

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/247162459>

State space orderings for Gauss–Seidel in Markov chains revisited

Article in *SIAM Journal on Scientific Computing* · January 1998

DOI: 10.1137/S1064827596303612

CITATIONS

17

READS

13

1 author:



Tuğrul Dayar

Bilkent University

74 PUBLICATIONS 636 CITATIONS

SEE PROFILE

STATE SPACE ORDERINGS FOR GAUSS–SEIDEL IN MARKOV CHAINS REVISITED*

TUĞRUL DAYAR†

Abstract. States of a Markov chain may be reordered to reduce the magnitude of the subdominant eigenvalue of the Gauss–Seidel (GS) iteration matrix. Orderings that maximize the elemental mass or the number of nonzero elements in the dominant term of the GS splitting (that is, the term approximating the coefficient matrix) do not necessarily converge faster. An ordering of a Markov chain that satisfies Property-R is semiconvergent. On the other hand, there are semiconvergent state space orderings that do not satisfy Property-R. For a given ordering, a simple approach for checking Property-R is shown. Moreover, a version of the Cuthill–McKee algorithm may be used to order the states of a Markov chain so that Property-R is satisfied. The computational complexity of the ordering algorithm is less than that of a single GS iteration. In doing all this, the aim is to gain insight into (faster) converging orderings.

Key words. state space ordering, Markov chains, Gauss–Seidel, Property-R, Cuthill–McKee algorithm

AMS subject classifications. 65U05, 60J10, 60J27, 65F10, 65F50, 65F30, 65B99

PII. S1064827596303612

1. Introduction. One of the problems in which iterative methods are employed is the computation of the stationary distribution vector of a large continuous-time Markov chain (CTMC). These chains arise, for instance, in reliability modeling, queuing network analysis, large scale economic modeling, and computer system performance evaluation. The problem amounts to finding a nontrivial solution to a homogeneous system of linear algebraic equations with a normalization constraint

$$(1.1) \quad \boldsymbol{\pi} \mathbf{Q} = \mathbf{0}, \quad \|\boldsymbol{\pi}\|_1 = 1,$$

where \mathbf{Q} is the $(n \times n)$ singular infinitesimal generator matrix (i.e., CTMC, transition rate matrix), $\boldsymbol{\pi}$ is the unknown $(1 \times n)$ stationary vector to be determined, and $\mathbf{0}$ represents a row vector of all zeros; in many important applications \mathbf{Q} is sparse. The off-diagonal elements of \mathbf{Q} are nonnegative and its diagonal elements are given by $q_{i,i} = -\sum_{j \neq i} q_{i,j}$. Throughout this paper, we adhere to the irreducibility assumption in \mathbf{Q} and add that $-\mathbf{Q}$ is a singular M-matrix. The discussion in this paper also applies to the one-step stochastic transition probability matrix \mathbf{P} if $\mathbf{P} - \mathbf{I}$ is used instead of \mathbf{Q} .

In what follows, boldface capital letters denote matrices, \mathbf{e} represents a column vector of all ones, $\boldsymbol{\pi}$ is a row vector, and \mathbf{q} is a column vector.

In conformity with [5], let us rewrite the generator matrix as

$$(1.2) \quad \mathbf{Q} = \mathbf{L} - (\mathbf{D} - \mathbf{U}),$$

where \mathbf{L} , $-\mathbf{D}$, and \mathbf{U} represent, respectively, strictly lower-triangular, diagonal, strictly upper-triangular parts of \mathbf{Q} . For the splitting in (1.2), $\mathbf{D} = \text{diag}((\mathbf{L} + \mathbf{U})\mathbf{e})$, and the

*Received by the editors May 15, 1996; accepted for publication (in revised form) January 16, 1997. This research was partially supported by Scientific and Technical Research Council of Turkey (TÜBİTAK) grant EEEAG-161.

<http://www.siam.org/journals/sisc/19-1/30361.html>

†Department of Computer Engineering and Information Science, Bilkent University, 06533 Bilkent, Ankara, Turkey (tugrul@cs.bilkent.edu.tr).

GS iteration may be expressed as

$$(1.3) \quad \boldsymbol{\pi}^{(k+1)}(\mathbf{D} - \mathbf{U}) = \boldsymbol{\pi}^{(k)}\mathbf{L}, \quad k = 0, 1, \dots,$$

or equivalently as (see [5, p. 357])

$$(1.4) \quad \boldsymbol{\pi}^{(k+1)} = \boldsymbol{\pi}^{(k)}\mathbf{T}_{GS}, \quad k = 0, 1, \dots,$$

where

$$(1.5) \quad \mathbf{T}_{GS} = \mathbf{L}(\mathbf{D} - \mathbf{U})^{-1}$$

and $\boldsymbol{\pi}^{(k)}$ is the approximate solution vector at the k th iteration.

It can be shown that the spectral radius of \mathbf{T}_{GS} is equal to 1; furthermore, $\boldsymbol{\pi}$ is the left eigenvector corresponding to the unit eigenvalue of \mathbf{T}_{GS} . The method of GS will converge to the stationary vector for all $\boldsymbol{\pi}^{(0)} \notin R(\mathbf{I} - \mathbf{T}_{GS})$ (i.e., the initial approximation is not in the range of $(\mathbf{I} - \mathbf{T}_{GS})$) if \mathbf{T}_{GS} does not have eigenvalues other than the unit eigenvalue on the unit circle (that is, if \mathbf{T}_{GS} is primitive) (see [6, pp. 128–130]). The asymptotic convergence rate of GS for a given ordering depends on the magnitude of the subdominant eigenvalue, γ , of the iteration matrix. For (1.5), the magnitude of the subdominant eigenvalue is given by $\gamma(\mathbf{T}_{GS}) := \max\{|\lambda| \mid \lambda \in \sigma(\mathbf{T}_{GS}), \lambda \neq 1\}$. Here $\sigma(\mathbf{T}_{GS})$ is the set of eigenvalues of \mathbf{T}_{GS} .

Orderings that maximize the probability mass or the number of nonzero elements in the dominant term (i.e., $\mathbf{D} - \mathbf{U}$ in (1.3)) of the GS splitting do not necessarily converge faster (see [4, p. 540]). We seek simple rules and/or algorithms that will identify (if possible) in a reasonable amount of time symmetric permutations of the generator matrix that are (faster) converging. The orderings for which the generator matrix has Property-R [5] are semiconvergent, and we use this as our starting point. The task is difficult because one needs to know the smallest γ to say something about the worth of an ordering at hand for a given problem. The results that appear in [3] related to forecasting the nonzero structure of the inverse of an unsymmetric matrix have helped us considerably.

2. Background material. In this section, an overview of some concepts discussed in [5] and other remarks are given. Wherever something has been taken from [5], the appropriate reference to the corresponding page(s) is given.

DEFINITION 2.1. *The fundamental matrix of the GS iteration described by (1.4) and (1.5) is the nonnegative unit upper-triangular matrix (see [5, pp. 397, 404])*

$$(2.1) \quad \mathbf{Z} = (\mathbf{D} - \mathbf{U})^{-1}\mathbf{D}.$$

Let \mathbf{B} denote the inverse of \mathbf{Z} . Then

$$(2.2) \quad \mathbf{B} = \mathbf{Z}^{-1} = \mathbf{D}^{-1}(\mathbf{D} - \mathbf{U}) = \mathbf{I} - \mathbf{D}^{-1}\mathbf{U};$$

elementwise,

$$(2.3) \quad b_{i,j} = \begin{cases} 1, & i = j, \\ \leq 0, & i < j, \\ 0, & i > j. \end{cases}$$

Remark 2.2. \mathbf{B} ($= \mathbf{Z}^{-1}$) is a nonsingular unit upper-triangular M-matrix with upper-triangular row sums that satisfy

$$\forall i < n, \quad -1 \leq \sum_{j=i+1}^n b_{i,j} \leq 0.$$

Remark 2.3. The elements of the fundamental matrix of the GS iteration, \mathbf{Z} , satisfy

$$\forall j > i, \quad 0 \leq z_{i,j} \leq 1.$$

Remark 2.2 follows from the fact that \mathbf{B} has nonpositive off-diagonal entries, positive diagonal entries, a nonnegative inverse (see Definition 2.1), and diagonal dominance with strict diagonal dominance in at least one row (due to the irreducibility assumption in \mathbf{Q}). Remark 2.3 is a result of the identity $\mathbf{BZ} = \mathbf{I}$.

The nonnegative vector that conveys much information about the nature of the GS iteration is (see [5, p. 401])

$$(2.4) \quad \mathbf{q} = \mathbf{Z}^{-1}\mathbf{e}.$$

Throughout this work, we refer to \mathbf{q} as the *fundamental vector of the GS iteration*.¹

Remark 2.4. For any ordering, $q_n = 1$ (see [5, p. 409]), $q_1 = 0$, and $0 \leq q_i \leq 1$ for $0 < i < n$.

That $z_{n,n} = 1$ is the only nonzero element in row n of \mathbf{Z} coupled with (2.4) and Remark 2.2 establishes Remark 2.4.

DEFINITION 2.5. *A matrix with a given ordering is said to have Property-R if the highest indexed state is accessible from each state by following high-stepping transitions only, i.e., transitions in the upper-triangular part of the matrix (see [5, p. 405]).*

Remark 2.6. \mathbf{Q} has Property-R iff $\forall i, z_{i,n} > 0$ (see [5, p. 406]).

Remark 2.7. Any GS iteration defined by (1.4) and (1.5) with Property-R is semiconvergent (see [5, p. 410]). However, as indicated in the next section, there are semiconvergent GS iterations of orderings that do not satisfy Property-R.

3. Checking and ordering for Property-R. In order to state the main result, we use two lemmas. However, before we proceed with the lemmas, we would like to call attention to the following form of the fundamental vector of the GS iteration defined in (2.4) (see [5, p. 401]).

$$(3.1) \quad \mathbf{q} = \mathbf{B}\mathbf{e} = (\mathbf{I} - \mathbf{D}^{-1}\mathbf{U})\mathbf{e} = \mathbf{D}^{-1}\mathbf{L}\mathbf{e}.$$

LEMMA 3.1. *If $\exists i \neq n, q_i = 1$, in (3.1), then the generator matrix \mathbf{Q} with the given ordering does not satisfy Property-R.*

Proof. Without loss of generality, assume $q_k = 1$, where $k \neq n$. Then from (3.1) and (2.3), we have

$$q_k = \sum_{j=k}^n b_{k,j} = b_{k,k} + \sum_{j=k+1}^n b_{k,j} = 1 + \sum_{j=k+1}^n b_{k,j}$$

implying, $\forall j > k, b_{k,j} = 0$. When $\forall j > k, b_{k,j} = 0$, from $\mathbf{BZ} = \mathbf{I}$, we have $\forall j > k, z_{k,j} = 0$ as well. Hence, no higher-indexed state (including state n) is accessible from state k by following high-stepping transitions in \mathbf{Q} . \square

LEMMA 3.2. *If $q_n = 1$ and $\forall i < n, q_i < 1$ in (3.1), then the generator matrix \mathbf{Q} with the given ordering satisfies Property-R.*

Proof. The proof follows a strong induction argument. If $q_{n-1} < 1$, then $b_{n-1,n} < 0$, which implies $z_{n-1,n} > 0$. This is the basis of the induction. Now, assume $\forall i \in$

¹At the risk of introducing some confusion with the elements of \mathbf{Q} we will refer to the i th element of the fundamental vector as q_i .

$\{k, k+1, \dots, n-2\}$, $q_i < 1$ where $2 \leq k \leq n-2$. That is, $\forall i \in \{k, k+1, \dots, n-2\}, \exists j > i, b_{i,j} < 0$. This is the induction hypothesis. A consequence of the induction hypothesis is that state n is accessible from all states with index k or higher. In other words, $\forall i \in \{k, k+1, \dots, n-2\}, z_{i,n} > 0$ (using the fact that $\mathbf{BZ} = \mathbf{I}$). Now, if $q_{k-1} \neq 1$, then $\exists j > k-1, b_{k-1,j} < 0$, implying the existence of one or more direct (i.e., one-step) transitions to states in $\{k, k+1, \dots, n\}$. Hence, state $k-1$ has a high-stepping random walk to the highest indexed state; consequently, $z_{k-1,n} > 0$. \square

THEOREM 3.3. *The generator matrix \mathbf{Q} with the given ordering satisfies Property-R for the GS iteration in (1.4) and (1.5) iff the fundamental vector \mathbf{q} in (3.1) satisfies*

$$(3.2) \quad q_n = 1 \quad \text{and} \quad \forall i < n, \quad q_i < 1.$$

Proof. One may view Lemma 3.2 as an implication of the form $s \Rightarrow t$. In that case, Lemma 3.1 is an implication of the form $\neg s \Rightarrow \neg t$. Combining the two implications, one obtains $s \Leftrightarrow t$. \square

From the last equality in (3.1), the i th element of \mathbf{q} is obtained as

$$(3.3) \quad \forall i < n, \quad q_i = \sum_{k=1}^{i-1} l_{i,k}/d_{i,i} = - \sum_{k=1}^{i-1} q_{i,k}/q_{i,i} = - \frac{1}{q_{i,i}} \sum_{k=1}^{i-1} q_{i,k}.$$

Example. Consider the following matrices that appear in [6, pp. 166–167] for $\lambda = \mu_1 = \mu_2 = 1$. Three different orderings, namely, lexicographical, antilexicographical, and “Marca,” are investigated. Diagonal elements of a generator matrix are the negated sums of the corresponding off-diagonal row elements, and for convenience they are denoted by asterisks. The moduli of the eigenvalues of the GS iteration matrix (in descending order) and the fundamental vector of the GS iteration are given in three decimal digits of precision beside the corresponding generator matrix.

$$\mathbf{Q}_{lex} = \begin{matrix} (0,0) \\ (0,1) \\ (0,2) \\ (1,0) \\ (1,1) \\ (1,2) \\ (2,0) \\ (2,1) \\ (2,2) \end{matrix} \begin{pmatrix} * & & & \lambda & & & & & \\ \mu_2 & * & & & \lambda & & & & \\ & \mu_2 & * & & & \lambda & & & \\ & & \mu_1 & * & & & \lambda & & \\ & & & \mu_1 & \mu_2 & * & & & \lambda \\ & & & & \mu_2 & * & & & \\ & & & & \mu_1 & * & & & \\ & & & & & \mu_1 & \mu_2 & * & \\ & & & & & & \mu_2 & * & \end{pmatrix}$$

$$|\lambda(\mathbf{T}_{GS})| = \begin{pmatrix} 1.000 \\ 1.000 \\ 0.500 \\ 0.500 \\ 0.289 \\ 0.289 \\ 0.000 \\ 0.000 \\ 0.000 \end{pmatrix} \quad \mathbf{q} = \begin{pmatrix} 0.000 \\ 0.500 \\ 0.500 \\ 0.500 \\ 0.667 \\ 0.500 \\ 1.000 \\ 1.000 \\ 1.000 \end{pmatrix},$$

$$\mathbf{Q}_{alex} = \begin{matrix} (0,0) \\ (1,0) \\ (2,0) \\ (0,1) \\ (1,1) \\ (2,1) \\ (0,2) \\ (1,2) \\ (2,2) \end{matrix} \begin{pmatrix} * & \lambda & & & & & & & \\ & * & \lambda & \mu_1 & & & & & \\ & & * & & \mu_1 & & & & \\ \mu_2 & & & * & \lambda & & & & \\ & \mu_2 & & & * & \lambda & \mu_1 & & \\ & & \mu_2 & & & * & & \mu_1 & \\ & & & \mu_2 & & & * & \lambda & \\ & & & & \mu_2 & & & * & \lambda \\ & & & & & \mu_2 & & & * \end{pmatrix}$$

$$|\lambda(\mathbf{T}_{GS})| = \begin{pmatrix} 1.000 \\ 0.250 \\ 0.083 \\ 0.000 \\ 0.000 \\ 0.000 \\ 0.000 \\ 0.000 \\ 0.000 \end{pmatrix} \quad \mathbf{q} = \begin{pmatrix} 0.000 \\ 0.000 \\ 0.000 \\ 0.500 \\ 0.333 \\ 0.500 \\ 0.500 \\ 0.500 \\ 1.000 \end{pmatrix},$$

$$\mathbf{Q}_{marca} = \begin{matrix} (1,1) \\ (2,1) \\ (0,2) \\ (1,0) \\ (1,2) \\ (2,0) \\ (0,1) \\ (2,2) \\ (0,0) \end{matrix} \begin{pmatrix} * & \lambda & \mu_1 & \mu_2 & & & & & \\ & * & & & \mu_1 & \mu_2 & & & \\ & & * & & \lambda & & \mu_2 & & \\ & & & * & & \lambda & \mu_1 & & \\ \mu_2 & & & & * & & & \lambda & \\ \mu_1 & & & & & * & & & \\ \lambda & & & & & & * & & \mu_2 \\ & \mu_2 & & & & & & * & \\ & & & \lambda & & & & & * \end{pmatrix}$$

$$|\lambda(\mathbf{T}_{GS})| = \begin{pmatrix} 1.000 \\ 0.250 \\ 0.083 \\ 0.000 \\ 0.000 \\ 0.000 \\ 0.000 \\ 0.000 \\ 0.000 \end{pmatrix} \quad \mathbf{q} = \begin{pmatrix} 0.000 \\ 0.000 \\ 0.000 \\ 0.000 \\ 0.500 \\ 1.000 \\ 0.500 \\ 1.000 \\ 1.000 \end{pmatrix}.$$

Among the three orderings, \mathbf{Q}_{alex} is the only ordering that satisfies Property-R. This example is illustrative of two results. First, Markov chains may have semiconvergent orderings that do not satisfy Property-R as in \mathbf{Q}_{marca} . Second, semiconvergent orderings that do not satisfy Property-R may very well have the same (if not smaller) value for the magnitude of the subdominant eigenvalue of the GS iteration matrix as a high-stepping ordering. On the other hand, \mathbf{Q}_{lex} does not give a semiconvergent ordering.

To recapitulate, it is possible to check whether the GS iteration for a given ordering satisfies Property-R by computing \mathbf{q} from (3.3). The fundamental matrix \mathbf{Z} of the GS iteration need not be explicitly computed. Assuming \mathbf{Q} has an average of r nonzero off-diagonal elements (uniformly and independently distributed) across each

row and each column, there are $nr/2$ nonzero elements in \mathbf{L} . Therefore, computing \mathbf{q} requires $nr/2$ floating-point additions and n floating-point divisions. If a given ordering does not satisfy Property-R, one may seek a guaranteed-to-converge ordering. The ordering algorithm for Property-R emerges from the following observation.

DEFINITION 3.4. *Any irreducible Markov chain may be symmetrically permuted to a form called the block normal form for Property-R. A matrix in this form is a block $(N \times N)$ lower Hessenberg matrix with square diagonal blocks of order n_i , where $n_N = 1$ and $\sum_{i=1}^N n_i = n$. State n can be any state in the original ordering. Furthermore, each row of the superdiagonal blocks has at least one nonzero element.*

A matrix \mathbf{Q} in block normal form for Property-R is therefore given by

$$(3.4) \quad \mathbf{Q}_{n \times n} = \begin{pmatrix} n_1 & n_2 & n_3 & \cdots & 1 \\ \mathbf{Q}_{1,1} & \mathbf{Q}_{1,2} & & & \\ \mathbf{Q}_{2,1} & \mathbf{Q}_{2,2} & \mathbf{Q}_{2,3} & & \\ \vdots & \vdots & \ddots & \ddots & \\ \mathbf{Q}_{N-1,1} & \mathbf{Q}_{N-1,2} & \mathbf{Q}_{N-1,3} & \cdots & \mathbf{Q}_{N-1,N} \\ \mathbf{Q}_{N,1} & \mathbf{Q}_{N,2} & \mathbf{Q}_{N,3} & \cdots & q_{n,n} \end{pmatrix} \begin{matrix} n_1 \\ n_2 \\ \vdots \\ n_{N-1} \\ 1 \end{matrix} .$$

Note that $\forall j < N$, $\mathbf{Q}_{N,j}$ are row vectors. Similarly, $\mathbf{Q}_{N-1,N}$ is a column vector. Each state of block i in the state space partition, where $1 \leq i < N$, is $N - i$ steps away from state n .

The form in (3.4) suggests that the Cuthill–McKee algorithm (see [1, p. 162], or [2, p. 153–157]) for reordering a sparse matrix with arbitrary resolution of ties, rather than the original tie-breaking rule (see [7, pp. 69–71]), may be used to arrange the states into levels of distances from any particular state. The ordering algorithm selects a final state, which becomes state n , and at step k marks the states that are k steps away from state n by placing them into the $(N - k)$ th block of states in (3.4). The ordering of states within each block is immaterial, hence arbitrary resolution of ties occurs in the Cuthill–McKee algorithm. The space requirement of the algorithm is an extra $O(n)$ integer locations other than the space taken by the matrix, whereas its time complexity is $O(nr)$. However, no floating-point operations are required; only conditional statements and assignments are performed.

Remark 3.5. Sparse matrices possess orderings that do not satisfy Property-R. The proof follows from the observation that it is possible to permute a zero off-diagonal element in \mathbf{Q} , say $q_{i,j}$ to row $n - 1$, column n in (3.4), thereby preventing state i from reaching state j using high-stepping transitions only.

4. Conclusion. This paper investigates the effects of high-stepping orderings on the GS iteration matrix in Markov chains. High-stepping orderings are those for which the coefficient matrix satisfies Property-R. Each sparse irreducible Markov chain has at least one ordering that is not high-stepping, whereas its number of high-stepping orderings depends on the nonzero structure of the chain. A high-stepping ordering of a Markov chain is semiconvergent. Orderings that are not high-stepping may or may not be semiconvergent. We have described a simple approach to check a given ordering for Property-R. If a given ordering of an irreducible Markov chain does not satisfy the property, it is possible to permute the chain using a version of the Cuthill–McKee algorithm so that convergence is ensured. However, high-stepping orderings in general are not superior to semiconvergent non-high-stepping orderings, and it is not clear how much will be gained by employing high-stepping orderings of large sparse irreducible Markov chains.

Acknowledgments. The author thanks the anonymous referees for their constructive reports and for pointing out the Cuthill–McKee algorithm.

REFERENCES

- [1] E. CUTHILL, *Several strategies for reducing the bandwidth of matrices*, in *Sparse Matrices and Their Applications*, D. J. Rose and R. A. Willoughby, eds., Plenum Press, New York, 1972, pp. 157–166
- [2] I. S. DUFF, A. M. ERISMAN, AND J. K. REID, *Direct Methods for Sparse Matrices*, Clarendon Press, Oxford, UK, 1986.
- [3] J. R. GILBERT, *Predicting structure in sparse matrix computations*, *SIAM J. Matrix Anal. Appl.*, 15 (1994), pp. 62–79.
- [4] L. KAUFMAN, *Matrix methods for queueing problems*, *SIAM J. Sci. Statist. Comput.*, 4 (1983), pp. 525–552.
- [5] D. MITRA AND P. TSOUKAS, *Relaxations for the numerical solutions of some stochastic problems*, *Comm. Statistics Stochastic Models*, 4 (1988), pp. 387–419.
- [6] W. J. STEWART, *Introduction to the Numerical Solution of Markov Chains*, Princeton University Press, Princeton, NJ, 1994.
- [7] R. P. TEWARSON, *Sparse Matrices*, Academic Press, New York, 1973.