Linear Algebra and its Applications 431 (2009) 1439–1442

Contents lists available at ScienceDirect

Linear Algebra and its Applications

journal homepage: www.elsevier.com/locate/laa

Comparison of subdominant eigenvalues of some linear search schemes

A.J. Pryde

School of Mathematical Sciences, Monash University, Victoria 3800, Australia

ARTICLE INFO

Article history:
Received 13 March 2008
Accepted 8 May 2009
Available online 12 June 2009

Submitted by R.A. Brualdi

AMS classification:
15A18
60J10
60K35
68P10

Keywords:
Subdominant eigenvalue
Move-to-front
Transposition

ABSTRACT

The subdominant eigenvalue of the transition probability matrix of a Markov chain is a determining factor in the speed of transition of the chain to a stationary state. However, these eigenvalues can be difficult to estimate in a theoretical sense. In this paper we revisit the problem of dynamically organizing a linear list. Items in the list are selected with certain unknown probabilities and then returned to the list according to one of two schemes: the move-to-front scheme or the transposition scheme. The eigenvalues of the transition probability matrix Q of the former scheme are well known but those of the latter T are not. Nevertheless the transposition scheme gives rise to a reversible Markov chain. This enables us to employ a generalized Rayleigh–Ritz theorem to show that the subdominant eigenvalue of T is at least as large as the subdominant eigenvalue of Q.

1. Introduction

Suppose we have a collection of n items $B_1, B_2, \ldots, B_n$, such as files in a computer, ordered linearly from “left” to “right”. These items are accessed, independently in a statistical sense, with probabilities $w_1, w_2, \ldots, w_n$. When an item is accessed the list is searched from left to right until the desired item is reached and then returned to the list according to various schemes. This problem of dynamically organizing a linear list has been studied by probability theorists and computer scientists for many years; see Hester and Hirschberg [1] for an early survey. Two schemes that are frequently mentioned in the literature are the move-to-front and the transposition schemes. In the move-to-front scheme the accessed item is returned to the front (left) of the list and all other items retain their relative

E-mail address: alan.pryde@sci.monash.edu.au

0024-3795/$ - see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.laa.2009.05.008

positions. In the transposition scheme, if the accessed item came from the front of the list then it is returned to the same position. Otherwise it is interchanged with the nearest item closer to the front of the list. The move-to-front scheme has been studied extensively. See for example [2–6] and other references mentioned therein. As recently as 2008 Jelenković and Radovanović [7] used the search cost distribution for a move-to-front scheme to compute the fault probabilities for a cache replacement heuristic. In contrast, the transposition scheme is more intractable and much less has been written on it. However, performance measures such as search costs and rate of convergence to stationarity have significance when algorithms are selected for real world systems. Hendricks [8] derived the stationary distribution of the transposition chain and Rivest [6] showed that it has a smaller expected stationary search cost than the move-to-front scheme. More recently Gamarnik and Momčilović [9] established asymptotic optimality for the transposition scheme when the w_j are distributed according to a power law or geometrically. For other references, see [9]. In this paper we compare the subdominant eigenvalues of the move-to-front and transposition schemes.

For each of these two schemes the successive configurations of the list of items forms a Markov chain whose state space is the symmetric group S_n of permutations of the numbers $1, 2, \ldots, n$. We write these permutations in the form $\sigma = (\sigma(1), \sigma(2), \ldots, \sigma(n))$ or $\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_n)$. The transition probability matrices for the move-to-front and transposition schemes, denoted Q and T respectively, are matrices indexed by the elements of S_n. Hence, for $\sigma, \tau \in S_n$ we have

$$Q(\sigma,\tau) = \begin{cases} w_{\sigma(1)} & \text{if } \sigma = \tau, \\ w_{\sigma(k)} & \text{if } \tau = (\sigma_k, \sigma_1, \ldots, \sigma_{k-1}, \sigma_{k+1}, \ldots, \sigma_n) \text{ for some } k > 1, \\ 0 & \text{otherwise} \end{cases}$$

and

$$T(\sigma,\tau) = \begin{cases} w_{\sigma(1)} & \text{if } \sigma = \tau, \\ w_{\sigma(k)} & \text{if } \tau = (\sigma_1, \ldots, \sigma_{k-2}, \sigma_k, \sigma_{k-1}, \sigma_{k+1}, \ldots, \sigma_n) \text{ for some } k > 1, \\ 0 & \text{otherwise.} \end{cases}$$

Using probabilistic arguments, Phatarfod [4] calculated the eigenvalues of Q. They are the numbers of the form $w_{i_1} + w_{i_2} + \cdots + w_{i_k}$ where $1 \leqslant i_1 < i_2 < \cdots < i_k \leqslant n$, $1 \leqslant k \leqslant n$ and $k \neq n-1$ together with 0. The latter is repeated with multiplicity $D(n)$, the number of derangements of n objects; the former have multiplicity $D(n-k)$. It was shown in [5] that these results remain valid for arbitrary complex weights $w_1, w_2, \ldots, w_n$ rather than probabilities.

In this paper it is convenient to treat the numbers $w_1, w_2, \ldots, w_n$ as arbitrary non-negative numbers. It follows that Q and T are non-negative matrices with all row sums equal to the Perron eigenvalue $w_1 + w_2 + \cdots + w_n$ which we denote by $\mu_1(Q)$ or $\mu_1(T)$. Note also that each row of both Q and T contains the weights $w_1, w_2, \ldots, w_n$ exactly once each, whereas the diagonals contain the weights exactly $(n-1)!$ times each.

It is well known that the Markov chain for the transposition scheme is reversible. Indeed, for $\sigma \in S_n$ define $\pi(\sigma) = w_{\sigma(1)}^{n-1} w_{\sigma(2)}^{n-2} \ldots w_{\sigma(n-1)}^{1}$. Then for all $\sigma, \tau \in S_n$ we have $\pi(\sigma)T(\sigma,\tau) = \pi(\tau)T(\tau,\sigma)$ which is the defining condition for reversibility. In particular, summing over σ, we obtain $\pi T = \mu_1(T)\pi$ and so π is a stationary distribution for T in the case of probabilities $w_1, w_2, \ldots, w_n$ summing to 1.

Let R denote the square matrix whose diagonal entries are the numbers $\sqrt{\pi(\sigma)}$ for $\sigma \in S_n$ and whose off-diagonal entries are zero. If all weights are positive, we may set $U = RTR^{-1}$. The reversibility condition becomes $T^* = R^2TR^{-2}$ and so $U^* = U$. Moreover, T and U are similar so they have the same characteristic polynomial. A simple calculation shows

$$U(\sigma,\tau) = \begin{cases} w_{\sigma(1)} & \text{if } \sigma = \tau, \\ \sqrt{w_{\sigma(k-1)} w_{\sigma(k)}} & \text{if } \tau = (\sigma_1, \ldots, \sigma_{k-2}, \sigma_k, \sigma_{k-1}, \sigma_{k+1}, \ldots, \sigma_n) \\ & \text{for some } k > 1, \\ 0 & \text{otherwise.} \end{cases}$$

For the general case of non-negative weights R^{-1} may not exist so we define U by this last identity. By a simple continuity argument, T and U again have the same characteristic polynomial. In particular T has real eigenvalues. We will refer to U as the symmetrized form of T and sometimes write $U = U(w_1, w_2, \ldots, w_n)$ to denote its dependence on the weights.

For any matrix A with real eigenvalues, of size m by m say, we denote its eigenvalues by $\mu_1(A), \ldots, \mu_m(A)$ when arranged in decreasing order and by $\lambda_1(A), \ldots, \lambda_m(A)$ when the order is increasing.

2. The results

Theorem 2.1. *Let Q and T be move-to-front and transposition matrices corresponding to non-negative weights $w_1, w_2, \ldots, w_n$. Then $\mu_2(T) \geqslant \mu_2(Q)$.*

Proof. Let U be the symmetrized form of T with row indices ordered so that for the first $(n-1)!$ indices $\sigma(n) = n$, for the next $(n-1)!$ indices $\sigma(n) = n-1$ and so on. Then U has a block decomposition $U = [U_{ij}]$ for $1 \leqslant i, j \leqslant n$ whose diagonal blocks are of the form $U_{ii} = U(w_1, w_2, \ldots, \widehat{w}_{n+1-i}, \ldots, w_n)$. The symbol $\widehat{w}_j$ is used to denote that w_j is omitted. So $\mu_1(U_{ii}) = w_1 + w_2 + \cdots + \widehat{w}_{n+1-i} + \cdots + w_n$. To simplify notation we will assume that $w_1 \leqslant w_2 \leqslant \cdots \leqslant w_n$. As each U_{ii} is Hermitian, there are unitary matrices V_i such that each $V_i^* U_{ii} V_i$ is a diagonal matrix. For $n-1 \leqslant i \leqslant n$ we remove the column of V_i corresponding to the Perron eigenvalue $\mu_1(U_{ii})$ and denote the resulting matrix by W_i. For other values of i set $W_i = V_i$ and let W denote the matrix with diagonal blocks $W_1, W_2, \ldots, W_n$ and off-diagonal blocks zero. Then $W^*W = I_k$, the identity matrix of order $k = n! - 2$ and $W^*UW = [W_i^* U_{ij} W_j]$ where $1 \leqslant i, j \leqslant n$. So

$$\begin{aligned} trace(W^*UW) &= \sum_{i=1}^{n} trace(W_i^* U_{ii} W_i) \\ &= \sum_{i=1}^{n} trace(U_{ii}) - \mu_1(U_{n-1,n-1}) - \mu_1(U_{nn}) \\ &= trace(U) - (w_1 + \widehat{w}_2 + w_3 + \cdots + w_n) - (\widehat{w}_1 + w_2 + \cdots + w_n) \\ &= trace(U) - (w_3 + \cdots + w_n) - (w_1 + w_2 + \cdots + w_n) \\ &= trace(Q) - \mu_2(Q) - \mu_1(Q) \\ &= \sum_{i=1}^{n!-2} \lambda_i(Q). \end{aligned}$$

By the generalized Rayleigh–Ritz theorem (see [10]) we have

$$\sum_{i=1}^{n!-2} \lambda_i(U) = \min\{trace(X^*UX) : X^*X = I_{n!-2}\}$$

and therefore

$$\sum_{i=1}^{n!-2} \lambda_i(T) = \sum_{i=1}^{n!-2} \lambda_i(U) \leqslant \sum_{i=1}^{n!-2} \lambda_i(Q).$$

Since T and Q have the same trace and the same Perron eigenvalue, we conclude that $\mu_2(T) \geqslant \mu_2(Q)$. □

Using similar techniques, further information can be readily gained about the eigenvalues of T. For example:

Theorem 2.2. *Let T be a transposition matrix corresponding to non-negative weights $w_1, w_2, \ldots, w_n$. Then $\sum_{i=1}^{k} \lambda_i(T) \leqslant 0$ for $1 \leqslant k \leqslant n!/2$.*

Proof. Since the result is trivially true when $n = 2$, we may proceed by induction on n. Assume it is valid for lists of length $n-1$ for some $n > 2$. Take matrices U_{ii} as in the proof of Theorem 2.1. By the induction hypothesis and the generalized Rayleigh–Ritz theorem, for $1 \leqslant i \leqslant n$ and $1 \leqslant h \leqslant (n-1)!/2$ there are matrices W_{ih} of size $(n-1)! \times h$ with orthonormal columns such that $trace(W_{ih}^* U_{ii} W_{ih}) \leqslant 0$. Given

$1 \leqslant k \leqslant n!/2$ choose integers $h_1, h_2, \ldots, h_m$ where $1 \leqslant m \leqslant n$, $1 \leqslant h_i \leqslant (n-1)!/2$ and $h_1 + h_2 + \cdots + h_m = k$. Let W be the $n! \times k$ block matrix whose diagonal blocks are $W_i = W_{ih_i}$ for $1 \leqslant i \leqslant m$ with zeros elsewhere. Then $W^*W = I_k$ and $trace(W^*UW) = \sum_{i=1}^{m} trace(W_i^* U_{ii} W_i) \leqslant 0$ so $\sum_{i=1}^{k} \lambda_i(T) \leqslant 0$. □

Example 2.3. The situation is different if negative weights are permitted. For example, consider the case $n = 3$ and weights $-1, 2, 4$. The eigenvalues of Q are $5, 4, 2, -1, 0, 0$ and those of T are approximately $5.0, -3.429, 3.128 \pm 1.283i, 1.086 \pm 1.643i$. So the eigenvalue with second largest modulus for Q is 4 and for T is -3.429.

Acknowledgement

The author thanks Ravi Phatarfod who posed this problem in a private communication some years ago.

References

[1] J.H. Hester, D.S. Hirschberg, Self-organising linear search, ACM Comput. Surveys 19 (1985) 295–311.
[2] J. Barrera, C. Paroissin, On the distribution of the search cost for the move-to-front rule with random weights, J. Appl. Prob. 41 (2004) 250–262.
[3] J. Fill, An exact formula for the move-to-front rule for self-organising lists, J. Theoret. Prob. 9 (1996) 113–159.
[4] R.M. Phatarfod, On the matrix occurring in a linear search problem, J. Appl. Prob., 28 (1991) 336–346.
[5] R.M. Phatarfod, A.J. Pryde, D. Dyte, On the move-to-front scheme with Markov dependent requests, J. Appl. Prob. 34 (1997) 790–794.
[6] R. Rivest, On self-organising sequential search heuristics, Commun. ACM 19 (1976) 63–67.
[7] P.R. Jelenković, A. Radovanović, The persistent-access-caching algorithm, Random Struct. Algor. 33 (2008) 219–251.
[8] W.J. Hendricks, An account of self-organising systems, SIAM J. Comput. 5 (1973) 715–723.
[9] D. Gamarnik, P. Momčilović, A transposition rule based on a particle process, J. Appl. Prob. 42 (2005) 235–246.
[10] R.A. Horn, C.R. Johnson, Matrix Analysis, Cambridge University Press, Cambridge, 1985.