Max-Balancing Hungarian Scalings

Hook, James and Pestana, Jennifer and Tisseur, Francoise

2015

MIMS EPrint: **2015.36**

Manchester Institute for Mathematical Sciences
School of Mathematics

The University of Manchester

Reports available from: http://eprints.maths.manchester.ac.uk/
And by contacting: The MIMS Secretary
School of Mathematics
The University of Manchester
Manchester, M13 9PL, UK

ISSN 1749-9097

MAX-BALANCING HUNGARIAN SCALINGS*

JAMES HOOK†, JENNIFER PESTANA†, AND FRANÇOISE TISSEUR†

Abstract. A Hungarian scaling is a diagonal scaling of a matrix that is typically applied along with a permutation to a sparse symmetric or nonsymmetric indefinite linear system before calling a direct or iterative solver. A Hungarian scaled and reordered matrix has all its entries of modulus less than or equal to 1 and entries of modulus 1 on the diagonal. We use max-plus algebra to characterize the set of all Hungarian scalings for a given matrix and show that max-balancing a Hungarian scaled matrix yields the most "diagonally dominant" Hungarian scaled matrix possible with respect to some ordering. We also propose a new scaling, called centre of mass scaling, which can be seen as an approximate max-balancing Hungarian scaling and whose computation is embarrassingly parallel. Numerical experiments illustrate the increased diagonal dominance produced by max-balancing and centre of mass scaling of Hungarian scaled matrices as well as the reduced need for pivoting in Gaussian elimination with partial pivoting and the improved stability of LU factorizations without pivoting.

Key words. max-plus algebra, diagonal scaling, Hungarian scaling, max-balancing, diagonal dominance, centre of mass scaling, linear systems of equations, sparse matrices.

AMS subject classifications. 65F35 15A12, 15A60, 15A80.

1. Introduction.

A Hungarian scaling is a two-sided diagonal scaling of a matrix that can be applied along with a permutation P to a linear system $Ax = b$, with $A \in \mathbb{C}^{n\times n}$ and $b \in \mathbb{C}^n$, yielding

$$H = PD_1AD_2, \quad Hy = PD_1b, \quad x = D_2y,$$

where $D_1, D_2 \in \mathbb{R}^{n\times n}$ are diagonal and nonsingular. The scaled and reordered matrix $H = (h_{ij})$ is such that $|h_{ij}| \le 1$ and $|h_{ii}| = 1$ for $i, j = 1, \ldots, n$.

Benzi, Haws, and Tůma [1] show that Hungarian scaling is an effective preprocessing step before applying BiCGSTAB, GMRES, or TFQMR to sparse indefinite nonsymmetric matrices. The scaled matrices require significantly fewer iterations for convergence. The authors explain this phenomenon by pointing out that the Hungarian scaled matrix H tends to be more diagonally dominant than the original matrix A. The authors also experiment with using Hungarian scaling as a preprocessing step before applying preconditioned BiCGSTAB with an ILU preconditioner. Without scaling they show that there are many problems for which attempts to compute a very sparse ILU preconditioner break down. In order to reliably compute effective ILU preconditioners they are therefore forced to compute less sparse ILU factors at a considerably increased cost. However they show that after Hungarian scaling has been applied they are able to reliably compute very sparse ILU preconditioners.

Olschowka and Neumaier [15] proposed applying the Hungarian scaling together with a permutation to matrices prior to performing Gaussian elimination. Duff and Koster [6], [7] provide an efficient implementation of the Hungarian algorithm for sparse matrices, on which the HSL code MC64 is based [6], [7], [14], and analyze its performance for Gaussian elimination and preconditioned iterative methods. In [10]

*Version of June 8, 2015. This work was supported by Engineering and Physical Sciences Research Council grant EP/I005293. The work of the third author was supported by a Royal Society-Wolfson Research Merit Award.

†School of Mathematics, The University of Manchester, Manchester, M13 9PL, UK. (james.hook@manchester.ac.uk, jennifer.pestana@manchester.ac.uk, francoise.tisseur@manchester.ac.uk).

and [11] Hogg and Scott show that Hungarian scaling improves the stability of the LU factorization of sparse indefinite symmetric matrices and significantly reduces the need for pivoting, so that after Hungarian scaling many example problems can be factorized without the need for any pivoting at all. This can also be explained by the fact that the Hungarian scaled matrix tends to be more diagonally dominant. The largest entries in the matrix are on the diagonal and as a result pivoting is often not necessary.

Hungarian scaling, or for that matter any diagonal scaling of a matrix, can be better understood using max-plus algebra. Schneider and his collaborators have developed a considerable theory for matrix scalings and weighted graph potentials, which are equivalent to similarity scalings of the graph's weighted adjacency matrix (see for example [3], [17], [18]). In this paper we adapt some of these ideas to the problem of finding optimal Hungarian scalings for the linear system $Ax = b$.

An important point that has been overlooked in the previous numerical linear algebra literature is that the Hungarian scaling associated with a matrix $A \in \mathbb{C}^{n\times n}$ is not unique. In general there will be many different diagonal matrix pairs $D_1, D_2 \in \mathbb{R}^{n\times n}$, resulting in different Hungarian scaled matrices that may behave quite differently in direct factorizations or iterative methods. What does the set of all Hungarian scalings of a matrix look like? How do we choose the best possible Hungarian scaling for a particular problem?

The Hungarian scalings of a matrix A are all related by diagonal similarities, so that if $H = D_1PAD_2$ and $H' = D_1'PAD_2'$ are both Hungarian scaled then there exists a diagonal matrix S such that $H' = S^{-1}HS$. Starting from H, we can generate new Hungarian scaled matrices by applying "special" diagonal similarities to H: S must be such that H' is a Hungarian scaled matrix, i.e., $|h_{ij}'| \leq 1$ and $|h_{ii}'| = 1$ for all i, j. The conditions on S are very naturally expressed in terms of max-plus algebra and this is why it proves so useful here.

Since the increased diagonal dominance of the Hungarian scaled matrices has been repeatedly cited as being responsible for their improved numerical characteristics we will focus on trying to obtain Hungarian scaled matrices that are as diagonally dominant as possible. To this end we consider max-balancing Hungarian scalings. Max-balancing graphs were introduced by Schneider and Schneider in connection with certain network flow problems [17]. A directed weighted graph is max-balanced if for any subset of vertices the maximum weight of an edge into that subset is equal to the maximum weight of an edge out of that subset. A matrix $A \in \mathbb{C}^{n\times n}$ is max-balanced if and only if its precedence graph is max-balanced. We can use the max-balancing algorithm of Schneider and Schneider to compute a diagonal matrix $S \in \mathbb{R}^{n\times n}$ such that the scaled matrix $B = S^{-1}AS$ is max-balanced. We show in Section 4 that max-balancing (a) preserves the property of a matrix being Hungarian scaled and (b) minimizes the entrywise p-norm over all diagonal similarity scalings of A in the limit as p tends to infinity. As a result, the max-balancing of a Hungarian scaled matrix produces, in the entrywise p-norm sense, the most "diagonally dominant" Hungarian scaled matrix.

The cost of computing the max-balancing Hungarian scaling of $A \in \mathbb{C}^{n\times n}$ is equal to the cost of computing the initial Hungarian scaling plus the cost of computing the max-balancing scaling. Both of these steps have worst case complexity $O\big(n\tau + n^2 \log n\big)$, where τ is the number of nonzero entries in A. In practice, the cost of computing the initial Hungarian scaling is typically much smaller, namely $O(\tau)$. It is not yet clear from our investigations whether the cost of the max-balancing al-

gorithm typically meets the worst case complexity bound. However it is clear that the algorithm is difficult to parallelize so that faster alternative scalings may be of practical interest. One possibility proposed in Section 5 is the centre of mass scaling, which is a diagonal similarity scaling that aims to approximate the max-balancing scaling. The cost of computing the centre of mass scaling is the same as that of the max-balancing scaling but the computation is comparatively straightforward and easily parallelized.

To demonstrate the effectiveness of max-balancing and centre of mass Hungarian scalings we include numerical experiments in Section 6. We focus on solving $Ax = b$ via LU factorization, where $A \in \mathbb{C}^{n\times n}$ is sparse and nonsymmetric. Our experiments confirm that max-balancing improves diagonal dominance and that the condition numbers and number of row interchanges in Gaussian elimination with partial pivoting reduced by Hungarian scaling are further reduced by max-balancing Hungarian scaling. The centre of mass scaling tends to do nearly as well as the max-balancing scaling.

2. Background material. Max-plus algebra concerns the max-plus semiring $\mathbb{R}_{\max} = \mathbb{R} \cup \{-\infty\}$ along with the binary operations max and plus,

$$a \oplus b = \max\{a, b\}, \quad a \otimes b = a + b, \quad \text{for all } a, b \in \mathbb{R}_{\max},$$

and additive and multiplicative identities $-\infty$ and 0 ($a \oplus -\infty = a$, $a \otimes 0 = a$).

Throughout this paper we use calligraphic letters for elements in $\mathbb{R}_{\max}$, including matrices. A max-plus matrix $\mathcal{A} \in \mathbb{R}_{\max}^{n\times m}$ is simply an array of elements from $\mathbb{R}_{\max}$. A max-plus diagonal matrix has all nondiagonal entries equal to minus infinity, a particular case being the max-plus identity matrix, denoted by $\mathcal{I}$, which has all its diagonal entries equal to zero. We write max-plus diagonal matrices as $\operatorname{diag}_{\infty}(d)$ for some vector $d \in \mathbb{R}_{\max}^n$ defining the diagonal entries. We denote by $\mathcal{O}_n$ the $n \times n$ matrix of zeros so that $\beta \otimes \mathcal{O}_n$ for some $\beta \in \mathbb{R}_{\max}$ is the $n \times n$ matrix with all entries equal to β.

Max-plus matrix multiplication is defined analogously to classical matrix multiplication so that if $\mathcal{A} \in \mathbb{R}_{\max}^{n\times m}$ and $\mathcal{B} \in \mathbb{R}_{\max}^{m\times \ell}$ then $\mathcal{A} \otimes \mathcal{B} \in \mathbb{R}_{\max}^{n\times \ell}$ with

$$\big(\mathcal{A} \otimes \mathcal{B}\big)_{ij} = \bigoplus_{k=1}^{m} a_{ik} \otimes b_{kj} = \max_{1\le k\le m} a_{ik} + b_{kj}.$$

For clarity we denote powers of $\mathcal{A} \in \mathbb{R}_{\max}^{n\times n}$ by the $\otimes$ symbol so that for example $\mathcal{A}^{\otimes 3} = \mathcal{A} \otimes \mathcal{A} \otimes \mathcal{A}$. The *Kleene star* of $\mathcal{A} \in \mathbb{R}_{\max}^{n\times n}$, denoted by $\mathcal{A}^\star$, is given by

$$\mathcal{A}^\star = \mathcal{I} \oplus \mathcal{A} \oplus \mathcal{A}^{\otimes 2} \oplus \cdots .$$

If $\mathcal{A}$ is irreducible then the entries of $\mathcal{A}^\star$ are all finite.

The *precedence graph* $\Gamma(\mathcal{A})$ of $\mathcal{A} = (a_{ij}) \in \mathbb{R}_{\max}^{n\times n}$ is the weighted directed graph with vertices $\{1, \ldots, n\}$ and an edge from i to j with weight a_{ij} whenever $a_{ij} \neq -\infty$.

The *maximum cycle mean* of $\mathcal{A} \in \mathbb{R}_{\max}^{n\times n}$ is defined by

$$\max_C W(C)/L(C) =: \lambda_{\max}(\mathcal{A}),$$

where the maximum is taken over all elementary cycles C through $\Gamma(\mathcal{A})$. Here $W(C)$ is the *weight of the cycle* C, that is, the sum of the weights of its constituent edges, and $L(C)$ is the *length of the cycle* C, that is, the number of edges C contains. An

elementary cycle C is *critical* in $\Gamma(\mathcal{A})$ if $W(C)/L(C) = \lambda_{\max}(\mathcal{A})$. It is known that the Kleene star $\mathcal{A}^\star$ exists if and only if $\lambda_{\max}(\mathcal{A}) \leq 0$ (see [2, Prop. 1.6.10] for example).

We use the partial ordering $\leq$ on $\mathbb{R}_{\max}^{n\times n}$ defined by $\mathcal{A} \leq \mathcal{B}$ if and only if $a_{ij} \leq b_{ij}$ for all $i, j = 1, \ldots, n$. For $\mathcal{A} \in \mathbb{R}_{\max}^{n\times n}$ and $\beta \in \mathbb{R}_{\max}$, a vector $\chi \in \mathbb{R}_{\max}^{n}$ with at least one finite entry satisfying $\mathcal{A} \otimes \chi \leq \beta \otimes \chi$ is called a *subeigenvector* of $\mathcal{A}$ associated with β. Subeigenvectors with finite entries play an important role in this paper. Their existence is addressed in the next lemma (see [2, Thm. 1.6.18 (a)] for example).

LEMMA 2.1. *Let $\mathcal{A} \in \mathbb{R}_{\max}^{n\times n}$, $\mathcal{A} \neq -\infty \otimes \mathcal{O}_n$ and $\beta \in \mathbb{R}_{\max}$. Then $\mathcal{A} \otimes x \leq \beta \otimes x$ has a finite solution $x \in \mathbb{R}^n$ if and only if $\beta \geq \lambda_{\max}(\mathcal{A})$ and $\beta > -\infty$.*

The *column space* of $\mathcal{A} \in \mathbb{R}_{\max}^{n\times n}$ is defined by

$$\operatorname{col}(\mathcal{A}) = \{\mathcal{A} \otimes \chi : \ \chi \in \mathbb{R}_{\max}^{n}\}.$$

For $\mathcal{A}, \mathcal{B} \in \mathbb{R}_{\max}^{n\times n}$ with $\mathcal{B}$ having finite entries, define $\mathcal{A}/\mathcal{B} \in \mathbb{R}_{\max}^{n\times n}$ by

$$(\mathcal{A}/\mathcal{B})_{ij} = a_{ij} - b_{ij}.$$

To link the classical algebra of complex matrices with standard addition and multiplication to the max-plus algebra, we use the non-Archimedean valuation

$$x \in \mathbb{C} \mapsto \mathcal{V}(x) = \log|x| \in \mathbb{R}_{\max} \tag{2.1}$$

with the convention that $\log 0 = -\infty$. For matrices, we apply the valuation componentwise, that is, for $A \in \mathbb{C}^{n\times n}$, $\mathcal{V}(A) = \mathcal{A} = \big(\log|a_{ij}|\big) \in \mathbb{R}_{\max}^{n\times n}$. For some $u \in \mathbb{R}^n$, we use the notation $\operatorname{diag}_0(u)$ to denote an $n \times n$ real diagonal matrix with ith diagonal entry equal to u_i. Then we have that

$$\mathcal{V}\big(\operatorname{diag}_0(u)\big) = \operatorname{diag}_\infty\big(\log(|u|)\big) \in \mathbb{R}_{\max}^{n\times n},$$

where the absolute value and the logarithm of a vector are taken componentwise.

We say that $\mathcal{B} \in \mathbb{R}_{\max}^{n\times n}$ is obtained from $\mathcal{A} \in \mathbb{R}_{\max}^{n\times n}$ by *diagonal similarity scaling* if $\mathcal{B} = \operatorname{diag}_\infty(-d) \otimes \mathcal{A} \otimes \operatorname{diag}_\infty(d)$ for some $d \in \mathbb{R}^n$.

LEMMA 2.2. *Let $\mathcal{A}$ be a max-plus matrix. Then the weight $W(C)$ of a cycle C in $\Gamma(\mathcal{A})$ is preserved by diagonal similarity scaling of $\mathcal{A}$.*

Proof. The weight of $C = \{(i_1, i_2), \ldots, (i_\ell, i_1)\}$ as a cycle in $\Gamma(\mathcal{A})$ is given by $W_{\Gamma(\mathcal{A})}(C) = a_{i_\ell i_1} + \sum_{k=1}^{\ell-1} a_{i_k i_{k+1}}$. Now let $\mathcal{B} = \operatorname{diag}_\infty(-d) \otimes \mathcal{A} \otimes \operatorname{diag}_\infty(d)$ for some $d \in \mathbb{R}^n$. Then the weight of C as a cycle in $\Gamma(\mathcal{B})$ is given by

$$\begin{aligned} W_{\Gamma(\mathcal{B})}(C) &= b_{i_\ell i_1} + \sum_{k=1}^{\ell-1} b_{i_k i_{k+1}} \\ &= a_{i_\ell i_1} + \sum_{k=1}^{\ell-1} a_{i_k i_{k+1}} - d_{i_\ell} + d_{i_1} + \sum_{k=1}^{\ell-1} -d_{i_k} + d_{i_{k+1}} = W_{\Gamma(\mathcal{A})}(C). \end{aligned}$$

□

Recall that, by Lemma 2.1, a max-plus matrix $\mathcal{A}$ with at least one finite entry has a finite subeigenvector associated with $\lambda_{\max}(\mathcal{A})$.

LEMMA 2.3. *Let $\mathcal{A} \in \mathbb{R}_{\max}^{n\times n}$ have at least one finite entry.*

(a) *A vector $s \in \mathbb{R}^n$ is a subeigenvector of $\mathcal{A}$ associated with the maximum cycle mean $\lambda_{\max}(\mathcal{A})$ if and only if $\operatorname{diag}_\infty(-s) \otimes \mathcal{A} \otimes \operatorname{diag}_\infty(s) \leq \lambda_{\max}(\mathcal{A}) \otimes \mathcal{O}_n$.*

(b) *If $s \in \mathbb{R}^n$ is a subeigenvector of $\mathcal{A}$ associated with $\lambda_{\max}(\mathcal{A})$ then $(\mathcal{S})_{ij} = \lambda_{\max}(\mathcal{A})$ for all i, j such that (i, j) is an edge in a critical cycle C in $\Gamma(\mathcal{A})$.*

(c) *If $s, s' \in \mathbb{R}^n$ are both subeigenvectors of $\mathcal{A}$ associated with $\lambda_{\max}(\mathcal{A})$ then $-s_i + s_j = -s'_i + s'_j$ whenever i and j are both vertices visited by a critical cycle C in $\Gamma(\mathcal{A})$.*

Proof. Let $\mathcal{S} = \operatorname{diag}_\infty(-s) \otimes \mathcal{A} \otimes \operatorname{diag}_\infty(s)$ so that $\mathcal{S}_{ij} = -s_i + a_{ij} + s_j$.

(a) By the definition of a subeigenvector we have $\max_j a_{ij} + s_j \leq \lambda_{\max}(\mathcal{A}) + s_i$ for all i, which is equivalent to $\mathcal{S}_{ij} \leq \lambda_{\max}(\mathcal{A})$ for all $i, j = 1, \ldots, n$.

(b) Lemma 2.2 states that similarity scaling does not affect cycle means, so that

$$\lambda_{\max}(\mathcal{A}) = W(C)/L(C) = \frac{1}{L(C)} \sum_{(i,j)\in C} (\mathcal{S})_{ij},$$

and since each of the terms in the sum is bounded above by $\lambda_{\max}(\mathcal{A})$ they must all be equal to $\lambda_{\max}(\mathcal{A})$.

(c) Let $\mathcal{S}' = \operatorname{diag}_\infty(-s') \otimes \mathcal{A} \otimes \operatorname{diag}_\infty(s')$. Now suppose that i and j are both vertices visited by a critical cycle C. There is a directed path through C from i to j given by $\sigma = (\sigma(1) = i, \ldots, \sigma(t) = j)$. Since the entries of $\mathcal{S}$ and $\mathcal{S}'$ on C are all equal to $\lambda_{\max}(\mathcal{A})$ we have $-s_{\sigma(k)} + a_{\sigma(k)\sigma(k+1)} + s_{\sigma(k+1)} = -s'_{\sigma(k)} + a_{\sigma(k)\sigma(k+1)} + s'_{\sigma(k+1)} = \lambda_{\max}(\mathcal{A})$, so that

$$-s_{\sigma(k)} + s_{\sigma(k+1)} = -s'_{\sigma(k)} + s'_{\sigma(k+1)}, \quad k = 1, \ldots, t. \tag{2.2}$$

Finally summing (2.2) over k yields $-s_i + s_j = -s'_i + s'_j$. ∎

3. Hungarian scaling. The *max-plus permanent* of $\mathcal{A} \in \mathbb{R}_{\max}^{n\times n}$ is given by

$$\operatorname{perm}(\mathcal{A}) = \max_{\pi\in\Pi(n)} \sum_{j=1}^{n} a_{\pi(j)j}, \tag{3.1}$$

where the maximum is taken over the set $\Pi(n)$ of all permutations of $\{1, \ldots, n\}$. We denote by $\pi = \mathrm{id}$ the identity permutation, i.e., $\mathrm{id} = \{1, \ldots, n\}$. A permutation π which attains the maximum in (3.1) is called an *optimal assignment* of $\mathcal{A}$. When $\operatorname{perm}(\mathcal{A}) \neq -\infty$, the max-plus permanent of $\mathcal{A}$ can be rewritten as a minimization problem (see for example [13]),

$$\operatorname{perm}(\mathcal{A}) = \min \left\{ \sum_{i=1}^{n} (u_i + v_i) : \; u, v \in \mathbb{R}^n, \; a_{ij} - u_i - v_j \leq 0 \right\}. \tag{3.2}$$

A *Hungarian pair* of $\mathcal{A}$ is an optimal solution (u, v) to (3.2). It is named after the Hungarian algorithm, which is a widely used primal-dual algorithm for solving the optimal assignment problem.

To any $\pi \in \Pi(n)$ we associate the max-plus permutation matrix $\mathcal{P}_\pi \in \mathbb{R}_{\max}^{n\times n}$ and the permutation matrix $P_\pi \in \mathbb{C}^{n\times n}$ defined by

$$(\mathcal{P}_\pi)_{ij} = \begin{cases} 0 & \text{for } j = \pi(i), \\ -\infty & \text{otherwise,} \end{cases} \qquad (P_\pi)_{ij} = \begin{cases} 1 & \text{for } j = \pi(i), \\ 0 & \text{otherwise,} \end{cases}$$

so that $\mathcal{V}(P_\pi) = \mathcal{P}_\pi$, where $\mathcal{V}(\cdot)$ is the valuation in (2.1).

The next theorem or more precisely, its corollary, appears in [15, Thm. 2.8]. We provide the proof for completeness.

THEOREM 3.1 (Hungarian scaling). *Let π be an optimal assignment of $\mathcal{A} \in \mathbb{R}_{\max}^{n\times n}$, let (u, v) be a Hungarian pair of $\mathcal{A}$, and let $\mathcal{P}_\pi$ be the max-plus permutation matrix defined by π. Then the max-plus Hungarian scaled and reordered matrix*

$$\mathcal{H} = \mathcal{P}_\pi \otimes \operatorname{diag}_\infty(-u) \otimes \mathcal{A} \otimes \operatorname{diag}_\infty(-v)$$

is such that $\mathcal{h}_{ij} \leq 0$ and $\mathcal{h}_{ii} = 0$ for all $i, j = 1, \ldots, n$. Moreover, $\lambda_{\max}(\mathcal{H}) = 0$.

Proof. Note that $\mathcal{H} = \big(\mathcal{P}_\pi \otimes \operatorname{diag}_\infty(-u) \otimes \mathcal{P}_\pi^T\big) \otimes \mathcal{P}_\pi \otimes \mathcal{A} \otimes \operatorname{diag}_\infty(-v)$ so that the entries of $\mathcal{H}$ are given by $\mathcal{h}_{ij} = (-u_{\pi(i)}) \otimes \mathcal{a}_{\pi(i)j} \otimes (-v_j) = \mathcal{a}_{\pi(i)j} - u_{\pi(i)} - v_j$. Now from the definition of a Hungarian pair and (3.2) we have $\mathcal{h}_{ij} \leq 0$ for all i, j. Also,

$$\sum_{i=1}^{n} \mathcal{h}_{ii} = \sum_{i=1}^{n} \mathcal{a}_{\pi(i)i} - u_{\pi(i)} - v_i = \sum_{i=1}^{n} \mathcal{a}_{\pi(i)i} - \sum_{i=1}^{n} u_i + v_i = \operatorname{perm}(\mathcal{A}) - \operatorname{perm}(\mathcal{A}) = 0,$$

which implies that $\mathcal{h}_{ii} = 0$ for all $i = 1, \ldots, n$ since $\mathcal{h}_{ii} \leq 0$ for all $i = 1, \ldots, n$.

Since the diagonal entries of $\mathcal{H}$ correspond to length one cycles of weight zero in $\Gamma(\mathcal{H})$ and no cycle in $\Gamma(\mathcal{H})$ can have strictly positive weight, it follows that $\lambda_{\max}(\mathcal{H}) = 0$. ∎

We refer to a max-plus matrix $\mathcal{H}$ such that $\mathcal{h}_{ij} \leq 0$ and $\mathcal{h}_{ii} = 0$ for all i, j as a *Hungarian matrix*. Since $\lambda_{\max}(\mathcal{H}) = 0$, the Kleene star of a Hungarian matrix always exists. The next result, which holds for complex matrices, is a direct consequence of Theorem 3.1.

COROLLARY 3.2. *Let $A \in \mathbb{C}^{n\times n}$ and let π and (u, v) be an optimal assignment and a Hungarian pair of $\mathcal{V}(A)$, respectively. Let $P_\pi \in \mathbb{C}^{n\times n}$ be the permutation matrix defined by π. Then the Hungarian scaled and reordered matrix*

$$H = P_\pi \operatorname{diag}_0\big(\exp(-u)\big) A \operatorname{diag}_0\big(\exp(-v)\big),$$

has all entries of modulus less than or equal to one and entries of modulus one on the diagonal, i.e., $|h_{ij}| \leq 1$ and $|h_{ii}| = 1$ for all $i, j = 1, \ldots, n$.

We note that the max-plus matrix $\mathcal{H}$ in Theorem 3.1 is the componentwise log-of-absolute-value of the matrix H in Corollary 3.2, that is, $\mathcal{H} = \mathcal{V}(H)$. The *max-plus singular values* of $\mathcal{H}$ are the points of non-differentiability of the max-plus characteristic polynomial $p(\chi) = \operatorname{perm}(\mathcal{H} \oplus \chi \otimes \mathcal{O})$ [12], [13]. Hook [12, Thm. 4.1] shows that the max-plus singular values of a Hungarian scaled matrix are all equal to zero and that, for the valuation (2.1), the exponential of the max-plus singular values of $\mathcal{H} = \mathcal{V}(H)$ tend to offer order of magnitude approximation to the singular values of H. As a result, we can expect the singular values $\sigma_1 \geq \cdots \geq \sigma_n$ of the Hungarian scaled matrix $H \in \mathbb{C}^{n\times n}$ to be close to 1 and therefore for H to have a small 2-norm condition number $\kappa_2(H)$ since $\kappa_2(H) := \|H\|_2\|H^{-1}\|_2 = \sigma_1/\sigma_n$.

EXAMPLE 3.3. *Let*

$$A = \begin{bmatrix} \exp(6) & \exp(2) & \exp(1) \\ 1 & \exp(-3) & \exp(-6) \\ 0 & \exp(-3) & 1 \end{bmatrix}, \qquad \mathcal{A} := \mathcal{V}(A) = \begin{bmatrix} 6 & 2 & 1 \\ 0 & -3 & -6 \\ -\infty & -3 & 0 \end{bmatrix}.$$

It is easy to check that the max-plus matrix $\mathcal{A}$ has a unique optimal assignment $\pi = (1, 2, 3)$ and that (u, v) with $u = [0, -5, -1]^T$ and $v = [6, 2, 1]^T$ is a Hungarian pair for $\mathcal{A}$ yielding the max-plus Hungarian scaled matrix

$$\mathcal{H} = \operatorname{diag}_\infty(-u) \otimes \mathcal{A} \otimes \operatorname{diag}_\infty(-v) = \begin{bmatrix} 0 & 0 & 0 \\ -1 & 0 & -2 \\ -\infty & -4 & 0 \end{bmatrix}.$$

Applying the Hungarian pair (u, v) of $\mathcal{A}$ to A yields the Hungarian scaled matrix

$$H = \operatorname{diag}_0\big(\exp(-u)\big) A \operatorname{diag}_0\big(\exp(-v)\big) = \begin{bmatrix} 1 & 1 & 1 \\ \exp(-1) & 1 & \exp(-2) \\ 0 & \exp(-4) & 1 \end{bmatrix}.$$

Note that $\mathcal{V}(H) = \mathcal{H}$. Hungarian scaling tends to significantly reduce the condition number of a matrix. For this example we have $\kappa_2(A) = 1.3 \times 10^4 \gg \kappa_2(H) = 6.2$.

Next we state a result from Butkovič and Schneider [3, Problem 3.1], which, given a max-plus matrix and some desired upper bounds on the entries of the matrix, enables us to compute diagonal similarity scalings so that the scaled matrix satisfies the desired bounds, or otherwise to prove that no such scaling exists. We note that although Butkovič and Schneider's results in [3] are stated for nonnegative matrices in the max-times algebra rather than max-plus matrices in the max-plus algebra, the transformation from one to the other is very straightforward. The solution to [3, Problem 3.1] we state below is for the max-plus algebra.

THEOREM 3.4 (One-sided inequality). *For $\mathcal{A}, \mathcal{B} \in \mathbb{R}_{\max}^{n\times n}$, $\mathcal{B}$ with finite entries,*

$$\big\{s \in \mathbb{R}^n : \operatorname{diag}_\infty(-s) \otimes \mathcal{A} \otimes \operatorname{diag}_\infty(s) \le \mathcal{B}\big\} = \begin{cases} \operatorname{col}\big((\mathcal{A}/\mathcal{B})^\star\big) \text{ if } \lambda_{\max}(\mathcal{A}/\mathcal{B}) \le 0, \\ \emptyset \text{ otherwise.} \end{cases}$$

Note that if $\mathcal{A}$ has optimal assignment $\pi \neq \mathrm{id}$ then $\mathcal{P}_\pi \otimes \mathcal{A}$ has the identity permutation as optimal assignment so there is no loss of generality in assuming that $\mathcal{A}$ has $\pi = \mathrm{id}$ as optimal assignment. We now provide a characterization for the set of all Hungarian pairs, which is new to the best of our knowledge.

THEOREM 3.5 (Set of all Hungarian pairs). *Assume that $\mathcal{A} \in \mathbb{R}_{\max}^{n\times n}$ has optimal assignment $\pi = \mathrm{id}$ and let (u, v) be a Hungarian pair for $\mathcal{A}$. Then the set of all Hungarian pairs* $\operatorname{Hung}(\mathcal{A})$ *of $\mathcal{A}$ is given by*

$$\operatorname{Hung}(\mathcal{A}) = \{(u + s, v - s) : s \in \operatorname{col}(\mathcal{H}^\star)\},$$

where $\mathcal{H} = \operatorname{diag}_\infty(-u) \otimes \mathcal{A} \otimes \operatorname{diag}_\infty(-v)$.

Proof. Let $(u', v') \in \operatorname{Hung}(\mathcal{A})$ and $\mathcal{H}' = \operatorname{diag}_\infty(-u') \otimes \mathcal{A} \otimes \operatorname{diag}_\infty(-v')$. From Theorem 3.1 we have

$$h_{ii} = h'_{ii} = 0 \iff a_{ii} - u'_i - v'_i = a_{ii} - u_i - v_i = 0$$

so that there exists $s \in \mathbb{R}^n$ such that $(u', v') = (u + s, v - s)$. Therefore $\mathcal{H}' = \operatorname{diag}_\infty(-s) \otimes \mathcal{H} \otimes \operatorname{diag}_\infty(s)$. Now from Theorem 3.1,

$$\mathcal{H}' = \operatorname{diag}_\infty(-s) \otimes \mathcal{H} \otimes \operatorname{diag}_\infty(s) \le \mathcal{O}_n. \tag{3.3}$$

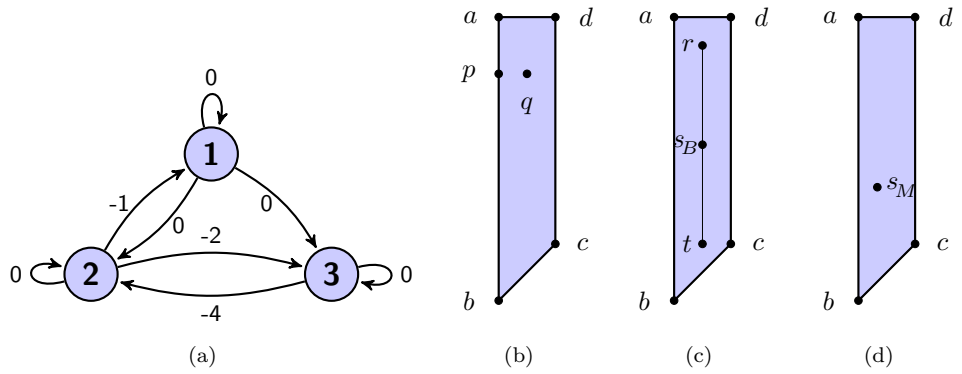
From Theorem 3.4 we know that inequality (3.3) is feasible provided $\lambda_{\max}(\mathcal{H}/\mathcal{O}_n) \le 0$, which is true since $\mathcal{H}/\mathcal{O}_n = \mathcal{H}$ and $\lambda_{\max}(\mathcal{H}) = 0$. Therefore, (3.3) is satisfied if and only if $s \in \operatorname{col}(\mathcal{H}^\star)$.

Conversely suppose that $s \in \operatorname{col}(\mathcal{H}^\star)$, then s is a solution to

$$\operatorname{diag}_\infty(-s) \otimes \mathcal{H} \otimes \operatorname{diag}_\infty(s) \le \mathcal{O}_n \iff a_{ij} - (u_i + s_i) - (v_j - s_j) \le 0$$

so that $(u', v') := (u + s, v - s)$ is a feasible solution to (3.2). Then since

$$\sum_{i=1}^n u'_i + v'_i = \sum_{i=1}^n (u_i + s_i) + (v_i - s_i) = \sum_{i=1}^n u_i + v_i = \operatorname{perm}(\mathcal{A}),$$



(a) (b) (c) (d)

FIG. 3.1. *For the matrix of Example* 3.6. (a) *precedence graph* $\Gamma(\mathcal{H})$. *Subfigures* (b), (c), *and* (d) *show* $\operatorname{col}(\mathcal{H}^\star) \cap \{s_1 = 0\}$ *and different scaling vectors.*

the pair (u', v') must also be an optimal solution to (3.2) and therefore a Hungarian pair of $\mathcal{A}$. □

EXAMPLE 3.6. *Consider the matrix* $\mathcal{A}$ *and the Hungarian scaled matrix* $\mathcal{H}$ *of Example* 3.3. *From Theorem* 3.5 *we know that the set of all Hungarian pairs of* $\mathcal{A}$ *is given by* $\operatorname{Hung}(\mathcal{A}) = \{\big((0,-5,-1)+s, (6,2,1)-s\big) : s \in \operatorname{col}(\mathcal{H}^\star)\}$. *The* (i,j) *entry of* $\mathcal{H}^\star$ *is equal to the weight of the maximally weighted path through the precedence graph* $\Gamma(\mathcal{H})$ *of* $\mathcal{H}$ *from* i *to* j *so, by looking at* $\Gamma(\mathcal{H})$ *in Figure* 3.1(a), *we find that*

$$\mathcal{H}^\star = \begin{bmatrix} 0 & 0 & 0 \\ -1 & 0 & -1 \\ -5 & -4 & 0 \end{bmatrix}.$$

Computing $\operatorname{col}(\mathcal{H}^\star)$ *generally requires some special knowledge of tropical geometry. However in this case, we know from Theorem* 3.4 *that*

$$s \in \operatorname{col}(\mathcal{H}^\star) \iff \mathcal{H}_s := \operatorname{diag}_\infty(-s) \otimes \mathcal{H} \otimes \operatorname{diag}_\infty(s) \leq \mathcal{O}_3,$$

which yields the following constraints on the entries of s:

$$-1 + s_1 - s_2 \leq 0, \quad s_2 - s_1 \leq 0, \quad -4 + s_2 - s_3 \leq 0, \quad s_3 - s_1 \leq 0. \tag{3.4}$$

These constraints define the boundary faces of $\operatorname{col}(\mathcal{H}^\star)$. *For all* $\alpha \in \mathbb{R}$, *a scaling parameter* $s \in \mathbb{R}^n$ *satisfies* (3.4) *if and only if* $\tilde{s} := s + [\alpha, \alpha, \alpha]^T = s \otimes \alpha$ *satisfies* (3.4). *Therefore* $\operatorname{col}(\mathcal{H}^\star)$ *is a prism extruded in the* $[1,1,1]^T$ *direction. Its intersection with the plane* $s_1 = 0$ *is the set of solutions to*

$$s_2 \geq -1, \quad s_2 \leq 0, \quad s_2 - s_3 \leq 4, \quad s_3 \leq 0,$$

which is given by the quadrilateral shown in Figure 3.1(b). *The three vertices* a, b, c *of the quadrilateral are columns of* $\mathcal{H}^\star$ *and the fourth vertex* d *is a max-plus linear combination of two columns of* $\mathcal{H}^\star$,

$$a = [0,-1,0]^T, \quad b = [0,-1,-5]^T, \quad c = [0,0,-4]^T, \quad d = a \oplus c = [0,0,0]^T.$$

Since for any $\alpha \in \mathbb{R}$ *the scaling vectors* s *and* $\tilde{s} = s \otimes \alpha$ *both give rise to the same scaling of* $\mathcal{H}$ *it follows that the set of all Hungarian scalings of* $\mathcal{A}$ *is isomorphic*

TABLE 3.1
2-norm condition number and Frobenius norm for the matrices of Examples 3.6, 4.4 and 5.3.

Matrix X	A	H (H_d)	H_a	H_b	H_c	H_p	H_q	H_{s_B}	H_{s_M}
$\kappa_2(X)$	1.27e4	6.19	6.56	6.98	6.40	4.96	4.27	**4.08**	4.18
$\|X\|_F$	4.04e2	2.27	2.30	2.27	2.27	2.07	1.97	**1.94**	1.95

to the intersection of $\operatorname{col}(\mathcal{H}^\star)$ *with the plane* $s_1 = 0$. *Each vertex of the quadrilateral therefore corresponds to an extremal Hungarian scaling of* $\mathcal{A}$ *given by*

$$\mathcal{H}_a = \begin{bmatrix} 0 & -1 & 0 \\ 0 & 0 & -1 \\ -\infty & -5 & 0 \end{bmatrix}, \qquad \mathcal{H}_b = \begin{bmatrix} 0 & -1 & -5 \\ 0 & 0 & -6 \\ -\infty & 0 & 0 \end{bmatrix},$$
$$\mathcal{H}_c = \begin{bmatrix} 0 & 0 & -4 \\ -1 & 0 & -6 \\ -\infty & 0 & 0 \end{bmatrix}, \qquad \mathcal{H}_d = \begin{bmatrix} 0 & 0 & 0 \\ -1 & 0 & -2 \\ -\infty & -4 & 0 \end{bmatrix}.$$

Note that each of these Hungarian scaled matrices contain precisely five entries equal to zero. If we scale using any parameter from the relative interior of an edge of the quadrilateral then we obtain a scaled matrix with exactly four entries equal to zero. If we take any scaling parameter from the interior of the quadrilateral then we obtain a scaled matrix with exactly three entries equal to zero. For example, $p = [0, -1, -1]^T$ *and* $q = [0, -0.5, -1]^T$ *yield*

$$\mathcal{H}_p = \begin{bmatrix} 0 & -1 & -1 \\ 0 & 0 & -2 \\ -\infty & -4 & 0 \end{bmatrix}, \qquad \mathcal{H}_q = \begin{bmatrix} 0 & -0.5 & -1 \\ -0.5 & 0 & -2.5 \\ -\infty & -3.5 & 0 \end{bmatrix}.$$

Let A *and* H *be as in Example* 3.3 *so that* $\mathcal{V}(A) = \mathcal{A}$ *and* $\mathcal{V}(H) = \mathcal{H}$. *For* $s = a, b, c, d, p, q$ *define* $H_s := \operatorname{diag}_0\big(\exp(-s)\big) H \operatorname{diag}_0\big(\exp(-s)\big) \in \mathbb{R}^{3\times 3}$. *The 2-norm and 2-norm condition number of these matrices are provided in Table* 3.1. *As in Example* 3.3 *we find that Hungarian scaling significantly reduces the matrix condition number. The scalings* a, b, c, d *which are taken from extreme points of* $\operatorname{col}(\mathcal{H}^\star)$ *(see Figure* 3.1 *(b)) all result in scaled matrices with five entries of modulus one and all with very similar condition numbers and norms. The scaling* p *which is taken from an edge of* $\operatorname{col}(\mathcal{H}^\star)$ *results in a scaled matrix with four entries of modulus one and a slightly smaller condition number and norm compared to the previous Hungarian scalings. The scaling* q *which is taken from the interior of* $\operatorname{col}(\mathcal{H}^\star)$ *results in a scaled matrix with three entries of modulus one and has a further reduced condition number and norm. Note that only* H_q *is diagonally dominant.*

In order to Hungarian scale a matrix $\mathcal{A} \in \mathbb{R}^{n\times n}_{\max}$ we must compute an optimal assignment and Hungarian pair for $\mathcal{A}$. The best known algorithms for this have worst case cost $O\big(n\tau + n^2 \log n\big)$, where τ is the number of finite entries in $\mathcal{A}$ (finite entries are the max-plus equivalent of nonzero entries). However, in practical numerical examples it is found that optimal assignment algorithms such as the Hungarian algorithm [8], the successive shortest paths algorithm [16] and the auction algorithm [11] have run times roughly linear in the number of finite entries in the matrix. It is only for some very special examples that the worst case complexity bound is attained.

In general the space $\operatorname{col}(\mathcal{H}^\star)$ contains more than one possible scaling so that different optimal assignment algorithms may return different Hungarian pairs, which result in different scalings that may have different properties. Theorem 3.5 tells us

that these different scalings are all related by similarity scalings. Moreover if we suppose that $\mathcal{A}$ has been Hungarian scaled and reordered into a Hungarian matrix $\mathcal{H}$ then Theorem 3.5 tells us that for $s \in \mathrm{col}(\mathcal{H}^\star)$, $\mathcal{H}_s = \mathrm{diag}_\infty(-s) \otimes \mathcal{H} \otimes \mathrm{diag}_\infty(s)$ is also a Hungarian matrix. In the remainder of this paper we investigate two approaches to construct a vector $s \in \mathrm{col}(\mathcal{H}^\star)$ that aims to make $\mathcal{H}_s$ more diagonally dominant than $\mathcal{H}$. The computed scalings can then be exponentiated and applied to the original problem matrix A as in Corollary 3.2.

4. Max-balancing. An irreducible matrix $\mathcal{A} \in \mathbb{R}_{\max}^{n\times n}$ is max-balanced if for any nontrivial subset $\mathcal{J} \subset \{1, \ldots, n\}$ we have

$$\max_{i\in\mathcal{J}, j\notin\mathcal{J}} a_{ij} = \max_{i\notin\mathcal{J}, j\in\mathcal{J}} a_{ij}. \tag{4.1}$$

We note that for a matrix to be max-balanced is a stronger condition than to be balanced in the max-norm sense as illustrated by the following example taken from [17]. Indeed, the max-norm of the ith column of the matrix

$$\mathcal{A} = \begin{bmatrix} -\infty & 2 & -\infty & -\infty \\ 2 & -\infty & 1 & -\infty \\ -\infty & 0 & -\infty & 2 \\ -\infty & -\infty & 2 & -\infty \end{bmatrix}$$

is equal to the max-norm of the ith row for all i. However the set $\mathcal{J} = \{1, 2\}$ does not satisfy the max-balancing condition (4.1).

Schneider and Schneider [17, Cor. 9] show that any irreducible nonnegative matrix can be max-balanced by a unique similarity scaling. It is trivial to rephrase their result for max-plus matrices.

THEOREM 4.1 (Uniqueness of max-balancing scaling). *For any irreducible $\mathcal{A} \in \mathbb{R}_{\max}^{n\times n}$ there exists a unique $s \in \mathbb{R}^n$ with $\sum_i s_i = 0$ such that $\mathrm{diag}_\infty(-s) \otimes \mathcal{A} \otimes \mathrm{diag}_\infty(s)$ is max-balanced.*

Note that the requirement that $\sum_i s_i = 0$ in Theorem 4.1 is a normalization condition, which is necessary for uniqueness. Indeed if s max-balances $\mathcal{A}$ so does $\alpha \otimes s$ for any $\alpha \in \mathbb{R}$. Now if $\mathcal{H}$ is a Hungarian scaled and reordered matrix obtained from an irreducible max-plus matrix $\mathcal{A}$ then the max-balancing scaling of $\mathcal{H}$ is the unique max-balancing Hungarian scaling of $\mathcal{A}$.

4.1. Max-balancing algorithm. Schneider and Schneider's description of the max-balancing algorithm is purely in terms of the precedence graph of the matrix [17]. Our description of the algorithm is in terms of matrices and this turns out to be useful for the proofs of the theorems in this section and Section 4.2.

ALGORITHM 4.2 (Max-balancing). *Given an irreducible matrix $\mathcal{A} \in \mathbb{R}_{\max}^{n\times n}$ this algorithm returns $s_B \in \mathbb{R}^n$ such that $\mathrm{diag}_\infty(-s_B) \otimes \mathcal{A} \otimes \mathrm{diag}_\infty(s_B)$ is max-balanced.*

1 Set $t = 1$, $m_0 = n$, $f_1 = \mathrm{id}$.
2 Let $\mathcal{A}_1 \in \mathbb{R}_{\max}^{n\times n}$ be such that $(\mathcal{A}_1)_{ij} = a_{ij}$ if $i \neq j$ and $(\mathcal{A}_1)_{ii} = -\infty$.
3 Compute $\beta_1 := \lambda_{\max}(\mathcal{A}_1)$ with critical cycle C_1.
4 Compute a subeigenvector $s_1 \in \mathbb{R}^n$ of $\mathcal{A}_1$ associated with β_1.
5 Let $m_1 := m_0 + 1 -$ number of vertices in C_1.
6 while $m_t > 1$
7 $\quad t = t + 1$
8 $\quad \mathcal{S}_t = \mathrm{diag}_\infty(-s_{t-1}) \otimes \mathcal{A}_{t-1} \otimes \mathrm{diag}_\infty(s_{t-1})$
9 $\quad$ Let $f_t : \{1, \ldots, m_{t-2}\} \mapsto \{1, \ldots, m_{t-1}\}$ be such that $f_t(i) = f_t(j)$ if

and only if i and j are both vertices of C_{t-1}. Let $\mathcal{A}_t \in \mathbb{R}_{\max}^{m_{t-1}\times m_{t-1}}$ be such that $(\mathcal{A}_t)_{\ell p} = \begin{cases} -\infty & \text{if } \ell = p, \\ \max\{(\mathcal{S}_t)_{ij}:\ f_t(i)=\ell, f_t(j)=p\} & \text{otherwise.} \end{cases}$

10 Compute $\beta_t := \lambda_{\max}(\mathcal{A}_t)$ with critical cycle C_t.

11 Compute a subeigenvector $s_t \in \mathbb{R}^{m_t}$ of $\mathcal{A}_t$ associated with β_t.

12 $m_t = m_{t-1} + 1 -$ number of nodes in C_t

13 end

14 $s_B = s_1(f_1) + s_2(f_2 \circ f_1) + \cdots + s_t(f_t \circ \cdots \circ f_1)$.

Note that since diagonal similarities do not affect diagonal entries, there is no harm in setting the diagonal entries of $\mathcal{A}$ to $-\infty$ in line 2 of Algorithm 4.2. On line 14, $s_\ell(g_\ell)$ with $g_\ell = f_k \circ \cdots \circ f_1$ is a vector of length n such that $\big(s_\ell(g_\ell)\big)_i = (s_\ell)_{g_\ell(i)}$, $\ell = 1, \ldots, k$. We show in Lemma A.1(a) that the maximum cycle means β_k are finite so that, by Lemma 2.1, the subeigenvectors s_t exist.

We say that the matrix $\mathcal{A}_t$ on line 9 is a *contraction* of $\mathcal{S}_t$ with respect to the projection f_t, which we denote by $\mathcal{A}_t = \operatorname{contr}(\mathcal{S}_t, f_t)$. Since the diagonal entries of the matrices $\mathcal{A}_t$ are equal to $-\infty$, the number of nodes in the critical cycles C_t is always strictly larger than 1 so the size of the matrix $\mathcal{A}_t$ decreases at each step. It is then easy to see that the algorithm terminates after at most n steps. Schneider and Schneider [17, Thm. 6] show that the vector s_B returned by Algorithm 4.2 defines the diagonal similarity scaling which max-balances $\mathcal{A}$. Now if we apply Algorithm 4.2 to a Hungarian scaled matrix $\mathcal{H}$ then the next result shows that s_B belongs to $\operatorname{col}(\mathcal{H}^\star)$. So max-balancing preserves the Hungarian scaled property.

THEOREM 4.3. *Assume that $\mathcal{A} \in \mathbb{R}_{\max}^{n\times n}$ has optimal assignment $\pi = \mathrm{id}$ and let (u, v) be a Hungarian pair for $\mathcal{A}$. Let s_B be the scaling vector returned by Algorithm 4.2 when applied to the max-plus Hungarian scaled matrix $\mathcal{H} = \operatorname{diag}_\infty(-u) \otimes \mathcal{A} \otimes \operatorname{diag}_\infty(-v)$. Then $s_B \in \operatorname{col}(\mathcal{H}^\star)$ and with the normalization $\sum_i (s_B)_i = 0$, the pair $(u + s_B, v - s_B)$ is the unique Hungarian pair of $\mathcal{A}$ that max-balances $\mathcal{A}$.*

Proof. See Appendix A.2. □

The max-balancing scaling of $A \in \mathbb{C}^{n\times n}$ is given by

$$A_{s_B} = \operatorname{diag}_0\big(\exp(-s_B)\big) A \operatorname{diag}_0\big(\exp(s_B)\big),$$

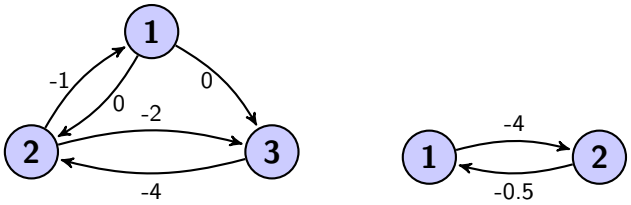
where $s_B \in \mathbb{R}^n$ is such that the max-plus matrix $\operatorname{diag}_\infty(-s_B) \otimes \mathcal{V}(A) \otimes \operatorname{diag}_\infty(s_B)$ is max-balanced, $\mathcal{V}(\cdot)$ being the valuation in (2.1).

Young, Tarjan and Orlin show that the max-balancing algorithm can be implemented with $O\big(n\tau + n^2 \log n\big)$ operations, where τ is the number of finite entries in $\mathcal{A}$ [19].

EXAMPLE 4.4. *Let us use Algorithm 4.2 to max-balance $\mathcal{A} = \mathcal{H}_d$, where $\mathcal{H}_d$ is one of the max-plus Hungarian-scaled matrices of Example 3.6.*
$t = 1$. We start by setting the diagonal entries of $\mathcal{A}$ to $-\infty$ to give

$$\mathcal{A}_1 = \begin{bmatrix} -\infty & 0 & 0 \\ -1 & -\infty & -2 \\ -\infty & -4 & -\infty \end{bmatrix}. \tag{4.2}$$

The precedence graph $\Gamma(\mathcal{A}_1)$ is shown in Figure 4.1(a). The maximum cycle mean β_1, a critical cycle C_1 and a subeigenvector s_1 for $\mathcal{A}_1$ associated with β_1 are given by $\beta_1 = -0.5$, $C_1 = \{(1,2),(2,1)\}$, $s_1 = [0, -0.5, -4]^T$ so that $m_1 = 2$.



(a) (b)

FIG. 4.1. (a) *is the precedence graph of* $\mathcal{A}_1$ *in* (4.2) *and* (b) *is that of* $\mathcal{A}_2$ *in* (4.3).

$t = 2$. *We compute*

$$\mathcal{S}_2 = \operatorname{diag}_\infty(-s_1) \otimes \mathcal{H}_1 \otimes \operatorname{diag}_\infty(s_1) = \begin{bmatrix} -\infty & -0.5 & -4 \\ -0.5 & -\infty & -5.5 \\ -\infty & -0.5 & -\infty \end{bmatrix}.$$

Next we set $f_2(1) = f_2(2) = 1$, $f_2(3) = 2$ *so that*

$$\mathcal{A}_2 = \begin{bmatrix} -\infty & \max\{-4, -5.5\} \\ \max\{-\infty, -0.5\} & -\infty \end{bmatrix} = \begin{bmatrix} -\infty & -4 \\ -0.5 & -\infty \end{bmatrix}. \tag{4.3}$$

The precedence graph $\Gamma(\mathcal{A}_2)$ *is shown in Figure* 4.1(*b*). *The maximum cycle mean, critical cycle and subeigenvector for* $\mathcal{H}_2$ *are given by* $\beta_2 = -2.25$, $C_2 = \{(1,2),(2,1)\}$, $s_2 = [0, 1.75]^T$ *so that* $m_2 = 2 - 2 + 1 = 1$ *and the algorithm terminates. The max-balancing scaling parameter* s_B *is then given by* $s_B = s_1 + s_2(f_2) = [0, -0.5, -4]^T + [0, 0, 1.75]^T = [0, -0.5, -2.25]^T$, *which results in the max-balanced Hungarian scaled max-plus matrix*

$$\mathcal{H}_{s_B} = \operatorname{diag}_\infty(-s_B) \otimes \mathcal{A} \otimes \operatorname{diag}_\infty(s_B) = \begin{bmatrix} 0 & -0.5 & -2.25 \\ -0.5 & 0 & -3.75 \\ -\infty & -2.25 & 0 \end{bmatrix}.$$

For the matrices $A, H \in \mathbb{C}^{n\times n}$ *of Example* 3.3, *max-balancing leads to the max-balanced Hungarian scaled matrix*

$$H_{s_B} = \operatorname{diag}_0\big(\exp(-s_B)\big) H \operatorname{diag}_0\big(\exp(s_B)\big) = \begin{bmatrix} 1 & \exp(-\frac{1}{2}) & \exp(-\frac{9}{4}) \\ \exp(-\frac{1}{2}) & 1 & \exp(-\frac{15}{4}) \\ 0 & \exp(-\frac{9}{4}) & 1 \end{bmatrix}.$$

Table 3.1 *shows that* H_{s_B} *has the smallest norm and condition number amongst all of the Hungarian scaled matrices obtained so far from* A. *Note that* H_{s_B} *is diagonally dominant by row and by column.*

4.2. Norm minimization of max-balancing scaling. The property that max-balancing scaling minimizes the larger entries in the matrix is formalized as follows. For $A \in \mathbb{C}^{n\times n}$ let

$$\|A\|_p := \|\operatorname{vec}(A)\|_p = \Big(\sum_{ij} |a_{ij}|^p\Big)^{1/p}, \quad 1 \le p \le \infty,$$

denote the entrywise p-norm of the matrix. Note that $p = 2$ corresponds to the Frobenius norm and if H is a Hungarian matrix then $1 \le \|H\|_p \le \tau$, where τ is the number of nonzeros in H.

Define the ordering $\prec$ by $A \prec B$ if there exists $p' \in \mathbb{R}_+$ such that $\|A\|_p < \|B\|_p$ for all $p > p'$. We are ready to state the main result of this paper.

THEOREM 4.5 (Norm minimization of max-balancing scaling). *Let A_s be the max-balancing scaling of an irreducible $A \in \mathbb{C}^{n\times n}$. Then A_s is the unique least element with respect to $\prec$ in the set of all diagonal similarity scalings of A.*

Proof. See Appendix A.3. ∎

The next result is a direct consequence of Theorem 3.5 and Theorem 4.5.

THEOREM 4.6 (Norm minimization of max-balancing Hungarian scaling). *Let H_{s_B} be the max-balancing Hungarian scaling of an irreducible $A \in \mathbb{C}^{n\times n}$. Then H_{s_B} is the unique least element, with respect to $\prec$, in the set of all Hungarian scalings of A.*

Theorem 4.6 says that max-balancing minimizes the entrywise p-norm over all diagonal similarity scalings of A that preserve the Hungarian scaling property in the limit as p tends to infinity. As a result, the max-balancing of a Hungarian scaled matrix produces, in the entrywise p-norm sense, the most "diagonally dominant" Hungarian scaled matrix.

4.3. Reducible case. If $\mathcal{A} \in \mathbb{R}_{\max}^{n\times n}$ is not irreducible then it is not diagonally similar to a max-balanced matrix. For example

$$\mathcal{A} = \begin{bmatrix} 0 & 0 \\ -\infty & 0 \end{bmatrix}, \quad \mathcal{B} = \operatorname{diag}_\infty(-s) \otimes \mathcal{A} \otimes \operatorname{diag}_\infty(s) = \begin{bmatrix} 0 & s_2 - s_1 \\ -\infty & 0 \end{bmatrix},$$

is not max-balanced unless $s_2 - s_1 = -\infty$.

We treat a reducible matrix $\mathcal{A} \in \mathbb{R}_{\max}^{n\times n}$ as follows. First we determine the strongly connected components of $\Gamma(\mathcal{A})$. Suppose that there are m of these of size $n_1, \ldots, n_m$. We then apply a similarity permutation to obtain a block upper triangular matrix $\mathcal{A}' = \mathcal{P}^{-1} \otimes \mathcal{A} \otimes \mathcal{P}$ with irreducible diagonal blocks $\mathcal{A}'_{ii} \in \mathbb{R}_{\max}^{n_i\times n_i}$, $i = 1, \ldots, m$. Next we compute the max-balancing scaling vector s_i for each of the diagonal blocks $\mathcal{A}'_{ii}$ and we let $\mathcal{B}' = \operatorname{diag}_\infty(-s) \otimes \mathcal{A}' \otimes \operatorname{diag}_\infty(s)$, where $s = [s_1^T, \ldots, s_m^T]^T \in \mathbb{R}^n$. The matrix $\mathcal{B}'$ is block upper triangular with irreducible max-balanced diagonal blocks. However it may have some large entries in its off diagonal blocks and these can be scaled down as follows. For $\epsilon \in \mathbb{R}_{\max}$, let $\mathcal{C} \in \mathbb{R}_{\max}^{m\times m}$ be such that

$$c_{ij} = \begin{cases} \epsilon + \max\limits_{kl} (\mathcal{B}'_{ij})_{kl} & \text{if } j > i, \\ -\infty & \text{otherwise.} \end{cases}$$

Let $r = \mathcal{C}^\star \otimes \mathcal{O}_{m\times 1}$ and $t = [r_1 \mathcal{O}_{1\times n_1}, \ldots, r_m \mathcal{O}_{1\times n_m}]^T \in \mathbb{R}^n$. We apply t as a scaling to obtain

$$\mathcal{B} = \operatorname{diag}_\infty(-t) \otimes \mathcal{B}' \otimes \operatorname{diag}_\infty(t) = \operatorname{diag}_\infty(-t - s) \otimes \mathcal{P}^{-1} \otimes \mathcal{A} \otimes \mathcal{P} \otimes \operatorname{diag}_\infty(s + t).$$

The matrix $\mathcal{B}$ is block upper triangular with irreducible max-balanced diagonal blocks and all entries in off-diagonal blocks less than or equal to ϵ. The smaller ϵ is, the more diagonally dominant is $\mathcal{B}$. However setting ϵ to be extremely small has little extra benefit and may lead to large errors when scaling back solutions of linear systems after Gaussian elimination. We choose ϵ to be the smallest of all of the maximum cycle means computed by the max-balancing algorithm for all the diagonal blocks $\mathcal{A}'_{ii}$, $i = 1 \ldots, m$.

5. Centre of mass scaling. A *visualization scaling* as introduced by Sergeev, Schneider and Butkovič [18] is a diagonal similarity scaling which results in a matrix whose entries are all less than or equal to the maximum cycle mean of the matrix, with a strict inequality for any entries that do not lie on critical cycles. Such scaled matrices are called *strictly visualized*. The motivation behind visualization scaling is to better understand connections between classical and tropical algebra and also to provide normal forms for certain tropical linear maps.

For a nonnegative matrix A, Sergeev, Schneider and Butkovič characterize the set of all diagonal similarities which strictly visualize A (see [18, Thm. 3.7]). It is straightforward to rephrase their result in terms of max-plus matrices.

THEOREM 5.1 (Visualization scalings). *Let $\mathcal{A} \in \mathbb{R}_{\max}^{n\times n}$. Then $\mathcal{A}_s = \operatorname{diag}_\infty(-s) \otimes \mathcal{A} \otimes \operatorname{diag}_\infty(s)$ is strictly visualized if and only if $s \in \operatorname{relint}\big(\operatorname{col}(\mathcal{A}^\star)\big)$, where $\operatorname{relint}(\mathcal{X})$ is the relative interior of $\mathcal{X}$.*

Now if $\mathcal{H} \in \mathbb{R}_{\max}^{n\times n}$ is a Hungarian scaled matrix which has been strictly visualized into $\mathcal{H}_s := \operatorname{diag}_\infty(-s) \otimes \mathcal{H} \otimes \operatorname{diag}_\infty(s)$ then the entries of $\mathcal{H}_s$ are such that $(\mathcal{H}_s)_{ij} = 0$ for all i, j such that $j = \pi(i)$ for some optimal assignment $\pi \in \Pi_n$ and $(\mathcal{H}_s)_{ij} < 0$ otherwise. So $\mathcal{H}_s$ is a Hungarian matrix which has as few zero entries as is possible. In particular, if $\mathcal{H} \in \mathbb{R}_{\max}^{n\times n}$ has a unique optimal assignment then a strict visualization scaling of $\mathcal{H}$ will have exactly n entries equal to zero.

As we saw in Example 3.6 each face of $\operatorname{col}(\mathcal{H}^\star)$ corresponds to a different linear constraint on the scaling parameter s for $\mathcal{H}_s$ to be a Hungarian scaling. Each of these constraints is of the form $(\mathcal{H}_s)_{ij} \le 0$ for some entry (i, j). Thus if the scaling parameter s is taken from the face of $\operatorname{col}(\mathcal{H}^\star)$ associated with the (i, j)th entry constraint then the scaled matrix will have $(\mathcal{H}_s)_{ij} = 0$. If the scaling parameter is a point outside of $\operatorname{col}(\mathcal{H}^\star)$ on the other side of this face then the scaled matrix will have $(\mathcal{H}_s)_{ij} > 0$ and if the scaling parameter is taken from $\operatorname{col}(\mathcal{H}^\star)$ but not from this face then $(\mathcal{H}_s)_{ij} < 0$. The further away from the face, towards the interior of $\operatorname{col}(\mathcal{H}^\star)$, the smaller the value of the (i, j)th entry in the scaled matrix. The max-balancing scaling attempts to minimize larger entries in the matrix and this is reflected in the observation that the max-balancing scaling parameter lies roughly in the centre of $\operatorname{col}(\mathcal{H}^\star)$. Points in the centre of $\operatorname{col}(\mathcal{H}^\star)$ maximize the minimum distance from the various faces and thus minimize larger entries in the scaled matrix.

Rather than using max-balancing scaling, we can instead try any scaling that takes its scaling parameter from a point close to the centre of $\operatorname{col}(\mathcal{H}^\star)$. One simple way of obtaining such a scaling is to take the centre of mass of the vertices of $\operatorname{col}(\mathcal{H}^\star)$. The *centre of mass scaling* of $\mathcal{H}$ is given by $\mathcal{C} = \operatorname{diag}_\infty(-s_M) \otimes \mathcal{H} \otimes \operatorname{diag}_\infty(s_M)$, where

$$s_M = \frac{1}{n} \sum_{i=1}^{n} \mathcal{H}^\star \otimes e_i \in \mathbb{R}^n.$$

Here $e_i \in \mathbb{R}_{\max}^n$ is the ith column of the max-plus identity matrix. If τ is the number of finite entries of $\mathcal{H}$ then the total cost of computing this scaling is $O\big(n\tau + n^2 \log n\big)$, which is the same as the max-balancing scaling. However computing the centre of mass scaling is embarrassingly parallel since each of the $\mathcal{H}^\star \otimes e_i$ can be computed independently using Dijkstra's algorithm with an individual cost of $O\big(\tau + n \log n\big)$.

THEOREM 5.2 (Uniqueness of centre of mass scaling). *Let $\mathcal{H}$ and $\mathcal{H}'$ be two different Hungarian scalings of a max-plus matrix $\mathcal{A}$. Then the centre of mass scalings of $\mathcal{H}$ and $\mathcal{H}'$ are the same.*

Proof. Let $\mathcal{C} = \operatorname{diag}_\infty(-s_M) \otimes \mathcal{H} \otimes \operatorname{diag}_\infty(s_M) \in \mathbb{R}_{\max}^{n\times n}$ be the centre of mass scaling of $\mathcal{H}$. Since $\mathcal{H}'$ is also a Hungarian scaling of $\mathcal{A}$, it must, by Theorem 3.5, be

a similarity scaling of $\mathcal{H}$, i.e., there exists $s \in \mathbb{R}^n$ such that $\mathcal{H}' = \mathrm{diag}_\infty(-s) \otimes \mathcal{H} \otimes \mathrm{diag}_\infty(s)$. The centre of mass scaling parameter for $\mathcal{H}'$ is then given by

$$s_{M'} = \frac{1}{n}\sum_{i=1}^{n}(\mathcal{H}')^\star \otimes e_i = \frac{1}{n}\sum_{i=1}^{n}\mathrm{diag}_\infty(-s)\otimes\mathcal{H}^\star\otimes\mathrm{diag}_\infty(s)\otimes e_i = -s+s_M+e\frac{1}{n}\sum_j s_j,$$

where e is a vector of ones and s_M is the centre of mass scaling parameter for $\mathcal{H}$. Thus the centre of mass scaling of $\mathcal{H}'$ is

$$\begin{aligned}\mathcal{C}' &= \mathrm{diag}_\infty(+s - s_M - e\frac{1}{n}\sum_j s_j) \otimes \mathcal{H}' \otimes \mathrm{diag}_\infty(-s + s_M + e\frac{1}{n}\sum_j s_j)\\ &= \mathrm{diag}_\infty(-e\frac{1}{n}\sum_j s_j) \otimes \mathrm{diag}_\infty(-s_M) \otimes \mathcal{H} \otimes \mathrm{diag}_\infty(s_M) \otimes \mathrm{diag}_\infty(e\frac{1}{n}\sum_j s_j)\\ &= \mathrm{diag}_\infty(-e\frac{1}{n}\sum_j s_j) \otimes \mathcal{C} \otimes \mathrm{diag}_\infty(e\frac{1}{n}\sum_j s_j) = \mathcal{C}. \qquad \square\end{aligned}$$

EXAMPLE 5.3. *Consider the matrices $\mathcal{H}$, $\mathcal{H}^\star$ and H of Examples* 3.3 *and* 3.6. *The centre of mass scaling parameter is given by*

$$s_M = \frac{1}{3}(\mathcal{H}^\star \otimes e_1 + \mathcal{H}^\star \otimes e_2 + \mathcal{H}^\star \otimes e_3) = [0, -2/3, -3]^T$$

yielding the Hungarian scaled matrix

$$\mathcal{H}_{s_M} = \mathrm{diag}_\infty(-s_M) \otimes \mathcal{H} \otimes \mathrm{diag}_\infty(s_M) = \begin{bmatrix} 0 & -2/3 & -3 \\ -1/3 & 0 & -13/3 \\ -\infty & -5/3 & 0 \end{bmatrix}.$$

Figure 3.1(*d*) *shows the position of the arithmetic centre of mass scaling parameter s_M in* $\mathrm{col}(\mathcal{H}^*)$. *Table* 3.1 *shows that* 2*-norm condition number of*

$$H_{s_M} = \mathrm{diag}_0\big(\exp(-s_M)\big) H \,\mathrm{diag}_0(\exp\big(s_M)\big)$$

is slightly larger than that for the max-balanced Hungarian matrix H_{s_B} of Example 4.4. *Note that, as for H_{s_B}, the centre of mass Hungarian scaled matrix H_{s_M} is diagonally dominant by row and by column for this example.*

6. Numerical results for linear system scalings. In this section we report on the effect of the max balancing and centre of mass scalings on classical matrices, and on the solution of linear systems $Ax = b$ by Gaussian elimination, where $A \in \mathbb{R}^{n\times n}$ is sparse and nonsymmetric, and $b \in \mathbb{R}^n$. Our test matrices are from the University of Florida Sparse Matrix Collection [4], and come from a variety of applications. We select all real nonsymmetric matrices in the collection for which $100 \le n \le 500$, excluding binary matrices and matrices with a two-norm condition number larger than 10^{15}. This gives us 80 matrices in total.

For each matrix A in our test set we use the HSL code MC64 [6], [7], [14] to compute a Hungarian pair and optimal assignment permutation. The resulting classical Hungarian scaled and reordered matrix is $H = D_1 A D_2 P$. Note that the permutation provided by MC64 should be applied to the columns of A, in contrast to the row

permutation in Corollary 3.2. The max balancing and centre of mass scalings are then applied to H to give

$$B = D_s^{-1} H D_s = (D_s^{-1} D_1) A (D_2 P D_s), \tag{6.1}$$

with D_s a nonsingular diagonal matrix. For comparison, we also report results for the original unscaled matrix A. The right-hand side b is chosen so that all components of the solution x are equal to two, i.e., $x_i = 2$, $i = 1, \ldots, n$.

We make use of performance profiles [5], that allow us to easily display, for all matrices in the test set, how the scalings affect a performance measure like the condition number. To obtain the performance profile we first define the performance ratio for the kth scaling on a given matrix to be the ratio of the performance for scaling k to that of the best possible scaling for that matrix. (Throughout, we assume that the performance measure of interest is one for which a smaller number is better.) The monotonically increasing function $f_k(\alpha)$, $\alpha \in [1, \infty)$ then measures the proportion of matrices for which the performance ratio for scaling k is at most α. Plotting the curves $f_k(\alpha)$ against α for the different scalings gives a performance profile, that shows which scaling performs best or joint-best ($\alpha = 1$) and which scalings are near-best (small α). Additionally, $\lim_{\alpha \to \infty}$ indicates when a scaling fails (say, to produce L and U factors without pivoting) on a matrix for which at least one other scaling does not fail.

6.1. Comparison of scalings. In this section we investigate the impact of our scalings on the diagonal dominance and conditioning of matrices in our test set, and on Gaussian elimination.

Recall that the diagonal similarity scalings tend to improve the diagonal dominance of H by reducing the magnitude of off-diagonal entries (cf. Theorem 4.5). To see this we first measure the number of rows of B in (6.1) that are not diagonally dominant, i.e., for which

$$|b_{ii}| \leq \sum_{\substack{j=1, \\ j \neq i}}^{n} |b_{ij}|,$$

and display the results in a performance profile (see Figure 6.1, left plot). We see that both the max balancing and centre of mass scalings improve the row-wise diagonal dominance, with the max balanced matrices tending to have the most diagonally dominant rows.

However, the left plot in Figure 6.1 does not tell us how close (or far) each row is from a diagonally dominant one. Consequently, the measure

$$\rho = \log \left(\prod_{i=1}^{n} \eta_i \right), \quad \eta_i = \max \left\{ \frac{\sum_{j=1, j \neq i}^{n} |b_{ij}|}{|b_{ii}|}, 1 \right\} \tag{6.2}$$

may be more instructive. (Note that for a Hungarian scaled matrix $|b_{ii}| = 1$.) The performance profile of ρ in Figure 6.1 shows clearly that the max balanced matrices tend to give lower values of ρ, i.e., are in some sense closest to diagonally dominant matrices. The 53 fails correspond to matrices A with at least one zero on the diagonal. The centre of mass scaled matrices tend to have lower values of ρ than the MC64 Hungarian scaled matrices but higher values than the max balanced matrices.

As a by-product of increased diagonal dominance the max balancing and centre of mass scalings may reduce the condition number. The left plot in Figure 6.2 shows

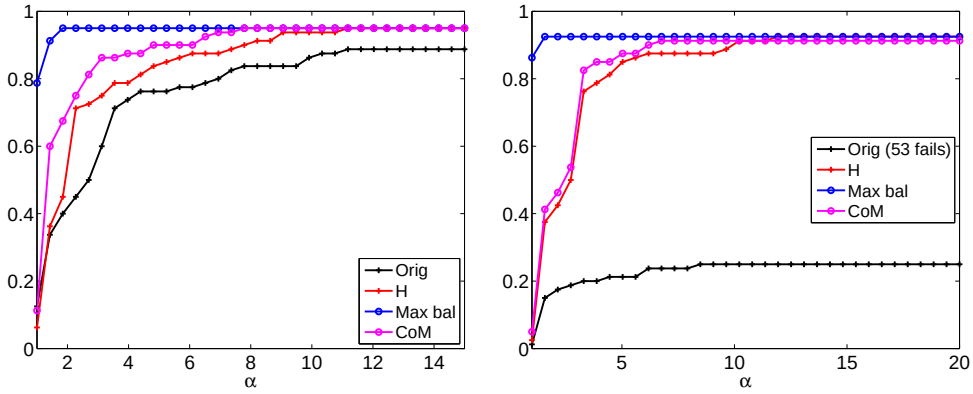


FIG. 6.1. *Performance profile of the number of rows that are not diagonally dominant (left plot) and ρ in (6.2) (right plot).*

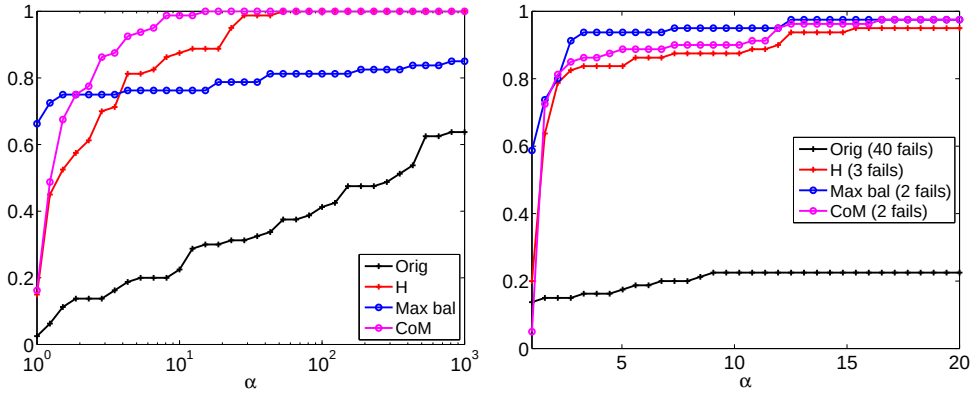


FIG. 6.2. *Performance profile of the 2-norm condition number (left plot) and of $\||L||U|\|_F$ for the L and U factors of $B = LU$ computed without pivoting (right plot).*

that for more than 60 per cent of matrices in our test set the max balanced matrices have the smallest condition number, but for a small number of matrices the condition number increases when we apply the max balancing scaling. However, even for these problems the condition number is still smaller than for the original matrix A. The centre of mass scaled matrices, however, have condition numbers that are always within a factor of 10 of the smallest condition number.

Diagonal dominance by row or column is also important for Gaussian elimination without pivoting since in that case an LU factorization exists and the growth factor is bounded by 2 (see for example [9, Thm. 9.9]). Consequently, the max balancing and centre of mass scalings may improve the stability of the computation of L and U factors when pivoting is not used. It may also reduce the number of interchanges when partial pivoting is employed. To investigate whether the scalings are beneficial for Gaussian elimination, we first measure the stability of the computed factors when no pivoting is used. Since stability is related to the size of $|L||U|$, where $|\cdot|$ is the componentwise absolute value [9, Section 9.3], here we compare the Frobenius norm of $|L||U|$, with a failure indicating that at least one entry of L or U is not finite. We see from Figure 6.2 (right plot) that applying the MC64 Hungarian scaling reduces the

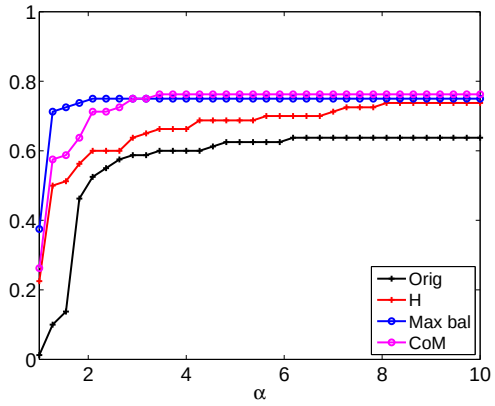


FIG. 6.3. *Performance profile of the number of interchanges needed to perform Gaussian elimination with partial pivoting.*

TABLE 6.1
Number of matrices amongst our set of 80 matrices for which no pivoting is required to perform Gaussian elimination.

Method	Number
Original	3
Hungarian scaled	10
Max balance	16
Centre of Mass	15

number of failures. However, the max balancing scaling improves stability further, most likely because of its effect on diagonal dominance. The value of $\||L||U|\|_F$ is rarely smallest for the centre of mass scaled matrices but is within a factor of 20 of the lowest.

In practice, however, for stability and accuracy we usually require pivoting, with some variant of partial pivoting often used. However, pivoting tends to increase the cost of performing Gaussian elimination, since data structures may require modification and, in multifrontal methods, pivots may be delayed. Accordingly, it is desirable to minimize the amount of pivoting performed. Table 6.1 and Figure 6.3 show that although the MC64 scaling and reordering reduces the number of pivots, as is well documented, we obtain further improvements by applying the max balancing and centre of mass scalings. This is most likely due to improved diagonal dominance. The max balanced matrices tend to require the fewest pivots but the centre of mass scaled matrices are almost as good.

6.2. Subset of matrices. In this section we explore the effect of the scalings in more detail for a subset of matrices from our test set. These matrices have been chosen because they exhibit the range of behaviours seen in the larger test set. The matrices are listed in Table 6.2.

We start by comparing the diagonal dominance of the scaled matrices. For most of the matrices, applying the MC64 Hungarian scaling increases the number of diagonally dominant rows, as shown in Table 6.3. However, the max balancing scaling tends to improve this further, and often by a large amount. For example, the number of diagonally dominant rows increases almost six times for `impcol_c` and almost eight times for `robot`. (Note that although `utm300` and `robot` have fewer diagonally

TABLE 6.2
Information on subset of test problems.

Matrix	n	Application
bwm200	200	chemical process simulation
cz148	148	closest point method for computing on surfaces
fs_183_1	183	nonsymmetric facsimile convergence matrix
impcol_c	137	chemical process simulation
oscil_dcop_01	430	circuit simulation
rajat05	301	circuit simulation
robot	120	robotics problem
str_200	363	linear programming
tols340	340	computational fluid dynamics
utm300	300	electromagnetics

TABLE 6.3
Number of diagonally dominant rows and the quantity ρ in (6.2).

Matrix	Orig	H	Max bal	CoM	Orig	H	Max bal	CoM
bwm200	4	**102**	**102**	**102**	1.8	1.9	**1.7**	1.9
cz148	1	13	52	**56**	4	5.7	**3.3**	5.9
fs_183_1	75	150	**180**	163	7e2	27	**2.7**	17
impcol_c	14	22	**114**	74	∞	62	**5.3**	26
oscil_dcop_01	172	220	**330**	284	∞	77	**27**	74
rajat05	91	151	**245**	189	∞	63	**16**	55
robot	26	12	**83**	41	∞	33	**14**	21
str_200	0	100	**327**	193	∞	90	**23**	69
tols340	204	267	**340**	322	∞	38	**0**	26
utm300	96	67	**100**	79	2.9e2	2.1e2	**1.8e2**	2.0e2

dominant rows after the MC64 scaling is applied, in both cases A has zeros on the diagonal, so that the MC64 Hungarian scaling and optimal assignment permutation may be beneficial overall.) With the exception of `cz148` the centre of mass scaled matrices have fewer diagonally dominant rows than the max balanced matrices, but more than the MC64 Hungarian scaled matrices.

The quantity ρ in (6.2) measures how close the rows of the scaled matrix are to diagonally dominant ones, and is tabulated in Table 6.3. We see that ρ is smallest for the max balanced matrices and is often much smaller than for the MC64 Hungarian scaled matrices. The centre of mass scaling also tends to reduce ρ, relative to the MC64 scaling, and for some problems this reduction is also large.

Since the matrix 2-norm is greatly reduced by the MC64 scaling, the max balancing and centre of mass scalings only offer small additional gains (see Table 6.4). However, the scalings have a significant impact on the 2-norm condition number, which is up to two orders of magnitude smaller for the max balanced matrices than for the MC64 scaled matrices. Note, however, that the max balancing scaling is not guaranteed to improve the condition number as `oscil_dcop_01` demonstrates. The centre of mass scaled matrices generally have smaller condition numbers than the MC64 Hungarian scaled matrices. However, these condition numbers are larger than for the max balanced matrices with the exception of `oscil_dcop_01`.

Finally, we examine in Table 6.5 the number of interchanges needed when Gaussian elimination with partial pivoting is applied to the scaled matrices. The max balanced matrices tend to require fewer interchanges, and for `cz148` and `tols340` no interchanges are required. On the other hand significantly more interchanges are needed for `rajat05` after the max balancing scaling is applied. For most matrices, the centre of mass scaled matrices require slightly more interchanges than the max

TABLE 6.4
The Frobenius norm (second column) and 2-norm condition number (third column).

Matrix	Orig	H	Max bal	CoM	Orig	H	Max bal	CoM
bwm200	8.5e3	**1.7e1**	**1.7e1**	**1.7e1**	2.4e3	**1.8e3**	**1.8e3**	**1.8e3**
cz148	1.4e3	**1.4e1**	**1.4e1**	**1.4e1**	2.5e3	8.9e2	**8.1e2**	8.4e2
fs_183_1	1.1e9	1.9e1	**1.4e1**	1.6e1	2.2e13	4.3e2	**1.7e1**	2.6e1
impcol_c	1.5e2	1.9e1	**1.3e1**	1.6e1	1.8e4	8e2	**2.1e2**	3.4e2
oscil_dcop_01	4.1e6	2.8e1	**2.5e1**	2.8e1	5.9e12	1.2e10	1.8e12	**1.1e10**
rajat05	2.4e1	2.4e1	**2.0e1**	2.3e1	5.4e4	2.5e4	**1.1e4**	1.5e4
robot	2.8e2	1.6e1	**1.3e1**	1.5e1	4.3e8	5.2e1	**2.1e1**	3.4e1
str_200	3.7e1	2.6e1	**2.1e1**	2.5e1	1.5e4	1.1e3	**2.9e2**	7.2e2
tols340	8e5	2.3e1	**1.8e1**	2.1e1	2e5	3.1e1	**1.1e0**	8.7e0
utm300	1.7e1	2.8e1	**2.5e1**	2.7e1	8.5e5	1.6e5	**7.6e3**	4.2e4

TABLE 6.5
The number of interchanges needed to perform Gaussian elimination with partial pivoting.

Matrix	Original	MC64	Max bal	CoM
bwm200	**236**	**236**	**236**	**236**
cz148	12	4	**0**	**0**
fs_183_1	8	**0**	**0**	**0**
impcol_c	194	38	**12**	24
oscil_dcop_01	384	278	242	**220**
rajat05	162	**76**	144	80
robot	102	122	**22**	26
str_200	724	144	**20**	58
tols340	272	64	**0**	**0**
utm300	312	254	**96**	190

balanced matrices but fewer than the MC64 scaled matrices.

7. Conclusion. The fact that a given matrix A usually has a range of possible Hungarian scalings has been overlooked in matrix computation. Given a Hungarian pair (u, v) of A, we have provided a characterization of the set of all Hungarian pairs of A in terms of (u, v) and a vector parameter s that belongs to a specific set of vectors. This gives us a way to build many Hungarian scaled matrices from a Hungarian scaled matrix and to look for Hungarian scaled matrices with particular properties such as diagonal dominance. We used our characterization to (a) show that the Schneider and Schneider's max-balancing algorithm preserves the Hungarian scaled property, to (b) construct a special visualization scaling, called centre of mass scaling, that is aimed at approximating the max-balancing scaling and whose computation is embarrassingly parallel, and to (c) show that max-balancing scaling and centre of mass scaling both produce a unique Hungarian scaled matrix. We have shown that max-balancing minimizes the entrywise p-norm according to the ordering $\prec$. As a result, the max-balancing of a Hungarian scaled matrix minimizes the off-diagonal entries while maintaining the weight on the diagonal (all diagonal entries have modulus one), leading to a matrix which is more diagonally dominant than the other Hungarian scalings of A. This was confirmed by our numerical experiments. The latter also showed that, when combined with Hungarian scaling, max-balancing and centre of mass scalings improve the stability of LU factorizations with no pivoting.

Appendix A. This appendix provides a technical lemma with its proof and the proofs of Theorems 4.3 and 4.5.

A.1. Properties of the max-balancing algorithm. Suppose that we apply Algorithm 4.2 to an irreducible matrix $\mathcal{A} \in \mathbb{R}_{\max}^{n\times n}$. We obtain a sequence of matrices $\mathcal{A}_1, \dots, \mathcal{A}_k$ of dimensions $n = m_0, \dots, m_k$, and sequences of cycle means $\beta_1, \dots, \beta_k$, critical cycles $C_1, \dots, C_k$, subeigenvectors $s_1, \dots, s_k$ and projections $f_1 = \mathrm{id}, \dots, f_k$, where $k \le n$ is the number of steps required for the max-balancing algorithm to terminate. For $1 \le \ell \le k$ define $g_\ell : \{1, \dots, n\} \mapsto \{1, \dots, m_{\ell-1}\}$ and $u_\ell \in \mathbb{R}^n$ by

$$g_\ell = f_\ell \circ \cdots \circ f_1, \qquad u_\ell = s_1(g_1) + s_2(g_2) + \cdots + s_\ell(g_\ell) \in \mathbb{R}^n,$$

where $\big(s_\ell(g_\ell)\big)_i = (s_\ell)_{g_\ell(i)}$. We let $g_{k+1} : \{1, \dots, n\} \mapsto \{1\}$ so that $g_{k+1}(i) = g_{k+1}(j)$ for all $i, j = 1, \dots, n$. Also define $\mathcal{M}_\ell, \mathcal{U}_\ell, \mathcal{Q}_\ell \in \mathbb{R}_{\max}^{n\times n}$ by

$$(\mathcal{M}_\ell)_{ij} = \begin{cases} \beta_\ell & \text{if } g_{\ell+1}(i) = g_{\ell+1}(j), \\ -\infty & \text{otherwise,} \end{cases} \qquad \ell = 1, \dots, k-1,$$

$$\mathcal{U}_1 = \beta_1 \otimes \mathcal{O}_n, \qquad \mathcal{U}_\ell = \mathcal{M}_1 \oplus \cdots \oplus \mathcal{M}_{\ell-1} \oplus \beta_\ell \otimes \mathcal{O}_n, \quad \ell = 2, \dots, k,$$

$$\mathcal{Q}_\ell = \mathcal{A}_1 / \mathcal{U}_\ell, \quad \ell = 1, \dots, k.$$

The next lemma gives a list of properties that the above quantities have.

LEMMA A.1. *With the above notation the following statements hold.*

(a) $\beta_1 \ge \cdots \ge \beta_k > -\infty$.
(b) $\mathcal{A}_\ell = \mathrm{contr}\big(\mathrm{diag}_\infty(-u_{\ell-1}) \otimes \mathcal{A}_1 \otimes \mathrm{diag}_\infty(u_{\ell-1}), g_\ell\big)$, $\ell = 2, \dots, k$.
(c) $\mathrm{diag}_\infty(-u_\ell) \otimes \mathcal{A}_1 \otimes \mathrm{diag}_\infty(u_\ell) \le \mathcal{U}_\ell$, $\ell = 1, \dots, k$.
(d) $u_\ell \in \mathrm{col}(\mathcal{Q}_\ell^\star)$, $\ell = 1, \dots, k$.
(e) $\mathrm{col}(\mathcal{Q}_k^\star) \subseteq \cdots \subseteq \mathrm{col}(\mathcal{Q}_1^\star)$.
(f) *Let $q \in \mathrm{col}(\mathcal{Q}_\ell^\star)$ for some ℓ such that $1 \le \ell \le k$. Then*

$$\big(\mathrm{diag}_\infty(-u_\ell) \otimes \mathcal{A} \otimes \mathrm{diag}_\infty(u_\ell)\big)_{ij} = \big(\mathrm{diag}_\infty(-q) \otimes \mathcal{A} \otimes \mathrm{diag}_\infty(q)\big)_{ij}$$

whenever $g_{\ell+1}(i) = g_{\ell+1}(j)$.

Proof. (a) See [17, Lem. 4], which proves that $\beta_1 \ge \cdots \ge \beta_k$. That these cycle means are all finite follows from the fact that any contraction of an irreducible matrix is also an irreducible matrix so that while $m_\ell > 1$ the graph $\Gamma(\mathcal{A}_\ell)$ contains at least one cycle of finite weight and therefore $\beta_\ell > -\infty$ for $\ell = 1, \dots, k$.

(b) By construction, $\mathcal{A}_\ell = \mathrm{contr}\big(\mathrm{diag}_\infty(-s_{\ell-1}) \otimes \mathcal{A}_{\ell-1} \otimes \mathrm{diag}_\infty(s_{\ell-1}), f_\ell\big)$, and since $u_1 = s_1$ and $g_2 = f_2 \circ f_1 = f_2$, the claim is true for $\ell = 2$. Now suppose that for $\ell > 2$, $\mathcal{A}_{\ell-1} = \mathrm{contr}\big(\mathrm{diag}_\infty(-u_{\ell-2}) \otimes \mathcal{A}_1 \otimes \mathrm{diag}_\infty(u_{\ell-2}), g_{\ell-1}\big)$ and, as in Algorithm 4.2, let $\mathcal{S}_\ell = \mathrm{diag}_\infty(-s_{\ell-1}) \otimes \mathcal{A}_{\ell-1} \otimes \mathrm{diag}_\infty(s_{\ell-1})$. For $i \ne j$, we have

$$(\mathcal{S}_\ell)_{ij} = -(s_{\ell-1})_i + (s_{\ell-1})_j + \max_{\substack{(i',j'):\\ g_{\ell-1}(i',j')=(i,j)}} \big(\mathrm{diag}_\infty(-u_{\ell-2}) \otimes \mathcal{A}_1 \otimes \mathrm{diag}_\infty(u_{\ell-2})\big)_{i'j'},$$

and since $\big(s_{\ell-1}(g_{\ell-1})\big)_i = \big(s_{\ell-1}(g_{\ell-1})\big)_j$ whenever $g_{\ell-1}(i) = g_{\ell-1}(j)$ we have

$$\begin{aligned}(\mathcal{S}_\ell)_{ij} &= \max_{(i',j'):g_{\ell-1}(i',j')=(i,j)} \big\{ -\big(s_{\ell-1}(g_{\ell-1})\big)_{i'} + \big(s_{\ell-1}(g_{\ell-1})\big)_{j'} \\ &\qquad + \big(\mathrm{diag}_\infty(-u_{\ell-2}) \otimes \mathcal{A}_1 \otimes \mathrm{diag}_\infty(u_{\ell-2})\big)_{i'j'} \big\} \\ &= \max_{(i',j'):g_{\ell-1}(i',j')=(i,j)} \big(\mathrm{diag}_\infty(-u_{\ell-1}) \otimes \mathcal{A}_1 \otimes \mathrm{diag}_\infty(u_{\ell-1})\big)_{i'j'} \\ &= \mathrm{contr}\big(\mathrm{diag}_\infty(-u_{\ell-1}) \otimes \mathcal{A}_1 \otimes \mathrm{diag}_\infty(u_{\ell-1}), g_{\ell-1}\big)\big)_{ij}.\end{aligned}$$

Finally, since $g_\ell = f_\ell \circ g_{\ell-1}$, we have

$$\begin{aligned}\mathcal{A}_\ell = \operatorname{contr}\big(\mathcal{S}_\ell, f_\ell\big) &= \operatorname{contr}\Big(\operatorname{contr}\big(\operatorname{diag}_\infty(-u_{\ell-1}) \otimes \mathcal{A}_1 \otimes \operatorname{diag}_\infty(u_{\ell-1}), g_{\ell-1}\big), f_\ell\Big) \\ &= \operatorname{contr}\big(\operatorname{diag}_\infty(-u_{\ell-1}) \otimes \mathcal{A}_1 \otimes \operatorname{diag}_\infty(u_{\ell-1}), g_\ell\big).\end{aligned}$$

(c) Define $\mathcal{B}_\ell := \operatorname{diag}_\infty(-u_\ell) \otimes \mathcal{A}_1 \otimes \operatorname{diag}_\infty(u_\ell)$, $\ell = 1, \dots, k$. For $\ell = 1$ we have $u_1 = s_1$, which is a subeigenvector of $\mathcal{A}_1$ corresponding to the maximum cycle mean β_1, so the claim follows from Lemma 2.3(a).

For $\ell \geq 2$ suppose that $\mathcal{B}_{\ell-1} \leq \mathcal{U}_{\ell-1}$. Now since $u_\ell = s_\ell(g_\ell) + u_{\ell-1}$ we have $\mathcal{B}_\ell = \operatorname{diag}_\infty\big(-s_\ell(g_\ell)\big) \otimes \mathcal{B}_{\ell-1} \otimes \operatorname{diag}_\infty\big(s_\ell(g_\ell)\big)$. First consider i, j such that $g_\ell(i) = g_\ell(j)$. We have $\big(s_\ell(g_\ell)\big)_i = \big(s_\ell(g_\ell)\big)_j$ so that $(\mathcal{B}_\ell)_{ij} = (\mathcal{B}_{\ell-1})_{ij}$ and from Lemma A.1(a) we also have $(\mathcal{U}_\ell)_{ij} = (\mathcal{U}_{\ell-1})_{ij}$ so that $(\mathcal{B}_\ell)_{ij} \leq (\mathcal{U}_\ell)_{ij}$.

Now consider i, j such that $g_\ell(i) \neq g_\ell(j)$. We have $(\mathcal{U}_\ell)_{ij} = \beta_\ell$ and

$$(\mathcal{B}_\ell)_{ij} \leq \operatorname{contr}(\mathcal{B}_\ell, g_\ell)_{g_\ell(i), g_\ell(j)}.$$

But, from Lemma A.1(b), we have $\operatorname{contr}(\mathcal{B}_\ell, g_\ell) = \operatorname{diag}_\infty(-s_\ell) \otimes \mathcal{A}_\ell \otimes \operatorname{diag}_\infty(s_\ell)$, where by construction s_ℓ is a subeigenvector for $\mathcal{A}_\ell$ corresponding to the maximum cycle mean β_ℓ so that, from Lemma 2.3(a), we have

$$(\mathcal{B}_\ell)_{ij} \leq \big(\operatorname{diag}_\infty(-s_\ell) \otimes \mathcal{A}_\ell \otimes \operatorname{diag}_\infty(s_\ell)\big)_{g_\ell(i) g_\ell(j)} \leq \beta_\ell = (\mathcal{U}_\ell)_{ij}.$$

(d) This follows immediately from Lemma A.1(c) and Theorem 3.4.

(e) It follows from Lemma A.1(c) that the inequality

$$\operatorname{diag}_\infty(-s_\ell) \otimes \mathcal{A}_1 \otimes \operatorname{diag}_\infty(s_\ell) \leq \mathcal{U}_\ell$$

is feasible for $\ell = 1, \dots, k$. Therefore by Theorem 3.4 we have $\lambda_{\max}(\mathcal{Q}_\ell) \leq 0$ and $\operatorname{col}(\mathcal{Q}_\ell^\star) = \{s \in \mathbb{R}^n : \operatorname{diag}_\infty(-s) \otimes \mathcal{A}_1 \otimes \operatorname{diag}_\infty(s) \leq \mathcal{U}_\ell\}$. Now from Lemma A.1(a) it follows that $\mathcal{U}_k \leq \cdots \leq \mathcal{U}_1$ so that $\operatorname{col}(\mathcal{Q}_\ell^\star) \subseteq \operatorname{col}(\mathcal{Q}_{\ell-1}^\star)$ for $\ell \geq 2$.

(f) Let $\mathcal{B}_\ell := \operatorname{diag}_\infty(-u_\ell) \otimes \mathcal{A}_1 \otimes \operatorname{diag}_\infty(u_\ell)$ and let $\mathcal{B}' := \operatorname{diag}_\infty(-q) \otimes \mathcal{A}_1 \otimes \operatorname{diag}_\infty(q)$.

First consider the case $\ell = 1$ and suppose that $q \in \operatorname{col}(\mathcal{Q}_1^\star)$. We have $\mathcal{B}_1, \mathcal{B}' \leq \mathcal{U}_1 = \beta_1 \otimes \mathcal{O}_n$, where β_1 is the maximal cycle mean of $\mathcal{A}_1$. Thus it follows from Lemma 2.3(a) that u_1 and q are both subeigenvectors of $\mathcal{A}_1$ and therefore from Lemma 2.3(c) that $-(u_1)_i + (u_1)_j = -q_i + q_j$, whenever i and j are both vertices on C_1 or equivalently whenever $g_2(i) = g_2(j)$.

Now assume that the statement is true for $\ell - 1$ and suppose that $q \in \operatorname{col}(\mathcal{Q}_\ell^\star)$. From Lemma A.1(e) we have $\operatorname{col}(\mathcal{Q}_\ell^\star) \subset \operatorname{col}(\mathcal{Q}_{\ell-1}^\star)$ so that $q \in \operatorname{col}(\mathcal{Q}_{\ell-1}^\star)$ and therefore from the induction hypothesis we have $-(u_{\ell-1})_i + (u_{\ell-1})_j = -q_i + q_j$ whenever $g_\ell(i) = g_\ell(j)$. Therefore $q = u_{\ell-1} + t$ for some $t \in \mathbb{R}^n$ such that $t_i = t_j$ whenever $g_\ell(i) = g_\ell(j)$. Thus there exists $r \in \mathbb{R}^{m_{\ell-1}}$ such that $t = r(g_\ell)$ and $q = u_{\ell-1} + r(g_\ell)$ and $\mathcal{B}' = \operatorname{diag}_\infty\big(-r(g_\ell)\big) \otimes \mathcal{B}_\ell \otimes \operatorname{diag}_\infty\big(r(g_\ell)\big)$.

From Lemma A.1(b) we have $\mathcal{A}_\ell = \operatorname{contr}(\mathcal{B}_{\ell-1}, g_\ell) \in \mathbb{R}_{\max}^{m_\ell \times m_\ell}$, where $\mathcal{A}_\ell$ has maximum cycle mean β_ℓ with C_ℓ a critical cycle and s_ℓ a subeigenvector. Also define

$$\begin{aligned}\mathcal{A}' = \operatorname{contr}\big(\mathcal{B}', g_\ell\big) &= \operatorname{contr}\big(\operatorname{diag}_\infty\big(-r(g_\ell)\big) \otimes \mathcal{B}_\ell \otimes \operatorname{diag}_\infty\big(r(g_\ell)\big), g_\ell\big) \\ &= \operatorname{diag}_\infty(-r) \otimes \operatorname{contr}(\mathcal{B}_{\ell-1}, g_\ell) \otimes \operatorname{diag}_\infty(r) \\ &= \operatorname{diag}_\infty(-r) \otimes \mathcal{A}_\ell \otimes \operatorname{diag}_\infty(r).\end{aligned}$$

Since $\mathcal{B}' \leq \mathcal{U}_\ell$ we have $\mathcal{A}' = \operatorname{contr}(\mathcal{B}', g_\ell) \leq \operatorname{contr}(\mathcal{U}_\ell, g_\ell)$, where $\operatorname{contr}(\mathcal{U}_\ell, g_\ell) = \beta_\ell \otimes \mathcal{O}_{m_{\ell-1}}$, so that $(\mathcal{A}')_{ij} \leq \beta_\ell$ for all $i, j = 1, \ldots, n$. Therefore from Lemma 2.3 (a), r must also be a subeigenvector of $\mathcal{A}_\ell$ and therefore from Lemma 2.3 (c), $-r_i + r_j = -(s_\ell)_i + (s_\ell)_j$ whenever i and j are both vertices visited by the cycle C_ℓ or equivalently whenever $f_{\ell+1}(i) = f_{\ell+1}(j)$.

Finally note that $g_{\ell+1} = f_{\ell+1} \circ g_\ell$ so that whenever we have $g_{\ell+1}(i) = g_{\ell+1}(j)$ we have $f_{\ell+1}\big(g_\ell(i)\big) = f_{\ell+1}\big(g_\ell(j)\big)$ and

$$\begin{aligned} -q_i + q_j &= -(u_{\ell-1})_i + (u_{\ell-1})_j - r(g_\ell)_i + r(g_\ell)_j \\ &= -(u_{\ell-1})_i + (u_{\ell-1})_j - s_\ell(g_\ell)_i + s_\ell(g_\ell)_j \\ &= -(u_\ell)_i + (u_\ell)_j. \qquad \square \end{aligned}$$

A.2. Proof of Theorem 4.3. We show that $s \in \operatorname{col}(\mathcal{H}^\star)$. For that, let us apply Algorithm 4.2 to $\mathcal{H}$. Then $\mathcal{A}_1$ is simply $\mathcal{H}$ with diagonal entries replaced by $-\infty$ and $\mathcal{Q}_1 = \mathcal{A}_1/\mathcal{U}_1 = \mathcal{A}_1/(\mathcal{O}_n \otimes \beta_1) = \mathcal{A}_1 \otimes -\beta_1$, where $\beta_1 = \lambda_{\max}(\mathcal{A}_1)$. By Theorem 3.1, $\lambda_{\max}(\mathcal{H}) = 0$ and from the definition of the maximum cycle mean, $\beta_1 = \lambda_{\max}(\mathcal{A}_1) \leq \lambda_{\max}(\mathcal{H}) = 0$ and $\lambda_{\max}(\mathcal{Q}_1) = 0$. From Lemma A.1 (d) and (e) we have that $s = u_k \in \operatorname{col}(\mathcal{Q}_k^\star) \subseteq \operatorname{col}(\mathcal{Q}_1^\star)$. Since $\lambda_{\max}(\mathcal{Q}_1) \leq 0$, Theorem 3.4 says that

$$\operatorname{col}(\mathcal{Q}_1^\star) = \{t \in \mathbb{R}^n : \operatorname{diag}_\infty(-t) \otimes \mathcal{A}_1 \otimes \operatorname{diag}_\infty(t) \leq \beta_1 \otimes \mathcal{O}_n\}.$$

Also, since $\lambda_{\max}(\mathcal{H}) \leq 0$

$$\begin{aligned} \operatorname{col}(\mathcal{H}^\star) &= \{t \in \mathbb{R}^n : \operatorname{diag}_\infty(-t) \otimes \mathcal{H} \otimes \operatorname{diag}_\infty(t) \leq \mathcal{O}_n\} \\ &= \{t \in \mathbb{R}^n : \operatorname{diag}_\infty(-t) \otimes \mathcal{A}_1 \otimes \operatorname{diag}_\infty(t) \leq \mathcal{O}_n\} \end{aligned}$$

so that $\operatorname{col}(\mathcal{Q}_1^\star) \subseteq \operatorname{col}(\mathcal{H}^\star)$ and hence, $s \in \operatorname{col}(\mathcal{H}^\star)$.

That $(u+s, v-s)$ is the unique Hungarian pair of $\mathcal{A}$ that max-balances $\mathcal{A}$ follows from Theorem 4.1.

A.3. Proof of Theorem 4.5. Define $A_t = \operatorname{diag}_0\big(\exp(-t)\big) A \operatorname{diag}_0\big(\exp(t)\big)$ for $t = s, r \in \mathbb{R}^n$ and let $\mathcal{A}_t := \mathcal{V}(A_t) = \operatorname{diag}_\infty(-t) \otimes \mathcal{A} \otimes \operatorname{diag}_\infty(t)$, where $\mathcal{A} = \mathcal{V}(A)$ and $\mathcal{V}(\cdot)$ is the valuation in (2.1). We use the notation introduced before Lemma A.1.

Suppose that $s \in \operatorname{col}(\mathcal{Q}_1^\star)$ and $r \notin \operatorname{col}(\mathcal{Q}_1^\star)$. It follows from Theorem 3.4 that $\mathcal{A}_s \leq \beta_1 \otimes \mathcal{O}_n$ and $\mathcal{A}_r \not\leq \beta_1 \otimes \mathcal{O}_n$. Thus all entries in A_s are of modulus less than or equal to $\exp(\beta_1)$ while A_r contains at least one entry of modulus greater than $\exp(\beta_1)$. Therefore we have $A_s \prec A_r$.

Now suppose that $s \in \operatorname{col}(\mathcal{Q}_\ell^\star)$ and $r \in \operatorname{col}(\mathcal{Q}_{\ell-1}^\star)/\operatorname{col}(\mathcal{Q}_\ell^\star)$ with $\ell > 1$. It follows from Lemma A.1 (e) that $s, r \in \operatorname{col}(\mathcal{Q}_{\ell-1}^\star)$ and therefore from Lemma A.1 (f) we have that $|(A_s)_{ij}| = |(A_r)_{ij}|$ whenever $g_\ell(i) = g_\ell(j)$. Thus

$$\|A_s\|_p^p - \|A_r\|_p^p = \sum_{(i,j):g_\ell(i) \neq g_\ell(j)} |(A_s)_{ij}|^p - \sum_{(i,j):g_\ell(i) \neq g_\ell(j)} |(A_r)_{ij}|^p.$$

It follows from Theorem 3.4 that

$$\begin{aligned} &\mathcal{A}_s \leq \mathcal{M}_1 \oplus \cdots \oplus \mathcal{M}_{\ell-1} \oplus \mathcal{O}_n \otimes \beta_\ell, \\ &\mathcal{A}_r \leq \mathcal{M}_1 \oplus \cdots \oplus \mathcal{M}_{\ell-2} \oplus \mathcal{O}_n \otimes \beta_{\ell-1}, \quad \mathcal{A}_r \not\leq \mathcal{M}_1 \oplus \cdots \oplus \mathcal{M}_{\ell-1} \oplus \mathcal{O}_n \otimes \beta_\ell. \end{aligned}$$

Therefore $(\mathcal{A}_s)_{ij} \leq \beta_\ell$ whenever $g_\ell(i) \neq g_\ell(j)$ but there exists at least one pair (i,j) such that $g_\ell(i) \neq g_\ell(j)$ and $(\mathcal{A}_r)_{ij} > \beta_\ell$ from which it follows that $A_s \prec A_r$.

Finally let $s \in \mathrm{col}(\mathcal{Q}_k^\star)$ be a max-balancing scaling and let $r \in \mathbb{R}^n$ be any other scaling parameter. If $r \notin \mathrm{col}(\mathcal{Q}_1^\star)$ then $A_s \prec A_r$. Otherwise if $r \in \mathrm{col}(\mathcal{Q}_1^\star)$ but $r \notin \mathrm{col}(\mathcal{Q}_2^\star)$ then $A_s \prec A_r$ and so on. Either $A_s \prec A_r$ or $r \in \mathrm{col}(\mathcal{Q}_k^\star)$ in which case $A_s = A_r$.

REFERENCES

[1] M. Benzi, J. C. Haws, and M. Tůma. Preconditioning highly indefinite and nonsymmetric matrices. *SIAM J. Sci. Comput.*, 22(4):1333–1353, 2000.

[2] P. Butkovič. *Max-Linear Systems: Theory and Algorithms*. Springer-Verlag, London, 2010. xvii+272 pp.

[3] P. Butkovič and H. Schneider. Applications of max algebra to diagonal scaling of matrices. *Electron. J. Linear Algebra*, 13:262–273, 2005.

[4] T. A. Davis and Y. Hu. The University of Florida sparse matrix collection. *ACM Trans. Math. Software*, 38(1):1:1–1:25, 2011.

[5] E. D. Dolan and J. J. Moré. Benchmarking optimization software with performance profiles. *Math. Programming*, 91:201–213, 2002.

[6] I. S. Duff and J. Koster. The design and use of algorithms for permuting large entries to the diagonal of sparse matrices. *SIAM J. Matrix Anal. Appl.*, 20:889–901, 1999.

[7] I. S. Duff and J. Koster. On algorithms for permuting large entries to the diagonal of a sparse matrix. *SIAM J. Matrix Anal. Appl.*, 22:973–996, 2001.

[8] A. Frank. On Kuhn's Hungarian method–a tribute from Hungary. Technical Report TR-2004-14, Egerváry Research Group, Budapest, 2004. `www.cs.elte.hu/egres`.

[9] N. J. Higham. *Accuracy and Stability of Numerical Algorithms*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, second edition, 2002. ISBN 0-89871-521-0. xxviii+680 pp.

[10] J. D. Hogg and J. A. Scott. The effects of scalings on the performance of a sparse symmetric indefinite solver. Report RAL-TR-2008-007, Atlas Centre, Rutherford Appleton Laboratory, Didcot, Oxon, UK, 2008.

[11] J. D. Hogg and J. A. Scott. On the efficient scaling of sparse symmetric matrices using an auction algorithm. Report RAL-P-2014-002, Atlas Centre, Rutherford Appleton Laboratory, Didcot, Oxon, UK, 2014.

[12] J. Hook. Max-plus singular values. MIMS EPrint 2014.7, Manchester Institute for Mathematical Sciences, The University of Manchester, UK, Mar. 2014.

[13] J. Hook and F. Tisseur. Max-plus eigenvalues and singular values: a useful tool in numerical linear algebra, 2015. In preparation.

[14] HSL. A collection of Fortran codes for large scale scientific computation. http://www.hsl.rl.ac.uk/.

[15] M. Olschowka and A. Neumaier. A new pivoting strategy for Gaussian elimination. *Linear Algebra Appl.*, 240:131–151, 1996.

[16] J. B. Orlin and Y. Lee. Quickmatch–a very fast algorithm for the assignment problem. Working papers 3547-93, Massachusetts Institute of Technology (MIT), Sloan School of Management, 1993.

[17] H. Schneider and M. H. Schneider. Max-balancing weighted directed graphs and matrix scaling. *Math. Oper. Res.*, 16(1):208–222, Feb. 1991.

[18] S. Sergeev, H. Schneider, and P. Butkovič. On visualization scaling, subeigenvectors and Kleene stars in max algebra. *Linear Algebra Appl.*, 431:2395–2406, 2009.

[19] N. E. Young, R. E. Tarjan, and J. B. Orlin. Faster parametric shortest path and minimum balance algorithms. *Networks*, 1(2):205–221, 1991.