Applied Mathematics and Computation 255 (2015) 58–65

Contents lists available at ScienceDirect

Applied Mathematics and Computation

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

Finding the stationary states of Markov chains by iterative methods

Yurii Nesterov [a,*,1], Arkadi Nemirovski [b,2]

[a] Center for Operations Research and Econometrics (CORE), Catholic University of Louvain (UCL), 34 voie du Roman Pays, 1348 Louvain-la-Neuve, Belgium
[b] Georgia Institute of Technology, Atlanta, GA 30332, USA

ARTICLE INFO

Keywords:
Google problem
Power Method
Stochastic matrices
Global rate of convergence
Gradient methods
Strong convexity

ABSTRACT

In this paper, we develop new methods for approximating dominant eigenvector of column-stochastic matrices. We analyze the Google matrix, and present an averaging scheme with linear rate of convergence in terms of 1-norm distance. For extending this convergence result onto general case, we assume existence of a positive row in the matrix. Our new numerical scheme, the Reduced Power Method (RPM), can be seen as a proper averaging of the power iterates of a *reduced* stochastic matrix. We analyze also the usual Power Method (PM) and obtain convenient conditions for its linear rate of convergence with respect to 1-norm.

1. Introduction

1.1. Motivation

Problem of finding stationary states in Markov chains arise in many application fields. Usually it is reduced to a problem of finding a dominant eigenvector of a stochastic matrix. The later problem is traditionally solved by the Power Method (PM). Recall that the convergence of the Power Method is related to ratio of modulus of the second and the first leading eigenvalues [5]. This ratio is not very visible from the initial data (coefficients of the matrix). Thus, for a particular matrix, an a priory estimate of the possible rate of convergence of the Power Method remains a nontrivial question.

On the other hand, from the theory of Discrete Dynamical Systems, we know that the best possible rate of convergence can be established only with respect to a proper Euclidean metric, defined by some Linear Matrix Inequality. Consequently, the corresponding results on the convergence rate are usually written in an implicit form.

In this paper, we show that for (column-) stochastic matrices the situation is different. For this class, the uniqueness of a dominant eigenvector can be guaranteed by some simple and verifiable conditions. One of them is the existence of a strictly positive row (e.g., p. 51 in [1]). It appears that the sum of the minimal elements of all rows defines a linear rate of convergence of a special version of the power-type method (we call it the *Reduced Power Method* (RPM)). For the standard Power

* Corresponding author.
E-mail addresses: nesterov@core.ucl.ac.be (Y. Nesterov), nemirovs@isye.gatech.edu (A. Nemirovski).

[1] The research of the first author has been supported by the Grant "Action de recherche concertè ARC 04/09-315" from the "Direction de la recherche scientifique – Communautè française de Belgique", and by Laboratory of Structural Methods of Data Analysis in Predictive Modeling, MIPT, through the RF government Grant, ag.11.G34.31.0073.
[2] Research of the second author was supported by the NSF Grants CMMI-1232623 and CCF-1415498. The scientific responsibility rests with the authors.

http://dx.doi.org/10.1016/j.amc.2014.04.053
0096-3003/

Method (PM), we derive its linear rate of convergence from a simple expression for 1-norm of stochastic matrix acting on the vectors with zero sum of coordinates.

Our results are motivated by the Page Rank problem (or Google problem [2,4]). In particular, we show that, for the suggested in [2] value of the damping coefficient $\alpha = 0.15$, the corresponding problems can be solved very easily. For this particular application, the above mentioned uniqueness result can be formulated as follows. *The existence of global authority in the network ensures uniqueness of the stationary state in the corresponding Markov chain (SSMC).* This state can be found by a random walk with a positive service rate provided by global authorities. We show that SSMC problem is easy also for the Power Method provided that for two agents located at any pair of states, the probability to come next step at the same state is positive.

1.2. Contents

The paper is organized as follows. In Section 2 we introduce the Google problem and derive an explicit representation for the dominant eigenvector of the damped stochastic matrix. This representation naturally leads to an approximation procedure, based on a proper averaging of the power series for initial matrix. We prove a linear rate of convergence in terms of 1-norm both for the residual of linear system and for the distance to exact solution. In Section 3, we extend the above technique onto the general column-stochastic matrices. For its applicability, it is enough to assume existence of a *positive row* in the matrix. Then the initial matrix can be represented as a convex combination of two stochastic matrices, such that the second matrix is of rank one. This feature is essential for constructing an efficient approximation scheme (RPM) based on the power series of a *reduced* stochastic matrix. The global rate of convergence of this process is again linear. In Section 4, we study the Power Method. Despite to the negative expectations derived from the Jordan-form representation, we show that this method converges linearly on SSMC problem. In Section 5, we present a better framework for its convergence analysis, and discuss some interpretation of characteristics responsible for its convergence rate.

1.3. Notation

For two vectors $x, y \in R^n$ we denote by $\langle x, y \rangle$ their scalar product:

$$\langle x, y \rangle = \sum_{i=1}^{n} x^{(i)} y^{(i)}.$$

Notation $\| \cdot \|_p$ with $p \geqslant 1$, is used for p-norms:

$$\|x\|_p = \left[\sum_{i=1}^{n} |x^{(i)}|^p \right]^{1/p}, \quad x \in R^n.$$

The positive orthant in R^n is denoted by R^n_+. Notation e_j is used for the jth coordinate vector in R^n, and $e \in R^n$ denotes the vector of all ones. By $R^{n \times n}$, we denote the space of real $n \times n$-matrices, and I denotes the unit matrix of an appropriate size. We write $A \geqslant 0$ if matrix A has all entries nonnegative.

2. Solving Google problem by power sequences

The Google problem (or Page Rank problem) consists in approximating an eigenvector of a very big stochastic matrix. Let $E \in R^{n \times n}$ be an incidence matrix of a graph. Let us make it stochastic by an appropriate column scaling:

$$A \overset{\text{def}}{=} ED^{-1}(E^T e), \quad e = (1, \ldots, 1)^T \in R^n,$$

where $D(x) \in R^{n \times n}$ is a diagonal matrix with vector $x \in R^n$ on its diagonal. Thus,

$$A^T e = e. \tag{2.1}$$

Now, each column $Ae_j \in \Delta_n \overset{\text{def}}{=} \{x \in R^n_+ : \langle e, x \rangle = 1\}, j = 1, \ldots, n$. It contains the *transition probabilities* of the corresponding node.

We need to find a vector $x^* \in R^n_+$ satisfying the following system of linear equations:

$$Ax^* = x^*, \quad \langle e, x^* \rangle = 1. \tag{2.2}$$

From Perron–Frobenius theorem, we know that such a solution always exists.

Usually, system (2.2) is solved by different versions of the *Power Method*. Indeed, let us fix a starting vector $x_0 \in \Delta_n$ (we allow it to have some zero components). Define the sequence

$$x_{k+1} = Ax_k, \quad k \geqslant 0. \tag{2.3}$$

Note that $\langle e, x_{k+1} \rangle = \langle e, Ax_k \rangle \overset{(2.1)}{=} \langle e, x_k \rangle$. Thus, $x_k \in \Delta_n$ for all $k \geqslant 0$. The following result is well known.

Lemma 1. *Define* $\bar{x}_N = \frac{1}{N+1}\sum_{k=0}^{N} x_k$. *Then* $\bar{x}_N \in \Delta_n$ *and*

$$\|\bar{x}_N - A\bar{x}_N\|_1 \leqslant \frac{2}{N+1}. \tag{2.4}$$

Proof. Indeed, $\bar{x}_N - A\bar{x}_N = \frac{1}{N+1}\left[\sum_{i=0}^{N} A^i x_0 - \sum_{i=1}^{N+1} A^i x_0\right] = \frac{1}{N+1}\left[x_0 - A^{N+1}x_0\right]$. Therefore,

$$\|\bar{x}_N - A\bar{x}_N\|_1 \leqslant \frac{1}{N+1}\left[\|x_0\|_1 + \|A^{N+1}x_0\|_1\right] = \frac{1}{N+1}\left[\langle e, x_0\rangle + \langle e, A^{N+1}x_0\rangle\right] \overset{(2.1)}{=} \frac{2}{N+1}. \qquad \square$$

Remark 1. The estimate (2.4) looks quite pessimistic. However, it is attained on permutation matrices. Indeed, let $A_{ij} = \begin{cases} 1, & i = j+1 (mod n) \\ 0, & otherwise \end{cases}$. If we choose $x_0 = e_1$, then $\|\bar{x}_N - A\bar{x}_N\|_1 = \frac{2}{N+1}$ for all $N < n-1$.

The rate of convergence (2.4) is quite slow. Therefore, in [2] it was suggested to modify the initial system. Let us interpret $x_0 \in \Delta_n$ as a vector of initial preferences for the starting websites. Let us fix a coefficient $\alpha \in (0,1)$. Define $A_\alpha = (1-\alpha)A + \alpha x_0 e^T$. Clearly, this matrix remains stochastic:

$$A_\alpha^T e = e. \tag{2.5}$$

On the other hand, the problem of finding its leading eigenvector becomes much simpler. We justify this claim by presenting an approximation scheme based on a power series of the initial matrix A.

Denote by $x_\alpha^* \in \Delta_n$ a solution of the equation $A_\alpha x_\alpha^* = x_\alpha^*$. Note that

$$x_\alpha^* = \left[(1-\alpha)A + \alpha x_0 e^T\right]x_\alpha^* = (1-\alpha)Ax_\alpha^* + \alpha x_0.$$

Thus,

$$x_\alpha^* = \alpha[I - (1-\alpha)A]^{-1}x_0 = \alpha\sum_{k=0}^{\infty}(1-\alpha)^k A^k x_0. \tag{2.6}$$

This representation of the solution suggests the following approximation strategy. Define

$$\hat{x}_N = \sum_{k=0}^{N}(1-\alpha)^k x_k, \tag{2.7}$$

where the sequence $\{x_k\}_{k=0}^{\infty}$ is formed by (2.3). Then

$$\langle e, \hat{x}_N\rangle = \left\langle e, \sum_{k=0}^{N}(1-\alpha)^k A^k x_0\right\rangle \overset{(2.1)}{=} \sum_{k=0}^{N}(1-\alpha)^k = \frac{1}{\alpha}(1-(1-\alpha)^{N+1}).$$

Thus, defining $\tilde{x}_N = \alpha\hat{x}_N/(1-(1-\alpha)^{N+1})$, we get $\tilde{x}_N \in \Delta_n$. On the other hand,

$$\begin{aligned}\hat{x}_N - A_\alpha\hat{x}_N &= \sum_{k=0}^{N}(1-\alpha)^k A^k x_0 - \sum_{k=0}^{N}(1-\alpha)^k[(1-\alpha)A + \alpha x_0 e^T]A^k x_0 = x_0 - (1-\alpha)^{N+1}A^{N+1}x_0 - \alpha\sum_{k=0}^{N}(1-\alpha)^k x_0 \\ &= (1-\alpha)^{N+1}\left[x_0 - A^{N+1}x_0\right].\end{aligned}$$

Thus, we come to the following result.

Lemma 2. *For any* $N \geqslant 0$ *we have* $\tilde{x}_N \in \Delta_n$ *and*

$$\|\tilde{x}_N - A_\alpha\tilde{x}_N\|_1 \leqslant \frac{2\alpha(1-\alpha)^{N+1}}{1-(1-\alpha)^{N+1}}, \tag{2.8}$$

$$\|\tilde{x}_N - x_\alpha^*\|_1 \leqslant 2(1-\alpha)^{N+1}. \tag{2.9}$$

Proof. As we have seen,

$$\|\tilde{x}_N - A_\alpha\tilde{x}_N\|_1 = \frac{\alpha(1-\alpha)^{N+1}}{1-(1-\alpha)^{N+1}}\|x_0 - A^{N+1}x_0\|_1 \leqslant \frac{2\alpha(1-\alpha)^{N+1}}{1-(1-\alpha)^{N+1}}.$$

Moreover, since $x^*_\alpha = \alpha\sum_{k=0}^{\infty}(1-\alpha)^k x_k$, we have

$$\|\tilde{x}_N - x^*_\alpha\|_1 = \alpha\left\|\frac{1}{1-(1-\alpha^{N+1})}\sum_{k=0}^{N}(1-\alpha)^k x_k - \sum_{k=0}^{\infty}(1-\alpha)^k x_k\right\|_1 = \alpha\left\|\frac{(1-\alpha)^{N+1}}{1-(1-\alpha)^{N+1}}\sum_{k=0}^{N}(1-\alpha)^k x_k - \sum_{k=N+1}^{\infty}(1-\alpha)^k x_k\right\|_1$$
$$\leqslant \frac{\alpha(1-\alpha)^{N+1}}{1-(1-\alpha)^{N+1}}\cdot\frac{1}{\alpha}(1-(1-\alpha)^{N+1}) + \alpha\cdot\frac{1}{\alpha}(1-\alpha)^{N+1} = 2(1-\alpha)^{N+1}. \qquad \square$$

Note that the rate of convergence (2.9) implicitly confirms that the absolute value of the second eigenvalue of matrix A_α does not exceed $1-\alpha$.

In the original paper [2], it is suggested to take rather big value of the damping coefficient α, namely $\alpha = 0.15$. For this choice, method (2.7) can ensure a very high accuracy in the residual after a small number of steps. Indeed, in view of (2.9), in order to get l_1-distance to the exact solution x^*_α smaller than ϵ, we need only

$$\frac{1}{\alpha}\ln\frac{2}{\epsilon} \tag{2.10}$$

iterations of the method (2.3) and (2.7). It is important that this estimate does not depend on the size of the network.

3. General stochastic matrices

The above technique can be used for approximating a leading eigenvector of an arbitrary stochastic matrix A satisfying the following condition:

The set of positive rows of matrix A is nonempty. (3.1)

Let us choose some nonnegative values

$$r^{(i)} \leqslant \bar{r}^{(i)} \stackrel{\text{def}}{=} \min_{1\leqslant j\leqslant n} A_{i,j}, \quad i = 1,\dots,n, \; r = (r^{(1)},\dots,r^{(n)})^T. \tag{3.2}$$

In view of (3.1), we can ensure $r \neq 0$. Define $x_0 = r/\langle e, r\rangle$.

Let us try to represent A in the following form:

$$A = (1-\alpha)\bar{A} + \alpha x_0 e^T, \tag{3.3}$$

where $\alpha \in (0,1)$ and matrix $\bar{A}$ is stochastic. In order to achieve this goal, we need to ensure that

$$\bar{A}_{i,j} = \frac{1}{1-\alpha}\left(A_{i,j} - \alpha\frac{r^{(i)}}{\langle e, r\rangle}\right) \geqslant 0, \quad i,j = 1,\dots,n.$$

Thus, we can take

$$\alpha = \langle e, r\rangle \in (0,1). \tag{3.4}$$

Consequently, the dominant eigenvector x^* of matrix A admits the following representation:

$$x^* = \alpha\sum_{k=0}^{\infty}(1-\alpha)^k\bar{A}^k x_0. \tag{3.5}$$

As a byproduct of our reasoning, we get the following result.

Lemma 3. *Let stochastic matrix A satisfy condition (3.1). Then it has unique dominant eigenvector $x^* \in \Delta_n$.*

Recall that the standard sufficient condition for uniqueness of the dominant eigenvector of a stochastic matrix A consists in positivity of all its elements. Note that this condition works for arbitrary positive matrices. It ensures a nonzero gap between the largest eigenvalue and the absolute value of all other ones. Lemma 3 significantly improves this condition for the class of stochastic matrices.

In accordance to (3.5), in order to approximate x^*, we need to form the vectors $\bar{A}^k x_0$ and generate the averaging points

$$\hat{x}_N = \sum_{k=0}^{N}(1-\alpha)^k\bar{A}^k x_0 = \sum_{k=0}^{N}[A - re^T]^k x_0. \tag{3.6}$$

Note that $\langle e, \hat{x}_N\rangle = \sum_{k=0}^{N}(1-\alpha)^k = \frac{1}{\alpha}(1-(1-\alpha)^{N+1})$. Thus, defining $\tilde{x}_N = \bar{x}_N/\langle e, \bar{x}_N\rangle$, and using Lemma 2, we get

$$\begin{aligned}\|\tilde{x}_N - A\tilde{x}_N\|_1 &\stackrel{(2.8)}{\leqslant} \frac{2\alpha(1-\alpha)^{N+1}}{1-(1-\alpha)^{N+1}},\\ \|\tilde{x}_N - x^*\|_1 &\leqslant 2(1-\alpha)^{N+1}.\end{aligned} \tag{3.7}$$

The right-hand side of the first inequality in (3.7) tends to $\frac{2}{N+1}$ as $r \to 0$ (compare with Lemma 1).

It is interesting that in (3.6) we construct the main sequence by a power scheme with a modified (*reduced*) matrix. We call this scheme the *Reduced Power Method*:

$$x_0 = \frac{r}{\langle e, r\rangle}, \quad x_{k+1} = [A - re^T]x_k, \ k \geqslant 0, \quad \hat{x}_N = \sum_{k=0}^{N} x_k, \quad \tilde{x}_N = \frac{\hat{x}_N}{\langle e, \hat{x}_N\rangle}. \tag{3.8}$$

In this method, we are not free in the choice of x_0. The main condition for its applicability is the existence of a positive row of matrix A. For Google problem, this means the existence of a webpage, which can be visited with nonzero probability from any other page of the network. Clearly, this condition is not binding. It is satisfied, for example, by the page representing the search engine of Google itself. The more pages of this type (*global authorities*) exist in the network, the higher is the value of α in (3.4), and consequently, the faster is the convergence of process (3.8). The only possible trouble could be the absence of references from these pages to themselves. However, this can be fixed by considering the matrix $A(\delta) = \delta I + (1-\delta)A, \delta \in (0,1)$. This transformation clearly does not change the dominant eigenvector. At the same time, if we take $\delta = \frac{\langle e,r\rangle}{1+\langle e,r\rangle}$, then for any $i = 1, \ldots, n$ we have

$$r^{(i)}(\delta) \stackrel{\text{def}}{=} \min_{1 \leqslant j \leqslant n} A_{ij}(\delta) \ \geqslant \ \min\{\delta, (1-\delta)r^{(i)}\} \ \geqslant \ \frac{r^{(i)}}{1+\langle e, r\rangle}.$$

Hence, $\alpha(\delta) \stackrel{\text{def}}{=} \langle e, r(\delta)\rangle \geqslant \frac{\langle e,r\rangle}{1+\langle e,r\rangle}$, and the convergence rate of method (3.8) as applied to matrix $A(\delta)$ stays on the same level (look at the change in the estimate (3.7)).

It would be interesting to find an interpretation of the above results in terms of random walks in a graph. For this framework, the following statement looks quite intriguing.

Theorem 1.

1. *Existence of a global authority implies uniqueness of stationary state in the corresponding Markov chain.*
2. *This state can be efficiently approximated by random walk (3.8) with a service rate offered by global authorities.*
3. *The converging approximations are proportional to the total historical occupancy of the nodes.*

Theorem 1 delivers a natural sufficient conditions for uniqueness of the stationary state in the Markov chain. It can be easily verified. Moreover, from this verification, we get some important information about the rate of convergence of the process (3.8).

4. Comparison with Power Method

The good convergence results (3.7) certify that the problem of finding the leading eigenvector of a stochastic matrix A satisfying condition (3.1) is easy. This easiness is traditionally explained by a good structure of the spectrum of matrix A. This feature should be also profitable for the standard Power Method (2.3). Let us check how it works.

For that, we represent A in the Jordan form: $A = VJV^{-1}$, where J is a block-diagonal matrix, composed by Jordan blocks $J_i = \lambda_i I_i + Z_i, i = 1, \ldots, m$. In this representation,

- λ_i is an eigenvalue of matrix A,
- I_i is an identity matrix of dimension $k_i \times k_i$, $\sum_{i=1}^{m} k_i = n$,
- $Z_i \in R^{k_i \times k_i}$ is the upper shift matrix, which has zeros everywhere except the first upper diagonal, at which it has all ones. Note that

$$Z_i^{k_i} = 0, \quad i = 1, \ldots, m. \tag{4.1}$$

If $k_i = 1$, then J_i reduces to a single value λ_i.

We need one auxiliary result.

Lemma 4. *Let stochastic matrix A satisfy condition (3.1). Then:*

1. *The geometric multiplicity of its dominant eigenvalue is equal to one.*
2. *Any other eigenvalue λ satisfies inequality*[3]

$$|\lambda| \leqslant 1 - \bar{\alpha}, \tag{4.2}$$

where $\bar{\alpha} = \langle e, \bar{r}\rangle$.

[3] This inequality was proved first in [1]. It can be also derived from Theorem 1 in [3]. However, for the reader convenience, we present here a simple direct justification.

Proof

1. Let x be any eigenvector of A, which corresponds to the unit eigenvalue. Then $x \stackrel{(3.3)}{=} \alpha(I - (1-\alpha)\bar{A})^{-1} x_0 e^T x$. Hence, the dimension of this eigenspace is equal to one.
2. Denote $\hat{A} = A - \bar{r}e^T$. Let v_λ be an eigenvector of matrix A, which corresponds to a nonzero eigenvalue λ. Note that

$$e^T v_\lambda \stackrel{(2.1)}{=} e^T A v_\lambda = \lambda e^T v_\lambda.$$

Thus, $e^T v_\lambda = 0$ and $\hat{A} v_\lambda = A v_\lambda = \lambda v_\lambda$.

At the same time, matrix $\hat{A}$ has nonnegative elements and $\hat{A}^T e = (1 - \bar{\alpha})e$. Therefore, the dominant eigenvalue of matrix $\hat{A}$ is $1 - \bar{\alpha}$, and (4.2) follows from Perron-Frobenius theorem. □

Let us show that the bound (4.2) cannot be improved.

Example 1. Consider a stochastic (2×2)-matrix

$$A = \begin{pmatrix} \alpha & 1-\beta \\ 1-\alpha & \beta \end{pmatrix}, \quad \alpha, \beta \in (0,1).$$

Clearly, $\lambda_1(A) = 1$ and $\lambda_2(A) = 1 - \alpha - \beta$. On the other hand,

$$\bar{\alpha} = \min\{\alpha, 1-\beta\} + \min\{1-\alpha, \beta\}.$$

It is easy to check that $|\lambda_2(A)| = 1 - \bar{\alpha}$ for all $\alpha, \beta \in (0,1)$. □

Since geometric multiplicity of dominant eigenvalue of matrix A is equal to one, in matrix J there exists only one corresponding block. Without loss of generality, we can assume that this is the first Jordan block in J, namely, J_1.

Lemma 5. *Let the geometric multiplicity of eigenvalue 1 of stochastic matrix be equal to one. Then the algebraic multiplicity of this eigenvalue is also one.*

Proof. Note that $VJV^{-1}(Ve_1) = VJe_1 = Ve_1$. Therefore, the first column of matrix V is v_1, the dominant eigenvector of matrix A. We can use one available degree of freedom for scaling the first k_1 columns of matrix V and ensure the normalizing condition $e^T v_1 = 1$.

On the other hand, $e^T \stackrel{(2.1)}{=} e^T A = e^T V J V^{-1}$. Therefore $e^T V = e^T V J$. Since the left dominant eigenvector of matrix J is also unique, we conclude that

$$e^T V = \tau e_{k_1}^T$$

with some $\tau \neq 0$. Multiplying both sides of this equation from the right by e_1, we come to the unique possible conclusion: $\tau = 1$ and $k_1 = 1$. □

Assuming, as before, that $A \in \mathcal{P}_n$ satisfies (3.1) and taking into account Item 1 in Lemma 4, we conclude that $J_1 = 1$ and

$$e^T V = e_1. \tag{4.3}$$

Let us look now at the result of k iterations of Power Method (2.3). Clearly,

$$x_k = VJ^k V^{-1} x_0.$$

Therefore,

$$x_k - v_1 = (VJ^k V^{-1} - v_1 e^T) x_0 \stackrel{(4.3)}{=} (VJ^k V^{-1} - v_1 e_1^T V^{-1}) x_0 = V(J^k - e_1 e_1^T) V^{-1} x_0.$$

Thus, the rate of convergence of Power Method is defined by the rate of vanishing of diagonal blocks in the matrix $J^k - e_1 e_1^T$. The first 1×1 diagonal block of this matrix is equal to zero. The other blocks are presented by matrices

$$J_i^k = \sum_{j=0}^{k} \lambda_i^j \binom{k}{j} Z_i^{k-j}.$$

Note that the first term in this sum is Z_i^k. If $k < k_i$, then it has $k_i - k$ nonzero entries, all equal to one. Hence, matrix J_i^k *cannot* be small if $k < k_i$. In other words, the guaranteed convergence rate of Power Method should not be faster than $O\left((1-\bar{\alpha})^{k-\hat{k}}\right)$, where $\hat{k} = \max_{1 \le i \le m} k_i$. However, for some blocks, the dimension k_i can be in the order of n. Thus, it seems that we have an evidence that the convergence rate of Power Method cannot be dimension-independent.

It is interesting that the above impression is absolutely wrong. Indeed, denote

$$E_0 = \{h \in R^n : \langle e, h\rangle = 0\}.$$

Since A is stochastic, we have $AE_0 = E_0$. On the other hand, for any $h \in E_0$ we have

$$\|Ah\|_1 \overset{(3.3)}{=} (1-\alpha)\|\bar{A}h\|_1 \leqslant (1-\alpha)\langle \bar{A}|h|, e\rangle \overset{(2.1)}{=} (1-\alpha)\|h\|_1,$$

where $|h|$ denotes the vector with coordinates $|h^{(i)}|, i = 1, \ldots, n$. Therefore, for any $x \in \Delta_n$ we have

$$\|Ax - x^*\|_1 \overset{(2.2)}{=} \|A(x - x^*)\|_1 \leqslant (1-\alpha)\|x - x^*\|_1.$$

Thus, each iteration of Power Method (2.3) reduces 1-distance to the solution by a factor of $(1-\alpha)$. Consequently, this method has the same rate of convergence as (3.7). In the next section, we provide PM with better complexity analysis.

5. Convergence of Power Method

For matrix $A \in R^{n\times n}$, define the seminorm $\|A\|_{01}$ as follows:

$$\|A\|_{01} = \max_h \{\|Ah\|_1 : h \in E_0, \ \|h\|_1 \leqslant 1\}. \tag{5.1}$$

This seminorm has several important properties. In what follows, we denote by $\mathcal{P}_n$ the set of stochastic $n \times n$-matrices.

1. For all $h \in E_0$ we have
$$\|Ah\|_1 \leqslant \|A\|_{01}\|h\|_1. \tag{5.2}$$
2. Since all vertices of the polytope $\{h \in R^n : \langle e, h\rangle = 0, \ \|h\|_1 \leqslant 1\}$ have the form $\frac{1}{2}(e_i - e_j), i \neq j$, we have
$$\|A\|_{01} = \frac{1}{2}\max_{1\leqslant i,j\leqslant n} \|Ae_i - Ae_j\|_1. \tag{5.3}$$
3. $\|A\|_{01} = 0$ if and only if $A = xe^T$ for some $x \in R^n$.
4. If $A \in \mathcal{P}_n$, then $\|A\|_{01} \leqslant 1$. If matrix B also belongs to $\mathcal{P}_n$, then
$$\|AB\|_{01} \leqslant \|A\|_{01}\|B\|_{01} \leqslant \min\{\|A\|_{01}, \|B\|_{01}\}. \tag{5.4}$$

One of the main advantages of the seminorm $\|\cdot\|_{01}$ is that it is easily computable by (5.3). On the other hand, it provides us with a convenient bound for the magnitude of non-dominant eigenvalues (compare with Lemma 4).

Lemma 6. *Let $A \in \mathcal{P}_n$ and $\|A\|_{01} < 1$. Then the multiplicity of dominant eigenvalue of A is one, and all other eigenvalues λ satisfy*

$$|\lambda| \leqslant \|A\|_{01}. \tag{5.5}$$

Proof. Denote $\eta = \|A\|_{01}$. Since the subspace E_0 is invariant for A, and the restriction $B \overset{\text{def}}{=} A|_{E_0}$ is a η-contraction in 1-norm, the spectral radius of B does not exceed η. Let $x^* \in \Delta_n$ be a dominant eigenvector of $A, Ax^* = x^*$. Then the representation $R^n = R \cdot x^* + E_0$ is a decomposition of R^n into direct sum of two invariant for A subspaces. Hence, the spectrum of A is a union of the singleton $\{1\}$ and the spectrum of B. The latter is contained in the circle $\{\lambda : |\lambda| \leqslant \eta\}$. □

Let us derive a convenient representation of seminorm $\|\cdot\|_{01}$ for stochastic matrices. Recall the identity

$$\min\{\tau_1, \tau_2\} = \frac{1}{2}(\tau_1 + \tau_2) - \frac{1}{2}|\tau_1 - \tau_2|, \quad \tau_1, \tau_2 \in R. \tag{5.6}$$

Then, for two vectors $x, y \in \Delta_n$, we have

$$\frac{1}{2}\|x - y\|_1 = \frac{1}{2}\sum_{i=1}^n |x^{(i)} - y^{(i)}| \overset{(5.6)}{=} \sum_{i=1}^n \left[\frac{1}{2}(x^{(i)} + y^{(i)}) - \min\{x^{(i)}, y^{(i)}\}\right] = 1 - \langle x, y\rangle_{\min},$$

where $\langle x, y\rangle_{\min} = \sum_{i=1}^n \min\{x^{(i)}, y^{(i)}\}$. Thus, we have proved that

$$\|A\|_{01} = 1 - \mu(A), \quad \mu(A) \overset{\text{def}}{=} \min_{1\leqslant i,j\leqslant n} \langle Ae_i, Ae_j\rangle_{\min}, \quad A \in \mathcal{P}_n. \tag{5.7}$$

Let us mention several important properties of $\mu(A)$.

1. For any $\tau \geqslant 0$ we have $\mu(\tau A) = \tau\mu(A)$.

2. Function $\mu(A)$ is concave on $\mathcal{P}_n$ (since $\|\cdot\|_{01}$ is convex).
3. For $A, B \in \mathcal{P}_n$, we have $\mu(AB) \overset{(5.4)}{\geqslant} \max\{\mu(A), \mu(B)\}$.

In the remaining part of this section, we work only with square stochastic matrices.

Clearly, for stochastic matrices $\mu(A) \leqslant 1$. This value plays a crucial role in convergence analysis of the Power Method (2.3) with respect to 1-norm. Indeed,

$$\|x_{k+1} - x^*\|_1 = \|A(x_k - x^*)\|_1 \overset{(5.7)}{\leqslant} (1 - \mu(A))\|x_k - x^*\|_1.$$

Thus, we need to guarantee that $\mu(A) > 0$. Let us look at some examples.

- All elements of matrix A^TA are positive if and only if $\mu(A) > 0$.[4] The condition $A^TA > 0$ has an interesting interpretation in terms of Markov chains. Namely, it ensures that for two agents sitting at any pair of different nodes, there is a nonzero probability to meet after the next step.
- Clearly, $\mu(A) \geqslant \bar{\alpha} = \langle e, \bar{r}\rangle$. Therefore, Assumption 3.1 implies $\mu(A) > 0$ (but not vise versa).

On the other hand, let $p_k(\tau), \tau \in R$, be a polynomial of degree k with nonnegative coefficients. Since matrix A has the same eigenvectors as $p_k(A)$, the uniqueness of its dominant eigenvector has more chances to be detected by applying Lemma 6 to matrix $p_k(A)/p_k(1)$. In order to show that this is always possible, we prove a stronger version of Lemma 3.

Theorem 2. *For any matrix $A \in \mathcal{P}_n$, its dominant eigenvector is unique if and only if there exists a polynomial $p_k(\tau), k \leqslant n-1$, with nonnegative coefficients, such that the matrix $p_k(A)$ has a positive row.*

Proof. In one direction, this statement follows from Lemma 3. Let us assume now that $x^* \in \Delta_n$ is the unique dominant eigenvector of matrix A.

Let us represent A in the Jordan form, $A = VJV^{-1}$ with $J_1 = 1$ (we use the notation of Section 4). Consider its characteristic polynomial

$$p(\tau) = \det(\tau I - A) = \prod_{i=1}^{m} (\tau - \lambda_i)^{k_i},$$

where λ_i are the eigenvalues of matrix A. Note that $\lambda_1 = 1$. Since the multiplicity of this eigenvalue is one, we have $p'(1) \neq 0$. Therefore, we can define the polynomial

$$\hat{p}(\tau) \overset{\text{def}}{=} \frac{p(\tau)}{p'(1)(t-1)} = \prod_{i=2}^{m} (\tau - \lambda_i)^{k_i} \Big/ \prod_{i=2}^{m} (1 - \lambda_i)^{k_i}.$$

Note that $\hat{p}(1) = 1$. Therefore, $\hat{p}(J) = \frac{1}{p'(1)} \prod_{i=2}^{m} (J - \lambda_i I)^{k_i} \overset{(4.1)}{=} e_1 e_1^T$, and we conclude that

$$\hat{p}(A) = V\hat{p}(J)V^{-1} \overset{(4.3)}{=} x^* e^T.$$

Representing now $\hat{p}$ as a difference of two polynomials $\hat{p}_+$ and $\hat{p}_-$, both with nonnegative coefficients, we obtain

$$\hat{p}_+(A) = \hat{p}_-(A) + x^* e^T \geqslant x^* e^T.$$

Thus, polynomial $\hat{p}_+$ has at least one positive row. Its degree does not exceed $n-1$. □

Note that Theorem 2 can be applied even to the matrices, which are not primitive (e.g., $A = xe^T$ with $x \in \Delta_n$, permutation matrices with full cycle, etc.).

Acknowledgement

The authors would like to thank Paul Van Dooren for interesting and useful discussions on numerical aspects of Linear Algebra. We are also indebted to the two anonymous referees for their useful comments.

References

[1] A. Berman, R. Plemmons, Nonnegative Matrices in the Mathematical Sciences, Academic Press, New York, 1979.
[2] S. Brin, L. Page, The anatomy of a large-scale hypertextual web search engine, Comput. Networks ISDN Syst. 30 (1–7) (1998) 107–117.
[3] T. Haveliwala, S. Kamvar, The Second Eigenvalue of the Google Matrix, Technical Report, Stanford University, 2003.
[4] A. Langville, C. Meyer, Google's PageRank and Beyond: The Science of Search Engine Rankings, Princeton University Press, 2006.
[5] J.H. Wilkinson, The Algebraic Eigenvalue Problem, Clarendon Press, Oxford, 1965.

[4] Indeed, for $\alpha, \beta \in [0, 1]$, we have $\sqrt{\alpha\beta} \geqslant \min\{\alpha, \beta\} \geqslant \alpha\beta$.