Available online at www.sciencedirect.com
SCIENCE DIRECT®
LINEAR ALGEBRA AND ITS APPLICATIONS
ELSEVIER Linear Algebra and its Applications 375 (2003) 291–297
www.elsevier.com/locate/laa

An alternative derivation of Birkhoff's formula for the contraction coefficient of a positive matrix

Rolando Cavazos-Cadena

Departamento de Estadística y Cálculo, Universidad Autónoma Agraria, Antonio Narro, Buenavista, Saltillo COAH, 25315, Mexico

Received 31 October 2002; accepted 20 May 2003

Submitted by R.A. Brualdi

Abstract

This note concerns the projective contraction coefficient $\tau(H)$ of a rectangular matrix H with positive entries. A simple proof of an explicit formula for $\tau(H)$, originally established by [Trans. Am. Math. Soc. 85 (1957) 219], is given. The motivation for this work comes from the area of Markov decision processes, and the argument is based on elementary differential calculus.

AMS classification: 15A48; 90C39

Keywords: Projective distance; Mean value theorem; Risk-sensitive dynamic programming; Value iteration algorithm

1. Introduction

Given an integer $n \geqslant 2$, let $\mathscr{P}_n$ be the positive cone in $\mathbb{R}^n$, so that $\mathscr{P}_n$ consists of all vectors $\mathbf{x} = (x_1, x_2, \ldots, x_n)'$ satisfying $x_i > 0$ for all i. The projective distance $d_n : \mathscr{P}_n \times \mathscr{P}_n \to [0, \infty)$ is defined by

$$d_n(\mathbf{x}, \mathbf{y}) := \max\left\{ \log\left(\frac{x_r}{x_s}\frac{y_s}{y_r}\right) \middle| r, s = 1, 2, \ldots, n,\ r \neq s \right\}, \quad \mathbf{x}, \mathbf{y} \in \mathscr{P}_n \quad (1.1)$$

and if $H = [h_{ij}]$ is an $m \times n$ matrix with positive components, it is not difficult to see that $d_m(H\mathbf{x}, H\mathbf{y}) < d_n(\mathbf{x}, \mathbf{y})$ for every $\mathbf{x}, \mathbf{y} \in \mathscr{P}_n$ satisfying $d_n(\mathbf{x}, \mathbf{y}) > 0$ [5, p. 100],

E-mail address: rcavazos@narro.uaaan.mx (R. Cavazos-Cadena).

0024-3795/$ - see front matter © 2003 Elsevier Inc. All rights reserved.
doi:10.1016/j.laa.2003.05.002

so that the *projective contraction coefficient associated to* H, which is denoted by $\tau(H)$ and is given by

$$\tau(H) := \sup_{\mathbf{x},\mathbf{y}\in\mathscr{P}_n : d_n(\mathbf{x},\mathbf{y})>0} \frac{d_m(H\mathbf{x}, H\mathbf{y})}{d_n(\mathbf{x}, \mathbf{y})}, \tag{1.2}$$

satisfies $\tau(H) \leqslant 1$. As the explicit formula below shows, the inequality is strict.

Theorem 1.1. *Let the integers $m, n \geqslant 2$ be arbitrary. For each matrix $H = [h_{i\,j}]$ of order $m \times n$ with positive components, $\tau(H)$ is given by the following Birkhoff's formula:*

$$\tau(H) = \frac{1-\phi(H)^{1/2}}{1+\phi(H)^{1/2}}, \quad \text{where } \phi(H) := \min_{r,s,j,k} \frac{h_{r\,j}h_{s\,k}}{h_{s\,j}h_{r\,k}}. \tag{1.3}$$

For square matrices, this theorem was first established by Birkhoff [1] using arguments relying on projective geometry, and extensions of this result to the setting of partially ordered vector spaces can be found in [3,4]. An elementary proof of Theorem 1.1 was given in [5, pp. 100–110] where it is shown that $\tau(\cdot)$ is an essential tool in the analysis of the asymptotic behavior of matrix products and, recently, the importance of Birkhoff's formula in the study of the value iteration algorithm for risk-sensitive average Markov decision chains was sparkled by Bielecki et al. [2], where (1.3) was used for rectangular matrices. Although it is not difficult to see that the proof in [5] is valid for non square matrices, the arguments in this reference are rather long and quite involved, so that it is interesting to look for a simpler proof of Theorem 1.1. *The objective of this note* is to give an alternative proof of Birkhoff's formula which is based on elementary differential calculus. The argument, relying on the mean value theorem, is presented in the following two sections.

2. Preliminaries

This section contains the technical tools that will be used to prove Theorem 1.1. In the following two lemmas the contraction coefficient of the matrix

$$H = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \\ 1 & 1 & \cdots & 1 \end{bmatrix} \tag{2.1}$$

is studied, where $n \geqslant 2$ and $a_1, a_2, \ldots, a_n$ are fixed positive numbers; set

$$A = \max\{a_1, a_2, \ldots, a_n\} \quad \text{and} \quad a = \min\{a_1, a_2, \ldots, a_n\}. \tag{2.2}$$

The argument uses the functions $g_\alpha : [0, \infty) \to \mathscr{R}$ determined by

$$g_\alpha(z) = \frac{1}{1+\alpha z} - \frac{1}{1+z}, \quad z \geqslant 0, \quad \alpha \in (0, 1); \tag{2.3}$$

it is not difficult to see that

$$0 \leqslant g_\alpha(z) \leqslant g_\alpha(1/\sqrt{\alpha}) = \frac{1-\sqrt{\alpha}}{1+\sqrt{\alpha}}, \quad z \geqslant 0, \quad \alpha \in [0, 1]. \tag{2.4}$$

Lemma 2.1. *Let H be as in* (2.1) *and let* $\mathbf{x}, \mathbf{y} \in \mathscr{P}_n$ *be arbitrary vectors satisfying* $d_n(\mathbf{x}, \mathbf{y}) > 0$. *In this case, there exists* $z \in (0, \infty)$ *such that*

$$\frac{d_2(H\mathbf{x}, H\mathbf{y})}{d_n(\mathbf{x}, \mathbf{y})} \leqslant g_\alpha(z), \quad \text{where } \alpha = \frac{a}{A}; \tag{2.5}$$

see (2.2) *for notation. Consequently,*

$$\tau(H) \leqslant \frac{1 - \sqrt{a/A}}{1 + \sqrt{a/A}}. \tag{2.6}$$

Proof. Notice that, since $g_\alpha(\cdot) \geqslant 0$, when $d_2(H\mathbf{x}, H\mathbf{y}) = 0$ the inequality in (2.5) holds with an arbitrary $z > 0$, so that it is sufficient to consider the case $d_2(H\mathbf{x}, H\mathbf{y}) > 0$. In this situation, interchanging $\mathbf{x}$ and $\mathbf{y}$, if necessary, it can be assumed that $[H\mathbf{x}]_1/[H\mathbf{x}]_2 \geqslant [H\mathbf{y}]_1/[H\mathbf{y}]_2$, so that

$$d_2(H\mathbf{x}, H\mathbf{y}) = \log\left(\frac{[H\mathbf{x}]_1}{[H\mathbf{x}]_2}\frac{[H\mathbf{y}]_2}{[H\mathbf{y}]_1}\right) = \log\left(\frac{\sum_i a_i x_i}{\sum_i x_i} \times \frac{\sum_i y_i}{\sum_i a_i y_i}\right); \tag{2.7}$$

see (1.1). Define

$$r_i = \log\left(\frac{x_i}{c y_i}\right), \quad i = 1, 2, \ldots, n, \quad \text{where } c = \min_{i=1,2,\ldots,n} \frac{x_i}{y_i} = \frac{x_{i^*}}{y_{i^*}} \tag{2.8}$$

and notice that $r_i \geqslant r_{i^*} = 0$. Thus, $x_i = c y_i \mathrm{e}^{r_i}$ for every i, and (2.7) yields

$$d_2(H\mathbf{x}, H\mathbf{y}) = \log\left(\frac{\sum_i a_i y_i \mathrm{e}^{r_i}}{\sum_i y_i \mathrm{e}^{r_i}} \times \frac{\sum_i y_i}{\sum_i a_i y_i}\right) = F(1) - F(0), \tag{2.9}$$

where

$$F(t) = \log\left(\frac{\sum_i a_i y_i \mathrm{e}^{t r_i}}{\sum_i y_i \mathrm{e}^{t r_i}}\right), \quad t \in \mathbb{R}. \tag{2.10}$$

On the other hand, (2.8) shows that $\log([x_i y_j]/[x_j y_i]) = r_i - r_j$, and using that the r_j's are nonnegative and $r_{i^*} = 0$, from (1.1) it follows that

$$d_n(\mathbf{x}, \mathbf{y}) = r^* = \max_{i=1,2,\ldots,n} r_i.$$

Combining this equality with (2.9) and (2.10), by the mean value theorem there exists $t \in (0, 1)$ such that

$$\begin{aligned}\frac{d_2(H\mathbf{x}, H\mathbf{y})}{d_n(\mathbf{x}, \mathbf{y})} &= \frac{1}{r^*}F'(t) = \frac{1}{r^*}\frac{\sum_i r_i a_i y_i \mathrm{e}^{t r_i}}{\sum_i a_i y_i \mathrm{e}^{t r_i}} - \frac{1}{r^*}\frac{\sum_i r_i y_i \mathrm{e}^{t r_i}}{\sum_i y_i \mathrm{e}^{t r_i}} \\ &= \sum_i \frac{r_i}{r^*} w_i \left(\frac{a_i}{\sum_k a_k w_k} - \frac{1}{\sum_k w_k}\right),\end{aligned} \tag{2.11}$$

where $w_i := y_i \mathrm{e}^{t r_i} > 0$ for each i. Let the set $\mathscr{I} \subset \{1, 2, \ldots, n\}$ be characterized by

$$i \in \mathscr{I} \iff r_i \left(\frac{a_i}{\sum_k a_k w_k} - \frac{1}{\sum_k w_k}\right) > 0.$$

Since $d_2(H\mathbf{x}, H\mathbf{y})/d_n(\mathbf{x}, \mathbf{y}) > 0$, (2.11) yields that $\mathscr{I}$ is nonempty, and $r_{i^*} = 0$ implies that $i^* \notin \mathscr{I}$, so that $\mathscr{I}$ is a proper subset of $\{1, 2, \ldots, n\}$. Using that r^* is the maximum value of the r_i's, (2.11) implies that

$$\frac{d_2(H\mathbf{x}, H\mathbf{y})}{d_n(\mathbf{x}, \mathbf{y})} \leqslant \sum_{i\in\mathscr{I}} \frac{r_i}{r^*} w_i \left(\frac{a_i}{\sum_k a_k w_k} - \frac{1}{\sum_k w_k} \right)$$
$$\leqslant \sum_{i\in\mathscr{I}} w_i \left(\frac{a_i}{\sum_k a_k w_k} - \frac{1}{\sum_k w_k} \right)$$

and then

$$\frac{d_2(H\mathbf{x}, H\mathbf{y})}{d_n(\mathbf{x}, \mathbf{y})} \leqslant \frac{\sum_{i\in\mathscr{I}} a_i w_i}{\sum_k a_k w_k} - \frac{\sum_{i\in\mathscr{I}} w_i}{\sum_k w_k}$$
$$= \frac{\sum_{i\in\mathscr{I}} a_i w_i}{\sum_{i\in\mathscr{I}} a_i w_i + \sum_{k\notin\mathscr{I}} a_k w_k} - \frac{\sum_{i\in\mathscr{I}} w_i}{\sum_{i\in\mathscr{I}} w_i + \sum_{k\notin\mathscr{I}} w_k}. \quad (2.12)$$

Observe now that, for each $\delta > 0$, the mapping $\beta \mapsto \beta/(\beta + \delta)$ is increasing in $(0, \infty)$. Since $\sum_{i\in\mathscr{I}} a_i w_i \leqslant A \sum_{i\in\mathscr{I}} w_i$ (see (2.2)), it follows that

$$\frac{\sum_{i\in\mathscr{I}} a_i w_i}{\sum_{i\in\mathscr{I}} a_i w_i + \sum_{k\notin\mathscr{I}} a_k w_k} \leqslant \frac{A \sum_{i\in\mathscr{I}} w_i}{A \sum_{i\in\mathscr{I}} w_i + \sum_{k\notin\mathscr{I}} a_k w_k}$$
$$\leqslant \frac{A \sum_{i\in\mathscr{I}} w_i}{A \sum_{i\in\mathscr{I}} w_i + a \sum_{k\notin\mathscr{I}} w_k},$$

where the second inequality used that $\sum_{k\notin\mathscr{I}} a_k w_k \geqslant a \sum_{k\notin\mathscr{I}} w_k$. Thus, (2.12) implies

$$\frac{d_2(H\mathbf{x}, H\mathbf{y})}{d_n(\mathbf{x}, \mathbf{y})} \leqslant \frac{A \sum_{i\in\mathscr{I}} w_i}{A \sum_{i\in\mathscr{I}} w_i + a \sum_{k\notin\mathscr{I}} w_k} - \frac{\sum_{i\in\mathscr{I}} w_i}{\sum_{i\in\mathscr{I}} w_i + \sum_{k\notin\mathscr{I}} w_k}$$
$$= \frac{1}{1 + (a/A)z} - \frac{1}{1+z},$$

where $z = (\sum_{k\notin\mathscr{I}} w_k)/(\sum_{k\in\mathscr{I}} w_k)$, establishing (2.5); see (2.3). Finally, combining (2.4) and (2.5), it follows that $d_2(H\mathbf{x}, H\mathbf{y})/d_n(\mathbf{x}, \mathbf{y}) \leqslant (1 - \sqrt{\alpha})/(1 + \sqrt{\alpha})$ whenever $d_n(\mathbf{x}, \mathbf{y}) > 0$, so that (2.6) is obtained from (1.2). □

The following lemma establishes that the equality holds in (2.6).

Lemma 2.2. *For the matrix H in* (2.1),

$$\tau(H) = \frac{1 - \sqrt{a/A}}{1 + \sqrt{a/A}}.$$

Proof. Observe that, since $n \geqslant 2$, there exist integers $j^*, k^* \in \{1, 2, \ldots, n\}$ such that (see (2.2))

$$a_{j^*} = A \text{ and } a_{k^*} = a, \quad \text{where } j^* \neq k^*. \tag{2.13}$$

Next, let $\mathbf{y} \in \mathscr{P}_n$ be arbitrary, and for each $t \geqslant 0$, define the vector $\mathbf{x}(t) = (x_1(t), \ldots, x_n(t))' \in \mathscr{P}_n$ by

$$x_i(t) = y_i, \ i \neq j^*, \quad \text{and} \quad x_{j^*}(t) = \mathrm{e}^t y_{j^*}.$$

In this case, from (1.1) it is not difficult to see that

$$d_n(\mathbf{x}(t), \mathbf{y}) = t,$$

whereas

$$\begin{aligned} d_2(H\mathbf{x}(t), H\mathbf{y}) &\geqslant \log\left(\frac{[H\mathbf{x}(t)]_1}{[H\mathbf{x}(t)]_2}\frac{[H\mathbf{y}]_2}{[H\mathbf{y}]_1}\right) \\ &= \log\left(\frac{a_{j^*}y_{j^*}\mathrm{e}^t + \sum_{i\neq j^*} a_i y_i}{y_{j^*}\mathrm{e}^t + \sum_{i\neq j^*} y_i} \times \frac{\sum_i y_i}{\sum_i a_i y_i}\right), \end{aligned}$$

an inequality that is equivalent to

$$d_2(H\mathbf{x}(t), H\mathbf{y}) \geqslant G(t) - G(0),$$

where

$$G(t) = \log\left(\frac{a_{j^*}y_{j^*}\mathrm{e}^t + \sum_{i\neq j^*} a_i y_i}{y_{j^*}\mathrm{e}^t + \sum_{i\neq j^*} y_i}\right).$$

Therefore, (1.2) yields

$$\tau(H) \geqslant \frac{d_2(H\mathbf{x}(t), H\mathbf{y})}{d_n(\mathbf{x}(t), \mathbf{y})} = \frac{G(t) - G(0)}{t}, \quad t > 0$$

and taking limit as $t \searrow 0$, this implies that

$$\tau(H) \geqslant \frac{a_{j^*}y_{j^*}}{\sum_i a_i y_i} - \frac{y_{j^*}}{\sum_i y_i} = G'(0).$$

Letting y_i decrease to zero for each $i \neq j^*, k^*$, it follows that for each $y_{j^*}, y_{k^*} > 0$,

$$\tau(H) \geqslant \frac{a_{j^*}y_{j^*}}{a_{j^*}y_{j^*} + a_{k^*}y_{k^*}} - \frac{y_{j^*}}{y_{j^*} + y_{k^*}} = \frac{Ay_{j^*}}{Ay_{j^*} + ay_{k^*}} - \frac{y_{j^*}}{y_{j^*} + y_{k^*}},$$

where (2.13) was used to set the equality. Thus

$$\tau(H) \geqslant \frac{1}{1 + (a/A)(y_{k^*}/y_{j^*})} - \frac{1}{1 + (y_{k^*}/y_{j^*})} = g_{a/A}(y_{k^*}/y_{j^*})$$

(see (2.3)) and setting $y_{k^*}/y_{j^*} = \sqrt{A/a}$, (2.4) yields that

$$\tau(H) \geqslant \frac{1 - \sqrt{a/A}}{1 + \sqrt{a/A}}.$$

The conclusion follows from this inequality and (2.6). $\square$

The final preliminary result concerns an arbitrary positive and rectangular matrix H; it establishes an invariance property of $\tau(H)$, and shows that this quantity can be determined from the contraction coefficients of submatrices of order $2 \times n$.

Lemma 2.3. *Let H be a matrix of order $m \times n$ with positive components, where $m, n \geqslant 2$.*

(i) *For integers r, s satisfying $1 \leqslant r < s \leqslant m$, let $H_{[r,s]}$ be the submatrix consisting of the rth and sth rows of H. In this case,*

$$\tau(H) = \max_{r,s:1\leqslant r<s\leqslant m} \tau(H_{[r,s]}). \tag{2.14}$$

(ii) *If D is a diagonal matrix of order $n \times n$ with positive elements $\delta_1, \delta_2, \ldots, \delta_n$ along the main diagonal, then $\tau(H) = \tau(HD)$.*

Proof. (i) For each $\mathbf{x} \in \mathscr{P}_n$ write $H\mathbf{x} = ([H\mathbf{x}]_1, [H\mathbf{x}]_2, \ldots, [H\mathbf{x}]_n)'$ and notice that $H_{[r\,s]}\mathbf{x} = ([H\mathbf{x}]_r, [H\mathbf{x}]_s)'$. Thus,

$$d_2(H_{[r,s]}\mathbf{x}, H_{[r,s]}\mathbf{y}) = \max\left\{\log\left(\frac{[H\mathbf{x}]_r}{[H\mathbf{x}]_s}\frac{[H\mathbf{y}]_s}{[H\mathbf{y}]_r}\right),\ \log\left(\frac{[H\mathbf{x}]_s}{[H\mathbf{x}]_r}\frac{[H\mathbf{y}]_r}{[H\mathbf{y}]_s}\right)\right\}$$

and then

$$\max_{1\leqslant r<s\leqslant m} d_2(H_{[r,s]}\mathbf{x}, H_{[r,s]}\mathbf{y}) = \max\left\{\log\left(\frac{[H\mathbf{x}]_r}{[H\mathbf{x}]_s}\frac{[H\mathbf{y}]_s}{[H\mathbf{y}]_r}\right)\Bigg|\, r, s = 1, 2, \ldots, m,\ r \neq s\right\} = d_m(\mathbf{x}, \mathbf{y});$$

see (1.1). Therefore, when $d_n(\mathbf{x}, \mathbf{y}) > 0$

$$\frac{d_m(H\mathbf{x}, H\mathbf{y})}{d_n(\mathbf{x}, \mathbf{y})} = \max_{1\leqslant r<s\leqslant m} \frac{d_2(H_{[r,s]}\mathbf{x}, H_{[r,s]}\mathbf{y})}{d_n(\mathbf{x}, \mathbf{y})}$$

and (2.14) follows from (1.2)

(ii) For each $\mathbf{x}, \mathbf{y} \in \mathscr{P}_n$, the equalities

$$\frac{[D\mathbf{x}]_r}{[D\mathbf{x}]_s}\frac{[D\mathbf{y}]_s}{[D\mathbf{x}]_r} = \frac{\delta_r x_r}{\delta_s x_s}\frac{\delta_s y_s}{\delta_r x_r} = \frac{x_r}{x_s}\frac{y_s}{x_r}$$

are always valid, and (1.1) yields that

$$d_n(D\mathbf{x}, D\mathbf{y}) = d_n(\mathbf{x}, \mathbf{y}). \tag{2.15}$$

Moreover, observing that $\mathbf{x} \mapsto D\mathbf{x}$ is a one-to-one function from $\mathscr{P}_n$ onto itself, it follows that

$$\tau(H) = \sup_{\mathbf{x},\mathbf{y}\in\mathscr{P}_n:d_n(\mathbf{x},\mathbf{y})>0} \frac{d_m(H\mathbf{x}, H\mathbf{y})}{d_n(\mathbf{x}, \mathbf{y})} = \sup_{\mathbf{x},\mathbf{y}\in\mathscr{P}_n:d_n(\mathbf{x},\mathbf{y})>0} \frac{d_m(HD\mathbf{x}, HD\mathbf{y})}{d_n(D\mathbf{x}, D\mathbf{y})}$$

and, together with (2.15) and (1.2), this leads to $\tau(H) = \tau(HD)$. $\square$

3. Proof of Theorem 1.1

Let H be a matrix as in the statement of Theorem 1.1. To establish formula (1.3) consider the following two exhaustive cases:

• $m = 2$. In this framework let D be the diagonal matrix with elements $1/h_{2j}$ along the main diagonal and observe that

$$HD = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \\ 1 & 1 & \cdots & 1 \end{bmatrix}, \quad \text{where } a_j = \frac{h_{1j}}{h_{2j}}, \quad j = 1, 2, \ldots, n.$$

On the other hand, from Lemma 2.3(ii) and Lemma 2.2, it follows that

$$\tau(H) = \tau(HD) = \frac{1 - \sqrt{a/A}}{1 + \sqrt{a/A}}, \tag{3.1}$$

where the notation is as in (2.2). Observing that the expression for $\phi(H)$ in (1.3) can be written as

$$\phi(H) = \min\left\{\frac{a_j}{a_k}, \frac{a_k}{a_j} \,\middle|\, j, k = 1, 2, \ldots, n\right\},$$

it follows that $\phi(H) = a/A$ and, together with (3.1), this establishes Birkhoff's formula.

• $m > 2$. In this context, Lemma 2.3(i) and the case $m = 2$ already established together yield

$$\tau(H) = \max_{r,s:1 \leqslant r < s \leqslant m} \tau(H_{[r,s]}) = \max_{r,s:1 \leqslant r < s \leqslant m} \frac{1 - \phi(H_{[r,s]})^{1/2}}{1 + \phi(H_{[r,s]})^{1/2}}.$$

Observing that the mapping $\alpha \mapsto (1 - \sqrt{\alpha})/(1 + \sqrt{\alpha})$ is decreasing this equality implies that

$$\tau(H) = \frac{1 - \psi(H)^{1/2}}{1 + \psi(H)^{1/2}}, \quad \text{where } \psi(H) = \min_{r,s:1 \leqslant r < s \leqslant m} \phi(H_{[r,s]}).$$

Using the expression for $\phi(H)$ in (1.3) it is not difficult to see that $\psi(H) = \phi(H)$, and Birkhoff's formula follows in the general case. □

References

[1] G. Birkhoff, Extensions of Jentzsch's theorem, Trans. Am. Math. Soc. 85 (1957) 219–227.

[2] T. Bielecki, D. Hernández–Hernández, S. Pliska, Risk sensitive control of finite state Markov chains in discrete time, with applications to portfolio management, Math. Methods Oper. Res. 50 (1999) 167–188.

[3] S.P. Evson, R.D. Nussbaum, An elementary proof of the Birkhoff–Hopf theorem, Math. Proc. Cambridge Philos. Soc. 117 (1995) 31–55.

[4] M.A. Krasnosel'skij, J.A. Lifshits, A.V. Sobolev, Positive Linear Systems—The Method of Positive Operators, Heldermann Verlag, New York, 1989.

[5] E. Seneta, Non–Negative Matrices and Markov Chains, Springer-Verlag, New York, 1980.