QÜESTIIÓ, Vol 13, n° 1,2,3 pp. 75-79, 1989

A MATRIX DERIVATION OF A REPRESENTATION THEOREM FOR $(\mathrm{tr}A^p)^{1/p}$

H. NEUDECKER
University of Amsterdam

A matrix derivation of a well-known representation theorem for $(trA^p)^{1/p}$ is given, which is the solution of a restricted maximization problem.
The paper further gives a solution of the corresponding restricted minimization problem.

Keywords: Optimization of the trace of a matrix; matrix differentiation

1. INTRODUCTION

In a recent contribution Magnus (1987) gave a representation theorem for $(\mathrm{tr}A^p)^{1/p}$, where A is a non-zero positive semi-definite $(n \times n)$ matrix and $p > 1$.

The result reads:

$$\mathrm{tr}AX \leq (\mathrm{tr}A^p)^{1/p} \tag{1.1}$$

for every positive semi-definite $(n \times n)$ matrix X satisfying $\mathrm{tr}X^q = 1$, where $q = p/(p-1) > 1$.
Equality (1.1) occurs if and only if

$$X^q = (\mathrm{tr}A^p)^{-1} Ap \tag{1.2}$$

The representation therorem was derived by applying Hölder's and Karamata's scalar inequelities. This can be established even more rapidly by using

–Department of Actuarial Science and Econometrics. University of Amsterdam. Jodenbreestraat 23. 1011 NH Amsterdam, The Netherlands.
–Article rebut el desembre de 1989.

a scalar inequality that goes back to Schur. It is, however, appropriate to derive the result by means of matrix calculus, even if the proof becomes lengthy. This is done in this note. For details of the technique used see Magnus and Neudecker (1988).

2. THE DERIVATION

The obvious aim is to maximize the function $\text{tr}AX$ subject to $\text{tr}X^q = 1$. Let us therefore consider the Lagrangean function

$$\phi(Y) = \text{tr}AY'Y - \lambda[\text{tr}(Y'Y)^q - 1] \tag{2.1}$$

where X was replaced by $Y'Y$, Y' being a general $(n \times n)$ matrix.

The surface S defined by $\text{tr}(Y'Y)^q = 1$ is compact. Hence the differentiable function $\text{tr}AY'Y$ assumes its maximum and minimum values at some points Y_0 and Y_1 of S. We must look among the critical points of ϕ to find those points. See for analytical details Marsden and Tromba (1988, sections 4.2 and 4.3).

From

$$\tfrac{1}{2}\, d\phi(Y) = \text{tr}AY'\,dY - \lambda q\,\text{tr}(Y'Y)^{q-1}\,Y'\,dY \tag{2.2}$$

we find as necessary conditions for an extremum:

$$AY' = \lambda q(Y'Y)^{q-1}\,Y' \tag{2.3}$$

and

$$\text{tr}(Y'Y)^q = 1. \tag{2.4}$$

Among the points Y satisfying (2.3) and (2.4) are the points at which the constrained extrema occur.

We shall express the equations equivalently in terms of X, viz.

$$AX = \lambda q X^q \tag{2.5}$$

and

$$\text{(2.6)} \qquad \mathrm{tr} X^q = 1.$$

Let us first solve equation (2.5). It is easy to see that

$$\text{(2.7)} \qquad r(X) \leq r(A),$$

where $r(\,.)$ denotes rank.

It also follows from (2.5) that A and X commute:

$$\text{(2.8)} \qquad AX = XA.$$

The reason is that $AX = (AX)' = X'A' = XA$, because of the symmetry of a A and X.

Both A and X will therefore be diagonalized by one orthogonal matrix. Leaving out the zero eigenvalues of A and X, we can write:

$$\text{(2.9)} \qquad T'AT = \Lambda,\ T_1'XT_1 = M,$$

where T_1 is a subset of T, and Λ and M contain the non-zero eigenvalues of A and X respectively.

In case $r(X) = r(A)$, we have $T_1 = T$.

We can now rewrite (2.5) as:

$$\text{(2.10)} \qquad T\Lambda T'T_1 M T_1' = \lambda q T_1 M^q T_1'.$$

From (2.10) we obtain

$$\text{(2.11)} \qquad T_1 \Lambda_1 T_1' = \lambda q X^{q-1},$$

where Λ_1 is a subset of Λ, corresponding to the eigenvectors in T_1.

In case $r(X) = r(A)$, we have $\Lambda_1 = \Lambda$.

We have now expressed the points X in a more efficient way.

Raising both sides of (2.11) to the power p yields

$$T_1\Lambda_1^p T_1' = (\lambda q)^p X^q, \tag{2.12}$$

because $p(q-1) = q$.

By virtue of (2.5) and (2.6) we can write (2.12) as

$$T_1\Lambda_1^p T_1' = (\mathrm{tr}AX)^p X^q. \tag{2.13}$$

Hence

$$\mathrm{tr}\Lambda_1^p = (\mathrm{tr}AX)^p, \tag{2.14}$$

because (2.6).

It is our aim to maximize $\mathrm{tr}AX$, which is equivalent to maximizing $(\mathrm{tr}AX)^p$. It is immediate from (2.14) that $(\mathrm{tr}AX)^p$ reaches a maximum when $\Lambda_1 = \Lambda$ or equivalently $r(X) = r(A)$. The maximum is then $\mathrm{tr}A^p$. Hence max $\mathrm{tr}AX = (\mathrm{tr}A^p)^{1/p}$. It follows from (2.11) and (2.6) that $X = (\mathrm{tr}A^p)^{-1/q} A^{p-1}$ maximizes $\mathrm{tr}AX$ subject to the constrain $\mathrm{tr}X^q = 1$.

The function clearly reaches a minimum when Λ_1 contains only the smallest (non-zero)eigenvalue of A, λ_s say.

Let an associated normalized eigenvector be t_1. We then find

$$X = t_1 t_1', \tag{2.15}$$

which minimizes $\mathrm{tr}AX$ subject to $\mathrm{tr}X^q = 1$, or min $\mathrm{tr}AX = \lambda_s$.

Restating our result we get

$$\lambda_s \le \mathrm{tr}AX \le (\mathrm{tr}A^p)^{1/p}, \tag{2.16}$$

for every positive semi-definite matrix X satisfying $\mathrm{tr}X^q = 1$, where $q = p(p-1) > 1$ and λ_s is the smallest (non-zero) eigenvalue of A.

3. APPENDIX

We show that

$$d\text{tr}X^{\alpha} = \alpha\text{tr}X^{\alpha-1}\,dX,$$

when X is positive semi-definite and $\alpha > 1$.

Proof.

We use the spectral decomposition of X, viz. $X = SMS'$, S orthogonal and M diagonal.

Then

$$\begin{aligned} d\text{tr}X^{\alpha} &= d\text{tr}M^{\alpha} = \alpha\text{tr}M^{\alpha-1}\,dM \\ &= \alpha\text{tr}S'X^{\alpha-1}S\,dM = \\ &= \alpha\text{tr}X^{\alpha-1}\{S(dM)S' + (dS)MS' + SM(dS)'\} \\ &= \alpha\text{tr}X^{\alpha-1}\,dX, \\ &\text{as } O = d(S'S) = (dS)'S + S'dS. \end{aligned}$$

4. ACKNOWLEDGEMENT

The author is grateful to I.J. Steyn for valuable advice.

5. REFERENCES

[1] **Magnus, J.R.** (1987). "A representation theorem for $(\text{tr}A^p)^{1/p}$". *Linear Algebra* Appl.95: 127-134.

[2] **Magnus, J.R. and Neudecker H.** (1988). *"Matrix Differential Calculus with Applications in Statistics and Econometrics"*. John Wiley & Sons, Chichester.

[3] **Marsden, J.E. and Tromba, A.J.** (1988). *"Vector Calculus"*. (Third Edition), W.H. Freeman and Company, San Francisco.