



The Jump Start Power Method: A New Approach for Computing the Ergodic Projector of a Finite Markov Chain

Joost Berkhout¹ · Bernd F. Heidergott² 

Received: 8 April 2016 / Revised: 20 September 2017 / Accepted: 7 September 2018 /
Published online: 20 September 2018
© The Author(s) 2018

Abstract

This article presents a new numerical method for approximately computing the ergodic projector of a finite Markov chain. Our approach requires neither structural information on the chain, such as, the identification of ergodic classes, transient states, or qualitative information, such as whether the chain is nearly decomposable or not. The theoretical deduction of the new method is corroborated by an extensive numerical study.

Keywords Power method · Numerical evaluation · Markov multi-chains · Transient states · Nearly decomposable

1 Introduction

This paper studies aperiodic Markov chains defined on discrete state space $\mathbb{S} = \{1, \dots, S\}$, with $S \in \mathbb{N}$. For ease of presentation we will consider the case of aperiodic Markov chains. The extension of our results to the case of periodic chains is postponed to the “Appendix”. The steady-state behavior of an aperiodic Markov chain P is characterized through its ergodic projector Π_P , where

$$\Pi_P = \lim_{n \rightarrow \infty} P^n,$$

see [17,23]. Computing Π_P in the above way, i.e., by taking powers of P , is known as the *power method* (PM). Any finite aperiodic Markov chain P is geometrically ergodic, i.e., there

Part of this work as done while Joost Berkhout was with the Department of Econometrics and Operations Research of the Vrije Universiteit Amsterdam.

✉ Bernd F. Heidergott
b.f.heidergott@vu.nl
Joost Berkhout
joost.berkhout@cwi.nl

¹ Centrum Wiskunde & Informatica (CWI), Science Park 123, 1098 XG Amsterdam, The Netherlands

² Department of Econometrics and Operations Research, Vrije Universiteit Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands

exist finite numbers r , c and $\beta \in (0, 1)$ such that

$$\forall n \geq r : \|P^n - \Pi_P\| \leq c\beta^n,$$

where $\|\cdot\|$ denotes the maximum absolute row sum norm (note that any matrix norm may be used), β is called the *rate* and r is called the *transient phase*, see, for example, [14,17]. Geometric ergodicity implies that PM enjoys a geometric rate of convergence once the powers exceed r .

The main advantages of the power method are that PM is easy to implement and that it requires no further information on P . In addition, PM can be efficiently implemented for large sparse matrices, which is the main reason why PM is used for the acclaimed Google PageRank algorithm introduced by Brin and Page [7], and for more detail see [4,9,20]. PM has two main versions. In the *vector-updating* version of PM, one computes μP^n for given vector μ . Vector-updating applies in case a given Markov chain \hat{P} with known stationary distribution $\pi_{\hat{P}}$ is updated (due to a change in the underlying hyperlink structure of the network) to a new Markov chain P . Then, computing $\pi_{\hat{P}} P^n$ converges faster towards π_P . An advantage of vector updating is that it only requires vector-matrix multiplications. The downside of this approach is that one cannot change the initial vector without a complete recalculation. The *matrix-updating* version directly computes P^n in order to approximate Π_P . The advantage of the matrix-updating PM is that by squaring a matrix power P^n , i.e., going from P^n to $(P^n)^2 = P^{2n}$, high powers of P can be relatively easily computed. Indeed, computing P^n only requires $\log_2(n)$ matrix multiplications. Moreover, applying different initial vectors to Π_P allows to model different initial distributions which is of particular interest in case of multi-chains, see the subsequent section for details. The downside is that even with the $\log_2(n)$ advantage, matrix updating may require a significant number of matrix multiplications and as the power increase these matrices are not sparse. In case P is periodic, both the vector-updating and the matrix-updating do not converge unless a convex combination of the original P with the identity matrix is used which comes at the expense of reduced convergence speed.

In this paper, we mainly focus on matrix-updating PM, from now on simply referred to as PM. Iterative methods, such as PM converge slowly in case the subdominant eigenvalue of P is close to 1, see [13,15]. This typically happens if either the P -chain only jumps with small probability from the transient states to (one of) the ergodic class(es) or if P is nearly decomposable. Roughly speaking, an irreducible chain P is called nearly decomposable if the state-space can be divided into classes so that the interactions between states are relatively frequent compared to interactions between the classes (a formal definition will be provided later in the text). It can be shown that an irreducible Markov chain without transient states is nearly decomposable if and only if the subdominant eigenvalue is close to 1, see [11]. A famous example of a nearly decomposable Markov chain is the so-called Courtois matrix, which is a 6×6 transition matrix for which PM requires $n \approx 69.000$ in order to provide an approximation of Π_P that is correct in first 6 digits, [25]. In case the ergodic classes and the transient states are known, one may compute the ergodic projector directly by first computing the equilibrium distribution for each ergodic class, and then the long-term behavior of the transient states, see [5,17] and the detailed discussion in Sect. 2. For a comprehensive overview of numerical methods for computing the ergodic projector of a finite Markov chain, we refer to [25].

Our research on Markov chains is stimulated by the growing interest in the analysis of social networks (where the Markov chain is used to model relationships among social agents, see [22]) and by the analysis of the world wide web, were based on the (bored-) random-surfer-concept, the Markov chain models the probability of randomly going from one page

to another, [9,20]. A key feature of these networks is that they are large and that neither their structure (transient states, ergodic classes) nor their balancedness (nearly decomposable or not) are known *a priori*. Other examples of these type of complex networks include telecommunication networks, cognitive and semantic networks and biological networks.

In this paper, we develop a novel approach for approximating the ergodic projector of an aperiodic Markov chain. We firstly establish a new representation of the ergodic projector of P through constructing an alternative Markov chain and then make this result useful for numerical computation. Starting point of our analysis is the known analytical relation

$$\Pi_P = \lim_{\alpha \downarrow 0} \alpha(I - (1 - \alpha)P)^{-1},$$

see, for example, Theorem 1.5 in [16], where the term $\alpha(I - (1 - \alpha)P)^{-1}$ is recognizable as the *resolvent kernel* of P . See also [19] for applications of the resolvent kernel in stability theory of Markov chains. We call the transformation

$$H_\alpha(P) = \alpha P(I - (1 - \alpha)P)^{-1},$$

for $\alpha \in (0, 1]$, the *modified resolvent kernel* of P . In particular, the resolvent kernel is modified so that it allows for efficient numerical evaluation. To see this note that in case of large P one solves a system of linear equations, for which P can be used as a basis, rather than computing the inverse explicitly. Letting X_α be a geometrically distributed random variable with parameter $\alpha \in (0, 1)$, the modified resolvent kernel can be written as

$$H_\alpha(P) = \mathbb{E}[P^{X_\alpha+1}],$$

since $\|(1 - \alpha)P\| < 1$, which suffices to show that $H_\alpha(P)$ is again a Markov transition matrix with the same ergodic projector as P for any $\alpha \in (0, 1)$, and since at $\alpha = 1$ it holds $H_\alpha(P) = P$, the statement holds for $\alpha = 1$ as well. In formula,

$$\Pi_P = \Pi_{H_\alpha(P)} = \lim_{k \rightarrow \infty} (H_\alpha(P))^k, \quad \alpha \in (0, 1]. \tag{1}$$

The main contribution of this paper is that we take (1) as starting point for developing a new numerical algorithm for computing Π_P . As main technical result we will show that

$$\|\Pi_P - (H_\alpha(P))^k\| \leq (\alpha\gamma(P))^k, \tag{2}$$

where $\gamma(P)$ a finite (possibly large) constant depending on P , to be defined later in the text.

Letting $\alpha < 1/\gamma(P)$, the result put forward in (2) implies that the Markov kernel $H_\alpha(P)$ is geometrically ergodic with transient phase $r = 1$ and rate $\alpha\gamma(P)$. Put differently, the transformation $P \mapsto H_\alpha(P)$ provides a jump start for PM as the desired contraction property is immediately effective. Moreover, we will show that iterating the transformation yields a geometric reduction in the geometric rate, so that, for example, $H_\alpha(H_\alpha(P))$ has a rate that is proportional to α^2 . The above theoretical results lead to a new numerical approach for approximately computing Π_P , called *jump start power method*.

The main contributions of this paper are as follows.

- The error of approximating Π_P by powers of the modified resolvent $(H_\alpha(P))^k$ is of order $(\alpha\gamma(P))^k$. We use this fact to introduce the jump start power method (JSPM) that enjoys the robustness of PM but overcomes the numerical deficiency of PM. JSPM works well for multi-chains, nearly decomposable chains, and chains that jump with small probability from the transient states to (one of) the ergodic class(es).

- An adapted version of JSPM is developed for large-scale Markov chains which utilizes the structure of the Markov chain and takes only ‘one jump’ towards Π_P , i.e., $k = 1$ together with a carefully chosen $\alpha \in (0, 1)$.
- An extensive numerical study is provided that corroborates the form of the analytical bound for the decay of the error and illustrates the numerical advantages of JSPM.

The article is organized as follows. Section 2 formally introduces Markov multi-chains, nearly decomposable Markov chains, and defines concepts used throughout the article. Section 3 presents the main technical results of the paper. Specifically, the approximate formula in (2) is derived. JSPM is presented in Sect. 4 together with a numerical study on the performance of the algorithm. The article concludes with a discussion of potential further research. The extension of our results to the case of periodic chains is presented in the ‘‘Appendix’’.

2 A Brief Review of Markov Chains

This paper studies aperiodic Markov chains defined on discrete state space $\{X_t\}_{t=0,1,\dots}$, with $\mathbb{S} = \{1, \dots, S\}$; see [17] for definitions. For ease of presentation we will first consider the case of aperiodic Markov chains. The extension of our results to the case of periodic chains is postponed to the ‘‘Appendix’’. For the (i, j) -th element of P it holds that $P(i, j) = \Pr(X_{t+1} = j | X_t = i)$ is independent of t and the past states, i.e., the probability distribution of the next state only depends on the current state. This leads to

$$(P^n)(i, j) = \Pr(X_n = j | X_0 = i), \quad \text{for all } (i, j) \in \mathbb{S} \times \mathbb{S},$$

which reads as the n -step transition probability of the Markov chain, where transition matrix P^n is simply obtained from taking the n -th matrix power of P . Taking n to infinity leads to the (i, j) -th element of the *ergodic projector*, denoted by Π_P , and defined by

$$\Pi_P(i, j) = \lim_{n \rightarrow \infty} (P^n)(i, j), \quad \text{for all } (i, j) \in \mathbb{S} \times \mathbb{S}.$$

Entry $\Pi_P(i, j)$ represents the probability of the chain being in state j in the long-run when starting in state i . For more details we refer to [17].

In case the Markov chain has only one closed irreducible set of states, also called *ergodic class*, and a (possibly empty) set of transient states, it is called a Markov uni-chain (in short: uni-chain). For uni-chains it holds that the chain will eventually be trapped in the (unique) ergodic class, independent of the initial state. The unique distribution to which a uni-chain converges is described by the stationary distribution of P denoted as π_P^\top which can be found by solving $\pi_P^\top P = \pi_P^\top$. Since the stationary distribution is independent of the initial state, all rows of Π_P equal π_P^\top in case P describes a Markov uni-chain.

Markov multi-chains (in short: multi-chains) have multiple ergodic classes and a (possibly empty) set of transient states. Other than for uni-chains, for multi-chains the initial state has an impact on the resulting limiting distribution, which stems from the fact that once the chain enters one of the several ergodic classes it remains there permanently. First of all, one has to uncover the ergodic classes and the transient states using, for example, the already mentioned algorithm in [8]. After possible relabelling of states, the transition matrix and the ergodic

projector of a multi-chain can be written in the following canonical forms, respectively,

$$P = \begin{bmatrix} P_1 & 0 & 0 & \cdots & 0 \\ 0 & P_2 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & P_I & 0 \\ P_{T1} & P_{T2} & \cdots & P_{TI} & P_{TT} \end{bmatrix} \quad \text{and} \quad \Pi_P = \begin{bmatrix} \Pi_1 & 0 & 0 & \cdots & 0 \\ 0 & \Pi_2 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \Pi_I & 0 \\ R_1 & R_2 & \cdots & R_I & 0 \end{bmatrix},$$

where I is the number of ergodic classes. For the i -th ergodic class, P_i gives the one step transition probabilities between ergodic states from the i -th ergodic class and Π_i gives a square matrix of which all rows equal the unique stationary distribution of the chain inside the i -th ergodic class. Specifically, all rows in Π_{P_i} equal $\pi_{P_i}^\top$, which is the unique probability vector satisfying $\pi_{P_i}^\top P_i = \pi_{P_i}^\top$. Note that all diagonal values of Π_P corresponding to ergodic states are non-zero contrary to the diagonal values of transient states which are zero, an insight that will be elaborated in Sect. 4.3. Hence, whether state i is ergodic or transient can be concluded from the value of entry (i, i) of Π_P . We call this criterion for ergodicity of a state the *diagonal criterion*.

Moreover, $R_i(j, k)$ gives the equilibrium probability of ending in ergodic state k (which is part of the i -th ergodic class) when starting in transient state j . In order to calculate R_i , define J as the number of transient states, I_T as the unity matrix of size J and $Z(j, i)$ as the probability of ending in the i -th ergodic class when starting in transient state j . Note that Z is a $J \times I$ matrix Z . It then holds that

$$Z = (I_T - P_{TT})^{-1} [P_{T1}e_1 \quad P_{T2}e_2 \quad \cdots \quad P_{TI}e_I],$$

where e_i is a column vector of ones of size equal to the number of states in ergodic class i ; see, e.g., [5]. Denote the i -th column of Z with $Z(\bullet, i)$, then it holds that $R_i = Z(\bullet, i)\pi_{P_i}^\top$.

In case there are multiple ergodic classes the stationary distribution fails to be unique. Indeed, any row of Π_P is a stationary distribution of the Markov chain. More specifically, denote the i -th row of Π_P by $\Pi_P(i, \bullet)$, then it holds that $\Pi_P(i, \bullet)$ is a probability distribution which satisfies $\Pi_P(i, \bullet)P = \Pi_P(i, \bullet)$. This implies that any convex combination of the rows is also a stationary distribution of the Markov chain, i.e., for $(\gamma_i)_{i \in \mathbb{S}} : \sum_{i=1}^S \gamma_i = 1$ and $\gamma_i \geq 0$, for all $i \in \mathbb{S}$, it holds that $\sum_{i=1}^S \gamma_i \Pi_P(i, \bullet)$ is a probability distribution which is invariant with respect to P . When an initial distribution μ^\top is considered, this convex combination is fixed (and given by μ^\top) meaning that there exists a unique stationary distribution for the chain started in μ^\top (describing the long-run behavior of the chain started in μ^\top), or, more formally, $\mu^\top \Pi_P$ is the unique stationary distribution satisfying $(\mu^\top \Pi_P)P = (\mu^\top \Pi_P)$ when starting in μ^\top . Literature concerning Markov multi-chains includes Markov decision processes from [23], series expansion of Markov chains [2,5,6] and singular perturbation analysis [1,12] where the underlying multi-chain structure is often known beforehand.

A Markov chain P is called *nearly decomposable* if P is irreducible and after possible relabeling of states can be written

$$P = \begin{bmatrix} P_{11} & P_{12} & \cdots & P_{1k} \\ P_{21} & P_{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & P_{(k-1)k} \\ P_{k1} & \cdots & P_{k(k-1)} & P_{kk} \end{bmatrix},$$

where the diagonal blocks $P_{ii}, i = 1, 2, \dots, k$, are square and have rows that sum up to $1 - \varepsilon$, with $\varepsilon > 0$ small.

A Markov chain may belong to all of the above types simultaneously. For example, a multi-chain with transient states may have an ergodic class that for itself constitutes a nearly decomposable chain. Below we illustrate this by means of a simple Markov chain.

Example 1 Let $p, q, r_1, r_2, r_3 \in (0, 1)$ and define transition matrix P on state-space $\{1, 2, 3, 4\}$ as

$$P = \begin{pmatrix} 1-p & p & 0 & 0 \\ q & 1-q & 0 & 0 \\ 0 & 0 & 1 & 0 \\ r_1 & r_2 & r_3 & 1 - \sum_{i=1}^3 r_i \end{pmatrix},$$

where $0 < \sum_{i=1}^3 r_i \leq 1$.

Markov chain P is noticeably a multi-chain with ergodic classes $\{1, 2\}$ and $\{3\}$. State 4 is transient. If p, q are small, then the submatrix describing the transitions within the ergodic class $\{1, 2\}$ becomes nearly decomposable. Similar when $\sum_{i=1}^3 r_i$ is small, state 4 is only weakly connected to states $\{1, 2, 3\}$.

The ergodic projector of P can be computed to be

$$\Pi_P = \begin{bmatrix} \frac{q}{p+q} & \frac{p}{p+q} & 0 & 0 \\ \frac{q}{p+q} & \frac{p}{p+q} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \frac{(r_1+r_2)q}{(\sum_{i=1}^3 r_i)(p+q)} & \frac{(r_1+r_2)p}{(\sum_{i=1}^3 r_i)(p+q)} & \frac{r_3}{\sum_{i=1}^3 r_i} & 0 \end{bmatrix}.$$

Note that when $p = q = \sum_{i=1}^3 r_i = 0$, we obtain $P = I$ and $\Pi_P = I$.

Recall that throughout this paper we let $\|\cdot\|$ denote the maximum absolute row sum norm. For any finite-state, aperiodic Markov chain P , there exists a finite number r such that

$$\forall n \geq r : \|P^n - \Pi_P\| \leq c\beta^n,$$

where $c = \sup_{l=0,1,\dots,r-1} \|P^l - \Pi_P\| < \infty$ and $\beta \in (0, 1)$; see for details [14]. This property is called *geometric ergodicity* and we will call r the *transient phase* and β the *rate*. For ease of references, we call r, c, β *ergodicity parameters* of P .

In this paper we will also study the impact of starting with a power P^q for the evaluation of Π_P and we introduce the following additional ergodicity parameters

$$\gamma(P, q) = \kappa(P, q)\phi(R, q) + \frac{c\beta^{q(\phi(P,q)+1)}}{1 - \beta^q}, \tag{3}$$

where

$$\kappa(P, q) = \sup_{n=q, q+1, \dots, q\phi(P,q)} \|P^n - \Pi_P\|,$$

and

$$\phi(P, q) = \left\lceil \frac{\max\{r - q, 0\}}{q} \right\rceil.$$

For simplicity, we wrote in the introduction $\gamma(P)$ instead of $\gamma(P, 1)$.

3 Bounding the Approximation Error

Starting point of our analysis is the equality

$$\Pi_P = (1 - \alpha)\Pi_P + \alpha\Pi_P, \quad \text{for } \alpha \in [0, 1].$$

Adding and subtracting αP^q to the right hand side gives

$$\Pi_P = \alpha P^q + (1 - \alpha)\Pi_P P^q + \alpha(\Pi_P - P^q),$$

where we used for the second term on the right hand side that $\Pi_P P = \Pi_P$. Inserting N times this last expression for Π_P into the first Π_P on the right hand side leads to

$$\Pi_P = \alpha P^q \sum_{n=0}^N ((1 - \alpha)P^q)^n + \Pi_P ((1 - \alpha)P^q)^{N+1} + \alpha(\Pi_P - P^q) \sum_{n=0}^N ((1 - \alpha)P^q)^n. \tag{4}$$

For $N, q \in \mathbb{N}$ and $\alpha \in (0, 1)$ let

$$G_\alpha(N, P^q) = \frac{\sum_{n=0}^N ((1 - \alpha)P^q)^n}{1 - (1 - \alpha)^{N+1}}.$$

Note that for $\alpha \in (0, 1)$ and $q \geq 1$ it holds that

$$G_\alpha(P^q) := \lim_{N \rightarrow \infty} G_\alpha(N, P^q) = \sum_{n=0}^{\infty} ((1 - \alpha)P^q)^n = (I - (1 - \alpha)P^q)^{-1},$$

where existence of the Neumann series is guaranteed since $\|(1 - \alpha)P^q\| < 1$, for $\alpha \in (0, 1)$. Equation (4) can be rewritten in succession via (i) simplifying the second term, (ii) bringing the second term of the right hand side to the other side, (iii) dividing by $1 - (1 - \alpha)^{N+1}$ and (iv) using the $G_\alpha(N, P^q)$ -notation:

$$\Pi_P = \alpha P^q G_\alpha(N, P^q) + \alpha(\Pi_P - P^q)G_\alpha(N, P^q). \tag{5}$$

Remark 1 The (i, j) -th element of $G_\alpha(N, P^q)$ gives the scaled $(1 - \alpha)$ -discounted expected number of visits of the Markov chain with transition matrix P^q to the j -th state in the first $N + 1$ number of discrete time steps (including the state i at time zero) when starting in state i . Intuitively, the discounting ensures that the weights of the visits after many discrete time steps of the Markov chain with transition matrix P^q becomes smaller and smaller, ensuring existence of $H_\alpha(P^q)$ since $\|(1 - \alpha)P^q\| < 1$, for $\alpha \in (0, 1)$.

Post-multiplying Eq. (5) with

$$(\alpha P^q G_\alpha(N, P^q))^{k-1}, \quad \text{for } k \in \mathbb{N}$$

i.e., the $k - 1$ power of the first term of the right hand side of (5), gives

$$\Pi_P = (\alpha P^q G_\alpha(N, P^q))^k + \alpha^k (\Pi_P - P^q)^k (G_\alpha(N, P^q))^k, \tag{6}$$

where we used that

$$(\alpha P^q G_\alpha(N, P^q))^{k-1} \Pi_P = \left(\frac{\alpha \sum_{n=0}^N (1 - \alpha)^n}{1 - (1 - \alpha)^{N+1}} \right)^{k-1} \Pi_P = \Pi_P.$$

Taking the limit $N \rightarrow \infty$ in (6) leads to

$$\Pi_P = (H_\alpha(P^q))^k + \alpha^k (\Pi_P - P^q)^k (G_\alpha(P^q))^k,$$

where we use the notation

$$H_\alpha(P^q) := \alpha P^q G_\alpha(P^q),$$

which is the modified resolvent kernel of P . Analogous to $G_\alpha(\cdot)$ let in the following

$$H_\alpha(N, P^q) := \alpha P^q G_\alpha(N, P).$$

Lemma 1 For $k, q \in \mathbb{N}$, $N \geq \phi(P, q)$, and $\alpha \in (0, 1)$ it holds that

$$\left\| \Pi_P - (H_\alpha(N, P^q))^k \right\| \leq \left(\frac{\alpha \gamma(P, q)}{1 - (1 - \alpha)^{N+1}} \right)^k.$$

Proof From (6) it follows that

$$\Pi_P - (H_\alpha(N, P^q))^k = \alpha^k (\Pi_P - P^q) P^{q(k-1)} (G_\alpha(N, P^q))^k. \tag{7}$$

Since $(\Pi_P - P^q)\Pi_P = 0$, it holds that

$$(\Pi_P - P^q) P^{q(k-1)} = (-1)^{k-1} (\Pi_P - P^q)^k,$$

so that (7) can be written as

$$\begin{aligned} \Pi_P - (H_\alpha(N, P^q))^k &= \alpha^k (-1)^{k-1} (\Pi_P - P^q)^k (G_\alpha(N, P^q))^k \\ &= - \left[-\alpha (\Pi_P - P^q) \frac{\sum_{n=0}^N ((1 - \alpha) P^q)^n}{1 - (1 - \alpha)^{N+1}} \right]^k, \end{aligned} \tag{8}$$

where the definition of $G_\alpha(N, P^q)$ is filled in in the last equation. Applying norms to Eq. (8) we get

$$\begin{aligned} &\| \Pi_P - (H_\alpha(N, P^q))^k \| \\ &\leq \left\| \alpha (\Pi_P - P^q) \frac{\sum_{n=0}^N ((1 - \alpha) P^q)^n}{1 - (1 - \alpha)^{N+1}} \right\|^k \\ &\leq \left(\frac{\alpha \sum_{n=0}^N (1 - \alpha)^n \| \Pi_P - P^{q(n+1)} \|}{1 - (1 - \alpha)^{N+1}} \right)^k \\ &= \left(\frac{1}{1 - (1 - \alpha)^{N+1}} \right)^k \left\{ \alpha \sum_{n_1=0}^{\min\{N, \phi(P, q)-1\}} (1 - \alpha)^{n_1} \| \Pi_P - P^{q(n_1+1)} \| \right. \\ &\quad \left. + \alpha \sum_{n_2=\phi(P, q)}^N (1 - \alpha)^{n_2} \| \Pi_P - P^{q(n_2+1)} \| \right\}^k, \end{aligned} \tag{10}$$

where the summation is split at $\phi(P, q)$ into two summations, one where geometric ergodicity does not apply and one where it does, respectively. Continuing calculations from (10) shows

$$\begin{aligned} &\| \Pi_P - (H_\alpha(N, P^q))^k \| \\ &\leq \left(\frac{1}{1 - (1 - \alpha)^{N+1}} \right)^k \left\{ \sup_{n=0,1,\dots,\min\{N, \phi(P, q)-1\}} \| \Pi_P - P^{q(n+1)} \| \left[1 - (1 - \alpha)^{\min\{N+1, \phi(P, q)\}} \right] \right. \\ &\quad \left. + \alpha c \beta^q ((1 - \alpha) \beta^q)^{\phi(P, q)} \frac{1 - ((1 - \alpha) \beta^q)^{\max\{N - \phi(P, q) + 1, 0\}}}{1 - (1 - \alpha) \beta^q} \right\}^k. \end{aligned} \tag{11}$$

So we may conclude from (11) that for

1. $N \leq \phi(P, q) - 1$ (geometric ergodicity does not apply):

$$\| \Pi_P - (H_\alpha(N, P^q))^k \| \leq \left(\sup_{n=0,1,\dots,N} \| \Pi_P - P^{q(n+1)} \| \right)^k \tag{12}$$

2. $N \geq \phi(P, q)$ (geometric ergodicity applies): since

$$1 - (1 - \alpha)^{\phi(P,q)} \leq \alpha \phi(P, q), \quad \text{for } \alpha \in (0, 1) \text{ and } \phi(P, q) = 0, 1, \dots,$$

it holds that

$$\| \Pi_P - (H_\alpha(N, P^q))^k \| \leq \left(\frac{\alpha \gamma(P, q)}{1 - (1 - \alpha)^{N+1}} \right)^k,$$

where $\gamma(P, q)$ is a finite constant defined in (3).

Note that it is necessary for the bound to be meaningful that $N \geq \phi(P, q)$ so that the geometric ergodicity applies. □

Remark 2 For notational easiness define the bound found in Lemma 1 in case $N \geq \phi(P, q)$ as $f(\alpha) = \left(\frac{\alpha \gamma(P,q)}{1 - (1 - \alpha)^{N+1}} \right)^k$. It holds that

$$\lim_{\alpha \downarrow 0} f(\alpha) = \left(\frac{\gamma(P, q)}{N + 1} \right)^k,$$

and

$$\lim_{\alpha \uparrow 1} f(\alpha) = (\gamma(P, q))^k,$$

so that $\lim_{\alpha \downarrow 0} f(\alpha) < \lim_{\alpha \uparrow 1} f(\alpha)$ for $k, q \in \mathbb{N}$ and $N \geq \phi(P, q)$. Furthermore, since¹

$$\frac{d}{d\alpha} f(\alpha) = k \alpha^{k-1} \gamma(q)^k \frac{1 - (1 + \alpha N)(1 - \alpha)^N}{[1 - (1 - \alpha)^{N+1}]^{k+1}} > 0, \quad \text{for } \alpha \in (0, 1),$$

it holds that for any choice of k, q and $N \geq \phi(P, q)$ it is optimal to choose $\alpha \in (0, 1)$ as small as possible.

The following theorem summarizes some properties of $H_\alpha(P^q)$.

Theorem 1 *It holds for $k, q \geq 1$ and $\alpha \in (0, 1)$ that*

$$\| \Pi_P - (H_\alpha(P^q))^k \| \leq (\alpha \gamma(P, q))^k \tag{13}$$

and

$$\Pi_P = \lim_{\alpha \downarrow 0} (H_\alpha(P^q))^k = \lim_{q \rightarrow \infty} (H_\alpha(P^q))^k = \lim_{k \rightarrow \infty} (H_\alpha(P^q))^k. \tag{14}$$

Proof Inequality (13) follows directly from Lemma 1 by letting $N \rightarrow \infty$. The first two equalities from (14) follow from Inequality (13) and the third equality from (1). □

¹ Note that $(1 + \alpha N)(1 - \alpha)^N \leq (1 + \alpha)^N (1 - \alpha)^N = (1 - \alpha^2)^N < 1$ for $\alpha \in (0, 1)$ and $N \in \mathbb{N}$.

Remark 3 Theorem 3 in the “Appendix” shows that the results put forward in Lemma 1 and Theorem 1 apply to periodic Markov chains with period d for $q = 1$ when $\gamma(P, 1)$ is replaced by $\bar{\gamma}(P, d)$ defined in the “Appendix”.

The result put forward in Theorem 1 shows that for $\alpha < \gamma(P, q)$ it holds that the modified resolvent $H_\alpha(P^q)$ is geometrically ergodic with rate $\alpha\gamma(P, q)$, transient phase $r = 1$, and ergodic projector Π_P .

As our numerical study in the second part of the paper shows, the modified resolvent is potentially more efficient than PM, which makes it, apart from the fact that it directly applies to multi-chains, an attractive alternative to PM. In the following, $H_\alpha(P)$ is illustrated for Example 1.

Example 2 We revisit Example 1 where we assume that p, q, r_1, r_2 and r_3 are non-zero probabilities. For this chain $H_\alpha(P)$ can be explicitly solved to be

$$H_\alpha(P) = \begin{bmatrix} \frac{\alpha(1-p)+(1-\alpha)q}{\alpha+(1-\alpha)(p+q)} & \frac{p}{\alpha+(1-\alpha)(p+q)} & 0 & 0 \\ \frac{q}{\alpha+(1-\alpha)(p+q)} & \frac{\alpha(1-q)+(1-\alpha)p}{\alpha+(1-\alpha)(p+q)} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \frac{\alpha r_1+(1-\alpha)q(r_1+r_2)}{(\alpha+(1-\alpha)(p+q))(\alpha+(1-\alpha)(\sum_{i=1}^3 r_i))} & \frac{\alpha r_2+(1-\alpha)p(r_1+r_2)}{(\alpha+(1-\alpha)(p+q))(\alpha+(1-\alpha)(\sum_{i=1}^3 r_i))} & \frac{r_3}{\alpha+(1-\alpha)(\sum_{i=1}^3 r_i)} & \frac{\alpha(1-\sum_{i=1}^3 r_i)}{\alpha+(1-\alpha)(\sum_{i=1}^3 r_i)} \end{bmatrix}.$$

Hence, letting α tend to zero yields element-wise convergence of $H_\alpha(P)$ to Π_P , which is in accordance with Theorem 1. For example, the absolute error of the (1, 1)-th element equals

$$\frac{\alpha p|p + q - 1|}{(p + q)(\alpha + (1 - \alpha)(p + q))}$$

so that the corresponding relative error is

$$\frac{|H_\alpha(P)(1, 1) - \Pi_P(1, 1)|}{\Pi_P(1, 1)} = \frac{\alpha p|p + q - 1|}{q(\alpha + (1 - \alpha)(p + q))},$$

where $H_\alpha(P)(i, j)$ indicates the (i, j) -th element of $H_\alpha(P)$. It shows that the relative error of $H_\alpha(P)(1, 1)$ can be bounded by the linear function $\alpha c_1(p, q)$, where $c_1(p, q) = p|p + q - 1|/\min\{q, q(p + q)\}$ is a (p, q) -dependent constant.

Furthermore, the asymptotic probabilities of going from one ergodic class to another (or to a transient state) are zero. This shows that $H_\alpha(P)$ uncovers the structure of the ergodic classes. In addition, the approximation assigns in general a positive mass to jumps from a transient state to itself, e.g.,

$$H_\alpha(P)(4, 4) = \frac{\alpha \left(1 - \sum_{i=1}^3 r_i\right)}{\alpha + (1 - \alpha) \left(\sum_{i=1}^3 r_i\right)},$$

which is strictly larger than zero if $\sum_{i=1}^3 r_i < 1$, while clearly $\Pi_P(4, 4) = 0$. Note that, besides the case $\sum_{i=1}^3 r_i = 1$, the approximation may give the wrong impression that a transient state, say i , is ergodic. However, when α is chosen sufficiently small (for example of the order 10^{-8}), together with the fact that there are no transitions from ergodic states towards i , the fact that i is transient becomes apparent.

In the following, we analyze the effect that taking a power of the modified resolvent has on the convergence. Denote $(H_\alpha(P))^2(1, 1)$ as the (1, 1)-th element from $(H_\alpha(P))^2$. It then holds that

$$(H_\alpha(P))^2(1, 1) = \frac{(\alpha(1 - p) + (1 - \alpha)q)^2 + pq}{(\alpha + (1 - \alpha)(p + q))^2},$$

the relative error of which can be computed to be equal to

$$\frac{|(H_\alpha(P))^2(1, 1) - \Pi_P(1, 1)|}{\Pi_P(1, 1)} = \frac{\alpha^2 |(p + q) ((1 - p)^2 + pq) - q|}{q(\alpha + (1 - \alpha)(p + q))^2},$$

which can be bounded by the quadratic function $\alpha^2 c_2(p, q)$ where

$$c_2(p, q) = \frac{\alpha^2 |(p + q) ((1 - p)^2 + pq) - q|}{q(\min\{1, p + q\})^2}.$$

Note that when $p + q = 1$ or $p = q = 0$ the relative errors of approximation $H_\alpha(P)(1, 1)$ and $(H_\alpha(P))^2(1, 1)$ are zero. For $p + q \in (0, 2] \setminus \{1\}$ and $\alpha \in (0, 1)$ the relative error of $(H_\alpha(P))^2(1, 1)$ is strictly smaller than that of $H_\alpha(P)(1, 1)$. Furthermore, comparing the relative error bounds $\alpha c_1(p, q)$ and $\alpha^2 c_2(p, q)$ shows the quadratic improvement of $(H_\alpha(P))^2(1, 1)$, which is in accordance with Theorem 1. This illustrates the improvement that can be achieved through the power k in the generalization. The other entries of $H_\alpha(P)$ can be analyzed along the same lines.

In the following example, we discuss the convergence of $H_\alpha(P)$ in case of a nearly decomposable Markov chain.

Example 3 In the light of nearly decomposable Markov chains it is interesting to see what happens in case $p + q \downarrow 0$ for the Markov chain in Example 1, i.e., when p and q are both close to 0. L'Hôpital's rule shows that the relative error of the $(1, 1)$ -th element for $p + q \downarrow 0$ in the limit equals

$$\frac{\alpha}{1 - \alpha} \frac{p}{q},$$

see also the relative errors from Example 1. Similar, the relative error of $H_\alpha(P)(1, 2)$ converges for $p + q \downarrow 0$ towards

$$\frac{\alpha}{1 - \alpha}.$$

Both relative errors show that arbitrary accuracy can be achieved by using the modified resolvent with α small even in case of nearly decomposable Markov chains.

Now consider the case where $\sum_{i=1}^3 r_i = \varepsilon$, for $\varepsilon > 0$ small. In that case the Markov chain breaks almost up into 3 ergodic classes. For the $(4, 4)$ -th element it holds that

$$H_\alpha(P)(4, 4) = \frac{\alpha(1 - \varepsilon)}{\alpha + (1 - \alpha)\varepsilon},$$

which equals the absolute error, since $\Pi_{P_1}(4, 4) = 0$. Choosing α such that

$$\alpha < \frac{\varepsilon\delta}{1 - \delta - (1 - \delta)\varepsilon}$$

leads to an absolute error smaller than δ , showing that arbitrary accurate precision can be achieved with $H_\alpha(P)$ even in case the Markov chain almost breaks up into 3 ergodic classes. Similar for the $(4, 3)$ -th element of $H_\alpha(P)$ it can be shown that the relative error equals

$$\frac{|H_\alpha(P)(4, 3) - \Pi_P(4, 3)|}{\Pi_P(4, 3)} = 1 - \frac{\varepsilon}{\alpha + (1 - \alpha)\varepsilon},$$

showing that in order to obtain a relative error smaller than η one should choose $\alpha \leq \varepsilon\eta$, again showing that any accuracy can be achieved in theory. Note that $H_\alpha(P)(4, 1) = H_\alpha(P)(4, 2) = H_\alpha(P)(4, 3) = 0$ in case $\sum_{i=1}^3 r_i = 0$, i.e., in case the Markov chain consists of three ergodic classes this is correctly detected.

Alternatively to PM applied to $H_\alpha(P)$, one may compute the modified resolvent of $H_\alpha(P)$. More specifically, one may construct recursively a sequence $\{H_\alpha(P; n) : n \in \mathbb{N}\}$ of *nested modified resolvents* with $H_\alpha(P; 0) = P$ and, for $n \geq 1$

$$H_\alpha(P; n) = H_\alpha(H_\alpha(P; n - 1)).$$

As the following theorem shows, the norm error of $H_\alpha(P^q; n)$ can be bounded by a geometric function with power n and rate α .

Theorem 2 For $\alpha \in (0, 1)$ such that $\alpha\gamma(P, q) < 1$ it holds that

$$\|\Pi_P - H_\alpha(P^q; n)\| \leq \frac{\gamma(P, q)\alpha^n}{1 - \alpha\gamma(P, q)(1 - \alpha^{n-1})}, \quad n \in \mathbb{N},$$

and

$$\lim_{n \rightarrow \infty} H_\alpha(P^q; n) = \Pi_P.$$

Proof Proof via mathematical induction. Because $\alpha\gamma(P, q) < 1$ it is clear that the bound holds true for $n = 1$ via Theorem 1. Now suppose it holds true for general $n - 1 \geq 1$, then

$$\|\Pi_P - H_\alpha(P^q; n)\| = \|\Pi_P - H_\alpha(H_\alpha(P^q; n - 1))\|$$

since $\|(1 - \alpha)H_\alpha(P^q; n - 1)\| < 1$ we can write out the inverse and bring Π_P inside the summation

$$= \left\| \alpha \sum_{l=0}^{\infty} (1 - \alpha)^l (\Pi_P - (H_\alpha(P^q; n - 1))^{l+1}) \right\|$$

straightforward bounding

$$\leq \alpha \sum_{l=0}^{\infty} (1 - \alpha)^l \|\Pi_P - (H_\alpha(P^q; n - 1))^{l+1}\|$$

since $\Pi_P H_\alpha(P^q; n - 1) = \Pi_P$

$$\leq \alpha \sum_{l=0}^{\infty} (1 - \alpha)^l \|\Pi_P - H_\alpha(P^q; n - 1)\|^{l+1}$$

filling in the induction hypothesis

$$= \frac{\gamma(P, q)\alpha^n}{1 - \alpha\gamma(P, q)(1 - \alpha^{n-2})} \sum_{l=0}^{\infty} \left(\frac{(1 - \alpha)\gamma(P, q)\alpha^{n-1}}{1 - \alpha\gamma(P, q)(1 - \alpha^{n-2})} \right)^l$$

for all $n - 1 \geq 1$ when $\alpha\gamma(P, q) < 1$ it holds that $\frac{(1 - \alpha)\gamma(P, q)\alpha^{n-1}}{1 - \alpha\gamma(P, q)(1 - \alpha^{n-2})} < 1$ and thus

$$= \frac{\gamma(P, q)\alpha^n}{1 - \alpha\gamma(P, q)(1 - \alpha^{n-2}) - (1 - \alpha)\gamma(P, q)\alpha^{n-1}}$$

taking out $\alpha\gamma(P, q)$ in the denominator gives

$$= \frac{\gamma(P, q)\alpha^n}{1 - \alpha\gamma(P, q)(1 - \alpha^{n-1})}$$

thereby showing that it holds for n and thus ends the proof. □

Corollary 1 For $\alpha \in (0, 1)$ such that $\alpha\gamma(P, q) < 1$ it directly follows from the above theorem that

$$\|\Pi_P - H_\alpha(P; n)\| \leq \frac{\varepsilon}{1 - \varepsilon} \alpha^{n-1}, \quad n \geq 1, \tag{15}$$

when we define $\varepsilon = \alpha\gamma(P, q)$. Furthermore, since $H_\alpha(P; n)\Pi_P = \Pi_P$ it holds for $k \geq 1$ that

$$\begin{aligned} \|\Pi_P - (H_\alpha(P; n))^k\| &= \|(\Pi_P - H_\alpha(P; n))^k\| \\ &\leq (\|\Pi_P - H_\alpha(P; n)\|)^k \\ &= \left(\frac{\varepsilon}{1 - \varepsilon} \alpha^{n-1}\right)^k, \end{aligned}$$

where in the last Eq. (15) is used.

Remark 4 The results put forward in Theorem 2 and Corollary 1 apply for the case $q = 1$ also to periodic Markov chains with period d when $\gamma(P, 1)$ is replaced by $\bar{\gamma}(P, d)$. For details see the ‘‘Appendix’’.

Theorem 2 shows that repeated application of the modified resolvent yields a geometric improvement of the rate of geometric ergodicity. Example 4 illustrates Theorem 2.

Example 4 We revisit the instance from Examples 1 and 2 where we assume that p, q, r_1, r_2 and r_3 are non-zero probabilities. For this chain $H_\alpha(P; 2)$ can be explicitly solved to be

$$H_\alpha(P; 2) = \begin{bmatrix} \frac{q+\alpha^2(1-p-q)}{p+q+\alpha^2(1-p-q)} & \frac{p}{p+q+\alpha^2(1-p-q)} & 0 & 0 \\ \frac{q}{p+q+\alpha^2(1-p-q)} & \frac{p+\alpha^2(1-p-q)}{p+q+\alpha^2(1-p-q)} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \frac{(r_1+r_2)q(1-\alpha^2)+r_1\alpha^2}{((p+q)(1-\alpha^2)+\alpha^2)\left((1-\alpha^2)\sum_{i=1}^3 r_i + \alpha^2\right)} & \frac{(r_1+r_2)p(1-\alpha^2)+r_2\alpha^2}{((p+q)(1-\alpha^2)+\alpha^2)\left((1-\alpha^2)\sum_{i=1}^3 r_i + \alpha^2\right)} & \frac{r_3}{(1-\alpha^2)\sum_{i=1}^3 r_i + \alpha^2} & \frac{(1-\sum_{i=1}^3 r_i)\alpha^2}{(1-\alpha^2)\sum_{i=1}^3 r_i + \alpha^2} \end{bmatrix}.$$

This shows that $H_\alpha(P; 2)$ converges with quadratic rate in terms of α towards Π_P . Consequently, a larger rate of convergence is achieved for $H_\alpha(P; 2)$ than for $H_\alpha(P)$; compare with Example 2. More specifically, the relative error of the (1, 1)-th element equals

$$\frac{|H_\alpha(P; 2)(1, 1) - \Pi_P(1, 1)|}{\Pi_P(1, 1)} = \left| \frac{p(1 - p - q)}{q(p + q) + \alpha^2 q(1 - p - q)} \right| \alpha^2,$$

where the absolute term converges to constant $\frac{p|1-p-q|}{q(p+q)}$ for α small. Similar, the relative error of $H_\alpha(P; 3)(1, 1)$ equals

$$\frac{|H_\alpha(P; 3)(1, 1) - \Pi_P(1, 1)|}{\Pi_P(1, 1)} = \left| \frac{p(1 - p - q)}{q(p + q) + \alpha^3 q(1 - p - q)} \right| \alpha^3,$$

showing that the relative error error is approximately a factor α smaller than that of $H_\alpha(P; 2)(1, 1)$ in accordance with what can be expected from Theorem 2.

Unfortunately, as $\gamma(P, q)$ is not available, it is neither clear what a good initial choice for α is, nor when to terminate $(H_\alpha(P))^k$ or the repeated application of $H_\alpha(P; n)$. In the following, we will address these two issues in more detail.

Remark 5 In [24] the resolvent of a Markov chain P is defined as

$$R^\rho = (\rho I + (I - P))^{-1} = ((1 + \rho)I - P)^{-1},$$

for $\rho > 0$. Let $\rho = \alpha/(1 - \alpha)$, then

$$R^\rho = (1 - \alpha)(I - (1 - \alpha)P)^{-1},$$

which yields

$$G_\alpha(P) = \frac{1}{1 - \alpha} R^\rho.$$

Puterman develops in [24] the resolvent into a Laurent series yielding

$$\rho R^\rho = \Pi_P - \sum_{n=0}^{\infty} (-\rho D_P)^{n+1},$$

in our notation with $\rho = \alpha/(1 - \alpha)$

$$\Pi_P - \alpha G_\alpha(P) = \sum_{n=0}^{\infty} \left(-\frac{\alpha}{1 - \alpha} D_P \right)^{n+1}.$$

Note that the series on the above right hand side involves the deviation matrix $D_P = (I - P + \Pi_P)^{-1} - \Pi_P$ and that deriving efficient bounds for the norm of the deviation is itself a demanding task in case of multi-chains. As the norm of the deviation matrix typically takes large values, the above series only applies for significantly small values of α .

4 The Jump Start Power Method

In the previous section, we have shown that going from P to the modified resolvent $H_\alpha(P)$ potentially yields a geometrically ergodic Markov chain with no transient phase (i.e., $r = 1$). In this section we show how this result can be made fruitful for numerical computations. In particular, Sect. 4.1 illustrates the modified resolvent theory through numerical experiments, Sect. 4.2 develops a practical method that exploits the developed theory by introducing the jump start power method (JSPM) and provides numerical results. Lastly, Sect. 4.3 discusses and numerically illustrates the use of JSPM in case of large (sparse) systems.

4.1 Motivating Numerical Experiments

As a first step we analyze the effect of mapping P to $H_\alpha(P)$ by comparing numerically P^n with $H_\alpha(P)$. The considered instances cover a wide range of Markov chains and for an overview, we refer to Table 1. Each row in Table 1 corresponds to an instance defined by its transition matrix (Tr. Matrix). The instances are based on random graph models that capture key properties of real-life networks. The instances vary in terms of size S (given in the 'S' column), structure (given in the 'Ergodic Structure' column), connectivity (as indication, column ' p^* ' gives the smallest non-zero element of P), and parameters used for

random graph models (given in the 'Description' column). The ergodic structure is denoted by $([v_1, v_2, \dots, v_I], T)$, where I is the number of ergodic classes, v_i the number of states in the i -th ergodic class and $T = S - \sum_{i=1}^I v_i$ is the number of transient states. In case of transient states, the corresponding part in P is randomly filled such that transient states most likely point towards each other and multiple ergodic classes (if possible). The description column gives the relevant reference of the instance together with parameters given in the same order as they appeared in the original reference, where altered labels are used in case of conflicting notation (e.g., if α is used as parameter in the original reference, we refer to this parameter as β). For the implementation of the different instances, the code provided in MATLAB toolbox CONTEST [26] was used. The CONTEST toolbox generates symmetric adjacency matrices which are in many cases periodic. In order to obtain the corresponding transition matrix P , the rows are first normalized ensuring that each row sums up to one. Afterwards, the transition matrix P is mixed with the identity matrix to achieve aperiodicity, which does not affect the ergodic behavior of the chain.

Comparison of P^n and $H_\alpha(P)$ In the first numerical experiment, we compute for a series of Markov chains P_i , with $1 \leq i \leq 4$ in Table 1, the power $n_\alpha(P)$ such that

$$\left| \|\Pi_P - H_\alpha(P)\| - \|\Pi_P - P^{n_\alpha(P)}\| \right| < 10^{-12}.$$

In words, $n_\alpha(P)$ is the power of P that is substituted by $H_\alpha(P)$. Note that a power $n_\alpha(P)$ can be obtained via $\log_2 n_\alpha(P)$ matrix multiplications. The numerical results depicted in Fig. 1 show that the modified resolvent can replace PM approximations for large powers.

In particular for the Courtois matrix P_1 , in order to approximately achieve a norm error of $7.92 \cdot 10^{-7}$ a power is needed of 2^{16} while the same norm error is obtained via the modified resolvent with $\alpha \approx 10^{-10}$. For P_4 the modified resolvent with $\alpha \approx 10^{-11.18}$ leads to the same norm error (of approximately $1.63 \cdot 10^{-5}$) as PM with power 20655175 ($\approx 2^{24.3}$). As for computation times, experiments showed that on average PM $(P_4)^k$ with power $k = 20655175$ takes on average 73.12 seconds in a sparse matrix setting whereas the modified resolvent $H_{\alpha=10^{-11.18}}(P_4)$ takes on average 2.68 seconds, i.e., a difference of factor 27.28 on average (the experiments were performed in MATLAB R2011b on a 64-bit Windows desktop PC with Intel(R) Core(TM) i5-2310 CPU @ 2.90GHz processor).

Length of Transient Phase for $H_\alpha(P)$ Figure 2 illustrates the effect that powers of $H_\alpha(P)$ have on the norm error for the Courtois matrix. Different values for α are considered and for each α the exponential decay location is determined and thereby the length of the transient phase is identified. Note that $H_{\alpha=1}(P)$ equals P . A heuristic approach is used to find the exponential decay location where for each α an exponential function is repeatedly fitted to the data until the coefficient of determination R^2 is close enough to 1 (where $R^2 = 1$ represents a perfect fit). After each fit which leads to an insufficient coefficient of determination, the dataset is reduced by increasing the value of the first considered power n and the fitting repeats. The found exponential decay locations (i.e., the smallest power in the dataset that led to R^2 sufficiently close to 1) are denoted with the large dots and labeled, where the labels correspond to the fitted functions given under the graph together with the R^2 in parenthesis behind the function.

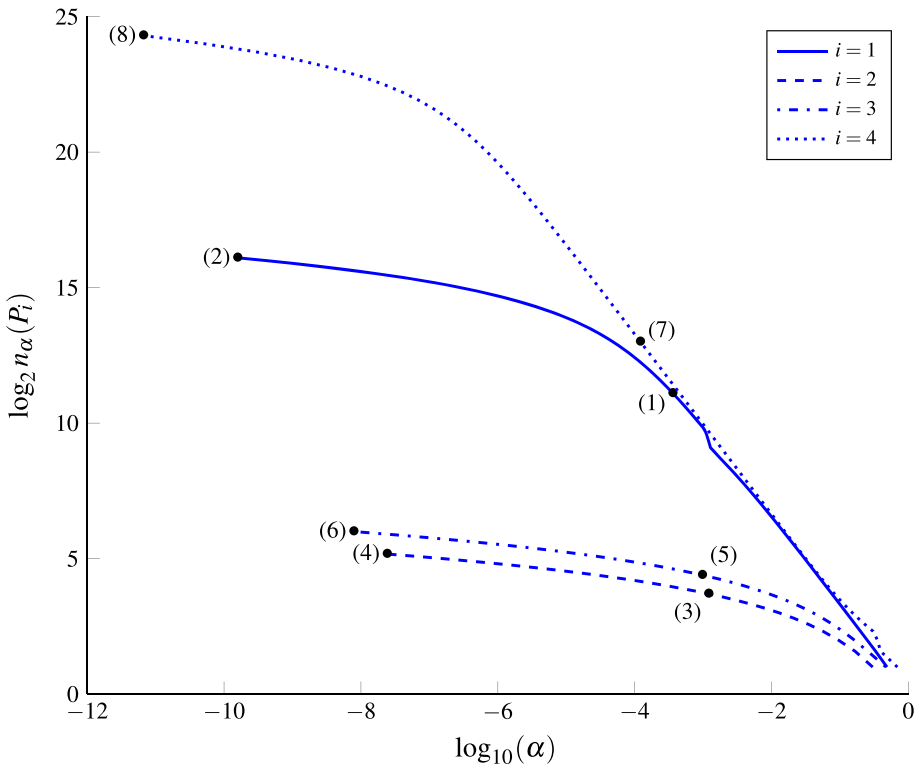
The main observation from Fig. 2 is that the exponential decay locations shift to the left (and thereby the transient phase becomes smaller) for decreasing values of α .

This phenomenon has been theoretically shown in the previous section. It is worth noting for $\alpha \leq 10^{-3}$ there is no transient phase, i.e., $r = 1$ in these cases. Furthermore, from the fitted functions it follows that smaller α values lead to stronger norm error reduction for increasing powers.

Table 1 Instances used for numerical experiments

Tr. Matrix	Description	S	Ergodic structure	p^*
P_1	Courtois matrix, [25].	8	(18], 0)	3e-5
P_2	Kleinberg's network, [18], with parameters $p = 5, q = 2$ and $\beta = 3$.	225	(225], 0)	1e-2
P_3	Kleinberg's network, [18], with parameters $p = 3, q = 6$ and $\beta = 1$.	900	(900], 0)	1e-2
P_4	Block diagonal transition matrix of P_2 and P_3 with weak connection between P_2 and P_3 . I.e., two random nodes from P_2 and P_3 , resp., are connected with probability $1e-4$.	1125	(1125], 0)	1e-4
P_5	Block diagonal transition matrix of P_2 and P_3 with 60 transient states.	1185	(225, 900], 60)	8.4e-6
P_6	Block diagonal transition matrix of P_1 and P_2 with 20 transient states.	253	(18, 225], 20)	3e-5
P_7	RENGA protein-protein interaction (PPI) network, [10], with parameters $\beta_1 = 0.4$ and $\beta_2 = 0.965$.	700	(202, 181, 67, 98, 152], 0)	1.7e-3
P_8	Preferential attachment network, [3], with parameter $d = 5$.	700	(700], 0)	4.7e-5
P_9	Lock and key PPI network, [21], with parameters $\beta_1 = 0.4$ and $\beta_2 = 0.965$.	700	(700], 0)	1.7e-4

For implementation the MATLAB code provided by CONTEST [26] was used



∞ -norm errors for some specific points:
 (1): 0.339; (2): $7.92e - 007$;
 (3): 0.000345; (4): $7.3e - 008$;
 (5): 0.00126; (6): $5.2e - 008$;
 (7): 1.58; (8): $1.63e - 005$;

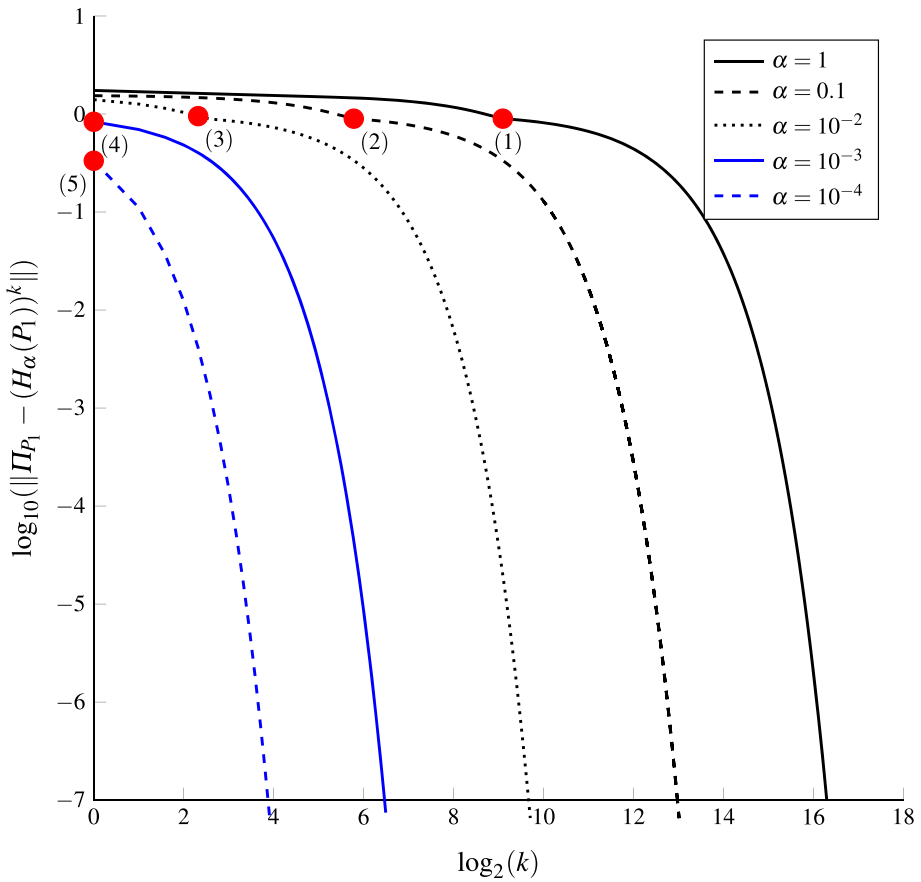
Fig. 1 Combinations of α and $n_\alpha(P_i)$ such that $\left\| \Pi_{P_i} - H_\alpha(P_i) \right\| - \left\| \Pi_{P_i} - (P_i)^{n_\alpha(P_i)} \right\| < 10^{-12}$ for $i = 1, 2, 3, 4$

The Nested Modified Resolvent $H_\alpha(P; n)$ Similar to Figs. 2, 3 illustrates the effect of the nested modified resolvent $H_\alpha(P; n)$ for varying n in case $P = P_1$.

It shows that relatively large values for α already lead to small norm errors after a few iterations. In particular, the fitted relation between the norm error of $H_{\alpha=0.01}(P; n)$ and n is approximately $4901e^{-4.6n}$ whereas that of $(H_{\alpha=0.01}(P))^n$ and n is approximately $e^{-0.0198n}$ (see also Fig. 2), showing that the effect of an increase in the number of iterations in the nested modified resolvent is far more effective than an increase in the power of the modified resolvent for the same α . It therefore illustrates the sharper bound found for the nested modified resolvent in comparison with powers of the modified resolvent.

4.2 Jump Start Power Method (JSPM)

In this section, we will develop a power-method like algorithm based on the theory established in the previous section. To that end we discuss how to choose α and we provide a stopping rule

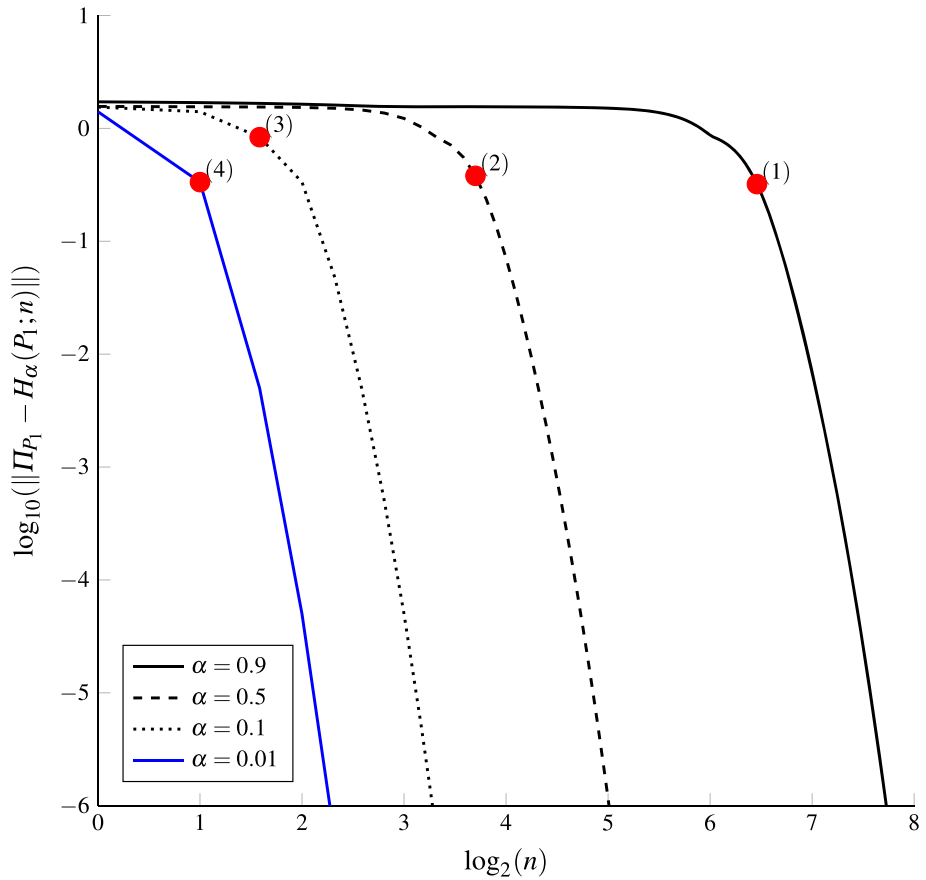


Fitted functions from exponential decay locations ● (R^2):
 (1): $1 \cdot e^{-0.00020002 \cdot k}$ (1)
 (2): $1 \cdot e^{-0.0019984 \cdot k}$ (1)
 (3): $1 \cdot e^{-0.019807 \cdot k}$ (1)
 (4): $1 \cdot e^{-0.18235 \cdot k}$ (1)
 (5): $1 \cdot e^{-1.0987 \cdot k}$ (1)

Fig. 2 Development norm errors for powers of modified resolvent in case of the Courtois matrix

for the algorithm. Our recommendations are based on numerical experiments and balance avoiding numerical issues with achieving good numerical approximations.

Numerical experiments indicate that in order to achieve high accuracy it is best to choose relatively large α and take powers of the resulting modified resolvent. Unfortunately, the modified resolvent is no sparse matrix and computing powers is rather costly. Alternatively, one may choose α (very) small and calculate the modified resolvent only once (without taking powers). This, however, may lead to numerical issues when approaching the machine precision. In particular, the condition number of matrix $(I - (1 - \alpha)P)$ grows for decreasing α , therefore, choosing α significantly small leads to an ill-conditioned matrix which is more difficult to invert accurately. For purpose of illustration consider the Courtois matrix P_1 . It then holds that



Fitted functions from exponential decay locations \bullet (R^2):
 (1): $926.3877 \cdot e^{-0.090026 \cdot n}$ (0.99903)
 (2): $1122.4366 \cdot e^{-0.60497 \cdot n}$ (0.9995)
 (3): $835.0589 \cdot e^{-1.9565 \cdot n}$ (0.99996)
 (4): $4901.0117 \cdot e^{-4.6002 \cdot n}$ (1)

Fig. 3 Development norm errors for nested modified resolvent in case of the Courtois matrix

$$\|\Pi_{P_1} - H_{\alpha=10^{-10}}(P_1)\| \approx 5.09 \cdot 10^{-7}.$$

According to the theoretical results decreasing α to, say, $\alpha = 10^{-12}$ should improve the quality of the approximation (see Theorem 1). But the contrary is true as

$$\|\Pi_{P_1} - H_{\alpha=10^{-12}}(P_1)\| \approx 3.48 \cdot 10^{-5}.$$

This shows that numerical issues come into play when computing the modified resolvent for $\alpha = 10^{-12}$ for P_1 . A similar effect can be observed for the nested modified resolvent

$$\|\Pi_{P_1} - H_{\alpha=10^{-3}}(P_1; 3)\| \approx 5.00 \cdot 10^{-6} \quad \text{compared to} \quad \|\Pi_{P_1} - H_{\alpha=10^{-4}}(P_1; 3)\| \approx 3.62 \cdot 10^{-5},$$

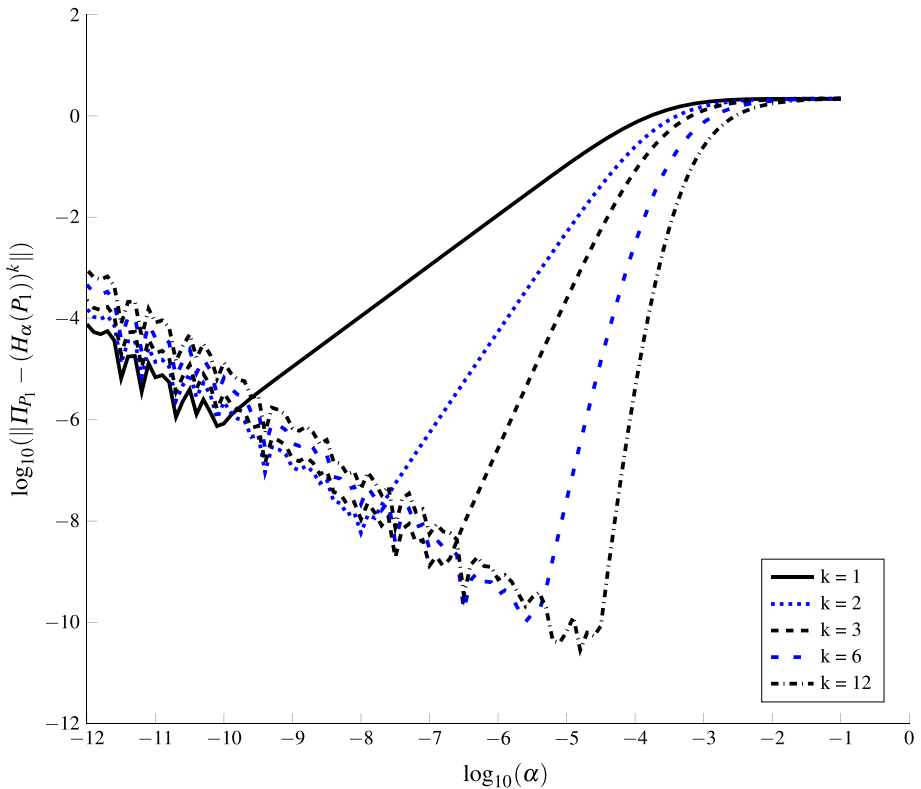


Fig. 4 Illustration of the norm errors for $(H_\alpha(P_1))^k$ with varying α and k

and even for PM with P_1 rounding errors can lead to numerical issues, i.e.,

$$\| \Pi_{P_1} - P_1^{10^5} \| \approx 2.06 \cdot 10^{-9} \quad \text{compared to} \quad \| \Pi_{P_1} - P_1^{10^{10}} \| \approx 1.22 \cdot 10^{-5}.$$

In Fig. 4 the norm errors of $(H_\alpha(P_1))^k$, i.e., $\| \Pi_{P_1} - (H_\alpha(P_1))^k \|$, are plotted for varying $\alpha \in (0, 1)$ and powers k . From Fig. 4 it follows for each k that choosing α too small leads to numerical issues and consequently leading to an increase of norm errors, contrary to what can be expected from theory. For example, for $k = 1$ numerical issues appear when choosing α smaller than 10^{-10} from where the norm errors start increasing in a zig-zag pattern. Furthermore, the figure shows that the smallest norm errors can be achieved by larger powers k and relatively larger α .

Based on the above results, we advise to use the modified resolvent in a PM framework with a carefully chosen α . When to terminate the power iterations is a delicate matter. A natural stopping rule is to terminate the algorithm when the improvement of an extra iteration becomes insignificant. More specifically, in order to find a power k such that $\| \Pi_P - P^k \| \leq \varepsilon$, one may terminate PM if $\| P^k P - P^k \| < \varepsilon$, for $\varepsilon > 0$ small. Unfortunately, this stopping rule may stop the algorithm too early as is illustrated in Example 5 below.

Example 5 For $\delta \in (0, 1)$ and $k \in \mathbb{N}$ let

$$P = \begin{bmatrix} 1 - \delta & \delta \\ 0 & 1 \end{bmatrix} \quad \text{so that} \quad \Pi_P = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad P^k = \begin{bmatrix} (1 - \delta)^k & 1 - (1 - \delta)^k \\ 0 & 1 \end{bmatrix}.$$

It can be shown that $\|P^k P - P^k\| = 2\delta(1 - \delta)^k$ and $\|\Pi_P - P^{k+1}\| = 2(1 - \delta)^k$, which implies that

$$\delta \|\Pi_P - P^{k+1}\| = \|P^k P - P^k\|.$$

Hence, $\|P^k P - P^k\| \leq \varepsilon$ only implies

$$\|\Pi_P - P^{k+1}\| \leq \frac{\varepsilon}{\delta},$$

which for small values of δ (e.g., $\delta < \varepsilon/2$) provides no insight and thereby showing that the stopping rule is insufficient.

To prevent a similar pitfall when using $\|(H_\alpha(P))^k H_\alpha(P) - (H_\alpha(P))^k\| < \varepsilon$ for $\alpha \in (0, 1)$ as stopping rule, we recommend choosing α such that

$$\alpha = \min\{\alpha_{\max}, p^*/N\}, \tag{16}$$

where α_{\max} is a user-defined upper bound for α , p^* denotes the minimal non-zero value of P , formally given by,

$$p^* := \min_{i,j} \{P(i, j) : P(i, j) > 0\},$$

and N is a user-defined scaling to ensure that α is significantly smaller than p^* . The intuition is that by choosing $\alpha \ll p^*$ the effect of the smallest transition is taken into account. As illustrated by our numerical examples, even for a nearly decomposable matrix, the minimal non-zero entry is typically not so small that α given in (16) leads to numerical instabilities for $H_\alpha(P)$. In addition, choosing α as in (16) typically leads to $H_\alpha(P)$ having no transient phase (i.e., $r = 1$).

The above considerations lead to the following *jump start power method* (JSPM).

- (1) CHOOSE $\alpha = \min\{\alpha_{\max}, p^*/N\}$ AND SELECT NUMERICAL PRECISION ε .
- (2) INITIALIZE $k = 1$ AND CALCULATE $H_\alpha(P)$.
- (3) SET $k = k + 1$.
- (4) IF

$$\|(H_\alpha(P))^{k-1} H_\alpha(P) - (H_\alpha(P))^{k-1}\| \geq \varepsilon$$

GO TO STEP 3. OTHERWISE GO TO STEP 5.

- (5) RETURN $(H_\alpha(P))^k$.

It is worth noting that rather than computing the resolvent in Step 2 of the JSPM directly, which requires evaluating the inverse of $I - (1 - \alpha)P$, it is numerically more efficient to solve $X(I - (1 - \alpha)P) = \alpha P$, which reduces the problem to that of solving systems of linear equations.

In Table 3 some numerical results for JSPM are shown. For an overview of the instances see Table 1. Two parameter choices for α and ε are considered, see Table 2. Parameter Setting 1 aims at achieving higher accuracy of the approximation (i.e., a small value for ε), which is numerically possible by choosing α not too small. Parameter Setting 2 focuses more on a quick convergence of the algorithm, i.e., a larger value for ε compared to setting 1 and small α .

From the results put forward in Table 3, it follows that significantly smaller norm errors can be achieved using Parameter Setting 1 but at the cost of more iterations (read, powers of $H_\alpha(P)$). Since the modified resolvent is typically not sparse taking powers of $H_\alpha(P)$ becomes impractical for large Markov chains. This issue will be the topic of the next subsection.

Table 2 Parameter settings used for JSPM in numerical experiments

	α	ε	Aim
Parameter setting 1:	$\min\{10^{-4}, p^*/100\}$	10^{-8}	High accuracy
Parameter setting 2:	$\min\{10^{-8}, p^*/S\}$	10^{-6}	Fast computation

Table 3 Results of JSPM for two parameter settings given in Table 2

Tr. Matrix	Parameter setting 1		Parameter setting 2	
	Norm error	# Iterations	Norm error	# Iterations
P_1	2.4461e−010	4	4.8780e−009	3
P_2	4.3119e−012	4	1.8567e−008	2
P_3	1.1869e−012	4	1.0155e−008	2
P_4	1.6029e−008	42	1.3002e−008	5
P_5	4.3520e−009	3	1.9519e−008	2
P_6	5.3403e−010	4	2.5750e−008	3
P_7	4.7504e−010	8	5.9758e−009	3
P_8	2.4177e−010	3	1.6035e−008	2
P_9	4.6657e−010	3	5.9593e−008	2

4.3 JSPM for Large Markov Chains

This final subsection discusses JSPM for large Markov chains. A common feature of large chains is that the transition matrix P is sparse but the ergodic projector Π_P is not due to connectivity [22]. This leads to numerical issues in approximating Π_P . In particular for JSPM: when the approximation $(H_\alpha(P))^k$ approaches Π_P as k is increasing, iterations become computational more expensive and a memory burden emerges due to the loss of sparsity.

Therefore, in case of large instances, our advice based on numerical experiments is to choose α significantly small and return $H_\alpha(P)$ as approximation, i.e., apply the JSPM for $k = 1$. In addition, instead of calculating $H_\alpha(P)$ as a whole, we recommend to calculate a concentrated version of $H_\alpha(P)$, denoted by $H_\alpha^c(P)$, where the computation of $H_\alpha^c(P)$ utilizes the structural properties of Π_P such as the fact that all rows corresponding to ergodic states from the same ergodic class are identical. In particular, when row i of $H_\alpha(P)$, denoted by $H_\alpha(P)(i, \bullet)$, is calculated, then based on this approximation it can be decided whether i is ergodic or transient by inspecting the value of $H_\alpha(P)(i, i)$. Indeed evoking the diagonal criterion, see Sect. 2, state i is ergodic if and only if $H_\alpha(P)(i, i)$ is significantly larger than 0. In case i is identified as ergodic, all indexes corresponding to (significantly) positive entries of $H_\alpha(P)(i, \bullet)$ are identified as belonging to the same ergodic class. Vector $H_\alpha(P)(i, \bullet)$ is saved in $H_\alpha^c(P)$ as approximation for the rows of the particular ergodic class and we are done considering all the indexes from this ergodic class. In case i is identified as transient, $H_\alpha(P)(i, \bullet)$ is saved in $H_\alpha^c(P)$ as approximation for the i -th row of Π_P . We will refer to this procedure as the *adapted JSPM* version for large instances.

In the following, we introduce the adapted JSPM algorithm, where E_j denotes the set of indexes identified as part of the j -th ergodic class, I denotes the number of identified ergodic classes, and C denotes the set of considered/evaluated indexes. The user-defined value to decide whether a state is ergodic or not is denoted by ι , with $\iota > 0$. For large instances P ,

Table 4 Large instances used for numerical experiments

Tr. Matrix	Description	S	Ergodic structure	p^*
P_{10}	Preferential attachment network, [3], with parameter $d = 2$.	10,000	([10,000],0)	$1.1e-3$
P_{11}	Kleinberg’s network, [18], with parameters $p = 1, q = 1$ and $\beta = 1.5$.	15,625	([15,625],0)	$4.6e-2$
P_{12}	Block diagonal transition matrix of P_{10} and P_{11} with weak connection between P_{10} and P_{11} . I.e., two random nodes from P_{10} and P_{11} , resp., are connected with probability $1.5e-9$.	25,625	([25,625],0)	$1.5e-9$
P_{13}	Block diagonal transition matrix of P_{10} and P_{11} with 30 transient states.	25,655	([10,000 15,625],30)	$1.9e-6$

It has the same setup as Table 1, see also the corresponding description in Sect. 4.1. For implementation the MATLAB code provided by CONTEST [26] was used

the adapted JSPM for computing $H_\alpha^c(P)$ can be summarized as follows (recall that \mathbb{S} is the state space of the Markov chain under consideration):

- (1) CHOOSE $\iota > 0$.
- (2) INITIALIZE $I = 0$ AND $C = \emptyset$.
- (3) IF $\mathbb{S} \setminus C \neq \emptyset$:
 - (3.1) SELECT $i \in \mathbb{S} \setminus C$.
 - (3.2) CALCULATE $H_\alpha(P)(i, \bullet)$.
 - OTHERWISE GO TO STEP 6.
- (4) IF $H_\alpha(P)(i, i) > \iota$:
 - (4.1) STATE i IS IDENTIFIED AS ERGODIC, SET $I = I + 1$.
 - (4.2) $E_I = \{j : H_\alpha(P)(i, j) > \iota\}$.
 - (4.3) $C = C \cup E_I$.
 - OTHERWISE i IS IDENTIFIED AS TRANSIENT, SET $C = C \cup \{i\}$.
- (5) SAVE $H_\alpha(P)(i, \bullet)$ IN $H_\alpha^c(P)$ AND GO TO STEP 3.
- (6) RETURN $H_\alpha^c(P)$.

For instances P_{10}, P_{11}, P_{12} and P_{13} from Table 4 the adapted JSPM is applied where we have chosen $\alpha = \min\{10^{-10}, (p^*)^2\}$ and $\iota = (1/S)^2$. The philosophy behind the choice of α is similar to Parameter Setting 2 in the previous section, i.e., small α is chosen such that one iteration is most likely sufficient. Our experience for real life networks is that $\iota = (1/S)^2$ ensures correct distinctions between transient and ergodic states. In Table 5 the norm errors and computation times in seconds (sec.) of the experiments can be found.

From the results it follows that high accuracy is achieved in a relatively small amount of time, i.e., a unique row of $H_\alpha(P)$ in case of 25625 states is calculated with MATLAB R2011b in 1.37 seconds on a 64-bit Windows desktop PC with Intel(R) Core(TM) i5-2310 CPU @ 2.90GHz processor with norm error $3.524 \cdot 10^{-8}$. To put the results into context,

Table 5 Results for adapted JSPM, with $\alpha = \min\{10^{-10}, (P^*)^2\}$ and $\iota = (1/S)^2$, in case of large instances

	$\ \Pi_{P_i} - H_\alpha^c(P_i)\ $	Computation time (in s) $H_\alpha^c(P_i)$
$i = 10$	2.0553e-010	0.24
$i = 11$	1.6724e-010	1.14
$i = 12$	3.5243e-008	1.37
$i = 13$	6.1897e-006	45.80

for instance P_{12} it takes on average 2.49 seconds to calculate $\mu^\top(P_{12})^4$, i.e., two sparse matrix multiplications, to obtain $(P_{12})^4$, with norm error $\|\mu^\top \Pi_{P_{12}} - \mu^\top(P_{12})^4\| = 0.1794$, where μ^\top equals the first row of an appropriate sized identity matrix. It becomes even more counterproductive if we calculate $\mu^\top(P_{12})^8$, which takes on average 593.75 seconds and leads to norm error $\|\mu^\top \Pi_{P_{12}} - \mu^\top(P_{12})^8\| = 0.079$. The significant increase in computation time is due to loss of sparsity. In this case, the vector-updating version of PM might be more efficient. Indeed, performing 7 sparse vector-matrix multiplications to obtain $\mu^\top(P_{12})^8 = P_{12}(1, \bullet)(P_{12})^7$, requires only 0.0109 seconds. However, evaluating $\mu^\top(P_{12})^{10000}$ in this way, leading to norm error $\|\mu^\top \Pi_{P_{12}} - \mu^\top(P_{12})^{10000}\| = 0.0048$, already requires 15.63 seconds. Clearly, this demonstrates the potential of the adapted JSPM in evaluating the ergodic projector. Similar observations can be expected for the other large instances.

A way to (most likely) improve accuracy of the adapted JSPM without significantly increasing computation time is to calculate $H_\alpha^c(P^q)$, for $q > 1$. The intuition is that for relatively small q , P^q may not affect the sparsity too much (increase in computation time is limited) but may increase the accuracy (which is likely according to the theory). Note that although it is common, it is not necessary that larger q increases accuracy, theory only provides upperbounds for the norm error. Example 6 provides an instance for which a larger q does not increase accuracy of $H_\alpha(P)$.

Example 6 Take

$$P = \begin{bmatrix} 0.1 & 0.45 & 0.45 \\ 0.9 & 0.1 & 0 \\ 0.9 & 0 & 0.1 \end{bmatrix} \text{ so that } \Pi_P = \begin{bmatrix} 1/2 & 1/4 & 1/4 \\ 1/2 & 1/4 & 1/4 \\ 1/2 & 1/4 & 1/4 \end{bmatrix} \text{ and}$$

$$P^2 = \begin{bmatrix} 0.82 & 0.09 & 0.09 \\ 0.18 & 0.415 & 0.405 \\ 0.18 & 0.405 & 0.415 \end{bmatrix}.$$

It holds that $\|\Pi_P - H_{\alpha=10^{-6}}(P)\| \approx 4.44 \cdot 10^{-7}$, whereas $\|\Pi_P - H_{\alpha=10^{-6}}(P^2)\| \approx 1.78 \cdot 10^{-6}$, which is due to the periodic behavior of P (visible by comparing P and P^2).

For the instances put forward in Table 4 we tested the effect of considering $(P_i)^2$ instead of P_i , where $i = 10, 11, 12, 13, \dots$. Most of the time taking a power increased the accuracy but more significantly increased computation time so that practical usability of taking powers is questionable.

5 Conclusion

This paper introduces JSPM which is a generalization of PM. JSPM is a highly accurate approximation method for the ergodic projector of a general finite Markov chain, including periodic Markov multi-chains. Convergence analysis and numerical experiments show

that it can provide a viable generalization of PM. Especially in case of large-scale Markov chains, JSPM works well and can deal with nearly decomposable chains without running into numerical instabilities.

Further research includes extending the techniques used for analyzing JSPM to the deviation matrix and achieving higher accuracy via numerical ingenuity.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Appendix: Proof of Results for Periodic Markov Chains

The main results of this paper can be extended to periodic Markov chains, also known as cyclic Markov chains. As standard literature is in general not concerned with cyclic Markov *multi*-chains, required technical toolbox with respect to cyclic Markov *multi*-chains will be developed first.

The Cesaro limit (of order one) version of the ergodic projector is given by

$$\Pi_P = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} P^n.$$

Property $\Pi_P P = \Pi_P$ also holds for the Cesaro limit version. When P is aperiodic, $\lim_{N \rightarrow \infty} P^n$ exists and equals Π_P , see, e.g., the appendix in [23].

For $i \in \mathbb{S}$, define

$$A(i) = \{n \in \mathbb{N} : P^n(i, i) > 0\}.$$

Using this notion, let the period of state $i \in \mathbb{S}$ be described as

$$d(i) = \begin{cases} \text{GCD}(A(i)), & A(i) \neq \emptyset \\ 1, & A(i) = \emptyset, \end{cases}$$

where $\text{GCD}(\cdot)$ denotes the greatest common divisor (GCD) of a set. The period d of a Markov *multi*-chain with P is defined as

$$d = \text{LCM}(\{d(i), i \in \mathbb{S}\}),$$

where $\text{LCM}(\cdot)$ denotes the least common multiple (LCM) of a set.

The communicating class of $i \in \mathbb{S}$ is denoted by $C(i)$ and can be formalized as

$$C(i) = \{j \in \mathbb{S} \mid \exists n_1, n_2 \geq 0, P^{n_1}(i, j) > 0, P^{n_2}(j, i) > 0\}.$$

Note that $A(i)$ is closed under addition, i.e.,

$$n_1 \in A(i), n_2 \in A(i) \implies n_1 + n_2 \in A(i),$$

and it follows from number theory that $A(i)$ contains all but a finite number of positive multiples of $d(i)$:

$$\exists K = K(i), \forall k \geq K : P^{kd(i)}(i, i) > 0. \tag{17}$$

This allows us to prove the following connection between periodic and aperiodic Markov chains.

Lemma 2 *Let P be a Markov multi-chain with period d . Then P^d is aperiodic.*

Proof Showing that

$$\forall i \in \mathbb{S}, \forall j \in C(i), \exists N, \forall n \geq N : (P^d)^n(i, j) > 0,$$

suffices to prove that P^d is aperiodic. To see this, choose $i \in \mathbb{S}$ and $j \in C(i)$. Then it holds that

$$\exists m : P^m(i, j) > 0$$

and furthermore due to (17)

$$\exists K = K(j), \forall k \geq K : P^{kd(j)}(j, j) > 0.$$

This means that

$$(P^d)^n(i, j) = P^{dn}(i, j) \geq P^m(i, j)P^{dn-m}(j, j) > 0$$

for

$$dn - m \geq Kd(j) \iff n \geq \frac{Kd(j) + m}{d} \xrightarrow{d \geq d(j), \forall j} n \geq N,$$

where $N = K + \frac{m}{d}$. □

Relying on Lemma 2, the following lemma establishes a geometric ergodicity result for periodic Markov chains.

Lemma 3 *For any finite-state Markov chain with P with period d , there exist a finite number r such that*

$$n \geq r : \left\| \sum_{l=0}^{d-1} (P^{n+l} - \Pi_P) \right\| \leq c\beta^n,$$

for $c < \infty$ and $\beta \in (0, 1)$. In particular, $r = r_{ad}$, $c = dc_d(\beta_d)^{\frac{1-d}{d}}$ and $\beta = (\beta_d)^{\frac{1}{d}}$, where r_d , c_d and β_d denote the geometric ergodicity parameters of (aperiodic) P^d .

Proof Assume without losing of generality that $n = xd + y$ for $0 \leq y \leq d - 1$, where d denotes the periodicity of P . Then,

$$\sum_{l=0}^{d-1} (P^{n+l} - \Pi_P) = \sum_{l=0}^{d-1} (P^{xd+y+l} - \Pi_P) = P^y \sum_{l=0}^{d-1} (P^{xd+l} - \Pi_P).$$

From Eq. (3) in the proof of Theorem 5.1.4 in [17], it follows that

$$\Pi_P = \frac{1}{d} \sum_{i=0}^{d-1} \Pi_{P^d} P^i.$$

Using this expression for Π_P gives

$$\begin{aligned} \sum_{l=0}^{d-1} (P^{n+l} - \Pi_P) &= P^y \sum_{l=0}^{d-1} \left(P^{xd+l} - \frac{1}{d} \sum_{i=0}^{d-1} \Pi_{P^d} P^i \right) \\ &= P^y \left(\sum_{l=0}^{d-1} P^{xd+l} - \sum_{i=0}^{d-1} \Pi_{P^d} P^i \right) \\ &= P^y \left(\sum_{l=0}^{d-1} (P^{xd} - \Pi_{P^d}) P^l \right) \\ &= P^y (P^{xd} - \Pi_{P^d}) \left(\sum_{l=0}^{d-1} P^l \right). \end{aligned}$$

Since P^d is aperiodic (and finite), see Lemma 2, geometric ergodicity of aperiodic Markov chains implies that there exists a finite r_d such that

$$x \geq r_d : \quad \|(P^d)^x - \Pi_{P^d}\| \leq c_d(\beta_d)^x,$$

where finite r_d, c_d and $\beta_d \in (0, 1)$ are the ergodicity parameters of P^d . Therefore, it holds for $x \geq r_d \implies n \geq r_d d$ that

$$\begin{aligned} \left\| \sum_{l=0}^{d-1} (P^{n+l} - \Pi_P) \right\| &\leq \|P^y\| c_d(\beta_d)^x \sum_{l=0}^{d-1} \|P^l\| \\ &= d c_d(\beta_d)^x, \end{aligned}$$

where the equality follows because $\|\cdot\|$ represents the ∞ -norm by assumption. Since $y \leq d - 1$,

$$n = xd + y \implies x \geq \frac{n}{d} + \frac{1-d}{d},$$

and thus

$$(\beta_d)^x \leq (\beta_d)^{\frac{n}{d} + \frac{1-d}{d}} = (\beta_d)^{\frac{1-d}{d}} ((\beta_d)^{\frac{1}{d}})^n,$$

note that $(\beta_d)^{\frac{1}{d}} \in (0, 1)$. This leads to

$$\left\| \sum_{l=0}^{d-1} (P^{n+l} - \Pi_P) \right\| \leq d c_d(\beta_d)^{\frac{1-d}{d}} ((\beta_d)^{\frac{1}{d}})^n, \quad \text{for } n \geq r_d d.$$

Labeling $r = r_d d, c = d c_d(\beta_d)^{\frac{1-d}{d}}$ and $\beta = (\beta_d)^{\frac{1}{d}}$ proves the result. □

Remark 6 Note that for $d = 1$, Lemma 3 reduces to the geometric ergodicity result for aperiodic Markov chains ($r = r_d, c = c_d$ and $\beta = \beta_d$ for $d = 1$).

In the following theorem, the above results will be applied to prove a sharp bound for the norm error also applicable to periodic Markov chains. As parameter q affects the periodicity of a Markov chain, only the case $q = 1$ will be considered for simplicity. The theorem uses

the following generalized geometric ergodicity parameters, indicated with a bar,

$$\bar{\gamma}(P, d) = \bar{\kappa}(d, P)d(\bar{\phi}(d, P) - 1) + \frac{c\beta^{d\bar{\phi}(d, P)+1}}{1 - \beta^d} + (d + 2)(d - 1),$$

$$\bar{\kappa}(d, P) = \sup_{n_1=0,1,\dots,\min\{\bar{\phi}(d, P)-1, x\}} \left\| \sum_{n_2=0}^{d-1} (\Pi_P - P^{n_1d+1+n_2}) \right\|,$$

and

$$\bar{\phi}(d, P) = \left\lceil \frac{r - 1}{d} \right\rceil.$$

The generalized geometric ergodicity parameters are ‘generalized’ in the sense that they reduce to $\gamma(P, q = 1)$, $\kappa(P, q = 1)$, and $\phi(P, q = 1)$, resp., in case of periodicity $d = 1$. As a result, the same holds for the following theorem, i.e., Theorem 3 reduces to the results from Lemma 1 and Theorem 2 when periodicity $d = 1$ (and $q = 1$).

Theorem 3 For $N \geq d\bar{\phi}(d, P)$,

$$\|\Pi_P - (H_\alpha(N, P))^k\| \leq \left(\frac{\alpha\bar{\gamma}(P, d)}{1 - (1 - \alpha)^{N+1}} \right)^k$$

so that for $N \rightarrow \infty$ it holds

$$\|\Pi_P - (H_\alpha(P))^k\| \leq (\alpha\bar{\gamma}(P, d))^k.$$

Proof Inequality (9) (which holds for a periodic Markov chain with P) for $q = 1$ gives

$$\|\Pi_P - (H_\alpha(N, P))^k\| \leq \left\| \alpha \frac{\sum_{n=0}^N (1 - \alpha)^n (\Pi_P - P^{n+1})}{1 - (1 - \alpha)^{N+1}} \right\|^k. \tag{18}$$

Without loss on generality, assuming that $N = N_d = xd + y$, for $0 \leq y \leq d - 1$ and d the periodicity of Markov chain P , allows to rewrite the summation from (18) as

$$\sum_{n=0}^N (1 - \alpha)^n (\Pi_P - P^{n+1}) = \overbrace{\sum_{n_1=0}^x (1 - \alpha)^{n_1d} \sum_{n_2=0}^{d-1} (1 - \alpha)^{n_2} (\Pi_P - P^{n_1d+1+n_2})}^{((19).i)} - \underbrace{\sum_{n_3=y+1}^{d-1} (1 - \alpha)^{xd+n_3} (\Pi_P - P^{xd+1+n_3})}_{((19).ii)}, \tag{19}$$

where ((19).ii) is a compensation term that allows the first term ((19).i) to sum up to $n_1 = x$. Equation (19) implies that

$$\left\| \sum_{n=0}^N (1 - \alpha)^n (\Pi_P - P^{n+1}) \right\| \leq \|((19).i)\| + \|((19).ii)\|, \tag{20}$$

where the two terms will be bounded sequentially. The bounding will utilize that $\|P_1 - P_2\| \leq 2$ for all stochastic matrices P_1 and P_2 (since $\|\cdot\|$ describes the ∞ -norm, i.e., ‘the maximum over all absolute row sums’).

For the first term $\|((19).i)\|$ it holds that

$$\|((19).i)\| \leq \sum_{n_1=0}^x (1 - \alpha)^{n_1 d} \left\| \sum_{n_2=0}^{d-1} (1 - \alpha)^{n_2} (\Pi_P - P^{n_1 d + 1 + n_2}) \right\|,$$

in order to apply the geometric ergodicity result for periodic Markov chains from Lemma 3, the norm of the inner summation can be bounded as

$$\begin{aligned} & \left\| \sum_{n_2=0}^{d-1} (1 - \alpha)^{n_2} (\Pi_P - P^{n_1 d + 1 + n_2}) \right\| \\ & \leq \left\| \sum_{n_2=0}^{d-1} (\Pi_P - P^{n_1 d + 1 + n_2}) \right\| + \left\| \sum_{i=0}^{d-1} ((1 - \alpha)^i - 1) (\Pi_P - P^{n_1 d + 1 + i}) \right\| \\ & \leq \left\| \sum_{n_2=0}^{d-1} (\Pi_P - P^{n_1 d + 1 + n_2}) \right\| + 2 \sum_{i=0}^{d-1} (1 - (1 - \alpha)^i) \\ & \leq \left\| \sum_{n_2=0}^{d-1} (\Pi_P - P^{n_1 d + 1 + n_2}) \right\| + \alpha(d - 1)d, \end{aligned}$$

where the last inequality follows from $1 - (1 - \alpha)^i \leq \alpha i$ for $\alpha \in (0, 1)$ and $i = 0, 1, \dots$, which leads to

$$\|((19).i)\| \leq \sum_{n_1=0}^x (1 - \alpha)^{n_1 d} \left\| \sum_{n_2=0}^{d-1} (\Pi_P - P^{n_1 d + 1 + n_2}) \right\| + \sum_{n_1=0}^x (1 - \alpha)^{n_1 d} \alpha(d - 1)d,$$

since, for $\alpha \in (0, 1)$ and $d \geq 1$,

$$\sum_{n_1=0}^x (1 - \alpha)^{n_1 d} \leq \sum_{n_1=0}^x (1 - \alpha)^{n_1} = \frac{1 - (1 - \alpha)^{x+1}}{\alpha} \leq \frac{1}{\alpha}$$

this means that

$$\|((19).i)\| \leq \sum_{n_1=0}^x (1 - \alpha)^{n_1 d} \left\| \sum_{n_2=0}^{d-1} (\Pi_P - P^{n_1 d + 1 + n_2}) \right\| + (d - 1)d.$$

Since $y \geq 0$ in $N = N_d = xd + y$ and $\alpha \in (0, 1)$, the second term (19).ii can be bounded as

$$\|((19).ii)\| \leq 2 \sum_{n_3=y+1}^{d-1} (1 - \alpha)^{x d + n_3} \leq 2 \sum_{n_3=1}^{d-1} 1 = 2(d - 1).$$

Inserting the boundings of $\|((19).i)\|$ and $\|((19).ii)\|$ into (20) leads to

$$\begin{aligned} \left\| \sum_{n=0}^N (1 - \alpha)^n (\Pi_P - P^{n+1}) \right\| & \leq \underbrace{\sum_{n_1=0}^x (1 - \alpha)^{n_1 d} \left\| \sum_{n_2=0}^{d-1} (\Pi_P - P^{n_1 d + 1 + n_2}) \right\|}_{(21).*} \\ & + (d + 2)(d - 1), \end{aligned} \tag{21}$$

in which the norm terms of ((21).★) can be replaced for the bound from Lemma 3 for n_1 large enough. In particular, for $n_1d + 1 \geq r$, Lemma 3 can be used, i.e., for

$$n_1 \geq \frac{r-1}{d} \implies n_1 \geq \left\lceil \frac{r-1}{d} \right\rceil =: \bar{\phi}(d, P),$$

where $\bar{\phi}(d, P)$ is defined as the nearest integer larger than $\frac{r-1}{d}$, denoted by $\lceil \frac{r-1}{d} \rceil$. Note that $\bar{\phi}(1, P) = \phi(P, 1)$. Similar as in the proof of Lemma 1, the summation of ((21).★) will be split into two parts: One part for which the geometric ergodicity result from Lemma 1 does not apply and one part where it does, respectively. Specifically,

$$\begin{aligned} ((21).★) = & \underbrace{\sum_{n_1=0}^{\min\{\bar{\phi}(d,P)-1, x\}} (1-\alpha)^{n_1d} \left\| \sum_{n_2=0}^{d-1} (\Pi_P - P^{n_1d+1+n_2}) \right\|}_{((21).★.i)} \\ & + \underbrace{\sum_{n_3=\bar{\phi}(d,P)}^x (1-\alpha)^{n_3d} \left\| \sum_{n_2=0}^{d-1} (\Pi_P - P^{n_3d+1+n_2}) \right\|}_{((21).★.ii)}, \end{aligned}$$

where parts ((21).★.i) and ((21).★.ii) will be bounded, respectively.

It holds that

$$((21).★.i) \leq \sup_{n_1=0,1,\dots,\min\{\bar{\phi}(d,P)-1, x\}} \left\| \sum_{n_2=0}^{d-1} (\Pi_P - P^{n_1d+1+n_2}) \right\| \frac{1 - (1-\alpha)^{d \min\{\bar{\phi}(d,P), x+1\}}}{1 - (1-\alpha)^d},$$

defining

$$\bar{\kappa}(d, P) = \sup_{n_1=0,1,\dots,\min\{\bar{\phi}(d,P)-1, x\}} \left\| \sum_{n_2=0}^{d-1} (\Pi_P - P^{n_1d+1+n_2}) \right\|$$

gives

$$((21).★.i) \leq \bar{\kappa}(d, P) \frac{1 - (1-\alpha)^{d \min\{\bar{\phi}(d,P), x+1\}}}{1 - (1-\alpha)^d}.$$

Regarding (21).★.ii, the result from Lemma 3 applies. This gives

$$\begin{aligned}
 (21).\star.ii) &\leq \sum_{n_3=\bar{\phi}(d,P)}^x (1-\alpha)^{n_3d} c\beta^{n_3d+1} \\
 &= c\beta \sum_{n_3=\bar{\phi}(d,P)}^x ((1-\alpha)^d \beta^d)^{n_3} \\
 &= c\beta((1-\alpha)^d \beta^d)^{\bar{\phi}(d,P)} \sum_{n_3=0}^{x-\bar{\phi}(d,P)} ((1-\alpha)^d \beta^d)^{n_3} \\
 &= c\beta((1-\alpha)^d \beta^d)^{\bar{\phi}(d,P)} \frac{1 - ((1-\alpha)\beta)^{d \max\{x-\bar{\phi}(d,P)+1, 0\}}}{1 - (1-\alpha)^d \beta^d} \\
 &\leq c\beta^{d\bar{\phi}(d,P)+1} \frac{1 - \beta^{d \max\{x-\bar{\phi}(d,P)+1, 0\}}}{1 - \beta^d},
 \end{aligned}$$

where property $\alpha \in (0, 1)$ was used in the last inequality.

Inserting ((21).★.i) and ((21).★.ii) into (21) gives

$$\begin{aligned}
 &\left\| \sum_{n=0}^N (1-\alpha)^n (\Pi_P - P^{n+1}) \right\| \\
 &\leq \bar{\kappa}(d, P)d \min\{\bar{\phi}(d, P), x + 1\} + c\beta^{d\bar{\phi}(d,P)+1} \frac{1 - \beta^{d \max\{x-\bar{\phi}(d,P)+1, 0\}}}{1 - \beta^d} \\
 &\quad + (d + 2)(d - 1).
 \end{aligned}$$

This result can be used in (18) to bound, for $x + 1 \leq \bar{\phi}(d, P)$ (geometric ergodicity does not apply),

$$\|\Pi_P - (H_\alpha(N, P))^k\| \leq \left(\alpha \frac{\bar{\kappa}(d, P) \frac{1-(1-\alpha)^{d(x+1)}}{1-(1-\alpha)^d} + (d+2)(d-1)}{1 - (1-\alpha)^{N+1}} \right)^k$$

which reduces to (12) in case $d = 1$ (and thus $N = x$). In case $x \geq \bar{\phi}(d, P)$ (geometric ergodicity does apply),

$$\begin{aligned}
 &\|\Pi_P - (H_\alpha(N, P))^k\| \\
 &\leq \left(\alpha \frac{\bar{\kappa}(d, P)d(\bar{\phi}(d, P) - 1) + c\beta^{d\bar{\phi}(d,P)+1} \frac{1-\beta^{d(x-\bar{\phi}(d,P)+1)}}{1-\beta^d} + (d+2)(d-1)}{1 - (1-\alpha)^{N+1}} \right)^k \\
 &\leq \left(\alpha \frac{\bar{\kappa}(d, P)d(\bar{\phi}(d, P) - 1) + \frac{c\beta^{d\bar{\phi}(d,P)+1}}{1-\beta^d} + (d+2)(d-1)}{1 - (1-\alpha)^{N+1}} \right)^k \\
 &\equiv \left(\frac{\alpha \bar{\gamma}(P, d)}{1 - (1-\alpha)^{N+1}} \right)^k
 \end{aligned}$$

where we defined

$$\bar{\gamma}(P, d) = \bar{\kappa}(d, P)d(\bar{\phi}(d, P) - 1) + \frac{c\beta^{d\bar{\phi}(d, P)+1}}{1 - \beta^d} + (d + 2)(d - 1).$$

Observing that

$$x + 1 \geq \bar{\phi}(d, P) \Leftrightarrow dx + d \geq d\bar{\phi}(d, P) \Leftrightarrow N \geq d\bar{\phi}(d, P) + y - d \implies N \geq d\bar{\phi}(d, P)$$

ends the proof. \square

References

- Altman, E., Avrachenkov, K.E., Núñez-Queija, R.: Perturbation analysis for denumerable Markov chains with application to queueing models. *Adv. Appl. Probab.* **36**, 839–853 (2004)
- Avrachenkov, K.E., Haviv, M.: The first Laurent series coefficients for singularly perturbed stochastic matrices. *Linear Algebra Its Appl.* **386**, 242–259 (2004)
- Barabási, A., Albert, R.: Emergence of scaling in random networks. *Science* **286**(5439), 509–512 (1999). <https://doi.org/10.1126/science.286.5439.509>
- Berkhin, P.: A survey on PageRank computing. *Internet Math.* **2**(1), 73–120 (2005). <https://doi.org/10.1080/15427951.2005.10129098>
- Berkhout, J., Heidergott, B.F.: A series expansion approach to risk analysis of an inventory system with sourcing. In: Proceedings of 12th International Workshop on Discrete Event Systems (WODES 2014), vol. 12, pp. 510–515 (2014). <http://www.ifac-papersonline.net/Detailed/65199.html>
- Berkhout, J., Heidergott, B.F.: Efficient algorithm for computing the ergodic projector of Markov multi-chains. *Procedia Comput. Sci.* **51**, 1818–1827 (2015). <https://doi.org/10.1016/j.procs.2015.05.403>
- Brin, S., Page, L.: The anatomy of a large-scale hypertextual Web search engine. *Comput. Netw. ISDN Syst.* **30**(1–7), 107–117 (1998)
- Fox, B.L., Landi, D.M.: Scientific applications: an algorithm for identifying the ergodic subchains and transient states of a stochastic matrix. *Commun. ACM* **11**(9), 619–621 (1968). <https://doi.org/10.1145/364063.364082>
- Franceschet, M.: PageRank: standing on the shoulders of giants. *Commun. ACM* **54**(6), 92–101 (2011)
- Grindrod, P.: Range-dependent random graphs and their application to modeling large small-world proteome datasets. *Phys. Rev. E* **66**(6), 066,702 (2002). <https://doi.org/10.1103/PhysRevE.66.066702>
- Hartfiel, D.J., Meyer, C.D.: On the structure of stochastic matrices with a subdominant eigenvalue near 1. *Linear Algebra Its Appl.* **272**(1–3), 172–193 (1998)
- Hassin, R., Haviv, M.: Mean passage times and nearly uncoupled Markov chains. *SIAM J. Discrete Math.* **5**(3), 386–397 (1992)
- Haviv, M., Rothblim, U.: Bounds on distances between eigenvectors. *Linear Algebra Its Appl.* **15**, 101–118 (1984)
- Heidergott, B.F., Hordijk, A., van Uitert, M.: Series expansions for finite-state Markov chains. *Probab. Eng. Inf. Sci.* **21**(3), 381–400 (2007). <https://doi.org/10.1017/S0269964807000034>
- Ipsen, I.C.F., Meyer, C.D.: Uniform stability of Markov chains. *SIAM J. Matrix Anal. Appl.* **15**, 1061–1074 (1994)
- Kartashov, N.V.: *Strong Stable Markov Chains*. TBIMC Scientific Publishers, Kiev (1996)
- Kemeny, J.G., Snell, J.L.: *Finite Markov Chains: With a New Appendix “Generalization of a Fundamental Matrix”*. Springer, New York (1976)
- Kleinberg, J.M.: Navigation in a small world. *Nature* **406**(6798), 845 (2000). <https://doi.org/10.1038/35022643>
- Kontoyiannis, I., Meyn, S.P.: Spectral theory and limit theorems for geometrically ergodic Markov processes. *Ann. Appl. Probab.* **13**(1), 304–362 (2013)
- Langville, A.N., Meyer, C.D.: *Google’s PageRank and Beyond: The Science of Search Engine Rankings*. Princeton University Press, Princeton (2011)
- Morrison, J.L., Breitling, R., Higham, D.J., Gilbert, D.R.: A lock-and-key model for protein-protein interactions. *Bioinformatics (Oxford, England)* **22**(16), 2012–2019 (2006). <https://doi.org/10.1093/bioinformatics/btl338>
- Newman, M.: *Networks: An Introduction*. Oxford University Press, Oxford (2010)
- Puterman, M.L.: *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. Wiley, New York (1994)

24. Puterman, M.L.: Markov Decision Processes: Discrete Stochastic Dynamic Programming. Wiley, New York (2014)
25. Stewart, W.J.: Introduction to the Numerical Solution of Markov Chains. Princeton University Press, Princeton (1994)
26. Taylor, A., Higham, D.J.: CONTEST. ACM Trans. Math. Softw. **35**(4), 1–17 (2009). <https://doi.org/10.1145/1462173.1462175>