



ELSEVIER

Contents lists available at ScienceDirect

# Linear Algebra and its Applications

journal homepage: [www.elsevier.com/locate/laa](http://www.elsevier.com/locate/laa)

## A preconditioning approach to the pagerank computation problem

Francesco Tudisco\*, Carmine Di Fiore

*Dipartimento di Matematica, Università di Roma "Tor Vergata", Via della Ricerca Scientifica, 00133 Rome, Italy*

### ARTICLE INFO

#### Article history:

Received 28 July 2009

Accepted 12 April 2011

Available online 14 May 2011

Submitted by E.E. Tyrtyshnikov

#### AMS classification:

65F05

65F10

65F15

65F50

65T50

#### Keywords:

Pagerank

Iterative methods

Preconditioning

Eigenvalues

Clustering

Fast discrete transforms

### ABSTRACT

Some spectral properties of the transition matrix of an oriented graph indicate the preconditioning of Euler–Richardson (ER) iterative scheme as a good way to compute efficiently the vertexrank vector associated with such graph. We choose the preconditioner from an algebra  $\mathcal{U}$  of matrices, thereby obtaining an  $ER_{\mathcal{U}}$  method, and we observe that  $ER_{\mathcal{U}}$  can outperform ER in terms of rate of convergence. The proposed preconditioner can be updated at a very low cost whenever the graph changes, as is the case when it represents a generic set of information. The particular  $\mathcal{U}$  utilized requires a surplus of operations per step and memory allocations, which make  $ER_{\mathcal{U}}$  superior to ER for not too wide graphs. However, the observed high improvement in convergence rate obtained by preconditioning and the general theory developed, are a reason for investigating different choices of  $\mathcal{U}$ , more efficient for huge graphs.

© 2011 Elsevier Inc. All rights reserved.

## 1. Introduction

The pagerank problem concerns the computation of the positive left eigenvector of an irreducible positive stochastic  $n \times n$  matrix  $W$ , i.e. the vector  $\mathbf{p}$  such that  $\mathbf{p} = W^T \mathbf{p}$ . The knowledge of  $\mathbf{p}$  allows to evaluate the importance of each information of a large set of information and thus facilitates the picking up in the same set of particular data. The matrix  $W$  is defined in terms of a sparse matrix  $P$  which

\* Corresponding author.

E-mail addresses: [tudisco.francesco@gmail.com](mailto:tudisco.francesco@gmail.com) (F. Tudisco), [difiore@mat.uniroma2.it](mailto:difiore@mat.uniroma2.it) (C. Di Fiore).

has a very particular structure and is the transition matrix of an oriented graph, usually describing the web hyperlink structure. Precisely  $W$  depends on  $P$ , on a parameter  $\alpha \in (0, 1)$  and on a positive *personalization* vector  $\mathbf{v}$  such that  $\|\mathbf{v}\|_1 = 1$  (see [1,2], for instance). The currently most competitive methods to determine  $\mathbf{p}$  are based on the power method, which generates a sequence  $\mathbf{p}_{k+1} = W^T \mathbf{p}_k$ ,  $k = 0, 1, 2, \dots$  convergent to  $\mathbf{p}$ . Such method, thanks to its simplicity, requires only two real vectors to store the data and a relatively small amount of operations per step, when implemented in terms of the sparse matrix  $P$  [1, pp. 77–78]. However, since the magnitude of all the eigenvalues of  $W$  (except the obvious eigenvalue 1) is less than or equal to  $\alpha$ , its convergence rate is  $O(\alpha^k)$ , and thus is slow for the usual exploited values of  $\alpha$  (f.i. the Google search engine sets  $\alpha = 0.85$ ).

In [3], it is proved that the pagerank vector  $\mathbf{p}$  is also solution of a linear system of type  $\gamma(I - \alpha P)^T \mathbf{p} = \mathbf{v}$ ,  $\gamma \in \mathbb{R}$ , referred as the pagerank system. Starting from this alternative formulation of the problem, some other possible methods to compute  $\mathbf{p}$  have been investigated and compared to power method, as in [3–6]. Such methods sometimes have a convergence rate greater than  $O(\alpha^k)$ , but this reverberates on the request of more operations each step. Other methods have been proposed, instead, regarding the problem in its original formulation (i.e. where  $\mathbf{p}$  is the dominant left eigenvector of  $W$ ), which are essentially extrapolation-based variants of power method [7–9].

In the present work, starting from a quasi-stochastic row-constant sparse matrix  $P$ , we first define a matrix  $W = \alpha P + R$ , with  $R$  a rank-one matrix, which encloses the essential characteristics of the matrix defining the pagerank vector, discussed above. For instance, we observe that such  $W$  still has the spectrum contained into an  $\alpha$ -neighborhood of the origin, with exception for the eigenvalue 1. Moreover, its left dominant eigenvector  $\mathbf{p}$  still is solution of a sparse linear system, involving the matrix  $P$  (see Section 2). Then, in the remaining part of the work, we study several aspects of this kind of *generalized* pagerank computation problem, and we indicate, in particular, the preconditioning as a possible way to improve the rate of convergence of power method in computing  $\mathbf{p}$ .

Let us describe in detail the contents of Sections 3–8. Section 3 is devoted to the investigation of the clustering spectral properties of the quasi-stochastic row-constant sparse matrix  $P$ , trying to give a theoretical justification of the following experimental observation: there exists  $\lambda_\alpha$ , eigenvalue of  $\alpha P$ , such that  $|\lambda_\alpha| \approx \alpha$  and the other eigenvalues of  $P$  fall into a relatively small cluster round the origin, whose size essentially does not depend on  $\alpha$  (see Section 8, for details). We establish a probabilistic relationship between eigenvalues and singular values clustering properties for  $P$ ; we give some bounds for the first and the second singular values of  $P$  and for the gap between them; we obtain an upper bound for the second eigenvalue of  $P$ . In Section 4, we introduce a way to define a sort of preconditioned power method. We observe that the complexity per step, the memory usage, and the convergence rate of power method and Euler–Richardson method, applied to solve the pagerank system, are the same, and thus such algorithms can be considered equivalent. In the same time, we introduce a basic Euler–Richardson type iterative scheme, defined in terms of a matrix  $Q$ , which turns out to be the way to precondition power method. In Sections 5 and 6, we investigate two different choices for the preconditioner  $Q$  to be used in such scheme: the first obtained as a polynomial of  $P$ , the second by considering a T. Chan-like approximation of  $P$  from a matrix algebra  $\mathcal{U}$ . Such second choice for  $Q$  gives rise, for  $\mathcal{U} = \text{circulants}$ , to a method whose convergence rate essentially does not depend on  $\alpha$  and is heavily greater than  $O(\alpha^k)$ . In fact, the experiments done show that only one eigenvalue of  $P$  is responsible of the slow convergence rate of power method and that the circulant preconditioner  $Q$  seems to remove such eigenvalue; since the magnitude of all the other eigenvalues of  $\alpha P$  essentially does not depend on  $\alpha$ , the gain in terms of convergence rate is high also, and primarily, for  $\alpha$  large. In Section 7, we introduce an useful and easily implementable procedure to update in real time the essential quantities defining  $Q$ . In fact, the graph defining the pagerank vector is generally subject to uninterrupted changes (think, for instance, to the web hyperlink graph); thus it is very important to be able to adapt easily the preconditioner  $Q$  to such changes. Finally, some numerical experiences are proposed and described in Section 8, which show how great advantages can be obtained by preconditioning. In spite of the fact that the novel method requires more operations and memory allocations per step than power method, and thus outperforms such method for not too large  $P$ , the theory developed in this work opens the way to future investigations looking for different choices of  $\mathcal{U}$  and of the preconditioner  $Q$  in  $\mathcal{U}$ , more efficient for huge  $P$ .

## 2. Preliminaries

### 2.1. Notation and definitions

We denote with  $\mathbf{M}_n(\mathcal{K})$  the ring of  $n \times n$  matrices with elements in the field  $\mathcal{K}$  ( $\mathcal{K}$  can be  $\mathbb{R}$  or  $\mathbb{C}$ ), with  $\mathbf{U}_n(\mathcal{K}) \subset \mathbf{M}_n(\mathcal{K})$  the group of unitary matrices and with  $(\mathbf{s})\mathbf{PD}_n(\mathcal{K}) \subset \mathbf{M}_n(\mathcal{K})$  the set of (semi)positive definite matrices. When it is clear from the context we omit to specify the field  $\mathcal{K}$ . With  $B_\rho(x)$  we indicate the open ball  $\{y \in \mathbb{C} \mid |x - y| < \rho\}$  and with  $\overline{B_\rho(x)}$  its closure. Given  $X \in \mathbf{M}_n(\mathcal{K})$  we shall denote with  $\Lambda(X)$  and  $\Sigma(X)$  the set of all the eigenvalues and of all the singular values of  $X$ , respectively; with  $\lambda_i(X)$  and  $\sigma_i(X)$  the  $i$ th eigenvalue and the  $i$ th singular value of  $X$ , respectively; with  $\rho(X)$  the spectral radius of  $X$ ; with  $X^*$  the Hermitian adjoint of  $X$  ( $X^T$  if  $X$  is real); with  $\text{diag}(x)$  the diagonal matrix obtained from  $x$ , distinguishing between the case in which  $x$  is a matrix and the case in which  $x$  is a vector: if  $x$  is a matrix then  $\text{diag}(x)_{ii} = x_{ii}$ , if  $x$  is a vector then  $\text{diag}(x)_{ii} = x_i$ , for  $i = 1, \dots, n$ . With  $\pi_{ij}(X)$  we denote the probability that  $(X)_{ij} \neq 0$  and with  $d_X$  the density of  $X$ , i.e. the ratio between the number of non-zero elements of  $X$  and the order  $n$ . So that  $d_X = n$  for a full matrix  $X$ .

Denote with  $P \in \mathbf{M}_n(\mathbb{R})$  a quasi-stochastic row-constant sparse matrix. That is,  $P = \Delta_p^{-1}E_p$ , where  $(E_p)_{ij} \in \{0, 1\}$ ,  $\forall i, j = 1, \dots, n$ ,  $\pi_{ij}(E_p) = d_{E_p}/n$  and  $d_{E_p}$  does not essentially depend on  $n$  and is heavily smaller than  $n$ ;  $\Delta_p = \text{diag}(\boldsymbol{\mu})$ ,  $\boldsymbol{\mu} = E_p \mathbf{e}$ ,  $\mathbf{e} = [1, \dots, 1]^T$  (we assume that the inverse of  $\Delta_p$  is done inverting its non-zero elements, taking no care of null ones). Set  $\Omega = \{i \mid \mu_i > 0\}$  assuming  $\Omega \neq \{1, \dots, n\}$ , and consider the map  $\varphi_\Omega : \mathbb{R}^n \rightarrow \mathbb{R}^n$ ,  $\varphi_\Omega(\mathbf{x})_i = x_i$  if  $i \in \Omega$  and  $\varphi_\Omega(\mathbf{x})_i = 0$ , otherwise. The cardinality  $|\Omega|$  coincides with the number of non-null rows of  $P$ . Let  $\nu(i)$  and  $\rho(i)$ , respectively, be the sets of indexes such that  $(P)_{ij} > 0, \forall j \in \nu(i)$ , and  $(P)_{ji} > 0, \forall j \in \rho(i)$ ,  $i = 1, \dots, n$ . Clearly  $\Omega^c \subset \rho(i)$  for any  $i = 1, \dots, n$ . Observe that by the above definitions it follows that  $(P)_{ij} \geq 0, \forall i, j = 1, \dots, n$ ,  $\rho(P) \leq 1$ , and  $P\mathbf{e} = \varphi_\Omega(\mathbf{e})$ .

Denote now with  $W \in \mathbf{M}_n(\mathbb{R})$  a positive stochastic irreducible matrix obtained by  $W = \alpha P + R$ , being  $\alpha \in (0, 1)$  and  $R \in \mathbf{M}_n(\mathbb{R})$  a suitable rank-one matrix, and let  $\mathbf{v}$  be a vector spanning range( $R^T$ ) such that  $\mathbf{v}^T \mathbf{e} = 1$ . We refer to such matrix as the *W-pagerank matrix*. Note that the definition of  $W$  is well posed since there exists at least one matrix  $R$  with the required properties, namely the rank-one correction matrix used to define the Google-pagerank matrix ([1], for instance). Observe that by requiring  $W$  to be stochastic, i.e. the identity  $W\mathbf{e} = \mathbf{e}$ , we implicitly require that  $R\mathbf{e} = \mathbf{e} - \alpha\varphi_\Omega(\mathbf{e})$ .

### 2.2. The W-pagerank problem

Observe that if we enumerate the web pages from 1 to  $n$ , then the matrix  $P$  we have defined has exactly the structure of a web hyperlink matrix, whose generic  $ij$  element is non-zero and equal to  $\frac{1}{\mu_i}$  only if there exists a hyperlink (the existence of such link depends randomly on  $i$  and  $j$ ) pointing from the page  $i$  to the page  $j$ , and the page  $i$  points to  $\mu_i$  pages. Anyway, we can assume in general that  $P$  is related to (is the transition matrix of) an oriented graph (see Section 7, for clearness).

The *W-pagerank computation problem* consists in the problem of computing the left dominant eigenvector  $\mathbf{p}$  of  $W$ , or rather the vector s.t.  $W^T \mathbf{p} = \mathbf{p}$ , what we call the *W-pagerank vector*. Note that the knowledge of  $\mathbf{p}$  allows to evaluate the importance of each vertex of the graph associated with  $P$  and thus facilitates the searching of a particular vertex (data) in the set of all vertices (all data), see, for instance [10–12]. Both huge and smaller dimension graphs have great interest in applications, but of course the most famous are the huge ones, i.e. Google, Yahoo, MSN.

Perron–Frobenius theorem ensures the existence and uniqueness of  $\mathbf{p}$ , and the well-known power method (i.e.  $\mathbf{p}_{k+1} = W^T \mathbf{p}_k, k = 0, 1, 2, \dots$ ) is a possible easily computable iterative way to obtain  $\mathbf{p}$ . Moreover such method converges as  $O(\alpha^k)$  to the solution. The following is a way to prove this fact.

**Theorem 2.1.**  $|\lambda_i(W)| \leq \alpha, i = 2, \dots, n$  and  $\lambda_1(W) = \rho(W) = 1$ . As a consequence the power method  $\mathbf{p}_{k+1} = W^T \mathbf{p}_k$  converges to  $\mathbf{p}$  such that  $W^T \mathbf{p} = \mathbf{p}$  with a rate of convergence  $O(\alpha^k)$ .

**Proof.** By the stochasticity of  $W$  we have just to prove  $|\lambda_i(W)| \leq \alpha$ , for  $i \neq 1$ . By its definition it follows that  $R = (\mathbf{e} - \alpha\varphi_{\Omega}(\mathbf{e}))\mathbf{v}^T$  where  $\mathbf{v}^T\mathbf{e} = 1$ . Let  $\Omega^c$  denote the complement of  $\Omega$ , then  $R = (\alpha\varphi_{\Omega^c}(\mathbf{e}) + (1-\alpha)\mathbf{e})\mathbf{v}^T$ . Therefore, if  $H = P + \varphi_{\Omega^c}(\mathbf{e})\mathbf{v}^T$  then  $\rho(H) = 1$  and  $W = \alpha H + (1-\alpha)\mathbf{e}\mathbf{v}^T$ . Consider a set of  $n-1$  vectors  $\mathbf{u}_2, \dots, \mathbf{u}_n$  such that  $\{\mathbf{e}, \mathbf{u}_2, \dots, \mathbf{u}_n\}$  are linearly independent and define the matrix  $U$  whose columns are  $\mathbf{e}, \mathbf{u}_2, \dots, \mathbf{u}_n$ . Then there exists  $Q \in \mathbf{M}_{n-1}(\mathbb{R})$  and a vector  $\mathbf{b}$  such that  $U^{-1}HU = \begin{pmatrix} 1 & \mathbf{b}^T \\ \mathbf{0} & Q \end{pmatrix}$ . As a consequence  $\frac{1}{\alpha}W$  is similar to a matrix  $B$  of the following form:

$$B = \frac{1}{\alpha}U^{-1}WU = U^{-1}HU + \frac{1-\alpha}{\alpha} \begin{pmatrix} 1 \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{pmatrix} (\mathbf{v}^T U) = \begin{pmatrix} 1/\alpha & \mathbf{b}^T \\ \mathbf{0} & Q \end{pmatrix}.$$

Thus, if  $\phi_A(\lambda)$  denotes the characteristic polynomial of  $A \in \mathbf{M}_n(\mathbb{C})$ , then  $\phi_H(\lambda) = (\lambda - 1)\phi_Q(\lambda)$  and  $\phi_{\alpha^{-1}W}(\lambda) = (\lambda - \alpha^{-1})\phi_Q(\lambda)$ . Therefore for any  $i = 2, \dots, n$ ,  $\lambda_i(W) = \alpha\lambda_i(Q) = \alpha\lambda_i(H)$  and  $|\lambda_i(W)| = \alpha|\lambda_i(H)| \leq \alpha$ .  $\square$

So, by Theorem 2.1 above, the power method, when applied to compute the  $W$ -pagerank vector, reduces the error essentially by an  $\alpha$  factor, for each iteration. Moreover, such method can be implemented in terms of the sparse matrix  $P$ , requiring a relatively small amount of operations and data to be stored per step. In fact  $\mathbf{p}_{k+1} = \alpha P^T \mathbf{p}_k + \gamma_k \mathbf{v}$ , being  $\gamma_k \in \mathbb{R}$  such that  $\mathbf{e}^T R^T \mathbf{p}_k = \gamma_k$ , and an efficient implementation of such scheme can be done as in Google case [1, pp. 77–78]. Note that, if  $p(n)$  denotes the number of multiplicative operations required to compute  $P^T \mathbf{z}$ , then in order to compute  $\mathbf{p}_{k+1}$  from  $\mathbf{p}_k$  one needs  $p(n) = O(n)$  multiplicative operations.

In [3], it is proved that the Google-pagerank vector is also solution of a sparse linear system. Such result can be easily extended to the  $W$ -pagerank vector, in fact:

**Theorem 2.2.** Let  $A = A(\alpha) = I - \alpha P$  and let  $\mathbf{v} \in \mathbb{R}^n$  be a vector spanning  $\text{range}(R^T)$  such that  $\mathbf{v}^T \mathbf{e} = 1$ . Then  $\det(A) \neq 0$  and  $\mathbf{p} = W^T \mathbf{p} = \mathbf{x} \|\mathbf{x}\|_1^{-1}$  where  $\mathbf{x}$  solves the system  $A^T \mathbf{x} = \mathbf{v}$ .

**Proof.** Of course  $A$  is non-singular because  $\rho(P) \leq 1$  implies  $|\lambda_i(A)| \geq |1 - \alpha\rho(P)| \geq 1 - \alpha$ . Since  $\mathbf{e}^T \mathbf{p} = 1$  and  $\text{range}(R) = \text{Span}(\mathbf{e} - \alpha\varphi_{\Omega}(\mathbf{e}))$ , we have  $R^T \mathbf{p} = \mathbf{v}(1 - \alpha\varphi_{\Omega}(\mathbf{e})^T \mathbf{p})$ . Then

$$W^T \mathbf{p} = \mathbf{p} \iff (A - R)^T \mathbf{p} = \mathbf{0} \iff (A^T - \alpha\mathbf{v}\varphi_{\Omega}(\mathbf{e})^T) \mathbf{p} = \mathbf{v}.$$

Therefore, by the Sherman–Morrison formula, there exists  $\gamma \in \mathbb{R}$  such that  $\mathbf{p} = \gamma(A^T)^{-1} \mathbf{v} = \gamma \mathbf{x}$  and  $\gamma = \|\mathbf{x}\|_1^{-1}$  since  $\mathbf{p}$  is non-negative and  $\sum_k p_k = 1$ .  $\square$

We refer to the system

$$A^T \mathbf{x} = \mathbf{v}, \quad A = A(\alpha) = I - \alpha P, \quad \mathbf{v}^T \mathbf{e} = 1 \tag{1}$$

as the  $W$ -pagerank system.

### 3. On the eigen and singular values of $P$

In this section, we investigate the spectral properties of the matrix  $P$  defining (via  $W$ ) the pagerank vector  $\mathbf{p}$ . In particular, by exploiting the particular structure of  $P$ , we state some interesting computable upper and lower bounds for its maximal singular value and for its second eigenvalue. Such studies have originated from attempts to give a theoretical justification to some experimental evidences (see also Section 8) about a clustering behavior of the spectrum of  $P$ , depending on the value of  $d_p$ , and can be useful to evaluate the convergence of iterative methods in solving the  $W$ -problem.

**Definition 3.1.** Accordingly with the notation used in [13,14] we say that  $\Gamma \subset \mathbb{C}$  is a eigen(singular) values general cluster for a sequence of  $n \times n$  matrices  $\{X_n\}_n$ , if the number of eigen(singular) values of  $X_n$  that fall outside the closure  $\overline{\Gamma}$  is  $o(n)$ .

It is useful to observe that the elements on the  $j$ th column of  $E_p, i = 1, 2, \dots, n$ , are  $n$  Bernoullian independent identical distributed random variables, with mean and variance, respectively, given by

$$m = \mathbb{E}[(E_p)_{1j}, \dots, (E_p)_{nj}] = d_p/n,$$

$$\sigma^2 = \text{Var}[(E_p)_{1j}, \dots, (E_p)_{nj}] = m(1 - m),$$

for  $j = 1, \dots, n$ . This observation yields the following:

**Theorem 3.2.** Let  $\{P_n\}_n$  be a sequence of  $n \times n$  matrices with the same structure of  $P$ , and assume that the sequence  $\{d_{p_n}\}_n$  of densities is bounded. Then  $\{\|P_n\|_1\}_n$  is bounded with probability one.

**Proof.** Let  $m, \sigma^2$  be as above and  $\mu_{\min}^{(n)} = \min\{(\Delta_{P_n})_{ii} \mid (\Delta_{P_n})_{ii} \neq 0, i = 1, \dots, n\}$ . Of course, if  $\frac{1}{\mu_{\min}^{(n)}} \sum_i (E_{P_n})_{ij} \leq M, \forall j = 1, \dots, n$ , then  $\|P_n\|_1 \leq M$ . Now let  $d$  be such that  $\sup_n d_{p_n} \leq d < \infty$ , we have  $(n\sigma^2)^{-1} > d^{-1}, \forall n \geq 1$ . Observe moreover that the sequence  $\{\mu_{\min}^{(n)}\}_n$  is bounded as well, i.e. there exists  $\mu_{\min}$  such that  $\sup_n \mu_{\min}^{(n)} \leq \mu_{\min} < \infty$ . Set  $S_n = \frac{\sum_{i=1}^n (E_{P_n})_{ij} - n \cdot m}{\sigma \sqrt{n}}$ , and let  $\mathbb{P}(\Omega)$  denote the probability measure of the set  $\Omega$ . Then

$$\mathbb{P}(\|P_n\|_1 \leq M) \geq \mathbb{P}\left(\sum_{i=1}^n (E_{P_n})_{ij} \leq M\mu_{\min}^{(n)}\right) \geq \mathbb{P}\left(S_n \leq \frac{M\mu_{\min} - d}{\sqrt{d}}\right).$$

Observe that  $S_n \leq S_{n+1}$ , thus by monotony  $\mathbb{P}(S_n \leq \delta) \geq \mathbb{P}(\lim_{n \rightarrow \infty} S_n \leq \delta) = \lim_{n \rightarrow \infty} \mathbb{P}(S_n \leq \delta)$ . Let  $\mathcal{G}$  denote a standard Gaussian random variable. The central limit theorem ensures  $\lim_{n \rightarrow \infty} \mathbb{P}(S_n \leq \delta) = \mathbb{P}(\mathcal{G} \leq \delta)$ . But  $\forall \varepsilon > 0, \forall \delta > 0 \exists M_\varepsilon > 0$  such that  $\mathbb{P}(\mathcal{G} \leq M_\varepsilon \delta) > \varepsilon$ . Therefore for any chosen  $\varepsilon > 0$  there exists an  $M_\varepsilon > 0$  such that

$$\mathbb{P}(\|P_n\|_1 \leq M_\varepsilon) > \varepsilon, \quad \forall n \geq 1.$$

As a consequence  $\exists M > 0$  such that the series  $\sum_{k=1}^{+\infty} \mathbb{P}(\|P_k\|_1 \leq M)$  diverges and by the second Borel–Cantelli lemma ([15], for instance) we have  $\mathbb{P}(\limsup_{n \rightarrow \infty} \|P_n\|_1 \leq M) = 1$ , which finally proves the claim.  $\square$

By using the above theorem we obtain now a clustering property for the eigenvalues of  $P$ , which holds with probability one. However one can state a deterministic analogous result but requiring a bit stronger assumption. See, respectively, the following Corollary 3.3 and Proposition 3.4.

**Corollary 3.3.** Let  $\{P_n\}_n$  be as in Theorem 3.2. If  $\Gamma = \{x \in \mathbb{R} \mid 0 \leq x \leq R\}$  is a singular values general cluster for  $P_n$ , then  $B_R(0)$  is an eigenvalues general cluster for  $P_n$  almost surely.

**Proof.** Note that  $\|P_n\|_2 \leq \sqrt{\|P_n\|_1 \|P_n\|_\infty} = \sqrt{\|P_n\|_1}$ , and by Theorem 3.2 there exists  $M > 0$  such that  $\log(\|P_n\|_2) \leq \frac{1}{2} \log(\|P_n\|_1) \leq M$  with probability one. Use Theorem 3.3 in [14] to derive the thesis.  $\square$

**Proposition 3.4.** Let  $\{P_n\}_n$  be as in Theorem 3.2. Let  $\Gamma = \{x \in \mathbb{R} \mid 0 \leq x \leq R\}$  a singular values general cluster for  $P_n$  and let  $\gamma(n)$  be the number of singular values of  $P_n$  in  $(R, +\infty)$ . If exists  $c \geq 0$  such that  $\gamma(n) \leq \frac{cn}{\log n}$  then  $B_R(0)$  is a general cluster for the eigenvalues of  $P_n$ .

**Proof.** By Theorem 3.3 in [14] it is enough to show that  $\log(\|P_n\|_2) = O\left(\frac{n}{\gamma(n)}\right)$ . But this is true since  $\|P_n\|_2 \leq \sqrt{n}\|P_n\|_\infty = \sqrt{n}$ , and therefore, by the hypothesis on  $\gamma(n)$ ,  $\log(\|P_n\|_2) \leq \frac{1}{2} \log n \leq \frac{c}{2} \frac{n}{\gamma(n)}$ .  $\square$

Corollary 3.3 and Proposition 3.4 affirm that if the singular values of  $P$  cluster in a neighborhood of the origin, then also the eigenvalues of  $P$  do the same. This is not surprising since very often the existence of a singular values cluster implies the existence of an eigenvalues one. However it is not always true, as observed in [14], where a counter-example is given. Moreover one can observe that a proper singular value cluster round zero always implies a proper eigenvalues cluster round zero. This fact has been proved for the first time in [16] and subsequently in [17].

As a consequence of the above results, it is useful to investigate the distribution (location) of the singular values of  $P$ , or, equivalently, of the spectrum of the matrices  $PP^T$  and  $P^T P$ , both in  $\mathbf{sPD}_n(\mathbb{R})$ . In fact, we now obtain upper and lower bounds for the maximal singular value  $\sigma_1(P)$ , and a lower bound for the gap  $\sigma_1(P) - \sigma_2(P)$ .

It is not difficult to prove the identities

$$\begin{aligned} (PP^T)_{ji} &= (PP^T)_{ij} = \frac{|v(i) \cap v(j)|}{\mu_i \mu_j} \\ (P^T P)_{ji} &= (P^T P)_{ij} = \sum_{k \in \rho(i) \cap \rho(j)} \frac{1}{\mu_k^2}, \quad \forall i, j = 1, \dots, n, \end{aligned}$$

and of course  $(PP^T)_{ij} = 0$ , whenever  $\mu_i = 0$  or  $\mu_j = 0$ . Note that  $\sigma_1(P)^2 = \rho(PP^T)$  and  $\lambda_i(PP^T) = \lambda_i(P^T P) = \sigma_i(P)^2, \forall i$ . By the definition of  $P$ ,  $PP^T$  has at least a null row, that is  $\min_i \lambda_i(PP^T) = \min_i \sigma_i(P)^2 = 0$ . From now on assume the singular and eigen values of  $P$  ordered so that  $\sigma_i(P) \geq \sigma_{i+1}(P)$  and  $|\lambda_i(P)| \geq |\lambda_{i+1}(P)|, i = 1, \dots, n - 1$ . Observe that

$$\begin{aligned} \|PP^T\|_\infty &= \max_i \sum_j |(PP^T)_{ij}| = \max_i \sum_j \frac{1}{\mu_i \mu_j} |v(i) \cap v(j)| \\ &= \max_i \frac{1}{\mu_i} \sum_{j: v(j) \neq \emptyset} \frac{1}{\mu_j} |v(i) \cap v(j)| \\ &\leq \max_i \frac{1}{\mu_i} \sum_{j: v(j) \neq \emptyset} \frac{1}{\mu_j} \\ &= \frac{|\{j: v(j) \neq \emptyset\}|}{\min_{i: v(i) \neq \emptyset} \mu_i}, \end{aligned}$$

but also

$$\begin{aligned} \|P^T P\|_\infty &= \max_i \sum_j |(P^T P)_{ij}| = \max_i \sum_j \sum_{k \in \rho(i) \cap \rho(j)} \frac{1}{\mu_k^2} \\ &\leq \max_i \frac{1}{\min_{k: v(k) \neq \emptyset} \mu_k^2} \sum_{j: \rho(j) \neq \emptyset} |\rho(i) \cap \rho(j)| \\ &\leq |\{j : \rho(j) \neq \emptyset\}| \frac{\max_k |\rho(k)|}{\min_{k: v(k) \neq \emptyset} \mu_k^2}. \end{aligned}$$

The latter upper bounds for the infinity norms of  $PP^T$  and  $P^T P$ , together with the remark that  $\sigma_1(P)^2 = \rho(PP^T) = \rho(P^T P) \leq \min\{\|PP^T\|_\infty, \|P^T P\|_\infty\}$ , let us state the proposition here below:

**Proposition 3.5.** *The following upper bounds hold for the greatest singular value of  $P$*

$$\sigma_1(P)^2 \leq \frac{|\{j : v(j) \neq \emptyset\}|}{\min_{i: v(i) \neq \emptyset} \mu_i} \leq \frac{|\{j : \rho(j) \neq \emptyset\}| \max_k |\rho(k)|}{\min_{i: v(i) \neq \emptyset} \mu_i^2}.$$

**Proof.** There follows the proof of the right inequality. Without loss of generality, assume that

$$\begin{aligned} \{j : v(j) \neq \emptyset\} &= \{1, 2, \dots, x\}, \quad x := |\{j : v(j) \neq \emptyset\}|, \\ \{j : \rho(j) \neq \emptyset\} &= \{1, 2, \dots, y\}, \quad y := |\{j : \rho(j) \neq \emptyset\}|. \end{aligned}$$

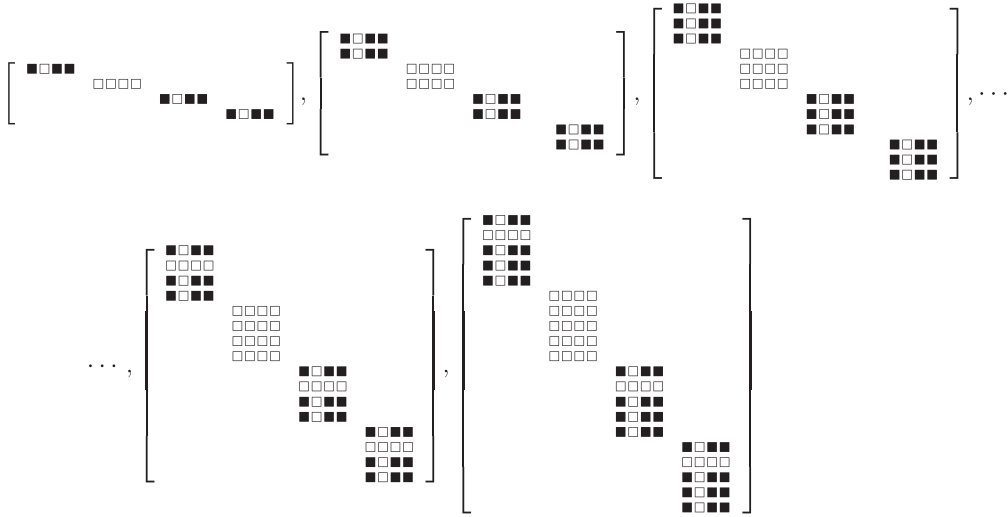


Fig. 1.  $Q$  for  $\mu_{\min}$  generic,  $y = \mu_{\min}^x, \mu_{\min}^{\frac{x}{2}}, \mu_{\min}^{\frac{x}{3}}, \dots, \mu_{\min}^{\frac{x}{i}}, \mu_{\min}^{\frac{x}{i+1}}$ .

Thus, the  $x \times y$  upper left submatrix of  $P$ , which we call  $Q$ , has no null row and no null column, whereas the entries of the remaining part of  $P$  are all zeros:

$$P = \begin{pmatrix} Q & 0 \\ 0 & 0 \end{pmatrix}.$$

This way we only need to prove that  $(\min_{k:v(k) \neq \emptyset} |v(k)|)x \leq (\max_k |\rho(k)|)y$ . Set  $\mu_{\min} = \min_{k:\mu_k > 0} \mu_k$ . We consider the two different cases  $y \geq \mu_{\min}^x$  and  $y < \mu_{\min}^x$ . It is useful to observe that, by the hypotheses, the non-zero entries of the matrix  $Q$  for  $y = \mu_{\min}^x, \mu_{\min}^{\frac{x}{2}}, \mu_{\min}^{\frac{x}{3}}, \dots, \mu_{\min}^{\frac{x}{i}}, \mu_{\min}^{\frac{x}{i+1}}$  must be (modulo permutations) in the positions shown in Fig. 1.  $y \geq \mu_{\min}^x$ : In this case  $\max_k |\rho(k)| \geq 1$  (see Fig. 1,  $y = \mu_{\min}^x$ ). Thus  $\mu_{\min}^x \leq 1 \cdot y \leq (\max_k |\rho(k)|)y$ .

$y < \mu_{\min}^x$ : If  $y < \mu_{\min}^x$ , then  $\max_k |\rho(k)| \geq 2$  (see Fig. 1,  $y = \mu_{\min}^x, \mu_{\min}^{\frac{x}{2}}$ ). If  $\mu_{\min}^{\frac{x}{2}} \leq y$  too, then  $\mu_{\min}^x \leq 2y \leq (\max_k |\rho(k)|)y$ . If  $y < \mu_{\min}^{\frac{x}{2}}$ , then  $\max_k |\rho(k)| \geq 3$  (see Fig. 1,  $y = \mu_{\min}^{\frac{x}{2}}, \mu_{\min}^{\frac{x}{3}}$ ). If  $\mu_{\min}^{\frac{x}{3}} \leq y$  too, then  $\mu_{\min}^x \leq 3y \leq (\max_k |\rho(k)|)y$ . In general, let  $i \in \{1, 2, \dots\}$ . If  $y < \mu_{\min}^{\frac{x}{i}}$ , then  $\max_k |\rho(k)| \geq i + 1$  (see Fig. 1,  $y = \mu_{\min}^{\frac{x}{i}}, \mu_{\min}^{\frac{x}{i+1}}$ ). If  $\mu_{\min}^{\frac{x}{i+1}} \leq y$  too, then  $\mu_{\min}^x \leq (i + 1)y \leq (\max_k |\rho(k)|)y$ .  $\square$

Observe that the particular structure of  $P$  let us also obtain lower bounds for  $\sigma_1(P)$ . These are obtained from the identities

$$\sigma_1(P)^2 = \rho(PP^T) = \rho(P^T P) = \max_{\|\mathbf{x}\|_2=1} \|P\mathbf{x}\|_2^2 = \max_{\|\mathbf{x}\|_2=1} \|P^T \mathbf{x}\|_2^2,$$

by choosing suitable vectors  $\mathbf{x}$ . For example, the choices  $\mathbf{x} = \mathbf{e}_i$  and  $\mathbf{x} = \frac{1}{\sqrt{2}}(\mathbf{e}_i + \mathbf{e}_j)$  (where  $\mathbf{e}_i$  is the  $i$ th canonical vector) yield the easily computable bounds in the proposition below. More generally, any choice  $\mathbf{x} = \frac{1}{\sqrt{k}}(\mathbf{e}_{i_1} + \dots + \mathbf{e}_{i_k})$ , gives significant lower bounds for  $\sigma_1(P)$ , but the greater is  $k$  the more difficult is their evaluations.

**Proposition 3.6.** *The following lower bounds hold for the greatest singular value of  $P$*

$$\sigma_1(P)^2 \geq \max \left\{ \max_{i:\rho(i) \neq \emptyset} \sum_{k \in \rho(i)} \frac{1}{\mu_k^2}, \frac{1}{\min_{i:\nu(i) \neq \emptyset} \mu_i} \right\}$$

$$\sigma_1(P)^2 \geq \frac{1}{2} \max \left\{ \max_{\substack{i \neq j \\ \nu(i) \neq \emptyset, \nu(j) \neq \emptyset}} \left[ \frac{1}{\mu_i} + \frac{1}{\mu_j} + \frac{2|\nu(i) \cap \nu(j)|}{\mu_i \mu_j} \right], \right.$$

$$\left. \max_{\substack{i \neq j \\ \rho(i) \neq \emptyset, \rho(j) \neq \emptyset}} \left[ \sum_{k \in (\rho(i) \cup \rho(j)) \setminus (\rho(i) \cap \rho(j))} \frac{1}{\mu_k^2} + \sum_{k \in \rho(i) \cap \rho(j)} \frac{4}{\mu_k^2} \right] \right\}$$

The upper and lower bounds obtained for  $\sigma_1(P)$  could be used in order to estimate the gap between the first and the second singular values of  $P$ , i.e. to observe the existence of the cluster of  $\Sigma(P)$  that experiments show (see also Section 8). For instance, one can use a formula in [18], which applied to  $P$  reads

$$\sigma_1(P) - \sigma_2(P) \geq \frac{1}{2} \left( \sigma_1(P) - \frac{1}{\min_i \mu_i} \right) \varepsilon(P, \mathbf{u})^2$$

where  $\mathbf{u}$  is the right (but also the left) maximal eigenvector of  $PP^T$  and

$$\varepsilon(P, \mathbf{u}) = \min_{S \subset \mathbf{N}, |S| \leq \frac{n}{2}} \sum_{j \in \mathbf{N} \setminus S} \frac{u_j}{\mu_j} \left( \frac{\sum_{i \in S} \frac{1}{\mu_i} |\nu(i) \cap \nu(j)| u_i}{\sum_{i \in S} \left( \sigma_1(P) - \frac{1}{\mu_i} \right) u_i^2} \right).$$

We conclude this section with the following theorem in which we give an upper bound for the second largest eigenvalue  $\lambda_2(S)$  of a generic stochastic row-constant matrix  $S$ .

**Theorem 3.7.** Let  $S = \Delta_S^{-1} E_S = \text{diag}(s_1^{-1}, \dots, s_n^{-1}) E_S$  be a row-constant stochastic  $n \times n$  matrix. Enumerate the eigenvalues of  $S$  so that  $1 = \rho(S) = \lambda_1(S) \geq |\lambda_2(S)| \geq \dots \geq |\lambda_n(S)|$ . Let  $\mathbf{N} = \{1, \dots, n\}$  and, as for the quasi-stochastic case, define  $\nu(i) = \{j \in \mathbf{N} \mid (E_S)_{ij} = 1\}, \forall i \in \mathbf{N}$ , so that  $|\nu(i)| = s_i$ . Set  $m = \arg(\min_{i \in \mathbf{N}} s_i)$ ,  $M = \arg(\max_{i \in \mathbf{N}} s_i)$  and  $\eta_S = \frac{|\nu(m) \cap \nu(M)|^2}{s_m s_M}$ . If  $\eta_S \geq 1 - \delta/2$  for  $0 < \delta < 1$  then

$$|\lambda_2(S)| \leq \delta < \rho(S) = 1.$$

In other words the eigenvalues of  $S$  (except for the spectral radius) cluster into a  $\delta$ -neighborhood of the origin.

**Proof.** In [19] was proved the following non-optimal bound for  $\lambda_2(S)$ , as discussed in [20],

$$|\lambda_2(S)| \leq \min_{j \in \mathbf{N}} \max_{i \in \mathbf{N}} \sum_{k \in \mathbf{N}} \sum_{r \in \mathbf{N}} |(S)_{ik}(S)_{jr} - (S)_{ir}(S)_{jk}|.$$

Therefore, if  $\varepsilon_{ij} = (E_S)_{ij}$ , we have

$$\begin{aligned} |\lambda_2(S)| &\leq \min_j \max_i \sum_k \sum_r \frac{1}{s_i s_j} |\varepsilon_{ik} \varepsilon_{jr} - \varepsilon_{ir} \varepsilon_{jk}| \\ &= \min_j \max_i \frac{1}{s_i s_j} \left[ |\nu(i) \cap \nu(j)| \sum_k |\varepsilon_{ik} - \varepsilon_{jk}| \right. \\ &\quad \left. + (|\nu(i)| - |\nu(i) \cap \nu(j)|) \sum_k \varepsilon_{jk} + (|\nu(j)| - |\nu(i) \cap \nu(j)|) \sum_k \varepsilon_{ik} \right] \end{aligned}$$



$$\begin{aligned}
 &= \min_j \max_i \frac{1}{s_i s_j} \left[ |\nu(i) \cap \nu(j)| (s_i + s_j - 2|\nu(i) \cap \nu(j)|) \right. \\
 &\quad \left. + s_i s_j - |\nu(i) \cap \nu(j)| s_j + s_i s_j - |\nu(i) \cap \nu(j)| s_i \right] \\
 &= \min_j \max_i \frac{1}{s_i s_j} (2s_i s_j - 2|\nu(i) \cap \nu(j)|^2) \\
 &= 2 - 2 \max_j \min_i \frac{|\nu(i) \cap \nu(j)|^2}{s_i s_j}. \tag{2}
 \end{aligned}$$

Let  $k \in \mathbf{N}$  and set  $f_k : \mathbf{N} \rightarrow \mathbf{N}, s_i \mapsto f_k(s_i) = |\nu(i) \cap \nu(k)|$ . The expression obtained in (2) can then be rewritten as

$$k \in \mathbf{N}, \quad \frac{f_k(s_i)^2}{s_i s_k} =: \psi_k(s_i), \quad |\lambda_2(S)| \leq 2 - 2 \max_{k \in \mathbf{N}} \min_{i \in \mathbf{N}} \psi_k(s_i).$$

Now consider the natural extension of  $f_k$  on  $\mathbb{R}$ , denote it still with  $f_k$ . Observe that in  $[0, n]$  one has  $0 \leq f_k(x) \leq x, f_k(0) = 0$  and  $y \leq x \Rightarrow f_k(y) \leq f_k(x)$ . This implies that  $f_k$  is monotone non-decreasing, hence differentiable a.e., precisely  $x > 0 \Rightarrow 0 < \frac{d}{dx} f_k(x) \leq 1$  a.e.. Extend also  $\psi_k$  on  $\mathbb{R}$ ; clearly it also is differentiable a.e., moreover it can be easily observed that  $x \geq f_k(x)$  implies  $\frac{d}{dx} \psi_k(x) \geq 0$  and thus also  $\psi_k$  is monotone non-decreasing. As a consequence

$$\max_{j \in \mathbf{N}} \min_{i \in \mathbf{N}} \psi_j(s_i) = \max_{j \in \mathbf{N}} \psi_j(s_m) = \max_{j \in \mathbf{N}} \frac{|\nu(m) \cap \nu(j)|^2}{s_m s_j}.$$

Finally, an analogous discussion yields  $\max_{j \in \mathbf{N}} \min_{i \in \mathbf{N}} \psi_j(s_i) = \eta_S$ , therefore by (2) we have the thesis.  $\square$

Observe that the formula obtained is of interest only if  $s_m/s_M > \frac{1}{2}$ , since  $\eta_S \leq s_m/s_M$ . Thus with the Theorem above we give a slack computable bound for the eigenvalue cluster of a generic row-constant stochastic  $n \times n$  matrix  $S$ , containing all the eigenvalues  $\lambda_i(S)$  with the only exception for  $\lambda_1(S) = 1$  (the largest one). Of course such formula should be adapted to estimate a bound for  $\Lambda(P)$ , removing the hypothesis of stochasticity.

#### 4. The power method embedded in a Euler–Richardson iterative scheme

Given a generic non-singular  $n \times n$  matrix  $B$  and a  $n$ -dimensional vector  $\mathbf{b}$ , let us consider the following vectorial application,

$$\Phi_{B, \mathbf{b}}(Q, \mathbf{z}) = \mathbf{z} + Q(\mathbf{b} - B\mathbf{z}), \tag{3}$$

where  $Q \in \mathbf{M}_n(\mathbb{C})$  and  $\mathbf{z}$  is any  $n \times 1$  vector. Observe that the vector  $\mathbf{y} = B^{-1}\mathbf{b}$ , is a fixed point for  $\Phi$ , independently from the choice of the matrix  $Q$ . So one may propose the iterative scheme  $\mathbf{x}_{k+1} = \Phi_{B, \mathbf{b}}(Q, \mathbf{x}_k)$  to solve the linear system  $B\mathbf{y} = \mathbf{b}$ . The convergence and the convergence rate of such scheme depends on the choice of  $Q$ , for this reason  $Q$  is referred as the *preconditioning* matrix.

In this work, we apply the above scheme to the linear system (1), i.e. we consider the method

$$\mathbf{x}_0 \in \mathbb{R}^n, \quad \mathbf{x}_{k+1} = \Phi_{A^T, \mathbf{v}}(Q, \mathbf{x}_k), \quad k \geq 0 \tag{4}$$

whose iteration matrix is

$$H_Q = I - QA^T$$

and we propose different choices for the preconditioner  $Q$ , looking for it through particular matrix structures. We will notice that (4) for  $Q = I$  has the same complexity per step, memory usage, and rate of convergence of the power method. For these reasons, in the following, we refer to (4),  $Q = I$ , and to power method indifferently, even if the sequences  $\{\mathbf{x}_k\}_k$  and  $\{\mathbf{p}_k\}_k$  they define may differ (see also [1]).

The idea behind our investigations is that a suitable preconditioner  $Q \neq I$  used in the scheme (4) could modify the spectrum of the iteration matrix  $H_Q$  contracting its spectral radius and then giving rise to a method whose convergence rate is higher than power method ( $Q = I$ ).

**Definition 4.1.** The matrix  $Q$  in (4) is said to be a *spectral preconditioner* or a  $\sigma$ -preconditioner, if  $\rho(I - QA^T) \leq \beta < \rho(I - A^T) = \rho(\alpha P^T)$ , for a positive  $\beta$  smaller than  $\alpha$ .

We expect that a  $\sigma$ -preconditioner can be produced by choosing  $Q$  as an approximation (in a sense that we will specify) of  $(A^T)^{-1}$ , since using exactly  $Q = (A^T)^{-1}$  into (4) ensures that the only eigenvalue of the iteration matrix is 0 with multiplicity  $n$ , or rather that (4) converges in one step.

Let us consider, first, the simplest case  $Q = \omega I$ , where  $\omega \in \mathbb{R}$ . This choice for  $Q$  gives rise to the Euler–Richardson (ER) method for linear systems (see [21], for instance). Since the eigenvalues of  $A$  have positive real parts, there exists  $\omega^* > 0$  such that for all  $\omega \in (0, \omega^*)$  such method is convergent. The optimal value for  $\omega$  is given by the following:

**Theorem 4.2.** Let  $A$ ,  $\mathbf{x}$ ,  $\alpha$  and  $\mathbf{v}$  be as in (1). Choose  $\mathbf{x}_0 \in \mathbb{R}^n$  and define the sequence

$$\mathbf{x}_{k+1} = \Phi_{A^T, \mathbf{v}}(\omega I, \mathbf{x}_k), \quad k = 0, 1, 2, \dots \tag{5}$$

Then the highest rate  $\mathbf{x}_k$  can reach in converging to  $\mathbf{x}$  is dominated by  $O(\alpha^k)$  and is obtained for  $\omega = 1$ .

**Proof.** Consider the iteration matrix  $H_{\omega I} = I - \omega A^T$ , let  $\lambda$  be the generic eigenvalue of  $A$ . Then  $\Lambda(A) \subseteq \bar{B}_\alpha(1)$ . The spectral radius  $\rho(H_{\omega I})$  is a holomorphic function of  $\lambda$ , thus it is dominated by the value it assumes at the boundary  $\partial B_\alpha(1)$ . It is possible to parametrize  $\lambda|_{\partial B_\alpha(1)}$  as a function of  $\phi \in \mathbb{R}$ ,  $\lambda|_{\partial B_\alpha(1)} = \ell(\phi) = 1 + \alpha e^{i\phi}$ . Remark that  $\rho(H_{\omega I})^2 \leq \max_{\phi \in \mathbb{R}} |1 - \omega \ell(\phi)|^2$  and that  $|1 - \omega \ell(\phi)|^2 = 2\omega\alpha(\omega - 1) \cos \phi + \omega^2\alpha^2 + (1 - \omega)^2$ . It is not difficult to observe that this implies

$$\min_{\omega \in \mathbb{R}} \rho(H_{\omega I})^2 \leq \min_{\omega \in \mathbb{R}} \max_{\phi \in \mathbb{R}} |1 - \omega \ell(\phi)|^2 = \alpha^2,$$

therefore the thesis.  $\square$

Theorem 4.2 states that the convergence rate  $\rho(H_I)^k$  of ER method, used to solve the  $W$ -pagerank system (1), is less than or equal to  $O(\alpha^k)$ , i.e. at most the convergence rate of the power method in computing  $\mathbf{p} = W^T \mathbf{p}$ . The ER method, referred in [6,4] as the Jacobi process, requires exactly  $O(p(n))$  multiplications at each step, and exactly two real vectors to be stored, the same as the power method. Therefore the choice of a preconditioning matrix  $Q$  in (4) different from the identity, is a way to introduce a sort of *preconditioned power method*.

### 5. Polynomial preconditioning

In this section, we study the case in which  $Q = f(\alpha P^T)$ , being  $f$  a polynomial function. We are interested in such a type of preconditioner since by the well-known Neumann series,<sup>1</sup> if  $\phi_\tau(X) = \sum_{\nu=0}^{\tau} X^\nu$  then  $\lim_{\tau \rightarrow \infty} \phi_\tau(\alpha P^T) = (A^T)^{-1}$ , for all  $0 \leq \alpha < 1$ . Therefore  $\phi_\tau(\alpha P^T)$  is a polynomial approximation of  $(A^T)^{-1}$  as better as greater  $\tau$  is. Thus, fixed the degree  $\tau$  of  $f$ , we look for the optimal

<sup>1</sup> We refer to the formula:  $\sum_{\nu=0}^{\infty} X^\nu = (I - X)^{-1}$  for all  $X \in \mathbf{M}_n(\mathbb{C})$  such that  $\rho(X) < 1$ .

choice of polynomial  $f$  in order to approximate  $(A^T)^{-1}$  or, similarly, for an optimal preconditioner  $Q$  of the form  $f(\alpha P^T)$ .

For the method defined by  $\mathbf{x}_{k+1} = \Phi_{A^T, \mathbf{v}}(f(\alpha P^T), \mathbf{x}_k)$ , the iteration matrix becomes

$$H_{f(\alpha P^T)} = I - f(\alpha P^T) (I - \alpha P^T). \tag{6}$$

Let  $g$  be the  $\tau + 1$  degree polynomial such that  $f(x)(1 - x) + \gamma = 1 - g(x)$ , for a suitable  $\gamma \in \mathbb{R}$ . If we indicate with  $\xi_i$  the generic eigenvalue of the matrix  $\alpha P$  (note that this implies  $\xi_i \in \overline{B_\alpha(0)}$  i.e.  $|\xi_i| \leq \alpha$ ), then it results

$$\rho(H_{f(\alpha P^T)}) = \max_i |q(\xi_i)| \leq \max_{|\xi| \leq \alpha} |g(\xi) + \gamma| = \max_{|\xi| \leq \alpha} |q(\xi)|$$

for a generic polynomial  $q$  such that  $q(1) = 1$  and  $\deg q = \tau + 1$ . It follows that if  $P_d^1$  denotes the set of polynomials  $P_d^1 = \{p \in \mathbb{R}[x] : \deg p = d, p(1) = 1\}$  then looking for the optimal preconditioner  $Q = f^*(\alpha P^T)$  can be reduced to the research of  $q^* \in P_{\tau+1}^1$  such that the modulus  $|q^*(x)|$  is minimum in the closed region  $|x| \leq \alpha$ .

**Theorem 5.1.** *Let  $T_{B_R(0)}^{(d)}(z) = T_d(z) = z^d$ . Then  $T_d(z)$  realizes the minimum*

$$\min_{p \in P_d^1} \|p(z)\|_{L^\infty(B_R(0))} = \|T_d(z)\|_{L^\infty(B_R(0))}. \tag{7}$$

for all  $0 < R < 1$ .

**Proof.** Consider, first, a monic polynomial  $p$  of degree  $d$ . If  $p \neq T_d(z)$  then  $p(z) = z^d + \sum_{j=1}^d a_j z^{d-j}$ ,  $\sum_j |a_j| \neq 0$ . Set  $\psi(z) = \frac{p(z)}{z^d}$ , then  $\psi$  is holomorphic over the open set  $\{\mathbb{C} \setminus \overline{B_R(0)}\}$ . Since  $\psi(z) \xrightarrow{|z| \uparrow \infty} 1$ , it follows from the maximum principle that  $\max_{z \in \partial B_R(0)} |\psi(z)| > 1$ . Let's indicate with  $\|\cdot\|_\infty$  the *sup-norm* over the compact set  $\overline{B_R(0)}$ . It results

$$\|p(z)\|_\infty = \max_{z \in \partial B_R(0)} |z^d| |\psi(z)| > \|T_d(z)\|_\infty.$$

Consider now a generic polynomial  $g \in P_d^1$ . Then it must exist a polynomial  $\ell(z) = c_{d-1}z^{d-1} + \dots + c_0$  such that  $c_{d-1} \neq 0$  so that  $g(z) = 1 + (1 - z)\ell(z)$ . The same considerations done for the case above imply

$$\|g(z)\|_\infty \geq |c_{d-1}| \|T_d(z)\|_\infty$$

therefore we can limit the attention to the case  $|c_{d-1}| < 1$ . Suppose  $\|g\|_\infty < \|T_d\|_\infty$ ; for the maximum principle it would imply that  $\max_{\phi \in \mathbb{R}} |g(\text{Re}^{i\phi})| < R^d$  or equivalently  $\max_{\phi \in \mathbb{R}} \frac{|g(\text{Re}^{i\phi})|}{R^d} < 1$ , but the latter inequality is impossible since

$$\begin{aligned} \max_{\phi \in \mathbb{R}} \frac{|g(\text{Re}^{i\phi})|}{R^d} &= \max_{\phi \in \mathbb{R}} \left| \frac{1}{R^d} + \frac{\ell(\text{Re}^{i\phi})(1 - \text{Re}^{i\phi})}{R^d} \right| \\ &\geq \max_{\phi \in \mathbb{R}} \left| \frac{1}{R^d} + c_{d-1} R^{d-1} e^{i\phi(d-1)} \frac{1 - \text{Re}^{i\phi}}{R^d} \right| \\ &\geq \left| \frac{1}{R^d} + \frac{1 - R}{R} c_{d-1} \right| > 1 \end{aligned}$$

for all  $R$  and  $c_{d-1}$  such that  $0 < R < 1, |c_{d-1}| < 1$ .  $\square$

We can now give an overestimation of the spectral radius for the iteration matrix (6). In fact, by Theorem 5.1 it follows that

$$f^\star(x) = \frac{1 - T_{B_\alpha(0)}^{(\tau+1)}(x)}{1 - x} = \sum_{j=0}^{\tau} x^j,$$

thus  $\rho(H_{f^\star(\alpha P^T)}) \leq \rho(H_{f(\alpha P^T)})$  for all polynomial  $f$  such that  $\deg f = \tau$ . Moreover,  $\rho(H_{f^\star(\alpha P^T)}) \leq \alpha^{\tau+1}$ .

This implies that each iteration of the method  $\mathbf{x}_{k+1} = \Phi_{A^T, \mathbf{v}}(f^\star(\alpha P^T), \mathbf{x}_k)$  is equivalent to iterate the Euler–Richardson method (5), with  $\omega = 1$ ,  $\tau + 1$  times, i.e. the method generates nothing else than a subsequence of the sequence generated by the ER method. It follows that a preconditioner of the form  $Q = f(\alpha P^T)$  is not of any interest for our scopes.

### 6. Preconditioning by algebras $\mathcal{U}$ : the ER $_{\mathcal{U}}$ algorithm

The second class of preconditioners for (4) is based on unitary matrix properties and fast transformations. Unlike the polynomial preconditioner studied in the previous section, this time we obtain an interesting method which exploits a possible  $\sigma$ -preconditioner  $Q$ .

Consider the subset  $\mathcal{U}$  of the  $n \times n$  complex matrices ring  $\mathbf{M}_n(\mathbb{C})$

$$\mathcal{U} = \{X \in \mathbf{M}_n(\mathbb{C}) : X = UDU^*, D \text{ diagonal}\}, \tag{8}$$

where  $U \in \mathbf{U}_n(\mathbb{C})$  is fixed. It is not difficult to prove that  $\mathcal{U}$  is a  $n$ -dimensional vectorial subspace of  $\mathbf{M}_n(\mathbb{C})$ , as well as a matrix algebra.

Given  $X \in \mathbf{M}_n(\mathbb{C})$ , we indicate with  $\mathcal{U}_X = UDU^*$  the matrix of  $\mathcal{U}$  that best approximates  $X$  in the Frobenius norm, i.e.

$$\|\mathcal{U}_X - X\|_F = \min_{Y \in \mathcal{U}} \|Y - X\|_F.$$

Note that  $\mathcal{U}_X$  is well defined since  $\mathbf{M}_n(\mathbb{C})$  is a Hilbert space with respect to the inner product  $\langle X, Y \rangle_F = \sum_{i,j} \bar{x}_{ij}y_{ij}$ , where  $\bar{x}$  denotes the complex conjugate of  $x$ , and  $\mathcal{U}$  is closed with respect to  $\|\cdot\|_F$ , which is exactly the norm induced by  $\langle \cdot, \cdot \rangle_F$ . In other words  $\mathcal{U}_X$  is the projection of  $X$  onto  $\mathcal{U}$  with respect  $\langle \cdot, \cdot \rangle_F$ . We also recall that  $\|\cdot\|_F$  is unitary transform invariant, i.e.  $\|MXQ\|_F = \|X\|_F$  for all unitary matrices  $M, Q$ .

The following result holds (see [22] or [23]):

**Proposition 6.1.** *Let  $X$  be a  $n \times n$  matrix, then  $(D_X)_{ii} = (U^*XU)_{ii}$ . Therefore  $\mathcal{U}_X = U \text{diag}(U^*XU)U^*$ . Moreover the operator  $X \mapsto \mathcal{U}_X$  is linear, i.e.*

$$\mathcal{U}_{\alpha A + \beta B} = \alpha \mathcal{U}_A + \beta \mathcal{U}_B.$$

Now consider the use of  $\mathcal{U}_{A^T}^{-1}$  as a preconditioner for the method (4), in order to solve the W-pagerank linear system (1):

$$\mathbf{x}_0 \in \mathbb{R}^n, \quad \mathbf{x}_{k+1} = \Phi_{A^T, \mathbf{v}}(\mathcal{U}_{A^T}^{-1}, \mathbf{x}_k). \tag{9}$$

Remark that the preconditioner we consider is the inverse of the best approximation of  $A^T$  (with respect  $\|\cdot\|_F$ ) into the subspace  $\mathcal{U}$ , thus we expect that it is a good choice. In particular, we will report some numerical experiences which show that at least one choice for  $U$  can make  $\mathcal{U}_{A^T}^{-1}$  a  $\sigma$ -preconditioner. We shall refer to (9) as the ER $_{\mathcal{U}}$  method.

Let us obtain an efficient implementation of the scheme (9). Observe that the linearity of  $X \mapsto \mathcal{U}_X$  implies the identity  $\mathcal{U}_{A^T} = I - \alpha \mathcal{U}_{P^T}$ . Thus the generic iteration (9) takes the form

$$\mathbf{x}_{k+1} = \mathbf{x}_k + U(I - \alpha D_{P^T})^{-1} U^* (\mathbf{v} - A^T \mathbf{x}_k).$$

Set  $\hat{\mathbf{u}} = U^* \mathbf{u}$  for a generic  $n \times 1$  vector  $\mathbf{u}$ . Then left-multiplying by  $U^*$  the expression obtained, we have the following updating formula

$$\hat{\mathbf{x}}_{k+1} = \hat{\mathbf{x}}_k + (I - \alpha D_{P^T})^{-1} (\hat{\mathbf{v}} - \hat{\mathbf{x}}_k + \alpha U^* P^T U \hat{\mathbf{x}}_k),$$

where  $D_{P^T}$  is the diagonal matrix that characterizes  $\mathcal{U}_{P^T}$  in  $\mathcal{U}$ . Note that if  $P$  changes and/or its dimension increases then the essential quantities to define  $D_P$  can be in real time updated at very low cost (this is true for a particular  $U$ , see Section 7).

Let  $s : \mathbb{C}^n \rightarrow \mathbb{R}^+$  indicate a stopping criterion for the  $k$ th approximant; the algorithm below is a possible implementation of the  $\text{ER}_{\mathcal{U}}$  method for the  $W$ -pagerank computation:

---

(A1) An implementation of the  $\text{ER}_{\mathcal{U}}$  algorithm (9)

---

**Compute**  $\hat{\mathbf{v}} = U^* \mathbf{v}$  and  $D_Q = (I - \alpha D_{P^T})^{-1}$ .

**Choose**  $\hat{\mathbf{x}} \in \mathbb{C}^n$  and a tolerance  $\tau \in \mathbb{R}^+$ .

**Set**

$$\mathbf{y} \leftarrow U \hat{\mathbf{x}}$$

$$\mathbf{z} \leftarrow P^T \mathbf{y}$$

$$\mathbf{y} \leftarrow U^* \mathbf{z}$$

$$\mathbf{y} \leftarrow \hat{\mathbf{v}} - \hat{\mathbf{x}} + \alpha \mathbf{y}$$

$$\hat{\mathbf{x}} \leftarrow \hat{\mathbf{x}} + D_Q \mathbf{y}$$

**until**  $s(\hat{\mathbf{x}}) < \tau$ . **Then** set  $\mathbf{y} = U \hat{\mathbf{x}}$ , and propose  $\frac{\mathbf{y}}{\|\mathbf{y}\|_1}$  as an approximation of the  $W$ -pagerank vector  $\mathbf{p} = \frac{\mathbf{x}}{\|\mathbf{x}\|_1}$

---

In order to estimate the number of multiplicative operations that each step of such Algorithm requires, we introduce the notation

$$\chi(A) = \# \text{ multiplicative operations required to compute } A.$$

Thus one can compute  $\hat{\mathbf{w}}$ , given  $\mathbf{w}$  and  $U$ , with  $\chi(U^* \mathbf{w})$  multiplicative operations, or  $D_{P^T}$  with  $\chi(D_{P^T})$ , given  $P$  and  $U$ . Then, if  $\mathbf{y}$  is a generic  $n \times 1$  vector, it follows that  $p(n) = \chi(P^T \mathbf{y}) = d_{P^T} n$  and that the complexity of Algorithm  $\text{ER}_{\mathcal{U}}$  is  $\chi(P^T \mathbf{y}) + \chi(U^* \mathbf{y}) + \chi(U \mathbf{y}) + O(n)$  for each step, plus  $\chi(D_{P^T}) + \chi(U^* \mathbf{y}) + \chi(U \mathbf{y}) + O(n)$  only once.

**Definition 6.2.** We say that  $Q$  is a *low complexity spectral preconditioner*, or briefly a *LC $\sigma$ -preconditioner*, if  $Q$  is a spectral preconditioner that can be computed with  $O(n)$  multiplications and such that the computation of  $\Phi_{A^T, \mathbf{v}}(Q, \mathbf{x}_k)$  also requires  $O(n)$  multiplications.

The question is to find an unitary matrix  $U$  such that the preconditioner  $\mathcal{U}_{A^T}^{-1}$  is a *LC $\sigma$ -preconditioner*. For instance, if  $U$  is any Householder unitary matrix then it is easy to show that  $\mathcal{U}_{A^T}^{-1}$  is a *LC-preconditioner*. Here we show, keeping in high esteem experimental results, that at least a choice of  $U$  makes the matrix  $\mathcal{U}_{A^T}^{-1}$  a  $\sigma$ -preconditioner (see Section 8) which, for any  $\theta > 1$ , can be computed with less than  $O(n^\theta)$  multiplicative operations and gives rise to a method whose computational complexity each iteration is less than  $O(n^\theta)$ , too. Let us prove this fact.

First observe that if  $\{\Psi_1^U, \dots, \Psi_n^U\} = \Psi^U$  is an orthogonal basis for  $\mathcal{U}$ , with respect to the inner product  $\langle \cdot, \cdot \rangle_F$ , then for any  $X \in \mathbf{M}_n(\mathbb{C})$ , there exist  $\phi_1^X, \dots, \phi_n^X$  such that  $\mathcal{U}_X = \sum_{i=1}^n \phi_i^X \Psi_i^U$ . Also observe that, by the known Hilbert projection theorem,  $\mathcal{U}_X$  can be characterized as the unique matrix

of  $\mathcal{U}$  such that

$$\langle \Psi_i^U, \mathcal{U}_X - X \rangle_F = 0, \quad i = 1, 2, \dots, n.$$

Therefore, by the orthogonality of the basis  $\Psi^U$  we have the following formula for  $\phi_i^X$ :

$$\phi_i^X = \frac{\langle \Psi_i^U, X \rangle_F}{\langle \Psi_i^U, \Psi_i^U \rangle_F}, \quad i = 1, 2, \dots, n. \tag{10}$$

Now introduce the Fourier matrix, i.e.

$$F = \frac{1}{\sqrt{n}} \left( \omega_n^{(k-1)(j-1)} \right)_{k,j=1}^n, \quad \omega_n = e^{-\frac{2\pi i}{n}}.$$

We have the following:

**Proposition 6.3.** Consider the case  $U = F$  or, in other words, the case in which  $\mathcal{U}$  is the space  $\mathcal{F}$  of circulant matrices. Then

1. For any  $n \times 1$  vector  $\mathbf{y}$ ,  $\chi(U^*\mathbf{y}) = \chi(U\mathbf{y}) = O(n \log_2 n)$ , by the well-known fast Fourier transform algorithm.
2.  $\Psi_i^U = \Pi^{i-1}$  where  $\Pi$  is the  $n \times n$  matrix such that  $\Pi^i = \Pi^{i \bmod n}$  and

$$(\Pi)_{ij} = \begin{cases} 1 & j = i + 1 \pmod n \\ 0 & \text{otherwise} \end{cases}.$$

3.  $\phi_i^X = \frac{1}{n} \langle \Pi^{i-1}, X \rangle_F$ ,  $i = 1, \dots, n$ . Moreover  $(D_X)_{ii}$  is the Discrete Fourier Transform of  $\phi_i^X$ , i.e.  $(D_X)_{ii} = \left( n^{\frac{1}{2}} U \phi^X \right)_i$ .
4. Using the previous formulas, by the sparsity of  $P$ , one can compute the set  $\{\phi_i^P\}_{i=1}^n$  with  $O(n)$  multiplications, and then  $D_{P^T}$  via the formula  $D_{P^T} = \sqrt{n} F^* \overline{\phi^P}$ , so that  $\chi(D_{P^T}) = O(n \log_2 n)$ .
5. The iterative  $ER_{\mathcal{U}}$  method can be implemented via Algorithm ( $\mathcal{A}1$ ) with

$$\chi(P^T \mathbf{y}) + \chi(U\mathbf{y}) + \chi(U^* \mathbf{y}) + O(n) = O(n \log_2 n)$$

multiplicative operations for each step, plus  $\chi(D_{P^T}) + \chi(U^* \mathbf{v}) + \chi(U\mathbf{y}) + O(n) = O(n \log_2 n)$  only once. Notice that  $\chi(U^* \mathbf{v}) = 0$  if  $\mathbf{v} = \frac{1}{n} \mathbf{e}$ .

**Proof.** Observe that, if  $\bar{x}$  is the conjugate of  $x$ , then  $(F^* P^T F)_{ii} = (F \bar{P} F)_{ii} = \overline{(F^* P F)_{ii}}$ . Moreover, since  $P$  is real and there exists a real basis for the circulant matrices (precisely  $\{\Pi^i\}_{i=0, \dots, n-1}$ ), the matrix  $\mathcal{F}_P$  is real too [24]. As a consequence  $\mathcal{F}_{P^T} = \mathcal{F}_P^* = \mathcal{F}_P^T$ , or, in other words,  $D_{P^T} = \overline{D_P}$ . This proves 4. Points 1, 2 and 3 are well-known facts, see, for instance, [24] or [25]. Finally, for 5, just note that  $F^* \mathbf{v} = \frac{1}{n} F^* \mathbf{e} = \frac{1}{\sqrt{n}} \mathbf{e}_1$ , thus no operation is required to compute such matrix-vector multiplication.  $\square$

Thus we propose the circulant matrix  $\mathcal{F}_{A^T}^{-1}$  as a preconditioner for (4).

### 6.1. On the complexity and the use of the circulant preconditioner

As a consequence of Proposition 6.3, by choosing  $U = F$  we obtain a preconditioner that can be computed with  $O(n \log_2 n)$  multiplications and such that the new approximation  $\mathbf{x}_{k+1} = \Phi_{A^T, \mathbf{v}} \left( \mathcal{F}_{A^T}^{-1}, \mathbf{x}_k \right)$

can be computed from  $\mathbf{x}_k$  with  $O(n \log_2 n)$  multiplications. Thus, applying the preconditioning matrix  $\mathcal{F}_{A^T}^{-1}$  as well as computing the eigenvalues of such preconditioner requires a surplus of operations. However, as underlined in [3] it is often convenient to spend more effort to decrease the running time of the solving method when the same matrix is used many times to solve different problems. In the Google case, for instance, the same matrix  $P$  is also required for personalized web search [26] or in the case of the heuristic for collusion-proof pagerank [27, 3]. Moreover the interest of the method just introduced, consists in its convergence rate, as discussed afterwards.

Let us obtain an estimate on how much the complexities of a single iteration of the  $ER_{\mathcal{F}}$  algorithm, implemented as in (A1), and of the power method differ. Suppose that  $n = 2^\gamma$  is the order of  $P$  or rather  $\log_2(n) = \gamma$ . Therefore the floating point multiplications required by the  $ER_{\mathcal{F}}$  algorithm each step are  $\chi(P^T \mathbf{y}) + 2\chi(F\mathbf{y}) = (2\gamma + d_p)n$ . Therefore the method requires about  $\left(1 + \frac{2\gamma}{d_p}\right)$  times the multiplicative operations required by the power method. For what concerns memory usage, four complex vectors are needed to implement  $ER_{\mathcal{F}}$ , thus four times the memory allocations required to implement the power method.

Further studies could reduce the complexity per step of the  $ER_{\mathcal{U}}$  method and the required memory allocations. In particular, a  $LC\sigma$ -preconditioner can be probably found into a subspace of  $\mathbf{M}_n(\mathbb{C})$ , for instance, there are lot of other interesting choices of algebras  $\mathcal{U}$  (different from circulants) associated with fast discrete transforms  $U$ . Besides the trigonometric and Hartley-type possible choices of  $U$ , all giving rise to fast discrete transforms (see, for instance, [24, 28–30]), one can simply consider Householder real unitary matrices. In such case it is easy to obtain a  $LC$ -preconditioner. Note that all such alternative choices of  $U$  reduce the required memory allocations from 4 complex to 4 real vectors.

6.2. About the convergence of the method

In the following theorem, we introduce a sufficient condition for the convergence of the  $ER_{\mathcal{U}}$  method, and in particular, of the  $ER_{\mathcal{F}}$  method. However, much more faithful convergence rate estimations will be deduced in Section 8, as a consequence of some observations and experimental results.

**Theorem 6.4.** *Let  $A, \mathbf{v}, \mathbf{x}$  and  $\alpha$  be as in (1). Let  $\beta$  be such that  $0 < \beta < 1, \{\mathbf{x}_k\}_k$  the sequence in (9), and*

$$\gamma_{A,U}^{(1)} = \frac{\|U^* A^T U\|_1}{\|A^T\|_1} \tag{11}$$

for a given unitary matrix  $U$ . If

$$\min_{i=1, \dots, n} |(U^* A^T U)_{ii}| \geq \gamma_{A,U}^{(1)} \frac{1 + \alpha}{1 + \beta} \tag{12}$$

then  $\mathbf{x}_k \xrightarrow{k \uparrow \infty} \mathbf{x}$  and the convergence rate of  $ER_{\mathcal{U}}$  algorithm, as implemented in (A1), is dominated by  $O(\beta^k)$ .

**Proof.** Given  $X \in \mathbf{M}_n(\mathbb{C})$ , set  $\widehat{X} = U^* X U$  and

$$\mathcal{G}_j(X) = \left\{ \mathbf{x} : |x - (X)_{jj}| \leq \sum_{\substack{i=1 \\ i \neq j}}^n |(X)_{ij}| \right\}$$

for  $j = 1, 2, \dots, n$ .

Left multiply  $\mathbf{x}_{k+1}$  (defined in (9)) by  $U^*$ , to obtain

$$\widehat{\mathbf{x}}_{k+1} = \Phi_{\widehat{A^T}, \widehat{\mathbf{v}}} \left( \text{diag}(\widehat{A^T})^{-1}, \widehat{\mathbf{x}}_k \right) \tag{13}$$

(note that  $U^*A^T\mathbf{x}_k = (U^*A^TU)(U^*\mathbf{x}_k) = \widehat{A^T}\widehat{\mathbf{x}}_k$ ). Moreover (9) converges if and only if (13) converges (as  $k$  diverges). Observe that the quasi-stochasticity of  $P$  implies

$$\|A\|_\infty \leq 1 + \alpha,$$

where the equality holds if and only if there exists at least one index  $i \in \{1, \dots, n\}$  such that  $(P)_{ii} = 0$  and the  $i$ th row of  $P$  is not null. As a consequence

$$\sum_{\substack{i=1 \\ i \neq j}}^n \left| (\text{diag}(\widehat{A^T})^{-1}\widehat{A^T})_{ij} \right| = \sum_{\substack{i=1 \\ i \neq j}}^n \left| \frac{(\widehat{A^T})_{ij}}{(\widehat{A^T})_{ii}} \right| \leq \frac{\|\widehat{A^T}\|_1}{\min_i |(\widehat{A^T})_{ii}|} - 1 \leq \beta$$

where  $(\text{diag}(\widehat{A^T})^{-1}\widehat{A^T})_{jj} = 1$  for all  $j = 1, \dots, n$ . It follows that

$$\mathcal{G} = \bigcup_{i=1}^n \mathcal{G}_i (\text{diag}(\widehat{A^T})^{-1}\widehat{A^T}) \subseteq \overline{B_\beta(1)}. \tag{14}$$

By the known Gershgorin theorem, all the eigenvalues of  $\text{diag}(\widehat{A^T})^{-1}\widehat{A^T}$  are in  $\mathcal{G}$ . Then by (14) it results that the spectral radius of the iteration matrix of the sequence (9) is less than or equal to  $\beta$ , or rather that such sequence converges with a  $O(\beta^k)$  rate. Finally observe that  $\widehat{\mathbf{x}} = \lim_{k \rightarrow \infty} \widehat{\mathbf{x}}_k$  solves the linear system  $\widehat{\mathbf{v}} = \widehat{A^T}\widehat{\mathbf{x}}$ . So the solution for (1) is obtained by the identity  $\mathbf{x} = U\widehat{\mathbf{x}}$ .  $\square$

Remark that under convergence hypothesis, the limit  $\mathbf{x}$  of the sequence  $\Phi_{A^T, \mathbf{v}}(\mathcal{U}_{A^T}^{-1}, \mathbf{x}_k)$  is solution of the real linear system (1); thus it will be a real vector, even if (9) defines a sequence of complex vectors, as in the case  $U = F$ .

Before to proceed with the study of the preconditioner  $\mathcal{U}_{A^T}^{-1}$  with  $U = F$ , let us observe that other approximations of  $A^T$  could be chosen in  $\mathcal{U}$ . In particular, in [31,32] it is shown that  $\mathcal{U}_X$  is not necessarily the best possible preconditioner in  $\mathcal{U}$  for a linear system  $X\mathbf{z} = \mathbf{b}$ . More precisely, denoted with  $\mathcal{R}^{(m)}$  the set of all rank  $m$  matrices ( $m$  constant with respect to  $n$ ), and with  $\widetilde{\mathcal{U}}_X + \widetilde{\mathcal{R}}_X^{(m)}$  the matrix solving the  $\mathcal{U} + \mathcal{R}^{(m)}$  approximation problem of  $X$  [31], the authors show that  $\widetilde{\mathcal{U}}_X$  works as a preconditioner better than  $\mathcal{U}_X$  provided  $U = F$  and  $X$  is Toeplitz. It would be interesting to investigate if the result is yet true when  $X$  is our matrix  $A^T = I - \alpha P^T$ . Moreover, for  $X = I - \alpha P^T$  one should also investigate if there are algebras  $\mathcal{U}$  closer to  $X$  than  $\mathcal{F}$ , and thus preconditioners  $\mathcal{U}_X$  ( $\widetilde{\mathcal{U}}_X$ ) better than  $\mathcal{F}_X$  ( $\widetilde{\mathcal{F}}_X$ ).

### 7. Updating the circulant preconditioner

Observe that, by the particular structure of  $P$ , the following formulas hold, for a generic  $n \times 1$  vector  $\mathbf{y}$

$$\mathbf{y}^*P\mathbf{y} = \sum_{k=1}^n \overline{y_k}(P\mathbf{y})_k = \sum_{k: v(k) \neq \emptyset} \overline{y_k} \frac{1}{\mu_k} \sum_{s \in v(k)} y_s, \tag{15}$$

$$\mathbf{y}^*P\mathbf{y} = (P^T\mathbf{y})^*\mathbf{y} = \sum_{k: \rho(k) \neq \emptyset} \overline{y_k} \sum_{s \in \rho(k)} \frac{1}{\mu_s} \overline{y_s}. \tag{16}$$

Let us associate to  $P$  its best circulant approximation  $\mathcal{F}_P$ . It follows from the previous section that

$$\mathcal{F}_P = F \text{diag}(F^*PF)F^* = \sqrt{n}F \text{diag}(F\phi^P)F^*. \tag{17}$$

As a consequence of Propositions 6.1 and 6.3, the computation of the eigenvalues of the projection  $\mathcal{F}_{A^T}$  can be brought back to the computation of the eigenvalues of  $\mathcal{F}_P$ . By (17) it is clear that there are two formulas for the eigenvalues of  $\mathcal{F}_P$ . Let us explicit these formulas.



For the first one just observe that (15) and (16),  $\mathbf{y} = \mathbf{e}_j$ , imply

$$(F^*PF)_{jj} = \frac{1}{n} \sum_{k:v(k) \neq \emptyset} \sum_{s \in v(k)} \frac{1}{\mu_k} \omega^{(s-k)(j-1)} = \frac{1}{n} \sum_{k:\rho(k) \neq \emptyset} \sum_{s \in \rho(k)} \frac{1}{\mu_s} \omega^{(k-s)(j-1)}.$$

For the second one, note that, if  $s_i$  denote the sums of the entries on the  $i$ th diagonal of  $P$ ,  $i = -n + 1, \dots, -1, 0, 1, \dots, n - 1$  (f.i.  $\dots, s_{-1} = \sum_{i=1}^{n-1} (P)_{i+1,i}, s_0 = \sum_{i=1}^n (P)_{ii}, s_1 = \sum_{i=1}^{n-1} (P)_{i,i+1}, \dots$ ), then we have  $s_0 = 0$  and, for  $i = 1, \dots, n - 1$ ,

$$s_i = \sum_{t=1 \dots n-i, t \in \rho(t+i)} \frac{1}{\mu_t} = \sum_{t=1 \dots n-i, t+i \in v(t)} \frac{1}{\mu_t},$$

$$s_{-i} = \sum_{t=1 \dots n-i, t \in v(t+i)} \frac{1}{\mu_{t+i}} = \sum_{t=1 \dots n-i, t+i \in \rho(t)} \frac{1}{\mu_{t+i}}.$$

In order to obtain the latter formulas, use, for instance, the following remark:

$$(P\mathbf{e}_{t+i})_k = \begin{cases} \frac{1}{\mu_k} \sum_{s \in v(k)} (\mathbf{e}_{t+i})_s & v(k) \neq \emptyset \\ 0 & v(k) = \emptyset \end{cases} = \begin{cases} \frac{1}{\mu_k} & v(k) \neq \emptyset, t+i \in v(k) \\ 0 & \text{otherwise} \end{cases}.$$

The elements on the first row of  $\mathcal{F}_P$ , say  $\mathbf{c}^T = [c_0 \ c_1 \ \dots \ c_{n-1}]$ , are the coefficients of  $\mathcal{F}_P$  when written in terms of the orthogonal basis  $\{\Psi_j^F = \Pi^{j-1}\}_j$  (see Section 6). In other words the  $c_i$  are equal to the  $\phi_i^X$  in (10) when  $X = P$  and  $U = F$ , so that  $\mathbf{c} = \phi^P$  and  $\mathcal{F}_P = \sum_{i=0}^{n-1} c_i \Pi^i$ . Thus the vector  $\mathbf{c}$  can be computed via the identities

$$c_0 = s_0/n = 0, \quad c_i = (s_i + s_{-n+i})/n, \quad i = 1, \dots, n - 1,$$

and we have the second formula for the eigenvalues of  $\mathcal{F}_P$ :

$$\mathcal{F}_P = F \text{diag}(\mathbf{z}_P) F^*, \quad \mathbf{z}_P = \sqrt{n} F \mathbf{c}. \tag{18}$$

Observe that the particular structure of  $E_P$  allows us to assume that it is the adjacency matrix of an oriented graph, for instance, describing the links among the web pages. Therefore we can say that a change on  $E_P$  is nothing else than a change of the graph it describes, and vice versa. It is also clear that whenever such matrix changes, also the matrix  $P$  changes, and the preconditioner  $\mathcal{F}_P$  (or  $\mathcal{U}_P$  in general) needs to be computed again. The observations done and the latter formula (18) allow us to propose an algorithm generating the preconditioner  $\mathcal{F}_P$  for the upgraded row-constant quasi-stochastic matrix  $P$ , known the preconditioner for the current  $P$ . Note that such algorithm can be implemented very cheaply if the eigenvalues of  $\mathcal{F}_P$  are computed only any time the dimension  $n$  of  $P$  doubles (see the end of this section).

Whenever something changes in the graph that  $E_P$  describes, we have to update  $P$  and  $\mathcal{F}_P$ . Let us denote with  $P_{old} \rightarrow P_{new}$  such elementary updating, where the matrix  $P_{old}$  is the current  $P$  (we assume the eigenvalues of  $\mathcal{F}_{P_{old}}$  have been already computed) and  $P_{new}$ , the updated  $P$ , has a dimension greater by one or equal to the dimension of  $P_{old}$ . For updating  $\mathcal{F}_P$  we have to compute the new  $s_i$  in order to define the new  $c_i$  and thus the new vector  $\mathbf{z}_{P_{new}}$  (of the eigenvalues of  $\mathcal{F}_{P_{new}}$ ).

Recall that by the sparsity of  $P$  the preconditioner  $\mathcal{F}_P$  can be computed in  $O(n)$  operations from scratch (as observed in Proposition 6.3). However, updating the essential quantities, i.e. the  $s_i$ , to define the eigenvalues of  $\mathcal{F}_{P_{new}}$  requires very few operations (we show this fact below describing, case for case, how the new  $s_i$  can be computed from the old  $s_i$ , in the main situations that may occur). So, it is convenient to do only these operations for many successive elementary updates  $P_{old} \rightarrow P_{new}$  until  $P$  has reached a previously established dimension. Then compute the eigenvalues of  $\mathcal{F}_P$ .

(1) *Adding a new vertex.* This is the case in which the graph described by  $E_P$  increases by one vertex, that is  $P_{new}$  increases by a new row (and thus a new column) with respect  $P_{old}$ :

$$P_{old} \longrightarrow P_{new} = \begin{pmatrix} P_{old} & \mathbf{0} \\ \dots & \frac{1}{\mu_{n+1}} \dots 0 \end{pmatrix}$$

We suppose to enumerate with  $n + 1$  the new vertex. We have to associate to it the new objects  $\mu_{n+1}, \nu(n + 1) \subset \{1, \dots, n\}, \rho(n + 1) \subset \{1, 2, \dots, n + 1\}$  (we need new memory allocations for them). Note that in the case in which  $E_P$  describes a web graph (as in the case of Google), updating  $P$  following the birth of a new vertex implies that  $\mu_{n+1} = |\nu(n + 1)|$  is small and that  $\rho(n + 1) = \emptyset$  (since the site just born is unknown to the others, then there are no edges pointing to the new vertex yet).

Then introduce the two new numbers  $s_n, s_{-n}$  (we need other two new cells in memory for them), and set:

$$s_n := 0, s_{-n} := 0, s_{j-n-1} := s_{j-n-1} + \frac{1}{\mu_{n+1}}, j \in \nu(n + 1).$$

After this, introduce another new number  $c_n$  (we need a further new cell in memory for it), and set:

$$c_n := 0, c_j := (s_j + s_{-(n+1)+j})/n, j \in \nu(n + 1).$$

Finally, set  $n := n + 1$

(2) *An element of  $\Delta_P$  changes.* By analogy with the case  $E_P$  describes a web graph, this is the situation in which the birth of a new site coincides with the death of an old one.

Assume that the  $r$ th element of  $\Delta_{P_{old}}$  changes. Then

$$P_{old} = \begin{pmatrix} \cdot \frac{1}{\mu_r} \cdot 0 \cdot \frac{1}{\mu_r} \cdot \end{pmatrix} \longrightarrow P_{new} = \begin{pmatrix} \mathbf{0} \\ \cdot \frac{1}{\mu_r^{new}} \cdot 0 \cdot \frac{1}{\mu_r^{new}} \cdot \\ \mathbf{0} \end{pmatrix}.$$

Set

$$s_{k-r} := s_{k-r} - \frac{1}{\mu_r}, k \in \nu(r),$$

and apply the operations described in the Case 4 below, for  $j = r$  and  $\forall i \in \rho(r)$ . Moreover we must define new  $\mu_r, \nu(r) \subset \{1, \dots, n\} \setminus \{r\}$ , and  $\rho(r), \rho(r) = \emptyset$ . Then set

$$s_{k-r} := s_{k-r} + \frac{1}{\mu_r}, k \in \nu(r),$$

$$c_{k-r} = (s_{k-r} + s_{-n+k-r})/n, k \in \nu(r), k - r > 0,$$

$$c_{n+k-r} = (s_{n+k-r} + s_{k-r})/n, k \in \nu(r), k - r < 0.$$

(3) *An element of  $E_P$  changes from 0 to 1.* Assume that the  $i$ th element of  $E_P$  changes from 0 to 1. This is the case in which a new edge from the vertex  $i$  to the vertex  $j$  is added to the graph. It is clear that this implies that also the  $i$ th element of  $\Delta_P$  changes, since  $(\Delta_P)_{ii} = (E_P \mathbf{e})_i = \mu_i$ . Precisely it increases by one, i.e.

$$P_{old} = \begin{pmatrix} \dots \frac{1}{\mu_i} \dots 0 \dots \end{pmatrix} \longrightarrow P_{new} = \begin{pmatrix} \dots \frac{1}{\mu_i+1} \dots \frac{1}{\mu_i+1} \dots \end{pmatrix}$$

The following operations must be done:

$$s_{j-i} := s_{j-i} + \frac{1}{\mu_i + 1}, \quad s_{k-i} := s_{k-i} + \frac{1}{\mu_i + 1} - \frac{1}{\mu_i}, \quad k \in \nu(i).$$

$$\begin{cases} c_{j-i} = (s_{j-i} + s_{-n+j-i})/n, & \text{if } j - i > 0, \text{ or} \\ c_{n+j-i} = (s_{n+j-i} + s_{j-i})/n, & \text{if } j - i < 0, \end{cases}$$

$$\begin{cases} c_{k-i} = (s_{k-i} + s_{-n+k-i})/n, & k \in \nu(i), \quad k - i > 0, \\ c_{n+k-i} = (s_{n+k-i} + s_{k-i})/n, & k \in \nu(i), \quad k - i < 0. \end{cases}$$

Then set  $\mu_i := \mu_i + 1$ ,  $\nu(i) := \nu(i) \cup \{j\}$ ,  $\rho(j) := \rho(j) \cup \{i\}$ . The latter identities require reordering (forward shift).

(4) An element of  $E_P$  changes from 1 to 0. This case, similarly to the case described before, corresponds to the situation in which the edge from  $i$  to  $j$  is removed from the graph, i.e. the  $ij$ th element of  $E_P$  changes from 1 to 0. As observed before, this implies that the  $i$ th element of  $\Delta_P$  changes too. Precisely it decreases by one:

$$P_{old} = \left( \begin{array}{cccc} \dots & \frac{1}{\mu_i} & \dots & \frac{1}{\mu_i} & \dots \end{array} \right) \longrightarrow P_{new} = \left( \begin{array}{cccc} \dots & \frac{1}{\mu_i - 1} & \dots & 0 & \dots \end{array} \right)$$

The operations required in this case are similar to the operations required by the Case 3. Precisely

$$s_{j-i} := s_{j-i} - \frac{1}{\mu_i}, \quad s_{k-i} := s_{k-i} + \frac{1}{\mu_i - 1} - \frac{1}{\mu_i}, \quad k \in \nu(i) \setminus \{j\},$$

$c_{j-i}$ ,  $c_{n+j-i}$  as in Case 3, and  $c_{k-i}$ ,  $c_{n+k-i}$  as in Case 3 except that  $k$  runs on  $\nu(i) \setminus \{j\}$ .

Then set  $\mu_i := \mu_i - 1$ ,  $\nu(i) := \nu(i) \setminus \{j\}$ ,  $\rho(j) := \rho(j) \setminus \{i\}$ . The latter identities require reordering (backward shift).

**Remark 7.1.** Note that, after Case 1 or Case 2 has been run, the site  $i^*$  which has created the new site ( $n$  in Case 1 and  $r$  in Case 2) must create a link to such new site. In other words, after Case 1 or Case 2 we have to run Case 3 for  $j = n$  and  $i = i^* \in \{1, \dots, n - 1\}$  or for  $j = r$  and  $i = i^* \in \{1, \dots, n\} \setminus \{r\}$ .

*The simultaneous updating of the matrix P and of its circulant projection  $\mathcal{F}_P$*

Let  $N$  be a fixed natural number. By the following procedure one could generate a transition matrix  $P$  of order  $N$  and, simultaneously, its best circulant approximation  $\mathcal{F}_P$  (or, more precisely: non-negative integers  $\mu_1, \dots, \mu_N$  and sets  $\nu(1), \dots, \nu(N)$  defining  $P$ ; and real numbers  $s_{-N+1}, \dots, s_0, \dots, s_{N-1}$ ,  $c_0, \dots, c_{N-1}$ ,  $z_1, \dots, z_N$  defining  $\mathcal{F}_P = F \text{diag}(\mathbf{z})F^*$ ).

Consider an initial matrix  $P$  of small size (i.e. choose a small  $n$  and define  $\mu_i, \nu(i), i = 1, \dots, n$ ). Then

- Apply to  $P$  the operators Case (1)–(3), Case (2)–(3),(3),(4) repeatedly, in suitable order, each possibly more times, until  $n$  is equal to  $N$ . During this phase, for what concerns  $\mathcal{F}_P$ , update only the  $s_i$ . (Note that the operator Case (2)–(3) requires  $|\rho(r)|$  applications of (4); so it is expensive if  $|\rho(r)|$  is large. However, the death of  $r$  means generally a small  $|\rho(r)|$ .)
- When  $n = N$  compute also the first row  $\mathbf{c}^T$  of  $\mathcal{F}_P$  and the vector  $\mathbf{z} = \sqrt{n}F\mathbf{c}$  of the eigenvalues of  $\mathcal{F}_P$

Note that it is advisable to choose  $N$  as a power of 2, so that the FFT involved (in computing the vector  $\mathbf{z}$  defining the preconditioner  $\mathcal{F}_P$ , and in each step of the ER $\mathcal{F}$  method) are more efficient. Note

also that the above procedure can be repeated starting from the new  $P$ . This procedure will be applied to generate test matrices  $P_N, P_{2N}, P_{4N}, \dots$ , in a future work, in order to make a deeper experimental investigation of the  $ER_{\mathcal{F}}$  efficiency in solving  $W$ -pagerank problems larger than those considered in the following section.

### 8. Numerical experiences and results

In this section, we study the efficiency of the  $ER_{\mathcal{F}}$  algorithm by the help of numerical experiences. In particular, we bear out the claim that  $\mathcal{F}_{A^T}^{-1}$  is a  $\sigma$ -preconditioner.

#### 8.1. Numerical estimations of the efficiency and of the convergence rate

The plots in Fig. 2 and the values in Table 1 show how the spectral radii of the iteration matrices  $H = H_I = I - A^T$  and  $H_{\mathcal{F}_{A^T}^{-1}} = I - \mathcal{F}_{A^T}^{-1}A^T$  differ. They are made using a randomly generated each time different matrix  $P$ , satisfying the properties requested.

In particular, they show that the particular quasi-stochastic row-constant structure of  $P$  seems to ensure that the spectrum  $\Lambda(H)$  consists of a small cluster round 0 and a single eigenvalue whose magnitude is close to  $\alpha$ , therefore only one eigenvalue of  $A$  is responsible of the low rate of convergence of power method (or rather ER method).

Call  $\lambda_\alpha$  ( $|\lambda_\alpha| \approx \alpha$ ), the maximal eigenvalue of  $H = \alpha P$  and  $\Gamma_\alpha(d_p)$  the smallest open disc in  $\mathbb{C}$  containing all the eigenvalues of  $H$  except for  $\lambda_\alpha$ ; then we have  $\Lambda(H) \subset \Gamma_\alpha(d_p) \sqcup \{\lambda_\alpha\}$ . The numerical experiments show that  $\Gamma_\alpha$  essentially does not depend on  $\alpha$  but just on the sparsity of the matrix, or rather on the density  $d_p$ . Namely we observe that  $\Gamma_\alpha(x) \subset \Gamma_\alpha(y)$  whenever  $x > y$ , and  $\Gamma_\alpha(x)$  essentially does not change if  $\alpha$  moves into  $[0.8, 1)$ . We believe that  $\Gamma(d_p)$  is a family of decreasing subsets such that  $\Gamma(d_p) \downarrow \{0\}$  as  $d_p \rightarrow n$ . See Fig. 2 or Table 1 for clearness. The slow convergence rate of Euler–Richardson method only depends on  $\lambda_\alpha$ ; but observe that such eigenvalue can be removed by preconditioning  $A$ , in fact the spectrum  $\Lambda\left(H_{\mathcal{F}_{A^T}^{-1}}\right)$  is all contained into a small cluster round 0, not much different from  $\Gamma_\alpha(d_p)$ . That is,  $\mathcal{F}_{A^T}^{-1}$  seems to be effectively a  $\sigma$ -preconditioner.

This implies that the  $ER_{\mathcal{F}}$  method is more efficient in terms of convergence rate than the power method. In particular, let  $\beta = \beta(\alpha)$  be the spectral radius of the preconditioned iteration matrix  $H_{\mathcal{F}_{A^T}^{-1}}$ . We underline that, accordingly with the experimental results, we have  $\beta(1 - \varepsilon) < 0.3$ , for all  $1 > \varepsilon > 0$  and that  $\beta(\alpha)$  is an increasing (bounded) function of  $\alpha$ , whereas  $|\lambda_\alpha| = \rho(H) = \rho(I - A)$  is always very close to  $\alpha$ .

To ensure a numerical precision  $\tau$  we need at least  $k_{\mathcal{F}} = \log(\tau) / \log(\beta)$  iterations of the  $ER_{\mathcal{F}}$  algorithm, or  $k_{\mathbf{R}} = \log(\tau) / \log(|\lambda_\alpha|)$  iterations of the ER (power) method. Therefore

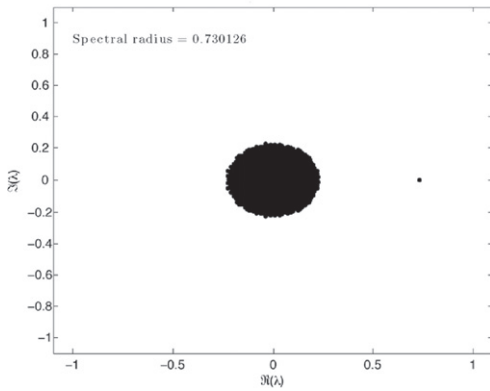
**Proposition 8.1.** *Let  $\chi(\mathbf{R})$  and  $\chi(\mathcal{F})$  denote, respectively, the number of multiplicative operations required for each step by the power method and by the  $ER_{\mathcal{F}}$  method. Let  $\beta$  and  $\lambda_\alpha$  be as defined above. If*

$$|\lambda_\alpha| \frac{\chi(\mathcal{F})}{\chi(\mathbf{R})} > \beta \tag{19}$$

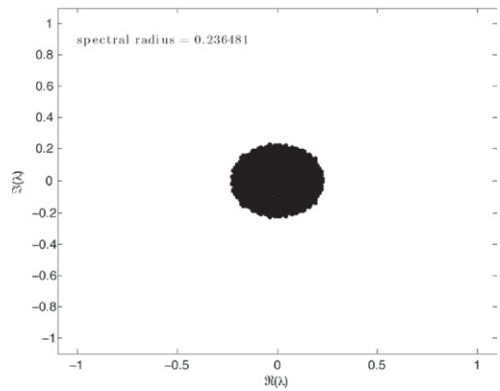
then the  $ER_{\mathcal{F}}$  method produces an approximation of the  $W$ -pagerank with less multiplicative operations than the power method.

**Proof.** It is an immediate consequence of the fact that (19) holds if and only if  $\chi(\mathcal{F})k_{\mathcal{F}} < \chi(\mathbf{R})k_{\mathbf{R}}$ .  $\square$

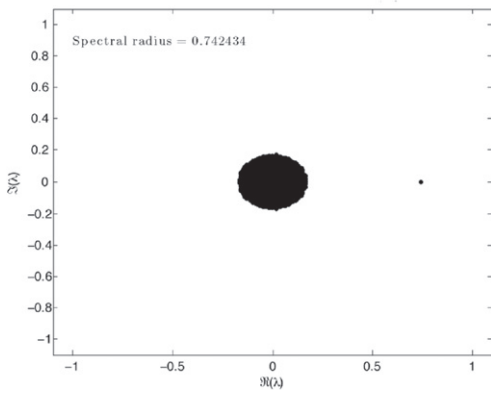
For instance (see Table 1) if  $\alpha = 9/10$  and  $d_p = 15$  then  $\beta \approx 0.23$  and  $|\lambda_\alpha| \approx 0.75$ . Taking no notice of additive operations, we observed in the previous section that  $\chi(\mathcal{F})/\chi(\mathbf{R}) \approx 1 + \frac{2\gamma}{d_p}$ , being  $\gamma = \lceil \log_2 n \rceil$ . Since  $\frac{\log \beta}{\log |\lambda_\alpha|} \approx 5.1$ , condition (19) is verified for all positive  $\gamma$  such that  $\gamma < 31$ .



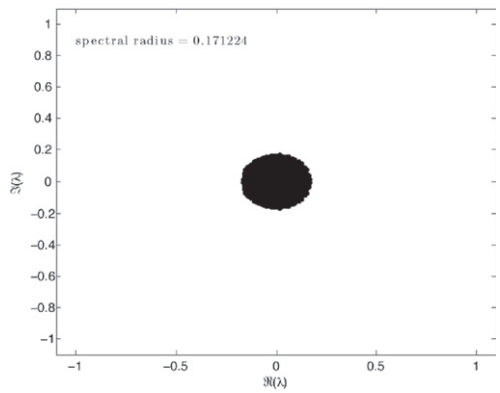
(a)  $\alpha = 0.9, n = 2000, d_P = 14$



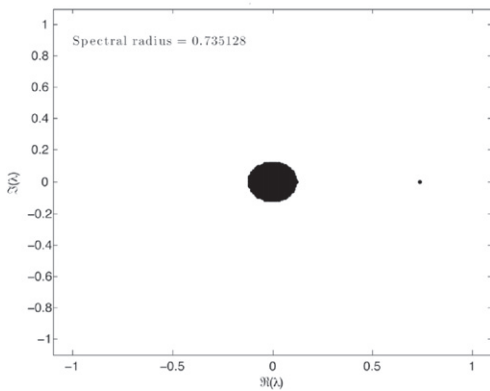
(b)  $\alpha = 0.9, n = 2000, d_P = 14$



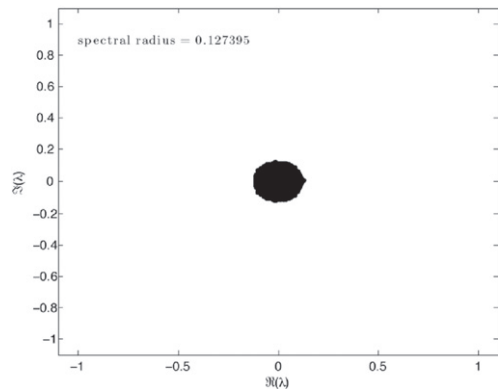
(a)  $\alpha = 0.9, n = 2000, d_P = 25$



(b)  $\alpha = 0.9, n = 2000, d_P = 25$

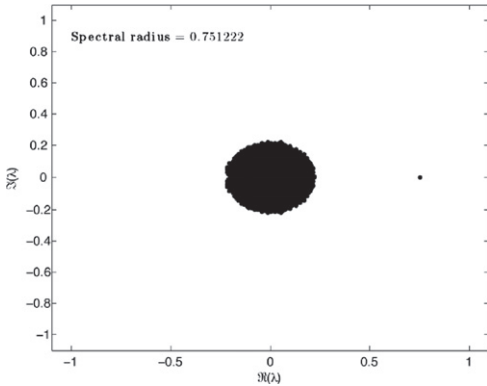


(a)  $\alpha = 0.9, n = 2000, d_P = 50$

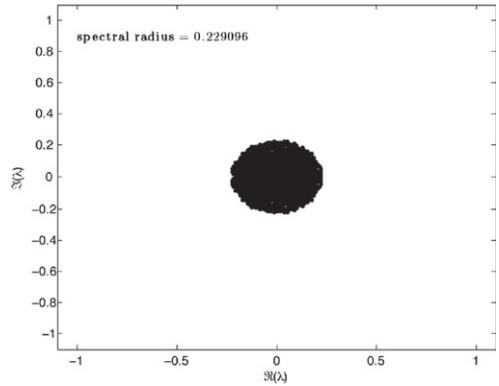


(b)  $\alpha = 0.9, n = 2000, d_P = 50$

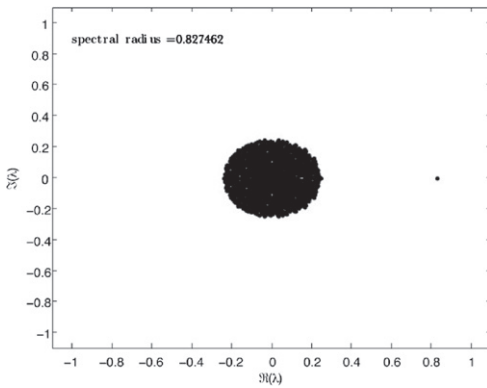
**Fig. 2.** The figures show the spectra of  $H = I - A^T$  (on the left) and of  $H_{\mathcal{F}_{A^T}^{-1}} = I - \mathcal{F}_{A^T}^{-1}A^T$  (on the right), for different values of  $\alpha, n$  and the density  $d_P$ . Note that the dimension of the cluster does not depend on  $n$  but only on  $d_P$ , nevertheless it nearly does not change when  $\alpha$  moves through 0.9, 0.99, and also (for brevity, not reported in figure) 0.85.



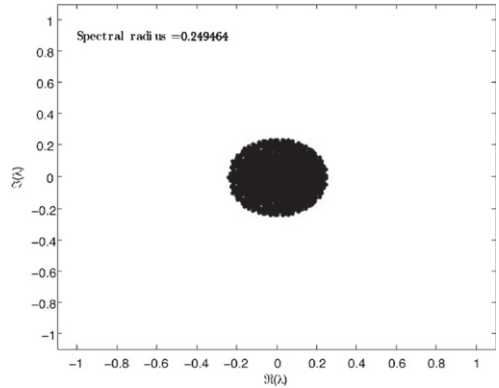
(a)  $\alpha = 0.9, n = 1500, d_P = 15$



(b)  $\alpha = 0.9, n = 1500, d_P = 15$



(a)  $\alpha = 0.99, n = 1500, d_P = 15$



(b)  $\alpha = 0.99, n = 1500, d_P = 15$

Fig. 2. (continued)

Moreover by the results in Table 1 it is clear that, fixed  $\gamma$  (i.e. fixed the order  $n$ ), condition (19) still holds when  $d_P$  or  $\alpha$  increase with respect the one here we consider. Nevertheless, the less null rows (or columns) are in  $P$ , the more  $|\lambda_\alpha|$  is close to  $\alpha$  and (19) is satisfied for smaller values of  $d_P$  or of  $\alpha$ .

Another nodal point concerns the memory allocations required by the method. In spite of the strong advantages in terms of convergence rate obtained by preconditioning power method as in the  $ER_{\mathcal{F}}$  algorithm, the latter novel method needs more data storage each iteration. It is well known that the low complexity implementation of power method needs only two real vector to be stored each iteration [1, pp. 77–78]. On the contrary, as well as Krylov subspace methods (applied to the Google case in [3,4,6], for instance), the  $ER_{\mathcal{F}}$  method requires more then two vectors to be stored. In the present form four complex vectors need to be stored, which results in a four-times increasing of memory usage. Note that the  $ER_{\mathcal{L}}$  algorithm can be implemented by using less memory allocations, for

Table 1

How the spectral radius of the iteration matrix changes due to preconditioning.

$n$	$\alpha$	$d_P$	$\rho(I - A^T)$	$\rho(I - \mathcal{F}_{A^T}^{-1}A^T)$
2000	0.9	14	0.730126	0.236481
2000	0.9	25	0.742434	0.171224
2000	0.9	50	0.735128	0.127395
1500	0.9	15	0.751222	0.229036
1500	0.99	15	0.827462	0.249464

instance, by choosing  $U$  as a real matrix. Anyway the aim of this work is to show how convergence rate advantages can be assured by preconditioning the power method. Finally, notice that the approach we are proposing (to precondition the power method) is substantially different from the residual Krylov subspace minimization idea [4,6].

### 8.2. Performance on test problems

This subsection is devoted to the choice of an easily computable stopping criterion for the  $ER_{\mathcal{F}}$  algorithm. Let us introduce the following vectors:

$$\begin{aligned} r_{\mathbf{R}}(k) &= \mathbf{v} - A^T \mathbf{x}_k, \\ r_{\mathcal{F}}(k) &= \mathbf{v} - A^T (F\widehat{\mathbf{x}}_k), \\ \widehat{r}_{\mathcal{F}}(k) &= \widehat{\mathbf{v}} - \widehat{A}^T \widehat{\mathbf{x}}_k. \end{aligned}$$

Note that  $r_{\mathbf{R}}$  and  $r_{\mathcal{F}}$  coincide. A classic way to stop the power method is to evaluate  $\|r_{\mathbf{R}}(k)\|$  for a suitable norm (generally  $L^1$  or  $L^2$ ). In order to establish a stopping criterion for the  $ER_{\mathcal{F}}$  method, it would be natural to consider  $\|r_{\mathcal{F}}\|$ , and to compare it with  $\|r_{\mathbf{R}}\|$ . However in the  $ER_{\mathcal{F}}$  algorithm the evaluation of  $r_{\mathcal{F}}$  is more expensive than the evaluation of  $\widehat{r}_{\mathcal{F}}$ . In particular, it requires one more discrete Fourier transformation each step. Anyway it can be easily observed that  $\|\widehat{r}_{\mathcal{F}}(k)\|_2 = \|r_{\mathcal{F}}(k)\|_2$ , and hence that

$$\max \left\{ \|r_{\mathcal{F}}(k)\|_2, \frac{\|r_{\mathcal{F}}(k)\|_1}{\sqrt{n}} \right\} \leq \|\widehat{r}_{\mathcal{F}}(k)\|_1 \leq \sqrt{n} \min \{ \|r_{\mathcal{F}}(k)\|_1, \|r_{\mathcal{F}}(k)\|_2 \}. \tag{20}$$

Therefore the use of  $L^2$  metric allows to avoid the additional operations that the computation of  $r_{\mathcal{F}}$  requires, i.e. one can check the value of  $\widehat{r}_{\mathcal{F}}$  in order to decide to stop the  $ER_{\mathcal{F}}$  algorithm. Alternatively, (20) shows that also the metric of  $L^1$  can ensure the same helpful property, at least for  $k$  large. In fact  $\|r_{\mathcal{F}}\|_1$  small enough (i.e.  $k$  large enough) implies  $\sqrt{n} \|r_{\mathcal{F}}\|_1$  small, especially for  $n$  not too large. Finally we underline that one can use other types of stopping criterion, for example the same used in solving the Google-pagerank problem. In particular, the classical step-reduction norm  $\|\widehat{\mathbf{x}}_k - \widehat{\mathbf{x}}_{k-1}\|$ , proposed in [3], would not require additional computations.

We have done several numerical tests, in which we simulate the linear system (1) and apply to it the method (5), with  $\omega = 1$  (i.e. the power method), and the  $ER_{\mathcal{F}}$  algorithm, as implemented in (A1), for different choices of the dimension  $n$  and the parameter  $\alpha$ . Since the results are all very similar, here we illustrate, in Fig. 3, only the results of one of them. We report the values of the iteration index  $k$  and the  $L^1$  norm of the three residuals  $r_{\mathbf{R}}(k)$  for Richardson method, and  $r_{\mathcal{F}}(k) = \mathbf{v} - A^T (F\widehat{\mathbf{x}}_k)$ ,  $\widehat{r}_{\mathcal{F}}(k) = \widehat{\mathbf{v}} - \widehat{A}^T \widehat{\mathbf{x}}_k$  for the  $ER_{\mathcal{F}}$  method. Since the dimension  $n$  in our experiments is much smaller than the inverse of the numerical accuracy, by (20) we can consider  $L^1$  residuals instead of  $L^2$ . The numerical results show that the  $L^1$  norms of  $r_{\mathcal{F}}$  and  $\widehat{r}_{\mathcal{F}}$  decrease closely, as (20) ensures.

From the numerical experiences we emphasize the following facts:

1. As we expect, the number of iterations required to reduce the residuals, for both the methods, does not depend on the dimension  $n$ .
2. The number of iterations that  $ER_{\mathcal{F}}$  needs to reduce  $\|\widehat{r}_{\mathcal{F}}\|_1$  under a certain threshold, is significantly less than the number that Richardson method requires for reducing  $\|r_{\mathbf{R}}\|_1$  (see Fig. 3). Moreover, the difference between the two methods becomes more evident as  $\alpha$  increases, in fact the spectral radius of the preconditioned iteration matrix does not depend on  $\alpha$  but only on the density of  $P$ , whereas the power method iteration matrix spectral radius is strictly dependent on  $\alpha$  and close to it. This is an interesting result recalling that the web hyperlink graph model is much more faithful to the original *Random surf model* as  $\alpha$  is close to 1 (see [1] or [2]).

It should be noticed that it is not easy to verify the convergence condition (12) in our case where  $U$  is  $F$ , the Fourier matrix. In fact, the number  $\gamma_{A,F}^{(1)}$  in (11) is easily valuable (and is equal to 1) only in the

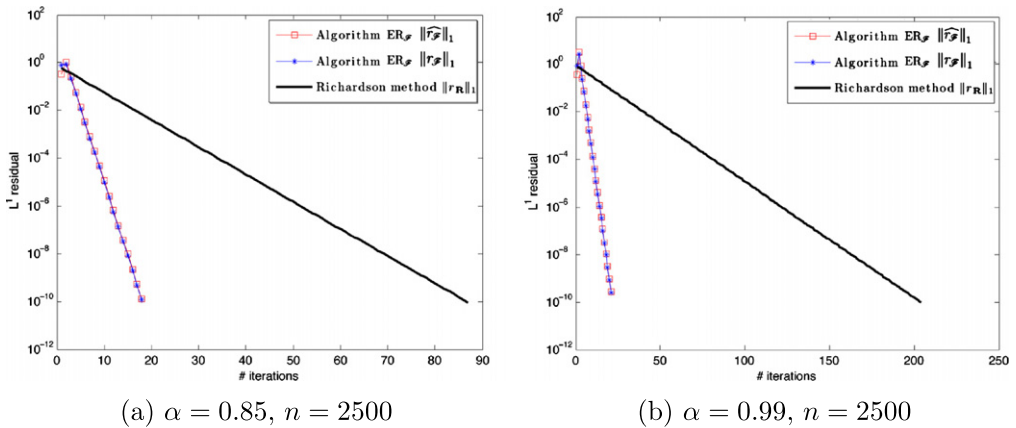


Fig. 3. The plots in logarithmic scale show both the three residuals  $\|r_R\|_1$ ,  $\|r_{\mathcal{F}}\|_1$  and  $\|\widehat{r}_{\mathcal{F}}\|_1$  over the number of iterations. It is noticeable how using the circulant  $\sigma$ -preconditioner assures a drastic increase of the convergence rate. Moreover observe how close are the residuals  $\|\widehat{r}_{\mathcal{F}}\|_1$  and  $\|r_{\mathcal{F}}\|_1$  (they are not distinguishable by the plots).

case  $U$  is a  $L^1$  isometry, or, equivalently,  $U = \Pi D$  where  $\Pi$  is a permutation matrix and  $D$  is a diagonal unitary matrix. Despite of this we finally underline that the  $ER_{\mathcal{F}}$  method converges all times we tested it, for all the values we gave to  $\alpha$ , for all the randomly generated matrices  $P$  that we considered (with the properties discussed in Section 1), and for all the values of the dimension  $n$  that our available means can support.

### 9. Conclusions

In this paper, we study several aspects of the  $W$ -pagerank computation problem, and we indicate, in particular, the preconditioning of power method as a way to improve its rate of convergence, in computing the  $W$ -pagerank vector  $\mathbf{p}$ . We introduce a quasi-stochastic row-constant sparse matrix  $P$ , and we define the  $W$ -matrix as a positive stochastic matrix of the form  $W = \alpha P + R$ ,  $R$  rank-one,  $\alpha \in (0, 1)$ , and the  $W$ -pagerank vector as the vector  $\mathbf{p} = W^T \mathbf{p}$ . So, we have the  $W$ -pagerank computation problem of calculating  $\mathbf{p}$ . All results obtained apply to the Google-pagerank computation problem [1], as well as to the problem of computing the vertexrank of the graph associated to any other set of information. We begin by investigating the (clustering) spectral properties of the matrix  $P$ , stating some bounds for its maximal singular value and for its second eigenvalue, useful to study the convergence of iterative methods in solving the  $W$ -problem. Then, we apply to the  $W$ -pagerank problem the Euler-Richardson (ER) iterative scheme, preconditioned by a matrix  $Q$ , with the aim of choosing  $Q$  so that the spectral radius of the ER iteration matrix is minimized. It is shown that only one eigenvalue of  $P$  seems to be responsible of the low rate of convergence of ER with  $Q = I$  (i.e. of the power method), and it is proposed at least one choice of  $Q$  that seems to remove such eigenvalue, remarkably improving ER convergence rate. In fact, such  $Q$  is defined in terms of the best approximation of  $P$  in a generic matrix algebra  $\mathcal{U} = \{UDU^* \mid D \text{ diagonal}\}$ ,  $U$  unitary, and the corresponding ER method is referred to as  $ER_{\mathcal{U}}$ . First we obtain a convergence result for  $ER_{\mathcal{U}}$  and then the choice  $\mathcal{U} = \text{circulants}$  is discussed in detail. In particular, taking into account the possible continuous changes in  $P$  and the continuous growth of its dimension, we introduce a useful procedure updating in real time the essential quantities defining the proposed circulant preconditioner  $Q$ . The arithmetic operations per step and the memory allocations required to implement  $ER_{\mathcal{U}}$  (any time  $\mathbf{p}$  must be recomputed) are more than those required by the power method ER, and thus  $ER_{\mathcal{U}}$  is superior to such method for not too large  $P$ . However, the several numerical experiences reported, showing the high improvement in convergence rate that can be obtained by preconditioning, and the general theory developed in this paper, are a reason to investigate different choices of  $\mathcal{U}$  and of  $Q \in \mathcal{U}$  more efficient in case of huge  $P$ . In fact, other



approximations in  $\mathscr{U}$  of  $P$ , which can be preconditioners better than the T. Chan-like approximation (as in [31,32]), and other algebras  $\mathscr{U}$ , closer to  $P$ , can be now investigated.

## Acknowledgement

The authors thank Professor E.E. Tyrtyshnikov and the Referees for their helpful remarks.

## References

- [1] P. Berkhin, A survey on PageRank computing, *Internet Math.* 2 (1) (2005) 73–120.
- [2] A.N. Langville, C.D. Meyer, *Google's PageRank and Beyond: The Science of Search Engine Rankings*, Princeton University Press, New Jersey, 2006.
- [3] G.M. Del Corso, A. Gulli, F. Romani, Fast PageRank computation via a sparse linear system, *Internet Math.* 2 (3) (2005) 259–281.
- [4] G.M. Del Corso, A. Gulli, F. Romani, Comparison of Krylov Subspace Methods on the PageRank Problem, University of Pisa, Technical Report, TR-05-20, 2005.
- [5] G.H. Golub, C. Greif, An Arnoldi-type algorithm for computing PageRank, *BIT* 46 (2006) 759–771.
- [6] D. Gleich, L. Zhukov, P. Berkhin, Fast Parallel PageRank: A Linear System Approach, Yahoo! Research Technical Report, YRL-2004-038, 2004.
- [7] S.D. Kamvar, T.H. Haveliwala, C.D. Manning, G.H. Golub, Extrapolation methods for accelerating PageRank computations, in: *Proceedings of the 12th International Conference on World Wide Web*, Budapest, Hungary, May 20–24, 2003.
- [8] C. Brezinski, M. Redivo-Zaglia, Extrapolation and minimization procedures for the PageRank vector, *Web Information Retrieval and Linear Algebra Algorithms*, Internationales Begegnungs- und Forschungszentrum für Informatik (IBFI), Schloss Dagstuhl, Germany, 07071, 2007, pp. 1862–4405.
- [9] C. Brezinski, M. Redivo-Zaglia, S. Serra-Capizzano, Extrapolation methods for PageRank computations, *Académie des sciences Published by Elsevier SAS, C.R. Math. Acad. Sci. Paris Ser. I* 340 (2005) 393–397.
- [10] D. Zhou, J. Huang, Schölkopf, Learning from labeled and unlabeled data on a directed graph, in: *Proceedings of the 22nd International Conference on Machine Learning*, 2005.
- [11] G.M. Del Corso, A. Giulli, F. Romani, Ranking a stream of news, in: *Proceedings of the 14th International Conference on World Wide Web*, 2005, pp. 97–106.
- [12] D.A. Bini, G.M. Del Corso, F. Romani, Evaluating scientific products by means of citation-based models: a first analysis and validation, *Electron. Trans. Numer. Anal.* 33 (2008) 1–16.
- [13] E.E. Tyrtyshnikov, *A Brief Introduction to Numerical Analysis*, Birkhäuser, Boston, 1997.
- [14] E.E. Tyrtyshnikov, N.L. Zamarashkin, On eigen and singular value clusters, *Calcolo* 33 (1996) 71–78.
- [15] David Williams, *Probability with Martingales*, Cambridge University Press, New York, 1991.
- [16] E.E. Tyrtyshnikov, A. Yu. Yeremin, N. Zamarashkin, Clusters, preconditioners, convergence, *Linear Algebra Appl.* 263 (1997) 25–48.
- [17] S. Serra-Capizzano, D. Bertaccini, G.H. Golub, How to deduce a proper eigenvalue cluster from a proper singular value cluster in the nonnormal case, *SIAM J. Matrix Anal. Appl.* 27 (2005) 82–86.
- [18] S. Friedland, L. Gurvits, An upper bound for the real part of nonmaximal eigenvalues of nonnegative irreducible matrices, *SIAM J. Matrix Anal. Appl.* 15 (1994) 1015–1017.
- [19] K.P. Hadeler, Abschätzungen für den zweiten Eigenwert eines positiven operators, *Aequationes Math.* 5 (1970) 227–235.
- [20] U.G. Rothblum, C.P. Tan, Upper bounds on the maximum modulus of subdominant eigenvalues of nonnegative matrices, *Linear Algebra Appl.* 66 (1985) 45–86.
- [21] Y. Saad, *Iterative Methods for Sparse Linear Systems*, SIAM, Philadelphia, PA, 2003.
- [22] R.H. Chan, X. Jin, M. Yeung, The circulant preconditioner in the Banach algebra of matrices, *Linear Algebra Appl.* 149 (1991) 41–53.
- [23] E.E. Tyrtyshnikov, Optimal and superoptimal circulant preconditioners, *SIAM J. Matrix Anal. Appl.* 13 (1992) 459–473.
- [24] C. Di Fiore, P. Zellini, Matrix algebras in optimal preconditioning, *Linear Algebra Appl.* 335 (2001) 1–54.
- [25] F. Tudisco, *Metodi analitico numerici per il problema del ranking delle pagine web*, Bachelor thesis, Dipartimento di Matematica, Università degli studi di Roma “Tor Vergata”, Rome, Italy, 2008.
- [26] T. Haveliwala, Topic-sensitive PageRank, in: *Proceedings of the 11th International Conference on World Wide Web*, ACM Press, 2002, pp. 517–526.
- [27] H. Zhang, A. Goel, R. Govindan, K. Mason, B. Van Roy, Making eigenvector-based reputation system robust to collusion, in: *Third International Workshop, WWW 2004, Rome, Italy, October 2004, Proceedings*, Lecture Notes in Computer Sciences, vol. 3243, Springer, 2004, pp. 92–104.
- [28] A. Bortoletti, C. Di Fiore, On a set of matrix algebras related to discrete Hartley-type transforms, *Linear Algebra Appl.* 366 (2003) 65–85.
- [29] C. Di Fiore, Matrix algebras and displacement decompositions, *SIAM J. Matrix Anal. Appl.* 21 (2000) 646–667.
- [30] E. Bozzo, C. Di Fiore, On the use of certain matrix algebras associated with discrete trigonometric transforms in matrix displacement decomposition, *SIAM J. Matrix Anal. Appl.* 16 (1995) 312–326.
- [31] N.L. Zamarashkin, I.V. Oseledets, E.E. Tyrtyshnikov, Approximation of Toeplitz matrices by sums of circulants and small-rank matrices, *Dokl. Math.* 73 (2006) 100–101.
- [32] I. Oseledets, E.E. Tyrtyshnikov, A unifying approach to the construction of circulant preconditioners, *Linear Algebra Appl.* 418 (2006) 435–449.