Representation of the function Tr(exp(A - λB)) as a Laplace transform with positive weight and some matrix inequalities

1980 J. Phys. A: Math. Gen. 13 3147

(http://iopscience.iop.org/0305-4470/13/10/012)

J. Phys. A: Math. Gen. **13** (1980) 3147–3159. Printed in Great Britain

Representation of the function Tr(exp($A-\lambda B$)) as a Laplace transform with positive weight and some matrix inequalities

K J Le Couteur

Research School of Physical Sciences, The Australian National University, Canberra, ACT 2600, Australia

Received 23 April 1979, in final form 15 April 1980

Abstract. The conjecture that $\mathrm{Tr}(\exp(A-\lambda B))$ can be written as a Laplace transform with positive measure ρ is considered for finite Hermitian matrices A and B by means of Bernstein's theorem. An explicit formula is given for the moments of ρ in terms of divided differences of $\exp(A)$ and elements of B. For a large class of matrices A and B the moments of ρ take their maximum and minimum values when A and B commute and so upper and lower bounds for the moments of ρ are established; further analysis suggests that this is generally true if B is positive definite and A and B are bounded.

Some inequalities for the divided differences of the exponential are derived. Also, if A and B are both positive definite, upper and lower bounds are derived for $\mathrm{Tr}(A^nB^n)$ and $\mathrm{Tr}(AB)^n$ in terms of the eigenvalues of A and B.

1. Introduction

Bessis *et al* (1975) conjectured that if A and B are Hermitian matrices of finite order N and λ is a real parameter, the trace of the exponential has a representation

$$\mathrm{Tr}(\exp(A-\lambda B))=\int_{b_-}^{b_+} \mathrm{e}^{-\lambda t}\rho(t)\,\mathrm{d}t \tag{1.1}$$

where $\rho(t)$ is a positive weight and b_- and b_+ are the lowest and highest eigenvalues of B.

The existence of the representation is obvious if A and B commute and has further been demonstrated by Mehta and Kumar (1976) in the special case that A is a tree matrix in the representation with B diagonal.

The main importance of this conjecture is that the existence of a representation with positive weight ρ implies that successive Padé approximants to the function (1.1) provide upper and lower bounds to the true value (Baker 1972). This allows, for example, approximations to the partition function of statistical mechanics as a function of a coupling parameter λ.

The matrix B can be expressed as

$$B=bI+B' \tag{1.2}$$

where $b<b_-$ and B' is a positive definite Hermitian matrix with eigenvalues lying between b_--b and b_+-b. Then since

$$\mathrm{Tr}(\exp(A-\lambda B))=\mathrm{e}^{-\lambda b}\mathrm{Tr}(\exp(A-\lambda B')) \tag{1.3}$$

0305-4470/80/103147+13$01.50

it is sufficient to prove that $\rho(t)$ is positive in the representation

$$\mathrm{Tr}(\exp(A-\lambda B)) = \int_0^\infty \mathrm{e}^{-\lambda t}\rho(t)\,\mathrm{d}t, \tag{1.4}$$

in which A is Hermitian and B positive definite.

The formal solution to (1.4) is given by the inverse Laplace transform

$$\rho(t) = \frac{1}{2\pi\mathrm{i}}\int_{c-1\infty}^{c+\mathrm{i}\infty} \mathrm{Tr}(\exp(A-\lambda B+\lambda t))\,\mathrm{d}\lambda. \tag{1.5}$$

Additional results were given by Le Couteur (1978). In the steepest descent approximation it was possible to deduce from (1.5) that $\rho(t)\geqslant 0$. An extensive discussion of the third moment of ρ was given in geometric terms and a positive lower bound to the third moment of ρ for three-dimensional matrices was established.

2. Conditions for the existence of the representation

The necessary and sufficient condition for the existence of (2.3) with positive weight ρ is given by Bernstein's theorem (Widder 1971) as

$$\frac{\mathrm{d}^n}{\mathrm{d}\lambda^n}\mathrm{Tr}(\exp(A-\lambda B))\begin{cases}>0 & \text{for } n \text{ even}\\ <0 & \text{for } n \text{ odd}\end{cases}. \tag{2.1}$$

Now, the representation of the exponential as

$$\mathrm{e}^{\Lambda} = \lim_{L\to\infty}(1+\Lambda/L)^L$$

implies immediately

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}(\exp(A-\lambda B)) = -\int_0^1 \mathrm{d}x\,\mathrm{e}^{x\Lambda}B\,\mathrm{e}^{y\Lambda} \qquad x+y=1 \tag{2.2}$$

where

$$\Lambda = A-\lambda B. \tag{2.3}$$

So,

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}\mathrm{Tr}(\exp(A-\lambda B)) = -\mathrm{Tr}(\exp(A-\lambda B))B \tag{2.4}$$

and by further differentiation of the exponential

$$\begin{aligned}\frac{\mathrm{d}^n}{\mathrm{d}\lambda^n}&\mathrm{Tr}(\exp(A-\lambda B))\\ &= (-1)^n(n-1)!\iint \mathrm{Tr}(\mathrm{e}^{x_1\Lambda}B\,\mathrm{e}^{x_2\Lambda}B\ldots\mathrm{e}^{x_{n-1}\Lambda}B\,\mathrm{e}^{x_n\Lambda}B)\,\mathrm{d}x_1\ldots\mathrm{d}x_{n-1}\\ &\quad x_r>0,\ x_1+\ldots+x_n=1\\ &= (-1)^n\mathcal{I}_n\end{aligned} \tag{2.5a}$$

$$= (-1)^n n! \iint \mathrm{Tr}(\mathrm{e}^{x_1\Lambda}B \ldots \mathrm{e}^{x_r\Lambda}B \ldots \mathrm{e}^{x_n\Lambda}B\, \mathrm{e}^{x_{n+1}\Lambda})\, \mathrm{d}x_1 \ldots \mathrm{d}x_n$$

$$x_r > 0, x_1 + \ldots + x_{n+1} = 1. \qquad (2.5b)$$

So, from (2.1) the existence of the representation (1.4) or (1.1) follows if the integrals $\mathcal{I}_n$ in (2.5) are all positive. As the matrix products in the integrands of (2.5*a* or *b*) are not positive definite the sign of $\mathcal{I}_n$ is not easily determined. The evaluation and bounds of $\mathcal{I}_n$ will be considered in the following sections.

3. Evaluation of derivatives and moments

The integral (2.5) for $\mathcal{I}_n$ is most simply evaluated in a representation with Λ diagonal, with eigenvalues Λ_i. Then the integral is, taking $n = 4$ for definiteness,

$$\mathcal{I}_4 = \sum_{ijkl} 3! \int \mathrm{e}^{x\Lambda_i}B_{ij}\, \mathrm{e}^{y\Lambda_j}B_{jk}\, \mathrm{e}^{z\Lambda_k}B_{kl}\, \mathrm{e}^{t\Lambda_l}B_{li}\, \mathrm{d}x\, \mathrm{d}y\, \mathrm{d}z \qquad (3.1)$$

with $x + y + z + t = 1$. The integral is given by Hermite's formula for divided differences (Milne-Thomson 1933) of e^{Λ} as

$$\mathcal{I}_n = (n-1)! \sum_{i,j\ldots} [\Lambda_i, \Lambda_j, \Lambda_k, \ldots, \Lambda_p, \Lambda_q] B_{ij}B_{jk} \ldots B_{pq}B_{qi} \qquad (3.2)$$

with n factors B_{ij}.

By a generalisation of Rolle's theorem the divided difference can be expressed as

$$[\Lambda_i, \Lambda_j, \Lambda_k, \ldots, \Lambda_q] = \mathrm{e}^{\Lambda''}/(n-1)! \qquad (3.3)$$

with Λ'' in the range $\Lambda_i, \Lambda_j, \Lambda_k, \ldots, \Lambda_q$. So in the special case when B commutes with A, and thus with Λ, we recover the obvious result

$$\mathcal{I}_n = \sum_i \mathrm{e}^{\Lambda_i} B_{ii}^n > 0. \qquad (3.4)$$

According to (1.4) for $\lambda = 0$ the integral $\mathcal{I}_n$ reduces to the nth moment of the distribution $\rho(t)$:

$$\mu_n = \int_0^\infty t^n \rho(t)\, \mathrm{d}t = (-1) \frac{\mathrm{d}^n}{\mathrm{d}\lambda^n} \mathrm{Tr}(\exp(A - \lambda B)) \qquad (\lambda = 0). \qquad (3.5)$$

Thus μ_n is given by putting $\lambda = 0$, $\Lambda = A$ in (3.2) as

$$\mu_n = \sum_{i,j\ldots} (n-1)! [A_i, A_j, \ldots, A_m, A_n] B_{ij}B_{jk} \ldots B_{mn}B_{ni}. \qquad (3.6)$$

The upper and lower bounds for μ_n are given by equations (4.6), (4.7) and (4.8) of the next section with A_i in place of C_i.

Bessis *et al* (1975) have shown how these moments can be used to give approximations to the trace (1.1). Equation (3.6) can also be derived from the work of Schafroth (1951).

Approximations to (3.1) are possible, for example by forming its Laplace or Fourier transform, but seem to offer no advantage over the steepest descent approximation to (1.5).

4. Special cases

In the representation with Λ diagonal

$$\mathcal{I}_n = (n-1)! \sum_i \int e^{x_1\Lambda_i} e^{x_2\Lambda_j} \dots e^{x_n\Lambda_q} B_{ij}B_{jk} \dots B_{qi} \, dx_1 \dots dx_{n-1}$$

$$x_i > 0, \qquad \sum x_i = 1. \tag{4.1}$$

Let Λ_+ and Λ_- be the greatest and least eigenvalues of Λ; then

$$e^{\Lambda_-} \leqslant (n-1)! \int e^{x_1\Lambda_i} \dots e^{x_n\Lambda_q} \, dx_1 \dots dx_{n-1} \leqslant e^{\Lambda_+}, \tag{4.2}$$

because

$$1 = (n-1)! \int dx_1 \dots dx_{n-1} \qquad x_i > 0, x_1 + \dots + x_n = 1 \tag{4.3}$$

so that the integral (4.2) is a weighted mean of exponentials. Therefore, if all circuit matrix elements $B_{ij}B_{jk} \dots B_{qi}$ are positive in the Λ representation

$$\sum_i e^{\Lambda_-} B_{ii}^n \leqslant \mathcal{I}_n \leqslant \sum_i e^{\Lambda_+} B_{ii}^n \tag{4.4}$$

and we are led to the following proposition.

Proposition 1. If C is a given Hermitian matrix with greatest and least eigenvalues C_+ and C_- and B is a positive definite matrix of finite norm, with positive circuit matrix elements in the representation with C diagonal, then

$$J_n = (n-1)! \operatorname{Tr} \int e^{x_1 C} B \dots e^{x_n C} B \, dx \dots dx_{n-1} \qquad \sum x_i = 1 \tag{4.5}$$

is positive with bounds

$$e^{C_-} \operatorname{Tr}(B^n) \leqslant J_n \leqslant e^{C_+} \operatorname{Tr}(B^n) \tag{4.6}$$

which are attained when B commutes with C.

The result (4.6) follows immediately from (4.4).

The positive circuit matrix elements are found in the following cases.

(i) If all B_{ij} are positive.

(ii) For 2×2 matrices, in which off-diagonal elements of B enter only as $|B_{ij}|^2$.

(iii) If B is approximately diagonal in the C representation, for to order $|B_{ij}|^2$, $i \neq j$, the only contributions to J_n from off-diagonal elements of B are of the form $B_{ii}^r B_{ij} B_{jj}^s B_{ji} B_{ii}^t$, $r + s + t + 2 = n$, which are positive since B is Hermitian.

(iv) If B is a 'tree' matrix in the representation with C diagonal.

Under the same conditions closer bounds can be derived.

Proposition 2. If B and C satisfy the conditions of proposition 1, then the upper and lower bounds of the integral J_n are

$$J_n \sup = \sum_i e^{C_i} B_i^n, \tag{4.7}$$

in which the eigenvalues C_i and B_i of C and B are both arranged in ascending order (similarly ordered), and

$$J_n \inf = \sum_i e^{C_i} B_i^n \tag{4.8}$$

with the eigenvalues C_i in ascending and B_i in descending order (oppositely ordered).

Proof. The convexity of the exponential function leads directly to the inequality for divided differences stated in appendix 1, from which it follows that

$$\frac{1}{n}(e^{C_i}+e^{C_j}+\ldots+e^{C_q})B_{ij}B_{jk}\ldots B_{qi} \geq J_n \geq e^{(C_i+C_j\ldots+C_q)/n}B_{ij}B_{jk}\ldots B_{qi} \tag{4.9}$$

or

$$\operatorname{Tr} e^{C}B^n \geq J_n \geq \operatorname{Tr}(e^{C/n}B)^n \tag{4.10}$$

and then, using the final result of appendix 2,

$$\sum_i e^{C_i}B_i^n \text{(similarly ordered)} \geq J_n \geq \sum_i e^{C_i}B_i^n \text{(oppositely ordered)}. \tag{4.11}$$

Note that proposition 1 is a consequence of the second, and is obtained by varying the B_i in (4.11). However, the first is more easily derived independently.

3 × 3 matrices. Without the assumption of positive circuit matrix elements, the analysis is much more difficult; however, by inequalities Le Couteur (1978) derived the lower bound

$$\mathcal{J}_3 \geq \sum_i e^{C_i}(B_{ii})^3 \tag{4.12}$$

for a general Hermitian A and positive definite B. This result is similar to (4.6) and (4.11) and the bound is realised by commuting matrices B and C.

5. General positive definite B: computational results

It is important to know whether the bounds on $\mathcal{J}_n$ or J_n given in §4 are generally valid for positive definite matrices B, without the restriction that the circuit matrix elements of B are positive. Because of the severe analytical difficulties of this non-linear problem, an extensive computer survey was made to determine the minima of $\mathcal{J}_n$ or J_n, using matrices B with elements B_{ij} of either sign.

In the first series of computations to test positivity of $\mathcal{J}_n$ matrices B normalised to $\operatorname{Tr} B = 1$ and matrices Λ with $\operatorname{Tr} \Lambda$ fixed were explored. The elements of B and Λ were varied to find the lowest $\mathcal{J}_n$, calculated from (3.2), by using the general minimisation code PRAXIS in double precision arithmetic. The search was limited to matrix dimensions $N \leq 7$ and orders 3, 4, 5 of $\mathcal{J}_n$, with many different randomly generated matrices Λ and B used as the starting point for minimisation. In each case the minimum of $\mathcal{J}_n$ was positive, typically of order 10^{-6} starting from an initial value of order 10^7.

The form of these results suggested the bounds stated in §4, and to test these directly matrices C with given eigenvalues and matrices B with given eigenvalues or given trace were examined in a second series of computations. All results were consistent with the bounds.

As the number of terms in equation (3.2) for $\mathcal{I}_n$ is N^n it is difficult to extend these results further. However, the structure of (4.1) shows that negative contributions to $\mathcal{I}_n$ can arise first when $N \geqslant 3$ and $n \geqslant 3$ so the results given are a significant test of the bounds stated, without the restriction to positive circuit matrix elements.

6. General positive definite *B*: variational methods

The results given in § 4 cover the main cases in which $\rho(t)$ and $\mathcal{I}_n$ are known to be positive. The results of § 5 suggest that the bounds stated in § 4 are generally true for all positive definite B and so an analytical investigation of the minima of $\mathcal{I}_n$ or J_n is appropriate.

We suppose B is scaled to satisfy the condition

$$\operatorname{Tr} B^n = 1, \tag{6.1}$$

and since B is positive definite it may be parametrised as

$$B = \beta^2 \tag{6.2}$$

where β is a Hermitian matrix, constrained only by (6.1).

Now consider a variation δB of B keeping C fixed; then using the cyclic properties of the trace

$$\delta J_n = n!\operatorname{Tr}\int (\mathrm{e}^{x_1 C} B C^{x_2 C} \ldots B\, \mathrm{e}^{x_n C})\, \delta B\, \mathrm{d}x_1 \ldots \mathrm{d}x_{n-1} = n \operatorname{Tr} Q\delta B \tag{6.3}$$

where the matrix Q is defined by (6.3) and formally

$$Q_{ij} = \frac{1}{n}\left(\frac{\delta J_n}{\delta B_{ji}}\right)_C. \tag{6.4}$$

Note that the definitions of J_n and Q imply generally

$$J_n = \operatorname{Tr} QB. \tag{6.5}$$

The condition for a turning value of J_n, subject to the auxiliary condition (6.1), is, with a Lagrangian multiplier μ,

$$\operatorname{Tr} Q\,\delta B = \mu \operatorname{Tr} B^{n-1}\,\delta B \tag{6.6}$$

or, from (6.2),

$$\operatorname{Tr}(Q\beta + \beta Q)\,\delta\beta = 2\mu \operatorname{Tr} B^{n-1}\beta\,\delta\beta$$

and since $\delta\beta$ is an arbitrary Hermitian matrix this requires the coefficient of $\delta\beta_{ij}$ to vanish, that is

$$Q\beta + \beta Q = 2\mu B^{n-1}\beta$$

or

$$QB = BQ = \beta Q\beta = \mu B^n \tag{6.7}$$

and so at the turning value

$$J_n = \mu. \tag{6.8}$$

It remains to establish bounds on J_n. The substitution

$$C \to \theta C \qquad 0 \leqslant \theta \leqslant 1 \tag{6.9}$$

in (4.5) leads, after using cyclic properties of the trace, to

$$\partial J_n(\theta)/\partial\theta = n!\,\mathrm{Tr}\, C \int x_1\, e^{x_1\theta C} B \ldots B\, e^{x_n\theta C} B\, dx_1 \ldots dx_{n-1} \tag{6.10}$$

$$= n!\,\mathrm{Tr}\, C \int e^{x_1\theta C} B \ldots B\, e^{x_n\theta C} B\, e^{x_{n+1}\theta C}\, dx_1 \ldots dx_n \qquad (\Sigma\, x_i = 1) \tag{6.11}$$

$$= \mathrm{Tr}\, CR$$

where the Hermitian matrix R, defined by (6.11), is the same as that introduced in § 10 of Le Couteur (1978) to handle general variations of Λ or C according to

$$\delta J_n = \mathrm{Tr}\, R\, \delta C. \tag{6.12}$$

As discussed in that paper, equation (6.7) has simple solutions in which B commutes with C, $J_n = \mathrm{Tr}\, e^A B^n > 0$, $Q = e^A B^{n-1}$, but the general solution of (6.7) or of (6.7) and (6.12) in combination is not known. However, in principle the most general solution is not required; we need only the solutions corresponding to the lowest (or highest) value of J_n. In the following these are examined by a heuristic argument.

As θ increases from 0 to 1 the curve along which, for each θ, $J(\theta)$ attains its lowest (highest) value must be the envelope of all solutions of (6.7) with C replaced by θQ: at each θ let $B(\theta)$ be the matrix B which gives the point on the envelope.

It is now necessary to assume that $B(\theta)$ varies continuously with θ. Then at each θ, $B(\theta)$ must minimise (maximise) $\partial J_n/\partial\theta$ as well as $J(\theta)$ so that in addition to

$$Q(\theta)B(\theta) = \mu(\theta)B^n(\theta) \tag{6.7a}$$

$B(\theta)$ must satisfy the similar equation derived by variation of (6.10) or (6.11). As

$$\delta(\partial J_n(\theta)/d\theta)$$

$$= n!\,\mathrm{Tr} \int (Cx_1\, e^{x_1\theta C} B \ldots e^{x_n\theta C} + e^{x_n\theta C} B C x_1\, e^{x_1\theta C} \ldots e^{x_{n-1}\theta C} + \ldots$$

$$+ e^{x_2\theta C} B \ldots C x_1\, e^{x_1\theta C})\, dx_1 \ldots dx_{n-1}\, \delta B$$

$$= \mathrm{Tr}\, S \times \delta B \tag{6.13}$$

the condition is

$$S(\theta)B(\theta) = \nu B^n(\theta). \tag{6.14}$$

Now consider how the envelope starts from $\theta = 0$. For $\theta \ll 1$

$$J_n(\theta) = \mathrm{Tr}\, B^n(\theta) + \theta\, \mathrm{Tr}\, CB^n(\theta) + O(\theta^2). \tag{6.15}$$

With the scaling (6.1) the upper and lower bounds of $\mathrm{Tr}\, CB^n(\theta)$ are C_+ and C_- which are attained when $B(\theta)$ commutes with C and has only one non-vanishing eigenvalue, 1, in correspondence with C_+ or C_- respectively, as follows by evaluating the trace in a representation with C diagonal. This defines the initial $B(\theta) = B(0)$. As θ increases the same $B(\theta)$ continues to satisfy (6.7a) and (6.14) and, as shown in § 4, for each θ makes $J_n(\theta)$ a local maximum (minimum) with respect to variation of B. The argument of § 4, applied to (6.10), shows that the same $B(\theta)$ makes $dJ_n/d\theta$ a local maximum (minimum) with respect to variation of B. Therefore with the above continuity assumption, the upper (lower) envelope is generated by $B(\theta)=B(0)$. Accordingly, putting $\theta=1$,

$$e^{C_+} \geqslant J_n \geqslant e^{C_-} \tag{6.16}$$

and as J_n scales as B^n this implies the bounds stated in proposition 1.

The other solutions of (6.7) and (6.14) in which B commutes with C are not true minima or maxima of J_n; as shown in appendix 3, they are saddle points and the results of § 4 make it clear that these solutions cannot cross the bounds given. At $\theta \approx 0$ these are the only solutions, but it has not been possible rigorously to exclude the possibility that new types of solution may appear for θ above some positive lower limit θ_0, and eventually cross the bounds for some $\theta > \theta_0$; this is why above continuity was assumed explicitly. Such new solutions are unlikely because the n^2 matrix element B_{ij} would have to satisfy the $2n^2$ conditions (6.7) and (6.14). This type of non-linear problem is usually handled by computation and the extensive search described in § 5 found no exception to the bounds proposed.

Proposition 2 can be examined similarly for general positive definite matrices B. Keeping C fixed, consider infinitesimal variations of B which leave its spectrum unchanged:

$$B \to (1+\mathrm{i}\epsilon)B(1-\mathrm{i}\epsilon) \qquad \delta B = \mathrm{i}(\epsilon, B) \tag{6.17}$$

where ϵ is an arbitrary infinitesimal Hermitian matrix. Then, from (6.3)

$$\delta J_n = \mathrm{Tr}\,\mathrm{i}\,Q(\epsilon, B) = \mathrm{i}\sum_{j,k}(BQ-QB)_{jk}\epsilon_{kj}. \tag{6.18}$$

The matrices B which yield turning values of J_n must make δJ_n vanish for an arbitrary Hermitian ϵ; since Q and therefore $\mathrm{i}(B, Q)$ are Hermitian this requires

$$(B, Q)=0. \tag{6.19}$$

Alternatively to (6.17) one could vary C by canonical transformation leaving B unchanged and then, using (6.13), the condition for a stationary value of J_n is

$$(C, R)=0. \tag{6.20}$$

But (6.20) does not yield independent information from (6.18) because simultaneous canonical transformation of B and C gives $\delta J_n \equiv 0$, so that

$$(B, Q)+(C, R)\equiv 0.$$

As in the previous case, we seek the upper and lower envelopes of the solutions of (6.19) with C replaced by θC; these must satisfy

$$(Q(\theta), B(\theta))=0 \qquad \text{and} \qquad (S(\theta), B(\theta))=0 \tag{6.21}$$

in place of (6.7a) and (6.14).

For $\theta \ll 1$, (6.15) shows that the upper (lower) bound of $J_n(\theta)$ corresponds to the extremum of $\mathrm{Tr}\, CB^n(\theta)$ which, according to a theorem of Marcus (1956), occurs when $B(\theta)$ commutes with C and the eigenvalues C_i and B_i are arranged in the same (opposite) order. As previously, it is expected that this $B(\theta)$ generates the whole upper (lower) envelope on which J_n is then trivially evaluated as

$$J_n = \sum_r e^{C_r} B_r^n, \tag{6.22}$$

which has upper and lower bounds when the eigenvalues of B and C are ordered as stated in proposition 2.

7. Discussion

As mentioned in the introduction, the existence of the representation (1.1) has important consequences for the application of Padé approximations to problems of statistical mechanics and possibly Euclidean field theory. Bessis *et al* (1975) showed how these approximations may be developed in terms of the moments of ρ. An advantage of the approach of this paper is that it provides explicit formulae for the moments in the general case of Hermitian A and positive definite B. The upper and lower bounds for the moments stated in § 4 hold strictly under the conditions given there and the close bounds will often provide useful approximations to the moments.

The moments $\mathcal{I}_n$ must be positive under much less restrictive conditions than given in § 4, because small negative contributions cannot change their sign. As shown in §§ 5 and 6, it is likely that the bounds stated are generally valid if A and B are bounded and B is positive definite. In any case, the problem of determining the bounds of $\mathcal{I}_n$ in terms of the eigenvalues or traces of A and B is well posed and the variational equations given in § 6 provide strong constraints on the matrices to be considered in a more complete solution.

It may be remarked that the difficulty of a more direct approach to the general problem is that $\mathcal{I}_n$ is a polynomial of degree n in the B_{ij} and, for $n > 2$, a positive definite polynomial cannot necessarily be expressed as a sum of squares.

The analysis leads to some inequalities for divided differences of the exponential and for $\mathrm{Tr}(AB)^n$. These are of general interest apart from the main problem and so have been separated into two appendices.

Acknowledgments

I wish to thank Drs K Kumar, W S Woolcock, M L Mehta and D Bessis for critical discussions.

Appendix 1. Inequality for divided differences of the exponential

Hermite's formula (Milne-Thomson 1933) expresses the nth divided difference of e^C as a weighted average:

$$(n-1)![C_iC_j \ldots C_q] = (n-1)! \int e^{x_1C_i} e^{x_2C_j} \ldots e^{x_nC_q} \, dx_1 \ldots dx_{n-1}$$

$$x_i > 0, \qquad \sum x_i = 1.$$

As e^C is a convex function of C, for any set of x_i in the integrand

$$x_1 e^{C_i} + x_2 e^{C_j} + \ldots + x_n e^{C_q} \geqslant \exp(x_1 C_i + \ldots + x_n C_q). \tag{A1.1}$$

By summing over the n cyclic permutations of the x_i we obtain, since $\Sigma x_i = 1$,

$$(e^{C_i} + e^{C_j} + \ldots + e^{C_q}) \geqslant \sum_{\text{perms}} \exp(x_1 C_i + \ldots + x_n C_q)$$
$$\geqslant n \exp(C_i + C_j + \ldots + C_q)/n, \tag{A1.2}$$

in which the last inequality follows from comparison of arithmetic and geometric means.

By applying this to Hermite's integral we find an inequality for divided differences:

$$(e^{C_i} + e^{C_j} \ldots + e^{C_q})/n \geqslant (n-1)![C_i C_j \ldots C_q] \geqslant \exp[(C_i + \ldots + C_q)/n]. \tag{A1.3}$$

Appendix 2. Inequality for matrix power products

If A, B are positive definite $N \times N$ matrices it was shown by Golden (1965), and Thompson (1965) for $n = 2^m$ and by Lieb and Thirring (1976) for general n, that

$$\operatorname{Tr} A^n B^n \geqslant \operatorname{Tr}(AB)^n \tag{A2.1}$$

and by Marcus (1956) that

$$\sum_i A_i^n B_i^n \geqslant \operatorname{Tr} A^n B^n \geqslant \sum_i A_i^n B_{N-i}^n \tag{A2.2}$$

where A_i, B_i are the eigenvalues of A, B ordered in the same sense. These results suggest

$$\operatorname{Tr}(AB)^n \geqslant \sum_i A_i^n B_{N-i}^n, \tag{A2.3}$$

which does not seem to be in the literature and is a simple example of the method of § 6.

Proof. Let $R = (AB)^n = ABAB \ldots AB$. Since the eigenvalues A_i, B_i are fixed, the turning values of $\operatorname{Tr} R$ with respect to canonical transformation of B as in (6.15) are determined from (6.17):

$$QB = BQ \qquad \text{with } Q = ABAB \ldots A. \tag{A2.4}$$

Thus, at a stationary value of $\operatorname{Tr} R$,

$$R = ABAB \ldots AB = BABA \ldots BA = (B^{1/2} A B^{1/2})^n \tag{A2.5}$$

where $B^{1/2}$ is the positive square root of B. It follows immediately from (5) that R is Hermitian and

$$BR = RB \qquad \text{and} \qquad AR = RA, \tag{A2.6}$$

of which the second is equation (6.20) for this special case. If the eigenvalues of R are all different, (6) implies that A and B are diagonal in the representation with R diagonal and therefore commute with each other. If the eigenvalues of R are degenerate there is a subspace in which R is a positive multiple r of the unit matrix and in this subspace

$B^{1/2}AB^{1/2}=R^{1/2}$ is a positive multiple of the unit matrix so that again A and B commute. Thus, at a turning value of Tr R,

$$\mathrm{Tr}(AB)^n=\sum A_j^n B_j^n \tag{A2.7}$$

and the upper and lower limits stated in (2) and (3) follow from classical inequalities for the ordering of the sets A_j, B_j. So, collecting results, we have

$$\sum_i A_i^n B_i^n \geq \mathrm{Tr}\, A^n B^n \geq \mathrm{Tr}(AB)^n \geq \sum A_i^n B_{N-i}^n. \tag{A2.8}$$

Appendix 3. Other stationary values

J_n is obviously stationary with respect to small variations of C or B whenever B and C commute and the nature of these $N!$ stationary values is of interest. Assuming given eigenvalues of C and B, let C be diagonal and let b_i denote eigenvalues of B. Then

$$B=\mathrm{e}^{-\mathrm{i}\epsilon}b\,\mathrm{e}^{\mathrm{i}\epsilon} \qquad \epsilon=\epsilon^+, \tag{A3.1}$$

and as we need consider only small variations of B from the diagonal form, it is sufficient to work to second order in ϵ, assumed small. Then

$$B=(1-\mathrm{i}\epsilon-\epsilon^2/2)b(1+\mathrm{i}\epsilon-\epsilon^2/2)$$

$$B_{ij}=\mathrm{i}\epsilon_{ij}(b_i-b_j)+\sum_k \epsilon_{ik}\epsilon_{kj}(b_k-\tfrac{1}{2}b_i-\tfrac{1}{2}b_j) \qquad i\neq j \tag{A3.2a}$$

$$B_{ii}=b_i+\sum_k |\epsilon_{ik}|^2(b_k-b_i) \tag{A3.2b}$$

and, since the ϵ transformation is unitary, transformation of B^n gives

$$(B^n)_{ii}=b_i^n+\sum_k |\epsilon_{ik}|^2(b_k^n-b_i^n). \tag{A3.3}$$

Equation (A3.3) has contributions both from the off-diagonal terms B_{ij} which must appear in pairs $B_{ij}B_{ji}$ and from the diagonal terms $(B_{ii})^n$; these are easily separated:

$$\begin{aligned}(B^n)_{ii}&=b_i^n+\sum_k b_i^{n-1}|\epsilon_{ik}|^2(b_k-b_i)\ \text{(diagonal terms)}\\ &\quad+\sum_k |\epsilon_{ik}|^2[b_k^n-b_i^n+nb_i^{n-1}(b_i-b_k)]\ \text{(off-diagonal terms)}.\end{aligned} \tag{A3.4}$$

The terms in (A3.4) can be derived directly:

$$\begin{aligned}(B^n)_{ii}&=(BB\ldots B)_{ii}\\ &=b_i^n+n\sum_j |\epsilon_{ij}|^2 b_i^{n-1}(b_j-b_i)\ \text{(diagonal terms)}\\ &\quad+\sum_{j\neq i}\sum_{r,s,t} b_i^r B_{ij} b_j^s B_{ji} b_i^t \qquad r+s+t+2=n\ \text{(off-diagonal terms)}.\end{aligned} \tag{A3.5}$$

Consider

$$J_n=(n-1)!\int \mathrm{e}^{x_1 C}B\ldots\mathrm{e}^{x_n C}B\,\mathrm{d}x_1\ldots\mathrm{d}x_{n-1} \qquad \sum x_i=1. \tag{A3.6}$$

When $\epsilon = 0$, C and B commute and $J_n = \Sigma_i \mathrm{e}^i B_i^n$, we have to calculate J_n to second order in ϵ to determine

$$\Delta J_n = J_n - \sum_i \mathrm{e}^{C_i} B_i^n$$

$$= \sum_i \mathrm{e}^{C_i}[(B_{ii})^n - b_i^n] + \sum_i \sum_{j\neq i} \sum_{r,s,t} (n-1)! \int \mathrm{e}^{x_1 C_i} b_i \ldots b_i \, \mathrm{e}^{x_{r+1} C_i} B_{ij} \tag{A3.7}$$

$$\times \mathrm{e}^{x_{r+2} C_j} b_j \ldots \mathrm{e}^{x_{r+1+s} C_j} B_{ji} \, \mathrm{e}^{x_{r+s+2} C_i} b_i \ldots \mathrm{e}^{x_n C_i} b_i \, \mathrm{d}x_1 \ldots \mathrm{d}x_{n-1}. \tag{A3.8}$$

For $j \neq i$, $B_{ij} B_{ji} = |\epsilon_{ij}|^2 (b_i - b_j)^2$ and the contributions to ΔJ_n from the various $|\epsilon_{ij}|^2$ are additive and may be considered separately. Suppose that just one ϵ_{ij} is non-vanishing. Note that if C_i and C_j were equal the integral in the last line of (A3.8) would exactly equal $\mathrm{e}^{C_i} \times$ the 'off-diagonal' contribution to $(B^n)_{ii}$ as given in (A3.5): quite generally, since the integrand is positive, we find, according as $C_i \gtrless C_j$, $\Delta J_n \gtrless$ RHS of (A3.8) with C_i replaced by C_j in the integrand

$$= \mathrm{e}^{C_i}[(B_{ii})^n - b_i^n] + \mathrm{e}^{C_j}[(B_{jj})^n - b_j^n]$$

$$+ \mathrm{e}^{C_j}(\text{'off-diagonal' contributions to } (B^n)_{ii} + (B^n)_{jj})$$

$$= (\mathrm{e}^{C_i} - \mathrm{e}^{C_j})[(B_{ii})^n - b_i^n] + \mathrm{e}^{C_j}[(B^n)_{ii} + (B^n)_{jj} - b_i^n - b_j^n] \tag{A3.9}$$

and the last term vanishes because the ϵ transformation preserves the trace of B^n. So finally, using (A3.2),

$$\Delta J_n \gtrless (\mathrm{e}^{C_i} - \mathrm{e}^{C_j}) n b_i^{n-1} (b_j - b_i) \tag{A3.10}$$

according as $C_i \gtrless C_j$. Then $\Delta J_n > 0$ and J_n is a true minimum if C_i and b_i are oppositely ordered and $\Delta J_n < 0$ and J_n is a true maximum if C_i and b_i are similarly ordered. For the other possible orderings of b_i relative to C_i, J_n is stationary only; these are saddle points.

Similarly, consider

$$\mathrm{d}J_n/\mathrm{d}\theta = n! \,\mathrm{Tr} \int C x_1 \, \mathrm{e}^{x_1 C_\theta} B \ldots \mathrm{e}^{x_n C_\theta} B \, \mathrm{d}x_1 \ldots \mathrm{d}x_{n-1}. \tag{A3.11}$$

When $\epsilon = 0$, Λ and B commute and $\partial J_n/\mathrm{d}\theta = \Sigma_i C_i \, \mathrm{e}^{C_i \theta} B_i^n$. With B given by (A3.2), to second order in ϵ we find, as in (A3.8),

$$\Delta(\partial J_n/\mathrm{d}\theta) = (\partial J_n/\mathrm{d}\theta) - \sum_i C_i \, \mathrm{e}^{C_i \theta} b_i^n$$

$$= \sum_i C_i \, \mathrm{e}^{\theta C_i}[(B_{ii})^n - b_i^n] + \sum_{i,j\neq i} \sum_{rst} C_i x_1 \, \mathrm{e}^{x_1 C_i} b_i \ldots b_i$$

$$\times \mathrm{e}^{x_{r+1} C_i \theta} B_{ij} \ldots B_{ji} \ldots \mathrm{e}^{x_n C_i \theta} b_i \, \mathrm{d}x_1 \ldots \mathrm{d}x_{n-1}. \tag{A3.12}$$

And so, as in (A3.10),

$$\Delta(\partial J_n/\mathrm{d}\theta) \gtrless C_i[(\mathrm{e}^{\theta C_i} - \mathrm{e}^{\theta C_j}) b_i^{n-1} (b_j - b_i)]$$

according as $C_i \gtrless C_j$. So $\partial J_n/\mathrm{d}\theta$ is a true minimum if C_i and b_i are oppositely ordered and $\partial J_n/\mathrm{d}\theta$ is a true maximum if C_i and b_i are similarly ordered.

References

Baker G A 1972 *Carnegie Lectures in Physics* (New York: Gordon and Breach) vol 5 pp 349–82
Bessis D, Moussa P and Villani M 1975 *J. Math. Phys.* **16** 2318
Golden S 1965 *Phys. Rev.* **137** B1127
Le Couteur K J 1978 *Proc. Int. Conf. and Winter School of Frontiers of Theoretical Physics* ed. F C Auluck, L S Kothari and V S Nanda (New Delhi: Indian National Science Academy) p 209
Lieb E and Thirring W 1976 *Studies in Mathematical Physics* (Princeton, NJ: Princeton University Press) p 269
Marcus M 1956 *Am. Math. Monthly* **63** 173
Mehta M L and Kumar K 1976 *J. Phys. A: Math. Gen.* **9** 117
Milne-Thomson L M 1933 *Calculus of Finite Differences* (London: Macmillan) p 10
Schafroth M R 1951 *Helv. Phys. Acta* **24** 645
Thompson C J 1965 *J. Math. Phys.* **6** 1812
Widder D 1971 *An Introduction to Transform Theory* (New York: Academic) p 155