed packing/covering was recently shown to have polylog-competitive algorithms [3].

Future work. For pure packing/covering, [18, 19] achieves time $O((n+m)\log(N)/\epsilon^2 + N)$, shifting the $1/\epsilon^2$ factor to a lower-order term for dense instances. Is this possible for mixed packing/covering or facility location? So far the primal/dual coupling used in [18, 19] eludes efficient extension to mixed packing/covering.

2 Mixed Packing and Covering

Theorem 1. *For mixed/packing covering, there are $(1 + \epsilon)$ -approximation algorithms running*

- (i) *in time $O(N \log(m)/\epsilon^2)$,*
- (ii) *in parallel time $O(\log N \log^2 m \log(n \log(m)/\epsilon)/\epsilon^4)$, doing work $O(N \log(n \log(m)/\epsilon) \log(m)/\epsilon^2)$.*

The rest of this section proves Thm. 1. The starting point is Alg. 1 (above), which is essentially a convenient reformulation of the generic “algorithm with phases” in [28, Fig. 2]. For part (i), we’ll describe how to implement it to run faster by only estimating the intermediate quantities of interest and updating the estimates periodically. We’ll need the following properties (essentially from [28]):

Lemma 1. *Given any feasible mixed packing/covering instance (P, C) , Alg. 1*

- (i) *returns a $(1 + O(\epsilon))$ -approximate solution (i.e., x such that $Cx \geq 1$ and $Px \leq 1 + O(\epsilon)$),*
- (ii) *scales λ_0 (lines 12–13) at most $O(U)$ times, where $U = O(\log(m)/\epsilon^2 + \max_i P_i x^0/\epsilon)$, and*
- (iii) *increments x (lines 5–11) at most $O(mU)$ times.*

See the appendix for a proof, which follows the proofs of Lemmas 1–5 of [28]. After initialization, Alg. 1 simply repeats one of two operations: **(a)** incrementing the current solution x by

Algorithm 1 Generic approximation algorithm for mixed packing/covering LPs

1 **function** PACKING-COVERING(matrices C, P ; initial solution $x = x^0 \in \mathbb{R}_+^n, \epsilon \in (0, 1/10)$)
2 Define: $U = (\max_i P_i x^0 + \ln m)/\epsilon^2$
 $p_i(x) = (1 + \epsilon)^{P_i x}$ $c_i(x) = (1 - \epsilon)^{C_i x}$ if $C_i x \leq U$, else $c_i(x) = 0$
 $|p(x)| = |p(x)|_1$ $|c(x)| = |c(x)|_1 = \sum_i c_i(x)$
 $\lambda(x, j) = P_j^T p(x)/C_j^T c(x)$ $\lambda^*(x) = \min_{j \in [n]} \lambda(x, j)$.
3 Initialize $\lambda_0 \leftarrow |p(x)|/|c(x)|$.
4 Repeatedly do **either** of the following two operations whose precondition is met:
5 **operation (a): increment** x \triangleright *precondition:* $\lambda^*(x) \leq (1 + 4\epsilon)\lambda_0$
6 Choose $\delta \in \mathbb{R}_+^n$ such that
7 (i) $\forall j \in [n]$, if $\delta_j > 0$ then $\lambda(x, j) \leq (1 + 4\epsilon)\lambda_0$, and
8 (ii) $\max\{\max_i P_i \delta, \max_{i: C_i x \leq U} C_i \delta\}$ is in $[1/2, 1]$
9 (the maximum increase in any $P_i x$ or active $C_i x$ is between $1/2$ and 1).
10 Let $x \leftarrow x + \delta$.
11 If $\min_i C_i x \geq U$ then return x/U .
12 **operation (b): scale** λ_0 \triangleright *precondition:* $\lambda^*(x) \geq (1 + \epsilon)\lambda_0$
13 Let $\lambda_0 \leftarrow (1 + \epsilon)\lambda_0$.

some vector δ , or **(b)** scaling λ_0 by $1 + \epsilon$. In each iteration, it can do either operation whose precondition is met, and when incrementing x there are many valid ways to choose δ .

2.1 Proof of part (i), sequential algorithm

Alg. 2, which we use to prove part (i) of Thm. 1, repeats these two operations in a particular way. To reduce the run time to nearly linear, instead of computing $Px, Cx, p(x)$ and $c(x)$ exactly as it proceeds, Alg. 2 maintains estimates: $\hat{P}, \hat{C}, \hat{p}$, and \hat{c} . To prove correctness, we show that the estimates suffice to ensure a correct implementation of Alg. 1, which is correct by Lemma 1 (i).

Lemma 2. *Given any feasible packing/covering instance (P, C) , provided the updates in lines 12–13 maintain Invariant (1):*

- (i) *Each operation (a) or (b) done by Alg. 2 is a valid operation (a) or (b) of Alg. 1, so*
- (ii) *Alg. 2 returns a $(1 + O(\epsilon))$ -approximate solution.*

The proof is in the appendix. The proof follows [28], but adds the idea from [18, 19] of maintaining estimates by sampling. (Here we maintain the estimates differently, though, using deterministic, periodic sampling as detailed in the next proof.)

Lemma 3. *Alg. 2 can do the updates in lines 11–12 so as to maintain Invariant (1) and take total time $O(N \log(m)/\epsilon^2)$.*

Proof. The algorithm maintains the following global data:

- the current solution x , and vectors $\hat{P}, \hat{C}, \hat{p}$ and \hat{c} satisfying Invariant (1);
- for each j , *column maxima:* $\max_i P_{ij}$ and $\max\{C_{ij} : C_i x \leq U\}$.

Initializing these items takes $O(N)$ (linear) time, with the exception of the column maxima. To initialize and maintain the maxima, the algorithm presorts the entries within each column of

Algorithm 2 Sequential implementation of Alg. 1 for mixed packing/covering LPs

```

1 function SEQUENTIAL-PACKING-COVERING( $P, C, \epsilon$ )
2   Initialize  $x_j \leftarrow 0$  for  $j \in [n]$ ,  $\lambda_0 \leftarrow |p(x)|/|c(x)| = m_p/m_c$ , and  $U = \ln(m)/\epsilon^2$ .
3   Maintain vectors  $\widehat{P}, \widehat{C}, \widehat{p}$ , and  $\widehat{c}$ , to satisfy invariant
4     For all  $i$  :  $\widehat{P}_i \in (P_i x - 1, P_i x]$   $\widehat{p}_i = (1 + \epsilon)^{\widehat{P}_i}$ ,
5      $\widehat{C}_i \in (C_i x - 1, C_i x]$   $\widehat{c}_i = (1 - \epsilon)^{\widehat{C}_i}$  if  $\widehat{C}_i \leq U$ , else  $\widehat{c}_i = 0$ . (1)
6   repeat
7     for each  $j \in [n]$  do ▷ do a run for  $x_j$ 
8     Compute values of  $P_j^\top \widehat{p}$  and  $C_j^\top \widehat{c}$  from  $\widehat{p}$  and  $\widehat{c}$ . Define  $\widehat{\lambda}_j = P_j^\top \widehat{p} / C_j^\top \widehat{c}$ .
9     while  $\widehat{\lambda}_j \leq (1 + \epsilon)^2 \lambda_0 / (1 - \epsilon)$  do
10    operation (a): increment  $x_j$  ▷ assertion:  $\lambda^*(x) \leq (1 + 4\epsilon)\lambda_0$ 
11    Let  $x_j \leftarrow x_j + z$ , choosing  $z$  so  $\max\{\max_i P_{ij} z, \max_{i: C_{ij} x \leq U} C_{ij} z\} = 1/2$ .
12    As described in text, to maintain Invariant (1):
13    For selected  $i$  with  $P_{ij} \neq 0$ , update  $\widehat{P}_i$  and  $\widehat{p}_i$ . Update  $P_j^\top \widehat{p}$  accordingly.
14    For selected  $i$  with  $C_{ij} \neq 0$ , update  $\widehat{C}_i$  and  $\widehat{c}_i$ . Update  $C_j^\top \widehat{c}$  accordingly.
15    If  $\min_i \widehat{C}_i \geq U$  then return  $x/U$  ▷ finished
16  operation (b): scale  $\lambda_0$  ▷ assertion:  $\lambda^*(x) \geq (1 + \epsilon)\lambda_0$ 
17  Let  $\lambda_0 \leftarrow (1 + \epsilon)\lambda_0$ .

```

P and C , in total time $O(N \log m)$, then, every time some covering constraint $C_i x \geq U$ becomes satisfied, updates the maxima of the columns j with $C_{ij} \neq 0$. (The total time for these updates is linear, as each can be charged to a non-zero C_{ij} .)

To maintain Invariant (1), the algorithm will actually guarantee something stronger: outside the **while** loop, each estimate \widehat{P}_i and \widehat{C}_i will be exactly accurate (that is, $\widehat{P}_i = P_i x$ and $\widehat{C}_i = C_i x$ for all i). Inside the **while** loop, during a run of increments for a particular x_j , for each i , only the contributions of $P_{ij} x_j$ to $P_i x$ (for the current j) will be underestimated in \widehat{P}_i . Likewise, only the contributions of $C_{ij} x_j$ to $C_i x$ will be underestimated in \widehat{C}_i . In line 12, the algorithm will update those \widehat{P}_i and \widehat{C}_i for which the under-estimation is in danger of exceeding 1, as follows.

Maintaining the estimates using periodic sampling. Define the *top* of any number $y > 0$ to be the smallest power of 2 greater than or equal to y . In a preprocessing step, within each column C_j^\top and P_j^\top separately, partition the non-zero entries C_{ij} and P_{ij} into equivalence classes according to their tops, and order the groups by decreasing top. (Use the presorted entries within each column to do this in $O(N)$ total time.)

Call a consecutive sequence of increments to x_j (done within a single iteration of the **for** loop for j) a *run* for x_j . During a run for x_j , say that a group G in P_j^\top with top 2^t is *eligible for update* if the increase δ_G in x_j since the last update of group G during the run (or, if none, the start of the run) is at least $1/2^{t+1}$.

Implement line 12 as follows. Starting with the group G in P_j^\top with largest top, Check the group to see if it's eligible for update ($\delta_G \geq 1/2^{t+1}$). If it is, then, for each i in the group G , increase \widehat{P}_i to $P_i x$ in constant time by adding $P_{ij} \delta_G$ to \widehat{P}_i . Update each scalar dependent of \widehat{P}_i ($\widehat{p}_i, P_j^\top \widehat{p}, \widehat{\lambda}_j$). Then, continue with the next group in P_j^\top (the one with next smaller top). Stop processing the groups in P_j^\top with the first group that is not eligible for update. — don't process any subsequent groups with smaller tops, regardless of eligibility.

Implement line 13 for \widehat{C} and its dependents likewise. (When updating some \widehat{C}_i , check whether

$\widehat{C}_i \geq U$, and if so, delete row i from C and associated data structures.² When the last row of C is deleted, stop and return x/U (line 14.)

Finally, at the end of the run for x_j , to maintain the invariant that all estimates are exact outside of the while loop, do the following. For each group G in P_j^T , for each i in G , update \widehat{P}_i to the exact value of $P_i x$ by increasing \widehat{P}_i by $P_{ij} \delta_G$ (for δ_G defined above), and update \widehat{p}_i accordingly. Likewise, update \widehat{C}_i (and its dependent \widehat{c}_i) for every i with $C_{ij} \neq 0$ to its exact value.

Correctness of periodic sampling. To show Invariant (1) holds during a run, we prove that, *if a given group G with top 2^t is not updated after a given increment of x_j , then $\delta_G \leq 1/2^t$.* (Invariant (1) follows, because, for $i \in G$, the increase $P_{ij} \delta_G$ in $P_i x$ since the last update of \widehat{P}_i is less than $2^t/2^t = 1$; similarly, the increase $C_{ij} \delta_G$ in $C_i x$ since the last update of \widehat{C}_i is less than 1.)

Suppose for contradiction that the claim fails. Consider the first increment of x_j for which it fails, and the group G with largest top 2^t for which $\delta_G > 1/2^t$ after that increment. Group G cannot be the group with maximum top in its column, because the algorithm considered that group after the increment. Let G' be the group with next larger top $2^{t'} > 2^t$. G' was not updated after the increment, because if it had been G would have been considered and updated. Let x_j denote the current value of x_j , and let $x'_j < x_j$ denote the value at the most recent update of G' .

When group G' was last updated, group G was considered but not updated (for, if G had been updated then, we would now have $\delta_G = \delta_{G'} \leq 1/2^{t'} < 1/2^t$). Thus, letting x''_j be the value of x_j at the most recent update of G , we have $x'_j - x''_j < 1/2^{t+1}$. Since group G' was not updated after the current increment, we have (by the choice of G) that $x_j - x'_j = \delta_{G'} \leq 1/2^{t'} \leq 1/2^{t+1}$. Summing gives $x_j - x''_j < 2/2^{t+1} = 1/2^t$, violating the supposition $\delta_G > 1/2^t$.

Time. At the start of each run for a given x_j , the time in line 7 is proportional to the number of non-zeroes in the j th columns of P and C , as is the time it spends at the end of the run updating all \widehat{P}_i (for $P_{ij} \neq 0$) and \widehat{C}_i (for $C_{ij} \neq 0$). Thus, the cumulative time spent on these actions during any single iteration of the **repeat** loop is $O(N)$. By Lemma 1 (ii), Alg. 2 does $O(U)$ iterations of its **repeat** loop, so the total time for the actions outside of increments is $O(NU)$, as desired.

Each increment to some x_j takes time proportional to the number of updates made to \widehat{P}_i 's and \widehat{C}_i 's. An update to \widehat{P}_i in group G with top 2^t increases \widehat{P}_i by $P_{ij} \delta_G \geq P_{ij}/2^{t+1} > 2^{t-1}/2^{t+1} = 1/4$. Throughout, \widehat{P}_i does not exceed $(1 + O(\epsilon))U$, so \widehat{P}_i is updated $O(U)$ times during increments. Likewise (using that \widehat{C}_i is updated only while $\widehat{C}_i \leq U$), each \widehat{C}_i is updated $O(U)$ times during increments. There are m \widehat{P}_i 's and \widehat{C}_i 's, so there are $O(mU) = O(NU)$ such updates. \square

2.2 Proof of part (ii), parallel algorithm

Next we prove part (ii) of Thm. 1, using Alg. 3. By careful inspection, Alg. 3 just repeats the two operations of Alg. 1 (increment x or scale λ_0), so is correct by Lemma 1 (i). To finish, we detail how to implement the steps so a careful accounting yields the desired time and work bounds.

Call each iteration of the repeat loop a *phase*. Each phase scales λ_0 , so by Lemma 1 (ii), there are $O(U)$ phases. By inspection, $\max_i P_i x^0 \leq 1$, so $U = O(\log(m)/\epsilon^2)$. Within any given phase, for the *first* increment, compute all quantities directly in $O(\log N)$ time and $O(N)$ total

²The condition " $\widehat{C}_i \geq U$ " differs from $C_i x \geq U$ in Alg. 1. We note without proof that this doesn't affect correctness.

Algorithm 3 Parallel implementation of Alg. 1 for mixed packing /covering LPs

```

1 function PARALLEL-PACKING-COVERING( $P, C, \epsilon$ )
2   Initialize  $x_j \leftarrow n^{-1} / \max_i P_{ij}$  for  $j \in [n]$ ,  $\lambda_0 \leftarrow |p(x)| / |c(x)| = m_p / m_c$ .
3   Define  $U, P_i x, C_i x, p_i(x), c_i(x), \lambda(x, j)$ , etc. per Alg. 1.
4   repeat
5     while  $\lambda^*(x) \leq (1 + \epsilon)\lambda_0$  do
6       operation (a): increment  $x$   $\triangleright$  assertion:  $\lambda^*(x) \leq (1 + 4\epsilon)\lambda_0$ 
7         Define  $J = \{j \in [n] : \lambda(x, j) \leq (1 + \epsilon)\lambda_0\}$ , and, for  $j \in J$ ,
           $I_j^p = \{i : P_{ij} \neq 0\}$  and  $I_j^c = \{i : C_{ij} \neq 0 \text{ and } C_i x \leq U\}$ .
8         For  $j \in J$ , let  $\delta_j = z x_j$  (and, implicitly,  $\delta_j = 0$  for  $j \notin J$ ),
9         choosing  $z$  such that  $\max\{\max_i P_i \delta, \max_{i: C_i x \leq U} C_i \delta\} = 1$ .
10        For  $j \in J$ , let  $x_j \leftarrow x_j + \delta_j$ .
11        For  $i \in \bigcup_{j \in J} I_j^p$ , update  $P_i x$  and  $p_i(x)$ . For  $i \in \bigcup_{j \in J} I_j^c$ , update  $C_i x$  and  $c_i(x)$ .
12        For  $j \in J$ , update  $C_j^T c(x), P_j^T p(x)$ , and  $\lambda(x, j)$ .
13        If  $\min_i C_i x \geq U$  then return  $x/U$ .  $\triangleright$  finished
14      operation (b): scale  $\lambda_0$   $\triangleright$  assertion:  $\lambda^*(x) \geq (1 + \epsilon)\lambda_0$ 
15      Let  $\lambda_0 \leftarrow (1 + \epsilon)\lambda_0$ .

```

work. In each subsequent increment within the phase, update all quantities incrementally, in time $O(\log N)$ and doing total work linear in the sizes of the sets $E_p = \{(i, j) : j \in J, i \in I_j^p\}$ and $E_c = \{(i, j) : j \in J, i \in I_j^c\}$ of *active edges*.

(For example: update each $P_i x$ by noting that the increment increases $P_i x$ by $\Delta_i^p = \sum_{j: i \in I_j^p} P_{ij} \delta_j$; update $P_j^T p(x)$ by noting that the increment increases it by $\sum_{i \in I_j^p} P_{ij} \Delta_i^p$. Update J by noting that $\lambda(x, j)$ only increases within the phase, so J only shrinks, so it suffices to delete a given j from J in the first increment when $\lambda(x, j)$ exceeds $(1 + \epsilon)\lambda_0$.)

Bounding the work and time. In each increment, if a given j is in J , then the increment increases x_j . When that happens the parameter z is at least $\Theta(1/U)$ (using $P_i x = O(U)$ and $C_i x = O(U)$) so x_j increases by at least a factor of $1 + \Theta(1/U)$. The value of x_j is initially at least $n^{-1} / \max_i P_{ij}$ and finally $O(U / \max_i P_{ij})$. It follows that j is in the set J during at most $O(U \log(nU))$ increments. Thus, for any given non-zero P_{ij} , the pair (i, j) is in E_p in at most $O(U \log(nU))$ increments. Likewise, for any given non-zero C_{ij} , the pair (i, j) is in E_c in at most $O(U \log(nU))$ increments. Hence, the total work for Alg. 3 is $O(NU \log(nU))$, as desired.

To bound the total time, note that, within each of the $O(U)$ phases, some j remains in J throughout the phase. As noted above, no j is in J for more than $O(U \log(nU))$ increments. Hence, each phase has $O(U \log(nU))$ increments. To finish, recall that each increment takes $O(\log N)$ time. This concludes the proof of Thm. 1. \square

3 Pure Covering

This section gives Alg's 4 and 5 for covering, and their performance guarantees, for use in the next section. The proofs (in the appendix) are similar to those of Lemmas 1 and 2.

Lemma 4. *Alg. 4 returns a solution x such that $w \cdot x \leq (1 + O(\epsilon)) \text{OPT}(A, w) + w \cdot x^0$, where x^0 is the initial solution given to the algorithm.*

Lemma 5. *(i) Alg. 5 is a specialization of Alg. 4 and (ii) scales λ_0 $O(U) = O(\log(m)/\epsilon^2)$ times.*

Algorithm 4 Generic approximation algorithm for covering LPs

```
1 function COVERING(matrix  $A$ , cost  $w$ , initial solution  $x = x^0 \in \mathbb{R}_+^n$ ,  $\epsilon \in (0, 1/10)$ )
2   Define       $U = \ln(m)/\epsilon^2$ , where  $m$  is the number of constraints,
                 $a_i(x) = (1 - \epsilon)^{A_i x}$  if  $A_i x \leq U$ , else  $a_i(x) = 0$ ,
                 $|a(x)| = \sum_i a_i(x)$ ,
                 $\lambda(x, j) = w_j / (A_j^\top a(x))$            ( $A_j^\top$  is column  $j$  of  $A$ ).
3   while  $\min_i A_i x \leq U$  do
4     Choose vector  $\delta \in \mathbb{R}_+^n$  such that
5     (i)  $\forall j \in [n]$  if  $\delta_j > 0$  then  $\lambda(x, j) \leq (1 + O(\epsilon)) \text{OPT}(A, w) / |a(x)|$ , and
6     (ii)  $\max\{A_i \delta : i \in [m], A_i x \leq U\} = 1$ .
7     Let  $x \leftarrow x + \delta$ .
8   return  $x/U$ 
```

Algorithm 5 Sequential implementation of Alg. 4

```
1 function SEQUENTIAL-COVERING( $A, w, \epsilon$ )
2   Define  $U, a_i, \lambda(x, j)$ , etc. as in Alg. 4 and  $\lambda^*(x) = \min_j \lambda(x, j)$ .
3   Initialize  $x_j \leftarrow 0$  for  $j \in [n]$  and  $\lambda_0 \leftarrow \max_i \min_{j \in [n]} w_j / (A_{ij} |a(x)|)$ .
4   Repeatedly do one of the following two operations whose precondition is met:
5     operation (a): increment  $x$                                  $\triangleright$  precondition:  $\lambda^*(x) \leq (1 + 4\epsilon)\lambda_0$ 
6       Choose  $j \in [n]$  such that  $\lambda(x, j) \leq (1 + 4\epsilon)\lambda_0$ .
7       Let  $x_j \leftarrow x_j + \min\{1/A_{ij} : A_i x \leq U\}$ .
8       if  $\min_i A_i x \geq U$ , then return  $x/U$ .
9     operation (b): scale  $\lambda_0$                                  $\triangleright$  precondition:  $\lambda^*(x) \geq (1 + \epsilon)\lambda_0$ 
10    Let  $\lambda_0 \leftarrow (1 + \epsilon)\lambda_0$ .
```

4 Facility Location

The facility location LP. Given a set C of m customers, a set F of n facilities, an opening cost $f_j \geq 0$ for each facility j , and a cost $c_{ij} \geq 0$ for assigning customer i to facility j , the standard *facility-location linear program* is

$$\begin{aligned} & \text{minimize}_{x,y} \text{cost}(x, y) = \sum_j f_j y_j + \sum_{ij} c_{ij} x_{ij} \\ & \text{subject to } \sum_j x_{ij} \geq 1 \text{ for } i \in C, \\ & y_j \geq x_{ij} \geq 0 \text{ for } i \in C, j \in F. \end{aligned} \tag{2}$$

For notational convenience, assume $c_{ij} = \infty$ if customer i may not be assigned to facility j . The input size is $N = \{(i, j) : c_{ij} < \infty\}$. A $(1 + \epsilon)$ -approximate solution is a feasible pair (x, y) whose cost is at most $1 + \epsilon$ times minimum.

Theorem 2. *For facility location LPs, there are $(1 + \epsilon)$ -approximation algorithms running*

- (i) *in time $O(N \log(m)/\epsilon^2)$, and*
- (ii) *in parallel time $O(\log^2(m) \log(N/\epsilon) \log(N)/\epsilon^4)$, doing work $O(N \log(N/\epsilon) \log(m)/\epsilon^2)$.*

Reducing to Set Cover. There are two standard reductions of facility location to set cover, both increasing the size of the LP super-linearly (see appendix Section 5.3). Our algorithms will

Algorithm 6 Sequential $(1 + \epsilon)$ -approximation algorithm for facility-location LPs

```

1 function SEQUENTIAL-FACILITY-LOCATION(facilities  $F$ , customers  $C$ , costs  $f, c, \epsilon$ )
2   Define  $U = \ln(m)/\epsilon^2$ ,
       $A_i x = \sum_{j \in F} x_{ij}$  (for  $i \in C$ ),
       $a_i(x) = (1 - \epsilon)^{A_i x}$  if  $A_i x \leq U$ , else  $a_i(x) = 0$ ,
       $\lambda(x, j, S) = (f_j + \sum_{i \in S} c_{ij}) / \sum_{i \in S} a_i(x)$  (for  $j \in F, S \subseteq C$ ).
3   Initialize  $y_j, x_{ij} \leftarrow 0$  for  $j \in F, i \in C$  and  $\lambda_0 \leftarrow (\max_{i \in C} \min_{j \in F} f_j + d_{ij}) / |a(x)|$ .
4   repeat
5     for each  $j \in F$  do
6       while  $\lambda(x, j, S_j) \leq (1 + \epsilon)\lambda_0$  where  $S_j = \{i \in C : c_{ij} < (1 + \epsilon)\lambda_0 a_j(x)\}$  do
7         operation (a): increment for  $j$   $\triangleright$  assert:  $\lambda(x, j, S_j) \leq (1 + 4\epsilon)\lambda_0$ 
8         Let  $y_j \leftarrow y_j + 1$ , and, for  $i \in S_j$ , let  $x_{ij} \leftarrow x_{ij} + 1$ .
9         If  $\min_{i \in C} A_i x \geq U$ , then return  $(x/U, y/U)$ .
10        operation (b): scale  $\lambda_0$   $\triangleright$  assert:  $\min_{j, S} \lambda(x, j, S) \geq (1 + \epsilon)\lambda_0$ 
11        Let  $\lambda_0 \leftarrow (1 + \epsilon)\lambda_0$ .

```

efficiently emulate Alg’s 4 and 5 on the set-cover LP produced by Hochbaum’s reduction, without explicitly building the set-cover LP. Given a facility-location instance (F, C, f, c) , the reduction gives the following LP (A, w) . For each facility $j \in F$ and subset $S \subseteq C$ of customers, there is a variable $x'_{j'}$, with cost $w_{j'} = f_j + \sum_{i \in S} c_{ij}$, where $j' = j'(j, S)$ is unique to the pair (j, S) . For each customer $i \in C$, there is a constraint $A_i x' \geq 1$, where $A_{ij'}$ is 1 if $i \in S$ and 0 otherwise (where $j' = j'(j, S)$). The resulting LP is $\min\{w \cdot x' : x' \in \mathbb{R}_+^\ell, Ax' \geq 1\}$, where $\ell \approx m2^n$.

This LP and the facility-location LP are equivalent: each feasible solution x' to the set-cover LP yields a feasible solution (x, y) of the facility-location LP of the same or lesser cost ($x_{ij} = \sum_{S \ni i} x'_{j'(j, S)}$ and $y_j = \max_i x_{ij}$), and the LPs have the same optimal cost.

Sequential algorithm. Alg. 6 is our sequential algorithm. To prove correctness, we show that it is a valid specialization of Alg. 5 as run on the set-cover LP given by the reduction. During the course of Alg. 6, given the current (x, y) and λ_0 , for any given facility $j \in F$, the following lemma justifies restricting attention to a single canonical subset $S_j = S_j(\lambda_0, x)$ of “nearby” customers.

Lemma 6. *Consider any x, λ_0 , and $i \in F$ during the execution of Alg. 6. Let $\lambda(x, j, S)$ be as defined there and $\lambda'_0 = (1 + \epsilon)\lambda_0$. Then $\min_{S \subseteq C} \lambda(x, j, S) \leq \lambda'_0$ iff $\lambda(x, j, S_j) \leq \lambda'_0$ where $S_j = \{i \in C : c_{ij} < \lambda'_0 a_i(x)\}$.*

Proof. $\lambda(x, j, S) \leq \lambda'_0$ iff $f_j + \sum_{i \in S} (c_{ij} - \lambda'_0 a_i(x)) \leq 0$, so $S = S_j$ is the best set for a given j . \square

Lemma 7. (i) Alg. 6 is a specialization of Alg. 4 on the set-cover LP.

(ii) Alg. 6 can be implemented to run in time $O(N \log(m)/\epsilon^2)$.

Proof. (i) Based on the reduction, doing an increment in Alg. 5 corresponds to choosing a facility $j \in F$ and set $S \subseteq C$, then incrementing y_j and x_{ij} for $i \in S$; the precondition for doing the increment (in Alg. 5) translates to $\lambda(x, j, S) \leq (1 + 4\epsilon)\lambda_0$. When Alg. 6 does an increment for (j, S_j) , this precondition is met because $\lambda(x, j, S_j) \leq (1 + \epsilon)\lambda_0$.

When Alg. 6 scales λ_0 , by inspection, Alg. 6 guarantees $\lambda(x, j, S_j) \geq (1 + \epsilon)\lambda_0$ for all $j \in F$. By Lemma 6 this implies $\lambda(x, j, S) \geq (1 + \epsilon)\lambda_0$ for all $S \subseteq C$, meeting the precondition in Alg. 5.

Algorithm 7 Parallel $(1 + \epsilon)$ -approximation algorithm for facility-location LPs

```

1 function PARALLEL-FACILITY-LOCATION(facilities  $F$ , customers  $C$ , costs  $f, c$ )
2   Define  $U, A_i x, a_i(x)$ , etc. as in Alg. 6, and  $\ell = \max_{i \in C} \min_{j \in F} f_j + c_{ij}$ .
3   Initialize  $x_{ij} \leftarrow \epsilon \ell / (f_j + c_{ij}) |F| |C|$  for  $j \in F, i \in C$ , and  $y_j \leftarrow \sum_{i \in C} x_{ij}$  for  $j \in F$ .
4   Initialize  $\lambda_0 \leftarrow \ell / |a(x)|$ .
5   repeat
6     operation (b): scale  $\lambda_0$ 
7     Let  $\lambda_0 \leftarrow (1 + \epsilon) \lambda_0$ .  $\triangleright$  assertion:  $\min_{j, S} \lambda(x, j, S) \geq (1 + \epsilon) \lambda_0$ 
8     Let  $x \leftarrow \text{TOP-UP}(x)$  (per Lemma 8). Guarantees  $\forall i, j. c_{ij} < \lambda_0 a_i(x)$  only if  $x_{ij} = y_j$ .
9     repeat
10    operation (a): increment  $x$   $\triangleright$  assertion:  $\min_j \lambda(x, j, S_j) \leq (1 + 4\epsilon) \lambda_0$ 
11    Define  $S_j = \{i \in C : c_{ij} < \lambda_0 a_i(x)\}$  and  $J = \{j \in F : \lambda(x, j, S_j) \leq \lambda_0\}$ .
12    For each  $j \in J$ , increase  $y_j$  by  $z y_j$ , and increase  $x_{ij}$  by  $z y_j$  for  $i \in S_j$ ,
13    choosing  $z$  s.t.  $z \max_{i \in C} \sum_{j \in J: i \in S_j} y_j = 1$  (the max. increase in any  $A_i x$  is 1).
14    if  $\min_{i \in C} A_i x \geq U$  then return  $x/U$ .
15  until  $J = \emptyset$ 

```

(ii) To implement the algorithm, maintain $x, y, A_i x$, and $a_i(x)$ for each $i \in C$. Within each iteration of the **for** loop for a given $j \in F$, call the sequence of increments done in the while loop a *run* for j . To do the first increment of each run, compute S_j and $\lambda(x, j, S_j)$ directly (from scratch). Then, in each subsequent increment in the run, update the relevant quantities incrementally: e.g., after doing an increment for (j, S_j) , update $A_i x$ and $a_i(x)$ (for $i \in S_j$ with $A_i x \leq U$) by noting that the increment increases $A_i x$ by 1 and decreases $a_i(x)$ by a factor of $1 - \epsilon$; delete from S_j any i 's that cease to satisfy $c_{ij} < \lambda_0 a_i(x)$.

The time for the run for j is proportional to (A) $|\{i \in C : c_{ij} < \infty\}|$ (the number of possible customers that j might serve), plus (B) the number of times y_j and any x_{ij} 's increase by 1 during the phase. By Thm. 7 (ii), there are $O(U)$ iterations of the outer **repeat** loop, so the total time for (A) is $O(UN)$. Since each y_i and each x_{ij} never exceeds U , the time for (B) is also $O(UN)$. \square

Parallel facility location. Lemma 7 proves Thm. 2 part (i). Next we prove part (ii).

Lemma 8. (i) Algorithm 7 is a $(1 + O(\epsilon))$ -approximation algorithm for facility-location LPs.
(ii) Algorithm 7 has a parallel implementation running in time $O(\log^2(m) \log(N/\epsilon) \log(N)/\epsilon^4)$ and doing work $O(N \log(N/\epsilon) \log(m)/\epsilon^2)$.

Proof. In line 8, TOP-UP(x) does the following for each customer $i \in C$ independently. Consider the facilities $j \in F$ such that $c_{ij} < \lambda_0 a_i(x)$ and $x_{ij} < y_j$, in order of increasing c_{ij} . When considering facility j , increase x_{ij} just until either $x_{ij} = y_j$ or $c_{ij} = \lambda_0 a_i(x)$ (recall $a_i(x) = (1 - \epsilon)^{A_i x}$). (Do this in parallel in $O(\log N)$ time and $O(N)$ work as follows. For each customer i , assume its facilities are presorted by increasing c_{ij} . Raise each x_{ij} by δ_{ij} , computed as follows. Compute the prefix sums $d_j = \sum_{j' \prec_i j} y_{j'} - x_{ij}$ where $j' \prec_i j$ if j' is before j in the ordering. Check for each j whether $c_{ij} < \lambda_0 (1 - \epsilon)^{A_i x + d_j}$. If so, then let $\delta_{ij} = y_j - x_{ij}$; otherwise, if $c_{ij} < \lambda_0 (1 - \epsilon)^{A_i x + d_{j'}}$ where j' is the facility preceding j , then choose δ_{ij} to make $c_{ij} = \lambda_0 (1 - \epsilon)^{A_i x + d_{j'} + \delta_{ij}}$; otherwise, take $\delta_{ij} = 0$.)

(i) We prove that, except for the call to TOP-UP in line. 8, Alg. 7 is a specialization of Alg. 4 on the cover LP. To verify, note that $\ell \leq \text{OPT}(F, C, f, c) \leq |C|\ell$, because the minimum cost to serve any single customer j is $\ell = \min_{j \in F} f_j + c_{ij}$, and each customer can be served at cost at most ℓ . Then (following the proof of correctness of Alg's 5 and 6) Alg. 7 maintains the invariant $\lambda_0 \leq (1 + \epsilon) \text{OPT}(F, C, f, c)/|a(x)|$. (Indeed, initially $\lambda_0 = \ell/|a(x)| \leq \text{OPT}(F, C, f, c)/|a(x)|$, because $\ell \leq \text{OPT}(F, C, f, c)$. Increasing x_{ij} 's and y_j 's only decreases $|a(x)|$, so preserves the invariant. When Alg. 7 increases λ_0 to $(1 + \epsilon)\lambda_0$, by inspection $\min_{j \in F} \lambda(x, j, S_j) > \lambda_0$. By Lemma 6, this ensures $\min_{j \in F, S \subseteq C} \lambda(x, j, S) > \lambda_0$, which by Lemma 10 implies $\lambda_0 < \text{OPT}(F, C, f, c)/|a(x)|$, so the invariant is preserved.) Since $\lambda_0 \leq (1 + \epsilon) \text{OPT}(F, C, f, c)/|a(x)|$, by inspection of the definition of J in Alg. 7, the increment to x and y corresponds to a valid increment in Alg. 4. So, except for the call to TOP-UP in line. 8, Alg. 7 is an implementation of algorithm Alg. 4.

Regarding the call to TOP-UP, we observe that it preserves Invariant 4 in the proof of correctness of Alg. 4. (To verify this, consider any x_{ij} with $x_{ij} < y_j$ and $c_{ij} < \lambda_0 a_i(x)$. Increasing x_{ij} increases $c \cdot x / \text{OPT}$ at rate c_{ij} / OPT which, by the assumption on c_{ij} and the invariant on λ_0 , is at most $(1 + \epsilon)a_i(x)/|a(x)|$. On the other hand, increasing x_{ij} increases $\text{lmin } a(x)$ at rate at least $(1 - O(\epsilon))a_i(x)/|a(x)|$ (see e.g. [27]). Hence, invariant 4 is preserved.) It follows that the performance guarantee from Lemma 4 (i) holds here. Since $\ell \leq \text{OPT}(F, C, f, c)$, the initial solution (x^0, y^0) costs at most $\epsilon \text{OPT}(F, C, f, c)$, so, by that performance guarantee, Alg. 7 returns a cover of cost $(1 + O(\epsilon)) \text{OPT}(F, C, f, c)$. This shows Lemma 8 part (i).

(ii) Call each iteration of the outer loop a *phase*. By Lemma 7 (ii), there are $O(U)$ phases. Consider any phase. In the first iteration of the inner loop, compute all quantities J , $A_i x$, and $a_i(x)$ for each $i \in C$, $\lambda(x, j, S_j)$ etc. directly, from scratch, in $O(N)$ work and $O(\log N)$ time. In each subsequent iteration within the phase, update each changing quantity incrementally (similarly to Alg. 3), doing work proportional to $\sum_{i \in F} |S_j|$, the number of pairs (i, j) where $i \in S_j$.

At the start of the phase, the call to TOP-UP ensures $x_{ij} = y_j$ if $i \in S_j$. Because each S_j decreases monotonically during the phase, this property is preserved throughout the phase. In the choice of z , each sum $\sum_{j \in J: i \in S_j} y_j$ is therefore equal to $\sum_{j \in J: i \in S_j} x_{ij}$, which is less than U (as $a_i(x) = 0$ if the sum exceeds U). Therefore, z is at least $1/U$. Hence, for each $i \in S_j$, the variable x_{ij} increases by at least a $1 + 1/U$ factor in the iteration. On the other hand, at the start of the algorithm $x_{ij} = \epsilon \ell / (f_i + c_{ij}) |F| |C|$, while at the end (by the performance guarantee (i) and $\ell \leq \text{OPT}(f, c)$), $x_{ij} \leq \ell / (f_i + c_{ij})$. Hence, x_{ij} increases by at most a factor of $|F| |C| / \epsilon$ throughout. Hence, the number of iterations in which i occurs in any S_j is $O(\log_{1+1/U} |F| |C| / \epsilon) = O(U \log(|F| |C| / \epsilon)) = O(U \log mn / \epsilon)$. In each such iteration, i contributes to at most $|\{j | c_{ij} < \infty\}|$ pairs. Hence, the total work is $O(NU \log(mn / \epsilon))$.

To bound the time, note that within each of the $O(U)$ phases, there is some pair i', j' such that i' is in $S_{j'}$ in the next-to-last iteration of the phase, and (since the sets J and $S_{j'}$ monotonically decrease during the phase) i' is in $S_{j'}$ in every iteration of the phase. As observed above, i' occurs in $\cup_j S_j$ in at most $O(U \log(mn / \epsilon))$ iterations. Hence, the number of iterations of the inner loop within each phase is $O(U \log(mn / \epsilon))$. Since each iteration can be implemented in $O(\log mn)$ time, (ii) follows. \square

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. *arXiv preprint arXiv:1407.1925*, 2014.

- [2] S. Arora, E. Hazan, and S. Kale. The multiplicative weights update method: A meta-algorithm and applications. *Theory of Computing*, 8:121–164, 2012.
- [3] Y. Azar, U. Bhaskar, L. Fleischer, and D. Panigrahi. Online mixed packing and covering. In *Proceedings of the Twenty-Fourth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 85–100. SIAM, 2013.
- [4] S. Basu and Y. Bresler. $O(N^2 \log^2 N)$ filtered backprojection reconstruction algorithm for tomography. *Image Processing, IEEE Transactions on*, 9(10):1760–1773, 2000.
- [5] D. Bienstock. *Potential function methods for approximately solving linear programming problems: theory and practice*. Kluwer Academic Publishers, 2002.
- [6] D. Bienstock and G. Iyengar. Approximating fractional packings and coverings in $O(1/\epsilon)$ iterations. *SIAM Journal on Computing*, 35(4):825–854, 2006.
- [7] P. Christiano, J. A. Kelner, A. Madry, D. A. Spielman, and S.-H. Teng. Electrical flows, Laplacian systems, and faster approximation of maximum flow in undirected graphs. In *Proceedings of the forty-third annual ACM symposium on Theory of computing*, pages 273–282. ACM, 2011.
- [8] F. A. Chudak and V. Eleutério. Improved approximation schemes for linear programming relaxations of combinatorial optimization problems. In *Integer Programming and Combinatorial Optimization*, pages 81–96. Springer, 2005.
- [9] L. Fleischer. Approximating fractional multicommodity flow independent of the number of commodities. In *Proceedings of the 40th Annual Symposium on Foundations of Computer Science*, page 24. IEEE Computer Society, 1999.
- [10] L. K. Fleischer. Approximating fractional multicommodity flow independent of the number of commodities. *SIAM J. Discrete Math.*, 13(4):505–520, 2000.
- [11] N. Garg and J. Koenemann. Faster and Simpler Algorithms for Multicommodity Flow and other Fractional Packing Problems. *Proceedings of the 39th Annual Symposium on Foundations of Computer Science*, pages 300–309, 1998.
- [12] 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.
- [13] M. X. Goemans and J. A. Kelner. The mathematical work of Daniel Spielman. *Notices of the AMS*, 58(9), 2011.
- [14] M. D. Grigoriadis and L. G. Khachiyan. A sublinear-time randomized approximation algorithm for matrix games. *Operations Research Letters*, 18(2):53–58, 1995.
- [15] D. S. Hochbaum. Heuristics for the fixed cost median problem. *Mathematical programming*, 22(1):148–162, 1982.
- [16] G. Kalai. The work of Daniel A. Spielman. In *Proceedings of the International Congress of Mathematicians*, pages 101–112, 2010. Hyderabad, India.
- [17] A. Kolen and A. Tamir. Covering problems. In R. L. Francis and P. B. Mirchandani, editors, *Discrete Location Theory (chapter 6)*. Wiley & Sons, Inc, 1990. Series in Discrete Mathematics and Optimization.
- [18] C. Koufogiannakis and N. E. Young. Beating Simplex for fractional packing and covering linear programs. In *Foundations of Computer Science, 2007. FOCS'07. 48th Annual IEEE Symposium on*, pages 494–504. IEEE, 2007.
- [19] C. Koufogiannakis and N. E. Young. A nearly linear-time PTAS for explicit fractional packing and covering linear programs. *Algorithmica*, 70(4):648–674, 2014.
- [20] I. Koutis, G. L. Miller, and R. Peng. A fast solver for a class of linear systems. *Communications of the ACM*, 55(10):99–107, 2012.

- [21] M. Luby and N. Nisan. A parallel approximation algorithm for positive linear programming. *Proceedings of the 25th annual ACM symposium on Theory of Computing*, pages 448–457, 1993.
- [22] Y. Nesterov. Rounding of convex sets and efficient gradient methods for linear programming problems. *Optimisation Methods and Software*, 23(1):109–128, 2008.
- [23] D. A. Spielman and S.-H. Teng. Nearly-linear time algorithms for graph partitioning, graph sparsification, and solving linear systems. In *Proceedings of the thirty-sixth annual ACM symposium on Theory of computing*, pages 81–90. ACM, 2004.
- [24] D. A. Spielman and S.-H. Teng. Nearly-linear time algorithms for preconditioning and solving symmetric, diagonally dominant linear systems. *arXiv preprint cs/0607105*, 2006.
- [25] M. J. Todd. The many facets of linear programming. *Mathematical Programming*, 91(3):417–436, 2002.
- [26] N. K. Vishnoi. $Lx = b$: Laplacian solvers and their algorithmic applications. *Foundations and Trends in Theoretical Computer Science*, 8(1-2):1–141, 2012. DOI: 10.1561/04000000054.
- [27] N. E. Young. K -medians, facility location, and the Chernoff-Wald bound. In *Proceedings of the eleventh annual ACM-SIAM symposium on Discrete algorithms*, pages 86–95. Society for Industrial and Applied Mathematics, 2000.
- [28] N. E. Young. Sequential and parallel algorithms for mixed packing and covering. In *Proceedings of IEEE Symposium on Foundations of Computer Science*, pages 538–546, 2001.

5 Appendix

5.1 Mixed packing and covering

Lemma 9. *In Alg. 1, if (P, C) is feasible, then, for any x , $\lambda^*(x) \leq |p(x)|/|c(x)|$.*

Proof. Let x^* be a feasible solution. Since x^* is feasible, it also satisfies $(Cx^*) \cdot c(x)/|c(x)| \geq 1 \geq (Px^*) \cdot p(x)/|p(x)|$, that is, $x^* \cdot (C^\top c(x)/|c(x)| - P^\top p(x)/|p(x)|) \geq 0$. Hence there exists a j such that $C_j^\top c(x)/|c(x)| - P_j^\top p(x)/|p(x)| \geq 0$, which is equivalent to $\lambda(x, j) \leq |p(x)|/|c(x)|$. \square

Reminder of Lemma 1. *Given any feasible instance (P, C) , Alg. 1*

(i) returns x such that $Cx \geq 1$ and $Px \leq 1 + O(\epsilon)$,

(ii) does step (a) at most $O(U)$ times, and

(iii) does step (b) at most $O(mU)$ times, where $U = O(\log(m)/\epsilon^2 + \max_i P_i x^0/\epsilon)$.

Proof. (i) Fix (P, C) . First observe that the algorithm maintains the invariant $\lambda_0 \leq |p(x)|/|c(x)|$. The initial choice of λ_0 guarantees that the invariant holds initially. By inspection, Line 13 is executed only when $\lambda^*(x) = \min_j \lambda(x, j) \geq (1 + \epsilon)\lambda_0$. This and Lemma 9 ($\lambda^*(x) \leq |p(x)|/|c(x)|$) imply that $(1 + \epsilon)\lambda_0 \leq |p(x)|/|c(x)|$, so that the invariant is maintained.

Define $\text{lmax } p(x) = \log_{1+\epsilon} \sum_i p_i(x)$ and $\text{lmin } c(x) = \log_{1-\epsilon} \sum_i c_i(x)$ for p and c as defined in the algorithm. We show that the algorithm maintains the invariant

$$(1 + O(\epsilon)) \text{lmin } c(x) - \log_{1-\epsilon} m \geq (1 - O(\epsilon)) (\text{lmax } p(x) - \max_i P_i x^0/\epsilon - \log_{1+\epsilon} m), \quad (3)$$

where x_0 is the initial solution given to the algorithm. By inspection, the invariant is initially true. In a given execution of step (ii), let x be as at the start of the step; let index j be the one chosen for the step. The step increases $\text{lmax } p(x)$ by at most $(1 + O(\epsilon)) \sum_j \delta_j P_j p(x)/|p(x)|$; it increases $\text{lmin } c(x)$ by at least $(1 - O(\epsilon)) \sum_j \delta_j C_j c(x)/|c(x)|$ (see e.g. [27]). Since $\delta_j > 0$ only if $\lambda(x, j) \leq (1 + \epsilon)\lambda_0$, the first invariant $\lambda_0 \leq |p(x)|/|c(x)|$ implies that the invariant is maintained.

Consider the step when the algorithm returns x/U . Just before the step, at least one $i \in [m]$ had $C_i x < U$, so $\text{lmin } c(x) < \log_{1-\epsilon}(1-\epsilon)^U = U$, and by the above invariant $\max_i P_i x \leq (1 + O(\epsilon))U$. During the step $\max_i P_i x$ increases by at most $1 = O(\epsilon U)$, so after the step $\max_i P_i x \leq (1 + O(\epsilon))U$ still holds. Part (i) follows.

(ii) The algorithm maintains $\lambda_0 \leq |p(x)|/|c(x)|$, with equality at the start. Each time λ_0 increases, it does so by a $1 + \epsilon$ factor, but throughout, $P_i x = O(U)$ and $\min_i C_i x = O(U)$, so $|p(x)|/|c(x)|$ is always at most $m(1 + \epsilon)^{O(U)}/(1 - \epsilon)^{O(U)}$. Part (ii) follows.

(iii) Each increment either increases some $P_i x$ by at least $1/2$, or increases some $C_i x$ by at least $1/2$ where $C_i x \leq U$. Since $P_i x = O(U)$ throughout, part (iii) follows. \square

Reminder of Lemma 2. *Given any feasible packing/covering instance (P, C) , provided the updates in lines 12–13 maintain Invariant (1):*

- (i) *Each operation (a) or (b) done by Alg. 2 is a valid operation (a) or (b) of Alg. 1, so*
- (ii) *Alg. 2 returns a $(1 + O(\epsilon))$ -approximate solution.*

Proof. (i) Alg. 2 only does increments for a given j when $\widehat{\lambda}_j \leq (1 + \epsilon)^2 \lambda_0 / (1 - \epsilon)$, which (with the definition of $\widehat{\lambda}_j$ and $\epsilon \leq 1/10$) ensures $\lambda(x, j) \leq (1 + 4\epsilon)\lambda_0$. Likewise, Alg. 2 only scales λ_0 when $\min_j \widehat{\lambda}_j > (1 + \epsilon)^2 \lambda_0 / (1 - \epsilon)$, which (with the guarantee) ensures $\min_j \lambda(x, j) \geq (1 + \epsilon)\lambda_0$. Thus, the precondition of each operation is appropriately met. (Alg. 2's termination condition is slightly different than that of Alg. 1, but this does not affect correctness.) This proves (i). Part (ii) follows from Lemma 1 part (i). \square

5.2 Covering

Lemma 10. *In Alg. 4, for any x , $\text{OPT}(A, w) \geq |a(x)| \lambda^*(x)$, where $\lambda^*(x) = \min_{j \in [n]} \lambda(x, j)$.*

Proof. Let x^* be a solution of cost $w \cdot x^* = \text{OPT}(A, w)$. Draw a single $j \in [n]$ at random from the distribution $x^*/|x^*|$. By calculation the expectation of the quantity $A_j^\top a(x)/|a(x)| - w_j/(w \cdot x^*)$ is proportional to $(Ax^*)^\top a(x)/|a(x)| - x^* \cdot w/(w \cdot x^*)$, which is non-negative (as $Ax^* \geq 1$), so with positive probability the quantity is non-negative, implying $w \cdot x^* \geq |a(x)| \lambda(x, j)$. \square

Reminder of Lemma 4. *Alg. 4 returns a solution x such that $w \cdot x \leq (1 + O(\epsilon)) \text{OPT}(A, w) + w \cdot x^0$, where x^0 is the initial solution given to the algorithm.*

Proof. First we observe that the algorithm is well-defined. In each iteration, by definition of $\lambda^*(x)$, there exists a $j \in [n]$ such that $\lambda(x, j) = \lambda^*(x)$. By Lemma 10, for this j , $\lambda(x, j) \leq \text{OPT}(A, w)/|a(x)|$. So, in each iteration there exists a suitable vector $\delta \in \mathbb{R}_+^n$. Next we prove the approximation ratio.

Define $\text{lmin } a(x) = \log_{1-\epsilon} |a(x)|$ for a as defined in the algorithm. We show that the algorithm maintains the invariant

$$(1 + O(\epsilon))(\text{lmin } a(x) - \log_{1-\epsilon} m) \geq \frac{w \cdot x - w \cdot x^0}{\text{OPT}(A, w)}. \quad (4)$$

The invariant is initially true by inspection. In a given iteration of the algorithm, let x be as at the start of the iteration, let vector δ be the one chosen in that iteration. The iteration increases $w \cdot x / \text{OPT}(A, w)$ by $\sum_j \delta_j w_j / \text{OPT}(A, w)$. It increases $\text{lmin } a(x)$ by at least $(1 - O(\epsilon)) \sum_j \delta_j A_j^\top a(x)/|a(x)|$ (see e.g. [27]). By the choice of δ , the definition of λ^* , and

Lemma 10, if $\delta_j > 0$ then $w_j / \text{OPT}(A, w) \leq (1 + O(\epsilon))A_j^\top a(x) / |a(x)|$, so the invariant is maintained.

Before the last iteration, at least one i has $A_i x \leq U$, so $\text{lmin cov}(x) \leq \log_{1-\epsilon}(1-\epsilon)^U = U$. This and the invariant imply that finally $(w \cdot x - w \cdot x^0) / \text{OPT}(A, w) \leq 1 + (1 + O(\epsilon))U + \log_{1-\epsilon} m$. By the choice of U this is $(1 + O(\epsilon))U$. \square

Reminder of Lemma 7. (i) Alg. 5 is a specialization of Alg. 4 and (ii) scales λ_0 $O(U) = O(\log(m)/\epsilon^2)$ times.

Proof. (i) Observe that the algorithm maintains the invariant $\lambda_0 \leq \text{OPT}(A, w) / |a(x)|$. The invariant is true for the initial choice of λ_0 because the minimum cost to satisfy just a single constraint $A_i x \geq 1$ is $\min_{j \in [n]} w_j / A_{ij}$. Scaling λ_0 only decreases $|a(x)|$, so maintains the invariant. increment is done only when $(1 + \epsilon)\lambda_0 \leq \lambda^*(x)$, which by Lemma 10 is at most $\text{OPT}(A, w) / |a(x)|$, so increment also preserves the invariant. Since the algorithm maintains this invariant, it is a special case of Alg. 4 with $\text{cost}(x^0) = 0$.

(ii) One way to satisfy every constraint $A_i x \geq 1$ is as follows: for every i , choose j minimizing w_j / A_{ij} then add $1 / A_{ij}$ to x_j . The cost of this solution is at most $m\lambda_0$. Hence, the initial value of λ_0 is at least $m^{-1} \text{OPT}(A, w) / |a(x)|$. At termination (by the invariant from part (i) above) λ_0 is at most $\text{OPT}(A, w) / |a(x)|$. Also, $|a(x)|$ decreases by at most a factor of $m / (1 - \epsilon)^U$ during the course of the algorithm, while each scaling of λ_0 increases λ_0 by a factor of $1 + \epsilon$. It follows that the number of scalings is at most $\log_{1+\epsilon} m^2 / (1 - \epsilon)^U = O(U)$. \square

5.3 Facility location

Hochbaum's reduction [15]. For every “star” (j, C) , where j is a facility and C is a subset of clients j might serve, create a set C_j containing those clients, whose cost is the cost of opening facility j and using j to serve the customers in C . There are exponentially many sets.

More efficient reduction [17, §6.2]. For every facility j , create a set F_j with cost equal to the cost of opening j . For every client i and facility j , create an element (i, j) and a set $S_{ij} = \{(i, j)\}$. For every i , let \prec_i order the facilities j by increasing distance from i , breaking ties arbitrarily, and make F_j be $\{(i, k) : j \prec_i k\}$. Give S_{ij} cost equal to the distance $d(i, j)$ from i to j , minus the distance $d(i, j')$ to i 's next closest facility j' , if any. For intuition, note that if a set F_j is chosen, then (for any given i) that set covers $\{(i, k) : j \prec_i k\}$, while i 's remaining elements can be covered using the sets $\{S_{ik} : k \preceq_i j\}$ at total cost $d(i, j)$. The resulting LP can have $\Omega(nm^2)$ non-zeros.