DIAGONAL SCALINGS FOR THE EIGENSTRUCTURE OF ARBITRARY PENCILS*

FROILÁN M. DOPICO†, MARÍA C. QUINTANA†, AND PAUL VAN DOOREN‡

Abstract. In this paper we show how to construct diagonal scalings for arbitrary matrix pencils $\lambda B - A$, in which both A and B are complex matrices (square or nonsquare). The goal of such diagonal scalings is to "balance" in some sense the row and column norms of the pencil. We see that the problem of scaling a matrix pencil is equivalent to the problem of scaling the row and colum sums of a particular nonnegative matrix. However, it is known that there exist square and nonsquare nonnegative matrices that can not be scaled arbitrarily. To address this issue, we consider an approximate embedded problem, in which the corresponding nonnegative matrix is square and can always be scaled. The new scaling method is then based on the Sinkhorn–Knopp algorithm for scaling a square nonnegative matrix with total support to be doubly stochastic. In addition, using results of U. G. Rothblum and H. Schneider (1989), we give sufficient conditions for the existence of diagonal scalings of square nonnegative matrices to be not only doubly stochastic but have any prescribed common vector for the row and column sums. We illustrate numerically that the new scaling techniques for pencils improve the sensitivity of the computation of their eigenvalues.

Key words. pencils, eigenvalue sensitivity, diagonal scaling, Sinkhorn-Knopp algorithm

AMS subject classifications. 15A18, 15A22, 65F15, 65F35

1. Introduction.

The problem of scaling an entrywise nonnegative $m \times n$ matrix A with diagonal transformations and prespecified vectors r and c for the row and column sums, respectively, consists of finding a matrix of the form $S = D_\ell A D_r$, where $D_\ell \in \mathbb{R}^{m\times m}$ and $D_r \in \mathbb{R}^{n\times n}$ are diagonal matrices having positive diagonal elements, and such that

$$S\mathbf{1}_n = r \quad \text{and} \quad \mathbf{1}_m^T S = c^T, \tag{1.1}$$

where $\mathbf{1}_i := [1, \ldots, 1]^T \in \mathbb{R}^i$ for $i = n, m$ [13]. When $r = \mathbf{1}_m$ and $c = \mathbf{1}_n$ the scaled matrix S is said to be doubly stochastic, i.e., its row and column sums are all equal to 1.

The related problem of scaling the rows and columns of a complex square matrix A (not necessarily nonnegative) using real and positive diagonal similarity transformations in order to compute more accurate eigenvalues, is a well established technique to improve the sensitivity of the eigenvalue problem of the matrix A [12]. This is known as *balancing* the matrix A and it amounts to minimizing the Frobenius norm distance of the scaled matrix $D^{-1}AD$ to the set of normal matrices [9], which is the set of matrices whose eigenvalues all have minimal condition number equal to 1. The method for computing the optimal scaling is a very simple cyclic procedure where at each step only a single diagonal element of D is updated. This method is implemented

*Submitted to the editors DATE.

Funding: Supported by "Ministerio de Economía, Industria y Competitividad (MINECO)" of Spain and "Fondo Europeo de Desarrollo Regional (FEDER)" of EU through grants MTM2015-65798-P and MTM2017-90682-REDT and by the "Proyecto financiado por la Agencia Estatal de Investigación (PID2019-106362GB-I00 / AEI / 10.13039/501100011033)" . The research of M. C. Quintana is funded by the "contrato predoctoral" BES-2016-076744 of MINECO. This work was partially developed while Paul Van Dooren held a "Chair of Excellence UC3M - Banco de Santander" at Universidad Carlos III de Madrid in the academic year 2019-2020.

†Departamento de Matemáticas, Universidad Carlos III de Madrid, Avda. Universidad 30, 28911 Leganés, Spain (dopico@math.uc3m.es, maquinta@math.uc3m.es).

‡Department of Mathematical Engineering, Université catholique de Louvain, Avenue Georges Lemaître 4, B-1348 Louvain-la-Neuve, Belgium (paul.vandooren@uclouvain.be).

in MATLAB [16] as a default option of the eigenvalue computation problem, which indicates that its effectiveness is well accepted. Notice that when restricting oneself to powers of 2 for the diagonal elements of D, the scaling does not produce any rounding errors and hence the eigenvalues are preserved *exactly* under such a scaling transformation.

The idea of performing positive diagonal scaling in order to improve the sensitity of eigenvalues was also extended to the generalized eigenvalue problem of a regular pencil $\lambda B - A$. In this case, the nonsingular diagonal matrices multiplying the pencil on the left and on the right are different. In [18], Ward describes a scaling technique which aims at making the pencil entries have magnitudes as close to unity as possible, whereas in [9], Lemonnier and Van Dooren propose a scaling that makes the pencil as close as possible to a so-called standardized normal pencil, which are pencils whose eigenvalues all have a condition number in the chordal metric that is smaller or equal to $\sqrt{2}$. This second approach is linked to the two-sided scaling of a related square nonnegative matrix in order to make it doubly stochastic.

We show in this paper that the link to the doubly stochastic scaling problem, implies that the scaling is essentially unique and bounded if and only if the corresponding nonnegative matrix satisfies certain conditions, namely *total support* and *full indecomposability*. Moreover, in that situation, the scaling can be found through the well-known Sinkhorn-Knopp algorithm [8, 14]. We then show how to extend this to singular or nonsquare pencils, which, to the best of our knowledge, has not been considered yet in the literature. For that, we introduce a regularization term into the original problem which ensures existence of a solution of an approximate problem with bounded diagonal scalings D_ℓ and D_r. In addition, the regularization term can be considered in both square or nonsquare cases.

These ideas are connected to the results of Rothblum and Schneider [13] about prespecified row and column sums, and the numerical solution we propose uses a Sinkhorn-Knopp-like algorithm. We then build on these ideas to further improve the scaling technique of Lemonnier and Van Dooren by introducing the regularization term as an additional cost. This cost can be viewed as a regularization to ensure existence and boundedness of our scaling, but it also ensures essential unicity of the computed scaling.

The paper is organized as follows. In Section 2, we give some basic notions about scaling pencils. In Sections 3 and 4, we study the diagonal scaling problem for square and nonsquare pencils, respectively. In Section 3, we will also recall the necessary and sufficient conditions for a square matrix to become doubly stochastic under diagonal scalings, and we give sufficient conditions for the existence of diagonal scalings having any prespecified common vector for the row and column sums. These results will be useful in Section 5. In that section, we develop a new scaling technique for generalized eigenvalue problems and show that it can be applied to any pencil, regular or singular, square or rectangular. For that, we introduce a regularization term into the original problem which guarantees existence, unicity and boundedness of the scaling. In addition, in Subsection 5.1, we consider a modified version of the new scaling technique that is better for scaling nonsquare pencils. In Section 6 we then illustrate the improved sensitivity of the computed eigenvalues using several numerical examples. In the last Section 7 we give some concluding remarks.

2. Preliminaries: Scaling pencils.

The standard techniques for computing eigenvalues of complex pencils of matrices guarantee that the backward errors corresponding to the computed spectrum is essentially bounded by the norm of the coefficients of the pencil, times the machine precision of the computer used. But one can improve this bound by reducing the norms of the coefficients without affecting the spectrum. This is where balancing using diagonal scaling comes in.

Two types of scalings can be applied to a pencil $\lambda A - B$.

The first one is a change of variable $\hat{\lambda} := d_\lambda \lambda$ to make sure that the scaled matrices A and $d_\lambda B$ have approximately the same norm. This can be done without introducing rounding errors, by taking d_λ equal to a power of 2. The staircase and the QZ algorithm work independently on both matrices and this scaling can be restored afterwards, again without introducing any additional errors. One could therefore argue that this scaling is irrelevant for these algorithms, but we will see that it affects the second scaling procedure we will discuss. Therefore we will assume in the sequel that both matrices A and B are of comparable norms, and that no such variable scaling needs to be applied.

The second type of scaling is based on multiplication on the left and on the right by positive diagonal matrices D_ℓ and D_r, respectively, that are chosen to "balance" in some sense the row and column norms of the complex matrices $\widetilde{A} := D_\ell A D_r$ and $\widetilde{B} := D_\ell B D_r$. We will see that balancing the row and column norms of the matrices $\widetilde{A}$ and $\widetilde{B}$ is equivalent to performing two-sided diagonal scalings to a particular real entrywise nonnegative matrix M. Therefore, we recall in the sequel some results on this problem.

We know by the work of Rothblum and Schneider in [13] that there exists at most one solution for the diagonal scaling problem of arbitrary real nonnegative matrices (square or nonsquare). The following Theorem 2.1 is a partial result of what is proven in [13].

THEOREM 2.1. *(Rothblum-Schneider) Let $M \in \mathbb{R}^{m\times n}$ be a nonnegative matrix and let $r \in \mathbb{R}^{m\times 1}$ and $c \in \mathbb{R}^{n\times 1}$ be strictly positive vectors satisfying $\mathbf{1}_m^T r = c^T \mathbf{1}_n$, then there exists at most one two-sided scaled matrix $S = D_{M,\ell} M D_{M,r}$ with row sums $S\mathbf{1}_n = r$ and column sums $\mathbf{1}_m^T S = c^T$, where $D_{M,\ell}$ and $D_{M,r}$ are diagonal matrices with positive main diagonals.*

Therefore, if a matrix M can be scaled, the scaled matrix is unique. However, necessary and sufficient conditions on M for the scaling to exist are not known. In addition, there are infinitely many examples of nonnegative matrices that cannot be scaled for prescribed r and c.

EXAMPLE 2.2. *For instance, one can easily check that the matrix*

$$M := \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

can not be scaled with prescribed vectors $r := [3, 3]^T$, for the row sums, and $c := [2, 2, 2]^T$, for the column sums.

In the next section, we will see conditions for diagonal scalings to exist with certain prescribed common vector for the row and column sums in the case of balancing square pencils and matrices.

3. Scaling square pencils and related problems. Let us first look at the case of square pencils. In [9, page 259], positive diagonal matrices D_ℓ and D_r are chosen to equilibrate the row and column norms of a $n \times n$ regular pencil $\lambda B - A$, by imposing

$$\|\mathrm{col}_j(\widetilde{A})\|_2^2 + \|\mathrm{col}_j(\widetilde{B})\|_2^2 = \|\mathrm{row}_i(\widetilde{A})\|_2^2 + \|\mathrm{row}_i(\widetilde{B})\|_2^2 = \gamma^2, \text{ for } i, j = 1, \ldots, n, \quad (3.1)$$

for some constant γ resulting from the balancing, where $\widetilde{A} := D_\ell A D_r$ and $\widetilde{B} := D_\ell B D_r$, and $\|\cdot\|_2$ denotes the standard Euclidean norm of a vector [5]. A pencil satisfying these conditions was called *balanced* and an algorithm was presented in [9] to compute a scaling to balance a regular pencil $\lambda B - A$. It was shown that this amounts to solving the following norm minimization problem

$$\inf_{\det D_\ell . \det D_r = 1} \|D_\ell(\lambda B - A)D_r\|_F^2, \tag{3.2}$$

using the so-called Frobenius norm of a pencil:

$$\|\lambda B - A\|_F^2 := \|B\|_F^2 + \|A\|_F^2,$$

where $\|A\|_F$ and $\|B\|_F$ are the matrix Frobenius norms of A and B [5]. Moreover, the following result was proven in [9].

THEOREM 3.1. *The minimization problem*

$$\inf_{\det T_\ell . \det T_r = 1} \|T_\ell(\lambda B - A)T_r\|_F^2, \tag{3.3}$$

where T_ℓ and T_r are arbitrary nonsingular matrices, has a so-called standardized normal pencil $\lambda\hat{B} - \hat{A}$ *as solution, satisfying*

$$U_\ell(\lambda\hat{B} - \hat{A})U_r = \lambda\Lambda_B - \Lambda_A, \quad U_\ell^* U_\ell = U_r^* U_r = I_n, \quad |\Lambda_B|^2 + |\Lambda_A|^2 = \gamma^2 I_n,$$

where Λ_B and Λ_A are diagonal. If the eigenvalues of the regular pencil $\lambda B - A$ are distinct, then T_ℓ and T_r have a bounded solution and the infimum is a minimum; otherwise they may be unbounded.

As shown in [9], the standardized normal pencils happen to have eigenvalues with condition number bounded by $\sqrt{2}$. This explains why performing the same minimization over the diagonal scalings is likely to improve the sensitivity of the eigenvalue computation. Moreover, if the transformation matrices are bounded then the eigenstructure of the regular pencil is preserved.

But the positive diagonal scalings that achieve the balancing in [9] are not unique, and they may be unbounded. In order to analyze this further we relate this problem to that of scaling a real nonnegative matrix by two-sided scalings to a doubly stochastic matrix, or in other words, to make the row sums and column sums equal to 1. An algorithm to solve this problem has been developed and analyzed by Sinkhorn and Knopp [14], and further analysis can be found in [8]. The link between both problems is the following. Let us define the nonnegative matrices

$$M := |A|^{\circ 2} + |B|^{\circ 2}, \quad \text{and} \quad \widetilde{M} := |\widetilde{A}|^{\circ 2} + |\widetilde{B}|^{\circ 2} \tag{3.4}$$

where $|X|$ indicates the element-wise absolute value of the matrix X, where $X^{\circ 2}$ indicates the elementwise square of the matrix X, and where D_ℓ and D_r satisfy the balancing equations (3.1). Then the scaled matrix $\widetilde{M} = D_\ell^2 M D_r^2$ satisfies

$$\widetilde{M}\mathbf{1}_n = D_\ell^2(|A|^{\circ 2} + |B|^{\circ 2})D_r^2\mathbf{1}_n = \gamma^2\mathbf{1}_n, \quad \mathbf{1}_n^T\widetilde{M} = \mathbf{1}_n^T D_\ell^2(|A|^{\circ 2} + |B|^{\circ 2})D_r^2 = \gamma^2\mathbf{1}_n^T$$

which implies that $\widetilde{M}/\gamma^2$ is doubly stochastic and that the two-sided scaling for the nonnegative matrix M satifies

$$\widetilde{M}/\gamma^2 = D_{M,\ell} M D_{M,r}, \quad \text{where} \quad D_{M,\ell} := D_\ell^2/\gamma, \; D_{M,r} := D_r^2/\gamma.$$

The only difference is that for balancing, we impose a scalar constraint $\det D_\ell . \det D_r = 1$, which is why the resulting row and column norms are equal to γ^2 rather than 1. In fact, the algorithm proposed in [9] was to alternately normalizing the rows and columns of M to 1 (rather than γ), and that is precisely the algorithm of Sinkhorn-Knopp. This connection was not established in [9].

It follows from this that the unicity or boundedness of the scalings are equivalent for the two problems.

We recall in Theorem 3.5 the results given for two-sided scaling in [14] for square nonnegative matrices $M \in \mathbb{R}^{n\times n}$ in order the corresponding matrix to become doubly stochastic. We notice that the doubly stochastic scaling problem of Theorem 3.5 is a special case of the scaling problem in Theorem 2.1, just by considering square matrices and $r = c = \mathbf{1}_n$. Before stating Theorem 3.5, we introduce the notions of total support and full indecomposability, that will be used.

Definition 3.2. *The sequence $m_{1,\sigma(1)}, m_{2,\sigma(2)}, \cdots, m_{n,\sigma(n)}$, where σ is a permutation of $\{1, 2, \cdots, n\}$, is called a diagonal of a $n \times n$ square matrix M. A nonnegative matrix $M \in \mathbb{R}^{n\times n}$ is said to have* total support *if every positive element of M lies on a positive diagonal.*

Definition 3.3. *A nonnegative matrix $M \in \mathbb{R}^{n\times n}$ is said to be* fully indecomposable *if there do not exist permutation matrices P_ℓ and P_r such that $P_\ell M P_r$ can be partitioned as*

$$P_\ell M P_r = \begin{bmatrix} M_{11} & M_{12} \\ 0 & M_{22} \end{bmatrix},$$

where M_{11} and M_{22} are square matrices.

Remark 3.4. It was proved in [1] that a fully indecomposable matrix has total support.

Theorem 3.5. *(Sinkhorn-Knopp) If $M \in \mathbb{R}^{n\times n}$ is a nonnegative matrix then a necessary and sufficient condition that there exists a doubly stochastic matrix S of the form $S = D_{M,\ell} M D_{M,r}$, where $D_{M,\ell}$ and $D_{M,r}$ are diagonal matrices with positive main diagonals, is that M has total support. If S exists, then it is unique. $D_{M,\ell}$ and $D_{M,r}$ are also unique up to a nonnegative scalar multiple if and only if M is fully indecomposable.*

The doubly stochastic matrix S can be obtained as a limit of a sequence of matrices generated by alternately normalizing the row and column sums of M. A necessary and sufficient condition that the iterative process of alternately normalizing the row and column sums of M will converge to a doubly stochastic limit of the form $D_{M,\ell} M D_{M,r}$ is that M has total support [8, 14]. See Appendix A for an extended version of this process.

We recall in the following Theorem 3.6 the particular case of having a symmetric and fully indecomposable matrix M. This case will be important in the new regularized scaling method developed in Section 5.

Theorem 3.6. *[8, Lemma 4.1] If $M \in \mathbb{R}^{n\times n}$ is a symmetric nonnegative and fully indecomposable matrix then there exists a unique diagonal matrix D with positive main diagonal such that DMD is doubly stochastic.*

Remark 3.7. When M is fully indecomposable, the solution set for the diagonal scalings is $\mathcal{S} := \{(D_{M,\ell}/c, cD_{M,r}) : c > 0\}$, for a given solution $(D_{M,\ell}, D_{M,r})$. To guarantee unicity for a solution in $\mathcal{S}$, one can consider a unique “normalized” scaling pair $(D_{M,\ell}, D_{M,r})$. For instance, by imposing that the solution satisfies $\det D_{M,\ell} =$

$\det D_{M,r}$ or $\max\limits_{i=1,\ldots,n}\{d_i^\ell\} = \max\limits_{i=1,\ldots,n}\{d_i^r\}$, where d_i^ℓ and d_i^r are the diagonal entries of $D_{M,\ell}$ and $D_{M,r}$, respectively. Then the pair $(D_{M,\ell}, D_{M,r})$ is unique in $\mathcal{S}$. Moreover, when M is symmetric, then these normalizations imply that $D_{M,\ell} = D_{M,r}$. In summary, one can always perform a normalization in order to obtain unicity for the diagonal scalings.

In the following examples, we illustrate what is happening when the conditions mentioned in Theorem 3.5 do not hold.

EXAMPLE 3.8. *Let us consider the regular pencil*

$$\lambda B_1 - A_1 := \begin{bmatrix} 1 & \lambda & 0 \\ \lambda & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \textit{and let} \quad M_1 := \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

be the corresponding matrix $M := M_1$ in (3.4). M_1 has no total support since the (1,1) entry is not on a positive diagonal. The Sinkhorn-Knopp algorithm does not converge for this example. In fact, any candidate pair of scalings $D_{M,\ell} = \operatorname{diag}(\ell_1, \ell_2, \ell_3)$, and $D_{M,r} = \operatorname{diag}(r_1, r_2, r_3)$, has to satisfy $\ell_1 r_2 = \ell_2 r_1 = \ell_3 r_3 = 1$ and $\ell_1 r_1 = 0$ which does not have a bounded solution.

Now, let us consider the regular pencil

$$\lambda B_2 - A_2 := \begin{bmatrix} 1 & \lambda & 0 \\ \lambda & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \textit{and let} \quad M_2 := \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

be the corresponding matrix $M := M_2$ in (3.4). In this case, M_2 has total support and, then, the Sinkhorn-Knopp algorithm converges. Indeed, the following positive diagonal scaling makes M doubly stochastic:

$$\begin{bmatrix} \sqrt{\frac{1}{2}} & 0 & 0 \\ 0 & \sqrt{\frac{1}{2}} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \sqrt{\frac{1}{2}} & 0 & 0 \\ 0 & \sqrt{\frac{1}{2}} & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

However, M_2 is not fully indecomposable, which implies that $D_{M,\ell}$ and $D_{M,r}$ are not unique up to a scalar multiple. In this case, the Sinkhorn-Knopp algorithm may converge to different diagonal scaling matrices for different starting diagonal initial conditions. Moreover, it may converge to unbounded $D_{M,\ell}$ and $D_{M,r}$. For instance, for the following scaling

$$\begin{bmatrix} t\sqrt{\frac{1}{2}} & 0 & 0 \\ 0 & t\sqrt{\frac{1}{2}} & 0 \\ 0 & 0 & 1/s \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{t}\sqrt{\frac{1}{2}} & 0 & 0 \\ 0 & \frac{1}{t}\sqrt{\frac{1}{2}} & 0 \\ 0 & 0 & s \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

the right diagonal matrix is unbounded as $t \to 0$ and the left one as $s \to 0$. Finally, let us consider the regular pencil

$$\lambda B_3 - A_3 := \begin{bmatrix} 1 & \lambda & 0 \\ \lambda & 0 & \lambda \\ 0 & \lambda & 1 \end{bmatrix}, \quad \textit{and let} \quad M_3 := \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$$

be the corresponding matrix $M := M_3$ in (3.4). In this case, M_3 has total support and is, in addition, fully indecomposable. Then the scaling procedure converges to bounded diagonal scaling matrices, that are essentially unique (up to a scalar multiple):

$$\begin{bmatrix} \sqrt{\frac{1}{2}} & 0 & 0 \\ 0 & \sqrt{\frac{1}{2}} & 0 \\ 0 & 0 & \sqrt{\frac{1}{2}} \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} \sqrt{\frac{1}{2}} & 0 & 0 \\ 0 & \sqrt{\frac{1}{2}} & 0 \\ 0 & 0 & \sqrt{\frac{1}{2}} \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix}.$$

For the general scaling problem in Theorem 2.1, with prespecified vectors for the row and column sums, sufficient conditions on M for the scaling to exist are not known in the literature, as far as we know, not even in the case of a square matrix M.

We now derive sufficient conditions for the existence of a diagonal scaling of a square matrix M by considering not only the vector $\mathbf{1}_n$ but any prescribed common vector r for the row and column sums. For that, we use the following Lemma 3.9, which is a partial result of [13, Theorem 2]. In what follows, the support of a matrix $A \in \mathbb{R}^{m\times n}$, denoted by $\operatorname{supp}(A)$, is defined as the set $\{(i,j) \,|\, a_{ij} \neq 0, i = 1, \cdots, m, \text{ and } j = 1, \cdots, n\}$.

LEMMA 3.9. *Let $M \in \mathbb{R}^{m\times n}$ be a nonnegative matrix and let $r \in \mathbb{R}^{m\times 1}$ and $c \in \mathbb{R}^{n\times 1}$ be strictly positive vectors such that $\mathbf{1}_m^T r = c^T \mathbf{1}_n$. Then there exists a scaled matrix $S = D_{M,\ell} M D_{M,r}$ with row sums $S\mathbf{1}_n = r$ and column sums $\mathbf{1}_m^T S = c^T$, where $D_{M,\ell}$ and $D_{M,r}$ are diagonal matrices with positive main diagonals, if and only if there exist no pair of vectors $(u,v) \in \mathbb{R}^m \times \mathbb{R}^n$ for which*

(a) *$u_i + v_j \leq 0$ for each pair $(i,j) \in supp(M)$,*
(b) *$r^T u = c^T v = 0$, and*
(c) *$u_{i_0} + v_{j_0} < 0$ for some pair $(i_0, j_0) \in supp(M)$.*

THEOREM 3.10. *Let $M \in \mathbb{R}^{n\times n}$ be a nonnegative matrix with $(i,i) \in supp(M)$ for all $i = 1, \cdots, n$ and such that $supp(M) = supp(M^T)$. Let $v \in \mathbb{R}^{n\times 1}$ be a strictly positive vector. Then there exists a scaled matrix $S = D_{M,\ell} M D_{M,r}$ with row sums $S\mathbf{1}_n = v$ and column sums $\mathbf{1}_n^T S = v^T$, where $D_{M,\ell}$ and $D_{M,r}$ are diagonal matrices with positive main diagonals. Moreover, S is unique. If, in addition, M is fully indecomposable then $D_{M,\ell}$ and $D_{M,r}$ are also unique up to a nonnegative scalar multiple.*

Proof. Consider a $n \times n$ nonnegative matrix M such that $\operatorname{supp}(M) = \operatorname{supp}(M^T)$ and $(i,i) \in \operatorname{supp}(M)$ for all $i = 1, \cdots, n$. By contradiction, let us assume that there exists no scaled matrix S with row sums $S\mathbf{1}_n = v$ and column sums $\mathbf{1}_n^T S = v^T$. Then, by Lemma 3.9, there exists a pair of vectors $(x,y) \in \mathbb{R}^n \times \mathbb{R}^n$ for which

(a) $x_i + y_j \leq 0$ for each pair $(i,j) \in \operatorname{supp}(M)$,
(b) $v^T x = v^T y = 0$, and
(c) $x_{i_0} + y_{j_0} < 0$ for some pair $(i_0, j_0) \in \operatorname{supp}(M)$.

Condition (b) implies that

$$v_1(x_1 + y_1) + \cdots + v_n(x_n + y_n) = 0. \tag{3.5}$$

In addition, since $(i,i) \in \operatorname{supp}(M)$ for all $i = 1, \cdots, n$, condition (a) implies that $x_i + y_i \leq 0$ for all $i = 1, \cdots n$. It then follows from (3.5) that $x_i + y_i = 0$ for all $i = 1, \cdots n$ since $v_i > 0$. Moreover, by condition (c), there exists a pair $(i_0, j_0) \in \operatorname{supp}(M)$ such that $x_{i_0} + y_{j_0} < 0$. Taking into account that $x_i + y_i = 0$ for all $i = 1, \cdots n$ we have that

$$(x_{i_0} + y_{i_0}) + (x_{j_0} + y_{j_0}) = 0. \tag{3.6}$$

∎

By equation (3.6) and the fact that $x_{i_0} + y_{j_0} < 0$, we obtain that $x_{j_0} + y_{i_0} > 0$. Therefore, by (a), $(j_0, i_0) \notin \operatorname{supp}(M)$, which is a contradiction since $(i_0, j_0) \in \operatorname{supp}(M)$ and $\operatorname{supp}(M) = \operatorname{supp}(M^T)$.

The uniqueness of S is a consequence of Theorem 2.1. Finally, if M is fully indecomposable its bipartite graph is connected [2, Theorem 1.3.7] and, thus, it is chainable [6, Theorem 1.2] (see [6] or [13] for the definition of "chainable"). Then, by [13, Theorem 4], $D_{M,\ell}$ and $D_{M,r}$ are also unique up to a nonnegative scalar multiple.

If M satisfies the conditions in Theorem 3.10, the scaled matrix S can be computed by using the algorithm in Appendix B with prescribed common vector v for the row and column sums, i.e., with $r = c = v$. Although we have not proved the convergence of the algorithm in this case, the method is analogous to the Sinkhorn-Knopp algorithm of alternately normalizing the row and column sums but making the row and column sums equal to v. We have checked that the algorithm works very well in practice under the conditions of Theorem 3.10.

In Section 5, we will present new cost functions for our minimization problem (3.2) to make sure that it always has a unique and bounded solution. This new approach will be based on the results presented in this section combined with regularization techniques. In addition, this new approach will be applied to arbitrary pencils (square or nonsquare). First, we study in Subsection 4 the nonsquare case.

4. Scaling nonsquare pencils and related problems. In the square case, we scaled the pencil so that its row norms and column norms were equal as in (3.1). However, this is no longer possible for $m \times n$ rectangular pencils since the numbers of rows and columns are different. But instead, one can try to balance the pencil $\lambda B - A$ by achieving the following equalities

$$\begin{aligned} &\|\mathrm{col}_j(\widetilde{A})\|_2^2 + \|\mathrm{col}_j(\widetilde{B})\|_2^2 = \gamma_\ell^2, \text{ for } j = 1, \ldots, n, \text{ and} \\ &\|\mathrm{row}_i(\widetilde{A})\|_2^2 + \|\mathrm{row}_i(\widetilde{B})\|_2^2 = \gamma_r^2, \text{ for } i = 1, \ldots, m, \end{aligned} \tag{4.1}$$

where $\widetilde{A} := D_\ell A D_r$ and $\widetilde{B} := D_\ell B D_r$ and $\|\lambda\widetilde{B} - \widetilde{A}\|_F^2 = n\gamma_\ell^2 = m\gamma_r^2$. For the nonsquare case, we also define the nonnegative matrices

$$M := |A|^{\circ 2} + |B|^{\circ 2}, \quad \text{and} \quad \widetilde{M} := |\widetilde{A}|^{\circ 2} + |\widetilde{B}|^{\circ 2}. \tag{4.2}$$

The scaling problem discussed in this section is a special case of the general scaling problem in Theorem 2.1, where we choose $r = \gamma_r^2 \mathbf{1}_m$ and $c = \gamma_\ell^2 \mathbf{1}_n$.

We now show that there is an optimization problem whose first order optimality conditions corresponds to the equalities in (4.1).

THEOREM 4.1. *The following minimization problem over the set of positive diagonal matrices $D_\ell = \operatorname{diag}(d_{\ell_1}, \ldots, d_{\ell_m})$ and $D_r = \operatorname{diag}(d_{r_1}, \ldots, d_{r_n})$:*

$$\inf_{\det D_\ell^2 = c_\ell, \det D_r^2 = c_r} (\|D_\ell A D_r\|_F^2 + \|D_\ell B D_r\|_F^2)$$

has the balancing equations (4.1) as first order optimality conditions.

Proof. If one makes the change of variables for the elements of D_ℓ and D_r as follows $d_{\ell_i}^2 = \exp(u_i)$, $d_{r_j}^2 = \exp(v_j)$, and introduce the notation $m_{ij} := |a_{ij}|^2 + |b_{ij}|^2$, then the above minimization is equivalent to a convex minimization problem with linear constraints :

$$\inf \sum_{i=1}^{m} \sum_{j=1}^{n} m_{ij} \exp(u_i + v_j), \quad \text{subject to} \quad \sum_{i=1}^{m} u_i = \ln c_\ell, \quad \sum_{j=1}^{n} v_j = \ln c_r. \tag{4.3}$$

The corresponding unconstrained problem with Lagrange multipliers Γ_ℓ and Γ_r, is

$$\inf \sum_{i=1}^{m}\sum_{j=1}^{n} m_{ij}\exp(u_i+v_j) + \Gamma_\ell(\ln c_\ell - \sum_{i=1}^{m} u_i) + \Gamma_r(\ln c_r - \sum_{j=1}^{n} v_j).$$

The first order conditions of optimality are the equality constraints of (4.3) and the equations

$$\sum_{j=1}^{n} d_{\ell_i}^2 m_{ij} d_{r_j}^2 = \Gamma_\ell, \quad \sum_{i=1}^{m} d_{\ell_i}^2 m_{ij} d_{r_j}^2 = \Gamma_r, \tag{4.4}$$

which express exactly that the row norms of $\widetilde{M} := D_\ell^2 M D_r^2$ are equal to each other and that its column norms are equal to each other. Since the Lagrange multipliers Γ_ℓ and Γ_r are clearly nonnegative, we can can write them as $\gamma_\ell^2 := \Gamma_\ell$ and $\gamma_r^2 := \Gamma_r$, which completes the proof. ∎

It is important to emphasize that unfortunately the optimization problem in Theorem 4.1 does not always have a solution. This happens, for instance, if the corresponding matrix $M := |A|^{\circ 2} + |B|^{\circ 2}$ is the matrix appearing in Example 2.2.

If there exists solution for the optimization problem in Theorem 4.1, it can be obtained by a sequence of alternating scalings D_ℓ^2 and D_r^2 that make the rows of $D_\ell^2(MD_r^2)$ have equal sum γ_r^2, and then the columns of $(D_\ell^2 M)D_r^2$ have equal sum γ_ℓ^2, while maintaining the constraints $\det D_\ell^2 = c_\ell$, $\det D_r^2 = c_r$. The cyclic alternation of row and column scalings, then amounts to coordinate descent applied to the minimization. This algorithm thus continues to decrease the cost function as long as the equalities (4.4) are not met. This is very similar to the Sinkhorn-Knopp algorithm, except that it is for a nonsquare matrix, and that there, one chooses $\gamma_\ell = \gamma_r = 1$ and one does not impose a determinant condition. A MATLAB code is given in Appendix A.

EXAMPLE 4.2. *Let us consider the pencil of a* 5×6 *Kronecker block*

$$\lambda B - A := \begin{bmatrix} \lambda & -1 & & & & \\ & \lambda & -1 & & & \\ & & \lambda & -1 & & \\ & & & \lambda & -1 & \\ & & & & \lambda & -1 \end{bmatrix}$$

then the scaled matrix $\widetilde{M}$ *and the corresponding diagonal scaling matrices* D_ℓ^2 *and* D_r^2 *look like*

$$\widetilde{M} := \begin{bmatrix} 5 & 1 & & & & \\ & 4 & 2 & & & \\ & & 3 & 3 & & \\ & & & 2 & 4 & \\ & & & & 1 & 5 \end{bmatrix}, \qquad \begin{matrix} D_\ell^2 = \mathrm{diag}(1,4,6,4,1), & \gamma_\ell^2 = 5, \\ D_r^2 = \mathrm{diag}(5,1,0.5,0.5,1,5), & \gamma_r^2 = 6. \end{matrix}$$

5. The regularized scaling method for pencils. The facts that for a non-square pencil the doubly stochastic scaling can not be applied anymore, that even for square pencils the corresponding matrix M may not have total support and that the optimization problem in Theorem 4.1 does not always have solution can be by-passed by introducing a regularization term which will ensure an essentially unique bounded solution for D_ℓ and D_r. The cost of introducing such a term is that we will obtain a solution of an approximate problem. Nevertheless, with the new approach we can always assure that we will find such a solution.

Given two matrices A, B of size $m \times n$, we consider the following constrained minimization problem over the set of nonnegative diagonal matrices $D_\ell = \operatorname{diag}(d_{\ell_1}, \ldots, d_{\ell_m})$ and $D_r = \operatorname{diag}(d_{r_1}, \ldots, d_{r_n})$:

$$\inf_{\det D_\ell^2 \det D_r^2 = c} 2(\|D_\ell A D_r\|_F^2 + \|D_\ell B D_r\|_F^2) + \alpha^2 \left(\frac{1}{m^2}\|D_\ell\|_F^4 + \frac{1}{n^2}\|D_r\|_F^4 \right), \tag{5.1}$$

for some real number $c > 0$ and a regularization parameter α. If we denote again the matrix $M := |A|^{\circ 2} + |B|^{\circ 2}$, then we can rewrite this as follows:

$$\inf_{\det D_\ell^2 \det D_r^2 = c} \mathbf{1}_{m+n}^T \begin{bmatrix} \frac{\alpha^2}{m^2} D_\ell^2 \mathbf{1}_m \mathbf{1}_m^T D_\ell^2 & D_\ell^2 M D_r^2 \\ D_r^2 M^T D_\ell^2 & \frac{\alpha^2}{n^2} D_r^2 \mathbf{1}_n \mathbf{1}_n^T D_r^2 \end{bmatrix} \mathbf{1}_{m+n}, \tag{5.2}$$

which suggests that there may be a link to doubly stochastic scaling. Indeed, let us consider the two-sided scaling problem $\widetilde{M}_\alpha := D_{\ell,r} M_\alpha D_{\ell,r}$, where

$$D_{\ell,r} := \begin{bmatrix} D_\ell & 0 \\ 0 & D_r \end{bmatrix},$$

subject to $\det D_\ell^2 \det D_r^2 = \det D_{\ell,r}^2 = c$, and

$$M_\alpha^{\circ 2} = \begin{bmatrix} \frac{\alpha^2}{m^2} \mathbf{1}_m \mathbf{1}_m^T & M \\ M^T & \frac{\alpha^2}{n^2} \mathbf{1}_n \mathbf{1}_n^T \end{bmatrix}. \tag{5.3}$$

Notice that both diagonal blocks in M_α have Frobenius norm α. We then prove in Theorem 5.2 that the optimization problem (5.1) can be solved by the Sinkhorn–Knopp algorithm in a unique way. We will need the following auxiliary Lemma 5.1 in our proof.

LEMMA 5.1. *Let $M_\alpha^{\circ 2}$ be the nonnegative matrix in (5.3) with $\alpha \neq 0$. Then $M_\alpha^{\circ 2}$ has total support. Moreover, if $M \neq 0$ then $M_\alpha^{\circ 2}$ is fully indecomposable.*

Proof. See Appendix C. ∎

THEOREM 5.2. *Let A and B be $m \times n$ complex matrices and $\alpha, c > 0$ be real numbers. Let us consider the constrained minimization problem (5.1) over the set $\{(D_\ell, D_r) : D_\ell := \operatorname{diag}(\delta_{\ell_1}, \ldots, \delta_{\ell_m}), D_r := \operatorname{diag}(\delta_{r_1}, \ldots, \delta_{r_n}), \delta_{\ell_i} > 0, \delta_{r_j} > 0\}$. Then the following statements hold:*

a) The optimization problem (5.1) is equivalent to the optimization problem (5.2).

b) The optimization problem (5.1) is equivalent to the optimization problem

$$\inf_{\det D_\ell^2 \det D_r^2 = c} \left\| \begin{bmatrix} D_\ell & 0 \\ 0 & D_r \end{bmatrix} M_\alpha \begin{bmatrix} D_\ell & 0 \\ 0 & D_r \end{bmatrix} \right\|_F^2,$$

where $M_\alpha^{\circ 2}$ is given in (5.3).

c) There exists a unique solution $(\widetilde{D}_\ell, \widetilde{D}_r)$ of (5.1). Moreover, $(\widetilde{D}_\ell, \widetilde{D}_r)$ is bounded and makes the matrix

$$\begin{bmatrix} \widetilde{D}_\ell^2 & 0 \\ 0 & \widetilde{D}_r^2 \end{bmatrix} M_\alpha^{\circ 2} \begin{bmatrix} \widetilde{D}_\ell^2 & 0 \\ 0 & \widetilde{D}_r^2 \end{bmatrix}$$

a scalar multiple of a doubly stochastic matrix. Therefore, $(\widetilde{D}_\ell, \widetilde{D}_r)$ *can be computed by applying the algorithm in Appendix A to* $M_\alpha^{\circ 2}$.

Proof. We have already seen statements *a*) and *b*) in this section because the optimization problem in *b*) is just (5.2). Then we only need to prove *c*). We make the change of variables $d_{\ell_i}^2 = \exp(u_i)$ and $d_{r_j}^2 = \exp(v_j)$ for the elements of D_ℓ and D_r, respectively. Then the optimization problem (5.1) is equivalent to the optimization problem:

$$
\begin{aligned}
&\inf\ 2\sum_{i=1}^{m}\sum_{j=1}^{n} m_{ij}\exp(u_i+v_j) + \alpha^2\left(\frac{1}{m^2}\left(\sum_{i=1}^{m}\exp(u_i)\right)^2 + \frac{1}{n^2}\left(\sum_{j=1}^{n}\exp(v_j)\right)^2\right),\\
&\text{subject to}\quad \sum_{i=1}^{m} u_i + \sum_{j=1}^{n} v_j = \ln c.
\end{aligned}
\tag{5.4}
$$

The corresponding unconstrained problem with Lagrange multiplier Γ is:

$$
\begin{aligned}
&\inf 2\sum_{i=1}^{m}\sum_{j=1}^{n} m_{ij}\exp(u_i+v_j) + \alpha^2\left(\frac{1}{m^2}\left(\sum_{i=1}^{m}\exp(u_i)\right)^2 + \frac{1}{n^2}\left(\sum_{j=1}^{n}\exp(v_j)\right)^2\right)\\
&\qquad + \Gamma\left(\ln c - \sum_{i=1}^{m} u_i - \sum_{j=1}^{n} v_j\right).
\end{aligned}
\tag{5.5}
$$

The first order conditions of optimality are the equality constraint of (5.4) and the equations

$\square$

$$
\frac{\alpha^2}{m^2} d_{\ell_i}^2 \sum_{i=1}^{m} d_{\ell_i}^2 + \sum_{j=1}^{n} d_{\ell_i}^2 m_{ij} d_{r_j}^2 = \frac{\Gamma}{2}, \quad \text{and} \quad \frac{\alpha^2}{n^2} d_{r_j}^2 \sum_{j=1}^{n} d_{r_j}^2 + \sum_{i=1}^{m} d_{\ell_i}^2 m_{ij} d_{r_j}^2 = \frac{\Gamma}{2},
$$

for $i = 1, \dots, m$ and $j = 1, \dots, n$, respectively, which express that the row sum and the column sum of

$$
\begin{bmatrix} D_\ell^2 & 0 \\ 0 & D_r^2 \end{bmatrix} M_\alpha^{\circ 2} \begin{bmatrix} D_\ell^2 & 0 \\ 0 & D_r^2 \end{bmatrix}
$$

are equal to $\dfrac{\Gamma}{2}$. By Lemma 5.1, we know that $M_\alpha^{\circ 2}$ is fully indecomposable. Then, by the Sinkhorn–Knopp theorem, there exists a unique and bounded scaling (E_ℓ, E_r) that makes the matrix

$$
\begin{bmatrix} E_\ell^2 & 0 \\ 0 & E_r^2 \end{bmatrix} M_\alpha^{\circ 2} \begin{bmatrix} E_\ell^2 & 0 \\ 0 & E_r^2 \end{bmatrix}
$$

doubly stochastic. Assume that $\det E_\ell^2 \det E_r^2 = k$. We define $\widetilde{D}_\ell := \left(\frac{c}{k}\right)^{\frac{1}{2(m+n)}} E_\ell$ and $\widetilde{D}_r := \left(\frac{c}{k}\right)^{\frac{1}{2(m+n)}} E_r$. Then $\det \widetilde{D}_\ell^2 \det \widetilde{D}_r^2 = c$ and $(\widetilde{D}_\ell, \widetilde{D}_r)$ is the solution of (5.1). We can again redefine $\gamma^2 := \Gamma/2$ since this quantity is nonnegative.

For completeness, we include the following result, which is a direct corollary of the proof of Theorem 5.2.

THEOREM 5.3. *Let A and B be $m \times n$ complex matrices and $\alpha, c > 0$ be real numbers. Then the constrained minimization problem*

$$\inf_{\det D_\ell^2 \det D_r^2 = c} 2(\|D_\ell A D_r\|_F^2 + \|D_\ell B D_r\|_F^2) + \alpha^2 \left(\frac{1}{m^2}\|D_\ell\|_F^4 + \frac{1}{n^2}\|D_r\|_F^4 \right),$$

over the set $\{(D_\ell, D_r) : D_\ell := \operatorname{diag}(\delta_{\ell_1}, \ldots, \delta_{\ell_m}), D_r := \operatorname{diag}(\delta_{r_1}, \ldots, \delta_{r_n}), \delta_{\ell_i} > 0, \delta_{r_j} > 0\}$ has a unique and bounded solution. Moreover, it satisfies the equations:

$$\begin{aligned} &\|col_j(\widetilde{A})\|_2^2 + \|col_j(\widetilde{B})\|_2^2 + \frac{\alpha^2}{n^2}\delta_{r_j}^2\|D_r\|_F^2 = \gamma^2, \text{ for } j = 1, \ldots, n, \text{ and} \\ &\|row_i(\widetilde{A})\|_2^2 + \|row_i(\widetilde{B})\|_2^2 + \frac{\alpha^2}{m^2}\delta_{\ell_i}^2\|D_\ell\|_F^2 = \gamma^2, \text{ for } i = 1, \ldots, m, \end{aligned}$$

for some nonzero scalar γ, where $\widetilde{A} := D_\ell A D_r$ and $\widetilde{B} := D_\ell B D_r$.

Remark 5.4. By Theorem 5.2, we know that the row sums and the column sums of the matrix

$$\begin{bmatrix} D_\ell^2 & 0 \\ 0 & D_r^2 \end{bmatrix} M_\alpha^{\circ 2} \begin{bmatrix} D_\ell^2 & 0 \\ 0 & D_r^2 \end{bmatrix}$$

are equal to each other, where (D_ℓ, D_r) is the solution in Theorem 5.3. The quantity of such row and column sums is the scalar γ^2 appearing in Theorem 5.3.

In the following example, we consider a square pencil and we illustrate the effect of choosing the value of α in (5.3) in order to make the row and column sums of $D_\ell^2 M D_r^2$ as equal as possible. That is, to scale M as a scalar multiple of an approximate doubly stochastic matrix. For that, we use the algorithm in Appendix A with the matrix $M_\alpha^{\circ 2}$, which is essentially the Sinkhorn-Knopp algorithm applied to $M_\alpha^{\circ 2}$.

EXAMPLE 5.5. *We consider the square pencil $\lambda B_1 - A_1$ in Example 3.8. The matrix*

$$M_1 := \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

has no total support and, thus, the Sinkhorn-Knopp algorithm does not converge. However, by using the regularized approach with the matrix $M_\alpha^{\circ 2}$ and considering $\alpha = 1$ and $\alpha = 0.5$ we obtain the following scaled solutions, where we "pulled out" a scalar factor to make the comparison easier :

$$\widetilde{M_1} := 4.4817 \begin{bmatrix} 0.3143 & 0.8345 & 0 \\ 0.8345 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

with $D_\ell^2 = D_r^2 = \operatorname{diag}(1.1869, 3.1511, 2.1170)$, and

$$\widetilde{M}_{0.5} := 5.4603 \begin{bmatrix} 0.1710 & 0.8869 & 0 \\ 0.8869 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

with $D_\ell^2 = D_r^2 = \operatorname{diag}(0.9665, 5.0109, 2.3367)$. Choosing a smaller α yields a better equilibration for the row and column sums, but at the cost of a worse conditioning of the scaling and of a slower convergence. The latter is to be expected since for $\alpha = 0$ the scaling does not exist.

Remark 5.6. One could also have considered for the regularization the cost function

$$\inf_{\det D_\ell^2 \det D_r^2 = c} 2(\|D_\ell A D_r\|_F^2 + \|D_\ell B D_r\|_F^2) + \alpha^2 \left(\|D_\ell^2\|_F^2 + \|D_r^2\|_F^2\right),$$

which would correspond to the matrix

$$M_\alpha^{\circ 2} := \begin{bmatrix} \alpha^2 I_m & M \\ M^T & \alpha^2 I_n \end{bmatrix}.$$

This matrix has total support for $\alpha > 0$. However, it is not necessarily fully indecomposable (assume for instance that M has a zero row or column).

5.1. The regularized method with prescribed nonhomogeneous common vector for the row and column sums. In the nonsquare case, we know from the discussions of Section 4 that making the column and row sums of $\widetilde{M} = D_\ell^2 M D_r^2$ become equal can not be achieved exactly, where M is the matrix in (4.2). In this case, we can use the regularized method in Theorem 5.2$-c$) in order to obtain a scaling that balances $\widetilde{M}$ approximately. We have used this approach on many problems and have obtained pretty satisfactory results. However, since by using this method we always obtain a scalar multiple of a doubly stochastic matrix as solution for $M_\alpha^{\circ 2}$, this method considers in some sense the rows and columns of M in the same way, which is not natural in the rectangular case. Thus, one possible strategy for improving this approach is not to request that $M_\alpha^{\circ 2}$ is scaled to be a scalar multiple of a doubly stochastic matrix but to impose a modified scaling with prescribed common vector

$$v := \begin{bmatrix} n\mathbf{1}_m \\ m\mathbf{1}_n \end{bmatrix}$$

for the row and column sums. The new regularized method is then described by :

$$\begin{bmatrix} D_\ell^2 & 0 \\ 0 & D_r^2 \end{bmatrix} M_\alpha^{\circ 2} \begin{bmatrix} D_\ell^2 & 0 \\ 0 & D_r^2 \end{bmatrix} \begin{bmatrix} \mathbf{1}_m \\ \mathbf{1}_n \end{bmatrix} = v \tag{5.6}$$

and

$$\begin{bmatrix} \mathbf{1}_m^T & \mathbf{1}_n^T \end{bmatrix} \begin{bmatrix} D_\ell^2 & 0 \\ 0 & D_r^2 \end{bmatrix} M_\alpha^{\circ 2} \begin{bmatrix} D_\ell^2 & 0 \\ 0 & D_r^2 \end{bmatrix} = v^T. \tag{5.7}$$

This is a problem that falls into the category of scalings considered in Theorem 2.1. In addition, notice that the matrix $M_\alpha^{\circ 2}$ satisfies the hypotheses in Theorem 3.10 if $\alpha \neq 0$, i.e., $\mathrm{supp}(M_\alpha^{\circ 2}) = \mathrm{supp}((M_\alpha^{\circ 2})^T)$ and $(i,i) \in \mathrm{supp}(M_\alpha^{\circ 2})$ for all $i = 1, \cdots, n+m$. Then, by considering $\alpha \neq 0$, we know by Theorem 3.10 that there always exists a solution for this modified problem with prescribed common vector for the row and column sums. Moreover, since $M_\alpha^{\circ 2}$ is fully indecomposable, the diagonal scaling matrix is unique up to a nonnegative scalar multiple and can be chosen to be bounded according to Remark 3.7. It can also be computed using a Sinkhorn-like algorithm. A MATLAB code is given in Appendix B. Though we have not proved yet that this algorithm converges under the conditions of Theorem 3.10, we have checked that it works very well in practice.

Notice that, when $\alpha = 0$, this problem reduces to the problem discussed in Section 4. Then, for very small α, the regularized scaling with prescribed row and column sums v tends to the scaling problem explained in Section 4, which does not always have a solution.

In the following example, we illustrate the effect of choosing different values of α and the row and column sum conditions (5.6) and (5.7).

EXAMPLE 5.7. *We remark that, for this example, the algorithm used in Section 4 (Appendix A) converges. Then there is no need to use the regularized method. Nevertheless, we use the regularized method developed in this section with two purposes: (1) for comparing the approximate regularized solution and the exact solution of the optimization problem in Theorem 4.1 and (2) for illustrating the effect of choosing different values of α. We will see that the regularized method yields very satisfactory results for small values of α. We consider again the nonsquare pencil $\lambda B - A$ in Example 4.2 but now with a preliminary diagonal scaling $\lambda\hat{B} - \hat{A} := \hat{D}_\ell(\lambda B - A)\hat{D}_r$ on the left and the right with condition numbers $\kappa(\hat{D}_\ell) \approx \kappa(\hat{D}_r) \approx 100$. The resulting matrix $M := \hat{A}^{\circ 2} + \hat{B}^{\circ 2}$ to be scaled is*

$$M = \begin{bmatrix} 8.617e-03 & 1.045e-01 & 0 & 0 & 0 & 0 \\ 0 & 2.125e-01 & 1.380e-03 & 0 & 0 & 0 \\ 0 & 0 & 3.386e-07 & 7.973e-07 & 0 & 0 \\ 0 & 0 & 0 & 1.087e-02 & 1.000e+00 & 0 \\ 0 & 0 & 0 & 0 & 3.191e-03 & 2.014e-05 \end{bmatrix}$$

which we normalized to have its largest element equal to 1. When applying the method described in Subsection 4 we obtained (with three digits of accuracy) the same result as in Example 4.2, but requiring scalings with condition numbers $\kappa(D_\ell) = 143$, $\kappa(D_r) = 28.1$:

$$\widetilde{M} = \begin{bmatrix} 5.000 & 1.000 & 0 & 0 & 0 & 0 \\ 0 & 4.000 & 2.000 & 0 & 0 & 0 \\ 0 & 0 & 3.000 & 3.000 & 0 & 0 \\ 0 & 0 & 0 & 2.000 & 4.000 & 0 \\ 0 & 0 & 0 & 0 & 1.000 & 5.000 \end{bmatrix}.$$

This indicates that the scaling method can compensate for a bad initial scaling.

We now apply the regularized method with the matrix $M_\alpha^{\circ 2}$ and prescribed common vector $v := [6,6,6,6,6,5,5,5,5,5]^T$ for the row and column sums.

First, we consider $\alpha = 0.001$. Then we obtain $\kappa(D_\ell) = 110$, $\kappa(D_r) = 20.0$ and

$$\widetilde{M}_{0.001} = \begin{bmatrix} 4.848e{+}00 & 1.152e{+}00 & 0 & 0 & 0 & 0 \\ 0 & 3.845e{+}00 & 2.155e{+}00 & 0 & 0 & 0 \\ 0 & 0 & 2.587e{+}00 & 2.307e{+}00 & 0 & 0 \\ 0 & 0 & 0 & 2.596e{+}00 & 3.404e{+}00 & 0 \\ 0 & 0 & 0 & 0 & 1.594e{+}00 & 4.392e{+}00 \end{bmatrix}.$$

With $\alpha = 0.025$ we obtain $\kappa(D_\ell) = 16.0$, $\kappa(D_r) = 13.1$ and

$$\widetilde{M}_{0.025} = \begin{bmatrix} 4.258e{+}00 & 1.587e{+}00 & 0 & 0 & 0 & 0 \\ 0 & 3.391e{+}00 & 2.446e{+}00 & 0 & 0 & 0 \\ 0 & 0 & 2.180e{-}02 & 4.827e{-}02 & 0 & 0 \\ 0 & 0 & 0 & 2.571e{+}00 & 3.406e{+}00 & 0 \\ 0 & 0 & 0 & 0 & 1.560e{+}00 & 1.115e{+}00 \end{bmatrix}.$$

These examples show that increasing α makes the scalings better conditioned for the regularization technique. Also one can see that for very small α, the regularized scaling with prescribed row and column sums v tends to the result of the scaling technique explained in Section 4.

In general, the necessary and sufficient conditions for the scaling technique in Section 4 (Appendix A) to converge are not known for the nonquare case. In contrast, the regularized method with the matrix $M_\alpha^{\circ 2}$ and prescribed common vector v always has a solution and the previous example, as well as many others, shows that it produces very satisfactory results. Therefore, using this new regularized method is a good option for scaling M in any case, i.e., either when the optimization problem in Theorem 4.1 has a solution or not.

In Example 5.7, we knew that the corresponding matrix M can be scaled with prescribed vectors $r := [6, 6, 6, 6, 6]^T$, for the row sums, and $c := [5, 5, 5, 5, 5, 5]^T$, for the column sums. We now consider the matrix M in Example 2.2 that can not be scaled, but we use the regularized method with prescribed common vector for the row and column sums to obtain an approximate scaling.

EXAMPLE 5.8. *We consider the nonsquare matrix*

$$M := \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

in Example 2.2, that can not be scaled with prescribed vectors $r := [3, 3]^T$*, for the row sums, and* $c := [2, 2, 2]^T$*, for the column sums. Therefore, the algorithm in Section 4 does not work on this matrix. Then we use the regularized approach with* $\alpha = 0.05$ *and prescribed common vector* $v := [3, 3, 2, 2, 2]^T$ *for the row and column sums of* $M_\alpha^{\circ 2}$*, and we obtain the following scaled approximate solution:*

$$\widetilde{M}_{0.05} := \begin{bmatrix} 1.4981 & 1.4981 & 0.0025 \\ 0 & 0 & 1.9967 \end{bmatrix},$$

with $D_\ell^2 = \operatorname{diag}(0.0499, 40.0414)$ *and* $D_r^2 = \operatorname{diag}(30.0913, , 30.0913, 0.0499)$*. Note that the row sums of* $\widetilde{M}_{0.05}$ *are much closer to each other than those of* M *and that the same happens for the column sums.*

Let us now look at the effect of the two sided scaling on the sensitivity of the underlying eigenvalue problem. In the case of regular pencils, we argued [9] that the minimization problem

$$\inf_{\det T_\ell^2 \det T_r^2 = 1} \|T_\ell(\lambda B - A)T_r\|_F^2$$

over the arbitrary nonsingular matrix pairs (T_ℓ, T_r), yielded nearly optimal sensitivity for the generalized eigevalues of the pencil. But since the eigenvalue problem for a singular pencil is known to be ill-conditioned, this may not make sense anymore. Nevertheless, if we constrain the transformations to be bounded, then the Kronecker structure can not change anymore, and it makes then sense to talk about the sensitivity of the eigenvalues again. In the numerical examples we show that the scaling also improves the sensitivity of the eigenvalues of the regular part of a singular pencil.

6. Numerical examples.

In this section we compare the precision of the computed eigenvalues of pencils without scaling and after applying one of our proposed scaling procedures. The first experiment is for regular pencils and the second one for singular pencils. In both cases the pencils are square and the scaling procedure used is the first one in Section 5, that is, scaling the matrix $M_\alpha^{\circ 2}$ to a multiple of a double stochastic matrix with the algorithm in Appendix A.

6.1. Regular example.

We generated random diagonalizable pencils of the form $T_\ell(\lambda\Lambda_B - \Lambda_A)T_r$ where $(\lambda\Lambda_B - \Lambda_A)$ is in standard normal form [9], i.e., Λ_A and Λ_B are diagonal, and $|\Lambda_A|^2 + |\Lambda_B|^2 = \gamma^2 I_n$ with γ a real number. The condition number of the random square nonsingular matrices T_ℓ and T_r was controlled by taking the kth power of normally distributed random numbers $r_{i,j}$ as their elements. A larger power k then typically yields a larger condition number for the diagonalizing transformation. In order to guarantee that the eigenvalues of $T_\ell(\lambda\Lambda_B - \Lambda_A)T_r$ and $(\lambda\Lambda_B - \Lambda_A)$ are "numerically" equal, the product $T_\ell(\lambda\Lambda_B - \Lambda_A)T_r$ was computed with a extended precision of 64 decimal digits by using the `vpa` command of MATLAB and, then, rounded to standard double precision [1].

We applied the QZ-algorithm [11] in MATLAB to such pencils of dimension 10×10, and for values of k going from 1 to 15. We compared the "exact" eigenvalues λ_i of the pencil $(\lambda\Lambda_B - \Lambda_A)$ with the computed eigenvalues $\tilde{\lambda}_i$ of the pencil $(\lambda B - A) := T_\ell(\lambda\Lambda_B - \Lambda_A)T_r$, and we did the same for the "scaled" pencil $D_\ell(\lambda B - A)D_r$. We constrained the diagonal elements of D_ℓ, D_r to be powers of two, and used $\alpha = 1$ for the regularizing parameter used to define $M_\alpha^{\circ 2}$. For the comparison of the eigenvalues, we used their chordal distances [15]

$$c_i := \chi(\lambda_i, \tilde{\lambda}_i) := \frac{|\lambda_i - \tilde{\lambda}_i|}{\sqrt{1+|\lambda_i|^2}\sqrt{1+|\tilde{\lambda}_i|^2}}.$$

We compared the quantities $c := \|[c_1, \ldots, c_n]\|_2$ for the original pencil $(\lambda B - A)$ (c_{orig}) and for the balanced pencil constructed by our algorithm (c_{bal}).

In Table 1 we give in each row the condition numbers $\kappa(T_\ell)$ and $\kappa(T_r)$ in the 2-norm, the 2-norms of the matrices M_{orig} and M_{bal}, the norms of the perturbation vectors c_{orig} and c_{bal}, and their ratio. This experiment shows that the scaling does improve the sensitivity of the eigenvalues, especially when the pencil has badly conditioned left and right diagonalizing transformations T_ℓ and T_r.

Table 1
Eigenvalue sensitivity of the QZ-algorithm for regular pencils

$\kappa(T_\ell)$	$\kappa(T_r)$	$\|M_{orig}\|_2$	$\|M_{bal}\|_2$	c_{orig}	c_{bal}	c_{bal}/c_{orig}
5.171e+01	8.484e+03	1.531e+02	1.870e+01	2.170e-12	6.045e-13	2.784e-01
6.742e+01	1.054e+02	1.401e+03	2.176e+01	4.594e-15	1.315e-14	2.863e+00
2.979e+02	4.080e+01	2.758e+04	2.261e+01	9.133e-15	6.545e-15	7.165e-01
1.394e+03	8.747e+02	1.044e+08	1.685e+01	1.478e-13	7.387e-14	4.996e-01
1.809e+05	7.192e+02	5.135e+08	1.387e+01	2.690e-12	2.579e-13	9.590e-02
1.682e+03	3.006e+04	5.364e+09	1.827e+01	3.104e-11	3.211e-14	1.034e-03
1.166e+04	1.450e+06	4.410e+12	2.785e+01	1.418e-10	4.825e-13	3.401e-03
2.433e+02	8.521e+05	8.136e+12	2.726e+01	9.471e-13	2.885e-15	3.046e-03
1.454e+06	1.634e+03	6.148e+14	3.724e+01	2.647e-11	4.300e-15	1.624e-04
2.570e+04	9.287e+03	3.807e+13	2.052e+01	1.484e-13	2.998e-14	2.019e-01
6.963e+03	4.208e+05	4.537e+16	3.019e+01	6.837e-12	3.725e-14	5.448e-03
2.594e+10	1.018e+07	9.784e+23	2.227e+01	2.308e-05	2.228e-12	9.651e-08
8.058e+05	4.516e+09	1.270e+15	2.187e+01	8.184e-12	8.085e-14	9.878e-03
2.248e+08	8.017e+07	7.765e+19	2.043e+01	2.783e-10	6.294e-13	2.261e-03
6.997e+08	2.613e+12	7.880e+22	2.395e+01	3.557e-03	2.839e-12	7.981e-10

6.2. Singular example. In the second experiment we replaced one of the diagonal pairs of the pencil $(\lambda\Lambda_B - \Lambda_A)$ generated in the regular example by two zeros, creating thus a singular pencil. Each transformed pencil $(\lambda B - A) := T_\ell(\lambda\Lambda_B - \Lambda_A)T_r$ is therefore also singular, but its left and right rational null spaces are both of dimension 1 and their minimal bases are formed by constant vectors [17]. For that reason,

[1] MATLAB Version: R2019a.

the regular part of that singular pencil has dimension 9×9 and its eigenvalues are the remaining 9 eigenvalues of $(\lambda\Lambda_B - \Lambda_A)$. If we follow the same procedure as in the first experiment, the QZ-algorithm applied to $(\lambda B - A)$ should in principle yield arbitrary eigenvalues, since it is known that the QZ-algorithm is backward stable and that there exist arbitrarily small perturbations of square singular pencils that make them regular, but with arbitrary spectrum in the complex plane [17]. However, it has been shown that such perturbations are very particular, and that, generically, tiny perturbations of a singular square pencil makes it regular with eigenvalues that are tiny perturbations of the eigenvalues of the unperturbed singular pencil, together with some other "arbitrary" eigenvalues determined by the perturbation [3, 4]. Even more, starting from these ideas, it has been shown very recently that it is possible to define sensible and useful "weak" condition numbers for the eigenvalues of a singular square pencil [10]. This explains the well-known fact that, in practice, the QZ-algorithm applied to a singular square matrix pencil finds almost always its eigenvalues, albeit with some loss of accuracy. Therefore, it makes sense to apply the QZ algorithm to our generated singular pencils as well as to their scaled versions. The numerical results are reported in Table 2. We generated the data just as in the previous experiment for regular pencils, except for the one eigenvalue replaced by 0/0 or, in other words, by NaN. When comparing the "original" spectrum with the computed one, we excluded NaN in the original set and looked for the best matching 9 eigenvalues in the "computed" spectrum. It appears from this Table that a few digits of accuracy may get lost, and that balancing still seems to improve the sensitivity and the accuracy of those eigenvalues in most cases.

TABLE 2
Eigenvalue sensitivity of the QZ-algorithm for singular pencils

$\kappa(T_\ell)$	$\kappa(T_r)$	$\|M_{orig}\|_2$	$\|M_{bal}\|_2$	c_{orig}	c_{bal}	c_{bal}/c_{orig}
5.171e+01	8.484e+03	1.453e+02	2.070e+01	2.770e-09	5.458e-15	1.970e-06
6.742e+01	1.054e+02	1.493e+03	1.306e+01	3.474e-14	1.794e-14	5.165e-01
2.979e+02	4.080e+01	2.757e+04	3.276e+01	1.296e-12	8.918e-14	6.877e-02
1.394e+03	8.747e+02	1.044e+08	2.205e+01	1.240e-12	9.857e-13	7.944e-01
1.809e+05	7.192e+02	5.162e+08	2.341e+01	1.810e-14	1.223e-14	6.760e-01
1.682e+03	3.006e+04	5.363e+09	2.303e+01	3.131e-12	2.375e-10	7.585e+01
1.166e+04	1.450e+06	4.410e+12	3.151e+01	3.565e-11	2.128e-14	5.970e-04
2.433e+02	8.521e+05	4.701e+12	2.134e+01	1.288e-10	7.383e-14	5.729e-04
1.454e+06	1.634e+03	6.148e+14	3.311e+01	8.683e-10	1.055e-14	1.215e-05
2.570e+04	9.287e+03	3.807e+13	2.783e+01	1.456e-10	1.545e-14	1.061e-04
6.963e+03	4.208e+05	4.537e+16	3.152e+01	1.056e-11	4.272e-15	4.044e-04
2.594e+10	1.018e+07	6.869e+17	3.483e+01	3.163e-07	7.220e-12	2.282e-05
8.058e+05	4.516e+09	1.270e+15	2.756e+01	9.551e-07	4.447e-14	4.656e-08
2.248e+08	8.017e+07	7.765e+19	4.495e+01	3.060e-10	4.231e-10	1.382e+00
6.997e+08	2.613e+12	7.880e+22	2.110e+01	1.757e-02	1.141e-11	6.495e-10

Though the direct use of the QZ-algorithm is a simple option for computing the eigenvalues of a singular square pencil when the accuracy requirements are moderate, the correct handling of a singular pencil is to first "deflate" its left and right null spaces, and then compute the spectrum of the regular part of that singular pencil, i.e., to apply the staircase algorithm (see [17]). In this experiment, it turns out that the left and right null spaces are one-dimensional and are given, respectively, by the left null vector of $[\,A\ B\,]$, and by the right null vector of $\left[\begin{smallmatrix} A \\ B \end{smallmatrix}\right]$, which we both computed using a singular value decomposition of these compound matrices. After this deflation applied to both pencils $(\lambda B - A)$ and the scaled pencil $D_\ell(\lambda B - A)D_r$, we again computed the spectrum of the deflated pencils with the QZ-algorithm. The results for the same data as reported in Table 2 are now reported in Table 3. Their comparison shows that the deflation of the singular spaces improves the sensitivity a lot and that balancing improves the sensitivity in virtually all cases, getting very often errors of order machine precision. We also added two columns with the sensitivities of the deflation in the original pencil γ_{orig} and of the balanced pencil γ_{bal}. We measured the sensitivity of the left and right null vectors defining the deflation of a singular pencil $\lambda B - A$, by

$$\gamma := \max\left(\frac{\sigma_n \left[\begin{smallmatrix} A \\ B \end{smallmatrix}\right]}{\sigma_{n-1} \left[\begin{smallmatrix} A \\ B \end{smallmatrix}\right]}, \frac{\sigma_n \left[\begin{smallmatrix} A & B \end{smallmatrix}\right]}{\sigma_{n-1} \left[\begin{smallmatrix} A & B \end{smallmatrix}\right]}\right),$$

i.e. the largest ratio between the two smallest singular values of the matrices that define these null vectors. It is an indication about how much these vectors can rotate when perturbing the pencil. It is easy to see from these data that the sensitivity of the eigenvalues of the deflated pencil is at least as bad as that of the deflation itself and that they are in fact closely related.

TABLE 3
Eigenvalue sensitivity for the regular part of singular pencils

c_{orig}	c_{bal}	c_{bal}/c_{orig}	γ_{orig}	γ_{bal}
5.7270e-16	9.0827e-16	1.5860e+00	1.4372e-15	1.0765e-15
1.5614e-15	9.6238e-16	6.1637e-01	3.7084e-15	2.8895e-15
1.6909e-15	8.2873e-16	4.9011e-01	6.2481e-15	2.7506e-15
8.2181e-14	5.7334e-15	6.9766e-02	1.8274e-13	2.9763e-14
3.1811e-15	2.7685e-15	8.7028e-01	1.8279e-14	7.0236e-15
2.6074e-12	1.8996e-14	7.2852e-03	3.9490e-12	1.2209e-13
9.9867e-14	5.7101e-15	5.7177e-02	1.2222e-10	5.6897e-14
2.0725e-13	1.6116e-15	7.7758e-03	1.3432e-11	9.9270e-15
6.4841e-13	6.5138e-16	1.0046e-03	3.2226e-13	3.2572e-15
1.7256e-13	9.2556e-16	5.3637e-03	1.8111e-14	1.0455e-15
1.9028e-13	9.8758e-16	5.1900e-03	7.1167e-12	2.9866e-15
1.4413e-13	5.8822e-13	4.0810e+00	2.5475e-10	1.2939e-12
1.5763e-13	1.1152e-15	7.0745e-03	1.6241e-15	2.7182e-15
9.1023e-13	5.9615e-13	6.5494e-01	1.7104e-12	9.6265e-14
1.1542e-06	6.6007e-16	5.7187e-10	1.9999e-12	1.2245e-15

These two experiments show that balancing improves the sensitivity of the eigenvalue computation of both regular and singular pencils as well as the sensitivity of the deflation of the regular part of a singular pencil. We briefly mention that recently an alternative robust method to the staircase algorithm has been proposed for computing the eigenvalues of singular pencils [7]. This new method is related to the ideas in [3, 4, 10] and its accuracy will also improve by using our scaling strategy.

7. Concluding remarks. In this paper, we developed a new scaling technique that applies to both regular and singular pencils. The method is a modified Sinkhorn-Knopp algorithm applied to a certain regularized problem and is guaranteed to have a unique and bounded solution, which also improves on earlier methods for the scaling of regular pencils. Finally, the algorithm computing this scaling has a complexity that is negligeable with respect to the complexity of the subsequent generalized eigenvalue problem. The method computes D_ℓ and D_r in an alternating fashion, until convergence is met.

Appendix A : Sinkhorn-Knopp-like algorithm MATLAB code.

```
function [Md,dleft,dright,error] = sinkhorn(M,maxiter,tol)
%
% [Md,dleft,dright,error] = sinkhorn(M,maxiter,tol)
%
% implements a Sinkhorn-Knopp-like algorithm for
% scaling a non-negative mxn matrix M such that
%
%        Md:=diag(dleft)*M*diag(dright)
%
%  has column sums equal to m and row sums equal to n
%
%  The iterative process is stopped as soon as the incremental
%  scalings are tol-close to the identity. The error vector
%  also shows the convergence pattern of the iterative scalings
%
%  Input : M, a nonnegative mxn matrix
%          maxiter, the maximum number of iterations
%          tol, a tolerance for the transformation updates
%  Output: Md, a matrix with equal row sums and equal column sums
%          dleft and dright, the diagonals of the left and right
%          scalings error, the convergence error
%
[m,n]=size(M);error=[];
% First scale the matrix to have the sum of all its entries 1
sumM=sum(sum(M));Md=M/sumM;
dleft=ones(m,1)/sqrt(sumM);dright=ones(1,n)/sqrt(sumM);
% Then scale left and right to make row and column sums 1
for i=1:maxiter;
dr=sum(Md,1)/m;Md=Md./dr;er=min(dr)/max(dr);dright=dright./dr;
dl=sum(Md,2)/n;Md=dl.\Md;el=min(dl)/max(dl);dleft=dleft./dl;
error=[error er el];if 2-(er+el) < tol, break; end
end
% Finally scale the two scalings to have equal maxima
scaled=sqrt(max(dright)/max(dleft));
dleft=dleft*scaled;dright=dright'/scaled;
end
```

Appendix B : Sinkhorn-Knopp-like algorithm MATLAB code with prescribed row sums and column sums.

```
function [Md,dleft,dright,error] = rowcolsums(M,r,c,maxiter,tol)
%
% [Md,dleft,dright,error] = rowcolsums(M,r,c,maxiter,tol)
%
% implements a Sinkhorn-Knopp-like algorithm for
% scaling a non-negative mxn matrix M such that
%
%        Md:=diag(dleft)*M*diag(dright)
%
%  has column sums equal to a row vector c
```

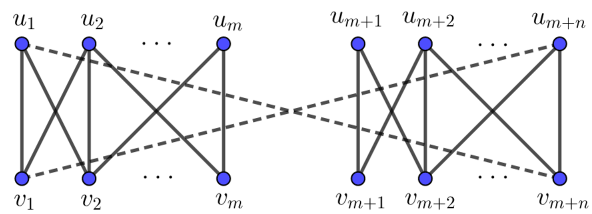
```
%  row sums equal to a column vector r where sum(c)=sum(r)
%
%  The iterative process is stopped as soon as the incremental
%  scalings are tol-close to the identity. The error vector
%  also shows the convergence pattern of the iterative scalings
%
%  Input : M, a nonnegative mxn matrix
%          r, a positive mx1 column vector and
%          c, a positive 1xn row vector
%               satisfying sum(c)=sum(r)
%          maxiter, the maximum number of iterations
%          tol, a tolerance for the transformation updates
%  Output: Md, a nonnegative matrix with row sums r and column sums c
%          dleft and dright, the diagonals of the left and right
%          scalings error, the convergence error
%
[m,n]=size(M);error=[];
% First scale the matrix to have total sum(sum(M))=sum(c)=sum(r);
sumcr=sum(c);sumM=sum(sum(M));Md=M*sumcr/sumM;
dleft=ones(m,1)*sqrt(sumcr/sumM);dright=ones(1,n)*sqrt(sumcr/sumM);
% Then scale left and right to make row and column sums equal to r
% and c
for i=1:maxiter;
dr=sum(Md,1)./c;Md=Md./dr;er=min(dr)/max(dr);dright=dright./dr;
dl=sum(Md,2)./r;Md=dl.\Md;el=min(dl)/max(dl);dleft=dleft./dl;
error=[error er el];if 2-(er+el) < tol, break; end
end
% Finally scale the two scalings to have equal maxima
scaled=sqrt(max(dright)/max(dleft));
dleft=dleft*scaled;dright=dright'/scaled;
end
```

Appendix C : Proof of Lemma 5.1.

Proof. $M_\alpha^{\circ 2}$ has total support for all $\alpha \neq 0$ since every nonzero element is an element of a positive diagonal. To see that $M_\alpha^{\circ 2}$ is fully indecomposable, we apply [2, Theorem 1.3.7]. This theorem states that a square matrix with total support is fully indecomposable if and only if its bipartite graph is connected. Then we consider the bipartite graph of $M_\alpha^{\circ 2}$, denoted by $BG(M_\alpha^{\circ 2})$. We assume without lost of generality that m_{1n} is a nonzero element of $M := [m_{ij}]$. Then we consider the matrix

$$N := \left[\begin{array}{c|c} \frac{\alpha^2}{m^2} 1_m 1_m^T & \begin{matrix} 0 & m_{1n} \\ 0 & 0 \end{matrix} \\ \hline \begin{matrix} 0 & 0 \\ m_{1n} & 0 \end{matrix} & \frac{\alpha^2}{n^2} 1_n 1_n^T \end{array}\right].$$

Notice that $BG(N)$ is a sub-graph of $BG(M_\alpha^{\circ 2})$. Moreover, if $\{u_1, u_2, \dots, u_{m+n}\}$ and $\{v_1, v_2, \dots, v_{m+n}\}$ are the sets of vertices associated with the rows and columns of N, respectively, then $BG(N)$ is of the form



where the left and right groups of solid edges are each bicliques (and hence connected) and where the two dashed edges correspond to the element m_{1n}. This proves that $BG(N)$ is connected, since the dashed edges make a connection between two connected components. Therefore, $BG(M_\alpha^{\circ 2})$ is connected and, by [2, Theorem 1.3.7], $M_\alpha^{\circ 2}$ is fully indecomposable. $\square$

REFERENCES

[1] R. A. Brualdi, Matrices of 0's and 1's with total support, J. Combin. Theory Ser. A, 28 (1980) 249–256.

[2] R. A. Brualdi, Combinatorial Matrix Classes, Cambridge University Press, New York, 2006.

[3] F. De Terán and F. M. Dopico, First order spectral perturbation theory of square singular matrix polynomials, Linear Algebra Appl., 432(4) (2010) 892–910.

[4] F. De Terán, F. M. Dopico and J. Moro, First order spectral perturbation theory of square singular matrix pencils, Linear Algebra Appl., 429 (2008) 548-576.

[5] G. H. Golub and C. F. Van Loan, Matrix Computations, Johns Hopkins Studies in the Mathematical Sciences, Johns Hopkins University Press, Baltimore, MD, 4th ed., 2013.

[6] D. J. Hartfiel and C. J. Maxson, The chainable matrix, a special combinatorial matrix, Discrete Math., 12 (1975) 245–256.

[7] M. E. Hochstenbach, C. Mehl, and B. Plestenjak, Solving singular generalized eigenvalue problems by a rank-completing perturbation, SIAM J. Matrix Anal. Appl., 40(3) (2019) 1022-1046.

[8] P. A. Knight, The Sinkhorn-Knopp Algorithm: Convergence and Applications, SIAM J. Matrix Anal. Appl., 30(1) (2008) 261–275.

[9] D. Lemonnier and P. Van Dooren, Balancing regular matrix pencils, SIAM J. Matrix Anal. Appl., 28(1) (2006) 253–263.

[10] M. Lotz and V. Noferini, Wilkinson's bus: Weak condition numbers, with an application to singular polynomial eigenproblems, published online in Foundations of Computational Mathematics `https://doi.org/10.1007/s10208-020-09455-y`

[11] C. B. Moler and G. W. Stewart, An algorithm for generalized matrix eigenvalue problems, SIAM J. Numer. Anal., 10(2) (1971) 241-256.

[12] B. Parlett and C. Reinsch, Balancing a matrix for the calculation of eigenvalues and eigenvectors, Numer. Math., 13 (1969) 293–304.

[13] U. G. Rothblum and H. Schneider, Scalings of matrices which have prespecified row sums and column sums via optimization, Linear Algebra Appl., 114/115 (1989) 737–764.

[14] R. Sinkhorn and P. Knopp, Concerning nonnegative matrices and doubly stochastic matrices, Pacific J. Math., 21(2) (1967) 343–348.

[15] G. W. Stewart and J.-G. Sun, Matrix Perturbation Theory, Academic Press, Boston, 1990.

[16] The MathWorks, Inc., Natick, MA: MATLAB Reference Guide (1992).

[17] P. Van Dooren, The computation of Kronecker's canonical form of a singular pencil. Linear Algebra Appl., 27 (1979) 103–141.

[18] R. C. Ward, Balancing the generalized eigenvalue problem, SIAM J. Sci. Stat. Comput., 2 (1981) 141–152.