A Matrix Kronecker Lemma*

Brian D. O. Anderson and John B. Moore
Department of Electrical Engineering
University of Newcastle
New South Wales, 2308, Australia

Submitted by Ingram Olkin

ABSTRACT

We show that a standard tool of probability theory, the Kronecker lemma, has matrix generalizations, but that one of these matrix generalizations is unsatisfactory, to the extent that unless certain extra conditions are placed on the matrix sequence appearing in the lemma statement, the lemma may fail to be true.

1. INTRODUCTION

A lemma due to Kronecker is a standard tool in probability theory; see [1, 2] for proof and applications of the lemma. A statement of the lemma is as follows:

KRONECKER LEMMA. *Let a_k be a sequence of real numbers for which $|\sum_{k=1}^{\infty} a_k| < \infty$, and q_k a monotone increasing sequence of positive real numbers such that $q_k \to \infty$ as $k \to \infty$. Then*

$$\lim_{n\to\infty} q_n^{-1}\left(\sum_{k=1}^{n} a_k q_k\right) = 0.$$

In examining a conjecture in martingale theory, the precise form of which is irrelevant here, we were led to seek a matrix generalization of the lemma. Section 2 states two forms of such a generalization, and it proves convenient in one of these generalizations to impose restrictions on the condition numbers of the members Q_k of the sequence of positive definite symmetric

*Work supported by the Australian Research Grants Committee.

matrices which replaces in the matrix version of the lemma the scalar sequence q_k of the scalar version. In Sec. 3, we show that if the condition number constraint is violated, there are sequences for which the lemma fails but there are still sequences for which it is true. The question is then raised as to whether or not there is a matrix version of the Kronecker lemma for which weaker restrictions are imposed on the Q_k matrices.

2. MATRIX KRONECKER LEMMA

We adopt the following notational conventions. All vectors and matrices will be real. A prime denotes matrix transposition. For a symmetric matrix A, $\lambda_{\max}(A)$ and $\lambda_{\min}(A)$ denote the maximum and minimum eigenvalues of A; $A>0$ $(A\geqslant 0)$ denotes that A is positive (nonnegative) definite; and for symmetric B, $A>B$ denotes that $A-B>0$ $(A\geqslant B$ that $A-B\geqslant 0)$. For a vector x, $\|x\|=(x'x)^{1/2}$, so that $\|A\|$ for any A is $[\lambda_{\max}(A'A)]^{1/2}$. For symmetric nonnegative A, $\|A\|=\lambda_{\max}(A)$, and if A is also nonsingular, $\|A^{-1}\|=\lambda_{\min}^{-1}(A)$; the condition number is $\|A\|\,\|A^{-1}\|$, or $\lambda_{\max}(A)/\lambda_{\min}(A)$. Finally, for arbitrary square A, $\lambda_i(A)$ denotes an eigenvalue of A.

In this section, we prove the following result:

THEOREM 2.1. *Let a_k be a sequence of real p-vectors for which $\|\sum_{k=1}^{\infty}a_k\|<\infty$, and let Q_k be a monotone increasing sequence of $p\times p$ nonnegative definite real symmetric matrices (i.e., Q_k-Q_{k-1} is nonnegative definite for all k) such that* $(\operatorname{trace} Q_k)^{-1}\to 0$ *as $k\to\infty$. Then*

$$\lim_{n\to\infty}(\operatorname{trace} Q_n)^{-1}\sum_{k=1}^{n}Q_k a_k=0. \tag{2.1}$$

If Q_n is nonsingular for all n and Q_n^{-1} trace Q_n *is bounded or (equivalently) $\lambda_{\max}(Q_n)/\lambda_{\min}(Q_n)$ is bounded, then*

$$\lim_{n\to\infty}Q_n^{-1}\sum_{k=1}^{n}Q_k a_k=0. \tag{2.2}$$

Notice that (2.1) could just as well be written with $\lambda_{\max}(Q_n)$ replacing trace Q_n.

Proof. By and large, we follow the proof for the scalar case. Set

$r_n = \sum_{k=n}^{\infty} a_k$. Then

$$(\operatorname{trace} Q_n)^{-1} \sum_{k=1}^{n} Q_k a_k = (\operatorname{trace} Q_n)^{-1} \sum_{k=1}^{n} Q_k (r_k - r_{k+1})$$

$$= (\operatorname{trace} Q_n)^{-1} \left[\sum_{k=1}^{n} (Q_k - Q_{k-1}) r_k - Q_n r_{n+1} \right]$$

$$= (\operatorname{trace} Q_n)^{-1} \sum_{k=1}^{N_0} (Q_k - Q_{k-1}) r_k$$

$$+ (\operatorname{trace} Q_n)^{-1} \sum_{N_0+1}^{n} (Q_k - Q_{k-1}) r_k$$

$$- (\operatorname{trace} Q_n)^{-1} Q_n r_{n+1}. \tag{2.3}$$

We take $Q_0 = 0$ in the second and third equalities. Now choose N_0 such that $\|r_n\| < \epsilon$ for all $n \geqslant N_0$. Then

$$\left\| \sum_{N_0+1}^{n} (Q_k - Q_{k-1}) r_k \right\| \leqslant \sum_{N_0+1}^{n} \|Q_k - Q_{k-1}\| \, \|r_k\|$$

$$\leqslant \sum_{N_0+1}^{n} \operatorname{trace}(Q_k - Q_{k-1})\epsilon$$

$$\leqslant \operatorname{trace} Q_n \epsilon.$$

So (2.3) yields for all $n \geqslant N_0$

$$\left\| (\operatorname{trace} Q_n)^{-1} \sum_{k=1}^{n} Q_k a_k \right\| \leqslant (\operatorname{trace} Q_n)^{-1} \left\| \sum_{k=1}^{N_0} (Q_k - Q_{k-1}) r_k \right\| + 2\epsilon$$

Letting $n \to \infty$ and using the arbitrariness of ϵ establishes (2.1). To establish (2.2), observe that premultiplication of (2.1) by $Q_n^{-1} \operatorname{trace} Q_n$ yields that (2.2) is satisfied if $Q_n^{-1} \operatorname{trace} Q_n$ is bounded. This is equivalent to boundedness of $\lambda_{\max}(Q_n)/\lambda_{\min}(Q_n)$, since

$$\frac{\lambda_{\max}(Q_n)}{\lambda_{\min}(Q_n)} \leqslant \|Q_n^{-1}\| \operatorname{trace} Q_n \leqslant p \frac{\lambda_{\max}(Q_n)}{\lambda_{\min}(Q_n)}.$$

■

3. NECESSITY OF THE CONDITION NUMBER ASSUMPTION

The proof of the second part of the Theorem 2.1 used the assumption that $\lambda_{\max}(Q_n)/\lambda_{\min}(Q_n)$ is bounded. One might suspect that the inequalities used in deducing (2.2) are simply too coarse to establish the theorem when the condition number assumption does not hold, while more sophisticated inequalities might do the trick. In this section, we shall show that if the condition number assumption fails, one can construct sequences a_k and Q_k satisfying the remaining condition of the theorem for which the quantity on the left side of (2.2) actually diverges. First however, we shall show trivially that (2.2) may hold even though the condition number assumption fails.

We take $a_k=[\alpha_k \alpha_k]'$, where α_k is a sequence of scalars with $\sum_1^\infty \alpha_k$ convergent, and $Q_k=\operatorname{diag}[kq_k, q_k]$, where $q_{k+1} \geqslant q_k > 0$ for all k, with $q_k \to \infty$ as $k \to \infty$. Then the matrix problem is essentially two decoupled scalar problems, and (2.2) is certainly true. On the other hand, $\lambda_{\max}(Q_k)/\lambda_{\min}(Q_k) = k$.

The construction of a counterexample to (2.2) will depend on the following proposition, the proof of which will proceed with the aid of two lemmas. The results of these lemmas are well known to many, but they are included for completeness.

Proposition 3.1. *There exist symmetric K and L with $0 < L < K$ such that $\lambda_{\max}(LK^{-2}L)$ takes on an arbitrary positive value.*

First we show that $\|N\|$ can be large without the eigenvalues of N necessarily being large.

Lemma 3.1. *There exist diagonalizable N with $0 < \lambda_i(N) < 1$ for all i and with $\lambda_{\max}(N'N)$ arbitrarily large.*

Proof. Let $\tilde{N} = \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix}$ with $0 < a < b < 1$. Then the two eigenvalues of $\tilde{N}$ are a and b. Let $M = \begin{bmatrix} 1 & -m \\ 0 & 1 \end{bmatrix}$, $m > 0$. Then if $N \equiv (n_{ij}) = M^{-1}\tilde{N}M$, we get $n_{11}=a$, $n_{22}=b$ and $n_{12}=m(b-a)$. Hence n_{12} can be made arbitrarily large by choice of m. Since for any $p \times p$ matrix $N=(n_{ij})$, $\operatorname{trace} N'N = \sum_{i,j}^{p} n_{ij}^2 \geqslant \lambda_{\max}(N'N) \geqslant p^{-1} \operatorname{trace} N'N$, it follows that $\lambda_{\max}(N'N)$ can be made arbitrarily large. ∎

Now we relate symmetric matrix pairs of the type occurring in Proposition 3.1 to diagonalizable matrices with eigenvalues lying in $(0,1)$.

LEMMA 3.2. *Let N be a $p \times p$ matrix. Then N is diagonalizable with $0 < \lambda_i(N) < 1$ for all i if and only if $N = K^{-1}L$ for symmetric K and L with $0 < L < K$.*

Proof. Assume N is diagonalizable with $0 < \lambda_i(N) < 1$ for all i. Set $N = T^{-1}\Lambda T$ with $\Lambda = \mathrm{diag}\lambda_i$. Then $N = (T'T)^{-1}(T'\Lambda T)$ and $K = T'T$, $L = T'\Lambda T$. Conversely, suppose $N = K^{-1}L$ with $0 < L < K$. Then $0 < K^{-\frac{1}{2}}LK^{-\frac{1}{2}} < I$, so $0 < \lambda_i(K^{-\frac{1}{2}}LK^{-\frac{1}{2}}) < 1$ for all i. But $K^{\frac{1}{2}}NK^{-\frac{1}{2}} = K^{-\frac{1}{2}}LK^{-\frac{1}{2}}$, so that $0 < \lambda_i(K^{\frac{1}{2}}NK^{-\frac{1}{2}}) = \lambda_i(N) < 1$ for all i. Moreover, $K^{\frac{1}{2}}NK^{-\frac{1}{2}}$, being symmetric, is diagonalizable, and therefore N is diagonalizable. ■

Lemmas 3.1 and 3.2 now tie together to establish the Proposition.

Proof of Proposition. We explain how to construct K, L with $\lambda_{\max}(LK^{-2}L) = \delta$ for arbitrary positive δ. If $\delta < 1$, take $L = \sqrt{\delta}\, I$, $K = I$. If $\delta \geqslant 1$, proceed as follows. Take $0 < a < b < 1$, and construct N as described in the proof of Lemma 3.1. The eigenvalues of $N'N$ vary continuously with m, and approach a, b when $m \to 0$, while $\lambda_{\max}(N'N)$ can be made arbitrarily large. Hence by appropriate choice of m, $\lambda_{\max}(N'N)$ can take any value in (b, ∞) and in particular any value $\delta \in [1, \infty)$. Now construct K, L as in Lemma 3.2 with $N = K^{-1}L$; then $\lambda_{\max}(LK^{-2}L) = \delta$. ■

We turn now to the construction of a counterexample to (2.2). Observing that, with $Q_0 = 0$,

$$Q_{2n}^{-1}\sum_{k=1}^{2n} Q_k a_k = Q_{2n}^{-1}\sum_{k=1}^{2n} Q_k(r_k - r_{k+1})$$

$$= Q_{2n}^{-1}\sum_{k=1}^{2n}(Q_k - Q_{k-1})r_k - r_{2n+1} \qquad (3.1)$$

(where $r_n = \Sigma_n^{\infty} a_k$), we see that if we can establish for one set of a_k, Q_k that the first quantity on the right side of (3.1) diverges, then (because $\|r_n\| \to 0$ as $n \to \infty$) the quantity on the left side of (3.1) will diverge. In order to prove the divergence, our strategy will be to select the Q_k sequence so that in the summation $Q_{2n}^{-1}\Sigma_{k=1}^{2n}(Q_k - Q_{k-1})r_k$, the last term, viz. $Q_{2n}^{-1}(Q_{2n} - Q_{2n-1})r_{2n}$, is dominant. (Basically, this is done by having the differences $Q_k - Q_{k-1}$ grow suitably fast.) We shall also arrange that the maximum eigenvalue of $(Q_{2n} - Q_{2n-1})Q_{2n}^{-2}(Q_{2n} - Q_{2n-1})$ grows with n at a rate faster than $\|r_{2n}\|$

decreases with n. Then by aligning r_{2n} with the eigenvector associated with this maximum eigenvalue, we ensure that $\|Q_{2n}^{-1}(Q_{2n}-Q_{2n-1})r_{2n}\|$ grows with n. Of course, in doing this, we must ensure the various side conditions on the Q_k and a_k sequences as provided in the Theorem statement are all fulfilled.

Construction of Q_k sequence. Let $\tilde{L}_n$, $\tilde{K}_n$ be symmetric matrices existing by Proposition 3.1 such that $0<\tilde{L}_n<\tilde{K}_n$ and $\lambda_{\max}(\tilde{L}_n\tilde{K}_n^{-2}\tilde{L}_n)=n^6$. Let $L_i=\alpha_i\tilde{L}_i$, $K_i=\alpha_i\tilde{K}_i$, where $\alpha_1=1$, and $\alpha_2,\alpha_3,\alpha_4,\ldots$ are chosen sequentially such that

$$\lambda_{\min}(K_n)\geqslant n^2\lambda_{\max}(K_{n-1}),$$

$$K_n-L_n\geqslant K_{n-1}.$$

Then set $Q_{2n}=K_n$, $Q_{2n-1}=K_n-L_n$. Observe that $Q_{2n}-Q_{2n-1}=L_n>0$ and $Q_{2n-1}-Q_{2n-2}=K_n-L_n-K_{n-1}\geqslant 0$. Further, in view of the eigenvalue inequality on the sequence K_n, we have $Q_{2n}^{-1}\to 0$ as $n\to\infty$. Notice that the condition numbers of the Q_{2n} sequence diverge, since

$$n^6=\|L_nK_n^{-2}L_n\|\leqslant\|L_n\|\,\|K_n^{-1}\|^2\|L_n\|\leqslant\|K_n\|^2\|K_n^{-1}\|^2$$

$$=\left[\frac{\lambda_{\max}(Q_{2n})}{\lambda_{\min}(Q_{2n})}\right]^2.$$

Construction of the a_n sequence. With $r_k=\sum_{j=k}^{\infty}a_j$, define r_{2n} by n^2r_{2n} = eigenvector of unit length corresponding to the maximum eigenvalue of $L_nK_n^{-2}L_n$, and define $r_{2n+1}=0$. It is easy to construct the a_k sequence and to verify that $\|\sum_{k=0}^{\infty}a_k\|<\infty$.

Now observe that

$$\left\|Q_{2n}^{-1}\sum_{j=1}^{2n}(Q_j-Q_{j-1})r_j\right\|=\left\|K_n^{-1}\sum_{k=1}^{n}L_kr_{2k}\right\| \tag{3.2}$$

Define this quantity as $g(n)$. In the light of the remarks following (3.1), observe that our task is now one of showing that the right hand side of (3.2) is approximately $\|K_n^{-1}L_n r_{2n}\|$, and that this quantity diverges with n. Now

$$\|K_n^{-1}L_n r_{2n}\| - \left\|K_n^{-1}\sum_{k=1}^{n-1} L_k r_{2k}\right\| \leqslant \left\|K_n^{-1}\sum_{k=1}^{n} L_k r_{2k}\right\| = g(n)$$

$$\leqslant \|K_n^{-1}L_n r_{2n}\| + \left\|K_n^{-1}\sum_{k=1}^{n-1} L_k r_{2k}\right\|$$

The first term in the first and third members of the inequality is n, as a result of the definitions of $\lambda_{\max}(L_n K_n^{-2} L_n)$ and r_{2n}. As for the second term, we have

$$\left(\sum_{k=1}^{n-1} L_k r_{2k}\right)' K_n^{-2}\left(\sum_{k=1}^{n-1} L_k r_{2k}\right) \leqslant \frac{1}{\lambda_{\min}^2(K_n)}\left\|\sum_{k=1}^{n-1} L_k r_{2k}\right\|^2$$

$$\leqslant \frac{1}{n^4\lambda_{\max}^2(K_{n-1})}\left\|\sum_{k=1}^{n-1} L_k r_{2k}\right\|^2$$

$$\leqslant \frac{1}{n^4}\left\|K_{n-1}^{-1}\sum_{k=1}^{n-1} L_k r_{2k}\right\|^2,$$

so that

$$\left\|K_n^{-1}\sum_{k=1}^{n-1} L_k r_{2k}\right\| \leqslant \frac{1}{n^2}\left\|K_{n-1}^{-1}\sum_{k=1}^{n-1} L_k r_{2k}\right\| = \frac{1}{n^2}\, g(n-1).$$

Thus we have

$$n - \frac{1}{n^2}\, g(n-1) \leqslant g(n) \leqslant n + \frac{1}{n^2}\, g(n-1). \tag{3.3}$$

This inequality suggests that $g(n)$ should grow at a rate of n. In fact we shall prove (by induction) that

$$\tfrac{1}{2}n < g(n) < 2n \qquad \text{for all } n. \tag{3.4}$$

First, $g(1) = \|K_1^{-1}L_1 r_2\| = 1$ by construction of r_2. Next, assume that (3.4) holds for $n=1,2,\ldots,r-1$. From (3.3), we have

$$r-\frac{2(r-1)}{r^2}\leqslant g(r)\leqslant r+\frac{2(r-1)}{r^2},$$

whence

$$r-\frac{2}{r}<g(r)<r+\frac{2}{r},$$

and then

$$\frac{r}{2}<g(r)<2r,$$

as required. In fact, a refinement of the above argument shows that $n^{-1}g(n)\to 1$ as $n\to\infty$.

The two results of this section suggest that some other condition than boundedness of $\lambda_{\max}(Q_k)/\lambda_{\min}(Q_k)$ might prove more effective for delineating those situations in which the Kronecker lemma holds. Not only does the condition number have to be unbounded for the lemma to fail, but also the sequence of orthogonal matrices T_k such that $T_k'Q_kT_k$ is diagonal must not be constant; it might be conjectured that some minimum rate of variation of T_k, determined somehow by the sequence of condition numbers, would divide the situations in which the lemma fails from those in which it is true. However, some further condition on the a_k will also be needed.

REFERENCES

1 H. G. Tucker, *A Graduate Course in Probability*, Academic, New York, 1967.
2 J. Neveu, *Mathematical Foundations of the Calculus Probability*, Holden-Day, San Francisco, 1965.

Received 13 May 1975; revised 1 December 1975