Hindawi Publishing Corporation
Fixed Point Theory and Applications
Volume 2010, Article ID 241908, 12 pages
doi:10.1155/2010/241908

Research Article

Trace-Inequalities and Matrix-Convex Functions

Tsuyoshi Ando

Hokkaido University (Emeritus), Shiroishi-ku, Hongo-dori 9, Minami 4-10-805, Sapporo 003-0024, Japan

Correspondence should be addressed to Tsuyoshi Ando, ando@es.hokudai.ac.jp

Received 8 October 2009; Accepted 30 November 2009

Academic Editor: Anthony To Ming Lau

A real-valued continuous function $f(t)$ on an interval (α, β) gives rise to a map $X \mapsto f(X)$ via functional calculus from the convex set of $n \times n$ Hermitian matrices all of whose eigenvalues belong to the interval. Since the subpace of Hermitian matrices is provided with the order structure induced by the cone of positive semidefinite matrices, one can consider convexity of this map. We will characterize its convexity by the following trace-inequalities: $\mathrm{Tr}(f(B) - f(A))(C - B) \leq \mathrm{Tr}(f(C) - f(B))(B - A)$ for $A \leq B \leq C$. A related topic will be also discussed.

1. Introduction and Theorems

Let $f(t)$ be a (real-valued) continuous function defined on an open interval (α, β) of the real line. The function $f(t)$ is said to be *convex* if

$$f(\lambda a + (1-\lambda)b) \leq \lambda f(a) + (1-\lambda) f(b) \quad \left(0 \leq \lambda \leq 1;\ \alpha < a, b < \beta\right). \tag{1.1}$$

We referee to [1] for convex functions. Under continuity the requirement (1.1) can be restricted only to the case $\lambda = 1/2$, that is,

$$f\left(\frac{a+b}{2}\right) \leq \frac{f(a) + f(b)}{2} \quad \left(\alpha < a, b < \beta\right). \tag{1.2}$$

It is well known that when $f(t)$ is a C^1-function, its convexity is characterized by the condition on the derivative

$$\frac{f(b) - f(b-t)}{t} \leq f'(b) \leq \frac{f(b+t) - f(b)}{t} \quad \left(\alpha < b - t < b + t < \beta\right), \tag{1.3}$$

and, further when $f(t)$ is a C^2-function, by the condition on the second derivative

$$f''(b) \geq 0 \quad (\alpha < b < \beta). \tag{1.4}$$

On the other hand, it is easy to see that (1.1) is equivalent to the following requirement on the divided difference:

$$\frac{f(b) - f(a)}{b - a} \leq \frac{f(c) - f(b)}{c - b} \quad (\alpha < a < b < c < \beta), \tag{1.5}$$

or even to the inequality

$$\big(f(b) - f(a)\big)(c - b) \leq \big(f(c) - f(b)\big)(b - a) \quad (\alpha < a \leq b \leq c < \beta). \tag{1.6}$$

Let $\mathbb{M}_n$ be the linear space of $n \times n$ complex matrices, and $\mathbb{H}_n$ its (real) subspace of $n \times n$ Hermitian matrices. The identity matrix I will be denoted simply by 1, and correspondingly, a scalar λ will represent λI. For Hermitian $A,\ B$ the order relation $A \leq B$ means that $B - A$ is *positive-semidefinite*, or equivalently

$$A \leq B \Longleftrightarrow \langle Ax, x\rangle \leq \langle Bx, x\rangle \quad (x \in \mathbb{C}^n), \tag{1.7}$$

where $\langle x, y\rangle$ denotes the inner product of vectors $x, y \in \mathbb{C}^n$. The strict order relation $A < B$ will mean that $B - A$ is positive definite; that is, $B \geq A$ and $B - A$ is invertible (see [2] for basic facts about matrices.)

Notice that for scalars α, β and Hermitian X the order relation $\alpha < X < \beta$ is equivalent to the condition that every eigenvalue of X is in the interval (α, β). Denote by $\mathbb{H}_n(\alpha, \beta)$ the convex set of Hermitian matrices X such that $\alpha < X < \beta$. A continuous function $f(t)$, defined on (α, β), induces a (nonlinear) map $X \mapsto f(X)$ from $\mathbb{H}_n(\alpha, \beta)$ to $\mathbb{H}_n$ through the familiar *functional calculus*, that is,

$$f(X) := U \operatorname{diag}\big(f(\lambda_1), \dots, f(\lambda_n)\big)U^* \tag{1.8}$$

with a unitary matrix U which diagonalizes X as

$$U^* X U = \operatorname{diag}(\lambda_1, \dots, \lambda_n). \tag{1.9}$$

The function $f(t)$ is said to be *matrix-convex of order* n, or simply n*-convex* on the interval (α, β) if the map $X \mapsto f(X)$ is *convex* on $\mathbb{H}_n(\alpha, \beta)$ or more exactly

$$f(\lambda A + (1 - \lambda)B) \leq \lambda f(A) + (1 - \lambda) f(B) \quad (0 \leq \lambda \leq 1;\ A, B \in \mathbb{H}_n(\alpha, \beta)) \tag{1.10}$$

(see [3, 4]). This is a formal matrix-version of (1.1). In view of (1.7) this convexity means that the numerical valued function $X \mapsto \langle f(X)x, x\rangle$ is convex for all vector $x \in \mathbb{C}^n$.

Just as in the scalar case for the matrix-convexity the following matrix-version of (1.2) is sufficient:

$$2f(B) \leq f(B+X) + f(B-X) \quad \left(B \pm X \in \mathbb{H}_n(\alpha,\beta)\right). \tag{1.11}$$

This means that $f(t)$ is n-convex if and only if the map $t \mapsto f(B+tX)$ is convex on $(-1,1)$ when $B \pm X \in \mathbb{H}_n(\alpha,\beta)$.

The 1-convexity is nothing but the usual convexity of the function $f(t)$. It is easy to see that n-convexity implies m-convexity for all $1 \leq m \leq n$.

It is known (see [3]) that if $f(t)$ is 2-convex then it is already a C^2-function, and (see [5, 6]) that for each n there is an n-convex function which is not $(n+1)$-convex.

It should be mentioned here that in his original definition of n-convexity Kraus [3] restricted the requirement (1.11) only for $X \geq 0$ with $\operatorname{rank}(X) = 1$. We will return to this point later.

The corresponding matrix-versions of (1.5) and (1.6) have no definite meaning because $(f(B) - f(A))(C - B)$ or $(f(B) - f(A))(B - A)^{-1}$ is no longer Hermitian.

On the space $\mathbb{M}_n$ the most useful linear functional is the *Trace*, in symbol, $\operatorname{Tr}(X)$, which is defined as the sum of diagonal entries of X with respect to any orhonormal basis. The useful properties of the trace are *commutativity*, $\operatorname{Tr}(XY) = \operatorname{Tr}(YX)$, and *positivity*, that is, $X \leq Y \Rightarrow \operatorname{Tr}(X) \leq \operatorname{Tr}(Y)$.

We will use a characterization of positive semidefiniteness $X \geq 0$ in terms of trace:

$$X \geq 0 \Longleftrightarrow \operatorname{Tr}(XY) \geq 0 \quad (0 \leq Y \text{ of rank-one}). \tag{1.12}$$

Notice in this connection that if both X, Y are Hermitian, then $\operatorname{Tr}(XY)$ is a real number.

Our main aim is to establish trace-versions of (1.5) and (1.6). The trace-version for (1.6) is quite natural.

Theorem 1.1. *A continuous function $f(t)$ on an interval (α,β) is n-convex if and only if*

$$Tr\left(f(B) - f(A)\right)(C-B) \leq Tr\left(f(C) - f(B)\right)(B-A) \quad \left(A \leq B \leq C \text{ in } \mathbb{H}_n(\alpha,\beta)\right). \tag{$\dagger_n$}$$

On the other hand, the trace-version for (1.5) *turns out quite restrictive.*

Theorem 1.2. *Let $n \geq 2$. A continuous function $f(t)$ on an interval (α,β) satisfies the condition*

$$Tr\left(f(B) - f(A)\right)(B-A)^{-1} \leq Tr\left(f(C) - f(B)\right)(C-B)^{-1} \quad \left(A < B < C \text{ in } \mathbb{H}_n(\alpha,\beta)\right) \tag{$\ddagger_n$}$$

if and only if it is of the form $f(t) = at^2 + bt + c$ with $a \geq 0$, and $b, c \in \mathbb{R}$.

2. Preliminary

In order to prove theorems, we use a well-established regularization technique (see [7] I-4). Take a nonnegative symmetric C^∞-function $\varphi(t)$ on $(-\infty,\infty)$ such that

$$\varphi(t) = 0 \quad (|t| \geq 1), \qquad \int_{-\infty}^{\infty} \varphi(t)dt = 1, \tag{2.1}$$

and for $\epsilon > 0$ let $\varphi_{(\epsilon)} := \varphi(t\epsilon^{-1})\epsilon^{-1}$. Then $\varphi_{(\epsilon)}(t)$ is a nonnegative, symmetric C^∞-function such that

$$\varphi_{(\epsilon)}(t) = 0 \quad (|t| \geq \epsilon), \qquad \int_{-\infty}^{\infty} \varphi_{(\epsilon)}(t)dt = 1. \tag{2.2}$$

Given a continuous function $f(t)$ on an interval (α,β), setting $f(t) = 0$ outside of the interval (α,β), define $f_\epsilon(t)$ as the *convolution* of this extended function f with $\varphi_{(\epsilon)}$, that is,

$$f_\epsilon(t) := \left(f \star \varphi_{(\epsilon)}\right)(t) = \int_{-\infty}^{\infty} f(t-s)\varphi_{(\epsilon)}(s)ds. \tag{2.3}$$

The following is well known.

Lemma 2.1. *The function f_ϵ is a C^∞-function, in fact,*

$$\frac{d^k}{dt^k} f_\epsilon = f \star \frac{d^k}{dt^k}\varphi_{(\epsilon)} \quad (k = 1,2,\ldots), \tag{2.4}$$

and $f_\epsilon(t)$ converges to $f(t)$ uniformly on each compact subset of the interval (α,β) as $\epsilon \to 0$.

Lemma 2.2. *Let $f(t)$ be a continuous function on an interval (α,β).*

(i) *$f(t)$ satisfies $(\dagger_n)$ on (α,β) if and only if for small $\epsilon > 0$ the function $f_\epsilon(t)$ satisfies $(\dagger_n)$ on $(\alpha+\epsilon,\beta-\epsilon)$.*

(ii) *$f(t)$ is n-convex on (α,β) if and only if for small $\epsilon >$ the function $f_\epsilon(t)$ is n-convex on $(\alpha+\epsilon,\beta-\epsilon)$.*

Proof. (i) Let $f(t)$ satisfy $(\dagger_n)$ on (α,β). Suppose that $\alpha+\epsilon < A \leq B \leq C < \beta-\epsilon$, then

$$\begin{aligned} &\mathrm{Tr}\left(f_\epsilon(C) - f_\epsilon(B)\right)(B-A) - \mathrm{Tr}\left(f_\epsilon(B) - f_\epsilon(A)\right)(C-B) \\ &\quad = \int_{-\epsilon}^{\epsilon} \left[\mathrm{Tr}\{f(C-s) - f(B-s)\}\{(B-s)-(A-s)\}\right. \\ &\qquad\qquad \left. -\mathrm{Tr}\{f(B-s) - f(A-s)\}\{(C-s)-(B-s)\}\right]\varphi_{(\epsilon)}(s)ds \geq 0, \end{aligned} \tag{2.5}$$

because

$$\alpha < A - s \leq B - s \leq C - s < \beta \quad (|s| \leq \epsilon). \tag{2.6}$$

The converse statement is clear by the second half of Lemma 2.1.

The proof of (ii) is also easy and omitted. □

In a similar way we have the following.

Lemma 2.3. *A continuous function $f(t)$ satisfies $(\ddagger_n)$ on (α, β) if and only if for small $\epsilon > 0$ the functin $f_\epsilon(t)$ satisfies $(\ddagger_n)$ on $(\alpha + \epsilon, \beta - \epsilon)$.*

When $f(t)$ is a C^1-function on (α, β), $B \in \mathbb{H}_n(\alpha, \beta)$ and $X \in \mathbb{H}_n$, the map $t \mapsto f(B + tX)$ is defined for small $|t|$ and differentiable at $t = 0$. The derivative of this map at $t = 0$ will be denoted by $\mathfrak{D}f(B; X)$, that is,

$$\mathfrak{D}f(B; X) := \left.\frac{d}{dt}\right|_{t=0} f(B + tX) \quad (X \in \mathbb{H}_n). \tag{2.7}$$

When B is *daigonal* as $B = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$, it is known (see [2] V-3 and [8] 6-6) that

$$\mathfrak{D}f(B; X) = \left[f^{[1]}(\lambda_i, \lambda_j)\right]_{i,j=1}^{n} \circ \left[x_{ij}\right]_{i,j=1}^{n} \quad \text{for } X = \left[x_{ij}\right]_{i,j=1}^{n}, \tag{2.8}$$

where $\circ$ denotes the *Schur product* (= entrywise product) and $f^{[1]}(s,t)$ is the *first divided difference* of f, defined as

$$f^{[1]}(s,t) = \begin{cases} \dfrac{f(s) - f(t)}{s - t}, & \text{if } s \neq t, \\ f'(s), & \text{if } s = t. \end{cases} \tag{2.9}$$

Notice that $[f^{[1]}(\lambda_i, \lambda_j)]_{i,j=1}^{n}$ is a real symmetric matrix.

In a similar way when $f(t)$ is a C^2-function, the *second derivative* of the map $t \mapsto f(B + tX)$ at $t = 0$ is written as (see [2] V-3 and [8] 6-6)

$$\left.\frac{d^2}{dt^2}\right|_{t=0} f(B + tX) = 2\left[\sum_{k=1}^{n} f^{[2]}(\lambda_i, \lambda_k, \lambda_j) x_{ik} x_{kj}\right]_{i,j=1}^{n} \quad \text{for } X = \left[x_{ij}\right]_{i,j=1}^{n}, \tag{2.10}$$

where $f^{[2]}(s,t,u)$ is the *second divided difference* of f, defined as

$$f^{[2]}(s,t,u) = \frac{f^{[1]}(s,t) - f^{[1]}(t,u)}{s - u} = \begin{cases} \dfrac{f^{[1]}(s,t) - f^{[1]}(t,u)}{s - u}, & \text{if } s \neq u, \\ \dfrac{f'(s) - f^{[1]}(t,s)}{s - t}, & \text{if } s = u \neq t, \\ \dfrac{f''(s)}{2}, & \text{if } s = t = u. \end{cases} \tag{2.11}$$

Since the functional calculus is invariant for unitary similarity, that is, $f(V^*XV) = V^*f(X)V$, the formulas (2.8) and (2.10) well determine the forms of derivatives.

Lemma 2.4. *If $f(t)$ is a C^1-function on an interval (α,β), then*

$$Tr\left(\mathfrak{D}f(B;X)\cdot Y\right) = Tr\left(\mathfrak{D}f(B;Y)\cdot X\right) \quad \left(B\in\mathbb{H}_n(\alpha,\beta);\ X,Y\in\mathbb{H}_n\right). \tag{2.12}$$

Proof. We may assume that $B = \operatorname{diag}(\lambda_1,\ldots,\lambda_n)$, then by (2.8) for $X = [x_{ij}]_{i,j=1}^n$ and $Y = [y_{ij}]_{i,j=1}^n$

$$\begin{aligned}\operatorname{Tr}\left(\mathfrak{D}f(B;X)\cdot Y\right) &= \sum_{i,j=1}^n f^{[1]}(\lambda_i,\lambda_j)x_{ij}y_{ji}\\ &= \sum_{i,j=1}^n f^{[1]}(\lambda_j,\lambda_i)y_{ji}x_{ij} = \operatorname{Tr}\left(\mathfrak{D}f(B;Y)\cdot X\right).\end{aligned} \tag{2.13}$$

□

3. Proofs of Theorems

By Lemmas 2.3 and 2.4 we may assume that $f(t)$ in theorems is a C^∞-function.

Proof of Theorem 1.1. Suppose that the function $f(t)$ satisfies $(\dagger_n)$ on (α,β). Take $B\in\mathbb{H}_n(\alpha,\beta)$ and $0\le X,Y$ of rank-one such that $C := B+X$ and $A := B-tY$ for small $t>0$ belong to $\mathbb{H}_n(\alpha,\beta)$. Since $A\le B\le C$, by assumption $(\dagger_n)$ we have

$$\operatorname{Tr}\frac{\left(f(B)-f(B-tY)\right)X}{t} \le \operatorname{Tr}\left(f(B+X)-f(B)\right)Y, \tag{3.1}$$

hence by (2.7)

$$\operatorname{Tr}\left(\mathfrak{D}f(B;Y)\cdot X\right) \le \operatorname{Tr}\left(f(B+X)-f(B)\right)Y. \tag{3.2}$$

Then it follows from Lemma 2.4 that

$$\operatorname{Tr}\left(\mathfrak{D}f(B;X)\cdot Y\right) \le \operatorname{Tr}\left(f(B+X)-f(B)\right)Y. \tag{3.3}$$

Since $0\le X,Y$ of rank-one are arbitrary, it follows from (3.3) and (1.12) that for any $0\le X$ of rank one such that $\alpha < B\pm X < \beta$

$$\mathfrak{D}f(B;X) \le f(B+X) - f(B), \tag{3.4}$$

and similarly

$$f(B) - f(B-X) \le \mathfrak{D}f(B;X). \tag{3.5}$$

Therefore

$$2f(B) \le f(B+X) + f(B-X) \quad \left(B \pm X \in \mathbb{H}_n(\alpha,\beta) \text{ and } 0 \le X \text{ of rank-one}\right). \tag{3.6}$$

This means that the matrix-valued function $t \mapsto f(B+tX)$ is convex under the condition that $0 \le X$ is of rank-one.

At this point we proved the n-convexity in the sense of Kraus [3] as mentioned in Section 1. The remaining part is essentially the same as Kraus' approach [3].

Since for $0 \le X$ of rank-one and small $t > 0$ by (3.6)

$$0 \le \frac{f(B+tX) + f(B-tX) - 2f(B)}{t^2} \longrightarrow \left.\frac{d^2}{dt^2}\right|_{t=0} f(B+tX), \tag{3.7}$$

we can conclude from (2.10) that for $0 \le X = [x_{ij}]_{i,j=1}^n$ of rank-one

$$\left[\sum_{k=1}^n f^{[2]}(\lambda_i,\lambda_k,\lambda_j)x_{ik}x_{kj}\right]_{i,j=1}^n \ge 0. \tag{3.8}$$

For each $t > 0$, consider a positive semidefinite matrix of rank-one

$$0 \le X = \left[x_{ij}\right]_{i,j=1}^n := \left[t^{i+j-2}\right]_{i,j=1}^n. \tag{3.9}$$

Then by (3.8)

$$\begin{aligned} 0 &\le \left[\sum_{k=1}^n f^{[2]}(\lambda_i,\lambda_k,\lambda_j)x_{ik}x_{kj}\right]_{i,j=1}^n \\ &= \operatorname{diag}\left(1,t,\ldots,t^{n-1}\right)\left[\sum_{k=1}^n f^{[2]}(\lambda_i,\lambda_k,\lambda_j)t^{2(k-1)}\right]_{i,j=1}^n \operatorname{diag}\left(1,t,\ldots,t^{n-1}\right), \end{aligned} \tag{3.10}$$

which implies

$$\left[\sum_{k=1}^n f^{[2]}(\lambda_i,\lambda_k,\lambda_j)t^{2(k-1)}\right]_{i,j=1}^n \ge 0. \tag{3.11}$$

Letting $t \to 0$, we have

$$C_1 := \left[f^{[2]}(\lambda_i,\lambda_1,\lambda_j)\right]_{i,j=1}^n \ge 0. \tag{3.12}$$

In a similar way we can see that

$$C_k := \left[f^{[2]}(\lambda_i, \lambda_k, \lambda_j)\right]_{i,j=1}^{n} \geq 0 \quad (k = 1, 2, \ldots, n). \tag{3.13}$$

Now since for any $X = [x_{ij}]_{i,j=1}^{n} \in \mathbb{H}_n$ each matrix $[x_{ik}x_{kj}]_{i,j=1}^{n}$ $(k = 1, 2, \ldots, n)$ is positive semidefinite and of rank-one, it follows from (3.8) that

$$\begin{aligned} \left.\frac{d^2}{dt^2}\right|_{t=0} f(B + tX) &= 2\left[\sum_{k=1}^{n} f^{[2]}(\lambda_i, \lambda_k, \lambda_j) x_{ik} x_{kj}\right]_{i,j=1}^{n} \\ &= 2\sum_{k=1}^{n} C_k \circ \left[x_{ik}\overline{x_{jk}}\right]_{i,j=1}^{n} \geq 0. \end{aligned} \tag{3.14}$$

Here we used the well-known fact that the Schur product of two positive semidefinite matrices is again positive semidefinite (see [2] I-6). Therefore

$$\left.\frac{d^2}{dt^2}\right|_{t=0} f(B + tX) \geq 0 \quad \left(B \in \mathbb{H}_n(\alpha, \beta); X \in \mathbb{H}_n\right), \tag{3.15}$$

which implies the convexity of the map $t \mapsto f(B + tX)$ whenever $B \pm X \in \mathbb{H}_n(\alpha, \beta)$. This completes the proof of the n-convexity of the function $f(t)$.

Suppose conversely that $f(t)$ is n-convex on the interval (α, β), then by (1.3)

$$\begin{aligned} f(C) - f(B) = f(B + (C - B)) - f(B) &\geq \left.\frac{d}{dt}\right|_{t=0} f(B + t(C - B)) \\ &= \mathfrak{D}f(B; C - B), \end{aligned} \tag{3.16}$$

so that by (1.12)

$$\operatorname{Tr}\big(f(C) - f(B)\big)(B - A) \geq \operatorname{Tr}\big(\mathfrak{D}f(B; C - B) \cdot (B - A)\big), \tag{3.17}$$

and similarly

$$\operatorname{Tr}\big(f(B) - f(A)\big)(C - B) \leq \operatorname{Tr}\big(\mathfrak{D}f(B; B - A) \cdot (C - B)\big). \tag{3.18}$$

Now by Lemma 2.4 we can conclude

$$\operatorname{Tr}\big(f(B) - f(A)\big)(C - B) \leq \operatorname{Tr}\big(f(C) - f(B)\big)(B - A), \tag{3.19}$$

which shows that the function $f(t)$ satisfies $(\dagger_n)$. This completes the whole proof of Theorem 1.1. □

In the above proof we really showed the following.

Theorem 3.1. *A continuous function $f(t)$ on an interval (α,β) is n-convex if and only if $Tr(f(B)-f(A))(C-B) \le Tr(f(C)-f(B))(B-A)$, whenever $A \le B \le C$ in $\mathbb{H}_n(\alpha,\beta)$ and* $\operatorname{rank}(B-A) = \operatorname{rank}(C-B) = 1$.

Notice that Kraus [3] (cf. [8, Theorem 6.6.52]) really showed, for $n \ge 2$, that $f(t)$ is n-convex on (α,β) if and only if it is a C^2-function and

$$\left[f^{[2]}(\lambda_i,\lambda_1,\lambda_j)\right]_{i,j=1}^{n} \ge 0 \quad \forall\lambda_1,\ldots,\lambda_n \in (\alpha,\beta). \tag{3.20}$$

For the proof of Theorem 1.2, let us start with an easy lemma.

Lemma 3.2. *If condition $(\ddagger_n)$ for $f(t)$ is valid on (α,β), so is condition $(\ddagger_m)$ for $1 \le m < n$.*

Proof. Given $A < B < C$ in $\mathbb{H}_m(\alpha,\beta)$, take $\lambda \in (\alpha,\beta)$ and small $\epsilon > 0$ and consider the $n \times n$ matrices

$$\widetilde{A} := \begin{bmatrix} A & 0 \\ 0 & (\lambda-\epsilon)I_{n-m}\end{bmatrix}, \qquad \widetilde{B} := \begin{bmatrix} B & 0 \\ 0 & \lambda I_{n-m}\end{bmatrix}, \qquad \widetilde{C} := \begin{bmatrix} C & 0 \\ 0 & (\lambda+\epsilon)I_{n-m}\end{bmatrix}, \tag{3.21}$$

where I_{n-m} is the $(n-m)\times(n-m)$ identity matrix. Then since $\widetilde{A} < \widetilde{B} < \widetilde{C}$ in $\mathbb{H}_n(\alpha,\beta)$ and

$$\begin{aligned}
&\operatorname{Tr}\left(f\left(\widetilde{B}\right)-f\left(\widetilde{A}\right)\right)\left(\widetilde{B}-\widetilde{A}\right)^{-1} \\
&\quad= \operatorname{Tr}(f(B)-f(A))(B-A)^{-1} + (n-m)\frac{f(\lambda)-f(\lambda-\epsilon)}{\epsilon}, \\
&\operatorname{Tr}\left(f\left(\widetilde{C}\right)-f\left(\widetilde{B}\right)\right)\left(\widetilde{C}-\widetilde{B}\right)^{-1} \\
&\quad= \operatorname{Tr}(f(C)-f(B))(B-A)^{-1} + (n-m)\frac{f(\lambda+\epsilon)-f(\lambda)}{\epsilon},
\end{aligned} \tag{3.22}$$

by letting $\epsilon \to 0$ it follows from $(\ddagger_n)$ that

$$\operatorname{Tr}(f(B)-f(A))(B-A)^{-1} \le \operatorname{Tr}(f(C)-f(B))(C-B)^{-1}. \tag{3.23}$$

This completes the proof. □

In view of Lemma 3.2 the essential part of the proof of Theorem 1.2 is in the next lemma.

Lemma 3.3. *If a C^1-function $f(t)$ satisfies $(\ddagger_2)$ on (α,β), then*

$$f'(s) + f'(t) = 2f^{[1]}(s,t) \quad (\alpha < s,t < \beta). \tag{3.24}$$

Proof. Take $B = \mathrm{diag}(t_1, t_2)$ with $t_1, t_2 \in (\alpha, \beta)$. Then for any 2×2 positive definite $X, Y > 0$ and small $\epsilon > 0$ we have by assumption

$$\mathrm{Tr}\frac{f(B) - f(B - \epsilon X)}{\epsilon} X^{-1} \le \mathrm{Tr}\frac{f(B + \epsilon Y) - f(B)}{\epsilon} Y^{-1}. \tag{3.25}$$

Letting $\epsilon \to 0$ by (2.8) this leads to the inequality

$$\mathrm{Tr}\left(\left[f^{[1]}(t_i, t_j) \right]_{i,j=1}^{2} \circ X \right) \cdot X^{-1} \le \mathrm{Tr}\left(\left[f^{[1]}(t_i, t_j) \right]_{i,j=1}^{2} \circ Y \right) \cdot Y^{-1}. \tag{3.26}$$

Replacing X and Y we have also

$$\mathrm{Tr}\left(\left[f^{[1]}(t_i, t_j) \right]_{i,j=1}^{2} \circ Y \right) \cdot Y^{-1} \le \mathrm{Tr}\left(\left[f^{[1]}(t_i, t_j) \right]_{i,j=1}^{2} \circ X \right) \cdot X^{-1}. \tag{3.27}$$

Those together show that

$$\mathrm{Tr}\left(\left[f^{[1]}(t_i, t_j) \right]_{i,j=1}^{2} \circ X \right) \cdot X^{-1} = \text{constant} \quad (0 < X \in \mathbb{H}_2). \tag{3.28}$$

It is easy to see that a 2×2 positive definite matrix X with $\mathrm{Tr}(X) = 1$ is of the form

$$X = \begin{bmatrix} a & u\sqrt{a(1-a)} \\ \overline{u}\sqrt{a(1-a)} & 1-a \end{bmatrix} \quad (0 < a < 1;\ |u| < 1). \tag{3.29}$$

Now it follows (3.28) and (3.29) that

$$\begin{aligned} &\mathrm{Tr}\left(\left[f^{[1]}(t_i, t_j) \right]_{i,j=1}^{2} \circ X \right) \cdot X^{-1} \\ &\quad = \frac{f'(t_1) + f'(t_2) - 2|u|^2 f^{[1]}(t_1, t_2)}{1 - |u|^2} = \text{constant} \quad (|u| < 1), \end{aligned} \tag{3.30}$$

which is possible only when (3.24) is valid. □

Proof of Theorem 1.2. Suppose that a C^2-function $f(t)$ satisfies $(\ddagger_n)$ on (α, β). By Lemmas 3.2 and 3.3 $f(t)$ satisfies the identity (3.24). Therefore we have

$$\{f'(t) + f'(s)\}(t - s) = 2\{f(t) - f(s)\} \quad (\alpha < s, t < \beta). \tag{3.31}$$

Twice differentiating both sides with respect to t we arrive at

$$f'''(t)(t - s) = 0 \quad (\alpha < s, t < \beta) \tag{3.32}$$

which is possible only when $f(t)$ is a quadratic function

$$f(t) = at^2 + bt + c. \tag{3.33}$$

Finally $a \geq 0$ follows from the usual convexity of $f(t)$.

Suppose conversely that $f(t)$ is of the form (3.33) with $a \geq 0$. Take $A < B < C$ in $\mathbb{H}_n$. Then

$$\mathrm{Tr}(f(B) - f(A))(B - A)^{-1} = a\mathrm{Tr}\left(B^2 - A^2\right)(B - A)^{-1} + nb, \tag{3.34}$$

and correspondingly

$$\mathrm{Tr}(f(C) - f(B))(C - B)^{-1} = a\mathrm{Tr}\left(C^2 - B^2\right)(C - B)^{-1} + nb. \tag{3.35}$$

Since

$$B^2 - A^2 = B(B - A) + (B - A)A, \tag{3.36}$$

we have

$$\mathrm{Tr}\left(B^2 - A^2\right)(B - A)^{-1} = \mathrm{Tr}(B) + \mathrm{Tr}(A) \tag{3.37}$$

and correspondingly

$$\mathrm{Tr}\left(C^2 - B^2\right)(C - B)^{-1} = \mathrm{Tr}(C) + \mathrm{Tr}(B). \tag{3.38}$$

Therefore we arrive at the inequality

$$\mathrm{Tr}(f(C) - f(B))(C - B)^{-1} - \mathrm{Tr}(f(B) - f(A))(B - A)^{-1} = a\{\mathrm{Tr}(C) - \mathrm{Tr}(A)\} \geq 0. \tag{3.39}$$

This shows that $f(t)$ satisfies $(\ddagger_n)$ for any n and on any interval (α, β). □

Acknowledgment

The author would like to thank Professor Fumio Hiai for his valuable comments on the original version of this paper.

References

[1] A. W. Roberts and D. E. Varberg, *Convex Functions*, vol. 57 of *Pure and Applied Mathematics*, Academic Press, New York, NY, USA, 1973.

[2] R. Bhatia, *Matrix Analysis*, vol. 169 of *Graduate Texts in Mathematics*, Springer, New York, NY, USA, 1997.

[3] F. Kraus, “Über konvexe Matrixfunktionen,” *Mathematische Zeitschrift*, vol. 41, no. 1, pp. 18–42, 1936.

[4] J. Bendat and S. Sherman, “Monotone and convex operator functions,” *Transactions of the American Mathematical Society*, vol. 79, pp. 58–71, 1955.

[5] F. Hansen and J. Tomiyama, “Differential analysis of matrix convex functions,” *Linear Algebra and Its Applications*, vol. 420, no. 1, pp. 102–116, 2007.

[6] F. Hansen and J. Tomiyama, “Differential analysis of matrix convex functions. II,” *Journal of Inequalities in Pure and Applied Mathematics*, vol. 10, no. 2, article 32, pp. 1–5, 2009.

[7] W. F. Donoghue Jr., *Monotone Matrix Functions and Analytic Continuation*, vol. 207 of *Die Grundlehren der mathematischen Wissenschaften*, Springer, New York, NY, USA, 1974.

[8] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*, Cambridge University Press, Cambridge, UK, 1991.