Finding the spectral radius of a large sparse non-negative matrix

R. J. Wood¹ M. J. O'Neill²

(Received 31 August 2006; revised 11 July 2007)

Abstract

A comparison of methods for finding the spectral radius of a large sparse non-negative matrix. The Arnoldi method is compared with a variation of the method of Collatz [Math Zeit, 48, 221–6, 1948], this method of Collatz being always convergent when finding the spectral radius of a non-negative matrix. The advantages and disadvantages of both methods are discussed, as well as a comparison with the methods of orthogonal iteration and simultaneous iteration. Comparisons are made using flop counts.

Contents

1 Introduction C331

See http://anziamj.austms.org.au/ojs/index.php/ANZIAMJ/article/view/117 for this article, © Austral. Mathematical Soc. 2007. Published August 1, 2007. ISSN 1446-8735

2 An always convergent method
C332
3 The Arnoldi method
C335
4 Other methods
C337
5 Comparison of the methods
C338

C331

C339

7 Conclusion C341

Practicalities for the method of Collatz

References C342

1 Introduction

Introduction

1

A non-negative matrix is a matrix in which all the elements are non-negative. Computation of the spectral radius of a non-negative matrix is important in showing the uniqueness of solution of a linear system which arises in Input-Output Analysis, a branch of Mathematical Economics [18]. It can also prove useful in the identification of an M-matrix [20]. Furthermore, computation of the eigenvalues and eigenvectors of a non-negative matrix is important in areas such as genetics, age-specific population growth and harvesting animal populations and in determining page rank for web pages on the Internet. The Power method for finding the dominant eigenvector of a matrix is used by Google in determining the page rank of a web page [1, 6]. Calculating the spectral radius of a matrix is also important in graph theory [9, 13]. The accepted method for finding all the eigenvalues of a full matrix is the QR method. However, if it is desired to find just the dominant eigenvalue of a large sparse, non-negative matrix there is an element of overkill in using the QR method. Furthermore, the requirement of the QR method that the subject matrix must first be converted to Hessenberg form, which is an $O(n^3)$ operation, may be prohibitively expensive and also will very likely destroy the sparsity of that matrix. A more promising approach is to use a method which calculates only the dominant, or the first few largest eigenvalues of the matrix, and a method which at the same time preserves the sparsity.

One such approach is to use an invariant sub-space method such as that of Arnoldi [2], Saad [12] and Stewart [16]. Saad [11] showed that the method could be used as a projection method to solve large eigenproblems and provided a number of variations of the method. Later more sophisticated versions of the method have been developed. Another approach is to develop further a result of Collatz [4] for bounding the spectral radius of an irreducible non-negative matrix. Various researchers contributed to the problem of bounding the spectral radius; among these are Frobenius [5], Ledermann [8], Ostrowski [10], and Brauer [3]. Wood and O'Neill [19] showed that the result of Collatz could be extended to provide converging bounds of uncertainty for the spectral radius of an irreducible non-negative matrix. Both of these methods will be discussed and a further comparison made with the methods of Orthogonal Iteration and Simultaneous Iteration.

2 An always convergent method

In 1942 Collatz [4] proved the equivalent of the following result.

Theorem 1 Let $A \ge 0$ be an $n \times n$ irreducible matrix and q_0 be an arbitrary positive n-dimensional vector. Defining $q_{\nu} = Aq_{\nu-1} = \cdots = A^{\nu}q_0$, $\nu \ge 1$, let

$$\underline{\lambda}_{\nu} = \min_{1 \leq i \leq n} \left(\frac{q_{\nu+1}^{(i)}}{q_{\nu}^{(i)}} \right) \quad and \quad \overline{\lambda}_{\nu} = \max_{1 \leq i \leq n} \left(\frac{q_{\nu+1}^{(i)}}{q_{\nu}^{(i)}} \right),$$

where the superscript i represents the ith component of a vector. Then, denoting the spectral radius of A by $\rho(A)$,

$$\underline{\lambda}_0 \le \underline{\lambda}_1 \le \underline{\lambda}_2 \le \dots \le \rho(A) \le \dots \le \overline{\lambda}_2 \le \overline{\lambda}_1 \le \overline{\lambda}_0$$
.

Furthermore, if it is known that the matrix A is primitive then the following may be deduced.

Theorem 2 In Theorem 1 both the sequences $\{\underline{\lambda}_{\nu}\}_{\nu=0}^{\infty}$ and $\{\overline{\lambda}_{\nu}\}_{\nu=0}^{\infty}$ converge to $\rho(A)$, from an arbitrary initial positive vector q_0 , if and only if the irreducible matrix $A \geq 0$ is primitive [19].

A matrix A is irreducible if and only if its directed graph G(A) is strongly connected. An irreducible matrix is primitive if it has only one eigenvalue of modulus $\rho(A)$ [17].

Thus when matrix A is primitive, the interval of uncertainty $(\underline{\lambda}_{\nu}, \overline{\lambda}_{\nu})$ must eventually be reducing and converge to the spectral radius of A. That is, when A is primitive, the method of Collatz is an always convergent method.

The further restriction that the matrix A be primitive and not just irreducible does not present any practical difficulties, for the following reason: An irreducible matrix is either primitive or cyclic, and, if cyclic there must exist a permutation matrix P such that PAP^T has diagonal submatrices which are null and square [17, page 38]. If then an irreducible matrix A presents with diagonal elements which are entirely zero, a simple diagonal shift qI, where q > 0 and I is the identity matrix, will ensure that the matrix A + qI is primitive and the spectral radius of this changed matrix is easily converted to the spectral radius of A. The issue of this shift is discussed by Wood and O'Neill [19].

Convergence of the method of Collatz can sometimes be very slow but a way to overcome this and accelerate convergence is to use a variant of the Inverse Power Method. The following theorem guarantees that, if q is chosen appropriately, $(qI - A)^{-1}$ is non-negative and primitive.

Theorem 3 If $A \ge 0$ is an $n \times n$ irreducible matrix with $\rho(A) < q$, then $(qI - A)^{-1}$ is a non-negative irreducible matrix. Furthermore, it is primitive [19].

The convergence rate when the method of Collatz is applied to the matrix A is $|\lambda_2/\lambda_1|$ where λ_1 and λ_2 are the dominant and sub-dominant eigenvalues respectively. When the method of Collatz is applied to the matrix $(qI - A)^{-1}$ the convergence rate is $|(q - \lambda_1)/(q - \lambda_2)|$, and if q is chosen sufficiently close to λ_1 , this will be a superior rate.

A commonly used set of test matrices developed by Stewart for large, sparse matrix computation is a set of transition matrices for a particular Markov chain consisting of a random walk on a $(k+1)\times(k+1)$ triangular grid. These matrices are usually denoted Mark(k), and are described more fully by Saad [12]. Mark(25) has dimension n=351 and contains 123201 non-zero elements. From the definition of the Mark matrices it is easily shown that they are irreducible. However, they are also cyclic; so to ensure primitivity the method of Collatz has been applied to the matrix $\text{Mark}(25) + I_{351}$. See Table 1 for results.

The Mark matrices are row stochastic and thus have a dominant eigenvalue of one. Consequently $Mark(25) + I_{351}$ will have a dominant eigenvalue of two. When the method of Collatz for this matrix was implemented using Matlab it was found that, with a randomly generated initial vector q_0 , the number of flops to achieve four decimal places of accuracy in the dominant eigenvalue was 216, 216 flops and 54 matrix-vector multiplications. The convergence criterion used was two successive values $\rho_{\rm est}$ agreeing to four decimal places of accuracy.

3

$\begin{array}{c ccccccccccccccccccccccccccccccccccc$		TABL		e 1: Collatz	z bounds for Mark(2	$(5) + I_{351}$
2 1.3747 5.8263 3.6005 1.6005	ν	$\underline{\lambda}_{ u}$	ν	$\overline{\lambda}_{ u}$	$\rho_{\rm est} = (\underline{\lambda}_{\nu} + \overline{\lambda}_{\nu})/2$	Approx error
	1	1.0704	1	146.3969	73.7336	71.7336
3 1.4351 3.0207 2.2279 0.2279 	2	1.3747	2	5.8263	3.6005	1.6005
	3	1.4351	3	3.0207	2.2279	0.2279
	:	÷	:	÷	:	<u>:</u>
53 1.9937 2.0062 2.0000 0.0000	53	1.9937	53	2.0062	2.0000	0.0000
54 1.9937 2.0062 2.0000 0.0000	_54	1.9937	_54	2.0062	2.0000	0.0000

3 The Arnoldi method

The procedure introduced by Arnoldi [2], begins by building an orthogonal basis $\{v_1, v_2, \ldots, v_m\}$ for the Krylov subspace K_m , where

$$K_m = \text{span}\{v_1, Av_1, \dots, A^{m-1}v_1\}$$

and v_1 is an arbitrarily chosen vector of norm one. The vectors in this basis then form the successive columns of the matrix V_m as shown in Figure ??, the so-called Arnoldi factorisation, where H_m is an $m \times m$ upper Hessenberg matrix and $E = w_m e_m^H$ is a rank one matrix. See Saad [12] for further details. For m sufficiently large the eigenvalues of H_m provide approximations to a set of the eigenvalues of A. The basic Arnoldi algorithm was applied to the test matrix Mark(25), with the results shown in Table 2. An estimate of the error at each step was calculated using $\|(A - \lambda_i^{(m)}I)u_i^{(m)}\|_2 = h_{m+1,m}|e_m^H y_i^{(m)}|$ where $u_i^{(m)}$ is the Ritz approximate eigenvector for $\lambda_i^{(m)}$ and $y_i^{(m)}$ is an eigenvector of H_m associated with the eigenvalue $\lambda_i^{(m)}$.

The total number of flops required to guarantee the dominant eigenvalue correct to four decimal places of accuracy was 4.08×10^6 and 120 matrixvector multiplications. The convergence criterion used in this example was two successive values of $\rho(A)$ approx agreeing to four decimal places of accuracy. We acknowledge that, since both Arnoldi and the method of Collatz 3 The Arnoldi method C336

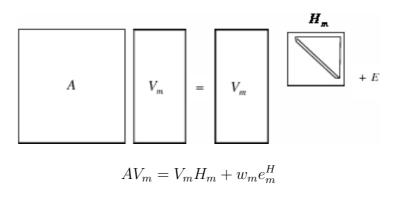


FIGURE 1: Arnoldi factorisation

were started with a random initial positive vector, the number of flops may vary depending on how close that initial vector is to the dominant eigenvector. To remove this effect, Arnoldi and the method of Collatz were repeated 100 times with different random initial vectors and the average number of flops calculated for each. The result was that, for Arnoldi, Mark(25) averaged 2.4×10^6 flops and, for Mark $(25) + I_{351}$, the method of Collatz averaged 1.1×10^5 flops. A further alternative method to find the dominant eigenvalue is to use the Matlab function eigs, which is an implementation of the Implicitly Restarted Arnoldi algorithm. When eigs was repeated 100 times for the matrix Mark(25) the average number of flops required to achieve four decimal accuracy in the dominant eigenvalue was approximately 3.54×10^7 flops $(1.4 \times 10^7 \text{ flops for Mark}(25) + I_{351})$. A further difficulty here was that sometimes eigs for Mark(25) converged to the eigenvalue -1 instead of 1. Also, when the QR method (applied using the Matlab function eig) was applied to Mark(25) in order to calculate all eigenvalues, the number of flops needed was approximately 4.2×10^8 (4.0×10^8 flops for Mark(25) + I_{351}).

Other methods C337

TAE	BLE 2:	Arnoldi	method applied	to $A = Mark(25)$
	m	flops	$\rho(A)$ approx.	Error
•	30	773100	0 9998	0.000759

\overline{m}	flops	$\rho(A)$ approx.	Error
30	773100	0.9998	0.000759
40	1311600	1.0000	0.00005
50	1990500	1.0000	4.23×10^{-7}

Table 3: Orthogonal and simultaneous iteration applied to A = Mark(25) + I_{351}

Method	Total no. of flops	Approx error in $\rho(A)$
Orthogonal	6.2×10^{6}	$< 10^{-4}$
Simultaneous	2×10^{6}	$< 10^{-4}$

Other methods 4

Other methods which are useful for large, sparse matrices are the methods of Orthogonal Iteration and Simultaneous Iteration. Orthogonal Iteration begins with an initial $n \times r$ matrix Q_0 , with orthonormal columns and generates $Z_1 = AQ_0$. The QR factorization of Z_1 , viz $Z_1 = Q_1R_1$, produces the next Q matrix, and the process is repeated to generate Q_2 , R_2 , Q_3 , R_3 , etc. Under certain conditions for convergence the diagonal entries of the R_k matrix will eventually approximate the r largest eigenvalues of A. Golub and VanLoan [7] gave details of the method in page 332 ff.

The method of Orthogonal Iteration can sometimes be very slow. Accordingly, the method of Simultaneous Iteration is designed to accelerate the convergence of Orthogonal Iteration by periodically performing a Schur decomposition. Stewart [15] described the technique. The two methods with r=1 were applied to Mark(25) + I_{351} . The results to achieve four decimal accuracy are shown in Table 3.

Larger values of r were tested, but these larger values of r resulted in the total number of flops increasing by a significant amount. Hence, if we

are only concerned with the dominant eigenvalue, increasing r is expensive in terms of the number of flops. The convergence criterion used in the above example was two successive values of $\rho(A)$ approximately agreeing to four decimal places of accuracy.

5 Comparison of the methods

For the test matrix $Mark(25) + I_{351}$ the method of Collatz recorded fewer flops than the Arnoldi method, eigs and eig for Mark(25) when finding the dominant eigenvalue. We acknowledge that the method of Arnoldi, and the Matlab routine eig, also provide some or all of the subdominant eigenvalues, and so more information than just the dominant eigenvalue. However, if the sole aim is to find the dominant eigenvalue of a non-negative matrix, then the method of Collatz would appear to be superior in terms of the number of flops.

For $\operatorname{Mark}(25) + I_{351}$, the methods of Orthogonal Iteration and Simultaneous Iteration require the order of 10 more flops than the method of Collatz. Furthermore for the method of Collatz the convergence criterion involving upper and lower bounds guarantees convergence. There are conditions under which Orthogonal Iteration can be guaranteed to converge to the dominant eigenvalue, but these are not easily verifiable a priori. For Simultaneous Iteration the condition $|\lambda_r| > |\lambda_{r+1}|$ does guarantee convergence for an irreducible matrix [15].

Because of its close relationship to the Power method, Arnoldi will typically generate a set of m eigenvalues among which will be the dominant eigenvalue, although convergence to this dominant eigenvalue is not guaranteed. A further matter of uncertainty with Arnoldi is the choice of m, the dimension of the Krylov subspace to be used. Choosing $m = 10, 11, 12, \ldots, 21$ was not effective for Mark(14), as Table 4 shows.

TABLE	4: A	Arnoldi ap	oplied to $A = M$	Iark(14)
	\overline{m}	flops	$\rho(A)$ approx.	•
	10	39640	0.9997	•
	11	46244	0.9965	
	:	÷	÷	
	20	127280	1.0001	
	21	138684	1.0001	

In this case, premature convergence to the wrong value 1.0001 occurred, and to achieve four decimal accuracy many more flops were required than $m = 30, 40, 50, \dots$

6 Practicalities for the method of Collatz

To ensure irreducibility of any non-negative matrix A, a slight perturbation of A will suffice. A suggested perturbation is the matrix

$$E = \begin{bmatrix} 0 & \varepsilon & 0 & \cdots & 0 \\ 0 & 0 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \varepsilon & 0 \\ 0 & 0 & \cdots & 0 & \varepsilon \\ \varepsilon & 0 & \cdots & 0 & 0 \end{bmatrix}$$
 (1)

where ε is a suitable small positive quantity. Then $||E||_2 = \varepsilon$ and A + E will be irreducible. Furthermore, if all the diagonal elements of the resulting matrix are zero, a diagonal shift with the identity matrix will ensure that we are not dealing with a cyclic matrix, and that primitivity is assured. A relevant concern when perturbing the elements of A by a small amount is its effect on the eigenvalues. If the matrix has an ill-conditioned dominant eigenvalue, a small perturbation may result in a very inaccurate largest eigenvalue. However, a result from Stewart [14] helpfully determines the impact of a small perturbation: if λ is a simple eigenvalue of the matrix A with corresponding right eigenvector x and left eigenvector y, with $\|x\|_2 = 1$ and $y^Tx = 1$, and A is deflated using an orthogonal matrix R such that

$$R^T A R = \left[\begin{array}{cc} \lambda & h^T \\ 0 & C \end{array} \right] ,$$

then $|\lambda - \lambda'| \leq \varepsilon ||y||_2 + \varepsilon^2/\delta + \eta O(\varepsilon^2)$, where λ' is the corresponding eigenvalue of the perturbed matrix A + E. Also, $\varepsilon = ||E||_2$, $\delta = ||(\lambda I - C)^{-1}||_2^{-1}$ and $\eta = ||h||_2$. So the numbers $||y||_2$, δ and η give a measure of the condition of the simple eigenvalue λ . Numerical experiments were carried out with 100 randomly generated sparse non-negative matrices for each of the orders 100, 200, 300 and 400. Sparsity density was randomly selected in the range (0, 0.5). In all but a very few exceptional cases, $||y||_2$ was less than 10, δ was greater than 0.1 and η less than 10, indicating that typically, the dominant eigenvalue of a non-negative matrix is not greatly affected if the values in E are appropriately small. The few exceptional cases occurred when the sparsity density was very low—less than 2%. So for matrices where sparsity is of such an order it would be wise to convert the matrix to normal form, which is block upper-triangular with the square diagonal blocks being irreducible matrices or null matrices. The method of Collatz may then be applied to those irreducible blocks. Varga [17] details the conversion to normal form. In cases where the method of Collatz produces very slow convergence a superior rate of convergence can usually be achieved by a hybrid method consisting of several steps of the method of Collatz applied to the matrix A, followed by the method of Collatz applied to the matrix $(qI - A)^{-1}$ with q chosen as the upper bound for the dominant eigenvalue calculated after several steps of the method of Collatz for A. Theorem 3 guarantees that $(qI - A)^{-1}$ is non-negative and primitive in this situation. When the Hybrid method was applied to Mark(25) + I_{351} , the method converged in 3.8×10^5 flops, which is of the same order as Collatz. In order to investigate performance of the methods for sparse matrices which are larger, Mark(50) and Mark(100) were 7 Conclusion C341

Table 5: Mark(50), dimensions 1326×132	Table 5:	Mark	(50),	dimensions	1326	X	1326.
---	----------	------	-------	------------	------	---	-------

Method	Total no. of flops
Collatz	2.6×10^{6}
Hybrid	5.3×10^{6}
Arnoldi	15.0×10^{6}
Orthogonal	5.7×10^{6}
Simultaneous	5.6×10^{6}

Table 6: Mark(100), dimensions 5151×5151 .

` /	•
Method	Total no. of flops
Collatz	1.2×10^{7}
Hybrid	6.5×10^{7}
Arnoldi	13.0×10^{7}
Orthogonal	4.4×10^{7}
Simultaneous	4.2×10^{7}

chosen and the results recorded in Tables 5 and 6.

For very large matrices the Hybrid method starts to lose its computational advantage, since it involves the solution of a large system of linear equations. Again the convergence criterion used in these examples was two successive values of $\rho(A)$ approximately agreeing to four decimal places of accuracy.

7 Conclusion

This article presents an always convergent method, the method of Collatz, for finding the spectral radius of a non-negative, irreducible matrix. It is a method closely related to the Power Method but makes use of a result by Collatz. The method has an advantage over the other methods presented in this article, viz Arnoldi, Orthogonal Iteration, and Simultaneous Iteration,

in that it bounds the dominant eigenvalue. For large, sparse matrices the number of flops is either of lower order or of a comparable order with the other methods. If the convergence of the method of Collatz is slow for A, the method can be applied to $(qI - A)^{-1}$ and this has superior convergence if q is chosen appropriately and sufficiently close to λ_1 , the dominant eigenvalue of A. Thus the method can be restarted in the case of very slow convergence, via the so-called Hybrid method. The method of Collatz and the Hybrid method also have an advantage over the Arnoldi method in that Arnoldi requires the initial size of the Krylov space to be predetermined. A variation of Arnoldi allows for the adaptive choice of the sub-space size, but we did not investigate this option. A poor choice of the initial sub-space dimension can affect performance of the Arnoldi method, in that more flops are needed, and in some cases this leads to premature convergence to the wrong value. The method of Collatz and the Hybrid method do not suffer from this difficulty.

References

- [1] E. Andersson and P. Ekstrom. Investigating Google's PageRank Algorithm. Technical report, Uppsala University, 2004. http://www.it.uu.se/edu/course/homepage/projektTDB/vt04/projekt5/website/report.pdf C331
- [2] W. E. Arnoldi. The principle of minimized iterations in the solution of the matrix eigenvalue problem. Q. Appl. Maths, 9:17–29, 1951. C332, C335
- [3] A. Brauer. The theorems of Ledermann and Ostrowski on positive matrices. *Duke Math. J.*, 24:265–274, 1957. C332
- [4] L. Collatz. Einschliessungenssatz für die characteristischen Zahlen von Matrizen. *Math Zeit*, 48:221–6, 1942. C332

[5] G Frobenius. Über Matrizen aus nicht negativen Elementen. Akad Wiss, Berlin, 1912. C332

- [6] G. H. Golub and C. Greif. An Arnoldi-type Algorithm for Computing PageRank, 2006. http://www.cs.ubc.ca/~greif/Papers/gg2006BIT.pdf C331
- [7] G. H. Golub and C. F. Van Loan. *Matrix Computations*. John Hopkins University Press, Baltimore, 1996. C337
- [8] W. Ledermann. Bounds for the greatest latent root of a positive matrix. J. London Math Soc, 25:265–268, 1950. C332
- [9] M. Lu, H. Liu, and F. Tian. A new upper bound for the Spectral Radius of graphs with girth at least 5. Lin Alg Appl, 414:512–516, 2006. C331
- [10] A. Ostrowski. Bounds for the greatest latent root of a positive matrix. J. London Math Soc, pages 253–256, 1952. C332
- [11] Y. Saad. Variations of Arnoldi's Method for computing Eigenelements of Large Unsymmetric Matrices. *Lin Alg. and its Applic*, 34:269–295. C332
- [12] Y. Saad. Numerical Methods for Large Eigenvalue Problems. Manchester University Press, New York, 1992. C332, C334, C335
- [13] J. Snellman. The Maximal Spectral Radius of a Digraph with $(m+1)^2-s$ Edges. Elec J. of Linear Algebra, 10:179–189, 2003. C331
- [14] G. W. Stewart. Introduction to Matrix Computations. Academic Press, New York, 1973. C340
- [15] G. W. Stewart. Simultaneous Iteration for Computing Invariant Subspaces of Non-Hermitian Matrices. *Numer. Math*, pages 123–136, 1976. C337, C338

[16] G. W. Stewart. *Matrix Algorithms*, volume 2: Eigensystems. SIAM, 2001. C332

- [17] R. Varga. *Matrix Iterative Analysis*. Prentice-Hall Inc, Englewood Cliffs, New Jersey, 1962. C333, C340
- [18] R. J. Wood and M. J. O'Neill. Using the spectral radius to determine whether a Leontief system has a unique positive solution. *Asia Pacific Journal of Operational Research, Operational Research Society of Singapore*, 19:233–247, 2002. C331
- [19] R. J. Wood and M. J. O'Neill. An always convergent method for finding the spectral radius of an irreducible non-negative matrix, 2004. http://anziamj.austms.org.au/V45/CTAC2003/Wood C332, C333, C334
- [20] R. J. Wood and M. J. O'Neill. A faster algorithm for identification of an M-Matrix, 2005.

http://anziamj.austms.org.au/V46/CTAC2004/Wood C331

Author addresses

1. R. J. Wood, School of Information Technology, Charles Sturt University, Bathurst, N.S.W. 2795, Australia.

2. M. J. O'Neill, School of Information Technology, Charles Sturt University, Bathurst, N.S.W. 2795, Australia.