# TRACE INEQUALITIES WITH APPLICATIONS TO ORTHOGONAL REGRESSION AND MATRIX NEARNESS PROBLEMS

I.D. COOPE AND P.F. RENAUD *

**Abstract.** Matrix trace inequalities are finding increased use in many areas such as analysis, where they can be used to generalise several well known classical inequalities, and computational statistics, where they can be applied, for example, to data fitting problems. In this paper we give simple proofs of two useful matrix trace inequalities and provide applications to orthogonal regression and matrix nearness problems.

**Key words.** Trace inequalities, projection matrices, total least squares, orthogonal regression, matrix nearness problems.

**AMS subject classifications.** 15A42, 15A45, 65F30

## 1. A Matrix Trace Inequality.

THEOREM 1.1. *Let $X$ be a $n \times n$ Hermitian matrix with* $\operatorname{rank}(X) = r$ *and let $Q_k$ be an $n \times k$ matrix, $k \leq r$, with $k$ orthonormal columns. Then, for given $X$,* $\operatorname{tr}(Q_k^* X Q_k)$ *is maximized when $Q_k = V_k$, where $V_k = [v_1, v_2, \ldots, v_k]$ denotes a matrix of $k$ orthonormal eigenvectors of $X$ corresponding to the $k$ largest eigenvalues.*

*Proof.* Let $X = VDV^*$ be a spectral decomposition of $X$ with $V$ unitary and $D = \operatorname{diag}[\lambda_1, \lambda_2, \ldots, \lambda_n]$, the diagonal matrix of (real) eigenvalues ordered so that

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n. \tag{1.1}$$

Then,

$$\operatorname{tr}(Q_k^* X Q_k) = \operatorname{tr}(Z_k^* D Z_k) = \operatorname{tr}(Z_k Z_k^* D) = \operatorname{tr}(PD), \tag{1.2}$$

where $Z_k = V^* Q_k$ and $P = Z_k Z_k^*$ is a projection matrix with $rank(P) = k$. Clearly, the $n \times k$ matrix $Z_k$ has orthonormal columns if and only if $Q_k$ has orthonormal columns. Now

$$\operatorname{tr}(PD) = \sum_{i=1}^{n} P_{ii} \lambda_i$$

*Department of Mathematics & Statistics, University of Canterbury, Private Bag 4800, Christchurch, New Zealand. ( I.Coope@math.canterbury.ac.nz, P.Renaud@math.canterbury.ac.nz).

with $0 \leq P_{jj} \leq 1,\ j = 1, 2, \ldots, n$ and $\sum_{i=1}^{n} P_{ii} = k$ because $P$ is an Hermitian projection matrix with rank $k$. Hence,

$$\operatorname{tr}(Q_k^* X Q_k) \leq L,$$

where $L$ denotes the maximum value attained by the linear programming problem:

$$\max_{p \in R^n} \left\{ \sum_{i=1}^{n} p_i \lambda_i \ :\ 0 \leq p_j \leq 1,\ j = 1, 2, \ldots, n;\ \sum_{i=1}^{n} p_i = k \right\}. \tag{1.3}$$

An optimal basic feasible solution to the LP problem (1.3) is easily identified (noting the ordering (1.1)) as $p_j = 1,\ j = 1, 2, \ldots, k;\ p_j = 0, j = k+1, k+2, \ldots, n,$ with $L = \sum_1^k \lambda_i$. But $P = E_k E_k^*$ gives $\operatorname{tr}(PD) = L$ where $E_k$ is the matrix with orthonormal columns formed from the first $k$ columns of the $n \times n$ identity matrix, therefore (1.2) provides the required result that $Q_k = V E_k = V_k$ maximizes $\operatorname{tr} Q_k^* X Q_k$. □

COROLLARY 1.2. *Let $Y$ be an $m \times n$ matrix with $m \geq n$ and* $\operatorname{rank}(Y) = r$ *and let $Q_k \in R^{n \times k}, k \leq r$, be a matrix with $k$ orthonormal columns. Then the Frobenius trace-norm $||YQ_k||_F^2 = \operatorname{tr}(Q_k^* Y^* Y Q_k)$ is maximized for given $Y$, when $Q = V_k$, where $USV^*$ is a singular value decomposition of $Y$ and $V_k = [v_1, v_2, \ldots, v_k] \in R^{n \times k}$ denotes a matrix of $k$ orthonormal right singular vectors of $Y$ corresponding to the $k$ largest singular values.*

COROLLARY 1.3. *If a minimum rather than maximum is required then substitute the $k$ smallest eigenvalues/singular values in the above results and reverse the ordering (1.1).*

Theorem 1.1 is a special case of a more general result established in Section 3. Alternative proofs can be found in some linear algebra texts (see, for example [3]). The special case above and the Corollary 1.2 have applications in total least squares data fitting.

**2. An Application to Data Fitting.** Suppose that data is available as a set of $m$ points in $R^n$ represented by the columns of the $n \times m$ matrix $A$ and it is required to find the best k-dimensional linear manifold $L_k \in R^n$ approximating the set of points in the sense that the sum of squares of the distances of each data point from its orthogonal projection onto the linear manifold is minimized. A general point in $L_k$ can be expressed in parametric form as

$$x(t) = z + Z_k t,\ t \in R^k, \tag{2.1}$$

where $z$ is a fixed point in $L_k$ and the columns of the $n \times k$ matrix $Z_k$ can be taken to be orthonormal. The problem is now to identify a suitable $z$ and $Z_k$. Now the orthogonal projection of a point $a \in R^n$ onto $L_k$ can be written as

$$\mathrm{proj}(a, L_k) = z + Z_k Z_k^T (a - z),$$

and hence the Euclidean distance from $a$ to $L_k$ is

$$\mathrm{dist}(a, L_k) = ||a - \mathrm{proj}(a, L_k)||_2 = ||(I - Z_k Z_k^T)(a - z)||_2.$$

Therefore, the total least squares data-fitting problem is reduced to finding a suitable $z$ and corresponding $Z_k$ to minimize the sum-of-squares function

$$SS = \sum_{j=1}^{m} ||(I - Z_k Z_k^T)(a_j - z)||_2^2,$$

where $a_j$ is the $j$th data point ($j$th column of $A$). A necessary condition for $SS$ to be minimized with respect to $z$ is

$$0 = \sum_{j=1}^{m} (I - Z_k Z_k^T)(a_j - z) = (I - Z_k Z_k^T) \sum_{j=1}^{m} (a_j - z).$$

Therefore, $\sum_{j=1}^{m}(a_j - z)$ lies in the null space of $(I - Z_k Z_k^T)$ or equivalently the column space of $Z_k$. The parametric representation (2.1) shows that there is no loss of generality in letting $\sum_{j=1}^{m}(a_j - z) = 0$ or

$$z = \frac{1}{m} \sum_{j=1}^{m} a_j. \tag{2.2}$$

Thus, a suitable $z$ has been determined and it should be noted that the value (2.2) solves the zero-dimensional case corresponding to $k = 0$. It remains to find $Z_k$ when $k > 0$, which is the problem:

$$\min \sum_{j=1}^{m} ||(I - Z_k Z_k^T)(a_j - z)||_2^2, \tag{2.3}$$

subject to the constraint that the columns of $Z_k$ are orthonormal and that $z$ satisfies equation (2.2). Using the properties of orthogonal projections and the definition of the vector 2-norm, (2.3) can be rewritten

$$\min \sum_{j=1}^{m} (a_j - z)^T (I - Z_k Z_k^T)(a_j - z). \tag{2.4}$$

Ignoring the terms in (2.4) independent of $Z_k$ then reduces the problem to

$$\min \sum_{j=1}^{m} -(a_j - z)^T Z_k Z_k^T (a_j - z),$$

or equivalently

$$\max \operatorname{tr} \sum_{j=1}^{m} (a_j - z)^T Z_k Z_k^T (a_j - z). \tag{2.5}$$

The introduction of the trace operator in (2.5) is allowed because the argument to the trace function is a matrix with only one element. The commutative property of the trace then shows that problem (2.5) is equivalent to

$$\max \operatorname{tr} \sum_{j=1}^{m} Z_k^T (a_j - z)(a_j - z)^T Z_k \equiv \max \operatorname{tr} Z_k^T \hat{A}\hat{A}^T Z_k$$

where $\hat{A}$ is the matrix

$$\hat{A} = [a_1 - z, a_2 - z, \ldots, a_m - z].$$

Theorem 1.1 and its corollary then show that the required matrix $Z_k$ can be taken to be the matrix of $k$ left singular vectors of the matrix $\hat{A}$ (right singular vectors of $\hat{A}^T$) corresponding to the $k$ largest singular values.

This result shows, not unexpectedly, that the best point lies on the best line which lies in the best plane, etc. Moreover, the total least squares problem described above clearly always has a solution although it will not be unique if the $(k+1)$st largest singular value of $\hat{A}$ has the same value as the $k$th largest. For example, if the data points are the 4 vertices of the unit square in $R^2$, $A = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}$, then any line passing through the centroid of the square is a best line in the total least squares sense because the matrix $\hat{A}$ for this data has two equal non-zero singular values.

The total least squares problem above (also referred to as orthogonal regression) has been considered by many authors and as is pointed out in [4, p 4]

> "... orthogonal regression has been discovered and rediscovered many times, often independently."

The approach taken above differs from that in [1], [2], and [4], in that the derivation is more geometric, it does not require the Eckart-Young-Mirsky Matrix Approximation Theorem (see, for example, [4]), and it uses only simple properties of projections and the matrix trace operator.

**3. A Stronger Result.** The proof of Theorem 1.1 relies on maximizing $\operatorname{tr}(DP)$ where $D$ is a (fixed) real diagonal matrix and $P$ varies over all rank $k$ projections. Since any two rank $k$ projections are unitarily equivalent the problem is now to maximize $\operatorname{tr}(DU^*PU)$ (for fixed $D$ and $P$) over all unitary matrices $U$. Generalizing from $P$ to a general Hermitian matrix leads to the following theorem.

THEOREM 3.1. *Let $A, B$ be $n \times n$ Hermitian matrices. Then*

$$\max_{U \text{unitary}} \operatorname{tr}(AU^*BU) = \sum_{i=1}^{n} \alpha_i \beta_i$$

*where $\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_n$ and $\beta_1 \geq \beta_2 \geq \cdots \geq \beta_n$ are the eigenvalues of $A$ and $B$ respectively, both similarly ordered.*

Clearly, Theorem 1.1 can be recovered since a projection of rank $k$ has eigenvalues 1, repeated $k$ times and 0 repeated $n-k$ times.

*Proof.* Let $\{e_i\}_{i=1}^n$ be an orthonormal basis of eigenvectors of A corresponding to the eigenvalues $\{\alpha_i\}_{i=1}^n$, written in descending order. Then

$$\operatorname{tr}(AU^*BU) = \sum_{i=1}^{n} e_i^* AU^*BUe_i = \sum_{i=1}^{n} (Ae_i)^* U^*BUe_i = \sum_{i=1}^{n} \alpha_i e_i^* U^*BUe_i$$

Let $B = V^*DV$ where $D$ is diagonal and $V$ is unitary. Writing $W = VU$ gives

$$\operatorname{tr}(AU^*BU) = \sum_{i=1}^{n} \alpha_i e_i^* W^*DWe_i = \sum_{i,j=1}^{n} p_{ij}\alpha_i\beta_j$$

where the $\beta_j$'s are the elements on the diagonal of $D$, i.e. the eigenvalues of $B$ and

$$p_{ij} = |(We_i)_j|^2 .$$

Note that since $W$ is unitary, the matrix $P = [p_{ij}]$, is doubly stochastic i.e. has non-negative entries and whose rows and columns sum to 1. The theorem will therefore follow once it is shown that for $\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_n$ and $\beta_1 \geq \beta_2 \geq \cdots \geq \beta_n$

$$\max_{[p_{ij}]} \sum_{i,j=1}^{n} \alpha_i\beta_j p_{ij} = \sum_{i=1}^{n} \alpha_i\beta_i \tag{3.1}$$

where the maximum is taken over all doubly stochastic matrices $P = [p_{ij}]$.

For fixed $P$ doubly stochastic, let

$$\chi = \sum_{i,j=1}^{n} \alpha_i\beta_j p_{ij}.$$

If $P \neq I$, let $k$ be the smallest index $i$ such that $p_{ii} \neq 1$. (Note that for $l < k, p_{ll} = 1$ and therefore $p_{ij} = 0$ if $i < k$ and $i \neq j$ and also if $j < k$ and $i \neq j$). Since $p_{kk} < 1$, then for some $l > k$, $p_{kl} > 0$. Likewise, for some $m > k$, $p_{mk} > 0$. These imply that $p_{ml} \neq 1$. The inequalities above mean that we can choose $\epsilon > 0$ such that the matrix $P'$ is doubly stochastic where

$$\begin{aligned} p'_{kk} &= p_{kk} + \epsilon \\ p'_{kl} &= p_{kl} - \epsilon \\ p'_{mk} &= p_{mk} - \epsilon \\ p'_{ml} &= p_{ml} + \epsilon \end{aligned}$$

and $p'_{ij} = p_{ij}$ in all other cases.
If we write

$$\chi' = \sum_{i,j=1}^{n} \alpha_i \beta_j p'_{ij}$$

then

$$\begin{aligned} \chi' - \chi &= \epsilon(\alpha_k\beta_k - \alpha_k\beta_l - \alpha_m\beta_k + \alpha_m\beta_l) \\ &= \epsilon(\alpha_k - \alpha_m)(\beta_k - \beta_l) \\ &\geq 0 \end{aligned}$$

which means that the term $\sum \alpha_i\beta_j p_{ij}$ is not decreased. Clearly $\epsilon$ can be chosen to reduce a non-diagonal term in $P$ to zero. After a finite number of iterations of this process it follows that $P = I$ maximizes this term. This proves (3.1) and hence Theorem 3.1. $\square$

Note that this theorem can also be regarded as a generalization of the classical result that if $\{\alpha_i\}_{i=1}^n$, $\{\beta_i\}_{i=1}^n$ are real sequences then $\sum \alpha_i \beta_{\sigma(i)}$ is maximized over all permutations $\sigma$ of $\{1, 2, \ldots, n\}$ when $\{\alpha_i\}$ and $\{\beta_{\sigma(i)}\}$ are similarly ordered.

**4. A Matrix Nearness Problem.** Theorem 3.1 also allows us to answer the following problem. If $A, B$ are Hermitian $n \times n$ matrices, what is the smallest distance between $A$ and a matrix $B'$ *unitarily equivalent* to B? Specifically, we have:

THEOREM 4.1. *Let $A, B$ be Hermitian $n \times n$ matrices with ordered eigenvalues $\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_n$ and $\beta_1 \geq \beta_2 \geq \cdots \geq \beta_n$ respectively. Let $||\cdot||$ denote the Frobenius norm. Then*

$$\min_{Q \text{unitary}} ||A - Q^*BQ|| = \sqrt{\sum_{i=1}^{n} (\alpha_i - \beta_i)^2}. \tag{4.1}$$

*Proof.*

$$\begin{aligned}||A - Q^*BQ||^2 &= \operatorname{tr}(A - Q^*BQ)^2 \\ &= \operatorname{tr}(A^2) + \operatorname{tr}(B^2) - 2\operatorname{tr}(AQ^*BQ)\end{aligned}$$

(Note that if $C, D$ are Hermitian, $\operatorname{tr}(CD)$ is real.)

So by Theorem 3.1

$$\begin{aligned}\min ||A - Q^*BQ||^2 &= \operatorname{tr}(A^2) + \operatorname{tr}(B^2) - 2\max_Q \operatorname{tr}(AQ^*BQ) \\ &= \sum \alpha_i^2 + \sum \beta_i^2 - 2\sum \alpha_i\beta_i \\ &= \sum (\alpha_i - \beta_i)^2\end{aligned}$$

and the result follows. ∎

An optimal $Q$ for problem (4.1) is clearly given by $Q = VU^*$ where $U, V$ are orthonormal matrices of eigenvectors of $A$, and $B$ respectively (corresponding to similarly ordered eigenvalues). This follows because $A = UD_\alpha U^*, B = VD_\beta V^*$, where $D_\alpha, D_\beta$ denote the diagonal matrices of eigenvalues $\{\alpha_i\}, \{\beta_i\}$ respectively and so

$$\begin{aligned}||A - Q^*BQ||^2 &= ||D_\alpha - U^*Q^*VD_\beta V^*QU||^2 \\ &= \sum (\alpha_i - \beta_i)^2 \quad \text{if} \quad Q = VU^*.\end{aligned}$$

Problem (4.1) is a variation on the well-known *Orthogonal Procrustes Problem* (see, for example, [2]) where an orthogonal (unitary) matrix is sought to solve

$$\min_{Q \text{unitary}} ||A - BQ||.$$

In this case $A$ and $B$ are no longer required to be Hermitian (or even square). A minimizing $Q$ for this problem can be obtained from a singular value decomposition of $B^*A$ [2, p 601].

## REFERENCES

[1] G. H. Golub and C. F. V. Loan, *An analysis of the total least squares problem*, SIAM J. Numer. Anal., 17 (1980), pp. 883–893.

[2] ———, *Matrix Computations*, Johns Hopkins University Press, Baltimore, MD, 3rd ed., 1996.

[3] R. O. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, New York, 1985.

[4] S. V. Huffel and J. Vandewalle, *The total least squares problem: computational aspects and analysis*, SIAM Publications, Philadelphia, PA, 1991.