An Inequality for a Sum of Quadratic Forms with Applications to Probability Theory*

T. W. Anderson
Department of Statistics
Stanford University
Stanford, California 94305

and

John B. Taylor
Department of Economics
Columbia University
New York, New York 10027

Submitted by Hans Schneider

ABSTRACT

Let $\{\mathbf{x}_t\}$ be a sequence of p-component vectors, and let $\mathbf{A}_t=\sum_{s=1}^{T}\mathbf{x}_t\mathbf{x}_t'$ with $\mathbf{A}_n$ nonsingular for some $n\geqslant p$. It is shown that $\sum_{t=q+1}^{t}\mathbf{x}_t'\mathbf{A}_t^{-2}\mathbf{x}_t\leqslant \operatorname{tr}\mathbf{A}_q^{-1}$ for $n\leqslant q<T$. An application of this proposition is the convergence of a certain martingale with probability 1. A matrix version of Kronecker's lemma then leads to strong consistency of least-squares estimates under a certain condition.

Consider a sequence of real numbers $\{x_t\}$, $t=1,2,\dots,$ with $x_n\neq 0$ for some $n\geqslant 1$. Then for $T>q\geqslant n$

$$\sum_{t=q+1}^{T}\left(\sum_{s=1}^{t}x_s^2\right)^{-2}x_t^2\leqslant\left(\sum_{s=1}^{q}x_s^2\right)^{-1}. \tag{1}$$

The inequality (1) has a number of applications in statistical estimation and control theory (see Anderson and Taylor [3,4] and Taylor [8], for example).

*This work is supported by National Science Foundation Grants SOC77-14944-A01 at the Institute for Mathematical Studies in the Social Sciences, Stanford University, and SOC78-07302 at Columbia University.

0024-3795/80/020093+7$01.75

Here we prove a vector generalization of (1) and show that, when combined with a matrix generalization of the Kronecker lemma, it provides a straightforward alternative method of proving multivariate generalizations of these statistical applications. The generalization of the Kronecker lemma involves a condition-number assumption, and we provide an elementary counterexample to show that this assumption is necessary. However, the counterexample does not apply to the statistical applications themselves; hence the question is open as to whether alternative vector generalizations of (1) and the Kronecker lemma can lead to further extensions of these results. In a recent paper, B. D. O. Anderson and J. B. Moore [1] have provided an alternative counterexample to show the necessity of the condition-number assumption in the same generalization of the Kronecker lemma.

The vector generalization of (1) is contained in the following

PROPOSITION 1. *Let* $\{\mathbf{x}_t\}$ *be a sequence of p-component vectors such that for some* $n \geqslant p$, $(\sum_{s=1}^{n}\mathbf{x}_s\mathbf{x}_s')^{-1}$ *exists. Then for* $T > q \geqslant n$,

$$\sum_{t=q+1}^{T} \mathbf{x}_t'\left(\sum_{s=1}^{t}\mathbf{x}_s\mathbf{x}_s'\right)^{-2}\mathbf{x}_t \leqslant \operatorname{tr}\left(\sum_{s=1}^{q}\mathbf{x}_s\mathbf{x}_s'\right)^{-1}. \tag{2}$$

Proof. Let $\mathbf{A}_t = \sum_{s=1}^{t}\mathbf{x}_s\mathbf{x}_s'$, and for a given $t \geqslant q+1$, let $\mathbf{A}_t = \mathbf{HDH}'$ and $\mathbf{A}_{t-1} = \mathbf{HH}'$, where $\mathbf{D}$ is a diagonal matrix. From $\mathbf{HDH}' = \mathbf{HH}' + \mathbf{x}_t\mathbf{x}_t'$ we have $\mathbf{D} = \mathbf{I} + \mathbf{yy}'$, where $\mathbf{y} = \mathbf{H}^{-1}\mathbf{x}_t$. Because the rank of $\mathbf{yy}'$ is at most one, and because $\mathbf{D}$ is diagonal, only one element of $\mathbf{y}$ can be nonzero. Let this be the ith element. Then all elements on the diagonal of $\mathbf{D}$ are equal to one except for the ith, which can be greater than one.

Let $\mathbf{z} = \mathbf{D}^{-1/2}\mathbf{y}$ and $\mathbf{G} = \mathbf{H}^{-1}(\mathbf{H}')^{-1}$; then

$$\begin{aligned} \mathbf{x}_t'\mathbf{A}_t^{-2}\mathbf{x}_t &= \mathbf{x}_t'(\mathbf{HDH}')^{-2}\mathbf{x}_t \\ &= \mathbf{z}'\mathbf{D}^{-1/2}\mathbf{G}\mathbf{D}^{-1/2}\mathbf{z} \\ &= z_i^2 g_{ii} d_{ii}^{-1}, \end{aligned} \tag{3}$$

where z_i is the ith element of $\mathbf{z}$ (all other elements being equal to zero) and where g_{ii} and d_{ii} are the ith diagonal elements of $\mathbf{G}$ and $\mathbf{D}$, respectively.

Further

$$\begin{aligned}\operatorname{tr}\left(\mathbf{A}_{t-1}^{-1}-\mathbf{A}_t^{-1}\right)&=\operatorname{tr}\left[(\mathbf{H}')^{-1}\mathbf{H}^{-1}-(\mathbf{H}')^{-1}\mathbf{D}^{-1}\mathbf{H}^{-1}\right]\\&=\operatorname{tr}(\mathbf{H}')^{-1}\mathbf{D}^{-1/2}\mathbf{H}^{-1}\left[\mathbf{HDH}'-\mathbf{HH}'\right](\mathbf{H}')^{-1}\mathbf{D}^{-1/2}\mathbf{H}^{-1}\\&=\operatorname{tr}(\mathbf{HD}^{1/2}\mathbf{H}')^{-1}\mathbf{x}_t\mathbf{x}_t'(\mathbf{HD}^{1/2}\mathbf{H}')^{-1}\\&=\mathbf{x}_t'(\mathbf{HD}^{1/2}\mathbf{H}')^{-2}\mathbf{x}_t\\&=\mathbf{z}'\mathbf{Gz}\\&=z_i^2g_{ii}.\end{aligned}\tag{4}$$

Because $d_{ii}\geqslant 1$, we have $z_i^2g_{ii}d_{ii}^{-1}\leqslant z_i^2g_{ii}$, and therefore from (3) and (4)

$$\mathbf{x}_t'\mathbf{A}_t^{-2}\mathbf{x}_t\leqslant\operatorname{tr}\left(\mathbf{A}_{t-1}^{-1}-\mathbf{A}_t^{-1}\right).\tag{5}$$

Summing both sides of (5) over t gives

$$\begin{aligned}\sum_{t=q+1}^{T}\mathbf{x}_t'\mathbf{A}_t^{-2}\mathbf{x}_t&\leqslant\sum_{t=q+1}^{T}\operatorname{tr}\left(\mathbf{A}_{t-1}^{-1}-\mathbf{A}_t^{-1}\right)\\&=\operatorname{tr}\mathbf{A}_q^{-1}-\operatorname{tr}\mathbf{A}_T^{-1}\\&\leqslant\operatorname{tr}\mathbf{A}_q^{-1}.\qquad\blacksquare\end{aligned}\tag{6}$$

The following proposition is a probabilistic application of Proposition 1 which is used in the statistical results reported below.

PROPOSITION 2. *Let $\mathbf{x}_1,\mathbf{x}_2,\ldots$ and $\mathbf{u}_1,\mathbf{u}_2,\ldots$ be two sequences of random vectors with p and m components, respectively. Let $\mathcal{F}_t$ be the σ-algebra generated by $(\mathbf{u}_1,\mathbf{u}_2,\ldots,\mathbf{u}_t,\mathbf{x}_1,\ldots,\mathbf{x}_{t+1})$, $t=1,2,\ldots,$ and let $\mathcal{F}_0$ be the σ-algebra generated by $\mathbf{x}_1$. Suppose that $\mathcal{E}(\mathbf{u}_t|\mathcal{F}_{t-1})=0$ and $\mathcal{E}(\mathbf{u}_t\mathbf{u}_t'|\mathcal{F}_{t-1})=\mathbf{\Sigma}$, and that $(\sum_{s=1}^{q}\mathbf{x}_s\mathbf{x}_s')$ is nonsingular with probability one and that $\mathcal{E}\,tr(\sum_{s=1}^{q}\mathbf{x}_s\mathbf{x}_s')^{-1}$ exists for some $q\geqslant p$. Then*

$$\sum_{t=q+1}^{T}\left(\sum_{s=1}^{t}\mathbf{x}_s\mathbf{x}_s'\right)^{-1}\mathbf{x}_t\mathbf{u}_t'\tag{7}$$

converges with probability one.

Proof. The ith column of (7) is

$$\mathbf{z}_T = \sum_{t=q+1}^{T} \left(\sum_{s=1}^{t} \mathbf{x}_s \mathbf{x}_s' \right)^{-1} \mathbf{x}_t u_{it}, \tag{8}$$

where u_{it} is the ith element of $\mathbf{u}_t$. The conditional expectation of the tth vector in the sum (8) is

$$\mathcal{E}\left[\left(\sum_{s=1}^{t} \mathbf{x}_s \mathbf{x}_s' \right)^{-1} \mathbf{x}_t u_{it} \middle| \mathcal{F}_{t-1} \right] = \left(\sum_{s=1}^{t} \mathbf{x}_s \mathbf{x}_s' \right)^{-1} \mathbf{x}_t \mathcal{E}(u_{it}|\mathcal{F}_{t-1})$$

$$= \mathbf{0} \tag{9}$$

with probability one, since each component of $(\Sigma_{s=1}^{t} \mathbf{x}_s \mathbf{x}_s')^{-1} \mathbf{x}_t$ is bounded with probability one. Thus the sum is a martingale (with expected value zero). The covariance matrix of the sum is

$$\begin{aligned} \mathcal{E}\,\mathbf{z}_T \mathbf{z}_T' &= \mathcal{E} \sum_{r,s=q+1}^{T} \mathbf{A}_r^{-1} \mathbf{x}_r \mathbf{x}_s' \mathbf{A}_s^{-1} u_{ir} u_{is} \\ &= \mathcal{E}\left[\sum_{s=q+1}^{T} \mathbf{A}_s^{-1} \mathbf{x}_s \mathbf{x}_s' \mathbf{A}_s^{-1} u_{is}^2 + \sum_{r=q+2}^{T} \sum_{s=q+1}^{r-1} \mathbf{A}_r^{-1} \mathbf{x}_r \mathbf{x}_s' \mathbf{A}_s^{-1} u_{ir} u_{is} \right. \\ &\qquad \left. + \sum_{s=q+2}^{T} \sum_{r=q+1}^{s-1} \mathbf{A}_r^{-1} \mathbf{x}_r \mathbf{x}_s' \mathbf{A}_s^{-1} u_{ir} u_{is} \right], \\ &= \mathcal{E} \sum_{s=q+1}^{T} \mathbf{A}_s^{-1} \mathbf{x}_s \mathbf{x}_s' \mathbf{A}_s^{-1} \sigma_{ii}, \end{aligned} \tag{10}$$

where σ_{ii} is the ith diagonal element of Σ. The last equality follows from the fact that for $s<r$

$$\mathcal{E}\,\mathbf{A}_r^{-1} \mathbf{x}_r \mathbf{x}_s' \mathbf{A}_s^{-1} u_{ir} u_{is} = \mathcal{E}\,\mathbf{A}_r^{-1} \mathbf{x}_r \mathbf{x}_s' \mathbf{A}_s^{-1} u_{is} \mathcal{E}(u_{ir}|\mathcal{F}_{r-1}) = \mathbf{0}. \tag{11}$$

Then

$$\mathcal{E}\,\mathbf{z}_T' \mathbf{z}_T = \mathcal{E}\,\operatorname{tr} \mathbf{z}_T \mathbf{z}_T' = \sigma_{ii} \mathcal{E}\,\operatorname{tr} \sum_{s=q+1}^{T} \mathbf{x}_s' \mathbf{A}_s^{-2} \mathbf{x}_s. \tag{12}$$

By Proposition 1 the right-hand side of (12) is bounded because, by assumption, $\mathcal{E}\,\mathrm{tr}\,A_q^{-1} < \infty$. Thus each component of $\mathbf{z}_T$ is a martingale with bounded second moment and, by the martingale convergence theorem, converges with probability one (Feller [5, p. 236]). ■

The statistical application of Proposition 1 requires the following generalization of the Kronecker lemma, which we state without proof. (B. D. O. Anderson and J. B. Moore [1] and T. W. Anderson and J. B. Taylor [2] have stated and proved this result.)

PROPOSITION 3. *Let* $\{\mathbf{z}_t\}$ *and* $\{\mathbf{x}_t\}$ *be two sequences of p-component vectors. Define* $\mathbf{A}_T = \sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t'$, *and suppose that* $\mathbf{A}_q$ *is nonsingular for some* $q \geqslant p$. *If*

(i) $\sum_{t=q+1}^{\infty}\mathbf{A}_t^{-1}\mathbf{z}_t$ *exists,*
(ii) $\lim_{T\to\infty}\mathbf{A}_T^{-1} = \mathbf{0}$, *and*
(iii) *the ratio of the largest to the smallest characteristic root of* $\mathbf{A}_T$ *is bounded by a number* $M < \infty$ *independent of* T,

then

$$\lim_{T\to\infty}\mathbf{A}_T^{-1}\sum_{t=1}^{T}\mathbf{z}_t = \mathbf{0}.$$

The main statistical application of Proposition 1 is in providing an alternative proof of the result that the least-squares estimates of $\mathbf{B}$ in the model $\mathbf{y}_t = \mathbf{B}'\mathbf{x}_t + \mathbf{u}_t$ converge to $\mathbf{B}$ with probability one under conditions described in the following proposition. This proposition is proved in T. W. Anderson and J. B. Taylor [4] by a different technique which does not make use of the multivariate generalization of (1). Because of the analogy with the scalar case, this alternative proof may suggest ways to improve on the result by modifying the condition-number assumption used in Proposition 3.

PROPOSITION 4. *Let* $\mathbf{y}_t = \mathbf{B}'\mathbf{x}_t + \mathbf{u}_t$, *where* $\mathbf{B}$ *is a* $p\times m$ *matrix of parameters,* $\mathbf{x}_t$ *is a p-component stochastic vector, and* $\mathbf{u}_t$ *is an m-component stochastic vector,* $t = 1,2,\ldots$. *Let* $\mathcal{F}_t$ *be the* σ*-algebra generated by* $(\mathbf{u}_1,\mathbf{u}_2,\ldots,\mathbf{u}_t,\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_{t+1})$, $t=1,2,\ldots$, *and let* $\mathcal{F}_0$ *be the* σ*-algebra generated by* $\mathbf{x}_1$. *Let* $\mathbf{A}_t = \sum_{s=1}^{t}\mathbf{x}_s\mathbf{x}_s'$, *and suppose that* $\mathbf{A}_q$ *is nonsingular with probability one and that* $\mathcal{E}\,\mathrm{tr}\,A_q^{-1} < \infty$ *for some* $q \geqslant p$. *Define* $\hat{\mathbf{B}}_T = \mathbf{A}_T^{-1}\sum_{t=1}^{T}\mathbf{x}_t\mathbf{y}_t'$. *If*

(i) $\mathcal{E}(\mathbf{u}_t|\mathcal{F}_{t-1}) = \mathbf{0}$ *and* $\mathcal{E}(\mathbf{u}_t\mathbf{u}_t'|\mathcal{F}_{t-1}) = \mathbf{\Sigma}$, $t = 1,2,\ldots$, *with probability one,*
(ii) $\lim_{T\to\infty}\mathbf{A}_T^{-1} = \mathbf{0}$ *with probability one, and*

(iii) *the ratio of the largest to the smallest characteristic root of* $\mathbf{A}_T$ *is bounded uniformly in* T *with probability one,*

then $\lim_{T\to\infty}\hat{\mathbf{B}}_T=\mathbf{B}$ *with probability one.*

Proof. The ith column of $\hat{\mathbf{B}}-\mathbf{B}$ is $\mathbf{A}_T^{-1}\sum_{t=1}^{T}\mathbf{z}_{it}$, where $\mathbf{z}_{it}=\mathbf{x}_t u_{it}$ and where u_{it} is the ith element of $\mathbf{u}_t$. By Propositions 1 and 2, $\sum_{t=q+1}^{T}\mathbf{A}_t^{-1}\mathbf{z}_{it}$ converges with probability one for $i=1,\ldots,m$. Therefore, with conditions (ii) and (iii), Proposition 3 implies that $\mathbf{A}_T^{-1}\sum_{t=1}^{T}\mathbf{z}_{it}$ converges to zero with probability one for each $i=1,\ldots,m$. ■

A useful extension of Proposition 4 would be to weaken the condition-number assumption (iii). However, condition (iii) is necessary for Proposition 3, as the following simple example illustrates. (See also B. D. O. Anderson and J. B. Moore [1].) Since the example violates the structure of the statistical model in Proposition 4, it does not prove the necessity of (iii) for that result. Lai, Robbins, and Wei [7] have recently shown that condition (iii) may be omitted if $\mathbf{x}_t$ is nonstochastic. However, Lai and Robbins [6] have given an example to show that (i) and (ii) are not sufficient in the stochastic case.

Example. Let $\mathbf{x}_t'=(1,\alpha^t)$ for $\alpha>1$, and define $\mathbf{z}_t$ so that

$$\mathbf{v}_t=\sum_{s=2}^{t}\mathbf{A}_s^{-1}\mathbf{z}_s=\mathbf{v}+\begin{pmatrix}0\\ \rho^t\end{pmatrix}, \tag{13}$$

where $\mathbf{v}$ is a constant vector, $0<\rho<1$, and $\alpha\rho>1$. Then $\mathbf{v}_t$ converges to $\mathbf{v}$ and

$$\mathbf{A}_T^{-1}=\left[T\alpha^2(\alpha^2-1)^{-1}(\alpha^{2T}-1)-\alpha^2(\alpha-1)^{-2}(\alpha^T-1)^2\right]^{-1} \times\begin{bmatrix}\alpha^2\dfrac{\alpha^{2T}-1}{\alpha^2-1} & -\alpha\dfrac{\alpha^T-1}{\alpha-1}\\ -\alpha\dfrac{\alpha^T-1}{\alpha-1} & T\end{bmatrix} \tag{14}$$

converges to $\mathbf{0}$. Hence conditions (i) and (ii) of Proposition 3 are satisfied while condition (iii) is not. From $\mathbf{z}_t=\mathbf{A}_t(\mathbf{v}_t-\mathbf{v}_{t-1})$ we find

$$\mathbf{z}_t=\begin{bmatrix}\alpha\dfrac{(\rho\alpha)^{t-1}\alpha-\rho^{t-1}}{\alpha-1}\\ \alpha^2\dfrac{(\rho\alpha^2)^{t-1}\alpha^2-\rho^{t-1}}{\alpha^2-1}\end{bmatrix}(\rho-1) \tag{15}$$

and

$$\sum_{t=1}^{T} \mathbf{z}_t = \begin{bmatrix} \dfrac{\alpha}{\alpha-1}\left(\alpha\dfrac{(\rho\alpha)^T-1}{\rho\alpha-1} - \dfrac{\rho^T-1}{\rho-1}\right) \\ \dfrac{\alpha^2}{\alpha^2-1}\left(\alpha^2\dfrac{(\rho\alpha^2)^T-1}{\rho\alpha^2-1} - \dfrac{\rho^T-1}{\rho-1}\right) \end{bmatrix}(\rho-1). \tag{16}$$

From (15) and (16) it follows that the dominant term of the first component of $\mathbf{A}_T^{-1}\Sigma_{t=1}^{T}\mathbf{z}_t$ is of the order $(\rho\alpha)^{\mathrm{T}}/\mathrm{T}$, which diverges as $T\to\infty$, since $\rho\alpha>1$. Therefore, when condition (iii) of Proposition 3 is not satisfied, the sequence in question may not converge at all, let alone converge to zero. However, this example does not capture the structure of the $\mathbf{z}_t$ in Proposition 4; that is, $\mathbf{z}_t$ is not proportional to $\mathbf{x}_t$.

REFERENCES

1 B. D. O. Anderson and J. B. Moore, A matrix Kronecker lemma, *Linear Algebra and Appl.* 15:227–234 (1976).

2 T. W. Anderson and J. B. Taylor, Some theorems concerning strong consistency of least squares estimates in linear regression, Working Paper No. 47, Institute for Mathematical Studies in the Social Sciences, Stanford Univ., Aug. 1974.

3 T. W. Anderson and J. B. Taylor, Strong consistency of least squares estimates in normal linear regression, *Ann. Statist.* 4:788–790 (1976).

4 T. W. Anderson and J. B. Taylor, Strong consistency of least squares estimates in dynamic models, *Ann. Statist.* 7:484–489 (1979).

5 W. Feller, *An Introduction to Probability Theory and Its Applications*, Vol. II, Wiley, New York, 1966.

6 T. L. Lai and Herbert Robbins, Consistency and asymptotic efficiency of slope estimates in stochastic approximation schemes, unpublished technical report, 1978.

7 T. L. Lai, Herbert Robbins, and C. Z. Wei, Strong consistency of least squares estimates in multiple regression, *Proc. Nat. Acad. Sci. U.S.A.* 75:3034–3036 (1978).

8 J. B. Taylor, Asymptotic properties of multiperiod control rules in the linear regression model, *Internat. Econom. Rev.* 15:472–484 (1974).

Received 1 August 1978; revised 9 May 1979