Journal of Computational and Applied Mathematics 295 (2016) 149–158

Contents lists available at ScienceDirect

Journal of Computational and Applied Mathematics

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

On perturbations of principal eigenvectors of substochastic matrices

Ludmila Bourchtein *, Andrei Bourchtein

Institute of Physics and Mathematics, Pelotas State University, Brazil

ARTICLE INFO

Article history:
Received 10 September 2014
Received in revised form 14 December 2014

MSC:
60J10
60J20
15A18
15A51

Keywords:
Substochastic matrices
Markov chains
Dangling nodes
Principal eigenvectors
PageRank vector

ABSTRACT

In this report we consider substochastic matrices with some zero rows and study the sensitivity of their eigenvectors to the modifications of zero entries. The analysis is prompted by and directly applied to the Google method of ranking the web sites, which substitutes the pages without outlinks (dangling nodes) in the original web link matrix by non-zero rows (dangling vectors). We present an analysis of the influence of artificial links attributed to the dangling nodes on the principal eigenvectors of the web matrix. We clarify when the choice of the dangling vector does not change the original eigenvectors and give an evaluation for perturbations of the principal eigenvectors when they are subject to modification.

1. Introduction

The methods of finding the PageRank vector, including formation of the original stochastic matrix, use of personalization vector, the reasoning behind the PageRank approach, and application of different numerical algorithms, are well documented in a number of sources (see, for example [1–6] and references therein). So we will recall below some points essential for our analysis, referring the interested reader to the above and other sources for additional information.

The methods of computing the PageRank vector can be divided in the following steps. First, the original matrix A_0 is defined considering the web of n pages as a directed graph whose vertices are web pages and edges are links. If d_i denotes the number of outlinks from page i, then the entries of A_0 are defined as follows:

$$A_0 = (a_{ij}), \quad a_{ij} = \begin{cases} 1/d_i, & \text{if there is outlink from } i \text{ to } j; \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

A page that has no outlink is called a dangling node (page). Evidently, A_0 is row-substochastic matrix, which we call the original link matrix. At the second step, the matrix A_0 is transformed to the row-stochastic matrix A_1 by substituting zero rows corresponding to the dangling nodes by a stochastic (row) vector w^T:

$$A_1 = (\tilde{a}_{ij}), \quad \tilde{a}_{ij} = \begin{cases} 1/d_i, & \text{if } d_i \geq 1 \text{ and there is outlink from } i \text{ to } j; \\ 0, & \text{if } d_i \geq 1 \text{ and there is not outlink from } i \text{ to } j; \\ w_j, & \text{otherwise.} \end{cases} \tag{2}$$

* Corresponding author.
E-mail address: ludmila.bourchtein@gmail.com (L. Bourchtein).

http://dx.doi.org/10.1016/j.cam.2015.01.013
0377-0427/

We call such vector w the dangling vector. Usually, $w = \frac{1}{n}e$, where e is the n-order (column) vector of ones. Third, the convex combination of the matrices A_1 and V gives the so-called Google matrix G:

$$G = \alpha A_1 + (1 - \alpha)V, \tag{3}$$

where $V = ev^T$ is a rank-one row-stochastic matrix based on a stochastic vector v (called personalization vector) and scalar $\alpha \in (0, 1)$ (called teleportation parameter). Frequently, $v = w$ and both vectors are also called personalization vectors. Finally, the PageRank vector is computed by finding the Perron vectors (left-hand positive normalized eigenvectors for eigenvalue $\lambda = 1$) for matrices G with different α and applying the limit to such vectors as α approaches 1. In practice, different approximations to that limiting vector are considered as the PageRank vector. For example, Google reports to use the Perron vector of the matrix (3), corresponding to the specific value of $\alpha = 0.85$, as the PageRank vector.

Let us note that substitution of the matrix A_0 subsequently by A_1 and G is a way to deal with technical problems of finding the principal eigenvectors of the original link matrix A_0. (Principal eigenvectors are the eigenvectors with nonnegative components corresponding to the principal eigenvalue $\lambda = \rho(A_0)$, where $\rho(A_0)$ stands for the spectral radius of A_0.)

Since the introduction by the Google co-founders [1], the study of properties of the PageRank algorithm is being active field of research in theoretical and computational mathematics, and different contributions, which contain results about the structure of the Google matrix and convergence of the PageRank algorithms, have been obtained (e.g., [2,3,7–11], just to mention a few). In this study we clarify some relations between the principal eigenvectors of the matrices A_0 and A_1.

The paper is structured as follows. In Section 2 we find the dangling vectors, which assure the same principal eigenvectors as in the original matrix A_0: first we transform the original matrix to a simpler form by lumping all the dangling nodes into a single node, and then we solve the problem separately for the cases when the principal submatrix of A_0 corresponding to the nondangling nodes has the spectral radius equal to 1 and less than 1. In Section 3 we study sensitivity of the principal eigenvectors to the choice of the dangling vector: first we derive a lower bound for perturbations of the principal eigenvectors and then compare the number of the principal eigenvectors of the original and perturbed matrices. Finally, in Section 4 we give a simple illustration by analyzing a web with two nondangling nodes. Although all the analysis is made for the case of the web link matrix, the same results are valid for any substochastic matrix with the described properties.

From now on, we will use the term stochastic to refer to the row-stochastic matrices. Any non-transposed vector is a column vector; the transpose to vector and matrix is superscripted with a T. Along all the text we use standard terminology and different results from matrix analysis and theory of Markov chains. All the classical definitions and statements used in this text can be found in [12–14].

2. Relations for dangling vector

The main point of this section is to clarify if the dangling vectors can be chosen in such a way that the matrix A_1 has the same principal eigenvectors as the matrix A_0.

2.1. Equivalent forms of the original matrix

Let us consider a web of n pages with n_d ($0 < n_d < n$) dangling nodes. According to (1), the original $n \times n$ link matrix A_0 of this web can be written in the following block form

$$A_0 = \begin{pmatrix} H & H_d \\ 0 & 0 \end{pmatrix}, \tag{4}$$

where H is $k \times k$ substochastic matrix ($k = n - n_d$), and H_d is nonzero $k \times n_d$ matrix containing the links to the dangling nodes.

In this study we will be interested in behavior of the first k components of the principal eigenvectors, that is, projection of the eigenvectors, associated with the eigenvalue equal to the spectral radius $\rho(A_0)$, onto the subspace $\mathbb{R}^k$. It means that the last n_d components of these eigenvectors will be of secondary (if any) interest for us. For this reason, we can reduce the number of considered dangling nodes to one, using the fact that matrix A_0 is lumpable with respect to these nodes, that is, the search for the eigenvectors of A_0 can be reduced to finding the eigenvectors of the lower order $(k+1) \times (k+1)$ matrix A in the form

$$A = \begin{pmatrix} H & g \\ 0 & 0 \end{pmatrix}, \tag{5}$$

where $g = H_d e$ is the column vector of order k, and e is the n_d-order vector of ones.

In fact, if $(\lambda = \rho(A_0), \sigma_0)$ is the (principal) eigenpair of the matrix A_0, with the first k components of the eigenvector placed in the subvector σ_p, such that

$$\sigma_0^T A_0 = \lambda \sigma_0^T, \qquad \sigma_0^T = (\sigma_p^T, \sigma_s^T), \qquad \sigma_p^T H = \lambda \sigma_p^T, \qquad \sigma_p^T H_d = \lambda \sigma_s^T, \tag{6}$$

then

$$\sigma^T A = \lambda \sigma^T, \qquad \sigma^T = (\sigma_p^T, \sigma_{k+1}), \qquad \sigma_{k+1} = \sigma_s^T e, \tag{7}$$

that is, the pair (λ, σ), is the (principal) eigenpair of the matrix A.