University of Maryland College Park

Institute for Advanced Computer Studies TR–92–23
Department of Computer Science TR–2847

Gaussian Elimination, Perturbation Theory and Markov Chains

G. W. Stewart[†]

ABSTRACT

The purpose of this paper is to describe the special problems that emerge when Gaussian elimination is used to determinin the steady-state vector of a Markov chain.

*This report is available by anonymous ftp from `thales.cs.umd.edu` in the directory `pub/reports`. It will appear in the Proceedings of the IMA Workshop on Linear Algebra, Markov Chains, and Queuing Models.

[†]Department of Computer Science and Institute for Advanced Computer Studies, University of Maryland, College Park, MD 20742. This work was supported in part by the Air Force Office of Scientific Research under Contract AFOSR-87-0188 and was done while the author was a visiting faculty member at the Institute for Mathematics and Its Applications, The University of Minnesota, Minneapolis, MN 55455.

Gaussian Elimination, Perturbation Theory, and Markov Chains

G. W. Stewart

1. Introduction

The purpose of this paper is to describe the special problems that emerge when Gaussian elimination is used to determinine the steady-state vector of a Markov chain. Although there are many iterative techniques for solving this problem, direct methods are appropriate when the problem is small or when it is sparse and unstructured. In such cases, Gaussian elimination, the simplest of the direct methods, is a natural candidate.

The analysis of direct methods for linear systems has traditionally combined rounding-error analysis with perturbation theory — the former to establish the stability of the algorithm in question and the latter to assess the accuracy of the solution. This fruitful interplay carries over to the solution of Markov chains, and will be one of the main themes of this paper.

The paper begins with a review of the basic facts about Gaussian elimination for general linear systems. We then turn to the application of the algorithm to general Markov chains. The theory here is quite satisfactory and justifies the algorithm for many, if not most, applications. However, there is an important class of chains for which the algorithm fails — the nearly completely decomposable (NCD) chains. Paradoxically, although the algorithm fails, the problem is well determined in the sense that all the information necessary to compute a solution is available at the outset of the computations. A closer study of this paradox motivates a variant of Gaussian elimination proposed by Grassmann, Taksar, and Heyman [6].

Throughout out this paper we will assume that the reader is familiar with Gaussian elimination and its relation to the LU decomposition. Treatments of this material can be found in most introductory books on numerical linear algebra (e.g., [5, 12, 17]). The symbol $\|\cdot\|$ denote a family of consistent matrix norms; i.e., one for which $\|AB\| \le \|A\|\|B\|$, whenever the product AB is defined.

2. Gaussian Elimination and Linear Systems

The need for a formal rounding-error analysis of Gaussian elimination became critical as the arrival of the electronic computer made it impossible for a human to monitor the individual numbers in a calculation. In an early analysis, the statistician Hotelling [8] arrived at the conclusion that the errors in Gaussian elimination could grow exponentially with the size of the matrix, which would have precluded its use for matrices of even modest size. A subsequent analysis by von Neumann and Goldstine [16] showed that the algorithm would solve positive definite systems accurately, provided they were what we now call well conditioned. Finally, in 1961 Wilkinson [18] provided a comprehensive error analysis of Gaussian elimination for general linear systems.

A key component of Wilkinson's treatment is the projection of the errors back onto the original problem, a procedure known as backward rounding-error analysis. Specifically, he showed that if Gaussian elimination is used to solve the system

$$Ax = b,$$

where A is of order n then the computed solution $\tilde{x}$ satisfies the equation

$$(A + E)\tilde{x} = b, \tag{2.1}$$

where

$$\|E\| \leq \varphi(n)\gamma\epsilon_{\mathrm{M}}. \tag{2.2}$$

Here $\varphi(n)$ is a slowly growing function of the size of the matrix that depends on the norm and details of the arithmetic used in the computations, and ϵ_{M} is the rounding unit for the arithmetic. The number γ is the largest matrix element encountered in the course of the elimination.

The analysis shows that provided γ is not large compared with A—i.e., there has been no undue growth of elements in the elimination—then the computed solution is the exact solution of a problem differing from the original by terms of order of magnitude of the rounding error. Since such errors are generally smaller than errors already present in the elements of A, the algorithm itself cannot be held responsible for inaccuracies in the solution. An algorithm with such a backward error analysis is said to be *stable*. By the above error analysis, partial pivoting (the practice of interchanging rows to bring the largest element in the column into the diagonal) is seen to be a device to ensure stability by limiting the growth of elements as reflected by γ.

Although stability is a compelling reason for using an algorithm, it is not sufficient for users who want to know the accuracy of their solutions. Consequently, it is customary to supplement a rounding-error analysis with a perturbation theory for the problem being solved. The basic perturbation theory for linear systems is particularly simple. If $Ax = b$ and $\tilde{x}$ satisfies (2.1), then

$$\tilde{x} = x - A^{-1}Ex + O(\|E\|^2).$$

It follows that

$$\frac{\|\tilde{x} - x\|}{\|x\|} \lesssim \kappa(A)\frac{\|E\|}{\|A\|}, \tag{2.3}$$

where $\kappa(A) = \|A\|\|A^{-1}\|$ is the *condition number* of A. The left-hand side of (2.3) is a relative error in the computed solution. The fraction on the right is the relative error in A due to the perturbation E. The number $\kappa(A)$, which is always greater than one, is a magnifying factor that says how much the error in A is magnified as it passes to an error in x.

Together the rounding-error analysis and the perturbation analysis form a neat summary of the properties of Gaussian elimination. The rounding error analysis shows that any inaccuracies in the solution must come from exceedingly small changes in the problem: the perturbation analysis gives us a means of assessing the errors in the solution due to these changes. We will now turn to a similar analysis of the use of Gaussian elimination to solve Markov chains.

3. Gaussian Elimination and Markov Chains

Now let P be the transition matrix of an irreducible Markov chain. Then up to normalization factors, P has unique, positive right and left eigenvectors corresponding to the eigenvector one. The right eigenvector is $\mathbf{e}$; i.e., the vector whose components are all one. Our problem is to compute the left eigenvector, which we will denote by y^{T}.

The problem can be cast in a form that is more convenient for Gaussian elimination. Let

$$Q = I - P.$$

Then

$$y^{\mathrm{T}}Q = y^{\mathrm{T}} - y^{\mathrm{T}}P = y^{\mathrm{T}} - y^{\mathrm{T}} = 0.$$

Thus the problem has been transformed from that of finding an eigenvector of P to that of finding a null vector of Q.[1] Note that $Q\mathbf{e} = 0$; i.e., the row sums of Q are zero.

[1]It is worth noting that this kind of reduction cannot be applied to the general eigenvalue problem, since we will not ordinarily know the eigenvalue a priori.

Gaussian elimination can be used to solve the equation $y^{\mathrm{T}}Q = 0$ as follows.

1. Use Gaussian elimination to decompose $Q^{\mathrm{T}} = LU$, where L is unit lower triangular and U is upper triangular.
2. Partition
$$U = \begin{pmatrix} U_* & u \\ 0 & 0 \end{pmatrix}.$$
The matrix U_* will be upper triangular and nonsingular.
3. Compute
$$y^{\mathrm{T}} = \frac{(-u^{\mathrm{T}}U_*^{-\mathrm{T}} \;\; 1)}{\|(-u^{\mathrm{T}}U_*^{-\mathrm{T}} \;\; 1)\|_1},$$
where $\|x\|_1 = \mathbf{e}^{\mathrm{T}}|x|$.

The third step of the algorithm is almost self-explanatory. Since L is nonsingular, any null vector of Q^{T} is a null vector of U and vice versa. The third step then amounts to computing a null vector of U by assuming its last component is one, solving the resulting triangular system, and normalizing.

The nonsingularity of U_*, which is not obvious, is closely bound up with the problem of pivoting. To see what is going on let us consider the first two columns of Q^{T}, which we write in the form

$$\begin{matrix} q_{11} & q_{12} \\ q_{21} & q_{22} \\ q_1 & q_2 \end{matrix}$$

The quantities q_{11} and q_{22} are nonnegative, while q_{21}, q_{12}, q_1, and q_2 are nonpositive. Moreover, since the components of the first column sum to zero, q_{11} can be zero only if q_{21} and q_2 are zero, in which case Q is reducible, contrary to assumption.

Thus the first step of Gaussian elimination can be performed on the second column, yielding a new column of the form

$$\begin{aligned} \bar{q}_{12} &= 0 \\ \bar{q}_{22} &= q_{22} - \frac{q_{21}q_{12}}{q_{11}} \\ \bar{q}_2 &= q_2 - \frac{q_1 q_{12}}{q_{11}} \end{aligned}$$

By considering the signs of the quantities involved, we find that

$$\begin{aligned} &1. \quad q_{22} \ge \bar{q}_{22}, \\ &2. \quad q_2 \ge \bar{q}_2. \end{aligned}$$

In other words, the diagonal element decreases while the off-diagonal elements increase in magnitude. Since the column sums are still zero, we have $q_{22} \ge \bar{q}_{22} \ge -\bar{q}_{i2}$ $(i > 2)$, which implies that all the elements in the reduced second column are bounded in magnitude by q_{22}; i.e., there is no net growth in the elements of the reduced column. Finally, the quantity $\bar{q}_{22}$ can be zero only if both q_1 and q_2 are zero, which contradicts irreducibility.

Since any column of Q^{T} can be symmetrically permuted into the second column, the above observations apply to any column. Thus the result of the first step of Gaussian elimination is an irreducible matrix with positive diagonal elements and nonpositive off-diagonal elements. By induction the same is true of the the subsequent steps, save the last, which must produce the single number $u_{nn} = 0$. Since there is no net growth in the elements, the reduction can be carried to completion, and the diagonal elements of U_*, which are the pivots in the elimination are positive. Since the growth factor γ in (2.2) is one, the algorithm is stable. This stability was first pointed out by Funderlic and Mankin [4].

Two further points. not only is there no need to pivot in the algorithm, but the usual form of partial pivoting is in some sense harmful. The reason is that interchanging two rows of Q^{T} (without interchanging the corresponding columns) destroys the properties that keep growth from occurring, and we are left with the (admittedly unlikely) possibility of instability. The second point is that symmetric pivoting, in which two rows and the same two columns are interchanged, does preserve the structure of Q and can be used with complete freedom. This fact is important in applications involving sparse matrices, in which pivoting is necessary to avoid fill-in [2].

Turning now to the perturbation theory of Markov chains, the basic theory goes back almost a quarter of a century and has been presented in a variety of ways [3, 7, 9, 10]. Here we give it in a form that will be useful in the sequel. For proofs see [14].

It can be shown that the matrix P has the spectral decomposition

$$P = \mathbf{e}y^{\mathrm{T}} + XBY^{\mathrm{T}}, \tag{3.1}$$

where

$$\begin{pmatrix} y^{\mathrm{T}} \\ Y^{\mathrm{T}} \end{pmatrix} = (\mathbf{e}\ X)^{-1}.$$

Alternatively,

$$(\mathbf{e}\ X)^{-1}P(\mathbf{e}\ X) = \mathrm{diag}(1, B);$$

that is, $(\mathbf{e}\ X)$ transforms P via a similarity transformation into a block-diagonal matrix. It follows that the eigenvalues of B must be those of P other than one, and hence $I - B$ is nonsingular.

Now assume that $\tilde{P} = P + F$ is an irreducible stochastic matrix, and let $\tilde{y}^{\mathrm{T}}$ be its steady-state vector. Then it can be shown that

$$\tilde{y}^{\mathrm{T}} \cong y^{\mathrm{T}} + y^{\mathrm{T}} F X (I - B)^{-1} Y^{\mathrm{T}}, \tag{3.2}$$

from which it follows that

$$\frac{\|\tilde{y}^{\mathrm{T}} - y^{\mathrm{T}}\|}{\|y^{\mathrm{T}}\|} \lesssim \|X(I - B)^{-1}Y^{\mathrm{T}}\| \|F\|. \tag{3.3}$$

This is the desired perturbation bound.

The matrix $(I - P)^{\#} \equiv X(I - B)^{-1}Y^{\mathrm{T}}$ is called the *Drazin pseudo-inverse* or the *group inverse* of the system. The bound (3.3) shows that the norm of the group inverse is a condition number for the steady-state vector of a Markov chain.

4. Nearly Completely Decomposable Chains

The results of the last section place Gaussian elimination for the solution of Markov chains on a par with Gaussian elimination for the solution of linear systems. The algorithms are backwards stable, and the problems have reasonable perturbation theories. However, in neither case does the analysis apply to matrices whose elements vary widely and systematically in size.

Consider, for example the following matrix:

$$Q = \begin{pmatrix} +0.75287 & -0.75283 & -0.00003 \\ -0.75283 & +0.75284 & -0.00001 \\ -0.00003 & -0.00001 & +0.00004 \end{pmatrix}.$$

Since the matrix is symmetric, the left and right eigenvector are both $\mathbf{e}$, and there is no difference between applying Gaussian elimination to Q or Q^{T}.

Working in five decimal digits, we compute the correctly rounded multipliers for the first step of Gaussian elimination as

$$0.99996 = \mathrm{fl}(-0.75283/0.75287) \quad \text{and} \quad 0.39849 \cdot 10^{-4} = \mathrm{fl}(-0.00003/0.75287).$$

(Here fl denotes a floating point operation.) If we use these multipliers and make no further rounding errors, the matrix assumes the form.

$$Q = \begin{pmatrix} +0.75287 & -0.75283 & -0.00003 \\ +0.0 & +0.000040113 & -0.000039999 \\ +0.0 & -0.000039999 & +0.000039999 \end{pmatrix} \tag{4.1}$$

This shows that the last two components of the computed null vector, which should be equal, will be in a ratio of 40113/39999. Thus, we have only two figures of accuracy.

Since Gaussian elimination is stable, the only way we can get an inaccurate answer is for the problem to be ill-conditioned. And indeed it is. For the problem is an example of a nearly completely decomposable (NCD) chain; that is, one which, after a suitable reordering of the states, is almost block diagonal. For the case of three blocks, such a chain has the form

$$P = \begin{pmatrix} P_{11} & E_{12} & E_{13} \\ E_{21} & P_{22} & E_{23} \\ E_{31} & E_{32} & P_{33} \end{pmatrix},$$

where the matrices E_{ij} are small. Such chains were introduced by Simon and Ando [11], and have been studied extensively since (e.g., see [1, 13]).

Since the E_{ij} are small, each of the matrices P_{ii} has an eigenvalue near one. Consequently the entire matrix, in addition to an eigenvalue of one, has $k-1$ eigenvalues near one, where k is the number of blocks. Consequently, the matrix B in (3.1) has $k-1$ eigenvalues near one and the condition number $\|(I-P)^{\#}\| = \|X(I-B)^{-1}Y^{\mathrm{T}}\|$ will be large.

In spite of this ill-conditioning, the problem can be solved by the following aggregation-disaggregation technique (here ϵ is the norm of the matrix consisting of the off-diagonal blocks):

1. Compute $\hat{y}_i^{\mathrm{T}}$, the Perron eigenvector[2] of P_{ii} normalized so that $\hat{y}_i^{\mathrm{T}}\mathbf{e} = 1$.
2. Compute $\epsilon_{ij} = \hat{y}_i^{\mathrm{T}} E_{ij}\mathbf{e}$ and $\pi_{ii} = \hat{y}_i^{\mathrm{T}} A_{ii}\mathbf{e}$.
3. The *coupling matrix*

$$C = \begin{pmatrix} \pi_{11} & -\epsilon_{12} & -\epsilon_{13} \\ -\epsilon_{21} & \pi_{22} & -\epsilon_{23} \\ -\epsilon_{31} & -\epsilon_{32} & \pi_{33} \end{pmatrix}$$

[2]The Perron eigenvector of an irreducible nonnegative matrix is the positive eigenvector corresponding to the largest positive eigenvalue.

is easily seen to be an irreducible stochastic matrix. Compute its steady-state vector (ν_1, ν_2, ν_3).

4. Then
$$y^{\mathrm{T}} = \left(\nu_1 \hat{y}_1^{\mathrm{T}}, \nu_2 \hat{y}_2^{\mathrm{T}}, \nu_3 \hat{y}_3^{\mathrm{T}}\right) + O(\epsilon).$$

The solution provided by the algorithm has two components: the vectors $\hat{y}_i^{\mathrm{T}}$ and the *coupling coefficients* ν_i. Provided the diagonal blocks A_{ii} are well behaved, the former will be insensitive to perturbations in P. On the other hand, unless the perturbation in P is *small compared to the* E_{ij}, the elements of C and hence the coupling coefficients will be poorly determined. Since the E_{ij} are small compared to the P_{ii}, small relative perturbations in P can be large compared to the E_{ij} and harm the solution.

All this agrees with the perturbation theory for NCD Markov chains. It can be shown [14] that under suitable regularity conditions the matrix P has a spectral decomposition of the form

$$P = \mathbf{1}y^{\mathrm{T}} + X_{\mathrm{s}} B_{\mathrm{s}} Y_{\mathrm{s}}^{\mathrm{T}} + X_{\mathrm{f}} B_{\mathrm{f}} Y_{\mathrm{f}}^{\mathrm{T}}, \tag{4.2}$$

where

$$\begin{pmatrix} y^{\mathrm{T}} \\ Y_{\mathrm{s}}^{\mathrm{T}} \\ Y_{\mathrm{f}}^{\mathrm{T}} \end{pmatrix} = \begin{pmatrix} \mathbf{1} & X_{\mathrm{s}} & X_{\mathrm{f}} \end{pmatrix}^{-1}.$$

The matrix B_{s} is a perturbation of the identity, while the matrix B_{f} has eigenvalues whose eigenvalues are bounded in magnitude away from one.[3] In analogy with (3.2) the perturbed steady-state vector, due to a perturbation F in P, can be written

$$\tilde{y} \cong y^{\mathrm{T}} + X_{\mathrm{s}} F (I - B_{\mathrm{s}})^{-1} Y_{\mathrm{s}}^{\mathrm{T}} + X_{\mathrm{f}} F (I - B_{\mathrm{f}})^{-1} Y_{\mathrm{f}}^{\mathrm{T}}. \tag{4.3}$$

Since the eigenvalues of B_{s} are near one while those of B_{f} are not, the second term in (4.3) will dominate; i.e., the perturbations will tend to lie in the space

[3]The subscripts "s" and "f", which stand for slow and fast, have the following origin. It is easy to see that

$$P^i = \mathbf{1}y^{\mathrm{T}} + X_{\mathrm{s}} B_{\mathrm{s}}^i Y_{\mathrm{s}}^{\mathrm{T}} + X_{\mathrm{f}} B_{\mathrm{f}}^i Y_{\mathrm{f}}^{\mathrm{T}}.$$

Since the eigenvalues of B_{s} are near one, B_{s}^i approaches zero more slowly that the ith power of B_{f}, whose eigenvalues are smaller. Consequently the decomposition (4.2) exhibits two transient behaviors: a fast transient associated with B_{f} and a slow transient associated with B_{s}. This behavior was noted by Simon and Ando.

spanned by the rows of Y_s^T. However, it can be shown that the row space of Y_s^T is essentially the same as the row space of

$$\begin{pmatrix} \hat{y}_1^T & 0 & 0 \\ 0 & \hat{y}_2^T & 0 \\ 0 & 0 & \hat{y}_3^T \end{pmatrix}.$$

Consequently, the components of $\tilde{y}^T$ will tend to lie, as they should, along the directions of the vector $\hat{y}_i^T$; however, their relative proportions, which correspond to the coupling coefficients, will change. Thus, it is the coupling coefficients that are sensitive to changes in P, a fact which agrees with our comments on the aggregation algorithm.

5. The GTH Algorithm

The analysis of the preceding section shows that unless we know the elements of the E_{ij} to high relative accuracy, the steady-state vector of the chain will be ill-determined. In situations where the E_{ij} must be determined empirically, this accuracy may be difficult to achieve, since their elements correspond to events that occur only infrequently. On the other hand, in parameter studies, where the behavior of a system is being modeled by a Markov chain, the E_{ij} can be taken as fully accurate. Thus it is reasonable to pose the following problem: How do we compute the steady-state vector of a NCD chain when the E_{ij} are known to high accuracy?

The obvious answer is to use the aggregation algorithm. However, this answer begs the question; for the coupling matrix C is itself nearly completely decomposable and hence ill-conditioned. For example, suppose that

$$C = \begin{pmatrix} 0.9999 & 0.1499\text{e}{-3} \\ 0.2499\text{e}{-3} & 0.9998 \end{pmatrix}.$$

Note that to four decimal digits, C is a correctly rounded stochastic matrix. However,

$$I - C = 10^{-3}\begin{pmatrix} 0.1000 & -0.1499 \\ -0.2499 & 0.2000 \end{pmatrix} \tag{5.1}$$

has the positive eigenvector

$$y = (0.4968\ 0.5032),$$

whereas the corresponding eigenvector of the exactly stochastic matrix

$$C_{\text{true}} = \begin{pmatrix} 0.9998501 & 0.1499\text{e}{-3} \\ 0.2499\text{e}{-3} & 0.9997501 \end{pmatrix}.$$

is

$$y_{\text{true}} = (0.5000\ 0.5000).$$

Thus the aggregation algorithm will produce coupling coefficients that are inaccurate in the third figure.

Looking carefully at this example, we see that the problem lies with the diagonal elements of $I - C$, which are inaccurate. However, since the rows of $I - C$ must sum to zero, we can restore the accuracy by replacing $I - C$ with

$$10^{-3} \begin{pmatrix} 0.1499 & -0.1499 \\ -0.2499 & 0.2499 \end{pmatrix}.$$

In the general aggregation algorithm, this amounts to computing the coupling matrix in the form

$$I - C = \begin{pmatrix} \epsilon_{12} + \epsilon_{13} & -\epsilon_{12} & -\epsilon_{13} \\ -\epsilon_{21} & \epsilon_{21} + \epsilon_{23} & -\epsilon_{23} \\ -\epsilon_{31} & -\epsilon_{32} & \epsilon_{31} + \epsilon_{32} \end{pmatrix}.$$

This procedure of adjusting the diagonals restores the figures that could not be represented in the diagonals of the original matrix C.

The idea of diagonal adjustment can be applied to Gaussian elimination. For example, the trailing 2×2 principal submatrix of (4.1) should have zero column sums. Since it does not, we force them to be by replacing the diagonals with the sum of the off diagonals, to get the matrix

$$Q = \begin{pmatrix} +0.75287 & -0.75283 & -0.00003 \\ +0.0 & +0.000039999 & -0.000039999 \\ +0.0 & -0.000039999 & +0.000039999 \end{pmatrix},$$

which gives the right answer.

In general, after k steps of Gaussian elimination applied to Q^{T}, the matrix assumes the form

$$\begin{pmatrix} U_{11}^{(k)} & U_{12}^{(k)} \\ 0 & U_{22}^{(k)} \end{pmatrix},$$

where $U_{11}^{(k)}$ is of order k. Since the row sums of $U_{22}^{(k)}$ are known to be zero, instead of using Gaussian elimination to compute its diagonal elements, we use the alternative formula

$$u_{jj}^{(k)} = -\sum_{\substack{i=k+1 \\ i \neq j}}^{i=n} u_{ij}, \qquad j = k+1, \ldots, n. \tag{5.2}$$

This, in essence, is the algorithm of Grassmann, Taksar, and Heyman [6] mentioned in the introduction. There are three points to be made about it.

In the first place, it is easy to implement and not very expensive. The sums (5.2) can be accumulated as the off-diagonal elements are generated during the elimination, which increases the work done in the inner loop by a single addition.[4]

Second, the numerical properties of the algorithm are not well understood. A possible justification is that the method never subtracts and hence cancellation cannot cause it to fail. Unfortunately, this line of reasoning, when formalized, is essentially the analysis of Hotelling mentioned in the introduction, and it leads to the same pessimistic conclusions. With G. Zhang, I have given a rounding-error analysis of a closely related algorithm [15]; however, the original algorithm remains unanalyzed. Nontheless, I believe that the GTH algorithm is stable and should be used routinely in the direct solution of Markov chains.

Third, the above discussion is a little unfair to Gaussian elimination, which is frequently asked to do the impossible: solve a problem that is not in the computer. The matrix (5.1) is a case in point. Owing to initial rounding errors, it is only approximately singular and Gaussian elimination will do a very good job of computing an approximate null vector. The fact that this vector is not what we want cannot be blamed on Gaussian elimination, which has no way of knowing it is dealing with a Markov chain. In this light, the GTH algorithm is seen to be an augmentation of Gaussian elimination that does know.[5]

References

[1] P.-J. Courtois. *Decomposability.* Academic Press, New York, 1977.

[4]This assumes that the exterior product form of Gaussian elimination is used. If inner-product forms of the kind associated with the names of Crout and Doolittle are used, the algorithm is a little more complicated, though the amount of extra work stays the same.

[5]Some of the examples appearing in the literature purporting to show that Gaussian elimination fails are of the same nature: the damage is done in the initial rounding of problem.

[2] I. S. Duff, A. M. Erisman, and J. K. Reid. *Direct Methods for Sparse Matrices.* Clarendon Press, Oxford, 1986.

[3] R. E. Funderlic. Sensitivity of the stationary distribution vector for an ergodic Markov chain. *Linear Algebra and Its Applications*, 76:1–17, 1986.

[4] R. E. Funderlic and J. B. Mankin. Solution of homogeneous systems of linear equations arising from compartmental models. *SIAM Journal on Scientific and Statistical Computing*, 2:375–109, 1981.

[5] G. H. Golub and C. F. Van Loan. *Matrix Computations.* Johns Hopkins University Press, Baltimore, Maryland, 2nd edition, 1989.

[6] W. K. Grassmann, M. I. Taksar, and D. P. Heyman. Regenerative analysis and steady state distributions. *Operations Research*, 33:1107–1116, 1985.

[7] M. Haviv and L. van der Heyden. Perturbation bounds for the stationary probabilities of a finite Markov chain. *Advances in Applied Probability*, 16:804–818, 1984.

[8] H. Hotelling. Some new methods in matrix calculations. *Annals of Mathematical Statistics*, 14:1–34, 1943.

[9] C. D. Meyer. The condition of a Markov chain and perturbations bounds for the limiting probabilities. *SIAM Journal on Algebraic and Discrete Methods*, 1:273–283, 1980.

[10] P. J. Schweitzer. Perturbation theory and finite Markov chains. *Journal of Applied Probability*, 5:401–413, 1968.

[11] H. A. Simon and A. Ando. Aggregation of variables in dynamic systems. *Econometrica*, 29:111–138, 1961.

[12] G. W. Stewart. *Introduction to Matrix Computations.* Academic Press, New York, 1973.

[13] G. W. Stewart. On the structure of nearly uncoupled Markov chains. In G. Iazeolla, P. J. Courtois, and A. Hordijk, editors, *Mathematical Computer Performance and Reliability*, pages 287–302, Elsevier, New York, 1984.

[14] G. W. Stewart. Perturbation theory for nearly uncoupled Markov chains. In W. J. Stewart, editor, *Numerical Methods for Markov Chains*, pages 105–120, North Holland, Amsterdam, 1990.

[15] G. W. Stewart and G. Zhang. On a direct method for the solution of nearly uncoupled Markov chains. *Numerische Mathematik*, 59:1–11, 1991.

[16] J. von Neumann and H. H. Goldstine. Numerical inverting of matrices of high order. *Bulletin of the American Mathematical Society*, 53:1021–1099, 1947.

[17] D. S. Watkins. *Fundamentals of Matrix Computations*. John Wiley & Sons, New York, 1991.

[18] J. H. Wilkinson. Error analysis of direct methods of matrix inversion. *Journal of the ACM*, 8:281–330, 1961.