



EULER-RICHARDSON METHOD PRECONDITIONED BY WEAKLY STOCHASTIC MATRIX ALGEBRAS: A POTENTIAL CONTRIBUTION TO PAGERANK COMPUTATION*

S. CIPOLLA[†], C. DI FIORE[†], AND F. TUDISCO[‡]

Abstract. Let S be a column stochastic matrix with at least one full row. Then S describes a Pagerank-like random walk since the computation of the Perron vector x of S can be tackled by solving a suitable M-matrix linear system $Mx = y$, where $M = I - \tau A$, A is a column stochastic matrix and τ is a positive coefficient smaller than one. The Pagerank centrality index on graphs is a relevant example where these two formulations appear. Previous investigations have shown that the Euler-Richardson (ER) method can be considered in order to approach the Pagerank computation problem by means of preconditioning strategies. In this work, it is observed indeed that the classical power method can be embedded into the ER scheme, through a suitable simple preconditioner. Therefore, a new preconditioner is proposed based on fast Householder transformations and the concept of low complexity weakly stochastic algebras, which gives rise to an effective alternative to the power method for large-scale sparse problems. Detailed mathematical reasonings for this choice are given and the convergence properties discussed. Numerical tests performed on real-world datasets are presented, showing the advantages given by the use of the proposed Householder-Richardson method.

Key words. Matrix algebras, Preconditioning, Nonnegative matrices, Pagerank.

AMS subject classifications. 15B48, 15B51, 65F08.

1. Introduction. Markov chains are used to model many different real world systems which evolve in time. When the total number of states which the system may occupy is finite, the chain is typically represented by a column stochastic matrix S . The state of equilibrium is described by the ergodic distribution p , defined as the solution of the eigenproblem $Sp = p$. Under suitable hypotheses on S , for example irreducibility, the solution p is unique and entry-wise positive. The problem of computing such p is one of the crucial issues in Markov processes analysis.

The power method is one of the simplest iterative schemes that converges to the solution p (provided that the eigenvalues of S different from one have absolute value smaller than one). The rate of convergence of this method is well known to be proportional to the magnitude of the subdominant eigenvalue of S . Due to its simplicity and its well understood limit behaviour, this method is often used in practice, especially for large-scale unstructured problems.

Examples of growing interest in recent literature are connected with the analysis of complex networks where the pattern of the edges of the network is used for localizing important nodes or group of nodes. Many important models, based on matrices or functions of matrices and describing certain features of the network, are related with a random walk defined on the graph and thus exploit extremal eigenvectors and eigenvalues of such matrices (see, for example, [20, 21, 31, 32, 33, 34]). A popular example to which we are particularly interested in is the centrality index problem on graphs known as the Pagerank problem (see [1] for instance). In that case the web surfer moves randomly on the web graph $W = (V, E)$ and the importance of each node

*Received by the editor on June 29, 2016. Accepted for publication on April 20, 2017. Handling Editor: Dario Bini. The work has been partially supported by INdAM-GNCS and, for F.T., by the ERC grant NOLEPRO.

[†]Department of Mathematics, University of Rome "Tor Vergata", Rome, Italy (stefano.cipolla87@gmail.com).

[‡]Department of Mathematics, University of Padua, Padua, Italy.

in V is given by the ergodic distribution $w = Gw$ of the random walk defined on W by the Google engine web matrix G (see Section 1.2 for more details). The dimension of w in that case is the number of web pages that populate the World Wide Web, thus w roughly has 10^9 entries. The power method can be performed on G in a relatively cheap way by means of the transition matrix of the graph, which is typically sparse. On the other hand, the original formula by Brin and Page [8] defines the same Pagerank vector w as the solution of a linear system whose coefficient matrix is an M-matrix and, as a consequence, the ergodic Pagerank distribution w can be computed either by solving the eigenproblem or by solving such linear system. Thus, one can use any linear system solver to approximate w , and several approaches have been investigated and compared to the power method, e.g., [13, 22, 23, 40]. Although such methods sometimes have a convergence rate greater than the one achieved by the power method, they are often more demanding in terms of memory storage and number of operations per step.

The equivalence between the eigenproblem $Sp = p$ and a linear system problem holds in general for a large set of stochastic matrices, not only the Google matrix. Indeed, it has been observed in [39] that, if S is a column stochastic matrix having at least one full row, then 1 is a simple and dominant eigenvalue of S , the ergodic distribution $Sp = p$ is well defined and p is also a solution of an M-matrix linear system problem associated to S . In this work, we propose a class of simple iterative schemes, named preconditioned Euler-Richardson, to solve such linear system. These methods can be seen as a subset of the class of stationary iterative methods often introduced in terms of a splitting of the coefficient matrix, e.g., [30, 43]. Here we observe that this kind of methods provides a natural generalization of the power method and of the well known Jacobi iterative scheme, which correspond to two particular choices of the preconditioner. Then we introduce the concept of weakly stochastic matrix algebra in order to define a new fast and efficient preconditioner, based on Householder unitary transformations. We discuss the relation among the new preconditioned method, the original power method and the Jacobi iterations by providing, in particular, an analysis of the convergence and a number of results on the spectral radius of the respective iteration matrices. Finally we present several numerical tests on synthetic datasets and matrices coming from real-world models. Although the proposed Householder preconditioner does not preserve the nonnegativity of the entries of the original matrix and despite we cannot provide an exhaustive convergence theorem when the coefficient matrix is not assumed symmetric, the analysis made in Section 5 and the experiments proposed in Section 6 show that the Householder preconditioner reduces significantly the number of iterations without significantly affecting the computational cost nor the memory storage. Thus, it stands as a preconditioned version of the power method, well suited for large-scale stochastic M-matrix problems with sparsity structure.

1.1. Preliminary notation. For an integer n , the linear space of square $n \times n$ real matrices is denoted by \mathbb{M}_n . The symbols O and I denote the zero and the identity matrices, respectively. A matrix is called nonnegative (resp. positive), if its entries are nonnegative (resp. positive) numbers, in symbols $A \geq O$ (resp., $A > O$); for real matrices A, B we write $A \geq B$ if $A - B \geq O$; the cone of nonnegative matrices is denoted by \mathbb{M}_n^+ , the one of nonnegative vectors by \mathbb{R}_+^n .

Given $A \in \mathbb{M}_n$ we use standard spectral notations, in particular $\sigma(A)$ denotes the spectrum of A , $\rho(A)$ denotes its spectral radius and $\lambda_i(A)$ its i -th eigenvalue, according to the reverse magnitude ordering

$$|\lambda_1(A)| \geq |\lambda_2(A)| \geq \dots \geq |\lambda_n(A)|.$$

The symbol $\mathbf{1}$ denotes the vector whose components are all ones, $\mathbf{1} = (1, \dots, 1)^T$, whereas e_i is the i -th canonical vector, for $i \in \{1, \dots, n\}$. Scalar products are always meant to be Euclidean (componentwise) products. Thus, $(A, B) = \sum_{ij} \bar{a}_{ij} b_{ij}$ for a generic pair of complex matrices A and B , and similarly for

vectors. When a matrix A satisfies the equality $A^T \mathbf{1} = \mathbf{1}$, we say that A is a weakly (column) stochastic matrix. If both A and A^T are weakly stochastic we say that A is doubly weakly stochastic. Note that a nonnegative weakly stochastic matrix is a stochastic matrix in the standard sense, that is a matrix having the set of discrete probability distributions as an invariant.

1.2. A generalization of the Pagerank linear system formulation. We say that $M \in \mathbb{M}_n$ is a (column) stochastic M-matrix if it can be decomposed as $M = I - \tau A$, with $A \geq O$, $A^T \mathbf{1} = \mathbf{1}$ and $0 < \tau < 1$. We let SK_n denote the set of such matrices, namely,

$$\text{SK}_n = \{I - \tau A \mid \tau \in (0, 1), A \geq O, A^T \mathbf{1} = \mathbf{1}\}.$$

If $S \in \mathbb{M}_n$ is any stochastic matrix having at least one full row we say that S belongs to Σ_n ,

$$\Sigma_n = \{A \in \mathbb{M}_n : A \geq O, A^T \mathbf{1} = \mathbf{1}, \max_i \min_j a_{ij} > 0\}.$$

The following theorem is a collection of results proved in [35, 39]. It shows that the two sets of matrices SK_n and Σ_n are strictly related.

Given S stochastic (nonnegative, weakly stochastic) define $\tau(S) \in \mathbb{R}_+$ and $y_S \in \mathbb{R}_+^n$ as

$$\tau(S) = 1 - \sum_{i=1}^n \min_{j=1, \dots, n} s_{ij}, \quad (y_S)_i = \min_{j=1, \dots, n} s_{ij}$$

and, for nonzero $\tau(S)$, let $A_S \in \mathbb{M}_n$ be

$$A_S = \tau(S)^{-1}(S - y_S \mathbf{1}^T).$$

THEOREM 1.1.

- (i) Let S be a stochastic matrix. The quantity $\tau(S)$ belongs to the interval $[0, 1]$ and $\tau(S) \geq |\lambda|$, for all $\lambda \in \sigma(S) \setminus \{1\}$.
- (ii) Let $S \in \Sigma_n$ and p be the ergodic distribution of S (i.e., $p \geq 0$, $p \neq 0$, $Sp = p$, $p^T \mathbf{1} = 1$). Then $\tau(S) \in [0, 1)$, $y_S \neq 0$, $(I - (S - y_S \mathbf{1}^T))p = y_S$. If moreover $\tau(S) > 0$ then $I - (S - y_S \mathbf{1}^T) = I - \tau(S)A_S$ with $\tau(S) \in (0, 1)$, $A_S \geq 0$ and A_S stochastic (by columns), that is, $I - (S - y_S \mathbf{1}^T) \in \text{SK}_n$.
- (iii) Let $M = I - \tau A \in \text{SK}_n$, $y \geq 0$ nonzero and let x be such that $Mx = y$. Define $\tilde{y} = (1 - \tau)/(y^T \mathbf{1})y$ and $\tilde{x} = (1 - \tau)/(y^T \mathbf{1})x$. Then

$$(1.1) \quad S := \tilde{y} \mathbf{1}^T + \tau A \in \Sigma_n,$$

$\tau(S) \in [0, \tau] \subset [0, 1)$, and \tilde{x} is the ergodic distribution of S (i.e. $\tilde{x} \geq 0$, $\tilde{x} \neq 0$, $S\tilde{x} = \tilde{x}$, $\tilde{x}^T \mathbf{1} = 1$).

If $S \in \Sigma_n$, then Theorem 1.1 shows that the eigenproblem $Sp = p$ can be solved by solving the linear system $(I - \tau(S)A_S)x = y_S$, and vice-versa, if $M \in \text{SK}_n$, then the solution of $Mx = y$ is a multiple of the ergodic distribution $p = Sp$ (where S is obtained from $M = I - \tau A$ through (1.1)).

It is worth noting that this generalizes to any matrix $S \in \Sigma_n$ a famous property of the Google's Pagerank index, where the particular structure of the problem allows to recast the stationary distribution problem in terms of a linear system problem [29].

Let $W = (V, E)$ be the direct graph where nodes correspond to web-pages and edges to hyperlinks between pages. The Pagerank index vector p of W is the solution of the equation

$$(1.2) \quad Gp = p,$$



where $G = \alpha T^T + (1 - \alpha)v\mathbb{1}^T$ is the Google engine web matrix, T is the row stochastic transition matrix of W , v is a real positive personalization vector such that $v^T\mathbb{1} = 1$ and $0 < \alpha < 1$. Due to the huge dimension of G , several algorithms essentially based on the power method have been proposed to compute the stationary distribution of (1.2). However the original formula by Brin and Page [29] defines the Pagerank vector p as the solution of a M-matrix linear system of the type

$$(1.3) \quad \gamma(I - \alpha T)^T p = v, \quad \gamma \in \mathbb{R}.$$

In fact, such system follows immediately from (1.2), by the particular form of G , but can be also recovered by means of Theorem 1.1, as we show now:

Since $\max_i \min_j (G)_{ij} \geq (1 - \alpha) \max_i v_i > 0$ we deduce that $G \in \Sigma_n$ and $p = (I - \tau(G)A_G)^{-1}y_G$. For the sake of simplicity, suppose that each column of T has at least one zero entry. Then $\tau(G) = 1 - \sum_i \min_j (G)_{ij} = \alpha$, $y_G = (1 - \alpha)v$ and $A_G = \alpha^{-1}(G - y_G\mathbb{1}^T) = T^T$. This shows, indeed, that p is both the solution of the eigenvector problem (1.2) and of the Pagerank linear system (1.3), with $\gamma = (1 - \alpha)^{-1}$.

1.3. The Euler-Richardson method. We assume from now on that any random walk considered is described by a stochastic matrix $S \in \Sigma_n$. We discuss a method which computes the solution of the eigenproblem $p = Sp$ by solving the associated stochastic M-matrix linear system. By virtue of Theorem 1.1 the two problems are equivalent, so for the sake of clarity and generality we always assume that a stochastic M-matrix $M \in \text{SK}_n$ and a nonnegative vector $y \geq 0$ are given, and we are interested in the solution of the equation $Mx = y$.

The preconditioned Euler-Richardson method (briefly, PER method) for the solution of $Mx = y$ is the stationary iterative scheme based on the splitting $M = P - (P - M)$ and defined by the following sequence (for example, see [43])

$$(1.4) \quad x_{k+1} = P^{-1}y + (I - P^{-1}M)x_k, \quad k = 0, 1, 2, 3, \dots,$$

where P is a suitable nonsingular preconditioner. The iteration matrix of such method is evidently $I - P^{-1}M$, thus we write

$$H(P) = I - P^{-1}M$$

to denote such matrix, underlining the dependence upon the chosen preconditioner P . Since the eigenvalues of any $M \in \text{SK}_n$ have positive real part, the standard Euler-Richardson method (ER), obtained by setting $P^{-1} = \omega I$ inside (1.4), is convergent for all $\omega \in (0, \min \frac{2\Re(\lambda_i)}{|\lambda_i|^2})$ and its rate of convergence is optimized by setting $\omega_0 = \arg \min_{\omega \in \mathbb{R}} \rho(H(\omega I))$. This is the simplest iterative method and it may not be the best choice in terms of efficiency. However, its simplicity allows its easy implementation for problems that are unstructured and have huge dimension, as for instance the Pagerank problem. In particular, as for the power method, the ER scheme requires only one real vector to store the data. Moreover, the analysis made throughout this paper shows that (1.4) can be seen as a preconditioned power method. This opens the way to a number of further investigations and improvements, as, for instance, the variety of techniques proposed to speed-up the power method for Pagerank computation can be potentially applied to (1.4) (e.g., extrapolation [6, 7, 9, 28] or structural adaptive methods [25, 26, 27]). More precisely, when $M \in \text{SK}_n$, one can show that $\omega_0 = 1$ (for example, see [40, Thm. 4.2]). It is therefore easy to realize that, if $M = I - \tau(S)A_S$ is defined as in Theorem 1.1, then the ER, with $\omega = 1$, and the power methods are very close. In particular, we show in the sequel that there exists a simple choice P_{pm} for the preconditioner P in (1.4) that gives rise exactly to the same convergent sequence as the one defined by the power method applied to the original eigenproblem

$Sp = p$. To this aim we initially devote Section 2 to define and investigate the concept of weakly stochastic matrix algebra, then we show that the power method preconditioner P_{pm} is indeed defined in terms of such algebras. In Section 4, we consider a new preconditioner chosen in a suitable weakly stochastic and low complexity Householder algebra, giving rise to a competitive method that can be implemented with linear memory storage allocation (two real vectors) and with the same order of operations per step of the power method. Besides its direct application to the iterative scheme (1.4), the analysis made in Sections 2 and 4 provides a number of interesting relations among matrix algebras, stochastic and nonnegative matrices, and in our opinion it is of self interest.

2. Low complexity matrix subspaces. Given $J_1, \dots, J_m \in \mathbb{M}_n$ linearly independent, the subspace $\mathcal{L} = \text{span}(J_1, \dots, J_m)$ is said to be of low complexity if for any $L \in \mathcal{L}$, the order of complexity required to multiply L times a vector or to solve a linear system with L as coefficient matrix, is much less than $O(n^2)$ (typically $O(n \log n)$, see examples in Section 2.2). A preconditioner for (1.4) can thus be chosen inside \mathcal{L} . A popular choice for $P \in \mathcal{L}$ is the so called optimal fit preconditioner obtained by projecting the coefficient matrix over \mathcal{L} . For any given matrix $X \in \mathbb{M}_n$, we write \mathcal{L}_X to denote its projection over \mathcal{L} . Note that, by definition of projection, one has that $\|\mathcal{L}_X - X\|_F \leq \|Y - X\|_F$, for any $Y \in \mathcal{L}$, with $\|\cdot\|_F$ being the Frobenius norm. A possible representation of \mathcal{L}_X is as follows

$$(2.5) \quad \mathcal{L}_X = \sum_{i=1}^m (B^{-1}c)_i J_i,$$

where B is the Gram matrix $b_{ij} = (J_i, J_j)$ and c is the vector $c_i = (J_i, X)$. Of course the number of arithmetic operations required to identify such \mathcal{L}_X in \mathcal{L} should be “not too large”, that is the linear system $Bx = c$ should be easily solvable. It is shown in [14, 18] that, under suitable hypotheses on \mathcal{L} , the matrix B is in \mathcal{L} itself or belongs to other low complexity classes, and thus, the projection \mathcal{L}_X can be obtained with a small amount of computations whenever \mathcal{L} is of low complexity. Typical examples of such spaces are the algebras of matrices simultaneously diagonalized by a unitary transform, henceforth briefly called *SDU spaces*. Fix any unitary matrix $U \in \mathbb{M}_n$, any such space is denoted by $\text{sd}U$ and is defined by

$$\mathcal{U} = \text{sd}U = \{U \text{diag}(\lambda)U^* \mid \lambda \in \mathbb{C}^n\}.$$

Any $\mathcal{U} = \text{sd}U$ is an n -dimensional matrix algebra, thus a commutative set of matrices closed under addition, multiplication and inversion. Moreover, the following further representation for \mathcal{U}_X holds for $\mathcal{U} = \text{sd}U$:

$$(2.6) \quad \mathcal{U}_X = U \text{diag}(U^* X U) U^*,$$

where, for a matrix M , $\text{diag}(M)$ denotes the diagonal matrix with diagonal entries m_{11}, \dots, m_{nn} . As $I \in \text{sd}U$, the linearity of the projection operator implies that $\mathcal{U}_M = I - \tau \mathcal{U}_A$, for any $M \in \text{SK}_n$. Therefore, the problem of defining a preconditioner for (1.4) reduces to the problem of identifying the projection of the nonnegative and weakly stochastic matrix A , and solving low complexity systems with $I - \tau \mathcal{U}_A$ as coefficient matrix.

Note that in many cases, if \mathcal{U}_A is a weakly stochastic matrix, then \mathcal{U}_M is invertible. Indeed if $\|A\|_2 \leq 1$, and this is true at least for all $A \geq O$ which are stochastic and normal, then, using the Cauchy-Schwarz inequality, we get

$$\rho(\mathcal{U}_A) = \max_{i=1, \dots, n} |u_i^T A u_i| \leq \max_{\|x\|_2=1} |x^T A x| \leq \max_{\|x\|_2=1} \|Ax\|_2 = \|A\|_2 \leq 1,$$

where u_i are the columns of U , defining \mathcal{U} . Thus, $\mathcal{U}_M = I - \tau \mathcal{U}_A$ is evidently invertible.



Next, Section 2.1 contains a theorem characterizing the SDU spaces \mathcal{U} such that A weakly stochastic implies \mathcal{U}_A weakly stochastic as well. Then in Section 4.1 we show that P_{pm} can be defined in terms of such spaces, and finally we introduce a new SDU space where we select a different preconditioner for (1.4).

2.1. Weakly stochastic matrix algebras. This subsection is devoted to characterize the SDU matrix algebras \mathcal{U} which preserve the weakly stochasticity of A , when projecting A on them. For a vector u such that $u^T U$ has no zero entries, define the map $L_u : \mathbb{C}^n \rightarrow \mathcal{U}$ that associates to a vector x the matrix $L_u(x)$ of \mathcal{U} such that $u^T L_u(x) = x^T$. As $u^T U$ has no zero entries, it is not difficult to see that L_u is a well defined bijection, for any SDU matrix algebra $\mathcal{U} = \text{sd}U$. However, it is worth pointing out that the class of spaces for which the operator L_u is a well defined bijection contains properly the set of SDU spaces (see [18]). A direct computation shows that the following representation of $L_u(x) \in \mathcal{U} = \text{sd}U$ holds

$$(2.7) \quad L_u(x) = U \text{diag}(U^T x) \text{diag}(U^T u)^{-1} U^*.$$

DEFINITION 2.1. If there exists a column of U which has all constant entries then we call $\mathcal{U} = \text{sd}U$ a *weakly stochastic SDU matrix algebra*.

The reason of such name is made evident by the following Theorem 2.2 which completely characterizes those SDU spaces \mathcal{U} with the property that the projection over \mathcal{U} of a weakly stochastic matrix, is still weakly stochastic.

THEOREM 2.2. Let $\mathcal{U} = \text{sd}U$ for some unitary U , and let u be any vector such that $u^T U$ has no zero entries. The following statements are equivalent.

1. There exists an index k s.t. the k^{th} column of U has constant entries.
2. $\mathbb{1}\mathbb{1}^T \in \mathcal{U}$.
3. For any vector $x \in \mathbb{C}^n$, it holds $L_u(x)\mathbb{1} = L_u(x)^T \mathbb{1} = (x^T \mathbb{1} / u^T \mathbb{1}) \mathbb{1}$.
4. For any matrix $A \in \mathbb{M}_n$, it holds $\mathcal{U}_A \mathbb{1} = \mathcal{U}_A^T \mathbb{1} = \frac{1}{n} (\mathbb{1}^T A \mathbb{1}) \mathbb{1}$.

In particular if A or A^T are weakly stochastic then \mathcal{U}_A is doubly weakly stochastic.

Proof. (1) \implies (2) Let D_i be the diagonal rank one matrix whose only nonzero entry is $(D_i)_{ii} = \mathbb{1}^T \mathbb{1}$. The rank one matrices $R_i = U D_i U^* = (\mathbb{1}^T \mathbb{1})(U e_i)(U e_i)^*$ clearly all belong to \mathcal{U} , and in particular $R_k = \mathbb{1}\mathbb{1}^T \in \mathcal{U}$. (2) \implies (3) Since L_u is a bijection and since $\mathbb{1}\mathbb{1}^T \in \mathcal{U}$, we have $L_u(\mathbb{1}) = \mathbb{1}\mathbb{1}^T / \mathbb{1}^T u$. Thus, formula (2.7) implies

$$\mathbb{1}^T \left(\frac{\mathbb{1}^T u}{\mathbb{1}^T x} \right) L_u(x) = \left(\frac{\mathbb{1}^T u}{\mathbb{1}^T x} \right) x^T L_u(\mathbb{1}) = \left(\frac{\mathbb{1}^T u}{\mathbb{1}^T x} \right) \left(\frac{x^T \mathbb{1}}{\mathbb{1}^T u} \right) \mathbb{1}^T = \mathbb{1}^T,$$

that is, $L_u(x)^T \mathbb{1} = (x^T \mathbb{1} / u^T \mathbb{1}) \mathbb{1}$, for any $x \in \mathbb{C}^n$. Now using the hypothesis $\mathbb{1}\mathbb{1}^T \in \mathcal{U}$ and the fact that matrices in \mathcal{U} commute, we have the equality $L_u(x)\mathbb{1}\mathbb{1}^T - \mathbb{1}\mathbb{1}^T L_u(x) = O$ implying that $L_u(x)\mathbb{1} = (x^T \mathbb{1} / \mathbb{1}^T u) \mathbb{1}$.

(3) \implies (4) Since $\mathcal{U}_A \in \mathcal{U}$, there exists a vector $z_A \in \mathbb{C}^n$ such that $\mathcal{U}_A = L_u(z_A)$. It is enough to show that $\mathbb{1}^T u / \mathbb{1}^T z_A = n / \mathbb{1}^T A \mathbb{1}$. Matching the representations (2.7) and $\mathcal{U}_A = U \text{diag}(U^* A U) U^*$ we get $z_A^T = u^T U \text{diag}(U^* A U) U^*$. Thus, since $U^* \mathbb{1} = \bar{\alpha} e_k$, with $|\alpha|^2 = 1/n$, it holds

$$z_A^T \mathbb{1} = \bar{\alpha} u^T U e_k (U^* A U)_{kk} = u^T \mathbb{1} (U^* A U)_{kk} = u^T \mathbb{1} (\mathbb{1}^T A \mathbb{1} / n).$$

(4) \implies (1) The eigenvectors of any matrix $L \in \mathcal{U} = \text{sd}U$ are the columns of U . The fact that $\mathbb{1}$ is an eigenvector of \mathcal{U}_A for any $A \in \mathbb{M}_n$, implies that there exists k such that $U e_k = \alpha \mathbb{1}$, with $|\alpha|^2 = 1/n$. \square

2.2. Examples. Low complexity matrix algebras have been studied extensively in relatively recent years in the context of preconditioning, displacement and optimization, for example, see [10, 14, 16, 17, 18,

[41, 42] and the references therein. Among the best known matrix algebras developed in past literature, we recognize several weakly stochastic matrix algebras. For the sake of completeness, we briefly discuss some relevant examples in the following. Other examples can be found among the Hartley-type algebras and the matrix algebras associated with trigonometric transforms, for example, see [3, 4, 5]. In particular, it is not difficult to observe that the Hartley [3] and the τ_{11} [5] algebras are weakly stochastic.

Circulants. The discrete Fourier transform is realized through the action of the Fourier matrix $F_n = 1/\sqrt{n}(\exp(2\pi\mathbf{i}(ij)/n))_{i,j=0}^{n-1}$. It is easy to check that F_n is unitary and that $F_n e_1 = \mathbf{1}/\sqrt{n}$, that is, the first column of F_n is constant. The matrix algebra $\mathcal{C} = \text{sd}(F_n)$ is usually referred to as the *circulant* algebra. If Σ is the modulo- n shift backward matrix

$$\Sigma = \begin{bmatrix} & & 1 & & \\ & & & \ddots & \\ & & & & 1 \\ 1 & & & & \end{bmatrix}$$

then $\{I, \Sigma, \Sigma^2, \dots, \Sigma^{n-1}\}$ is a basis for \mathcal{C} and this basis is made by nonnegative and mutually orthogonal matrices. It follows that the computation of the projection \mathcal{C}_A only requires additive operations among the entries of A .

Haar. The discrete Haar transform is realized through the action of the Haar matrix W_n , which can be described recursively in terms of two matrices Q_n and D_n . Let $S_1 = Q_1 = D_1 = 1$, then for $n = 2, 4, 8, \dots, 2^m$ let $S_n = \text{diag}(1/\sqrt{2}, 1, \dots, 1)$ and

$$Q_n = \begin{bmatrix} \mathbf{1}e_1^T & Q_{n/2} \\ Q_{n/2} & -\mathbf{1}e_1^T \end{bmatrix}, \quad D_n = \begin{bmatrix} D_{n/2}S_{n/2} & \mathbf{0} \\ \mathbf{0} & D_{n/2}S_{n/2} \end{bmatrix}.$$

Thus, the Haar matrix is given by

$$W_n = Q_n D_n = \begin{bmatrix} \frac{1}{\sqrt{n}}\mathbf{1}e_1^T & W_{n/2}S_{n/2} \\ W_{n/2}S_{n/2} & -\frac{1}{\sqrt{n}}\mathbf{1}e_1^T \end{bmatrix}.$$

Other equivalent definitions of this matrix can be found in the literature (e.g., [24, §1.4.3]) obtained by permuting rows and columns of W_n . Using the proposed construction, it is not difficult to check that such W_n is unitary, that its first column is $(1/\sqrt{n})\mathbf{1}$ and that its multiplication times a generic vector can be performed very cheaply. The generated algebra $\mathcal{W} = \text{sd}(W_n)$ is called Haar matrix algebra.

Eta. Consider the matrix U defined as follows:

$$\begin{cases} U_{i,1} = 1/\sqrt{n}, \\ U_{ij} = \sqrt{\frac{2}{n}} \left\{ \cos\left(\frac{(2i-1)(j-1)\pi}{n}\right) + \sin\left(\frac{(2i-1)(j-1)\pi}{n}\right) \right\}, & j = 2, \dots, \lceil \frac{n}{2} \rceil \\ U_{i, \frac{n}{2}+1} = \frac{(-1)^{i-1}}{\sqrt{n}}, & \text{if } n \text{ is even} \\ U_{ij} = \sqrt{\frac{2}{n}} \left\{ \cos\left(\frac{(2i-1)(j-1)\pi}{n}\right) + \sin\left(\frac{(2i-1)(j-1)\pi}{n}\right) \right\}, & j = \lfloor \frac{n}{2} \rfloor + 2, \dots, n \end{cases}$$

for $i = 1, \dots, n$. The matrix U is unitary and real, moreover $Ue_1 = \frac{1}{\sqrt{n}}\mathbf{1}$. As a consequence, setting $\eta = \text{sd}U$, we obtain a weakly stochastic SDU low complexity matrix algebra (see [4, 18, 36, 37] for the fast sine-cosine

transforms used in the computations involving U). For the sake of completeness let us recall the following representation [4]: $\eta = \mathcal{C}_s + J\mathcal{C}_s$ where \mathcal{C}_s is the algebra of symmetric circulant matrices and J is the reverse identity matrix, $(J)_{ij} = 1$ if $i + j = n + 1$, and $(J)_{ij} = 0$ otherwise. It follows that any matrix A in η is symmetric and persymmetric and satisfies the cross-sum rule

$$a_{i-1,j} + a_{i+1,j} = a_{i,j-1} + a_{i,j+1}, \quad i, j = 1, \dots, n$$

with border conditions $a_{0,i} = a_{1,n+1-i}$, $i = 1, \dots, n$. See also [14, 18, 41] and the references therein.

Hadamard. The Sylvester-Hadamard orthogonal matrix of order $n = 2^m$ is defined recursively by the rule $H_m = H_1 \otimes H_{m-1}$, where

$$H_1 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

and \otimes is the Kronecker product. As the first column of H_1 has all ones entries, we immediately see that the first column of H_m has constant entries as well. The associated unitary matrix is $\frac{1}{\sqrt{2^m}}H_m$ and the associated matrix algebra $\text{sd}(\frac{1}{\sqrt{2^m}}H_m)$ is therefore a weakly stochastic SDU algebra. The Hadamard matrix has relevant applications both in statistics, where it is used, for instance, to uncover the dependencies in a multivariable data set, and in error correcting codes theory. We refer to [44] for a detailed description of applications and properties of the Hadamard matrix.

3. Preserving the nonnegativity of the entries. As the original matrix A is both nonnegative and weakly stochastic, one ideally would like to have \mathcal{L}_A both nonnegative and weakly stochastic. The following theorem characterizes the subspaces \mathcal{L} which preserve the nonnegativity of a matrix A , when projecting A onto them.

DEFINITION 3.1. We say that \mathcal{L} is a *nonnegative matrix space* if $\mathcal{L}_A \geq O$ for any $A \geq O$, i.e., if the projection $A \mapsto \mathcal{L}_A$ preserves the cone of nonnegative matrices.

THEOREM 3.2. \mathcal{L} is a *nonnegative matrix space* if and only if \mathcal{L} has a basis of orthonormal nonnegative matrices.

Proof. It is straightforward to see that if \mathcal{L} has a basis of orthogonal nonnegative matrices, then \mathcal{L} is a nonnegative space. Let $\pi : \mathbb{M}_n \rightarrow \mathcal{L}$ be the projection operator. If \mathcal{L} is a nonnegative matrix space, we have that $\pi(\mathbb{M}_n^+) \subseteq \mathbb{M}_n^+$, that is the projection leaves the cone of nonnegative matrices invariant. For the sake of simplicity let us consider the vectorization operator which realizes the standard isomorphism between \mathbb{M}_n and \mathbb{R}^{n^2} . We have $\text{vec}(\mathbb{M}_n^+) \equiv \mathbb{R}_+^{n^2}$, $\Pi = \text{vec}(\pi) \in \mathbb{M}_{n^2}^+$ and $\sigma(\Pi) = \{0, 1\}$. The multiplicity of $1 = \rho(\Pi)$ is the dimension of \mathcal{L} . If $\dim(\mathcal{L}) = 1$ then the proof is trivial. If $\dim(\mathcal{L}) = k > 1$ then $\rho(\Pi)$ is not simple; thus, due to the Perron-Frobenius theorem, Π is reducible and there exists a permutation matrix $Q \in \mathbb{M}_{n^2}^+$ such that

$$Q\Pi Q^T = \Pi_1 \oplus \dots \oplus \Pi_k \oplus N,$$

where each Π_i is irreducible and N is nilpotent. Therefore, there exist k positive vectors x_1, \dots, x_k such that $\Pi_i x_i = x_i$. Let \tilde{x}_i be the embedding of x_i into \mathbb{R}^{n^2} , obtained by filling x_i with zero entries, and set $y_i = Q^T \tilde{x}_i$. Then $\Pi y_i = y_i$ and $y_i^T y_j = 0$ for $i = 1, \dots, k$ and any $j \neq i$. Finally note that, if $Y_i \in \mathbb{M}_n$ is the matrix such that $y_i = \text{vec}(Y_i)$, we have $Y_i \geq O$, $Y_i \in \mathcal{L}$ and $(Y_i, Y_j) = y_i^T y_j$ so that Y_1, \dots, Y_k is a basis for \mathcal{L} made by orthogonal nonnegative matrices. \square

It is immediate to note that the algebra of diagonal matrices $\mathcal{D} = \text{sd}I$ is a nonnegative matrix space. If we define the preconditioner for PER as the projection of M on \mathcal{D} then we have $P = \mathcal{D}_M = \text{diag}(M)$ and the

method in (1.4) coincides with the Jacobi iterative scheme. The convergence properties of such method are well known (see for instance [30, 43]). We analyze them in more details in Section 5, taking into account the structure of the matrices $M \in \text{SK}_n$. Despite its simple formulation and cheap implementation, the diagonal preconditioner does not preserve the weakly stochasticity of the original matrix. This has two drawbacks: on the one hand the use of $P = \mathcal{D}_M$ in (1.4) has a less clear relation to the power method, since the power method preconditioner P_{pm} is weakly stochastic, as we will show in the next section; on the other hand, the numerical implementations in Section 6 show that the property of being weakly stochastic ensures faster convergence.

To our knowledge, if we exclude the multilevel generalizations, the only low complexity matrix algebra satisfying the hypothesis of both Theorems 2.2 and 3.2 is the circulant algebra. The use of a circulant preconditioner in this context has been analyzed in details in [40]. Although the analysis in [40] shows a reduction of the number of iterations with respect the classical power method, the use of the circulant algebra requires one Fourier transform and two complex vectors to be stored per each step. In Section 4.3 we will introduce a new preconditioner based on a weakly stochastic matrix algebra diagonalized by a suitable Householder transform. We propose a convergence analysis of PER with such novel preconditioner under the assumption that A is symmetric. Although an exhaustive convergence analysis for the more realistic case where A is generic is still missing, we point out that the low memory storage and the linear order of operations per step required by the new technique make it effectively applicable also when the dimension of the problem is huge. This is further highlighted by the numerical tests proposed in Section 6.

4. The choice of the preconditioner. In this section, we show that the power method preconditioner P_{pm} for the PER scheme is a matrix belonging to a class of weakly stochastic SDU matrix algebras. Then we develop a new matrix algebra defined in terms of Householder unitary transformations, leading to a new cheap preconditioner for PER. In the subsequent sections we analyze the convergence of PER method and we provide numerical evidences of the advantages obtained by using the new Householder preconditioner.

4.1. The power method embedded into a PER iterative scheme. Recall that the unpreconditioned Euler-Richardson method is obtained by choosing $P = I$. Given $M = I - \tau A \in \text{SK}_n$, the power method preconditioner P_{pm} for the solution of $Mx = y$, instead, is the following rank-one correction of the identity matrix

$$P_{\text{pm}} = I - \frac{\tau}{n} \mathbb{1}\mathbb{1}^T.$$

We observe that P_{pm} belongs to any weakly stochastic SDU matrix algebra. In other words such preconditioner belongs to the intersection $\cap \{\text{sd}U \mid U \text{ has a constant column}\}$. Indeed, let U be any unitary matrix such that Ue_k has constant entries, and consider the diagonal matrix $D = \text{diag}(1, \dots, 1, 1 - \tau, 1, \dots, 1)$, where $1 - \tau$ lies in the k -th diagonal position, then $P_{\text{pm}} = UDU^*$. It is worth noting that this is somehow analogous to the property shown in point 2 of Theorem 2.2. Indeed observe that, as for the projection $\mathcal{U}_M = I - \tau \mathcal{L}_A$, the matrix P_{pm} has the structure $P_{\text{pm}} = I - \tau E$, where E is the weakly stochastic matrix $E = \mathbb{1}\mathbb{1}^T/n$ which indeed belongs to any weakly stochastic SDU algebra. Note moreover that

$$(4.8) \quad P_{\text{pm}}^{-1} = I + \left(\frac{\tau}{1 - \tau} \right) \frac{\mathbb{1}\mathbb{1}^T}{n}.$$

In view of Theorem 1.1, we can show the connection between the PER method applied to $M \in \text{SK}_n$ and the power method for the ergodic distribution of a Markov chain described by $S \in \Sigma_n$.

THEOREM 4.1. *Given $S \in \Sigma_n$ let $\tau(S)$, A_S and y_S be defined as before Theorem 1.1. When the preconditioner is $P = P_{\text{pm}}$, the PER method for the solution of $Mx = (I - \tau(S)A_S)x = y_S$ coincides with the power method applied to S .*

Proof. Let $\{x_k\}$ be the sequence defined by the PER method (1.4). By Theorem 1.1 and equation (4.8), it follows that $y_S^T \mathbf{1} = 1 - \tau(S)$ and $P_{\text{pm}}^{-1} \mathbf{1} = (1 - \tau(S))^{-1} \mathbf{1}$. As a consequence we observe that $\mathbf{1}^T x_k = 1$ implies

$$\mathbf{1}^T x_{k+1} = \mathbf{1}^T P_{\text{pm}}^{-1} y_S + \mathbf{1}^T (I - P_{\text{pm}}^{-1} M) x_k = 1.$$

Therefore, we can assume that the entries of the initial x_0 sum up to 1, and that $\mathbf{1}^T x_k = 1$ for all $k \geq 0$. We have

$$\begin{aligned} x_{k+1} &= P_{\text{pm}}^{-1} y_S + (I - P_{\text{pm}}^{-1} M) x_k \\ &= y_S + \frac{\tau(S)}{n} \mathbf{1} + x_k - M x_k - \frac{\mathbf{1}^T M x_k}{1 - \tau(S)} \frac{\tau(S)}{n} \mathbf{1} \\ &= y_S + \tau(S) A_S x_k \end{aligned}$$

and, by Theorem 1.1, $S = \tau(S)A_S + y_S \mathbf{1}^T$, therefore $x_{k+1} = Sx_k$, and the proof is complete. \square

Let $\lambda_i(X)$ for $i = 1, 2, \dots, n$ be the eigenvalues of the matrix X ordered as $|\lambda_1(X)| \geq |\lambda_2(X)| \geq \dots \geq |\lambda_n(X)|$. As $S \in \Sigma_n$, by Theorem 1.1, we have $|\lambda_2(S)| < 1$ and, by the well known behaviour of the power method, we have that x_k converges to the solution of $Mx = y_S$ as $O(|\lambda_2(S)|^k)$. However, a different bound can be observed by using the equivalence shown in Theorem 4.1 as indeed we will show in Section 5 that x_k converges to x as $O(\tau(S)^k |\lambda_2(A_S)|^k)$. Note that $S \in \Sigma_n$ implies, by Theorem 1.1, that $\tau(S) < 1$ and A_S is stochastic too, thus $|\lambda_2(S)| < 1$ and $\tau(S) |\lambda_2(A_S)| < 1$. However in several cases (for instance if A_S is primitive, i.e., $A_S^k > O$ for some integer power $k > 0$), one has $|\lambda_2(A_S)| < 1$, thus $\tau(S) |\lambda_2(A_S)| < \tau(S)$ whereas $|\lambda_2(S)| \leq \tau(S)$, thus suggesting that $\tau(S) |\lambda_2(A_S)|$ could provide a better bound on the convergence rate.

4.2. The Householder weakly stochastic matrix algebra. Let \mathcal{U} be any weakly stochastic SDU algebra. The power method is obtained by applying PER and choosing P inside \mathcal{U} as the matrix with the following eigenvalues: $1 - \tau$, with multiplicity one, and 1 with multiplicity $n - 1$. To improve the performances of the power method, we define a new preconditioner by replacing the eigenvalues 1 with the spectrum of the projection \mathcal{U}_M of M onto \mathcal{U} . Note that the eigenvalue $1 - \tau$ is an invariant, that is $1 - \tau \in \sigma(\mathcal{U}_M)$ for any weakly stochastic SDU algebra \mathcal{U} . In fact, as $A^T \mathbf{1} = \mathbf{1}$ implies $\mathcal{U}_A^T \mathbf{1} = \mathbf{1}$, we have $\mathcal{U}_M^T \mathbf{1} = (I - \tau \mathcal{U}_A)^T \mathbf{1} = (1 - \tau) \mathbf{1}$. Also note that \mathcal{U}_M minimizes the distance $\|X - M\|_F$ among the matrices $X \in \mathcal{U}$, with $\|X\|_F = \sqrt{\langle X, X \rangle}$ being the Frobenius norm. Therefore, $\|\mathcal{U}_M - M\|_F \leq \|P_{\text{pm}} - M\|_F$ for any weakly stochastic SDU space \mathcal{U} , and the inequality is strict up to trivial cases. This motivates the choice $P = \mathcal{U}_M$, in place of the classical $P = P_{\text{pm}}$, to improve the performances of the method.

The example spaces shown in Section 2.2 are defined in terms of fast transformations U whose space and time complexities are $O(n \log n)$. In order to keep the complexity of the PER iterations as low as possible, we define a weakly stochastic SDU algebra diagonalized by a Householder transformation. As we prove in the next section this allows us to keep the time and space complexity per step linear in n . It is worth mentioning that, due to their linear computational complexity, matrix algebras diagonalized by Householder unitary transforms have been already involved in a number of applications. In particular they have been recently used to define competitive iterative optimization algorithms, whose space and time per step complexity is

$O(n)$, c.f. [11, 19]. Let us introduce a linear space \mathcal{H} of the form

$$\mathcal{H} = \{H(w)\text{diag}(z)H(w) \mid z \in \mathbb{C}^n\}, \quad H(w) = I - 2ww^*, \quad \|w\| = 1,$$

where $H(w)$ is a Householder unitary matrix such that $H(w)e_k = (e^{i\theta}/\sqrt{n})\mathbf{1}$, for some $\theta \in \mathbb{R}$. We shall observe that all the Householder matrices of this kind are of the form $H(w^\pm)$, where w^+ and w^- , are two suitable vectors in \mathbb{R}^n . We firstly look at the k -th column of $H(w)$, and we get $(I - 2ww^*)e_k = e_k - 2w(w^*e_k) = (e^{i\theta}/\sqrt{n})\mathbf{1}$. Therefore,

$$(4.9) \quad 2\overline{w}_k w = e_k - \frac{e^{i\theta}}{\sqrt{n}}\mathbf{1}.$$

The k -th component of equation (4.9) implies $2\overline{w}_k w_k = 1 - (e^{i\theta}/\sqrt{n})$, so that $w_k \neq 0$, $\theta \in \{0, \pi\}$, and thus, $|w_k|^2 = (1 \pm 1/\sqrt{n})/2$. As a consequence the k -th entry of w is given by either of the two following formulas, corresponding to $\theta = 0$ and $\theta = \pi$, respectively:

$$w_k = \left(\frac{\sqrt{n}-1}{2\sqrt{n}}\right)^{1/2} e^{i\phi}, \quad \text{or} \quad w_k = \left(\frac{\sqrt{n}+1}{2\sqrt{n}}\right)^{1/2} e^{i\phi}, \quad \phi \in \mathbb{R}.$$

Writing now the j -th component of (4.9) for $j \neq k$ we obtain an analogous formula also for the other entries of w :

$$w_j = -\frac{e^{i\phi}}{\sqrt{2}\sqrt[4]{n}\sqrt{\sqrt{n}-1}}, \quad \text{or} \quad w_j = \frac{e^{i\phi}}{\sqrt{2}\sqrt[4]{n}\sqrt{\sqrt{n}+1}}, \quad \phi \in \mathbb{R}.$$

The previous relations can be written in compact form, showing that any vector w such that $H(w)$ defines a weakly stochastic algebra is either $w = w_\phi^+$ or $w = w_\phi^-$, where

$$w_\phi^- = e^{i\phi} \cdot \beta_n^-(\sqrt{n}e_k + \mathbf{1}), \quad w_\phi^+ = e^{i\phi} \cdot \beta_n^+(\sqrt{n}e_k - \mathbf{1}), \quad \beta_n^\pm = \frac{1}{\sqrt{2}\sqrt[4]{n}\sqrt{\sqrt{n} \mp 1}}.$$

This finally shows an explicit formula for all the possible weakly stochastic Householder algebras. Note indeed that the Householder matrices $H(w_\phi^+)$ and $H(w_\phi^-)$ do not depend on $\phi \in \mathbb{R}$ therefore, setting $w^\pm = w_0^\pm$, we see that $H(w^+)$ and $H(w^-)$ are the only two Householder matrices which define a weakly stochastic SDU algebra. They are both real unitary matrices and such that $H(w^\pm)e_k = (\pm 1/\sqrt{n})\mathbf{1}$.

4.3. The Householder PER method.

We use the notation

$$\chi(B) = \text{computational cost of the product } B \times \text{vector}.$$

It is not difficult to check that, when $P = \mathcal{U}_M$ and $\mathcal{U} = \text{sd}U$, the overall computational cost of the PER method for the solution of $Mx = y$, $M = I - \tau A \in \text{SK}_n$, is

$$(4.10) \quad \chi(A) + \chi(U) + \chi(U^*) + O(n)$$

for each step, plus a preprocessing phase which is required essentially for the computation of the eigenvalues $\lambda_1, \dots, \lambda_n$ of \mathcal{U}_A , and whose computational complexity highly depends on the chosen U as, by (2.6), $\lambda_i = (U^*AU)_{ii}$.

The Householder PER method (HPER in short) is obtained by projecting M over one of the two Householder SDU algebras introduced above. As we can freely choose either w^+ or w^- , in what follows we

set $w = w^+$ and assume for simplicity that the constant column of $H(w)$ is the first one (i.e., $k = 1$ in the construction of Section 4.2). Then we let $\mathcal{H} = \text{sd}H(w)$. Let us briefly analyze the computational cost of HPER.

Set $H(w) = I - 2ww^T$, where $w = \beta_n(\sqrt{n}e_1 - \mathbb{1})$ and $\beta_n^2 = \frac{1}{2} \frac{1}{\sqrt{n}(\sqrt{n}-1)}$. We immediately see that $\chi(H(w)) = O(n)$. Therefore, for this choice, even if A is a sparse or a strongly structured matrix, the complexity per step of HPER is dominated by $\chi(A)$ as the estimate (4.10) becomes $\chi(A) + O(n)$ when $U = H(w)$. Note that this is the same complexity required by the standard power method iterations.

Note that a preprocessing phase is required for the computation of the diagonal entries of $H(w)AH(w)$ (that is, the eigenvalues of \mathcal{H}_A), as well as to compute Aw , $A^T w$ and $H(w)y = y + 2\beta_n(n^{-1/2} - y_1)w$. Observe that

$$(4.11) \quad H(w)AH(w) = A - 2(Aww^T + ww^T A - 2\gamma_n ww^T),$$

where

$$(4.12) \quad \gamma_n = w^T A w = n\beta_n^2 \left((A)_{11} - \frac{1}{\sqrt{n}} - \frac{1}{\sqrt{n}}(A\mathbb{1})_1 + 1 \right).$$

Therefore, from (4.11), we obtain the equalities

$$\text{diag}(H(w)AH(w))_{ii} = (A)_{ii} - 2w_i \left((Aw)_i + (w^T A)_i - 2\gamma_n w_i \right),$$

$i = 1, \dots, n$, which, together with (4.12), show that $\chi(A)$ operations are sufficient to compute the diagonal entries of $H(w)AH(w)$. We conclude that the overall cost of the preprocessing phase is $\chi(A) + \chi(A^T) + O(n)$. Let us point out that even in the worst case, when A is a general, non structured and dense matrix, by the particular form of w , $O(n^2)$ additive operations and $O(n)$ multiplicative operations are sufficient to compute Aw and $w^T A$.

5. Convergence analysis. In this section, we analyze the convergence of the preconditioned Euler-Richardson method applied to the linear system $Mx = y$ when $M \in \text{SK}_n$ and the preconditioner P is the optimal fit of M onto a matrix algebra $\text{sd}U$. First of all we state the following simple but somewhat general theorem. We shortly outline a possible proof.

THEOREM 5.1. *Let $M \in \text{SK}_n$ and \mathcal{L} be a subspace of \mathbb{M}_n such that $I \in \mathcal{L}$ and $A \geq \mathcal{L}_A \geq O$, then $\rho(H(\mathcal{L}_M)) < 1$. That is the PER scheme (1.4) with $P = \mathcal{L}_M$ converges.*

Proof. The linearity of the projection and the fact that $I \in \mathcal{L}$ imply that $\mathcal{L}_M = I - \tau\mathcal{L}_A$. Since $A \geq \mathcal{L}_A \geq O$, the Perron-Frobenius theorem implies $\rho(\mathcal{L}_A) \leq \rho(A) = 1$. Thus, \mathcal{L}_M is an M-matrix as well, and $\mathcal{L}_M^{-1} \geq O$. Moreover clearly $\mathcal{L}_M - M = \tau(A - \mathcal{L}_A) \geq O$. We have as a consequence that both $H(\mathcal{L}_M)$ and $(I - H(\mathcal{L}_M))^{-1}$ are nonnegative matrices. Indeed $H(\mathcal{L}_M) = \mathcal{L}_M^{-1}(\mathcal{L}_M - M) \geq O$ and $(I - H(\mathcal{L}_M))^{-1} = M^{-1}(\mathcal{L}_M - M + M) = M^{-1}(\mathcal{L}_M - M) + I \geq O$. Let $\rho = \rho(H(\mathcal{L}_M))$ and let $z \geq 0$, $z \neq 0$ be such that $H(\mathcal{L}_M)z = \rho z$. We have that $y = (I - H(\mathcal{L}_M))^{-1}z$ is nonzero and nonnegative, and $z = (1 - \rho)(I - H(\mathcal{L}_M))^{-1}z = (1 - \rho)y$. But z is nonzero and nonnegative, that is $(1 - \rho) > 0$. \square

It is worth noting that Theorem 5.1 is related with the concept of regular splitting of M-matrices, for example, see [43]. In fact, under the hypothesis of Theorem 5.1, letting $N = \mathcal{L}_M - M$, we observe that $M = \mathcal{L}_M - N$ is a regular splitting of M , that is $N \geq 0$ and $\mathcal{L}_M^{-1} \geq 0$.

COROLLARY 5.2. *Let $M = I - \tau A \in \text{SK}_n$ and let $\{J_1, \dots, J_m\}$ be a set of nonnegative mutually orthogonal matrices such that*

1. $\#\{\text{nonzero entries of } J_k\} = 1$ for any $k \in \{1, \dots, m\}$, and
2. $I \in \mathcal{L} = \text{span}\{J_1, \dots, J_m\}$.

Then \mathcal{L}_M is invertible and the PER method with $P = \mathcal{L}_M$ is convergent.

Proof. To prove this corollary we simply show that the hypothesis of the previous theorem are all satisfied. First of all, since J_1, \dots, J_m are orthogonal and nonnegative, \mathcal{L} satisfies the hypothesis of Theorem 3.2 by construction, thus $\mathcal{L}_A \geq O$.

Now, for any $k \in \{1, \dots, m\}$, let $(J_k)_{i_k, j_k}$ be the unique nonzero element of J_k . We can obviously assume $(J_k)_{i_k, j_k} = 1$ without losing generality. Then, in the notation of (2.5), we have $B = I$, $c_k = a_{i_k, j_k}$ and $\mathcal{L}_A = \sum_k a_{i_k, j_k} J_k$, implying that for any $i \in \{1, \dots, n\}$ it holds

$$\sum_{j=1}^n |(\mathcal{L}_A)_{ij}| \leq \sum_{j=1}^n \left| \left(\sum_{k=1}^m a_{i_k, j_k} J_k \right)_{ij} \right| = \sum_{j, k: (i_k, j_k) = (i, j)} |a_{i_k, j_k}| \leq \sum_{j=1}^n |a_{i, j}|,$$

and hence,

$$\rho(\mathcal{L}_A) \leq \|\mathcal{L}_A\|_\infty = \max_i \sum_{j=1}^n |(\mathcal{L}_A)_{ij}| \leq \max_i \sum_{j=1}^n |a_{i, j}| \leq \|A\|_\infty = 1 = \rho(A).$$

Therefore, $\rho(\mathcal{L}_A) \leq \rho(A)$. It follows that $\mathcal{L}_M = I - \tau \mathcal{L}_A$ is invertible and $\mathcal{L}_M^{-1} \geq O$. Finally, note that

$$(A - \mathcal{L}_A)_{ij} = \begin{cases} a_{ij} & ij \notin \{i_1 j_1, \dots, i_m j_m\} \\ 0 & \text{otherwise} \end{cases}$$

which implies $A \geq \mathcal{L}_A$. The thesis follows. \square

The set $\{e_1 e_1^T, e_2 e_2^T, \dots, e_n e_n^T\}$ is a simple example of nonnegative matrices satisfying the hypothesis of Corollary 5.2. Their linear span is the algebra of diagonal matrices $\mathcal{D} = \text{sd}I$, the PER method applied with $P = \mathcal{D}_M$ coincides with the classical Jacobi method and it is well known to be convergent (e.g., see [43, §3]). Nonetheless the next Theorem 5.3 shows that a more precise control on the rate of convergence can be achieved. Recall that the classical unpreconditioned ER scheme is obtained for $P = I$ and one has $\rho(H(I)) \leq \tau$.

THEOREM 5.3. *Let $\mathcal{D} = \text{sd}I$ be the algebra of diagonal matrices and let $M = I - \tau A \in \text{SK}_n$. If $\varepsilon = \arg \max \left\{ \lambda \geq 0 \mid \min_i a_{ii} \geq \frac{1-\tau^\lambda}{1-\tau^{1+\lambda}} \right\}$, then $\rho(H(\mathcal{D}_M)) \leq \tau^{1+\varepsilon}$. In particular $\rho(H(\mathcal{D}_M)) \leq \tau$ for any $M \in \text{SK}_n$ and if $\min_i a_{ii} \geq (1 + \tau)^{-1}$ then $\rho(H(\mathcal{D}_M)) \leq \tau^2$.*

Proof. By exploiting the entries of $H = I - \mathcal{D}_M^{-1} M$ we have

$$(H)_{ij} = \begin{cases} \frac{\tau a_{ij}}{1-\tau a_{ii}} & \text{if } i \neq j, \\ 0 & \text{if } i = j. \end{cases}$$

By Gershgorin localization theorem, the eigenvalues of H are contained inside the ball in \mathbb{C} centered over the origin and with radius $R = \max_i \sum_{j \neq i} (H)_{ij}$. By using the identity $\sum_j a_{ij} = 1$ and by observing that $a_{ii} \geq \frac{1-\tau^\varepsilon}{1-\tau^{1+\varepsilon}}$ if and only if $\frac{1-a_{ii}}{1-\tau a_{ii}} \leq \tau^\varepsilon$, we get $\sum_{j \neq i} (H)_{ij} = \tau \frac{1-a_{ii}}{1-\tau a_{ii}} \leq \tau^{1+\varepsilon}$, thus $\rho(H) \leq R \leq \tau^{1+\varepsilon}$. \square

Note that the Theorem 5.3 shows that the larger the diagonal entries of A are, the smaller $\rho(H(\mathcal{D}_M))$ can be.

For the sake of completeness let us point out that other examples of a set of matrices satisfying the hypotheses of Corollary 5.2 have been considered in the literature. For example the authors of [15] define the set of matrix spaces (and matrix algebras) $\{U\Delta U^* : \Delta \in \mathcal{M}(E)\}$, where U is a unitary matrix, $E \in \{0, 1\}^{n \times n}$ is a non-degenerate mask matrix, $\mathcal{M}(E) = \{E \circ A \text{ s.t. } A \in \mathbb{M}_n\}$ and \circ is the Hadamard entry-wise product. It is not difficult to observe that any space $\mathcal{L} = \mathcal{M}(E)$ indeed admits a basis of matrices $\{J_1, \dots, J_m\}$ satisfying the hypothesis of the corollary. Of course, for such space \mathcal{L} , the matrix \mathcal{L}_M is $\mathcal{L}_M = E \circ M$.

5.1. Projecting over a weakly stochastic algebra. The results presented in this section give a further and more detailed intuition on why, when $M \in \text{SK}_n$, a preconditioner for (1.4) based on weakly stochastic matrix algebras behaves well. We assume for the remaining part of this section that any stochastic matrix A , defining the given stochastic M-matrix $M = I - \tau A \in \text{SK}_n$, has a simple dominant eigenvalue $\rho(A) = 1$.

Let $\mathcal{U} = \text{sd}U$ be a weakly stochastic matrix algebra. For a matrix $M = I - \tau A \in \text{SK}_n$ we have $\mathcal{U}_M = I - \tau \mathcal{U}_A = I - \tau U \text{diag}(z_A) U^*$ (which we assume invertible) and

$$H(\mathcal{U}_M) = I - (I - \tau U \text{diag}(z_A) U^*)^{-1} (I - \tau A),$$

where z_A is the vector whose entries are the eigenvalues of \mathcal{U}_A , ordered as usual. Recall that the components of z_A are the diagonal entries of U^*AU .

We claim that, for any such algebra \mathcal{U} , the spectrum of the iteration matrix $H(\mathcal{U}_M)$ only depends on the eigenvalues $\lambda_i(A)$ and $\lambda_i(\mathcal{U}_A)$, for $i \neq 1$. In other words when the preconditioner is chosen projecting M over a weakly stochastic algebra, the leading eigenvalues $\lambda_1(A) = \lambda_1(\mathcal{U}_A) = 1$ of A and its projection, are not involved in the analysis of the convergence.

To this end let us observe that, since \mathcal{U} is weakly stochastic, there exists an index $k \in \{1, \dots, n\}$ such that Ue_k has constant elements, that is $Ue_k = \alpha \mathbf{1}$ for some $\alpha \in \mathbb{C}$ such that $n\alpha\bar{\alpha} = 1$. Therefore, the k -th entry of z_A is

$$(z_A)_k = (U^*AU)_{kk} = e_k^T U^*AUe_k = \bar{\alpha} \mathbf{1}^T AUe_k = \bar{\alpha} \mathbf{1}^T Ue_k = \bar{\alpha} \alpha n = 1.$$

Observe analogously that $e_k^T U^*(I - \tau A)U = (1 - \tau)e_k^T$, that is, the k -th entry of the k -th row of U^*MU is $1 - \tau$ and the remaining components are all zeros. The same holds for $U^*\mathcal{U}_M U$. It follows that the two matrices U^*MU and $U^*\mathcal{U}_M U$ have the same block structure, which we represent here when $k = 1$ for ease of notation:

$$(5.13) \quad U^*MU = \begin{pmatrix} 1 - \tau & 0^T \\ f & I - \tau V^*AV \end{pmatrix}, \quad U^*\mathcal{U}_M U = \begin{pmatrix} 1 - \tau & 0^T \\ 0 & I - \tau \text{diag}(V^*AV) \end{pmatrix},$$

where V is the partial isometry given by the last $n - 1$ columns of U and f is a suitable $n - 1$ vector. Of course, also $U^*H(\mathcal{U}_M)U = I - (U^*\mathcal{U}_M U)^{-1}(U^*MU)$ has the same structure.

Now consider the matrix $A_2 = A - q\mathbf{1}^T/\sqrt{n}$, where q is any vector such that $q \geq 0$, $q^T \mathbf{1} = \sqrt{n}$. Also, let $A = XJX^{-1}$ be the Jordan decomposition of A . As $\rho(A) = 1$ is a simple eigenvalue of A , the k -th row of X^{-1} is constant. That is $e_k^T X^{-1} = \mathbf{1}^T/\sqrt{n}$ and $Je_k = J^T e_k = e_k$. Note moreover that $e_k^T X^{-1}q = 1$. Then

$$(5.14) \quad X^{-1}A_2X = X^{-1}(A - qe_k^T X^{-1})X = J - X^{-1}qe_k^T = \begin{pmatrix} 1 & 0^T \\ 0 & \tilde{J} \end{pmatrix} - \begin{pmatrix} 1 & 0^T \\ \tilde{q} & O \end{pmatrix},$$

where \tilde{J} is the Jordan form of A except for the 1×1 block associated with $\rho(A)$, \tilde{q} is the $n - 1$ vector made by the entries of $X^{-1}q$ except for the k -th one, and the right most block representation has been shown for the case where $k = 1$, for notational convenience.

We deduce that, for any $q \geq 0$ with $q^T \mathbb{1} = \sqrt{n}$, we have $\sigma(A_2) = \sigma(X^{-1}A_2X) = \sigma(\tilde{J}) \cup \{0\} = \{\sigma(A) \setminus \{1\}\} \cup \{0\}$. Moreover note that this shows that the eigenvalues of A_2 and V^*AV coincide except for the 0 eigenvalue. In fact U^*MU and $X^{-1}MX$ are similar and thus, comparing (5.13) and (5.14), the blocks $I - \tau V^*AV$ and $I - \tau \tilde{J}$ have same eigenvalues. The same holds for the eigenvalues of \mathcal{U}_{A_2} and $\text{diag}(V^*AV)$. Therefore, for any weakly stochastic algebra \mathcal{U} , the spectrum of the iteration matrix $H(\mathcal{U}_M)$ of the PER method (1.4), can be decomposed as

$$\sigma(H(\mathcal{U}_M)) = \sigma(I - (I - \tau \mathcal{U}_{A_2})^{-1}(I - \tau A_2))$$

which finally shows our claim.

It is worth noting that the observations did so far apply to the choice $P = P_{\text{pm}}$. Precisely, from (4.8) we get

$$H(P_{\text{pm}}) = I - P_{\text{pm}}^{-1}M = \tau(A - \mathbb{1}\mathbb{1}^T/n)$$

and therefore, choosing $q = \mathbb{1}/\sqrt{n}$,

$$\sigma(H(P_{\text{pm}})) = \sigma(\tau A - \frac{\tau}{n} \mathbb{1}\mathbb{1}^T) = \sigma(\tau X^{-1}AX - \frac{\tau}{n} X^{-1} \mathbb{1}\mathbb{1}^T X) = \sigma(\tau A_2).$$

By using Theorems 1.1 and 4.1, we deduce a new upper-bound on the convergence rate of the power method applied to $S \in \Sigma_n$. Namely, we have that the sequence Sx_k converges to the ergodic distribution of S with a rate of convergence bounded by $O(\tau(S)|\lambda_2(A_S)|)$, where $\tau(S)$ and A_S are defined in terms of S as in Theorem 1.1.

As a matter of fact, when the preconditioner in \mathcal{U} is not P_{pm} , but instead is chosen in \mathcal{U} as the matrix with smaller Euclidean distance from M , we cannot provide a theoretical control on the eigenvalues of $I - (I - \tau \mathcal{U}_{A_2})^{-1}(I - \tau A_2)$. However, both intuition and numerical tests shown in Section 6 suggest that the spectral radius of $H(\mathcal{U}_M)$ is significantly smaller than the one of $H(P_{\text{pm}})$.

For the sake of completeness, we observe that, when A is stochastic nonnegative and symmetric, further results hold as stated in the following Theorem 5.4, where the eigenvalues of W (symmetric) are ordered as $\lambda_1(W) \geq \dots \geq \lambda_n(W)$.

THEOREM 5.4. *Let $M = I - \tau A \in \text{SK}_n$ be such that A is symmetric and $\rho(A)$ is simple. Let $\mathcal{U} = \text{sd}U$ be a weakly stochastic matrix algebra. Then*

$$\rho(H(\mathcal{U}_M)) \leq \tau \max \left\{ \frac{\lambda_2(\mathcal{U}_A) - \lambda_n(A)}{1 - \tau \lambda_2(\mathcal{U}_A)}, \frac{\lambda_2(A) - \lambda_n(\mathcal{U}_A)}{1 - \tau \lambda_n(\mathcal{U}_A)} \right\}.$$

Proof. To lighten the notation, we denote with M_2 the matrix $I - \tau A_2$, where A_2 is the matrix such that $A = \mathbb{1}\mathbb{1}^T/n + A_2$. Note that the symmetry of A implies that both M and \mathcal{U}_M are positive definite matrices. In fact, M is clearly real symmetric and $\sigma(M) \in \mathbb{R}_+$, whereas \mathcal{U}_M has the form $\mathcal{U}_M = U \text{diag}(U^*MU)U^*$. Therefore, the eigenvalues of \mathcal{U}_M are inside the convex hull of $\sigma(M)$, so they are real and positive, implying that \mathcal{U}_M is positive definite. A known consequence of the Weyl's inequalities states that, for any two positive definite matrices X and Y , the following inequalities hold (see for instance [2])

$$\lambda_n(X)\lambda_n(Y) \leq \lambda_n(XY) \leq \lambda_1(XY) \leq \lambda_1(X)\lambda_1(Y).$$

Collecting such inequalities, the considerations we did shortly above the statement of this theorem, and the fact that $\lambda_i(\mathcal{U}_M^{-1}) = \lambda_i(\mathcal{U}_M)^{-1}$ for any $i \in \{1, \dots, n\}$, we get

$$\begin{aligned} \rho(H(\mathcal{U}_M)) &= \max \left\{ \lambda_1(\mathcal{U}_{M_2}^{-1} M_2) - 1, 1 - \lambda_n(\mathcal{U}_{M_2}^{-1} M_2) \right\} \\ &\leq \max \left\{ \frac{\lambda_2(M)}{\lambda_n(\mathcal{U}_M)} - 1, 1 - \frac{\lambda_n(M)}{\lambda_2(\mathcal{U}_M)} \right\} \\ &= \tau \max \left\{ \frac{\lambda_2(\mathcal{U}_A) - \lambda_n(A)}{1 - \tau \lambda_2(\mathcal{U}_A)}, \frac{\lambda_2(A) - \lambda_n(\mathcal{U}_A)}{1 - \tau \lambda_n(\mathcal{U}_A)} \right\} \end{aligned}$$

and the thesis follows. □

In particular, under the same hypotheses of Theorem 5.4, we have

$$\rho(H(\mathcal{U}_M)) < \frac{2\tau}{1 - \tau},$$

and hence, if τ is small enough, precisely $\tau \leq 1/3$, then PER with $P = \mathcal{U}_M$ and A symmetric, converges for any choice of the weakly stochastic algebra \mathcal{U} .

6. Numerical comparisons. In this final section we present a number of numerical tests comparing the behaviour of three methods on several synthetic and real-world datasets. The linear system solved is $Mx = (I - \tau A)x = y$ where y is a random vector with entries in $[0, 1]$, $\tau = 0.9$ and A is a stochastic matrix defined as follows. We consider X , the adjacency matrix of the dataset, then we normalize it into the associated transition matrix $T = D^{-1}X$, with D being the diagonal matrix $d_{ii} = e_i^T X \mathbf{1}$. In order to force a lower bound on the diagonal entries, we introduce a further parameter $0 < \beta < 1$, and finally define the matrix A as the convex combination $A = \beta I + (1 - \beta)T^T$. The standard Pagerank random walk is retrieved for $\beta = 0$. The methods are defined by different choices of the preconditioner P in (1.4):

HPER. The Euler-Richardson method preconditioned via the optimal fit $P = \mathcal{H}_M$, where $\mathcal{H} = \text{sd}H(w)$ is the Householder weakly stochastic matrix algebra discussed in Section 4.3. This method consists of a preprocessing phase in which the quantities β_n , γ_n , Aw , $H(w)y$ and $\text{diag}(H(w)AH(w))$ must be computed. The overall cost of this initial computation is $O(\chi(A) + \chi(A^T)) + O(n)$. Then each step of the method is performed by the recursive computation of $x_{k+1} = \mathcal{H}_M^{-1}y + (I - \mathcal{H}_M^{-1}M)x_k$, and thus requires $O(\chi(A)) + O(n)$ operations.

Jacobi. The Euler-Richardson method preconditioned with the diagonal optimal fit matrix $P = \mathcal{D}_M = I - \tau \text{diag}(A)$. The rate of convergence is given here by Theorem 5.3. Note that \mathcal{D} can be seen as the span of n orthogonal nonnegative rank-one matrices, and thus, accordingly with Theorem 3.2, both \mathcal{D}_A and \mathcal{D}_M^{-1} maintain the nonnegativity of the entries. On the other hand this choice for P does not ensure any stochasticity property of P^{-1} , in the general case. We recall that this method coincides with the standard Jacobi iterations.

Power method. The Euler-Richardson method preconditioned with the power method matrix P_{pm} defined as the following rank-one correction of the identity $P_{\text{pm}} = I - \frac{\tau}{n} \mathbf{1}\mathbf{1}^T$. As discussed in Section 4.1, this method coincides with the power method applied to the stochastic eigenproblem $S p = p$, where S is obtained from A as discussed in Theorem 1.1.

It is worth mentioning that both the Jacobi and the power methods can count on a convergence theorem with an explicit upper bound on the convergence rate. The spectral radius of the iteration matrix for the

Random matrix of order $n = 10^7$			
β	HPER	Jacobi	power method
0.1	11	104	37
0.2	8	99	32
0.5	6	60	54
0.9	4	19	141

TABLE 1

The table shows the number of iterations required by the three methods to achieve a precision of 10^{-7} on the residual $\|Mx - y\|$, when β ranges from 0.1 to 0.9. The coefficient matrix here is defined in terms of a random binary matrix of order 10^7 . The number of iterations shown is the median over 10 tests.

Jacobi method is upper-bounded by $\tau^{1+\varepsilon}$, where ε is defined as in Theorem 5.3 and increases with the magnitude of the diagonal entries of A . We have introduced the parameter β to appreciate the acceleration gained by this method when the a_{ii} are close to 1.

Similarly, the spectral radius of the iteration matrix for the power method is upper-bounded by $\tau|\lambda_2(A)|$. Note that this convergence rate is linear in τ , but the method can be sensibly faster than the Jacobi one, when the second eigenvalue of A is small. This property is essentially given by the use of a weakly stochastic preconditioner.

Unfortunately, we do not have an explicit convergence theorem for HPER for non-symmetric problems. However note that the use of a weakly stochastic preconditioner combines somehow the two previous convergence properties. In fact, on the one hand, as for the power method, the dominant eigenvalue of A is deflated and does not influence the spectral radius of the iteration matrix, on the other hand, as for the Jacobi scheme, the preconditioner is related with the diagonal entries of the matrix U^*AU , that is similar to A . The tests that we present in Tables 1 and 2 show that HPER runs faster than the Jacobi and power methods and, in particular, its convergence rate increases with β (as the Jacobi iterations do) and increases when the magnitude of the subdominant eigenvalue of A decreases (as for the power method).

We point out that the choice $\tau = 0.9$ has been done accordingly with typical network applications, as for instance the Google's Pagerank centrality. It is worth pointing out that the smaller τ is, the simpler the problem is, thus we do not consider small values of τ in the numerical experiences.

Table 1 shows the results for a randomly generated binary matrix X . The eigenvalues of X in this case cluster around the origin (e.g., [38, 40]). As the value of β increases, the number of iterations shown is the median over 10 tests. The HPER method significantly outperforms the other ones. The subsequent Table 2 shows, instead, the behaviour of the three methods on a number of real world datasets. The test matrices considered are part of the University of Florida sparse matrix collection [12]. The matrices considered are both symmetric (undirected) and unsymmetric (directed).

Acknowledgement. The authors are grateful to two anonymous referees for their careful reading of the paper and their insightful suggestions.

	Network	n	β	# iterations		
				HPER	Jacobi	power method
Undirected	Delaunay19	524 288	0.1	156	184	172
	Delaunay21	2 097 152	0.2	178	206	221
	Italy OSM	6 686 493	0.2	201	214	241
	Europe OSM	50 912 018	0.2	159	167	186
Directed	Indian web crawl	1 382 908	0.2	187	209	235
	Wikipedia 2006	3 148 440	0.1	159	177	175
	Wikipedia 2007	3 566 907	0.1	157	178	173
	LJournal 2008	5 363 260	0.1	154	192	185

TABLE 2

The table shows the number of iterations required by the three methods to achieve a precision of 10^{-7} on the residual $\|Mx - y\|$. Tests here have been made on real world matrices of different sizes, and the value of β has been chosen between 0.1 and 0.2. Tests for larger values of β (here omitted) show a significant acceleration of HPER over the other two methods.

REFERENCES

- [1] P. Berkhin. A survey on PageRank computing. *Internet Mathematics*, 2(1):73–120, 2005.
- [2] R. Bhatia. *Matrix analysis*. Graduate Texts in Mathematics, Springer Verlag, New York, 1997.
- [3] D. Bini and P. Favati. On a matrix algebra related to the discrete Hartley transform. *SIAM Journal on Matrix Analysis and Applications*, 14(2):500–507, 1993.
- [4] A. Bortoletti and C. Di Fiore. On a set of matrix algebras related to discrete Hartley-type transforms. *Linear Algebra and its Applications*, 366:65–85, 2003.
- [5] E. Bozzo and C. Di Fiore. On the use of certain matrix algebras associated with discrete trigonometric transforms in matrix displacement decomposition. *SIAM Journal on Matrix Analysis and Applications*, 16(1):312–326, 1995.
- [6] C. Brezinski and M. Redivo-Zaglia. Rational extrapolation for the PageRank vector. *Mathematics of Computation*, 77(263):1585–1598, 2008.
- [7] C. Brezinski, M. Redivo-Zaglia, and S. Serra-Capizzano. Extrapolation methods for PageRank computations. *Comptes Rendus Mathématique*, 340(5):393–397, 2005.
- [8] S. Brin and L. Page. The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems*, 30(1):107–117, 1998.
- [9] A.Z. Broder, R. Lempel, F. Maghoul, and J. Pedersen. Efficient PageRank approximation via graph aggregation. *Information Retrieval*, 9(2):123–138, 2006.
- [10] R.H. Chan, X.Q. Jin, and M.C. Yeung. The circulant operator in the Banach algebra of matrices. *Linear Algebra and its Applications*, 149:41–53, 1991.
- [11] S. Cipolla, C. Di Fiore, F. Tudisco, and P. Zellini. Adaptive matrix algebras in unconstrained minimization. *Linear Algebra and its Applications*, 471:544–568, 2015.
- [12] T.A. Davis and Y. Hu. The University of Florida sparse matrix collection. *ACM Transactions on Mathematical Software (TOMS)*, 38(1):1–25, 2011.
- [13] G.M. Del Corso, A. Gulli, and F. Romani. Fast PageRank computation via a sparse linear system. *Internet Mathematics*, 2(3):251–273, 2005.
- [14] F. Di Benedetto. Gram matrices of fast algebras have a rank structure. *SIAM Journal on Matrix Analysis and Applications*, 31(2):526–545, 2009.
- [15] F. Di Benedetto and S. Serra-Capizzano. A note on the superoptimal matrix algebra operators. *Linear and Multilinear Algebra*, 50(4):343–372, 2002.
- [16] F. Di Benedetto and S. Serra-Capizzano. A unifying approach to abstract matrix algebra preconditioning. *Numerische Mathematik*, 82(1):57–90, 1999.
- [17] F. Di Benedetto and S. Serra-Capizzano. Optimal multilevel matrix algebra operators. *Linear and Multilinear Algebra*, 48(1):35–66, 2000.

- [18] C. Di Fiore and P. Zellini. Matrix algebras in optimal preconditioning. *Linear Algebra and its Applications*, 335(1–3):1–54, 2001.
- [19] C. Di Fiore, S. Fanelli, and P. Zellini. Low complexity minimization algorithms. *Numerical Linear Algebra with Applications*, 12(8):755–768, 2005.
- [20] D. Fasino and F. Tudisco. An algebraic analysis of the graph modularity. *SIAM Journal on Matrix Analysis and Applications*, 35(3):997–1018, 2014.
- [21] D. Fasino and F. Tudisco. Generalized modularity matrices. *Linear Algebra and its Applications*, 502:327–345, 2016.
- [22] D. Gleich, L. Zhukov, and P. Berkhin. Fast parallel PageRank: A linear system approach. Yahoo! Research Technical Report, 13–22, 2004.
- [23] G.H. Golub and C. Greif. An Arnoldi-type algorithm for computing PageRank. *BIT Numerical Mathematics*, 46(4):759–771, 2006.
- [24] G.H. Golub and C.F. Van Loan. *Matrix Computations*, fourth edition. Johns Hopkins University Press, Baltimore, MD, 1983.
- [25] C. Gu and L. Wang. On the multi-splitting iteration method for computing PageRank. *Journal of Applied Mathematics and Computing*, 42(1–2):479–490, 2013.
- [26] S. Kamvar, T. Haveliwala, and G. Golub. Adaptive methods for the computation of PageRank. *Linear Algebra and its Applications*, 386:51–65, 2004.
- [27] S. Kamvar, T. Haveliwala, C. Manning, and G. Golub. Exploiting the block structure of the web for computing pagerank. Technical Report, Stanford University, 2003.
- [28] S.D. Kamvar, T.H. Haveliwala, C.D. Manning, and G.H. Golub. Extrapolation methods for accelerating PageRank computations. *Proceedings of the 12th International Conference on World Wide Web*, ACM, 261–270, 2003.
- [29] A.N. Langville and C.D. Meyer. *Google’s PageRank and beyond: The science of search engine rankings*. Princeton University Press, Princeton, NJ, 2012.
- [30] I. Marek and D.B. Szyld. Iterative and semi-iterative methods for computing stationary probability vectors of Markov operators. *Mathematics of Computation*, 61(204):719–731, 1993.
- [31] M.E. Newman. A measure of betweenness centrality based on random walks. *Social Networks*, 27(1):39–54, 2005.
- [32] J.D. Noh and H. Rieger. Random walks on complex networks. *Physical Review letters*, 92(11):118701, 2004.
- [33] P. Pons and M. Latapy. Computing communities in large networks using random walks. *Journal of Graph Algorithms and Applications*, 10(2):191–218, 2006.
- [34] X. Qi, E. Fuller, Q. Wu, Y. Wu, and C.Q. Zhang. Laplacian centrality: A new centrality measure for weighted networks. *Information Sciences*, 194:240–253, 2012.
- [35] U.G. Rothblum and C.P. Tan. Upper bounds on the maximum modulus of subdominant eigenvalues of nonnegative matrices. *Linear Algebra and its Applications*, 66:45–86, 1985.
- [36] V. Sanchez, P. Garcia, A.M. Peinado, J.C. Segura, and A.J. Rubio. Diagonalizing properties of the discrete cosine transforms. *IEEE transactions on Signal Processing*, 43(11):2631–2641, 1995.
- [37] V. Sanchez, A.M. Peinado, J.C. Segura, P. Garcia, and A.J. Rubio. Generating matrices for the discrete sine transforms. *IEEE Transactions on Signal Processing*, 44(10):2644–2646, 1995.
- [38] T. Tao. *Topics in Random Matrix Theory*. American Mathematical Society, Providence, RI, 2012.
- [39] F. Tudisco. A note on certain ergodicity coefficients. *Special Matrices*, 3(1):175–185, 2015.
- [40] F. Tudisco and C. Di Fiore. A preconditioning approach to the PageRank computation problem. *Linear Algebra and its Applications*, 435(9):2222–2246, 2011.
- [41] F. Tudisco, C. Di Fiore, and E.E. Tyrtyshnikov. Optimal rank matrix algebras preconditioners. *Linear Algebra and its Applications*, 438(1):405–427, 2013.
- [42] E.E. Tyrtyshnikov. A unifying approach to some old and new theorems on distribution and clustering. *Linear Algebra and its Applications*, 232:1–43, 1996.
- [43] R.S. Varga. *Matrix Iterative Analysis*. Prentice-Hall, Englewood Cliffs, New Jersey, 1962.
- [44] R.K. Yarlagadda and J.E. Hershey. *Hadamard Matrix Analysis and Synthesis: With Applications to Communications and Signal/Image Processing*. Springer Science & Business Media, 2012.