Cubo Matemática Educacional
Vol., 4. Nº 2, Junio 2002

CONVEX MATRIX FUNCTIONS

Frank Hansen
Institute of Economics, University of Copenhagen,
Studiestraede 6, DK-1455 Copenhagen K, Denmark.

1 Convex functions and Jensen's inequality

Let I be a real interval. A function $f : I \longrightarrow \mathbf{R}$ is said to be convex, if

$$f(\lambda t + (1-\lambda)s) \leq \lambda f(t) + (1-\lambda) f(s) \tag{1.1}$$

for all $t, s \in I$ and every $\lambda \in [0, 1]$. Notice that the definition, in order to be meaningful, requires that f can be evaluated in $\lambda t + (1-\lambda)s$, or equivalently that I is convex. But this is satisfied because the convex subsets of $\mathbf{R}$ are the intervals. If f satisfies (1.1) just for $\lambda = 1/2$, then f is said to be mid-point convex. It is easy to establish that a continuous and mid-point convex function is convex. The geometric interpretation of (1.1) is that the graph of f is below the chord and consequently above the extensions of the chord. This entails that a convex function defined on an open interval is continuous. Condition (1.1) can be reformulated as

$$\frac{f(t)}{(t-s)(t-r)} + \frac{f(s)}{(s-t)(s-r)} + \frac{f(r)}{(r-t)(r-s)} \geq 0 \tag{1.2}$$

for all mutually different numbers $t, s, r \in I$. If we for such numbers define the divided difference $[ts]$ of f taken in the points t, s as

$$[ts]_f = \frac{f(t) - f(s)}{t - s}$$

and the second divided difference $[tsr]$ of f in the points t, s, r as

$$[tsr]_f = \frac{[ts]_f - [sr]_f}{t - r},$$

then the left hand side of (1.2) is equal to $[tsr]_f$. The function f is thus convex, if and only if its second divided differences (evaluated in mutually different points) are nonnegative, or equivalently that the slope of its chords are increasing to the right. If f is differentiable, this is equivalent to the requirement that f' is non-decreasing, and if f is twice differentiable, it is equivalent to $f'' \geq 0$. The form of condition (1.2) shows that the set of convex functions on I is a convex cone which is closed in the weak topology of pointwise convergence. A function f is said to be concave if $-f$ is convex. J.L.W.V. Jensen (1905) [13] proved the following inequality:

Theorem 1.1 (Jensen's inequality) *Let $f : I \longrightarrow \mathbf{R}$ be a convex function defined on a real interval I and let n be any natural number. The inequality*

$$f\left(\sum_{i=1}^{n} \lambda_i t_i\right) \leq \sum_{i=1}^{n} \lambda_i f(t_i) \tag{1.3}$$

is valid for any set of nonnegative real numbers $\lambda_1, \ldots, \lambda_n$ with sum one and all points $t_1, \ldots, t_n \in I$.

The inequality (1.3) reduces for $n = 2$ to the convexity condition (1.1) and it follows in general by induction. The opposite inequality is obtained for concave functions. Jensen realized the importance of his inequality as a vehicle to collect a number of known, but seemingly unrelated inequalities under the same umbrella as well as a generator of many new inequalities, each generated simply by choosing appropriate convex (or concave) functions. The function $f(t) = t^2$ defined on the real line is convex, and Jensen's inequality for this function gives Cauchy's inequality [13, p. 181]

$$\left(\sum_{i=1}^{n} a_i b_i\right)^2 \leq \sum_{i=1}^{n} a_i^2 \sum_{i=1}^{n} b_i^2$$

stated for real numbers $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$. The function $f(t) = \log t$ defined on the real positive half-line is concave, and Jensen's inequality for this function gives Cauchy's inequality

$$(t_1 \cdots t_n)^{1/n} \leq \frac{t_1 + \cdots + t_n}{n}$$

between the geometric and arithmetic means of real positive numbers.

2 The functional calculus

2.1 Functions of one variable

Consider a quadratic matrix A. It is natural to define A^2 as the matrix AA and A^3 as the matrix AAA. In continuation of this idea we set $Ak = A^{k-1}A$ for $k \geq 2$ and

$$p(A) = a_0 E + a_1 A + a_2 A^2 + \cdots + a_k A^k$$

for a polynomium $p(t) = a_0 + a_1 t + a_2 t^2 + \cdots + a_k t^k$ where E denotes the identity matrix. We have thus learned to take a polynomium of a quadratic matrix. Notice that $p(A)$ and A commute.

If in particular A is a real symmetric (and thus hermitian) matrix of order n, we may apply the spectral theorem and write A on the form

$$A = QDQ^{-1}$$

where Q is an orthogonal matrix, that is $Q^{-1} = Q^t$ where Q^t denotes the tranposed of Q, and D is a diagonal matrix with the eigenvalues of A counting multiplicity as diagonal elements. Setting

$$D = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}$$

we calculate

$$A^2 = QDQ^{-1}QDQ^{-1} = QD^2Q^{-1} = Q \begin{pmatrix} \lambda_1^2 & & \\ & \ddots & \\ & & \lambda_n^2 \end{pmatrix} Q^{-1}$$

and more generally

$$A^k = QD^kQ^{-1} = Q \begin{pmatrix} \lambda_1^k & & \\ & \ddots & \\ & & \lambda_n^k \end{pmatrix} Q^{-1}$$

and

$$p(A) = Qp(D)Q^{-1} = Q \begin{pmatrix} p(\lambda_1) & & \\ & \ddots & \\ & & p(\lambda_n) \end{pmatrix} Q^{-1}$$

for any polynomium. We may use this construction to define the functional calculus for any function f defined on the spectrum of A simply by setting

$$f(A) = Q \begin{pmatrix} f(\lambda_1) & & \\ & \ddots & \\ & & f(\lambda_n) \end{pmatrix} Q^{-1}.$$

If f is a polynomium, then this definition of $f(A)$ coincides with the elementary calculation given above.

There is a certain ambiguity in the diagonalisation of A because the diagonal elements in D can be permutated corresponding to permutations of the columns in Q. However, the definition of $f(A)$ is unaffected by this ambiguity. It becomes easier if we consider the spectral representation of A given by

$$A = \sum_{i=1}^{p} \lambda_i P_i$$

where $\lambda_1, \ldots, \lambda_p$ are the eigenvalues of A (not counting multiplicity) and $P_1, \ldots, P_p$ are orthogonal projections with the identity matrix as sum. This representation is unique and

$$f(A) = \sum_{i=1}^{m} f(\lambda_i) P_i.$$

The functional calculus can be extended to self-adjoint operators acting on an infinite-dimensional Hilbertspace, but we will consider the theory only for matrices in order to avoid unnecessary complications.

Definition 2.1 A real function f defined on a real interval I is said to be matrix monotone of order n, if

$$x \leq y \Longrightarrow f(x) \leq f(y)$$

for all hermitian $n \times n$ matrices x, y with spectra contained in I.

We say that a function is operator monotone, if it is matrix monotone of arbitrary order.

2.2 Functions of several variables

Let us consider the function $f(t,s) = ts$ of two variables and two quadratic matrices A and B of orders n and m. We would like to define the matrix $f(A,B)$. What would be a good definition? Korányi [14] proposed that the definition should be the tensor product

$$f(A,B) = A \otimes B$$

of A and B. If

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix}$$

then the tensor product is given by the block matrix

$$A \otimes B = \begin{pmatrix} a_{11}B & \cdots & a_{1n}B \\ \vdots & \ddots & \vdots \\ a_{n1}B & \cdots & a_{nn}B \end{pmatrix}$$

which is of order nm. If $f = f_1 f_2$ can be written as the product of two functions each depending of only one variable, then we set

$$f(A,B) = f_1(A) \otimes f_2(B).$$

The definition can then be extended by linearity in f (we want the mapping $f \longrightarrow f(A,B)$ to be linear) and continuity. We may also extend the definition to functions of more than two variables and obtain:

Definition 2.2 Let $f : I_1 \times \cdots \times I_k \longrightarrow \mathbf{R}$ be a real function (of any kind) defined on a product of real intervals, and let $x = (x_1, \ldots, x_k)$ be a k-tuple of real symmetric matrices such that the eigenvalues of x_i are contained in I_i for $i = 1, \ldots, k$. We say that such a k-tuple is in the domain of f. If

$$x_i = \sum_{t_i=1}^{p_i} \lambda_{t_i i} P_{t_i i} \quad i = 1, \ldots, k$$

is the spectral resolution of x_i, we define

$$f(x) = \sum_{t_i=1}^{p_1} \cdots \sum_{t_k=1}^{p_k} \cdot f(\lambda_{t_1}, \ldots, \lambda_{t_k}) P_{t_1 1} \otimes \cdots \otimes P_{t_k k}$$

as the function f applied to the k-tuple $x = (x_1, \ldots, x_k)$.

If the k-tuple $x = (x_1, \ldots, x_k)$ is of order $(n_1, \ldots, n_k)$, then $f(x)$ is a real symmetric matrix of order $n_1, \ldots, n_k$.

It is not obvious to extend the concept of monotonicity from functions of one variable to functions of several variables. This is so because there is no natural order structure on tuples of matrices. It is completely trivial to define the notion of matrix convexity for functions of several variables, simply because the definition of matrix convexity only involves the order structure for matrices.

Definition 2.3 A function $f : I_1 \times \cdots \times I_k \longrightarrow \mathbf{R}$ defined on a product of real intervals is said to be matrix convex of order $(n_1, \ldots, n_k)$, if the matrix inequality

$$f(\lambda x + (1 - \lambda) y) \leq \lambda f(x) + (1 - \lambda) f(y)$$

holds for any $\lambda \in [0, 1]$ and all k-tuples of matrices $x = (x_1, \ldots, x_k)$ and $y = (y_1, \ldots, y_k)$ of order $(n_1, \ldots, n_k)$ in the domain of f.

The definition is meaningful since also $\lambda x + (1 - \lambda) y$ is contained in the domain of f. We say that f is operator convex, if f is matrix convex of every order $(n_1, \ldots, n_k)$.

3 Some matrix inequalities

Löwner (1934) [16] proved that a function defined on an open interval is operator monotone, if and only if it allows an analytic continuation into the complex upper half-plane with nonnegative imaginary part, that is an analytic continuation to a Pick function [3]. The function $f(t) = t^2$ is not even matrix monotone of order 2 on any interval, while the function $f(t) = t^\alpha$ is operator monotone on the positive half-axis, that is

$$x \leq y \Longrightarrow x^\alpha \leq y^\alpha \qquad \alpha \in [0,1] \tag{3.4}$$

for positive matrices x, y. The result follows from Löwner's theorem, but this was not realized at the time when Löwner published his article. The inequality was independently proved by Heinz (1951) [12], and it is today known as the Löwner-Heinz inequality.

A number of different proofs of the Löwner-Heinz inequality are known. The following proof gives a useful integral decomposition of the inequality in terms of functions belonging to the extremal rays in the convex set of operator monotone functions defined on $[0,\infty[$. Since inversion is matrix decreasing, the functions $t \longrightarrow t\,(t+\lambda)^{-1}$ are operator monotone on $[0,\infty[$ for $\lambda \geq 0$. The identity

$$t^\alpha = \frac{\sin \alpha\pi}{\pi} \int_0^\infty \frac{t}{t+\lambda} \lambda^{\alpha-1} d\lambda \qquad 0 < \alpha < 1 \tag{3.5}$$

follows by making the substitution $x \longrightarrow t^{-1}\lambda$ in [5, Integral no. 3.222 (2)]. We thus obtain that $f(t) = t^\alpha$ is operator monotone for $\alpha \in [0,1]$. Pedersen (1972) [17] gave a truly elementary proof of the Löwner-Heinz inequality.

Let f be a real function defined on the positive half-line with $f(0) \geq 0$. If f is matrix monotone of order $2n$, then the inequality between $n \times n$ matrices

$$f(a^* x a) \geq a^* f(x) a \tag{3.6}$$

is valid [6] for any contraction a and any real symmetric matrix x in the domain of f. If on the other hand the same inequality is satisfied for $2n \times 2n$ matrices, then f is matrix monotone of order n, cf. [10, p. 233]. Notice that the function $f(t) = t^2$ is matrix monotone of order 1 on $[0,\infty[$ but does not satisfy the inequality for $x = 1$ and $a = 1/2$.

Davis (1957) [2] proved that a continuous function defined on an interval containing zero satisfies the pinching inequality

$$pf(pxp)p \leq pf(x)p \tag{3.7}$$

for orthogonal projections p and real symmetric matrices x in the domain of f, if and only if it is operator convex.

Theorem 3.1 *A function f defined on an interval I containing zero is operator convex with $f(0) \le 0$, if and only if it satisfies the matrix inequality*

$$f(a^*xa) \ge a^*f(x)a \tag{3.8}$$

for contractions a and real symmetric matrices x in the domain of f.

It is essential in the two above theorems that no bounds are given on the order of the matrices. The last result is known as Jensen's matrix inequality for functions of one variable [10, 8]. The above matrix inequalities have extensions to functions of several variables [1].

4 Differentiable matrix functions

Let X and Y be Banach spaces. We say that a function $f : A \longrightarrow Y$ defined on a subset A of X is Fréchet differentiable at an inner point $x_0 \in A$, if there exists a bounded linear operator $df(x_0) \in B(X,Y)$ such that

$$\lim_{h \to 0} \|h\|^{-1} \left(f(x_0 + h) - f(x_0) - df(x_0)h \right) = 0.$$

Likewise f is said to be Fréchet differentiable in an open set A, if f is Fréchet differentiable at every point $x_0 \in A$. We say that f is continuously Fréchet differentiable, if the differential mapping $Ax \longrightarrow df(x) \in B(X,Y)$ is continuous. The exponential function in a Banach algebra is an example of a continuously Fréchet differentiable function.

Proposition 4.1 *If A is a Banach algebra, then the exponential function $x \longrightarrow exp(x)$ is continuously Fréchet differentiable, and*

$$d\exp(x)h = \int_0^1 \exp(sx)h\exp((1-s)x)\,ds$$

for all x and h in A.

An interesting and non-trivial question is to specify conditions under which the map $T \longrightarrow f(T)$ is Fréchet differentiable, where f is a real function defined on an open interval I and T is in the domain of f. It is a necessary but not sufficient condition that f is continuously differentiable, cf. [19]. However, if we restrict ourselves to consider hermitian matrices of a certain fixed order, then the function $T \longrightarrow f(T)$ is Fréchet differentiable if and only if f is continuously differentiabe and

$$df(T)S = f^{[1]}(T) \circ S \tag{4.9}$$

where $\circ$ is the Hadamard product and

$$f^{[1]}(T) = ([\lambda_i \lambda_j]_f)_{i,j=1}^n \tag{4.10}$$

is the Löwner matrix [11] defined from the (not necessarily distinct) eigenvalues $\lambda_1, \ldots, \lambda_n$ of T. The definition of the divided differences introduced in section 1 for an arbitrary real function defined on an interval I and mutually different real numbers in I can for a differentiable function defined on an open interval be extended by setting

$$[ts]_f = \begin{cases} \dfrac{f(t)-f(s)}{t-s} & \text{for} \quad t \neq s \\ f'(t) & \text{for} \quad t = s \end{cases} \tag{4.11}$$

for any$t, s \in I$. It is tacitly assumed in equation (4.9) that the Fréchet differential is identified with its matrix representation in a basis of eigenvectors for T corresponding to the eigenvalues $\lambda_1, \ldots, \lambda_n$.

The notion of higher order Fréchet differentiability can be defined in a very natural way. The Fréchet differential df of a Fréchet differentiable function $f : A \longrightarrow Y$ defined on an open subset $A \subseteq X$ is a function from A into the Banach space $B(X, Y)$ of bounded linear functions from X to Y.

If df is Fréchet differentiable, then we define the second Fréchet differential of f, denoted by $d^2 f$, to be the Fréchet differential of df. The second order Fréchet differential can be considered as a function $d^2 f : A \longrightarrow B_2(X, Y)$ from A into the Banach space of bounded bilinear functions from X to Y. The second Fréchet differential $d^2 f(x)$ is symmetric in the sense that $d^2 f(x)(h,k) = d^2 f(x)(k,h)$, cf. [4].

Proposition 4.2 *Let A be an open convex subset of a real Banach space X and let H be a Hilbert space. A twice Fréchet differentiable function $f : A \longrightarrow B(H)_{sa}$ is convex, if and only if*

$$d^2 f(x)(h,h) \geq 0$$

for each $x \in A$ and all $h \in X$.

The result follows by adapting the reasoning of classical analysis to the present situation and can be found in [4, Exercises 3.1.8 and 3.6.4]. The author [7] gave the following sufficient conditions for second order Fréchet differentiability.

Theorem 4.3 *Let $f \in C^p(I)$ where $I = I_1 \times \cdots \times I_k$ is a product of open intervals and $p > 2 + k/2$. The function $x \longrightarrow f(x)$ defined on k-tuples of self-adjoint matrices $x = (x_1, \ldots, x_k)$ in the domain of f is twice continuously Fréchet differentiable.*

The above theorem can be extended, ad verbatim, from matrices to self-adjoint operators on Hilbert spaces. For matrices of a fixed order the condition $p > 2 + k/2$ may be somewhat relaxed. However, we shall make use of a type of Fourier expansion of the second Fréchet differential for which the stated condition is the proper one.

5 The second Fréchet differential

Let f be a twice differentiable real function defined on an open interval $I \subseteq \mathbf{R}$. The divided difference $[\lambda\mu]_f$ of f in the points $\lambda, \mu \in I$ defined in equation (4.11) is a symmetric function of the two arguments with partial derivatives in each of the two variables. The second divided difference $[\lambda\mu\zeta]_f$ of f in the points λ, μ, $\zeta \in I$ is defined as

$$[\lambda\mu\zeta]_f = \begin{cases} \dfrac{[\lambda\mu]_f - [\mu\zeta]_f}{\lambda - \zeta} & \text{for } \lambda \neq \zeta \\ \dfrac{\partial}{\partial\lambda}[\lambda\mu]_f & \text{for } \lambda = \zeta \neq \mu \\ \dfrac{1}{2} f''(\lambda) & \text{for } \lambda = \zeta = \mu. \end{cases}$$

It is a symmetric and continuous function of the three arguments.

Let $f : I_1 \times \cdots \times I_k \longrightarrow \mathbf{R}$ be a function defined on a product of open intervals with continuous partial derivatives up to the second order. We define the partial divided difference of f in the points μ_1 and μ_2 by setting

$$[\lambda_1|\cdots|\mu_1\mu_2|\cdots|\lambda_k]_f^i = [\mu_1\mu_2]_g$$

where

$$g(\mu) = f(\lambda_1, \ldots, \lambda_{i-1}, \mu, \lambda_{i+1}, \ldots, \lambda_k).$$

We similarly define the second partial divided difference of f in the points μ_1, μ_2 and μ_3 by setting

$$[\lambda_1|\cdots|\mu_1\mu_2\mu_3|\cdots|\lambda_k]_f^i = [\mu_1\mu_2\mu_3]_g.$$

Partial divided differences are similar to partial derivatives, and we may define also mixed second partial divided differences for a function of several variables. They are defined by setting

$$[\lambda_1|\cdots|\mu_1\mu_2|\cdots|\xi_1\xi_2|\cdots|\lambda_k]_f^{ij}$$

$$= \frac{[\lambda_1|\cdots|\mu_1|\cdots|\xi_1\xi_2|\cdots|\lambda_k]_f^j - [\lambda_1|\cdots|\mu_2|\cdots|\xi_1\xi_2|\cdots|\lambda_k]_f^j}{\mu_1 - \mu_2}$$

for $\mu_1 \neq \mu_2$ and otherwise

$$[\lambda_1|\cdots|\mu\mu|\cdots|\xi_1\xi_2|\cdots|\lambda_k]_f^{ij} = \frac{\partial}{\partial\mu}[\lambda_1|\cdots|\mu|\cdots|\xi_1\xi_2|\cdots|\lambda_k]_f^j$$

where the μ's are in position i and the ξ's are in position j for $i \neq j$. The notation does not imply any particular order of the coordinates i and j. We have defined the mixed second partial divided differences by first dividing in coordinate j and then in coordinate i, but we get the same result by reversing the order. This can be considered as a generalization of Young's theorem for partial differentials.

An element $\Lambda \in I_1^{n_1} \times \cdots \times I_k^{n_k}$ is called a data set of order $(n_1, \ldots, n_k)$ for f. We may write it in the form

$$\Lambda = \left((\lambda_{m_1}(1))_{m_1=1}^{n_1}, \ldots, (\lambda_{m_k}(k))_{m_k=1}^{n_k}\right). \tag{5.12}$$

A data set $\Lambda(x) \in I_1^{n_1} \times \cdots \times I_k^{n_k}$ can be obtained from a k-tuple $x = (x_1, \ldots, x_k)$ of self-adjoint matrices of order $(n_1, \ldots, n_k)$ in the domain of f by choosing the numbers $\lambda_1(i), \ldots, \lambda_{n_i}(i)$ as the (possibly) degenerate eigenvalues of x_i for each $i = 1, \ldots, k$.

5.1 Generalized Hessian matrices

Definition 5.1 Let $f : I_1 \times \cdots \times I_k \longrightarrow \mathbf{R}$ be a function defined on a product of open intervals with continuous partial derivatives up to the second order, and let

$$\Lambda = \left((\lambda_{m_1}(1))_{m_1=1}^{n_1}, \dots, (\lambda_{m_k}(k))_{m_k=1}^{n_k}\right)$$

be a data set of order $(n_1, \dots, n_k)$ for f. We define for each k- tuple of natural numbers $(m_1, \dots, m_k) \le (n_1, \dots, n_k)$ a generalized Hessian matrix $H(m_1, \dots, m_k)$ associated with f and the data set Λ by setting

$$H(m_1, \dots, m_k) = (H_{us}(m_1, \dots, m_k))_{u,s=1,\dots,k}$$

where $H_{us}(m_1, \dots, m_k)$ is a $n_u \times n_s$ matrix with entry

$$H_{us}(m_1, \dots, m_k)_{p_u j_s}$$

$$= [\lambda_{m_1}(1)| \cdots |\lambda_{m_s}(s)\lambda_{j_s}(s)| \cdots |\lambda_{p_u}(u)\lambda_{m_u}(u)| \cdots |\lambda_{m_k}(k)]_f^{su}$$

for $s \neq u$ and entry

$$H_{ss}(m_1, \dots, m_k)_{p_s j_s} = 2[\lambda_{m_1}(1)| \dots |\lambda_{m_s}(s)\lambda_{p_s}(s)\lambda_{j_s}(s)| \dots |\lambda_{m_k}(k)]_f^s$$

for $u = s$.

The generalized Hessian $H(m_1, \dots, m_k)$ is a real and symmetric matrix of order $n_1 + \cdots + n_k$. If the order of the data set Λ is $(1, \dots, 1)$, then there is only one generalized Hessian matrix $H(1, \dots, 1)$ and the data set Λ is reduced to the k numbers $\lambda_1(1), \dots, \lambda_1(k)$. The submatrix $H_{us}(1, \dots, 1)$ is a 1×1 matrix with the partial derivative $f''_{us}(\lambda_1(1), \dots, \lambda_1(k))$ as entry, and $H(1, \dots, 1)$ identifies with the usual Hessian matrix of f at the point $(\lambda_1(1), \dots, \lambda_1(n))$.

Let us look at some examples. For typographic reasons we will in some formulas write λ_i^s as shorthand for $\lambda_i(s)$ and ignore the intrinsic ambiguity in the symbol. We may also omit the subscript representing the function f in the divided differences when there is no possibility of confusion. If we set $k = 2$ and $n_1 = n_2 = 2$, then

$$H(1,1) = \begin{pmatrix} f''_{11}(\lambda_1^1,\lambda_1^2) & 2[\lambda_1^1\lambda_1^1\lambda_2^1|\lambda_1^2] & f''_{12}(\lambda_1^1,\lambda_1^2) & [\lambda_1^1\lambda_2^1|\lambda_1^2\lambda_1^2] \\ 2[\lambda_1^1\lambda_2^1\lambda_1^1|\lambda_1^2] & 2[\lambda_1^1\lambda_2^1\lambda_2^1|\lambda_1^2] & [\lambda_1^1\lambda_1^1|\lambda_2^2\lambda_1^2] & [\lambda_1^1\lambda_2^1|\lambda_2^2\lambda_1^2] \\ f''_{21}(\lambda_1^1,\lambda_1^2) & [\lambda_1^1\lambda_1^1|\lambda_1^2\lambda_2^2] & f''_{22}(\lambda_1^1,\lambda_1^2) & 2[\lambda_1^1|\lambda_1^2\lambda_1^2\lambda_2^2] \\ [\lambda_2^1\lambda_1^1|\lambda_1^2\lambda_1^2] & [\lambda_2^1\lambda_1^1|\lambda_1^2\lambda_2^2] & 2[\lambda_1^1|\lambda_1^2\lambda_2^2\lambda_1^2] & 2[\lambda_1^1|\lambda_1^2\lambda_2^2\lambda_2^2] \end{pmatrix}$$

and

$$H(1,2) = \begin{pmatrix} f''_{11}(\lambda_1^1,\lambda_1^2) & 2[\lambda_1^1\lambda_1^1\lambda_2^1|\lambda_2^2] & [\lambda_1^1\lambda_1^1|\lambda_2^2\lambda_1^2] & f''_{12}(\lambda_1^1,\lambda_2^2) \\ 2[\lambda_1^1\lambda_2^1\lambda_1^1|\lambda_2^2] & 2[\lambda_1^1\lambda_2^1\lambda_2^1|\lambda_2^2] & [\lambda_2^1\lambda_1^1|\lambda_2^2\lambda_1^2] & [\lambda_2^1\lambda_1^1|\lambda_2^2\lambda_2^2] \\ [\lambda_1^1\lambda_1^1|\lambda_1^1\lambda_2^2] & [\lambda_1^1\lambda_2^1|\lambda_1^2\lambda_2^2] & 2[\lambda_1^1|\lambda_2^2\lambda_1^2\lambda_1^2] & 2[\lambda_1^1|\lambda_2^2\lambda_1^2\lambda_2^2] \\ f''_{21}(\lambda_1^1,\lambda_2^2) & [\lambda_1^1\lambda_2^1|\lambda_2^2\lambda_2^2] & 2[\lambda_1^1|\lambda_2^2\lambda_2^2\lambda_1^2] & f''_{22}(\lambda_1^1,\lambda_2^2) \end{pmatrix}$$

Similarly for $H(2,1)$ and $H(2,2)$. Notice that the ordinary Hessian

$$H = \begin{pmatrix} f''_{11}(\lambda_1^1,\lambda_1^2) & f''_{21}(\lambda_1^1,\lambda_2^2) \\ f''_{12}(\lambda_1^1,\lambda_2^2) & f''_{22}(\lambda_1^1,\lambda_2^2) \end{pmatrix}$$

is the principal submatrix of $H(1,1)$ obtained by selecting rows and columns number 1 and 3. Likewise, it appears as the principal submatrix of $H(1,2)$ obtained by selecting rows and columns number 1 and 4. This is a reflection of a completely general phenomenon as we shall see in the next section. If we consider a generalized Hessian matrix $H(m_1,\dots,m_k)$ associated with a data set Λ written on the form 5.12, then the ordinary Hessian of the function f at the point $(H_{m_1}(1),\dots,H_{m_k}(k))$ appears as the $k\times k$ principal submatrix of $H(m_1,\dots,m_k)$ obtained by selecting the rows and columns numbered by $m_1, n_1+m_2, n_1+n_2+m_3,\dots,n_1+\cdots+n_{k-1}+m_k$.

We have the following structure theorem for the second Fréchet differential of the functional calculus mapping.

Theorem 5.2 *Let $f \in C^p(I_1\times\cdots\times I_k)$ where $I_1,\dots,I_k$ are open intervals and $p > 2+k/2$. The function $x \longrightarrow f(x)$ defined on k-tuples of matrices in the domain of f is twice Fréchet differentiable. Suppose that $x=(x_1,\dots,x_k)$ acts on fixed Hilbert spaces $H_1,\dots,H_k$ of dimensions $(n_1,\dots,n_k)$. A vector φ in the tensor product $H_1\otimes\cdots\otimes H_k$ is specified by a function $\varphi(m_1,\dots,m_k)$ of k natural numbers by setting*

$$\varphi = \sum_{m_1=1}^{n_1}\cdots\sum_{m_k=1}^{n_k}\varphi(m_1,\dots,m_k)e^1_{m_1}\otimes\cdots\otimes e^k_{m_k}$$

where $(e_1^s, \ldots, e_{n_s}^s)$ *is an orthonormal basis for* H_s *consisting of eigenvectors for* x_s *for each* $s = 1, \ldots, k$. *The expectation value is given by*

$$
\begin{aligned}
&(d^2 f(x)(h,h)\varphi|\varphi) \\
&= \sum_{m_1=1}^{n_1} \cdots \sum_{m_k=1}^{n_k} \left(H(m_1, \ldots, m_k)\Phi^h(m_1, \ldots, m_k) | \Phi^h(m_1, \ldots, m_k)\right)
\end{aligned}
$$

where $H(m_1, \ldots, m_k)$ *are the generalized Hessian matrices associated with* f *and* $\Lambda(x)$. *The vectors*

$$
\Phi^h(m_1, \ldots, m_k) = \begin{pmatrix} \Phi_1^h(m_1, \ldots, m_k) \\ \vdots \\ \Phi_k^h(m_1, \ldots, m_k) \end{pmatrix}
$$

are for each $s = 1, \ldots, k$ *and* $m_s = 1, \ldots, n_s$ *given by setting the entry*

$$
\Phi_s^h(m_1, \ldots, m_k)_{j_s} = h_{m_s j_s}^s \varphi(m_1, \ldots, m_{s-1}, j_s, m_{s+1}, \ldots, m_k)
$$

for $j_s = 1, \ldots, n_s$.

5.2 Variant Hessian matrices

Let $f \in C^2(I_1, \ldots, I_k)$ where $I_1, \ldots, I_k$ are open intervals. We consider a generalized Hessian matrix

$$
H(m_1, \ldots, m_k) = (H_{us}(m_1, \ldots, m_k))_{u,s=1}^k
$$

associated with f and matrices $x = (x_1, \ldots, x_k)$ of order $(n_1, \ldots, n_k)$ in the domain of f for a k-tuple $(m_1, \ldots, m_k) \le (n_1, \ldots, n_k)$. The possibly degenerate eigenvalues of x_s are denoted $\lambda_1(s), \ldots, \lambda_{n_s}(s)$ for $s = 1, \ldots, k$. Each entry $H_{us}(m_1, \ldots, m_k)$ is an $n_u \times n_s$ matrix with its entry labelled by the indices p_u and j_s ($p_u = 1, \ldots, n_u, j_s = 1, \ldots, n_s$).

We first introduce a principal submatrix of $H(m_1, \ldots, m_k)$ by selecting row m_u in the blocks $H_{u1}(m_1, \ldots, m_k), \ldots, H_{uk}(m_1, \ldots, m_k)$ for $u = 1, \ldots, k$ and column m_s in the blocks $H_{1s}(m_1, \ldots, m_k), \ldots, H_{ks}(m_1, \ldots, m_k)$ for $s = 1, \ldots, k$. Indeed, these rows and colunms are numbered $m_1, n_1 + m_2, n_1 + n_2 + m_3, \ldots, n_1 + \cdots + n_{k-1} + m_k$ in $H(m_1, \ldots, m_k)$. Namely, for each u, s we retain from the $n_u \times n_s$ matrix $H_{us}(m_1, \ldots, m_k)$ just one entry with the index $(p_u, j_s) = (m_u, m_s)$. The resulting principal submatrix is a $k \times k$ matrix with entries labelled by u, s and given by

$$\begin{cases} [\lambda^1_{m_1}|\cdots|\lambda^s_{m_s}\lambda^s_{m_s}|\cdots|\lambda^u_{m_u}\lambda^u_{m_u}|\cdots|\lambda^k_{m_k}]^{su} & s \neq u \\ 2[\lambda^1_{m_1}|\cdots|\lambda^s_{m_s}\lambda^s_{m_s}\lambda^s_{m_s}|\cdots|\lambda^k_{n_k}]^s & s = u \end{cases}$$

$$= \begin{cases} f''_{us}\left(\lambda^1_{m_1}, \ldots, \lambda^k_{m_k}\right) & s \neq u \\ f''_{ss}\left(\lambda^1_{m_1}, \ldots, \lambda^k_{m_k}\right) & s = u \end{cases}$$

It is thus the usual Hessian matrix for f at the point $(\lambda^1_{m_1}, \ldots, \lambda^k_{m_k})$.

Next we define a so called variant Hessian matrix as the block matrix

$$V(m_1, \ldots, m_k) = (V_{us}(m_1, \ldots, m_k))^k_{u,s=1}$$

where $V_{us}(m_1, \ldots, m_k)$ for each u, s is obtained from $H_{us}(m_1, \ldots, m_k)$ by retaining the following entries and replacing all other entries by zero. The retained entries are all diagonal entries of the diagonal blocks $H_{ss}(m_1, \ldots, m_k)$ and one entry $(p_u, j_s) = (m_u, m_s)$ from each of the off-diagonal blocks $H_{us}(m_1, \ldots, m_k)$. Notice that the entry $(p_s, j_s) = (m_s, m_s)$ in the diagonal block $H_{ss}(m_1, \ldots, m_k)$ is a diagonal entry and thus retained. The resulting variant Hessian matrix is an orthogonal direct sum of the principal submatrix constructed above and a diagonal matrix. The variant Hessians are symmetric matrices of order $n_1 + \cdots + n_k$. If we set $k = 2$ and $n_1 = n_2 = 2$, then

$$V(1,1) = \begin{pmatrix} f''_{11}(\lambda^1_1, \lambda^2_1) & 0 & f''_{12}(\lambda^1_1, \lambda^2_1) & 0 \\ 0 & 2[\lambda^1_1\lambda^1_2\lambda^1_2|\lambda^2_1] & 0 & 0 \\ f''_{21}(\lambda^1_1, \lambda^2_1) & 0 & f''_{22}(\lambda^1_1, \lambda^2_1) & 0 \\ 0 & 0 & 0 & 2[\lambda^1_1|\lambda^2_1\lambda^2_2\lambda^2_2] \end{pmatrix}$$

and

$$V(1,2) = \begin{pmatrix} f''_{11}(\lambda^1_1, \lambda^2_2) & 0 & 0 & f''_{12}(\lambda^1_1, \lambda^2_2) \\ 0 & 2[\lambda^1_1\lambda^1_2\lambda^1_2|\lambda^2_2] & 0 & 0 \\ 0 & 0 & 2[\lambda^1_1|\lambda^2_2\lambda^2_1\lambda^2_1] & 0 \\ f''_{21}(\lambda^1_1, \lambda^2_2) & 0 & 0 & f''_{22}(\lambda^1_1, \lambda^2_2) \end{pmatrix}$$

Similarly for $V(2,1)$ and $V(2,2)$.

6 Convexity theorems

6.1 Operator convex functions

The Combination of Theorem 5.2 and Proposition 4.2 entails:

Proposition 6.1 *Let $f \in C^p(I_1 \times \cdots \times I_k)$ where $I_1, \ldots, I_k$ are open intervals and $p > 2 + k/2$. If for a k-tuple $(n_1, \ldots, n_k)$ of natural numbers all of the generalized Hessian matrices associated with f and any data set $\Lambda \in I_1^{n_1} \times \cdots \times I_k^{n_k}$ are positive semi-definite, then f is matrix convex of order $(n_1, \ldots, n_k)$.*

It is thus a sufficient condition for f to be operator convex that all of the generalized Hessian matrices associated with f and any data set of any order are positive semi-definite.

The conditions given in Proposition (6.1) for matrix convexity of a fixed order $(n_1, \ldots, n_k)$ of the function f are also necessary, if either $(n_1, \ldots, n_k) = (1, \ldots, 1)$ or $k = 1$. The former result is a well known part of classical analysis, while the latter is due to Kraus [15]. It is unknown whether the conditions of Proposition 6.1 are necessary for the matrix convexity of f in any other case.

Let $\mu_1, \ldots, \mu_k \in [-1, 1]$ and consider the function

$$f(t_1, \ldots, t_k) = \prod_{i=1}^{k} \frac{1}{1 - \mu_i t_i} \qquad t_1, \ldots, t_k \in]-1, 1[$$

of k variables. It is an exercise [7, p. 461] to calculate the generalized Hessian matrices associated with f and any data set

$$\Lambda \in]-1, 1[^{n_1} \times \cdots \times]-1, 1[^{n_k} \qquad n_1, \ldots, n_k \in \mathbf{N}$$

given on the form (5.12). Indeed, the blocks in the generalized Hessian matrices are given by

$$H_{us}(m_1, \ldots, m_k) = f(\lambda_{m_1}(1), \ldots, \lambda_{m_k}(k)) a(u)^t a(s)$$

for $s \neq u$, and

$$H_{ss}(m_1, \ldots, m_k) = 2 f(\lambda_{m_1}(1), \ldots, \lambda_{m_k}(k)) a(s)^t a(s)$$

for $s = u$, where the vector $a(i) \in \mathbf{R}^{n_i}$ is given by

$$a(i) = \mu_i \left((1-\mu_i\lambda_1(i))^{-1}, \ldots, (1-\mu_i\lambda_{n_i}(i))^{-1}\right)$$

for $i = 1, \ldots, k$ and $a(i)^t$ denotes the transpose of $a(i)$. Consequently, for each k-tuple $(m_1, \ldots, m_k) \le (n_1, \ldots, n_k)$ the generalized Hessian matrix $H(m_1, \ldots, m_k)$ is the product of $f(\lambda_{m_1}(1), \ldots, \lambda_{m_k}(k))$, which is a positive factor depending on the k-tuple $(m_1, \ldots, m_k)$, and the matrix

$$\begin{pmatrix} 2a(1)^t a(1) & a(1)^t a(2) & \ldots & a(1)^t a(k) \\ a(2)^t a(1) & 2a(2)^t a(2) & \ldots & a(2)^t a(k) \\ \vdots & \vdots & \ddots & \vdots \\ a(k)^t a(1) & a(k)^t a(2) & \ldots & 2a(k)^t a(k) \end{pmatrix}$$

which is independent of $(m_1, \ldots, m_k)$. The latter matrix is bounded from below by the positive semi-definite matrix

$$\begin{pmatrix} a(1)^t a(1) & \ldots & a(1)^t a(k) \\ \vdots & \ddots & \vdots \\ a(k)^t a(1) & \ldots & a(k)^t a(k) \end{pmatrix}$$

$$= (a(1) \cdots a(k))^t \, (a(1) \cdots a(k))$$

We conclude from Proposition 6.1 that f is matrix convex of every order $(n_1, \ldots, n_k)$ and hence operator convex. The next result is an immediate consequence.

Theorem 6.2 *Let k be a natural number, and let ν be a nonnegative Borel measure on the compact cube $[-1,1]^k$. The function*

$$f(t_1, \ldots, t_k) = \int_{-1}^{1} \cdots \int_{-1}^{1} \prod_{i=1}^{k} \frac{1}{1-\mu_i t_i} d\nu(\mu_1, \ldots, \mu_k).$$

is operator convex on the open cube $]-1,1[^k$.

6.2 Convex trace functions

Proposition 6.3 *A function $f \in C^2(I_1, \ldots, I_k)$ defined on a product of open intervals $I_1, \ldots, I_k$ is convex, if and only if all variant Hessian matrices asociated with f are positive semi-definite.*

Proof: The variant Hessian matrix is, as explained above, an orthogonal direct sum of the principal Hessian matrix constructed above and a diagonal matrix. The positive semi-definiteness of the variant Hessian matrix, therefore, is equivalent to the same for the two (mutually orthogonal) submatrices. The first one is the usual Hessian matrix of an ordinary function f as displayed above and hence its positive semi-definiteness is equivalent to the ordinary convexity of the function f, as is well-known. On the other hand, the (diagonal) entries of the diagonal submatrix are partial second divided differences which are all non-negative if f is convex. Therefore the assertion follows. **QED**

Theorem 6.4 *Let $f \in C^p(I_1 \times \cdots \times I_k)$ where $I_1, \ldots, I_k$ are open intervals and $p > 2 + k/2$. The function $x \longrightarrow f(x)$ is twice Fréchet differentiable in the domain of f. We define to each k-tuple $h = (h_1, \ldots, h_k)$ of self-adjoint matrices of order $(n_1, \ldots, n_k)$ and each k-tuple of natural numbers $(m_1, \ldots, m_k) \le (n_1, \ldots, n_k)$ the block vector*

$$\Psi^h(m_1, \ldots, m_k) = \begin{pmatrix} \Psi_1^h(m_1) \\ \vdots \\ \Psi_k^h(m_k) \end{pmatrix}$$

where $\Psi_s^h(m_s)_{j_s} = h^s_{m_s j_s}$ is the (m_s, j_s)-entry of the matrix variable h^s for $j_s = 1, \ldots, n_s$ and $s = 1, \ldots, k$. The trace of the second Fréchet differential for each k-tuple $x = (x_1, \ldots, x_k)$ of self-adjoint matrices of order $(n_1, \ldots, n_k)$ in the domain of f is given by

$$\begin{gathered} Tr\,(d^2 f(x)(h,h)) \\ = \sum_{m_1=1}^{n_1} \cdots \sum_{m_k=1}^{n_k} \left(V(m_1, \ldots, m_k)\Psi^h(m_1, \ldots, m_k) \mid \Psi^h(m_1, \ldots, m_k)\right) \end{gathered}$$

where $V(m_1, \ldots, m_k)$ for $m_s = 1, \ldots, n_s$ and *$s = l, \ldots, k$ are the variant Hessian matrices associated with f and the matrices $x = (x_1, \ldots, x_k)$.*

Theorem 6.5 *Let f be a convex function defined on a product $I_1 \times \cdots \times I_k$ of open intervals. The trace function*

$$(x_1, \ldots, x_k) \longrightarrow Tr\,(f(x_1, \ldots, x_k))$$

is convex on k-tuples of symmetric matrices in the domain of f.

Proof: Suppose that f is continuously differentiable of order $p > 2+k/2$. It is not difficult [9] to establish that the trace function above is twice Fréchet differentiable with $Tr(d^2 f(x_1, \ldots, x_k)))$ as second Fréchet differential. We then combine Theorem 6.4 and Proposition 6.3 to obtain the desired result. In the general case we approximate f with a sequence of convex C^∞-functions (f_n), which may be chosen of the form $f_n = f * e_n$ for a C^∞-approximate unit e_n with compact support.

Bibliography

[1] **H. Araki and F. Hansen.** Jensen's operator inequality for functions of several variables. *Proc. Amer. Math- Soc.*, 128:2075-2084, 2000.

[2] **C. Davis.** A Schwarz inequality for convex operator functions. *Proc. Amer. Math. Soc.*, 8:42-44, 1957.

[3] **W. Donoghue.** *Monotone matrix functions and analytic continuation.* Springer, Berlin, Heidelberg, New York, 1974.

[4] **T.M. Flett.** *Differential Analysis.* Cambridge University Press, Cambridge, 1980.

[5] **I.S. Gradshteyn and I.M. Ryzhik.** *Table of Integrals, Series, and Products.* Academic Press, 1965.

[6] **F. Hansen.** An operator inequality. *Math. Ann.*, 246:249-250, 1980.

[7] **F. Hansen.** Operator convex functions of several variables. *Publ. RIMS*, 33:443-463, 1997.

[8] **F. Hansen.** Operator inequalities associated with Jensen's inequality. In T.M. Rassias, editor, *Survey on Classical Inequalities*, pages 67-98. Kluwer Academic Publishers, 2000.

[9] **F. Hansen.** Convex trace functions of several variables. *Linear Algebra Appl.*, to appear,2001.

[10] **F. Hansen and G.K. Pedersen.** Jensen's inequality for operators and Löwner's theorem. *Math. Ann.*, 258:229-241, 1982.

[11] **F. Hansen and G.K. Pedersen.** Perturbation formulas for traces on C*-algebras. *Publ. RIMS, Kyoto Univ.*, 31:169-178, 1995.

[12] **E. Heinz.** Beiträge zur Störungstheorie der Spectralzerlegung. *Math. Ann.*, 123:415-438, 1951.

[13] **J.L.W.V. Jensen.** Sur les fonction convexes et les inégalités entre les valeurs moyennes. *Acta Sci. Math-(Szeged)*, 30:175-193,1906.

[14] **A. Korányi.** On some classes of analytic functions of several variables. *Trans Amer. Math. Soc.*, 101:520-554, 1961.

[15] **F. Kraus.** Über konvekse Matrixfunktionen. *Math. Z.*, 41:18-42,1936.

[16] **K. Löwner.** Über monotone Matrixfunktionen. *Math. Z.*, 38:177-216, 1934.

[17] **G.K. Pedersen.** Some operator monotone functions. *Proc. Amer. Math. Soc.*, 36:309-310,1972.

[18] **G.K. Pedersen.** Operator differentiable functions. *Math. Inequalities and Appl., to appear,* 1999.

[19] **H. Widom.** When are differentiable functions differentiable? *In Linear and Complex Analysis Problem Book 3, Part 1*, volume 1573 of *Lecture Notes in Mathematics*, pages 266-271. Springer-Verlag, Berlin-Heidelberg-New York, 1994.