Matrix Scaling by Network Flow

Günter Rote* Martin Zachariasen†

Abstract

A given nonnegative $n \times n$ matrix $A = (a_{ij})$ is to be scaled, by multiplying its rows and columns by unknown positive multipliers λ_i and μ_j, such that the resulting matrix $(a_{ij}\lambda_i\mu_j)$ has specified row and column sums r_i and s_j.

We give an algorithm that achieves the desired row and column sums with a maximum absolute error ε in $O(n^4(\log n + \log \frac{h}{\varepsilon}))$ steps, where h is the overall total of the result matrix.

Our algorithm is a scaling algorithm. It solves a sequence of more and more refined discretizations. The discretizations are minimum-cost network flow problems with convex piecewise linear costs. These discretizations are interesting in their own right because they arise in proportional elections.

1 Introduction

The Matrix Scaling Problem. The input of the (real-valued) *matrix scaling problem* is a non-negative $n \times n$ matrix $A = (a_{ij})$, and two positive n-vectors $r = (r_i)$ and $s = (s_j)$. The problem is to find positive multipliers λ_i for scaling the rows and μ_j for scaling the columns, such that the resulting matrix $(a_{ij}\lambda_i\mu_j)$ has the specified values r_i and s_j as row and column sums:

$$(1) \qquad \sum_{j=1}^{n} a_{ij}\lambda_i\mu_j = r_i, \qquad \sum_{i=1}^{n} a_{ij}\lambda_i\mu_j = s_j$$

Of course, we must have $\sum_{i=1}^{n} r_i = \sum_{j=1}^{n} s_j$. We denote this common value by h, the *total sum* of the desired matrix.

A special case is when all r_i and s_j are 1. This is the *doubly stochastic* scaling problem. The problem also makes sense for rectangular matrices, but for simplicity, we restrict our attention to the square case.

The problem (1) is a non-linear system of equations, and there is no hope to solve it exactly in reasonable time. Thus we settle for an ε-approximate solution:

$$(2) \quad \left|\sum_{j=1}^{n} a_{ij}\lambda_i\mu_j - r_i\right| \le \varepsilon, \text{ and } \left|\sum_{i=1}^{n} a_{ij}\lambda_i\mu_j - s_j\right| \le \varepsilon$$

*Freie Universität Berlin, Institut für Informatik, Takustraße 9, 14195 Berlin, Germany, rote@inf.fu-berlin.de.

†University of Copenhagen, Department of Computer Science, Universitetsparken 1, DK-2100 Copenhagen Ø, Denmark, martinz@diku.dk.

Applications. There is an abundance of literature about the problem, see for example Bacharach [1] or Rothblum and Schneider [11] and the references given there. Linial, Samorodnitsky, and Wigderson [10] mention such diverse applications as statistics and image reconstruction. In numerical analysis, matrix scaling is used as a simple preconditioning operation to improve the numerical stability for solving linear equations. Linial et al. [10] have used doubly stochastic matrix scaling as a tool for approximating the permanent of a nonnegative matrix with a multiplicative error of e^n.

A fast *parallel* algorithm for doubly stochastic matrix scaling (with $\varepsilon = 1/n$) could be used for the bipartite matching problem, which is a major open problem in parallel complexity, see [10, Section 5.1] or [5].

Previous Results. The problem (1) has a solution if and only if the system

$$(3) \qquad \sum_{j=1}^{n} x_{ij} = r_i, \qquad \sum_{i=1}^{n} x_{ij} = s_j$$

has a solution (x_{ij}) with the same sign structure as A: $x_{ij} > 0$ if $a_{ij} > 0$ and $x_{ij} = 0$ otherwise. This condition can be tested in polynomial time because it is just a network flow problem, see [11].

If a solution exists, the values $a_{ij}\lambda_i\mu_j$ are unique. Obviously, one has the freedom of shifting a common factor between the λ_i's and the μ_j's.

There are several different well-known formulations of the matrix scaling problem as a nonlinear optimization problem: the minimum of

$$(4) \qquad \sum_{i,j:a_{ij}>0} x_{ij}\Big(\log\frac{x_{ij}}{a_{ij}} - 1\Big)$$

under the constraints (3) yields the solution $x_{ij} = a_{ij}\lambda_i\mu_j$ of the continuous matrix scaling problem (1). The dual problem is an unconstrained convex optimization problem.

$$\text{minimize} \sum_{i,j:a_{ij}>0} a_{ij} \cdot e^{\alpha_i+\beta_j} - \sum_i r_i\alpha_i - \sum_j s_j\beta_j$$

It yields the correct multipliers by the relations $\lambda_i := e^{\alpha_i}$ and $\mu_j := e^{\beta_j}$.

Algorithms. The original paper by Sinkhorn [12] introduced the simple algorithm of alternatively scaling the rows and columns to the desired sums, and he proved (originally only for the case of doubly stochastic scaling and for a positive matrix A) that the process converges to the solution. Relatively little work has been done on the complexity of computing a solution. Most of this research concentrated on investigating the convergence of the *Sinkhorn scaling* approach. If a solution exists, convergence to the correct solution is linear (i. e., the error decreases geometrically), but the rate of convergence deteriorates when entries of the scaled matrix approach zero [14].

For doubly stochastic scaling, an algorithm of Kalantari and Khachian [8] uses the ellipsoid method and takes $O(\log R(A) \cdot n^4 \log \frac{n}{\varepsilon})$ arithmetic operations, where $R(A)$ is the ratio between the largest and the smallest positive entry in the matrix. The only strongly polynomial running time so far is achieved by Linial et al. [10]. It takes $O(n^7 \log \frac{h}{\varepsilon})$ operations.[1] For the case of doubly stochastic scaling, they have a different algorithm which takes $O(n^3 \log n/\varepsilon^2)$ operations.[2]

The Integer Matrix Scaling Problem. In the integer version of the matrix scaling problem, the resulting values are rounded to integers:

$$\sum_{j=1}^{n}[a_{ij}\lambda_i\mu_j] = r_i, \qquad \sum_{i=1}^{n}[a_{ij}\lambda_i\mu_j] = s_j \tag{5}$$

Here, $[\cdot]$ denotes the rounding-down operation: $x = [u]$ denotes an integer value with

$$u - 1 \le x \le u, \tag{6}$$

i. e., $x = \lfloor u \rfloor$ if u is not integral. If u is integral, there is a *tie*: $x = u$ or $x = u - 1$. Other rounding functions, like rounding to the nearest integer, can be handled as well, but we stick to this simple rounding function in order to simplify the discussion.

This problem arises in *biproportional voting systems*, where the allocated seats are classified by two independent categories: the total number of seats s_j in each *district* j is specified in advance, according to the population; and the total number of seats r_i for each *party* i is calculated in proportion to the total vote for that party. Moreover, the seats should be proportional both in rows (parties) and columns (districts). Such a model (with a different rounding function: rounding to the nearest integer) was recently used for electing the 125 members of the municipal council of the city of Zurich on February 12, 2006.[3]

Previous Results on the Integer Matrix Scaling Problem. The biproportional scaling problem with rounding was first investigated by Balinski and Demange [2]. The conditions for the existence of a solution of the system (5) are the same as for the real version (1). The solution values $x_{ij} = [a_{ij}\lambda_i\mu_j]$ are unique except for the presence of ties.[4] (The multipliers λ_i and μ_j can of course vary in a certain range where they don't modify the rounded values x_{ij}.)

Balinski and Demange [2] have given an algorithm that computes a solution in $O(n^3 h)$ time (see also [3]). With hindsight, this algorithm can be interpreted as a special instance of a standard primal-dual algorithm for our network flow model in Section 2.

Gaffke and Pukelsheim [6] have investigated the relation between the primal and dual version of these problems (with generalized rounding functions) which is expressed in Theorem 1 below.

Our Contribution. We observe that the formulation (7–11) for the *integer* matrix scaling problem can in fact be modeled as a standard minimum-cost network flow problem. This opens the way to use any of numerous alternative methods for network flow problems, leading to better runtimes (Theorems 2 and 3 in Sections 2 and 3).

We then apply our result as an approximation algorithm for the *real-number* matrix scaling problem in a straightforward way, leading to a simpler and faster algorithm (Theorem 4 in Section 4).

By applying the algorithm of Karzanov and McCormick [9] for convex-cost flows, we can solve the real-number scaling problem with integral row and column sums in $O(n^3 \log n \cdot (\log \frac{h}{\varepsilon} + \log n + \log\log R(A))$ time, where $R(A)$ is the ratio between the largest and the smallest positive entry in the matrix A (Theorem 6 in Section 6).

[1] The running time claimed in [10] has an extra factor of $\log n$ which comes from their Theorem 4.3. We don't see how this factor arises in the proof.

[2] This result does not appear explicitly in [10]. A different definition of the error is used there: the squared L_2 norm ε' of the discrepancies in the row and column sums. The time bound is $O(n^3 \log n/\varepsilon')$, and it is stated only for $\varepsilon' = 1/(n \log n)$. An $O(n^5 \log n)$ additive term for preprocessing in the algorithm can be reduced to $O(n^3)$, see footnote 5.

[3] See `http://daten.wahlen.stzh.ch/`

[4] Some occasional ambiguity as in (6) is inherent in the model, for otherwise there might be no solution: Consider the 2×2 problem where all entries of A are identical, and the row and column sums are 1. The two possible solutions are $X = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ and $X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, but no "consistent" way of rounding fractions will produce these matrices, whatever the choice of λ_i and μ_j might be.

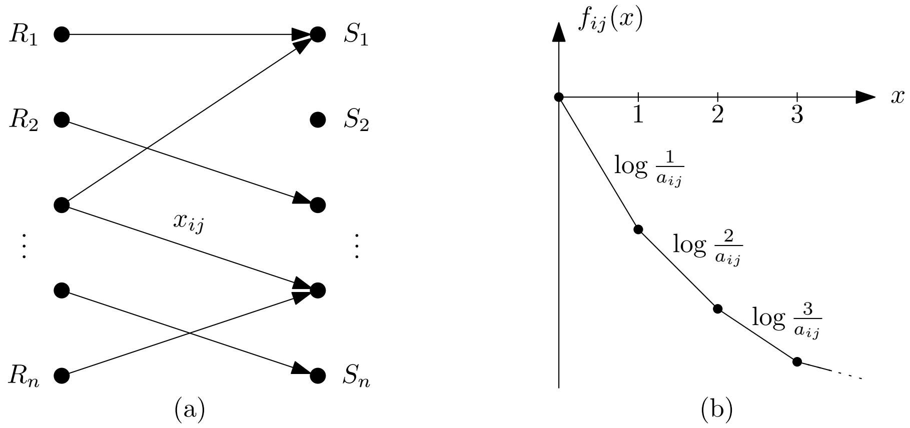

Figure 1: (a) The graph $G(A)$. (Only a few arcs are shown.) (b) A schematic example of the piecewise linear cost function $f_{ij}(x)$. Between successive breakpoints, the function makes vertical steps of size $\log \frac{t}{a_{ij}}$, for $t = 1, 2, 3, \ldots$

2 A Network Flow Model for Matrix Scaling

The optimization formulation presented in this section has been suggested by Gaffke and Pukelsheim [6].

The following constraints (8–11) define a network flow problem on the bipartite graph $G(A)$ with n source vertices R_i and n sink vertices S_j, see Figure 1a.

$$\text{(7)} \qquad \text{minimize} \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t=1}^{x_{ij}} \log \frac{t}{a_{ij}}$$

$$\text{(8)} \qquad \text{subject to} \sum_{j=1}^{n} x_{ij} = r_i \text{ for } i = 1, \ldots, n$$

$$\text{(9)} \qquad \sum_{i=1}^{n} x_{ij} = s_j \text{ for } j = 1, \ldots, n$$

$$\text{(10)} \qquad x_{ij} \geq 0,\ x_{ij} \text{ integer}$$

$$\text{(11)} \qquad x_{ij} = 0 \text{ if } a_{ij} = 0$$

The variable x_{ij} is the flow value in arc (R_i, S_j). If $a_{ij} = 0$, there is no arc (R_i, S_j). The cost function for an arc with $a_{ij} > 0$,

$$f_{ij}(x) := \sum_{t=1}^{x} \log \frac{t}{a_{ij}}$$

is piecewise linear and convex, as illustrated in Figure 1b.

A piecewise linear convex cost function can be modeled by linear cost functions on a sequence of parallel arcs: between A_i and B_j, we introduce parallel arcs with capacity 1 and cost coefficients $c_{ij}^t = \log \frac{t}{a_{ij}}$, for $t = 1, 2, \ldots$. Clearly, a flow of some integral value x from R_i to S_j will fill the x cheapest arcs, and hence the total cost equals $f_{ij}(x)$.

This is then a classical minimum-cost network flow problem, and the standard theory can be applied. In particular, we know that there is an integral optimum solution if there is a solution at all.

The following connection to the matrix scaling problem was proved in [6, Theorem 8.1], for more general rounding functions, without exploiting the connection to network flows, however.

Theorem 1. *An optimum solution $X = (x_{ij})$ of the flow problem* (7–11) *is a solution $x_{ij} = [a_{ij}\lambda_i\mu_j]$ of the integer matrix scaling problem* (5).

Proof. For an optimum solution (x_{ij}), there are dual variables α_i, β_j for which all reduced costs of arcs in the residual network are nonnegative. In particular, looking at the forward arcs (R_i, S_j), we have the reduced costs

$$c_{ij}^t - \alpha_i - \beta_j = \log \frac{t}{a_{ij}} - \alpha_i - \beta_j \geq 0, \text{ for all } t \geq x_{ij} + 1,$$

which is equivalent to

$$\log \frac{x_{ij}+1}{a_{ij}} - \alpha_i - \beta_j \geq 0$$

or

$$x_{ij} + 1 \geq a_{ij} \cdot e^{\alpha_i} e^{\beta_j}.$$

Similarly, from the backward arcs (S_j, R_i), we have $-c_{ij}^t + \alpha_i + \beta_j \geq 0$ for $t \leq x_{ij}$, or

$$x_{ij} \leq a_{ij} \cdot e^{\alpha_i} e^{\beta_j}.$$

With $\lambda_i := e^{\alpha_i}$ and $\mu_j := e^{\beta_j}$ this means that x_{ij} is the correctly rounded value of $a_{ij}\lambda_i\mu_j$:

$$\text{(12)} \qquad a_{ij}\lambda_i\mu_j - 1 \leq x_{ij} \leq a_{ij}\lambda_i\mu_j \qquad \square$$

Thus, we can solve the integer matrix scaling problem by solving a minimum-cost network flow problem with any of the standard algorithms. A simple algorithm is the least-cost augmenting path method, cf. [4, Section 4.3]: start with the zero flow and successively augment the flow along minimum-cost augmenting paths, using the labels from the shortest path computation to update dual variables α_i and β_j.

In order not to distract from the algorithm, we assume in this section that all calculations (including logarithms and exponentials) are computed exactly. We will analyze the necessary precision in Section 5.

For the row sums and column sums of a matrix $X = (x_{ij})$ we use the notations $x_{i.} := \sum_{j=1}^n x_{ij}$ and $x_{.j} := \sum_{i=1}^n x_{ij}$. We define

$$\Delta(X) := \sum_{i=1}^n |x_{i.} - r_i| + \sum_{j=1}^n |x_{.j} - s_j|$$

as the total *discrepancy* from the desired row and column sums.

LEMMA 1. *Suppose we have a set of values α_i, β_j and a set of integer values x_{ij} with*

$$\log x_{ij} \le \log a_{ij} + \alpha_i + \beta_j \le \log(x_{ij} + 1) \tag{13}$$

and $\Delta(X) > 0$. Then, in $O(n^2)$ time, we can find a new solution where Δ is decreased by 2, or we establish that no solution with smaller Δ exists, by solving a single-source shortest path problem. With the new solution, we also find new values α_i, β_j such that (13) is maintained.

Proof. (cf. [4, Lemma 4.16]) Suppose w.l.o.g. that there is a row i in which $x_{i.} < r_i$ holds. (The case $x_{i.} > r_i$ is similar.) We compute a shortest path in the residual network, starting from the row vertex R_i and ending at any column vertex S_j whose in-flow is smaller than the demand s_j, using the non-negative reduced costs $\bar{c}^t_{ij} = c^t_{ij} - \alpha_i - \beta_j$ and $\bar{c}^t_{ji} = -c^t_{ij} + \alpha_i + \beta_j$. Such a shortest path can be computed in $O(n^2)$ time by Dijkstra's algorithm. The shortest distance labels can be used to update the dual variables α_i, β_j such that (13) is maintained. If no path is found, it means that there is no possibility to increase $x_{i.}$. Dijkstra's algorithm only considers at most one forward and one backward arc for each pair of nodes, and not all parallel arcs. To achieve the claimed running time, we have to start the shortest path computation simultaneously at all rows with $x_{i.} < r_i$. Then, if no path is found, the discrepancy cannot be decreased. □

The total desired flow is h. Incrementing the flow in steps of 1 unit leads to h shortest path computations. Thus we get:

THEOREM 2. *The integer matrix scaling problem can be solved in $O(n^2 h)$ steps.* □

3 Scaling the Data

The pseudo-polynomial complexity of Theorem 2 can be improved to a polynomial complexity by a standard scaling approach. (Here, the term *scaling* is used in a different sense than in the name of the problem, *matrix scaling*.) For the general problem of convex minimum-cost flows, a similar scaling algorithm with the same running time is sketched in Hochbaum [7, Chapter 4].

We solve a sequence of more and more refined discretizations. For an integer $k \ge 0$ we define the modified problem P^k with $\hat{r}_i := \lfloor r_i/2^k \rfloor$ and $\hat{s}_j := \lfloor s_j/2^k \rfloor$. Since $\sum \hat{r}_i = \sum \hat{s}_j$ is no longer guaranteed, we have to modify the concept of a solution for such a problem instance. (Even if $\sum \hat{r}_i = \sum \hat{s}_j$, a solution of P^k in the sense of (3) might not exist, due to the presence of zeros in A.) A *solution* x_{ij} with $x_{ij} = [a_{ij}\lambda_i\mu_j]$ must satisfy

$$\sum_{j=1}^n x_{ij} \le r_i \quad \text{and} \quad \sum_{i=1}^n x_{ij} \le s_j, \tag{14}$$

while achieving the maximum value of $\sum_{i=1}^n \sum_{j=1}^n x_{ij}$ (i.e., the minimum discrepancy) under the constraints (14).

We measure the violation of (14) by the *positive discrepancy*

$$\Delta^+(X) := \sum_{i=1}^n \max\{x_{i.} - r_i, 0\} + \sum_{j=1}^n \max\{x_{.j} - s_j, 0\}$$

The following lemma deals with the reduction of $\Delta^+(X)$, in analogy with Lemma 1.

LEMMA 2. *Suppose we have a set of values α_i, β_j and a set of integer values x_{ij} fulfilling (13), with $\Delta^+(X) > 0$. Then, in $O(n^2)$ time, we can find a new solution where Δ^+ is decreased by 1 or 2, or we establish that no solution with smaller Δ^+ exists, by solving a single-source shortest path problem. With the new solution, we also find new values α_i, β_j such that (13) is maintained. The discrepancy Δ is not increased in this process.* □

LEMMA 3. *Suppose that the original problem P^0 is feasible.*

1. *Then the problem P^k has a solution X with $\Delta(X) \le 2n^2$.*
2. *Let λ_i, μ_j and $x_{ij} = [a_{ij}\lambda_i\mu_j]$ be a solution to a problem instance P^k ($k \ge 1$). Then the values*

$x'_{ij} := [2a_{ij}\lambda_i\mu_j]$ satisfy the row and column sum requirements for the problem instance P^{k-1} with a discrepancy $\Delta(X') \le 6n^2$.

Proof. 1. Let (x^0_{ij}) be a solution to the original problem. Then $\hat{x}_{ij} := \lfloor x^0_{ij}/2^k \rfloor$ satisfies (14) for problem P^k, and an easy calculation shows that $\Delta(\hat{X}) \le 2n^2$.

2. By part 1, the solution X has $\Delta(X) \le 2n^2$. It is easy to check that $x'_{ij} = 2x_{ij}$ or $x'_{ij} = 2x_{ij}+1$. (If there is a tie in defining x'_{ij}, we can make an appropriate choice.) The new row sum requirements are $\hat{r}'_i = 2\hat{r}_i$ or $\hat{r}'_i = 2\hat{r}_i + 1$, and similarly for the column sums. Thus, the discrepancy in each row is twice the old discrepancy plus at most n. For all rows and columns, this adds up to $\Delta(X') \le 2\Delta(X) + 2n \cdot n \le 6n^2$. □

We thus solve the sequence P^k, P^{k-1}, … of more and more refined problems until the original problem P^0 is solved. We multiply the input matrix A by a scalar such that the largest entry is $\frac{1}{2}$, without changing the problem. Then, for $k := 1 + \lceil \log_2 h \rceil$, we have $\hat{r}_i = \hat{s}_j = 0$, and with $\lambda_i = \mu_j = 1$ (or $\alpha_i = \beta_j = 0$), we already have $x_{ij} = 0$ as a valid solution for P^k.[5] Going from P^k to P^{k-1} incurs $O(n^2)$ shortest path computations, by Lemma 3: We first have to ensure that no row or column sum exceeds the desired value (14). This might involve at most $6n^2$ flow changes according to Lemma 2, and and does not increase the discrepancy above $6n^2$. Then with at most $3n^2$ additional applications of Lemma 1, $\Delta(X)$ is decreased to the minimum. (If this minimum is bigger than $2n^2$, we can abort since, by Lemma 3.1, there can then be no feasible solution for the original problem.) Thus, the transition from P^k to P^{k-1} takes $O(n^4)$ time, and we have:

THEOREM 3. *The integer matrix scaling problem can be solved in $O(n^4 \log h)$ steps.* □

4 Scaling of Real Matrices

The application to the approximate scaling of real matrices (2) is straightforward. To get a starting solution, we multiply all row and column sums r_i and s_j by the constant $F := \frac{n}{\varepsilon}$ and round them down to integers $\hat{r}_i$ and $\hat{s}_j$, respectively. We get the total sum $H \le hn/\varepsilon$. We solve the integer scaling problem for these data, using Theorem 2 or 3, in $O(n^2 H)$ or $O(n^4 \log H)$ time, respectively. In the end, we have to use a modification of Lemma 1 to get a solution in the sense of (14) in which the *maximum* discrepancy $\Delta_{\max}(X) := \max_{i,j}\{\, \hat{r}_i - x_{i.}, \hat{s}_j - x_{.j} \,\}$ is as small as possible. Taking the real-valued solution of (1), scaling it by F, and rounding everything down, we see that an integer solution with $\Delta_{\max} \le n-1$ exists. Let us now estimate the error in (1), looking for example at row i:

$$\sum_{j=1}^{n} a_{ij}\lambda_i\mu_j - Fr_i = \left(\sum_{j=1}^{n}[a_{ij}\lambda_i\mu_j] - \hat{r}_i\right) + \sum_{j=1}^{n}\big(a_{ij}\lambda_i\mu_j - [a_{ij}\lambda_i\mu_j]\big) + (\hat{r}_i - Fr_i) \tag{15}$$

The first term in parentheses is between $-n+1$ and 0, by the bound on $\Delta_{\max}$ for the solution $x_{ij} = [a_{ij}\lambda_i\mu_j]$; the second sum is between 0 and n, and the last term is between -1 and 0. This accumulates to a total error between $-n$ and $+n$ in each row sum, and similarly in each column sum. In terms of the original (unscaled) data, this is an absolute error of at most ε.

THEOREM 4. *The (real-valued) matrix scaling problem can be solved to an accuracy ε in $O(n^3 \frac{h}{\varepsilon})$ or in $O(n^4(\log n + \log \frac{h}{\varepsilon}))$ time.* □

The algorithms are simple, and they beat the best previous algorithms of Linial et al. [10] by several orders of magnitude. Only for the special case of doubly stochastic scaling (where $h = n$) and for moderate error bounds ε, the $O(n^3 \log n/\varepsilon^2)$ algorithm in [10], which is just Sinkhorn scaling with a special preprocessing step, is faster. It might be interesting to investigate hybrid algorithms that combine different approaches in various phases.

5 Numerical Accuracy

For the quantities related to the costs, we use fixed-point arithmetic with fixed precision, i. e. all numbers $\log a_{ij}$, $\log x_{ij}$, α_i and β_j are approximately represented as integer multiples of some small value δ, with an absolute error of $O(\delta)$. This means that, after approximating the "input data", shortest paths are computed exactly, as they involve only additions and comparisons.

A nice feature of the flow problem (7–11) is that we can use Lemma 1 without propagating errors. We can take *arbitrary* values α_i and β_j and define x_{ij} by $x_{ij} := \lfloor a_{ij} \cdot e^{\alpha_i} e^{\beta_j} \rfloor$ or more precisely, by the relation (13), These values can be used as a starting solution for Lemma 1.

Let us analyze what happens if (13) is only satisfied approximately:

$$\overline{\log}\, x_{ij} \le \overline{\log}\, a_{ij} + \alpha_i + \beta_j \le \overline{\log}(x_{ij}+1) \tag{16}$$

[5] We could have defined the scaled problems by rounding upward: $\hat{r}_i := \lceil r_i/2^k \rceil$, etc. Then the first problem $P^{1+\lceil \log_2 h \rceil}$ has all row and column sums $\hat{r}_i$ and $\hat{s}_j$ equal to 1: This is an assignment problem with cost matrix $\log a_{ij}$. Incidentally, this problem is precisely equivalent to the problem that is solved as the preprocessing step in the first algorithm of Linial et al. [10, Theorem 3.1].

where $\overline{\log} x$ is an approximation of $\log x$. Note that $\overline{\log}(x_{ij}+1)$ is a different (approximate) calculation from $\overline{\log} x_{ij}$, but all computations of the same number $\overline{\log} x$ must yield the same approximate result. Conceptually we locate the value $\overline{\log} a_{ij} + \alpha_i + \beta_j$ in the sorted list of approximate values $\overline{\log} 1$, $\overline{\log} 2$, $\overline{\log} 3$, ... To be specific, let us use natural logarithms in this section unless otherwise stated. Suppose that all values x_{ij} that occur in the computation are bounded by some bound M. Then, if we calculate $\overline{\log} x_{ij}$ with an absolute error $\le 0.1/M$ we may incur an error of approximately ± 0.1 in terms of x_{ij}. In other words, (13) holds for some *approximation* $\tilde{x}_{ij}$ instead of the integer value x_{ij}, with $|\tilde{x}_{ij} - x_{ij}| \le e^{0.1} < 0.11$. This has the effect that, in the "optimum" solution that we compute, (12) will hold with an additional error term of ± 0.11 on both sides. If $\overline{\log} a_{ij}$ is also computed with an error of at most $\pm 0.1/M$, we get that (12) is satisfied with an error term of ± 0.23 on both sides. This is good enough for achieving an ε-approximation of the real-valued matrix scaling problem. The scaling factor F in the proof of Theorem 4 must be adjusted a little, but the asymptotic runtime is unaffected.

In Lemma 3, the discrepancy after using the solution of P^k as the initial solution for P^{k-1} may be larger than $6n^2$ but it is still $O(n^2)$. Again, this does not affect the asymptotic running time.

We analyze the fractional and the integer part that is necessary for the variables x_{ij}, α_i, β_j. The flow values x_{ij} that occur during the course of the computation are bounded by $M := H + O(n^2) = hn/\varepsilon + O(n^2)$. We set $\delta = 0.1/M$, and this means that we need $\log_2 \frac{h}{\varepsilon} + 2\log_2 n + O(1)$ bits for the fractions. of x_{ij}, $\overline{\log} a_{ij}$, α_i, and β_j.

To estimate the maximum size of the numbers, one has to observe that, during the algorithm, the new values of α_i and β_j are always determined by setting some of the inequalities (16) to equations, corresponding to arcs of zero reduced cost in the residual network. These equations form a forest. The values α_i and β_j are connected to other values α_i and β_j by chains of equations. We can select one "free" value from every component that is formed in this way, and we can take care that the "free" value is the initial value of 0. Then, each value α_i or β_j is connected to a the free value α_i or β_j by a chain of equations, going over at most $2n$ arcs. Thus, it is an alternating sum of at most $2n$ terms of the form $\overline{\log} x_{ij} - \overline{\log} a_{ij}$ or $\overline{\log}(x_{ij}+1) - \overline{\log} a_{ij}$. Thus, the maximum absolute value of the numbers α_i and β_j is bounded by $2n \cdot (\log M + \log \max_{a_{i,j} \ne 0} |\log a_{ij}|)$. As in the proof leading to Theorem 3, we can scale the input matrix A such that the largest entry is $\frac{1}{2}$. Thus we need at most $1 + \log_2 n + \log_2 \log M + \log_2 \log(2R(A))$ bits to represent the integer part of the numbers α_i and β_j, where $R(A)$ is the ratio between the largest and the smallest positive entry in A. Thus, we have:

THEOREM 5. *For achieving Theorem 4, arithmetic operations on numbers with $O(\log n + \log\log R(A) + \log \frac{h}{\varepsilon})$ bits are sufficient, where $R(A)$ is the ratio between the largest and the smallest positive entry in the input matrix.* □

For example, if $R(A) \le 10^{10}$, $\varepsilon = 10^{-6}h$, and $n \le 1000$, the algorithm can work with IEEE double-precision floating point numbers.

The reader may wonder why only $O(\log\log R(A))$ bits are needed when one must, of course, *look* at the $\log_2 R(A)$ bits of the input numbers a_{ij} in the first place. But since we are interested in an approximate solution, excessive trailing bits of very long input numbers can be ignored.

An obvious alternative is to eliminate the logarithms altogether, to use directly the exponentiated variables λ_i and μ_j instead of α_i and β_j, and to work with a multiplicative version of Dijkstra's algorithm. However, this would incur a propagation of errors in the shortest path calculations, which needs to be analyzed.

The approximate calculations discussed in this section are of course not appropriate for carrying out an election (Theorems 2 and 3), where the tiniest fraction can be important to decide about one seat. We are planning to investigate the necessary precision to compute an exact result in a future work.

6 Faster Scaling of Real Matrices

Shortly before preparing the final version of this paper, we became aware of algorithms that solve the minimum-cost flow problem with convex costs directly (without scaling).

If one goes to the limit $\varepsilon \to 0$, the objective function (7), with appropriate rescaling, converges to (4). The convex cost function for an arc with $a_{ij} > 0$ is $f_{ij}(x_{ij}) = x_{ij}(\log \frac{x_{ij}}{a_{ij}} - 1)$, with derivative $f'_{ij}(x_{ij}) = \log x_{ij} - \log a_{ij}$.

Karzanov and McCormick [9] have given fast algorithms for directly solving minimum-cost flow problems with convex costs. The algorithms always maintain a feasible flow satisfying (8–9), and they measure progress by the quantity

$$\delta := \max_{i,j} |f'_{ij}(x_{ij}) - \alpha_i - \beta_j|$$

(In [9], this parameter is denoted by λ.) To achieve a reduced value of δ, starting from some initial value δ_0, the Cancel-and-Tighten algorithm of Karzanov and

McCormick [9, entry 1c of Table 1, lower part] needs

$$O\Big(\log \frac{\delta_0}{\delta} \cdot mn \log n\Big)$$

steps, in a graph with n vertices and m edges. Since $f'_{ij}(x_{ij}) = \log x_{ij} - \log a_{ij}$, an analysis similar to the one the leads to Theorem 4 shows that

$$\delta := \frac{\varepsilon}{nh}$$

is sufficient to guarantee an ε-approximate solution in the sense of (2).

It remains to find an initial solution with small $\delta_0 := \max_{i,j} |\log x_{ij} - \alpha_i - \beta_j - \log a_{ij}|$. If the row and column sums r_i and s_j are integral, we can define a starting solution as follows. Since the problem is feasible, we know that there is a flow in which all flow values x_{ij} with $a_{ij} > 0$ are positive. For each arc ij with $a_{ij} > 0$, there exists therefore some (integer) flow with $x_{ij} \geq 1$. Taking the average of these flows (one flow for every arc) leads to a flow in which $x_{ij} \geq 1/n^2$ for all arcs ij with $a_{ij} > 0$. This starting flow can be computed by putting a lower flow bound of $1/n^2$ on every arc, and solving the feasible flow problem, in $O(n^3)$ time. Setting $\alpha_i = \beta_j = 0$ yields a starting solution with $\delta_0 \leq \log n^2 + \log R(A)$.

Theorem 6. *The (real-valued) matrix scaling problem with integer row and column sums r_i and s_j can be solved to an accuracy ε in*

$$O\Big(n^3 \log n \cdot \log \frac{hn \log R(A)}{\varepsilon}\Big)$$

time, where $R(A)$ is the ratio between the largest and the smallest positive entry in A. □

To achieve the cubic complexity in the dominating term, the algorithm has to use the dynamic tree data structure of Sleator and Tarjan [13]. Thus, in contrast to the algorithms in the previous sections, this algorithm is not easy to implement.

If the row and column sums r_i and s_j are not integral, it seems difficult to find a starting solution with a good bound on δ_0. The quality of the solution may depend very much on the numbers r_i and s_j and on the zero pattern of the matrix a_{ij}. Of course, if the row and column sums are rational, one can compute their greatest common divisor, scale them to integers, and then apply the above theorem.

If the matrix a_{ij} is positive, however, a starting solution can always be found by the proportional solution $x_{ij} := r_i s_j/h$, $\alpha_i = \log r_i$, $\beta_j := \log(s_j/h)$, which yields $\log x_{ij} - \alpha_i - \beta_j = 0$, and $\delta_0 = \log R(A)$. In this case, the initial overhead reduces from $O(n^3)$ to $O(n^2)$.

Acknowledgements. We thank Dorit Hochbaum for helpful remarks, in particular for pointing out the work of Karzanov and McCormick [9].

References

[1] M. Bacharach. *Biproportional Matrices and Input-Output Change*. Cambridge University Press, 1970.

[2] Michel Balinski and Gabrielle Demange. Algorithms for proportional matrices in reals and integers. *Math. Programming*, 45:193–210, 1989.

[3] Michel Balinski and Svetlozar T. Rachev. Rounding proportions: methods of rounding. *Math. Scientist*, 22:1–26, 1997.

[4] William R. Cook, William H. Cunningham, William R. Pulleyblank, and Alexander Schrijver. *Combinatorial Optimization*. Wiley, 1998.

[5] Martin Fürer. Quadratic convergence for scaling of matrices. In Lars Arge, Giuseppe F. Italiano, and Robert Sedgewick, editors, *Proc. ALENEX/ANALC 2004, 6th Workshop on Algorithm Engineering and Experiments and 1st Workshop on Analytic Algorithmics and Combinatorics, New Orleans*, pages 216–223. SIAM, January 2004.

[6] Norbert Gaffke and Friedrich Pukelsheim. Divisor methods for proportional representation systems: an optimization approach to vector and matrix problems. Technical Report (Preprint) 06-18, Universität Magdeburg, Fakultät für Mathematik, March 2006. `http://www.math.uni-magdeburg.de/preprints/shadows/06-18report.html`.

[7] Dorit Hochbaum. Complexity and algorithms for convex network optimization and other nonlinear problems. *4OR*, 3(3):171–216, 2005.

[8] Bahman Kalantari and Leonid Khachiyan. On the complexity of nonnegative-matrix scaling. *Linear Algebra Appl.*, 240:87–104, 1996.

[9] Alexander V. Karzanov and S. Thomas McCormick. Polynomial methods for separable convex optimization in unimodular linear spaces with applications. *SIAM J. Comput.*, 26(4):1245–1275, 1997.

[10] Nathan Linial, Alex Samorodnitsky, and Avi Wigderson. A deterministic strongly polynomial algorithm for matrix scaling and approximate permanents. *Combinatorica*, 20(4):545–568, 2000.

[11] U. Rothblum and H. Schneider. Scaling of matrices which have prespecified row sums and column sums via optimization. *Linear Algebra Appl.*, 114–115:737–764, 1989.

[12] Richard Sinkhorn. A relationship between arbitrary positive matrices and doubly stochastic matrices. *Ann. Math. Statist.*, 35:876–879, 1964.

[13] D. D. Sleator and R. E. Tarjan. A data structure for dynamic trees. *J. Comput. Syst. Sci.*, 26(3):362–381, 1983.

[14] George W. Soules. The rate of convergence of Sinkhorn balancing. *Linear Algebra Appl.*, 150:3–40, 1991.