CONVERGENCE ANALYSIS OF AN IMPROVED PAGERANK ALGORITHM

ILSE C.F. IPSEN* AND STEVE KIRKLAND†

Abstract. The iterative aggregation/disaggregation (IAD) method is an improvement of the PageRank algorithm used by the search engine Google to compute stationary probabilities of very large Markov chains.

In this paper the convergence, in exact arithmetic, of the IAD method is analyzed. The IAD method is expressed as the power method preconditioned by an incomplete LU factorization. This leads to a simple derivation of the asymptotic convergence rate of the IAD method.

It is shown that the power method applied to the Google matrix always converges, and that the convergence rate of the IAD method is at least as good as that of the power method. Furthermore, by exploiting the hyperlink structure of the web it can be shown that the convergence rate of the IAD method applied to the Google matrix can be made strictly faster than that of the power method.

Key words. Google, PageRank, Markov chain, power method, stochastic complement, aggregation/disaggregation

AMS subject classification. 15A51, 65F15, 65F10, 15A18, 15A42

1. Introduction.

We analyse the convergence of the IAD algorithm [7, 8] which is an improvement of the PageRank algorithm [2] used by the search engine Google.

Google maintains a stochastic matrix[1] G, whose element in position (i, j) represents the probability of following a link from web page i to web page j. Google determines the PageRank of web page i, which is the probability that a random surfer visits web page i. Google uses the PageRank to determine the citation importance of web pages and to order the results of Web keyword searches. It turns out that the PageRank of web page i equals element i of the stationary distribution π of G, i.e. $\pi^T G = \pi^T$, where the superscript T denotes the transpose.

Thus Google faces the problem of having to compute the stationary probability of a large stochastic matrix (the dimension of the matrix G is several billion [12],[9, §3.1]). Direct methods [15, §2] are too time consuming for large matrices. It turns out that G is primitive, so that the power method applied to G converges, in exact arithmetic, to a multiple of π, for any starting vector [15, §3.1]. The asymptotic convergence rate of the power method is given by the modulus of the second largest eigenvalue, in magnitude, of G. The original PageRank algorithm [2] amounts to the power method [2, §2.1.1], [9, §3.1]. But the convergence can be slow [7, §6, §8], and the power method is often accelerated [5], [15, §3.4]. Meyer and Langville improve the PageRank algorithm [2] by accelerating the power method with an aggregation/disaggregation step similar to [15, §6.3]. Their method is called IAD for iterative aggregation/disaggregation [7, 8]. Comparison of operation counts and numerical experiments illustrate that the IAD method is faster than the power method [7, §6, §8].

*Center for Research in Scientific Computation, Department of Mathematics, North Carolina State University, P.O. Box 8205, Raleigh, NC 27695-8205, USA (ipsen@math.ncsu.edu, http://www4.ncsu.edu/~ipsen/). This research was supported in part by NSF grants DMS-0209931 and DMS-0209695.

†Department of Mathematics and Statistics, University of Regina, Regina, Saskatchewan S4S 0A2, Canada (kirkland@math.uregina.ca, http://www.math.uregina.ca/~kirkland/). This research was supported in part by NSERC grant OGP0138251

[1] A stochastic matrix P is a real square matrix with elements $0 \leq p_{ij} \leq 1$, whose rows sum to one, $\sum_j p_{ij} = 1$ for all i.

Overview. Here we analyse, in exact arithmetic, the convergence of the IAD method. In particular we show that the IAD method amounts to a power method preconditioned by an incomplete LU (ILU) factorization (§3, §4, §5) and give a simple derivation of the asymptotic convergence rate (§5). We consider the choice of different stochastic complements for better convergence (§6) and computable upper bounds on the convergence rate (§7). In particular we show that the power method applied to the Google matrix G always converges, and that the convergence rate of the IAD method applied to G is at least as good as that of the power method. We consider conditions that assure convergence of the IAD method in general (§8). At last we prove a stronger result for convergence of the IAD method applied to G (§9). By exploiting the hyperlink structure of the web we conclude that the convergence rate of the IAD method applied to G can be made strictly better than that of the power method. We also show how one can save computations when applying the IAD method to G.

Notation. All vectors and matrices are real. The identity matrix is I with columns e_j, $\mathbf{1} = (1 \quad \dots \quad 1)^T$ is the column vector of all ones. For a matrix P, the inequality $P \geq 0$ means that each element of P is non-negative. Similarly, $P > 0$ means that each element of P is positive.

A stochastic matrix P is a square matrix with $P \geq 0$ and $P\mathbf{1} = \mathbf{1}$. A probability vector v is a column vector with $v \geq 0$ and $v^T\mathbf{1} = 1$. The one-norm of a column vector v is $\|v^T\| \equiv |v|^T\mathbf{1}$. The eigenvalues $\lambda_i(P)$ of P are labelled in order of decreasing magnitude, $|\lambda_1(P)| \geq |\lambda_2(P)| \geq \dots$. The directed graph associated with P is denoted by $\Delta(P)$.

2. The Problem. Let P be a stochastic matrix. We want to compute the stationary distribution π of P, that is a column vector π such that [1, Theorem 2.(1.1)] $\pi^T P = \pi^T$ or $\pi^T(I - P) = 0$ where $\pi \geq 0$ and $\pi^T\mathbf{1} = 1$. If P is irreducible then the eigenvalue 1 is distinct and it is the spectral radius, and the stationary distribution $\pi > 0$ is unique [1, Theorem 2.(1.3)], [16, Theorem 2.1].

The matrix P may be large, so the idea is to work with only a small part of the matrix. That is, partition

$$P = \begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix},$$

where P_{11} and P_{22} are square. Ideally the dimension of P_{11} is small compared to the dimension of P. The methods considered here approximate the stationary distribution π of P by operating on matrices derived from P_{11} whose dimension is, essentially, the same as that of P_{11}.

3. Exact Aggregation/Disaggregation. We present an aggregation/disaggregation method that is the basis for the IAD method in [7, §8.2.2] and [8, §3.2.1], an improvement of the PageRank algorithm [2]. Our presentation differs from traditional ones [7, 10, 15] because we start from a block LDU decomposition. The aggregation algorithm computes the components of π as stationary distributions of smaller matrices.

Partition the irreducible stochastic matrix

$$P = \begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix}, \qquad \pi = \begin{pmatrix} \pi_1 \\ \pi_2 \end{pmatrix}$$

so that P_{11} and P_{22} are square. Since $I - P$ is an M-matrix [1, Theorem 8.(4.2)], as well as singular and irreducible, the non-trivial leading principal submatrix $I - P_{11}$ is non-singular [1, Theorem 6.(4.16)]. Hence we can factor $I - P = LDU$, where [10, Proof of Theorem 2.3]

$$(3.1) \qquad L \equiv \begin{pmatrix} I & 0 \\ -P_{21}(I - P_{11})^{-1} & I \end{pmatrix}, \quad D \equiv \begin{pmatrix} I - P_{11} & 0 \\ 0 & I - S \end{pmatrix}, \quad U \equiv \begin{pmatrix} I & -(I - P_{11})^{-1}P_{12} \\ 0 & I \end{pmatrix},$$

and

$$S \equiv P_{22} + P_{21}(I - P_{11})^{-1}P_{12},$$

so that $I-S$ is the Schur complement of $I-P_{11}$ in $I-P$ [3, §1]. The matrix S is also known as the stochastic complement of P_{22} in P [10, Definition 2.1]. It is a special case of a Perron complement in the context of non-negative matrices [11, Definition 2.1].

Since U is non-singular we have $\pi^T(I-P)=0$ if and only $\pi^T LD=0$. Hence

$$\pi_2^T S = \pi_2^T \qquad \pi_1^T = \pi_2^T P_{21}(I-P_{11})^{-1}, \tag{3.2}$$

which means that π_2 is a stationary distribution for the smaller matrix S. The expressions (3.2) represent a partial version of the coupling theorem for 2×2 block matrices [10, Corollary 4.1].

Since P is irreducible and stochastic, so is S [10, Theorem 2.3]. Hence S has a unique, positive stationary distribution σ,

$$\sigma^T S = \sigma^T, \qquad \sigma^T\mathbf{1} = 1, \qquad \sigma > 0.$$

Therefore we can determine π_2 from the stationary distribution σ of S and then set $\pi_2 = \rho\sigma$ where the, as yet unknown, factor ρ is responsible for the normalization $\pi^T\mathbf{1}=1$.

The component π_1 and the factor ρ can also be expressed as components of the stationary distribution of a smaller matrix, the 'aggregated' matrix

$$A \equiv \begin{pmatrix} P_{11} & P_{12}\mathbf{1} \\ \sigma^T P_{21} & \sigma^T P_{22}\mathbf{1} \end{pmatrix}.$$

This is because (3.2), $\pi_2 = \rho\sigma$ and $\sigma^T\mathbf{1}=1$ imply

$$\begin{pmatrix} \pi_1^T & \rho \end{pmatrix}(I-A) = 0, \qquad \begin{pmatrix} \pi_1^T & \rho \end{pmatrix}\mathbf{1} = 1.$$

Thus $\begin{pmatrix} \pi_1^T & \rho \end{pmatrix}$ is a stationary distribution of A. Since A is stochastic and irreducible [10, Theorem 4.1], it has a unique, positive stationary distribution α,

$$\alpha^T A = \alpha^T, \qquad \alpha^T\mathbf{1} = 1, \qquad \alpha > 0.$$

Uniqueness implies $\alpha^T = \begin{pmatrix} \pi_1^T & \rho \end{pmatrix}$.

We have shown that a block LDU decomposition of $\pi^T(I-P)=0$ leads to the expression of π_1 and π_2 as stationary distributions of two smaller matrices.

Algorithm 1 (Exact Aggregation/Disaggregation). *Determine the stationary distribution π of a stochastic irreducible matrix P.*

1. *Determine the stationary distribution σ of $S \equiv P_{22} + P_{21}(I-P_{11})^{-1}P_{12}$.*
2. *Determine the stationary distribution α of $A \equiv \begin{pmatrix} P_{11} & P_{12}\mathbf{1} \\ \sigma^T P_{21} & \sigma^T P_{22}\mathbf{1} \end{pmatrix}$.*
3. *Partition $\alpha = \begin{pmatrix} \pi_1 \\ \rho \end{pmatrix}$, and set $\pi \equiv \begin{pmatrix} \pi_1 \\ \rho\sigma \end{pmatrix}$.*

The first two steps of Algorithm 1 can be interpreted as an aggregation because they take place on 'aggregated' matrices of smaller size, while the third step represents a disaggregation that produces a long vector π of original size.

Example 3.1. *Consider the stochastic irreducible matrix $P = aI_n + (1-a)\mathbf{1}\pi^T$ of order n, where $0<a<1$ is a scalar and π is a vector with $\pi^T\mathbf{1}=1$ and $\pi>0$.*

Then π is the stationary distribution of P. Partitioning gives

$$P = \begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix} = \begin{pmatrix} aI_m + (1-a)\mathbf{1}\pi_1^T & (1-a)\mathbf{1}\pi_2^T \\ (1-a)\mathbf{1}\pi_1^T & aI_{n-m} + (1-a)\mathbf{1}\pi_2^T \end{pmatrix}.$$

The stochastic complement is

$$S = P_{22} + P_{21}(I - P_{11})^{-1}P_{12} = aI_{n-m} + (1-a)\mathbf{1}\sigma^T,$$

where $\sigma^T = \pi_2^T/\pi_2^T\mathbf{1}$ is the stationary distribution of S. The aggregation matrix is

$$A = \begin{pmatrix} P_{11} & P_{12}\mathbf{1} \\ \sigma^T P_{21} & \sigma^T P_{22}\mathbf{1} \end{pmatrix} = \begin{pmatrix} aI_m + (1-a)\mathbf{1}\pi_1^T & (1-a)(\pi_2^T\mathbf{1})\,\mathbf{1} \\ (1-a)\pi_1^T & a + (1-a)(\pi_2^T\mathbf{1})\,\mathbf{1} \end{pmatrix} = aI_{m+1} + (1-a)\mathbf{1}\alpha^T,$$

where $\alpha^T = \begin{pmatrix} \pi_1^T & \pi_2^T\mathbf{1} \end{pmatrix}$ is the stationary distribution of A.

In this particular example, the aggregation matrix A happens to have the same structure as the original matrix P.

Since exact aggregation/disaggregation is too time consuming for large matrices, it makes sense to consider approximate aggregation/disaggregation methods. To understand what happens in the approximate aggregation/disaggregation method in §4 we express π in terms of the block LDU decomposition (3.1).

PROPOSITION 3.1. *Let P be stochastic and irreducible. Then*

$$\pi^T = \rho \begin{pmatrix} * & \sigma^T \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & I \end{pmatrix} L^{-1},$$

where the scalar ρ ensures $\pi^T\mathbf{1} = 1$, and $*$ represents an arbitrary vector of appropriate length.

Proof. This follows from (3.2) and Algorithm 1. ☐

4. Approximate Aggregation/Disaggregation. The approximate aggregation/disaggregation algorithm does away with the time consuming computations involving the stochastic complement S. Instead of the exact stationary distribution σ of the stochastic complement S, we pick any positive probability vector $\tilde{\sigma}$, i.e. $\tilde{\sigma} > 0$ and $\tilde{\sigma}^T\mathbf{1} = 1$. The approximate aggregation matrix is

$$\tilde{A} \equiv \begin{pmatrix} P_{11} & P_{12}\mathbf{1} \\ \tilde{\sigma}^T P_{21} & \tilde{\sigma}^T P_{22}\mathbf{1} \end{pmatrix},$$

and it differs from the exact matrix A only in the last row.

The matrix $\tilde{A}$ is stochastic because P is. The matrix $\tilde{A}$ is also irreducible for the following reason. If P_{11} is of order k, then the directed graph $\Delta(\tilde{A})$ of $\tilde{A}$ contains $k+1$ vertices. The fact that $\tilde{\sigma}$ and $\mathbf{1}$ are positive vectors implies the following: $\Delta(\tilde{A})$ contains an edge from vertex $k+1$ to i, $1 \le i \le k$, if and only if $\Delta(P)$ contains an edge from vertex j, $j \ge k+1$, to i. Also, $\Delta(\tilde{A})$ contains an edge from vertex i, $1 \le i \le k$, to $k+1$ if and only if $\Delta(P)$ contains an edge from vertex i to j, $j \ge k+1$. Finally, $\Delta(\tilde{A})$ contains an edge from i to j, $1 \le i, j \le k$, if and only if $\Delta(P)$ contains an edge from vertex i to j. Since $\Delta(P)$ is strongly connected, $\Delta(\tilde{A})$ must be strongly connected as well. Hence $\tilde{A}$ is irreducible. Thus it has a unique, positive stationary distribution $\tilde{\alpha}$,

$$\tilde{\alpha}^T\tilde{A} = \tilde{\alpha}^T, \qquad \tilde{\alpha}^T\mathbf{1} = 1, \qquad \tilde{\alpha} > 0.$$

All by itself this approximate aggregation makes no progress. However following it up by a power method iteration is known to have 'a very salutory effect' [15, §6.3.1]. Below is a single iteration of the iterative aggregation/disaggregation algorithm. It makes up for the approximation $\tilde{\sigma}$ by appending one iteration of the power method with P to produce an approximation $\tilde{\pi}$ to π.

ALGORITHM 2 (Approximate Aggregation/Disaggregation). *Determine an approximation $\tilde{\pi}$ to the stationary distribution π of a stochastic irreducible matrix P, in one iteration.*

1. *Select a vector $\tilde{\sigma}$ with $\tilde{\sigma} > 0$ and $\tilde{\sigma}^T\mathbf{1} = 1$.*
2. *Determine the stationary distribution $\tilde{\alpha}$ of $\tilde{A} \equiv \begin{pmatrix} P_{11} & P_{12}\mathbf{1} \\ \tilde{\sigma}^T P_{21} & \tilde{\sigma}^T P_{22}\mathbf{1} \end{pmatrix}$.*
3. *Partition $\tilde{\alpha} = \begin{pmatrix} \omega_1 \\ \tilde{\rho} \end{pmatrix}$, and set $\omega \equiv \begin{pmatrix} \omega_1 \\ \tilde{\rho}\,\tilde{\sigma} \end{pmatrix}$.*
4. *Multiply $\tilde{\pi}^T \equiv \omega^T P$.*

If it so happens that $\tilde{\sigma} = \sigma$ then, according to Proposition 3.1, $\omega^T = \tilde{\pi}^T = \omega^T P$, and ω is the desired stationary distribution.

The approximate aggregation/disaggregation Algorithm 2 amounts, in exact arithmetic, to one iteration of an ILU preconditioned power method.

REMARK 4.1. *Let P be stochastic and irreducible and $\tilde{\sigma}$ a vector with $\tilde{\sigma} > 0$ and $\tilde{\sigma}^T\mathbf{1} = 1$. Then*

$$\tilde{\pi}^T = \tilde{\rho}\,\begin{pmatrix} * & \tilde{\sigma}^T \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & I \end{pmatrix} L^{-1} P,$$

where $\tilde{\rho}$ ensures $\tilde{\pi}^T\mathbf{1} = 1$.

Algorithm 2 can be viewed as applying one step of the power method, with the preconditioned matrix

$$\tilde{P} \equiv \begin{pmatrix} 0 & 0 \\ 0 & I \end{pmatrix} L^{-1} P$$

to the vector $\begin{pmatrix} * & \tilde{\sigma}^T \end{pmatrix}$ and then normalizing it.

In the next section we execute Algorithm 2 repeatedly.

5. Iterative Aggregation/Disaggregation (IAD). The iterative aggregation/disaggregation (IAD) method [7, §8.2.2], [8, §3.2.1] improves the PageRank algorithm [2]. It consists of repeated application of Algorithm 2. We show that the IAD method can be viewed as a preconditioned power method and give a simple derivation of the asymptotic convergence rate. For simplicity we view the IAD method as an alternative to the power method rather than an updating algorithm.

ALGORITHM 3 (Iterative Aggregation/Disaggregation (IAD)). *Determine an approximation $\pi^{(k)}$ to the stationary distribution π of a stochastic irreducible matrix P, in k iterations.*

1. *Select a vector $\pi^{(0)} = \begin{pmatrix} \pi_1^{(0)} & \pi_2^{(0)} \end{pmatrix}$ with $\pi_2^{(0)} > 0$.*
2. *Do $k = 1, 2 \ldots$*
   (a) *Normalize $\sigma^{(k)} \equiv \pi_2^{(k-1)}/[\pi_2^{(k-1)}]^T\mathbf{1}$.*
   (b) *Determine the stationary distribution $\alpha^{(k)}$ of $A^{(k)} \equiv \begin{pmatrix} P_{11} & P_{12}\mathbf{1} \\ [\sigma^{(k)}]^T P_{21} & [\sigma^{(k)}]^T P_{22}\mathbf{1} \end{pmatrix}$.*
   (c) *Partition $\alpha^{(k)} = \begin{pmatrix} \omega_1^{(k)} \\ \rho_k \end{pmatrix}$, and set $\omega^{(k)} \equiv \begin{pmatrix} \omega_1^{(k)} \\ \rho_k \sigma^{(k)} \end{pmatrix}$.*
   (d) *Multiply $[\pi^{(k)}]^T \equiv [\omega^{(k)}]^T P$.*

The IAD Algorithm 3 performs, in exact arithmetic, an ILU preconditioned power method.

PROPOSITION 5.1. *Let P be stochastic and irreducible and $\pi^{(0)} = \begin{pmatrix} \pi_1^{(0)} & \pi_2^{(0)} \end{pmatrix}$ a vector with $\pi_2^{(0)} > 0$. Then Algorithm 3 produces iterates*

$$[\pi^{(k)}]^T = \tilde{\rho}\, [\pi^{(0)}]^T \tilde{P}^k, \qquad \text{where} \quad \tilde{P} \equiv \begin{pmatrix} 0 & 0 \\ 0 & I \end{pmatrix} L^{-1} P,$$

and $\tilde{\rho}$ ensures $[\pi^{(k)}]^T \mathbf{1} = 1$.

Proof. This follows from Remark 4.1. □

To determine the asymptotic convergence rate of Algorithm 3, label the eigenvalues $\lambda_i(P)$ of a matrix P in order of decreasing magnitude, $|\lambda_1(P)| \geq |\lambda_2(P)| \geq \ldots$.

THEOREM 5.2 (Convergence Rate). *If P is stochastic and irreducible then the asymptotic convergence rate of Algorithm 3 is $|\lambda_2(S)|$.*

Proof. According to Proposition 5.1, Algorithm 3 represents a power method with the matrix

$$\tilde{P} = \begin{pmatrix} 0 & 0 \\ P_{21}(I - P_{11})^{-1} & S \end{pmatrix}.$$

Because $\tilde{P}$ is block triangular, its non-zero eigenvalues are those of S. Since P is stochastic and irreducible, so is S [10, Theorem 2.3]. Hence its spectral radius is $|\lambda_1(S)| = 1$. Therefore the power method applied to $\tilde{P}$ has asymptotic convergence rate $|\lambda_2(S)|/|\lambda_1(S)| = |\lambda_2(S)|$. □

As a consequence, the IAD Algorithm 3 converges like the power method applied to the stochastic complement S. The asymptotic convergence rate of Algorithm 3 is equal to the second eigenvalue of S. The example below illustrates the effect of the asymptotic convergence rate.

EXAMPLE 5.1. *Consider the stochastic irreducible matrix $P = aI + (1-a)\mathbf{1}\pi^T$ from Example 3.1, with $0 < a < 1$, $\pi^T\mathbf{1} = 1$ and $\pi > 0$*

The stochastic complement is $S = aI_{n-m} + (1-a)\mathbf{1}\sigma^T$ with $\sigma^T = \pi_2^T / \pi_2^T \mathbf{1}$. The subdominant eigenvalue of S is $\lambda_2(S) = a$. Hence the asymptotic convergence rate of Algorithm 3 is a.

What does this mean for the iterates $\pi^{(k)}$? Any positive probability vector $\tilde{\sigma}$ produces $\tilde{A} = A$, thus $\alpha^{(k)} = \alpha$ for all k. This means after one iteration the leading components of $\pi^{(k)}$ have converged, i.e. $\pi_1^{(k)} = \pi_1$ for all k. As for the trailing components of $\pi^{(k)}$, $[\pi_2^{(k)}]^T$ is a multiple of $[\pi_2^{(0)}]^T S^k$, where $S^k = a^k I_{n-m} + (1-a^k)\mathbf{1}\sigma^T$. Hence

$$[\sigma^{(k)}]^T = a^{k-1}\, \frac{[\pi_2^{(0)}]^T}{\gamma} + (1 - a^{k-1})\, \frac{\pi_2^T}{\tilde{\rho}}, \qquad \tilde{\rho} \equiv \pi_2^T \mathbf{1}, \quad \gamma \equiv [\pi_2^{(0)}]^T \mathbf{1}.$$

The iterates are then

$$[\pi^{(k)}]^T = \begin{pmatrix} \pi_1^T & a^k\, \frac{\tilde{\rho}}{\gamma}[\pi_2^{(0)}]^T + (1-a^k)\pi_2^T \end{pmatrix}.$$

The leading component π_1 is exact, and the trailing component $a^k\, \frac{\tilde{\rho}}{\gamma}[\pi_2^{(0)}]^T + (1-a^k)\pi_2^T$ shows clearly that the contribution of the initial vector $\pi_2^{(0)}$ decreases as a power of $a = \lambda_2(S)$, while the contribution of the target vector π_2 increases at the same rate.

6. Stochastic Complements. In this section we consider how different choices of stochastic complements affect the convergence rate of the IAD Algorithm 3.

Remember that the asymptotic convergence rate of the power method applied to a primitive stochastic matrix P is $|\lambda_2(P)|$. In Theorem 5.2 we showed that the asymptotic convergence rate of Algorithm 3 applied to

$$P = \begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix}$$

is $|\lambda_2(S)|$, where S is the stochastic complement of P_{22}. Unfortunately we cannot take for granted that Algorithm 3 has a better convergence rate than the power method, i.e. that $|\lambda_2(S)| < |\lambda_2(P)|$. In fact, the opposite is possible, as we illustrate in one of the examples below.

One remedy for accelerating Algorithm 3 in case of slow convergence is to be more flexible about the choice of stochastic complements and not to limit oneself to stochastic complements from *leading* principal submatrices. One can form a stochastic complement from any principal submatrix [10, Definition 2.1]. One way to think about this (to keep the notation simple) is to apply a permutation similarity transformation to P [10, Lemma 2.1], i.e.

$$QPQ^T = \begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix},$$

where Q is a permutation matrix, and use the stochastic complement of the permuted matrix

$$S \equiv P_{22} + P_{21}(I - P_{11})^{-1} P_{12}.$$

Note that a permutation similarity merely reorders the components of the stationary distribution and preserves the eigenvalues. That is, we can choose different stochastic complements by thinking of applying Algorithm 3 to a matrix QPQ^T that is permutationally similar to P. The question is, can we find permutations Q such that the stochastic complement S of QPQ^T has $|\lambda_2(S)| \leq |\lambda_2(P)|$? That is, can Algorithm 3 converge faster than the power method on a suitably reordered matrix?

The examples below illustrate different convergence possibilities for the power method and Algorithm 3.

EXAMPLE 6.1. *This example illustrates that Algorithm 3 can converge in one iteration, while the power method fails.*

Let v be a positive probability vector, and

$$P = \begin{pmatrix} 0 & v^T \\ \mathbf{1} & 0 \end{pmatrix}$$

a stochastic matrix. The eigenvalues of P are 1, -1 and 0. Hence $|\lambda_2(P)| = 1$, which means that the power method applied to P diverges (unless the initial vector is the stationary distribution). The stochastic complement that arises from partitioning off the first row and column of P is $S = \mathbf{1}v^T$. Its eigenvalues are 1 and 0. Thus $|\lambda_2(S)| = 0$, so that Algorithm 3 converges in a single iteration.

EXAMPLE 6.2. *This example illustrates that the power method and Algorithm 3 can have the same rate of convergence, regardless of the choice of stochastic complement.*

Consider the matrix from Examples 3.1 and 5.1, $P = aI + (1-a)\mathbf{1}\pi^T$, where π is a positive probability vector and $0 < a < 1$. The matrix P has just two distinct eigenvalues, $\lambda_1(P) = 1$ and $\lambda_i(P) = a$, $i > 1$. Each stochastic complement of P has the form $S = aI + (1-a)\mathbf{1}\sigma^T$, where σ is a positive probability vector. The eigenvalues of S are also 1 and a. In particular, $\lambda_2(S) = \lambda_2(P)$, so that the power method and Algorithm 3 converge at the same rate.

Example 6.3. *This example illustrates that the convergence rate of Algorithm 3 can be significantly worse than that of the power method, regardless of the choice of stochastic complement.*

The irreducible stochastic matrix

$$P = \begin{pmatrix} 5/6 & 0 & 1/6 \\ 3/4 & 1/6 & 1/12 \\ 2/3 & 1/3 & 0 \end{pmatrix}.$$

has eigenvalues $\lambda_1(P) = 1$ *and* $\lambda_2(P) = \lambda_3(P) = 0$*. It turns out that* 0 *is a defective eigenvalue, corresponding to a Jordan block of order* 2*. Thus the power method converges in at most two iterations for any initial probability vector.*

Now let's consider all possible stochastic complements that can arise during Algorithm 3. Denote by S_i *the stochastic complement that arises from* P *by simultaneously permuting row and column* i *to the first position, and then forming the corresponding* 2×2 *stochastic complement. The stochastic complements have the following subdominant eigenvalues:*

$$\lambda_2(S_1) = -1/6, \qquad \lambda_2(S_2) = -2/15, \qquad \lambda_2(S_3) = 5/36.$$

In all cases $|\lambda_2(S_i)| > |\lambda_2(P)| = 0$*,* $1 \le i \le 3$*. Thus for any choice of stochastic complement, the convergence rate of Algorithm 3 is slower than that of the power method.*

7. An Upper Bound on the Convergence Rate.

We discuss an upper bound on the subdominant eigenvalue $|\lambda_2|$, which also furnishes upper bounds for the asymptotic convergence rates of the power method and the IAD Algorithm 3. The upper bound can be easier to compute than $|\lambda_2|$ and has some properties convenient for a convergence analysis.

Example 6.3 shows that for any stochastic complement S of an irreducible stochastic matrix P, $|\lambda_2(S)| > |\lambda_2(P)|$ is possible, which means that for such a matrix the convergence rate of Algorithm 3 is always slower than that of the power method. As far as Algorithm 3 is concerned, this situation is undesirable. We can make things look better for Algorithm 3 if we replace $|\lambda_2|$ by a suitably chosen upper bound τ for which $\tau(S) > \tau(P)$ is impossible. Denote by $\|v^T\| \equiv |v|^T \mathbf{1}$ the one-norm of a column vector v, and by e_j the jth column of the identity matrix. With

$$\tau(P) \equiv \frac{1}{2} \max_{i,j} \|(e_i - e_j)^T P\|,$$

we have for any stochastic matrix P [14, §4],

$$|\lambda_2(P)| \le \tau(P). \tag{7.1}$$

Observe that $\tau(P) \le 1$ for any stochastic matrix P. The quantity $\tau(P)$ represents an upper bound on the convergence rate of the power method, while $\tau(S)$ is an upper bound on the convergence rate of Algorithm 3. In contrast to the eigenvalues, where $|\lambda_2(S)| > |\lambda_2(S)|$ is possible, this cannot happen for the upper bounds τ. The following result shows that $\tau(S) \le \tau(P)$, under appropriate conditions.

Proposition 7.1. *If*

$$P = \begin{array}{c} \\ k \\ n-k \end{array} \begin{array}{c} \begin{array}{cc} k & n-k \end{array} \\ \begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix} \end{array}$$

is stochastic and irreducible, and $S \equiv P_{22} + P_{21}(I - P_{11})^{-1} P_{12}$ *then*

$$\tau(S) \le \tau(P).$$

Proof. We claim that $\|u^T(I-P_{11})^{-1}P_{12}\| \leq \|u\|$ for *all* vectors u. To see that the claim is true, observe that $P\mathbf{1} = \mathbf{1}$ implies $(I-P_{11})^{-1}P_{12}\mathbf{1} = \mathbf{1}$. With $v_j \equiv |u_j|$, $1 \leq j \leq k$, the triangle inequality implies

$$|u^T(I-P_{11})^{-1}P_{12}e_i| \leq v^T(I-P_{11})^{-1}P_{12}e_i, \qquad 1 \leq i \leq n-k.$$

Hence

$$\|u^T(I-P_{11})^{-1}P_{12}\| \leq \sum_{i=1}^{n-k} v^T(I-P_{11})^{-1}P_{12}e_i = v^T(I-P_{11})^{-1}P_{12}\mathbf{1} = v^T\mathbf{1} = \|u\|,$$

which proves the claim.

For any two indices $1 \leq i, j \leq n-k$, the triangle inequality and the above claim imply

$$\begin{aligned}\|(e_i-e_j)^T S\| &\leq \|(e_i-e_j)^T P_{22}\| + \|(e_i-e_j)^T P_{21}(I-P_{11})^{-1}P_{12}\| \\ &\leq \|(e_i-e_j)^T P_{22}\| + \|(e_i-e_j)^T P_{21}\| = \|(e_{k+i}-e_{k+j})^T P\|.\end{aligned}$$

Hence

$$\tau(S) \leq \frac{1}{2}\max_{k+1\leq i,j\leq n} \|(e_i-e_j)^T P\| \leq \tau(P).$$

□

Proposition 7.1 shows that the upper bound on the convergence rate of Algorithm 3 is never worse than that of the power method.

The result below implies that there are matrices (including the Google matrix in §9) for which the power method always converges, and for which the convergence rate of Algorithm 3 is at least as good as that of the power method.

Corollary 7.2. *Let $G \equiv cP + (1-c)\mathbf{1}v^T$, where P is a stochastic matrix, $0 < c < 1$, and v is a positive probability vector. If $|\lambda_2(P)| = 1$ then for any stochastic complement S of G*

$$|\lambda_2(S)| \leq |\lambda_2(G)| = c < 1.$$

Proof. Since v is positive and $c < 1$, G is irreducible. Inequality (7.1) and Proposition 7.1 imply $|\lambda_2(S)| \leq \tau(S) \leq \tau(G)$. The results [4, Theorem 2], [9, Theorem 4.1] and (7.1) imply $c = |\lambda_2(G)| \leq \tau(G)$. At last, $\tau(G) \leq c$ follows from $\tau(G) = c\tau(P) \leq c$. □

Note that $|\lambda_2(G)| = c$ in Corollary 7.2 was already shown in [4, Theorem 2], [9, Theorem 4.1]. The crucial inequality in Corollary 7.2 is $|\lambda_2(S)| \leq |\lambda_2(G)|$. The matrices in Examples 3.1, 5.1 and 6.2 are special cases of the matrices in Corollary 7.2 with $P = I$.

For certain matrices, the upper bound τ provides easily an computable upper bound on the convergence rate of Algorithm 3. These include matrices where P_{21} or P_{22} has a column with all elements positive.

Remark 7.1. *If P is irreducible and stochastic, and $p_{im} > 0$ for some m and $k+1 \leq i \leq n$ then*

$$|\lambda_2(S)| \leq 1 - \min_{k+1\leq i\leq n} p_{im} < 1.$$

To see why this is true, note that rows i and j, $k+1 \leq i, j \leq n$, have positive entries in position m and

$$\|(e_i-e_j)^T P\| = \sum_{l=1}^{n} |p_{il}-p_{jl}| \leq 1 - p_{im} + 1 - p_{jm} + |p_{im}-p_{jm}| = 2(1-\min\{p_{im}, p_{jm}\}).$$

Proposition 7.1 implies $|\lambda_2(S)| \leq 1 - \min_{k+1\leq i\leq n} p_{im} < 1$.

8. Conditions for Convergence. We derive conditions that guarantee convergence, in exact arithmetic, of the IAD Algorithm 3. Operation counts and computation times for the power method, Algorithm 3 and other methods are compared in [7, §6, §8].

The power method with a stochastic irreducible matrix P converges if $|\lambda_2(P)| < 1$, or equivalently if P is primitive [1, Definition 2.(1.8)]. However, just because P is primitive, a stochastic complement S is not necessarily primitive [10, §5].

REMARK 8.1. *Let*

$$P = \begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix}$$

be stochastic and irreducible. If P_{22} *is primitive or* P_{22} *has at least one non-zero diagonal element then Algorithm 3 converges.*

This follows from [10, Theorems 5.1 and 5.2].

REMARK 8.2. *Primitivity of a stochastic complement* S *can also be related to properties of the directed graph* $\Delta(S)$*, which is also known as the compressed directed graph [6, §1, §4]. A useful property of the compressed directed graph is the following: for any pair of indices* i*,* j *with* $1 \leq i, j \leq n-k$*, if there is a path from vertex* $k+i$ *to vertex* $k+j$ *in* $\Delta(P)$*, then there is a path from vertex* i *to vertex* j *in* $\Delta(S)$*.*

THEOREM 8.1. *If* P *is irreducible and stochastic then the following holds:*

P *has at least one primitive stochastic complement of order at least two* $\Longleftrightarrow$ P *is not a cyclic permutation matrix.*

Proof. First, we claim that P is a cyclic permutation matrix if and only if each column of P has at most one positive entry. One direction of the claim is clear. To see the other direction, suppose that each column of P has at most one positive entry. Because P is irreducible, each column must have exactly one positive entry. Further, since the entries of P are at most 1, and the sum of all entries in P is equal to the order of the matrix, the single positive entry in each column of P must be a 1. Thus, P is an irreducible stochastic $(0, 1)$ matrix, hence P is a cyclic permutation matrix.

With the claim established, we now proceed with the proof of the result.

$\Longleftarrow$: Suppose that P is not a cyclic permutation matrix. Then some column m of P has at least two positive entries, i.e. $p_{i_1,m}, \ldots, p_{i_q,m} > 0$ for $i_1, \ldots, i_q$ with $q \geq 2$. With a suitable permutation Q, partition

$$QPQ^T = \begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix}$$

so that P_{22} is $q \times q$, and P_{21} or P_{22} has a positive column. Remark 7.1 implies that $|\lambda_2(S)| < 1$, so that S is primitive.

$\Longrightarrow$: Suppose that

$$P = \begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix}$$

is a cyclic permutation matrix. Since $\Delta(P)$ has exactly one cycle involving every vertex, and since $\Delta(P_{11})$ is a proper subgraph of $\Delta(P)$, $\Delta(P_{11})$ contains no cycles. Hence P_{11} is nilpotent, so

$(I-P_{11})^{-1}$ is a finite sum of powers of P_{11}. In particular, $(I-P_{11})^{-1}$ is a matrix with nonnegative integer entries. Hence the stochastic complement $S = P_{22} + P_{21}(I-P_{11})^{-1}P_{12}$ is also a matrix of nonnegative integers. Since S is stochastic, it is a $(0,1)$ matrix, and since S is irreducible, it is a cyclic permutation matrix. Therefore S is not primitive.

□

COROLLARY 8.2. *If the irreducible stochastic matrix P is not a cyclic permutation matrix, then P has at least one stochastic complement S for which Algorithm 3 converges, i.e. $|\lambda_2(S)| < 1$.*

9. The Google Matrix. We show that, when Algorithm 3 is applied to the Google matrix G with an appropriate choice of stochastic complement, then the convergence rate is better than that of the power method. We also show how one can save computations when applying Algorithm 3 to G.

The Google matrix is constructed as follows [9, §4.2, §4.5], [4, §2]. The hyperlink structure of the web can be represented as a directed graph, whose vertices are web pages and whose edges are links [7, §2]. If web page i has $d_i \geq 1$ outgoing links, then for each link from web page i to page j, entry (i,j) of the matrix P is $1/d_i$. If there is no link from page i to page j, then the (i,j) entry of P is 0. Finally, if page i has no outgoing links at all, then row i of P equals $\frac{1}{n}\mathbf{1}^T$, where n is the order of P. Thus P is a stochastic matrix, and in order to construct a primitive stochastic matrix, one forms the convex combination

$$G = cP + (1-c)\mathbf{1}v^T,$$

where v is a positive probability vector and $0 < c < 1$ a scalar. This is the Google matrix [9, §4.2, §4.5], [4, §2]. Element i of the stationary probability of G is the PageRank of page i [7, §2]; it is the probability that a random surfer visits web page i [2, §2.1.2]. The PageRank is used by the search engine Google to estimate the citation importance of a web page and to order the results of web keyword searches [2, §2.1].

In Corollary 7.2 we showed that if $|\lambda_2(P)| = 1$ then $|\lambda_2(S)| \leq |\lambda_2(G)| = c < 1$ for any stochastic complement S. This means the power method applied to G converges, and the convergence rate of Algorithm 3 is at least as good as that of the power method. Now we prove a stronger result by exploiting the structure of the Google matrix and by choosing a particular stochastic complement: In this case the convergence rate of Algorithm 3 is strictly better than that of the power method. We make the following assumptions about the web.

ASSUMPTIONS 1.

1. *The stochastic matrix P contains $k \geq 2$ essential index classes $C_1, \ldots, C_k$. The class of inessential indices (if any) is called D.*
   Recall that an index i is inessential if for some index j there is a chain of links from page i to page j, but there is no chain of links back from page j to page i (i.e. $(P^m)_{ij} > 0$ for some power P^m, but $(P^k)_{ij} = 0$ for all powers P^k); otherwise the index i is essential. An essential index class C is a set of essential indices, such that there is a chain of links between any two indices in C, but no outgoing link to an index outside of C [13, §1.2].
2. *Each essential index class C_j contains an index i_j whose corresponding diagonal entry in P is 0, i.e. at least one web page in each C_j does not link back to itself.*

These assumptions imply that the rows of P that are equal to $\frac{1}{n}\mathbf{1}^T$ all correspond to inessential indices in D. Also, $|\lambda_2(P)| = 1$ because P has at least 2 essential index classes.

When P satisfies Assumptions 1 there is a permutation matrix Q that orders the rows and colums of QPQ^T into the following form: $i_1, \ldots, i_k, C_1 \setminus \{i_1\}, \ldots, C_k \setminus \{i_k\}, D$. Partition

$$QGQ^T = \begin{pmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{pmatrix},$$

where G_{11} is $k \times k$ and corresponds to the indices $i_1, \ldots, i_k$, and let $S_G \equiv G_{22} + G_{21}(I - G_{11})^{-1}G_{12}$ be the corresponding stochastic complement.

THEOREM 9.1. *Let $G = cP + (1-c)\mathbf{1}v^T$ where $0 < c < 1$, v is a positive probability vector, and the stochastic matrix P satisfies Assumptions 1. Then*

$$|\lambda_2(S_G)| < |\lambda_2(G)|.$$

Proof. We exploit the structure of P, in several stages, to capture the structure of the stochastic complement S_G. Partition

$$QPQ^T = \begin{pmatrix} 0 & P_{12} \\ P_{21} & P_{22} \end{pmatrix}$$

where the leading diagonal block of order k is zero because the indices $i_1, \ldots, i_k$ belong to different essential classes, and each index corresponds to a zero diagonal element of P. While QPQ^T is presented as a 2×2 block matrix, whose leading principal block corresponds to indices $i_1, \ldots, i_k$, we will have occasion to further partition the second block into two classes, $C_1 \setminus \{i_1\}, \ldots, C_k \setminus \{i_k\}$, and D.

Partitioning $v = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}$ conformally with P gives for the partition of QGQ^T,

$$G_{11} = (1-c)\mathbf{1}v_1^T, \qquad G_{12} = cP_{12} + (1-c)\mathbf{1}\,(\, v_2^T \quad v_3^T\,),$$
$$G_{21} = cP_{21} + (1-c)\mathbf{1}v_1^T, \quad G_{22} = cP_{22} + (1-c)\mathbf{1}\,(\, v_2^T \quad v_3^T\,).$$

We claim that $|\lambda_2(S_G)| < c$, from which the conclusion follows with Corollary 7.2.

To establish this claim, write $S_G = G_{22} + G_{21}(I - G_{11})^{-1}G_{12}$ as a convex combination $S_G = cT + (1-c)\mathbf{1}z^T$, where

$$T \equiv P_{22} + P_{21}(I - G_{11})^{-1}G_{12}, \qquad z^T \equiv (\, v_2^T \quad v_3^T\,) + v_1^T(I - G_{11})^{-1}G_{12}.$$

Using the three facts that $(I - G_{11})^{-1} = I + \beta\mathbf{1}v_1^T$, where

$$\beta \equiv \frac{1-c}{1-(1-c)\delta}, \qquad \delta \equiv v_1^T\mathbf{1};$$

the stochastic complement $P_{22} + P_{21}P_{12}$ of P is again a stochastic matrix; and $P_{12}\mathbf{1} = \mathbf{1}$; we conclude that T is stochastic and z is a positive probability vector. Since $\lambda_2(S_G) = c\lambda_2(T)$ [9, Theorem 4.1] it suffices to prove $|\lambda_2(T)| < 1$. To this end, we will show that T has at least one positive row and is irreducible. This implies T is primitive [1, Corollary 2.(4.8)], hence $|\lambda_2(T)| < 1$ [1, Definition 2.(1.8)].

The principal submatrix of P corresponding to the essential index class C_j can be written as

$$M_j = \begin{pmatrix} 0 & x_j^T \\ y_j & B_j \end{pmatrix},$$

where the leading row and column correspond to index i_j, and the remaining rows and columns correspond to the indices of $C_j \setminus \{i_j\}$. Since C_j is an essential index class, y_j is not the zero vector. Hence M_j is irreducible and stochastic, and so is its stochastic complement $B_j + y_j x_j^T$.

With

$$B \equiv \begin{pmatrix} B_1 & & \\ & \ddots & \\ & & B_k \end{pmatrix}, \quad X \equiv \begin{pmatrix} x_1^T & & \\ & \ddots & \\ & & x_k^T \end{pmatrix}, \quad Y \equiv \begin{pmatrix} y_1 & & \\ & \ddots & \\ & & y_k \end{pmatrix}.$$

it follows that

$$QPQ^T = \begin{pmatrix} 0 & P_{12} \\ P_{21} & P_{22} \end{pmatrix} = \left(\begin{array}{c|cc} 0 & X & 0 \\ \hline Y & B & 0 \\ R_1 & R_2 & R_3 \end{array}\right).$$

Hence

$$T = \underbrace{P_{22} + cP_{21}P_{12}} + c\beta P_{21}\mathbf{1}v_1^T P_{12} + \underbrace{(1-c)(1+\beta\delta)P_{21}\mathbf{1}\begin{pmatrix} v_2^T & v_3^T \end{pmatrix}}.$$

Consider the last summand in T. The row $\begin{pmatrix} v_2^T & v_3^T \end{pmatrix}$ is positive because v is. Moreover

$$P_{21}\mathbf{1} = \begin{pmatrix} y_1 \\ \vdots \\ y_k \\ R_1 \end{pmatrix},$$

where each vector y_j has at least one positive element. This means, T has at least one positive row corresponding to each $C_j \setminus \{i_j\}$, $1 \le j \le k$.

We still need to prove that T is irreducible. T being irreducible is equivalent to the directed graph $\Delta(T)$ being strongly connected [1, 2.(2.7)]. We will show that for any vertex in $\Delta(T)$ there is a path to any other vertex.

1. Paths within $C_j \setminus \{i_j\}$, $1 \le j \le k$:
   Consider the first two summands in T,

   $$P_{22} + cP_{21}P_{12} = \begin{pmatrix} B + cYX & 0 \\ R_2 + cR_1X & R_3 \end{pmatrix}.$$

   The leading principal submatrix is

   $$B + cYX = \begin{pmatrix} B_1 + cy_1x_1^T & & \\ & \ddots & \\ & & B_k + cy_kx_k^T \end{pmatrix}.$$

   Every diagonal element $B_j + cy_jx_j^T$ is irreducible because its directed graph is the same as that of the irreducible matrix $B_j + y_jx_j^T$, i.e. $\Delta(B_j + cy_jx_j^T) = \Delta(B_j + y_jx_j^T)$. Therefore $\Delta(T)$ contains a path from every vertex in $C_j \setminus \{i_j\}$ to all other vertices in $C_j \setminus \{i_j\}$.
2. Paths leaving $C_j \setminus \{i_j\}$, $1 \le j \le k$:
   The above fact that T contains at least one positive row for each $C_j \setminus \{i_j\}$ implies that $\Delta(T)$ contains a path from every vertex in $C_j \setminus \{i_j\}$ to any other vertex in $\Delta(T)$. Together with 1. this implies $\Delta(T)$ contains a path from every vertex in $C_j \setminus \{i_j\}$ to any other vertex in $\Delta(T)$, $1 \le j \le k$.
3. Paths leaving D:
   From the leading two summands of T, $P_{22} + cP_{21}P_{12}$ with $c > 0$, we see that $\Delta(T)$ contains $\Delta(P_{22} + cP_{21}P_{12}) = \Delta(S_P)$, where

   $$S_P \equiv P_{22} + P_{21}P_{12} = \begin{pmatrix} B & 0 \\ R_2 & R_3 \end{pmatrix} + \begin{pmatrix} Y \\ R_1 \end{pmatrix} \begin{pmatrix} X & 0 \end{pmatrix}$$

   is a stochastic complement of P. Since every inessential index $d \in D$ has a path to some essential index class C_j, $\Delta(P)$ contains a path from d to every vertex in C_j. Remark 8.1 implies that therefore $\Delta(S_P)$ also contains a path from d to every vertex of $C_j \setminus \{i_j\}$. Since $\Delta(T)$ contains $\Delta(S_P)$, every vertex in D must have a path to some vertex of $C_j \setminus \{i_j\}$. With 2. we conclude that $\Delta(T)$ contains a path from every vertex in D to any other vertex in D.

□

Theorem 9.1 shows that for the Google matrix, there is always a choice of stochastic complement so that the convergence rate of Algorithm 3 is strictly smaller than that of the power method. We give a simple illustration below.

Example 9.1. *This example illustrates that the convergence rate of Algorithm 3 applied to the Google matrix G can be the square of that of the power method.*

Let $G = cP + (1-c)\mathbf{1}v^T$, where $0 < c < 1$, v is a positive probability vector and

$$P = \begin{pmatrix} 0 & I \\ I & 0 \end{pmatrix}.$$

If P is of order $n = 2k$ then its directed graph $\Delta(P)$ is a union of k cycles, each of length 2.

Partitioning $v^T = (\, v_1^T \quad v_2^T \,)$ conformally with P gives for the stochastic complement from Theorem 9.1

$$S_G = (1-c)\mathbf{1}v_2^T + \left(cI + (1-c)\mathbf{1}v_1^T\right)\left(I - (1-c)\mathbf{1}v_1^T\right)^{-1}\left(cI + (1-c)\mathbf{1}v_2^T\right).$$

Multiplying out yields

$$S_G = c^2 I + \frac{1-c^2}{1-(1-c)\delta}\mathbf{1}(cv_1^T + v_2^T), \qquad \text{where} \quad \delta \equiv v_1^T\mathbf{1}.$$

The eigenvalues of S_G are 1 and c^2.

Hence the convergence rate of Algorithm 3 for this matrix G is c^2, while the convergence rate of the power method is only c.

Below we illustrate that when Algorithm 3 is applied to the Google matrix G, the computation of the stationary vectors of the aggregated matrices $A^{(k)}$ is cheap.

Remark 9.1. *Algorithm 3 applied to the Google matrix G involves the computation of stationary vectors of aggregated matrices*

$$A^{(k)} = \begin{pmatrix} G_{11} & G_{12}\mathbf{1} \\ [\sigma^{(k)}]^T G_{21} & [\sigma^{(k)}]^T G_{22}\mathbf{1} \end{pmatrix}.$$

For the stochastic complement S_G from Theorem 9.1, $G_{11} = (1-c)\mathbf{1}v_1^T$. Hence

$$A^{(k)} = \begin{pmatrix} (1-c)\mathbf{1}v_1^T & \gamma\mathbf{1} \\ [z^{(k)}]^T & 1 - [z^{(k)}]^T\mathbf{1} \end{pmatrix}, \qquad [z^{(k)}]^T \equiv [\sigma^{(k)}]^T G_{21}, \quad \gamma \equiv 1 - (1-c)\delta, \quad \delta \equiv v_1^T\mathbf{1}.$$

The simple structure of $A^{(k)}$ yields an explicit expression for its stationary vector,

$$\alpha^{(k)} = \frac{1}{\epsilon_k + \gamma}\left(\, \epsilon_k(1-c)v_1^T + \gamma[z^{(k)}]^T \quad \gamma \,\right), \qquad \text{where} \quad \epsilon_k \equiv [z^{(k)}]^T\mathbf{1}.$$

Thus, when Algorithm 3 is applied to the Google matrix G, the computation of the stationary vectors $\alpha^{(k)}$ is cheap.

REFERENCES

[1] A. Berman and R. Plemmons, *Nonnegative Matrices in the Mathematical Sciences*, SIAM Classics In Applied Mathematics, SIAM, Philadelphia, 1994.

[2] S. Brin and L. Page, *The anatomy of a large-scale hypertextual web search engine*, Comput. Networks and ISDN Systems, 30 (1998), pp. 107–17.

[3] R. Cottle, *Manifestations of the Schur complement*, Linear Algebra Appl., 8 (1974), pp. 189–211.

[4] T. Haveliwala and S. Kamvar, *The second eigenvalue of the Google matrix*, Tech. Rep. 2003-20, Stanford University, http://dbpubs.stanford.edu/pub/2003-20, March 2003.

[5] T. Haveliwala, S. Kamvar, D. Klein, C. Manning, and G. Golub, *Computing pagerank using power extrapolation*, Tech. Rep. 2003-45, Stanford University, http://dbpubs.stanford.edu/pub/2003-45, July 2003.

[6] C. Johnson and C. Xenophontos, *Irreducibility and primitivity of Perron complements: Application of the compressed directed graph*, in Graph Theory and Sparse Matrix Computation, vol. 56 of IMA Vol. Math. Appl., Springer, New York, 1993, pp. 101–6.

[7] A. Langville and C. Meyer, *Updating PageRank using the group inverse and stochastic complementation*, (2002). CRSC-TR02-32 at www.ncsu.edu/crsc/reports/reports02.html.

[8] ———, *Updating the stationary vector of an irreducible Markov chain*, (2002). CRSC-TR02-33 at www.ncsu.edu/crsc/reports/reports02.html.

[9] ———, *Fiddling with PageRank*, (2003). CRSC-TR03-34 at www.ncsu.edu/crsc/reports/reports03.html.

[10] C. Meyer, *Stochastic complementation, uncoupling Markov chains, and the theory of nearly reducible systems*, SIAM Rev., 31 (1989), pp. 240–72.

[11] ———, *Uncoupling the Perron eigenvector problem*, Linear Algebra Appl., 114/115 (1989), pp. 69–94.

[12] C. Moler, *The world's largest matrix computation*, Matlab News and Notes, (2002).

[13] E. Seneta, *Non-Negative Matrices and Markov Chains*, Springer-Verlag, New York, 1981.

[14] ———, *Explicit forms for ergodicity coefficients and spectrum localization*, Linear Algebra Appl., 60 (1984), pp. 187–97.

[15] W. Stewart, *Introduction to the Numerical Solution of Markov Chains*, Princeton University Press, Princeton, 1994.

[16] R. Varga, *Matrix Iterative Analysis*, Prentice Hall, 1962.