Journal of Computational and Applied Mathematics 296 (2016) 397–409

Contents lists available at ScienceDirect

Journal of Computational and Applied Mathematics

journal homepage: www.elsevier.com/locate/cam

FOM accelerated by an extrapolation method for solving PageRank problems[☆]

Hong-Fan Zhang, Ting-Zhu Huang [*], Chun Wen, Zhao-Li Shen

School of Mathematical Sciences, University of Electronic Science and Technology of China, Chengdu, Sichuan, 611731, PR China

ARTICLE INFO

Article history:
Received 1 July 2014
Received in revised form 18 May 2015

MSC:
60J22
65C40

Keywords:
PageRank
FOM
Krylov subspace method
Extrapolation
Ritz value

ABSTRACT

This paper formulates the PageRank problem $Ax = x$ into a consistent singular linear system $(I - A)x = 0$, and applies the full orthogonalization method (FOM) to solve it. This singular system is characterized by index one, namely $index(I - A) = 1$. We analyze the breakdown performance of FOM on a general singular linear system, and conclude that FOM can determine a solution if it converges, without any unfortunate breakdowns for our target problem. Then we propose to use a vector extrapolation method to speed up the convergence performance of FOM. This extrapolation procedure is based on Ritz values, which directly stems from the Arnoldi-Extrapolation algorithm (Wu and Wei, 2010). Eventually numerical experiments are presented to illustrate the effectiveness of our approaches.

1. Introduction

Solving Google's PageRank problems is categorized as the issue of calculating the stationary probability vector of large, sparse and irreducible Markov chains. The Google web matrix is defined as a combination of a column stochastic matrix $\widetilde{P}$ and a rank one matrix ve^T:

$$A = \alpha\widetilde{P} + (1 - \alpha)ve^T.$$

$\widetilde{P}$ comes from a real graph matrix P by replacing its zero rows corresponding to dangling nodes with constant elements (generally take $1/n$). d is the dangling nodes indicator and it gives rise to $\widetilde{P} = (P + de^T/n)^T$. The vector v is called a personalization vector whose elements are positive and sum to 1. It stands for some probability distribution over the original web pages. The damping factor α $(0 \le \alpha < 1)$ simulates the case in which a web surfer takes a random link jump. It is the linear combination specified by the above two parameters that force the final Google matrix A to be irreducible, which ensures the existence and uniqueness of the stationary probability solution. We just desire the principal eigenvector of A,

$$Ax = x, \quad \|x\|_1 = 1, \; x \ge 0. \tag{1}$$

The value of α plays a significant role in the PageRank model. Low values (such as 0.85) indicate the second largest eigenvalue λ_2 is well separated from the largest one, namely 1, allowing for some simple iterative methods such as the power iteration.

☆ This research is supported by 973 Program (2013CB329404), NSFC (61170309, 11501085), Chinese Universities Specialized Research Fund for the Doctoral Program (20110185110020), and the Fundamental Research Funds for the Central Universities (ZYGX2014J084).

* Corresponding author. Tel.: +86 28 61831608; fax: +86 28 61801280.
E-mail addresses: zhfbeyond@126.com (H.-F. Zhang), tingzhuhuang@126.com (T.-Z. Huang), wchun17@163.com (C. Wen), szlxiaoyao@163.com (Z.-L. Shen).

http://dx.doi.org/10.1016/j.cam.2015.09.027
0377-0427/

For all the web matrices, an upper bond holds that $0 < \lambda_2 \leq \alpha < 1$, and this is the origin of approximating λ_2 by α in our later work. We refer readers to [1–6] for more detailed properties of Google matrices.

When the damping factor α is close to 1, more sophisticated and efficient techniques should be applied. The restarted Arnoldi-type algorithm has been leading the main trend in terms of high parallelization and low saving. Jia proposed a new algorithm that computes refined approximate eigenvectors by small sized singular value decompositions, which has urged more rapid convergence if Ritz values converge [7,8]. Golub and Greif applied the refined Arnoldi process to PageRank by forcing a relevant shift to be 1, which not only circumvents drawbacks of complex arithmetic but also improves the whole algorithm performance [9]. Recently abundant advanced strategies have sprung up to accelerate the celebrated Arnoldi-type method. The Power-Arnoldi algorithm [10] and the Arnoldi-Extrapolation algorithm [11], both proposed by Wu and Wei, were concentrating on periodically combining the power iterations with the Arnoldi processes. The thick restarting strategy and the Ritz value based extrapolation technique were fully utilized respectively. In fact the fine accelerating effect of that extrapolation procedure also motivates us to use it to speed up FOM in this paper.

Here we remind readers that the aforementioned algorithms are all aimed at searching the maximal eigenvector, rather than a linear equation's solution. To be specific, the desired eigenvector is searched in the following Krylov subspace which is induced by A and a nonzero initial vector v_0,

$$K_m(A, v_0) = span\{v_0, Av_0, \ldots, A^{m-1}v_0\}, \tag{2}$$

where m is the dimension of subspace. However FOM in our work does formulate the PageRank problem into a singular linear system

$$(I - A)x = 0, \tag{3}$$

and the associated Krylov subspace is

$$K_m(I - A, r_0) = span\{r_0, (I - A)r_0, \ldots, (I - A)^{m-1}r_0\}, \tag{4}$$

where r_0 is the residual vector of a given initial approximation. Obviously these are two completely different approaches. Solving the singular linear system (3) for PageRank has never been considered before and this is the first try in the literature. However transforming PageRank into a linear system is far from novel. Gleich et al. first put GMRES into effect on the nonsingular linear system

$$(I - \alpha P^T)x = v. \tag{5}$$

Premise is declared that the solution of the above nonsingular system makes little difference to the final page ranking. So some Krylov subspace techniques such as the preconditioned GMRES was developed for solving (5) [12,13].

Now focus on (3). The coefficient matrix $I - A$ is singular since 1 is the largest eigenvalue of A. More attention should be paid when some specific Krylov subspace methods are employed for solving such a singular linear system, taking account of possibility of unexpected breakdowns caused by rank deficiencies [14,15]. Fortunately, the singular matrix $I - A$ is characterized by index one, $index(I - A) = 1$, from which unexpected breakdowns can survive when FOM is applied. Here *index* denotes the size of the largest Jordan block corresponding to zero eigenvalues of a matrix. We make efforts to discuss the breakdown performance of FOM on a general singular linear system, and conclude that FOM is invariably capable of determining an approximate solution. These discussions in essence provide a theoretical guarantee for feasibility of applying FOM. A specific FOM implementation will be described at length in the later sections.

The remaining paper is organized as follows. In Section 2, we briefly introduce the classical refined Arnoldi-type algorithm for PageRank. Arnoldi process and Ritz values are involved. Section 3 mainly studies FOM for PageRank. Before that, we would like to work on the discussions of breakdown performances of FOM on singular linear systems. In order to speed up the convergence behavior of FOM, a vector extrapolation method based on Ritz values is proposed. The resulting algorithm is called FOM-EXT. We also present a brief convergence discussion. All of these are in Section 4. Numerical experiments are carried out to illustrate high effectiveness of our approaches in Section 5. Finally we close with concluding remarks in Section 6.

2. The refined Arnoldi-type method for PageRank

This section briefly introduces Golub and Greif's refined Arnoldi-type algorithm, which is based on small sized singular value decompositions with a shift being 1. The Arnoldi process is a cornerstone of all the Krylov subspace methods, no matter for eigen-problems or for linear systems. Given a general square matrix $G \in \mathcal{R}^{n\times n}$ and a nonzero initial vector $v_0 \in \mathcal{R}^n$, the Arnoldi process generates a canonical orthogonal basis $\{v_1, v_2, \ldots, v_m\}$,

$$span\{v_1, v_2, \ldots, v_m\} = K_m(G, v_0) = span\{v_0, Gv_0, \ldots, G^{m-1}v_0\}. \tag{6}$$

The modified *Gram–Schmidt* variant of the Arnoldi process is outlined as follows [16]: