Journal of Computational and Applied Mathematics 309 (2017) 219–229

Contents lists available at ScienceDirect

Journal of Computational and Applied Mathematics

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

An Arnoldi-Inout algorithm for computing PageRank problems[☆]

Chuanqing Gu[*], Wenwen Wang

Department of Mathematics, Shanghai University, Shanghai 200444, China

ARTICLE INFO

Article history:
Received 27 October 2015
Received in revised form 14 February 2016

Keywords:
PageRank
Inner–outer iteration
PIO iteration
Thick restarted Arnoldi algorithm

ABSTRACT

The PageRank algorithm plays an important role in the web search engines. To speed up the convergence behavior for computing PageRank, we propose a new method, called as Arnoldi-Inout, which is the inner–outer iteration method modified with the thick restarted Arnoldi method. The description and convergence of the new algorithm are discussed in detail. Numerical results are given to illustrate the efficiency of the new algorithm.

1. Introduction

Recently, with the booming development of the internet, web search engines have become one of the most important internet tools for information retrieval. The core of the web search engines is the design of the search algorithms. The PageRank algorithm [1] is one of the most popular search algorithms.

The PageRank problem is to compute the principal eigenvector of the Google matrix A defined by

$$A = \alpha P + (1-\alpha)E, \tag{1.1}$$

where $\alpha \in (0, 1)$ is the damping factor, $E = ve^T$, $e = (1, 1, \ldots, 1)^T \in \mathbb{R}^n$, $v = e/n$, P is a column-stochastic matrix, see [2] for the details, and n is the dimension of P. The PageRank vector is defined as the vector satisfying

$$x = Ax, \tag{1.2}$$

where A is defined as in (1.1).

The PageRank problem can be converted into the solution of the largest eigenvalues of the matrix corresponding to eigenvectors. The Power method [1] is one of the oldest methods for computing the dominant eigenvector of a given matrix. It converges for every choice of a nonnegative starting vector and it stands out for its reliable performance. However, the low efficiency in convergence is its fatal flaw. When the damping factor is close to 1, the Power method converges very slowly. Thus it is necessary to develop more efficient techniques for PageRank computation.

We note that the smaller the damping factor is, the easier it is to solve the problem. By converting the PageRank problem to the solution of the system of linear equations, Gleich et al. [3] proposed an inner–outer iteration method combined with

☆ The work are supported by National Natural Science Foundation (11371243), Innovation Program of Shanghai Municipal Education Commission (13ZZ068), Key Disciplines of Shanghai Municipality (S30104).

* Corresponding author.
E-mail address: cqgu@staff.shu.edu.cn (C. Gu).

http://dx.doi.org/10.1016/j.cam.2016.05.026
0377-0427/

Richardson iterations in which each iteration requires solving a linear system similar in its algebraic structure to the original one. This approach is very efficient as it can solve the PageRank problem with a suitable damping factor β.

On the other hand, A number of iterative methods based on Arnoldi process were proposed for solving the large sparse system of linear equations. They are also good alternatives for computing the dominant eigenvector. However, the Krylov subspace methods are not satisfactory due to its heavy cost used per iteration.

The PIO algorithm [4] proposed by Gu and Xie combines the Power method and the inner–outer iteration method, modifying the inner–outer iteration method preconditioned with the Power iteration. The Power-Arnoldi algorithm [5] proposed by Wu and Wei is a hybrid algorithm that is based on a periodic combination of the power method with the thick restarted Arnoldi algorithm. In their approach, the thick restarted Arnoldi method was used.

In this paper, we present a new method called as Arnoldi-Inout, which modifies the inner–outer method preconditioned with the thick restarted Arnoldi method. Numerical examples illustrate the efficiency of the new algorithm.

The remainder of the paper is organized as follows. In Section 2, we briefly introduce the inner–outer iteration method and the thick restarted Arnoldi algorithm for PageRank problem. In Section 3, we first present the main algorithm of this paper, that is, the Arnoldi-inner–outer (Arnoldi-Inout) algorithm. In Section 4, we analyze its convergence properties. Numerical tests and comparisons are reported in Section 5. Finally, some brief concluding remarks are presented in Section 6.

2. The inner–outer method and thick restarted Arnoldi method for computing PageRank

In this section, we briefly introduce the inner–outer iteration method [3] and the thick restarted Arnoldi method [5] for computing PageRank problem.

2.1. The inner–outer iteration method

The inner–outer iteration method for computing PageRank problem was proposed in [3]. Firstly, we summarize the inner–outer iteration method for computing PageRank.

From (1.1) and (1.2), the eigenvalue problem

$$x = Ax = (\alpha P + (1-\alpha)ve^T)x, \tag{2.1}$$

can be rewritten as the linear system

$$(I - \alpha P)x = (1-\alpha)v, \tag{2.2}$$

since $e^T x = 1$.

We note that the PageRank problem is easier to solve when the damping vector is small. Instead of solving the (2.2) directly, Gleich et al. [3] defined the outer iteration with a lower damping factor β, i.e., $0 < \beta < \alpha$. Then the linear system (2.2) is reformulated as the following equivalent equations

$$(I - \beta P)x = (\alpha - \beta)Px + (1-\alpha)v. \tag{2.3}$$

Then they consider the following iteration referred as the outer iteration

$$(I - \beta P)x^{(k+1)} = (\alpha - \beta)Px^{(k)} + (1-\alpha)v, \quad k = 0, 1, 2, \ldots, \tag{2.4}$$

for solving (2.2). For computing $x^{(k+1)}$, they use an inner Richardson iteration as follows. By setting

$$f = (\alpha - \beta)Px^{(k)} + (1-\alpha)v, \tag{2.5}$$

they define the inner linear system as

$$(I - \beta P)y = f, \tag{2.6}$$

and compute $x^{(k+1)}$ via the Richardson inner iteration

$$y^{(j+1)} = \beta P y^{(j)} + (\alpha - \beta)Px^{(k)} + (1-\alpha)v, \quad j = 0, 1, 2, \ldots, l-1, \tag{2.7}$$

where $y^{(0)} = x^{(k)}, y^{(l)} = x^{(k+1)}$.

The stopping criteria are given as follows. The outer iteration (2.4) terminates if

$$\|(1-\alpha)v - (I - \alpha P)x^{(k+1)}\|_2 < \tau, \tag{2.8}$$

while the inner iteration (2.6) terminates if

$$\|f - (I - \beta P)y^{(j+1)}\|_2 < \eta, \tag{2.9}$$

where η and τ are the inner and outer tolerances respectively.