Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Linear Algebra and its Applications 393 (2004) 47–54

LINEAR ALGEBRA AND ITS APPLICATIONS

www.elsevier.com/locate/laa

Positive matrices associated with synchronised communication networks

Abraham Berman [a,*], Robert Shorten [b], Douglas Leith [b]

[a] *Department of Mathematics, Technion-Israel Institute of Technology, Haifa 32000, Israel*
[b] *Hamilton Institute, NUI Maynooth, Co. Kildare, Ireland*

Received 14 October 2003; accepted 31 July 2004

Submitted by J. McDonald

Abstract

We study the spectrum of a positive matrix that arises in the study of certain communication networks. Bounds are given on the rate of convergence of the networks to the equilibrium condition.

AMS classification: 93A30

Keywords: Second eigenvalue of a positive matrix; Interlacing theorem for rank-1 perturbed matrices; Network congestion control; Communication networks

1. Introduction

In this paper we study properties of a positive matrix that arises in the study of certain communication networks. This matrix is important as it is used to model the dynamic properties of a class of communication network that employs TCP-like congestion control mechanisms. The problem of analysing and designing such networks is of considerable practical importance and has recently attracted much

* Corresponding author. Tel.: +9724 8294101; fax: +9724 8293388.
E-mail addresses: berman@tx.technion.ac.il (A. Berman), robert.shorten@may.ie (R. Shorten), doug.leith@may.ie (D. Leith).

0024-3795/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.laa.2004.07.016

attention in the context of internet congestion control [1–6]. In this paper we report results that represent a preliminary step toward the mathematical design of such networks. We use the theory of positive linear systems to relate important properties of synchronised communication networks to properties of a positive matrix. In particular, by characterising the spectra of this matrix, we establish bounds on the rate of convergence of the network to the equilibrium state.

This paper is structured as follows. We use positive linear systems theory to model TCP-based communications network in Section 2. In Section 3 we give bounds on the eigenvalues of the network matrix and relate its Perron eigenvector to the network equilibrium. Finally, the implications of our results are discussed in Section 4.

2. A model of TCP congestion control

We consider a network of n-sources competing for shared bandwidth. A communication network consists of a number of sources and sinks connected together via links and routers. We assume that these links can be modelled as a constant propagation delay together with a queue to buffer bursty traffic, and that all of the sources are operating a TCP-like congestion control algorithm.

TCP operates a window based congestion control algorithm. The TCP standard defines a variable *cwnd* called the congestion window. Each source uses this variable to track the number of sent unacknowledged packets that can be in transit at any time, i.e. the number of packets in the 'pipe' formed by the links and buffers in a transmission path. When the window size is exhausted, the source must wait for an acknowledgement (ACK) before sending a new packet. Congestion control is achieved by dynamically adapting the window size according to an additive-increase multiplicative-decrease (AIMD) law. The basic idea is for a source to gently probe the network for spare capacity and rapidly backoff its send rate when congestion is detected. A typical window evolution is depicted in Fig. 1 ($cwnd_i$ at the time of detecting congestion is denoted by w_i in this figure). Over the kth congestion epoch three important events can be discerned: $t_a(k)$, $t_b(k)$ and $t_c(k)$ in Fig. 1. The time $t_a(k)$ is the time at which the number of unacknowledged packets in the pipe equals $\beta_i w_i(k)$; $t_b(k)$ is the time at which the pipe is full; and $t_c(k)$ is the time at which packet drop is detected by the sources. Note that we measure time in units of round-trip time (RTT).[1]

We consider a network of sources operating AIMD congestion control algorithms. Each source is parameterized by an additive increase parameter and a multiplicative decrease factor, denoted α_i and β_i respectively. These parameters satisfy $\alpha_i \geqslant 1$ and $0 < \beta_i < 1\ \forall i \in \{1, \dots, n\}$. We assume that the event times t_a, t_b and t_c indicated in Fig. 1 are the same for every source, i.e. that the sources are synchronised.

[1] RTT is the time taken between a source sending a packet and receiving the corresponding acknowledgement, assuming no packet drop.

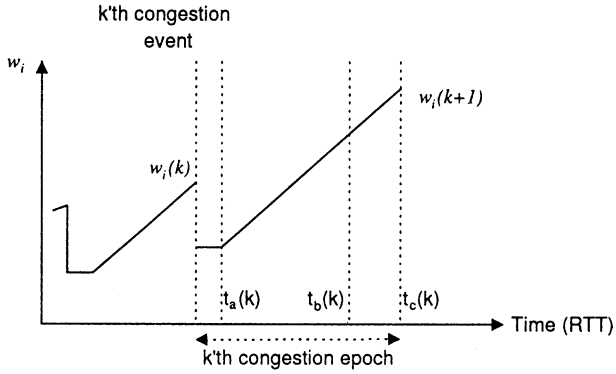

Fig. 1. Evolution of window size.

Let $w_i(k)$ denote congestion window size of source i immediately before the kth network congestion event is detected by the sources; see Fig. 1. It follows from the definition of the AIMD algorithm that the window evolution is completely defined over all time instants by knowledge of the $w_i(k)$ and the event times $t_a(k)$, $t_b(k)$ and $t_c(k)$ of each congestion epoch. We therefore only need to investigate the behaviour of these quantities.

We have that $t_c(k) - t_b(k) = 1$; namely, each source is informed of congestion exactly one RTT after the first dropped packet was transmitted. Also,

$$w_i(k) \geqslant 0, \quad \textstyle\sum_{i=1}^n w_i(k) = P + \sum_{i=1}^n \alpha_i, \qquad \forall k > 0, \tag{1}$$

where P is the maximum number of packets which can be held in the 'pipe'; this is usually equal to $q_{\max} + BT$ where $q_{\max}$ is the maximum queue length of the congested link, B is the service rate in packets per second and T is the round-trip time. At the $(k+1)$th congestion event

$$w_i(k+1) = \beta_i w_i(k) + \alpha_i[t_c(k) - t_a(k)] \tag{2}$$

and

$$t_c(k) - t_a(k) = \frac{1}{\sum_{i=1}^n \alpha_i}\left[P - \sum_{i=1}^n \beta_i w_i(k)\right] + 1. \tag{3}$$

Hence, it follows that

$$w_i(k+1) = \beta_i w_i(k) + \frac{\alpha_i}{\sum_{j=1}^n \alpha_i}\left[\sum_{i=1}^n (1-\beta_i) w_i(k)\right] \tag{4}$$

and that the dynamics of the entire network can be written in matrix form as

$$W(k+1) = AW(k), \tag{5}$$

where $W^{\mathrm{T}}(k) = [w_1(k), \dots, w_n(k)]$, and

$$A = \begin{bmatrix} \beta_1 & 0 & \cdots & 0 \\ 0 & \beta_2 & 0 & 0 \\ \vdots & 0 & \ddots & 0 \\ 0 & 0 & \cdots & \beta_n \end{bmatrix} + \frac{1}{\sum_{j=1}^{n} \alpha_i} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \cdots \\ \alpha_n \end{bmatrix} [1-\beta_1 \ 1-\beta_2 \ \cdots \ 1-\beta_n]. \tag{6}$$

In the sequel it is convenient to write A in the form

$$A = I - Y + xy^{\mathrm{T}}, \tag{7}$$

where $Y = \mathrm{diag}([1-\beta_1, \dots, 1-\beta_n])$, $x^{\mathrm{T}} = \frac{1}{\sum_{i=1}^{n} \alpha_i}[\alpha_1, \dots, \alpha_n]$ and $y^{\mathrm{T}} = [1-\beta_1, \dots, 1-\beta_n]$, and where we assume that the entries of $I - Y$ have been ordered:

$$\begin{aligned} \beta_1 = \cdots = \beta_{k_1} &= \gamma_1, \\ \beta_{k_1+1} = \cdots = \beta_{k_1+k_2} &= \gamma_2, \\ &\vdots \\ \beta_{k_1+k_2+\cdots+k_{s-1}+1} = \cdots = \beta_{k_1+k_2+\cdots+k_s} &= \gamma_s, \end{aligned}$$

with $k_1 + k_2 + \cdots + k_s = n$ and $\gamma_1 < \gamma_2 < \cdots < \gamma_s$.

Comment 1. The matrix A is strictly positive and it follows that the synchronised network (5) is a positive linear system.

Comment 2. The vector x is a probability vector and the matrix A is a column stochastic matrix. Thus, the Perron eigenvalue $\rho(A) = 1$ and the all the eigenvectors of A, except the Perron vector, are orthogonal to $e^{\mathrm{T}} = [1, \dots, 1]$.

3. The spectrum of the network matrix

We now present the main mathematical results of the paper. The following theorem and corollary are easily derived [7] and establish basic properties of the communication networks under study in this paper.

Theorem 3.1. *Let A be defined as in Eq.* (6). *Then, a Perron eigenvector of A is given by* $x_p^{\mathrm{T}} = [\frac{\alpha_1}{1-\beta_1}, \dots, \frac{\alpha_n}{1-\beta_n}]$.

The following corollary follows from Theorem 3.1 and properties of non-negative matrices [8,9].

Corollary 3.1. *For a network of synchronised time-invariant AIMD sources*: (i) *the network has a Perron eigenvector* $x_p^{\mathrm{T}} = [\frac{\alpha_1}{1-\beta_1}, \ldots, \frac{\alpha_n}{1-\beta_n}]$; *and* (ii) *the Perron eigenvalue is* $\rho(A) = 1$. *It follows that all other eigenvalues of A satisfy* $|\lambda_i(A)| < \rho(A)$. *The network possesses a unique stationary point* $W_{ss} = \Theta x_p$, *where* Θ *is a positive constant such that the constraint* (1) *is satisfied;* $\lim_{k\to\infty} W(k) = \Theta x_p$, *and the rate of convergence of the network to* W_{ss} *depends upon the second largest eigenvalue of* A ($\max|\lambda|, \lambda \neq 1 \in \mathrm{spec}(A)$).

It follows from the corollary that the second largest eigenvalue of the matrix A determines the convergence properties of the entire network. It is therefore important to determine this eigenvalue. Theorem 3.2, which is the main result of the paper, provides a characterisation of all the eigenvalues of the matrix A. It shows that all the eigenvalues of A are real and positive and lie in the interval $[\beta_1, 1]$. In particular, the second largest eigenvalue is bounded above by β_n.

We now present the main result of the paper. In the following discussion it is useful to study the eigenvalues of A by considering A^{T}. To aid exposition we begin by stating the following lemma before proceeding with the main result.

Lemma 3.1. *Let A be the matrix defined in* (7). *Then*, *if* k_i *is greater than one*, γ_i *is an eigenvalue of A with a geometric multiplicity of at least* $k_i - 1$.[2]

Proof. For every vector in the $k_i - 1$ dimensional subspace $\{z | x^{\mathrm{T}}z = 0 \text{ with } z_j = 0\ \forall j \notin \{k_1 + \cdots + k_{i-1} + 1, \ldots, k_1 + \cdots + k_i\}$, we have $A^{\mathrm{T}}z = (I - Y + yx^{\mathrm{T}})z = (I - Y)z = \gamma_i z$. $\square$

Theorem 3.2. *Consider the matrix* (7). *Then*, *the following statements are true*.

(a) *The matrix A is diagonally similar to a* (*real*) *positive diagonal matrix*.
(b) *Except for the Perron eigenvalue*, *all of the eigenvalues of A lie in the interval* $[\beta_1, \beta_n]$.
(c) *More specifically*, *if* $k_i > 1$, *then* γ_i *is an eigenvalue of A of multiplicity* $k_i - 1$, *and the remaining eigenvalues are simple*, *and with the exception of* 1 *lie in the intervals*, (γ_1, γ_2), (γ_2, γ_3), $\ldots$, (γ_{s-1}, γ_s).
(d) *In particular*, *if all the* β'*s are distinct*, *then* $\beta_1 < \lambda_1 < \beta_2 < \cdots < \beta_{n-1} < \lambda_{n-1} < \beta_n < \lambda_n = 1$.

Proof. Let v be an eigenvector of A^{T} with corresponding eigenvalue λ. Then,

$$A^{\mathrm{T}}v = (I - Y)v + yx^{\mathrm{T}}v = \lambda v. \tag{8}$$

We consider two cases: (i) $x^{\mathrm{T}}v = 0$; (ii) $x^{\mathrm{T}}v \neq 0$.

---

[2] In fact, we shall see that the geometric multiplicity of γ_i is exactly $k_i - 1$.

Case (i) $x^{\mathrm{T}}v = 0$:

It follows from Eq. (8) that

$$A^{\mathrm{T}}v = (I - Y)v = \lambda v. \tag{9}$$

So $v_i \neq 0$ implies that $\beta_i = \lambda$. Since v is an eigenvector there is an index i such that $v_i \neq 0$. In fact, at least two coordinates must be non-zero since v is orthogonal to the positive vector x. Thus, the solutions of (9) are $\lambda = \gamma_i$ whenever $k_i > 1$. Denoting the γ_i's in (9) for which $k_i > 1$ by $\gamma_{i_1}, \ldots, \gamma_{i_t}$, we see that they are eigenvalues of A^{T}.

Case (ii) $x^{\mathrm{T}}v \neq 0$:

Now consider Eq. (8) component-wise. Then,

$$v_i - y_i v_i + y_i x^{\mathrm{T}}v = \lambda v_i, \quad \forall\, 1 \leqslant i \leqslant n. \tag{10}$$

Assume that v is chosen such that $x^{\mathrm{T}}v = 1$. Then,

$$v_i - y_i v_i + y_i = \lambda v_i, \quad \forall\, 1 \leqslant i \leqslant n. \tag{11}$$

But $\beta_i \in (0, 1)$. Hence, v_i cannot be zero as this would imply that $y_i = (1 - \beta_i) = 0$.

Hence,

$$\lambda = 1 - y_i + \frac{y_i}{v_i}, \ \forall\, 1 \leqslant i \leqslant n \tag{12}$$

and denoting $r = -y_i + \frac{y_i}{v_i}\ \forall i$ we have

$$1 - y_1 + \frac{y_1}{v_1} = 1 - y_i + \frac{y_i}{v_i} = 1 + r = \lambda, \quad \forall\, 1 \leqslant i \leqslant n. \tag{13}$$

But $x^{\mathrm{T}}v = 1$. Hence,

$$\sum_{i=1}^{n} x_i v_i = \sum_{i=1}^{n} \frac{x_i y_i}{y_i + r} = 1. \tag{14}$$

The solutions to Eq. (14) determine the remaining eigenvalues of A^{T}. Observe that $r = 0$ is a solution corresponding to $\lambda = 1$ (the Perron eigenvalue of A^{T}). The function

$$g(r) = \sum_{i=1}^{n} \frac{x_i y_i}{y_i + r} \tag{15}$$

is continuous and decreasing except for $r \in \{-(\gamma_1 - 1), \ldots, -(\gamma_s - 1)\}$. For $r > 0$ we have that $0 < g(r) < 1$ and there is no solution to (14). This is consistent with the fact that A^{T} is a positive matrix and $\rho(A^{\mathrm{T}}) = 1$. For $\gamma_s - 1 < r < 0$ we have that $g(r) > 1$ and for $r < \gamma_1 - 1$ we have that $g(r) < 0$. Hence, there are no real solutions of (14) outside of the interval $[\gamma_1 - 1, \gamma_s - 1]$ other than the solution at $r = 0$. Finally, by considering the limit as r approaches $\gamma_i - 1$ from the right, and $\gamma_{i+1} - 1$ from the left, we conclude that (14) has one solution for $\gamma_i - 1 < r < \gamma_{i+1} - 1$,

$\forall i \in \{1, \ldots, s-1\}$. It follows that A^{T} has exactly one eigenvalue, denoted δ_i, in every interval (γ_i, γ_{i+1}) for all $i \in \{1, \ldots, s-1\}$.

We now complete the proof by showing that $\gamma_{i_1}, \ldots, \gamma_{i_t}, \delta_1, \ldots, \delta_{s-1}, 1$ are all the eigenvalues of A^{T}, that the multiplicities of $\gamma_i, \ldots, \gamma_{i_t}$ are $k_{i_1} - 1, \ldots k_{i_t} - 1$ respectively, that the $\delta_1, \ldots, \delta_{s-1}$ are simple, and that A is diagonalisable.

Suppose that the geometric multiplicities of $\gamma_{i_1}, \ldots, \gamma_{i_t}$ are $l_1, \ldots, l_t$ respectively, and that the geometric multiplicities of $\delta_1, \ldots, \delta_{s-1}$ are $m_1, \ldots, m_{s-1}$ respectively. By the lemma, $l_j \geqslant k_{i_j} - 1$; $j = 1, \ldots, t$. Also, $k_{i_1} + k_{i_2} + \cdots + k_{i_t} = n - (s - t)$. The sum of the geometric multiplicities satisfy, $n \geqslant l_1 + \cdots + l_t + m_1 + \cdots + m_{s-1} + 1 \geqslant k_{i_1} + \cdots + k_{i_t} - t + s = n - (s-t) - t + s = n$, so the sum must be n. Hence, A must be diagonalisable, $l_j = k_{i_j} - 1$, $j \in \{1, \ldots, t\}$ and $m_i = 1$ for all $i \in \{1, \ldots, s-1\}$. $\square$

4. Discussion and concluding remarks

In this paper we have analysed the spectrum of a positive matrix that arises in the context of a certain type of communication network. Specifically, our matrix describes the dynamic behaviour of synchronised communication networks where each source operates an AIMD congestion control algorithm and where each of the sources share the same RTT. While these assumptions do not apply to general communication networks, they are valid for important network types; in particular, for long-distance high-speed networks [10,11,7]. A basic problem in the design of these networks is to ensure fairness of the equilibrium condition, good throughput of data, and to ensure rapid convergence to the equilibrium condition in the presence of network disturbances. Our results show that: (i) fairness[3] is ensured when A is symmetric, i.e. by choosing the $\frac{\alpha_i}{1-\beta_i}$ to be constant for all sources in the network (Theorem 3.1); and (ii) the rate of convergence of the network to its equilibrium state is determined by β_n in the case when β_n is an eigenvalue of $I - Y$ with multiplicity greater than one, and is bounded above and below by β_n and β_{n-1} respectively when β_n has multiplicity 1 (Theorem 3.2). Importantly, our results indicate that good data throughput, which is normally achieved by ensuring that the β_i's are set to large values, cannot be achieved without adversely affecting the network convergence properties. In particular, knowledge of the eigenvalue locations of the matrix A (Theorem 3.2), and the fact that the non-Perron eigenvectors of this matrix satisfy $e^{\mathrm{T}}z = 0$ (Comment 2), provides a basis for understanding the transient response of such networks, and may provide a network for identification of network parameters from measured data. In this context our results may even provide a basis for network operators to detect of malicious network attacks (users setting their value of β to large values). Finally, we note that our results are also likely to provide

[3] When defined to be an equal share of the network 'pipe' for all sources.

valuable insights into the design of adaptive congestion control algorithms and this is currently an active area of research. We believe that the results presented here represent a small, but nevertheless important step in the mathematical design of communication networks. Future work will involve extending our results to the case of non-synchronised networks.

Acknowledgments

This work was supported by Science Foundation Ireland grant 00/PI.1/C067. This work was also partially supported by the European Union funded research training network *Multi-Agent Control*, HPRN-CT-1999-00107[4] and by the Enterprise Ireland grant SC/2000/084/Y. Neither the European Union or Enterprise Ireland is responsible for any use of data appearing in this publication.

References

[1] V. Jacobson, Congestion avoidance and control, in: Proceedings of SIGCOMM, 1988.
[2] S. Floyd, K. Fall, Promoting the use of end-to-end congestion control in the internet, IEEE/ACM Trans. Network. 7 (4) (1999) 458–472.
[3] Special issue on TCP performance in future networking environments, IEEE Commun. Mag. 39(4) (2001).
[4] S. Low, F. Paganini, J. Doyle, Internet congestion control, IEEE Control Syst. Mag. 32 (1) (2002) 28–43.
[5] Special issue on internet technology and convergence of communication services, Proc. IEEE 90(9) (2002).
[6] J. Hespanha, S. Hohacek, K. Obrarzka, J. Lee, Hybrid model of TCP congestion control, Hybrid Syst.: Comput. Control (2001) 291–304.
[7] R. Shorten, D. Leith, J. Foy, R. Kilduff, Analysis and design of synchronised communication networks, in: Proceedings of 12th Yale Workshop on Adaptive and Learning Systems, 2003.
[8] R. Horn, C. Johnson, Matrix Analysis, Cambridge University Press, 1985.
[9] A. Berman, R. Plemmons, Nonnegative matrices in the mathematical sciences, SIAM, 1979.
[10] S. Floyd, High speed TCP for large congestion windows, Tech. Rep., Internet draft draft-fioyd-tcp-highspeed-02.txt, work in progres, February 2003.
[11] T. Kelly, On engineering a stable and scalable TCP variant, Tech. Rep., Cambridge University Engineering Department Technical Report CUED/F-INFENG/TR.435, 2002.

[4] This work is the sole responsibility of the authors and does not reflect the European Union's opinion.