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

A R T I C L E   I N F O

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

A B S T R A C T

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:

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

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 $Ae_j \in \Delta_n \stackrel{\mathrm{def}}{=} \{x \in R^n_+ : \langle e, x\rangle = 1\}, j = 1, \ldots, n$. 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 $\langle e, x_{k+1}\rangle = \langle e, Ax_k\rangle \stackrel{(2.1)}{=} \langle e, x_k\rangle$. Thus, $x_k \in \Delta_n$ for all $k \geqslant 0$. The following result is well known.