数理解析研究所講究録
979 巻 1997 年 22-28

Characterizations of operator convex functions of several variables

Frank Hansen

1 Introduction

Let $f: I_1\times\cdots\times I_k\rightarrow\mathbf{R}$ be a real function of k variables defined on the product of k intervals, and let $x=(x_1,\ldots,x_k)$ be a tuple of selfadjoint matrices of order $n_1,\ldots,n_k$ such that the eigenvalues of x_i are contained in I_i for each $i=1,\ldots,k$. We say that such a tuple is in the domain of f and define $f(x)=f(x_1,\ldots,x_k)$ to be the matrix of order $n_1\cdots n_k$ constructed in the following way. For each $i=1,\ldots,k$ we consider the possibly degenerate spectral resolution

$$x_i=\sum_{m_i=1}^{n_i}\lambda_{m_i}(i)e_{m_im_i}^{i}$$

where $\{e_{s_iu_i}\}_{s_iu_i=1}^{n_i}$ is the corresponding system of matrix units and let the formula

$$f(x_1,\ldots,x_n)=\sum_{m_1=1}^{n_1}\cdots\sum_{m_k=1}^{n_k}f(\lambda_{m_1}(1),\ldots,\lambda_{m_k}(k))e_{m_1m_1}^{1}\otimes\cdots\otimes e_{m_km_k}^{k}$$

define the functional calculus. If f can be written as a product of k functions $f=f_1\cdots f_k$ where f_i is a function only of the i'th coordinate, then $f(x_1,\ldots,x_k)=f_1(x_1)\otimes\cdots\otimes f_k(x_k)$. The given definition is readily extended to bounded normal operators on a Hilbert space, cf. [7].

The above function f of k real variables is said to be matrix convex of order $(n_1,\ldots,n_k)$, if

$$(*)\qquad f(\lambda x_1+(1-\lambda)y_1,\ldots,\lambda x_k+(1-\lambda)y_k)\leq\lambda f(x_1,\ldots,x_k)+(1-\lambda)f(y_1,\ldots,y_k)$$

for every $\lambda\in[0,1]$ and all tuples of selfadjoint matrices $(x_1,\ldots,x_k)$ and $(y_1,\ldots,y_k)$ such that the orders of x_i and y_i are n_i and their eigenvalues are contained in I_i for $i=1,\ldots,k$. The definition is meaningful since also the spectrum of $\lambda x_i+(1-\lambda)y_i$ is contained in the interval I_i for each $i=1,\ldots,k$. It is clear that the pointwise limit of a sequence of matrix convex functions of order $(n_1,\ldots,n_k)$ is again matrix convex of order $(n_1,\ldots,n_k)$. If f is matrix convex of order $(n_1,\ldots,n_k)$, then it is also matrix convex of any order $(n_1',\ldots,n_k')$ such that $n_i'\leq n_i$ for $i=1,\ldots,k$. If f is matrix convex of all orders, then we say that f is operator convex. If $I_1,\ldots,I_k$ are open intervals, then it is enough to assume that f is mid-point matrix convex of arbitrary order. This follows because such a function is real analytic and hence continuous, cf. the discussion in the introduction of [4].

2 Jensen's operator inequality

The following theorem for functions of one variable were proved in [6].

Theorem 2.1 *If f is a continuous, real function on the half-open interval $[0, \alpha[$ (with $\alpha \leq \infty$), the following conditions are equiavalent:*

(1) f is operator convex and $f(0) \leq 0$.

(2) $f(a^*xa) \leq a^*f(x)a$ for all a with $\|a\| \leq 1$ and every self-adjoint x with spectrum in $[0, \alpha[$.

(3) $f(pxp) \leq pf(x)p$ for every projection p and every selfadjoint x with spectrum in $[0, \alpha[$.

Aujla [1] extended the previous result in 1993 and essentially proved the following theorem:

Theorem 2.2 *If f is a real continuous function of two variables defined on the domaine $[0, \alpha[\times[0, \beta[$ (with $\alpha, \beta \leq \infty$), the following conditions are equivalent:*

(1) f is separately operator convex, and $f(t, 0) \leq 0$ and $f(0, s) \leq 0$ for all $(t, s) \in [0, \alpha[\times[0, \beta[$.

(2) $f(a^*xa, a^*ya) \leq (a^* \otimes a)f(x, y)(a \otimes a)$ for all a with $\|a\| \leq 1$ and all selfadjoint x, y with spectra contained in $[0, \alpha[$ and $[0, \beta[$ respectively.

(3) $f(pxp, pyp) \leq (p \otimes p)f(x, y)(p \otimes p)$ for every projection p and all selfadjoint x, y with spectra contained in $[0, \alpha[$ and $[0, \beta[$ respectively.

The above operator inequality is equivalent to

$$f(a^*xa, b^*yb) \leq (a^* \otimes b^*)f(x, y)(a \otimes b)$$

for arbitrary contractions a and b, but this generalization is not essential. The class of separately operator convex functions is evidently not of much importance, but Aujla's result paved the road for further progress. The next result [4] followed in 1996.

Theorem 2.3 *If f is a real continuous function of two variables defined on the domaine $[0, \alpha[\times[0, \beta[$ (with $\alpha, \beta \leq \infty$), the following conditions are equivalent:*

(1) f is operator convex, and $f(t, 0) \leq 0$ and $f(0, s) \leq 0$ for all $(t, s) \in [0, \alpha[\times[0, \beta[$.

(2) The operator inequality

$$\begin{pmatrix} f(a^*xa, a^*ya) & 0 \\ 0 & f(b^*xb, b^*yb) \end{pmatrix} \leq \begin{pmatrix} (a \otimes a)^*f(x, y)(a \otimes a) & (a \otimes a)^*f(x, y)(b \otimes b) \\ (b \otimes b)^*f(x, y)(a \otimes a) & (b \otimes b)^*f(x, y)(b \otimes b) \end{pmatrix}$$

is valid for all selfadjoint operators x and y with spectra in $[0, \alpha[$ and $[0, \beta[$ respectively, and all pairs of operators (a, b) such that $aa^* + bb^* = 1$ and b is normal.

(3) The operator inequality

$$\begin{pmatrix} f(pxp,pyp) & 0 \\ 0 & f((1-p)x(1-p),(1-p)y(1-p)) \end{pmatrix}$$

$$\leq \begin{pmatrix} (p\otimes p)f(x,y)(p\otimes p) & (p\otimes p)f(x,y)((1-p)\otimes(1-p)) \\ ((1-p)\otimes(1-p))f(x,y)(p\otimes p) & ((1-p)\otimes(1-p))f(x,y)((1-p)\otimes(1-p)) \end{pmatrix}$$

is valid for all selfadjoint operators x and y with spectra in $[0,\alpha[$ and $[0,\beta[$ respectively, and every ortogonal projection p.

The characterization of operator convexity by a suitable generalization of Jensen's operator inequality has recently been extended to functions of several variables by H. Araki and the author.

3 Generalized Hessian matrices

The notion of partial divided differences plays an important role in differential analysis of matrix and operator convexity. The first divided difference of a differentible function of one variable goes back to Newton. It is defined as

$$[\lambda\mu] = \begin{cases} \dfrac{f(\lambda)-f(\mu)}{\lambda-\mu} & \text{for } \lambda \neq \mu \\ f'(\lambda) & \text{for } \lambda = \mu \end{cases}$$

and it is a symmetric function of the two arguments. If f is twice differentiable, then the second divided difference $[\lambda\mu\zeta]$ is defined as

$$[\lambda\mu\zeta] = \begin{cases} \dfrac{[\lambda\mu]-[\mu\zeta]}{\lambda-\zeta} & \text{for } \lambda \neq \zeta \\ \dfrac{\partial}{\partial\lambda}[\lambda\mu] & \text{for } \lambda = \zeta \end{cases}$$

and it is a symmetric function of the three arguments, cf. [2] for a more systematic introduction to divided differences for functions of one variable.

If f is a real function defined on the product $I_1 \times I_2$ of two open intervals with continuous partial derivatives up to the second order, then we can consider the divided differences $[\lambda\mu|\,\xi]$ and $[\lambda\mu\zeta|\,\xi]$ which are just the previously defined divided differences for the function of one variable obtained by fixing the second variable to ξ. We define the divided differences $[\xi|\lambda\mu]$ and $[\xi|\lambda\mu\zeta]$ similarly. There are, however, also mixed second derivatives defined as

$$[\lambda\mu|\zeta\xi] = \begin{cases} \dfrac{[\lambda|\zeta\xi]-[\mu|\zeta\xi]}{\lambda-\mu} & \text{for } \lambda \neq \mu \\ \dfrac{\partial}{\partial\lambda}[\lambda|\zeta\xi] & \text{for } \lambda = \mu. \end{cases}$$

We could have defined the mixed derivatives by dividing to the right instead of dividing to the left, but this gives the same result. Finally, if f is a real function defined on the product $I_1 \times \cdots \times I_k$ of k open intervals with continuous partial derivatives up to the second order, then we consider the second divided differences that appear by fixing all but one or two of the k coordinates of f. They are labeled as

$$[\lambda_1|\cdots|\mu_1\mu_2\mu_3|\cdots|\lambda_k]^i$$

where the superscript i indicates that the partial divided difference of the second order is taken at the i'th coordinate and all other coordinates are fixed at the values $\lambda_1,\ldots,\lambda_{i-1}$ and $\lambda_{i+1}\ldots,\lambda_k$ or as

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

where the superscripts ij indicate that the mixed partial divided difference of the second order is taken at the distinctly different coordinates i and j and all other coordinates are fixed at the values $\lambda_1,\ldots,\lambda_{i-1}$, $\lambda_{i+1},\ldots,\lambda_{j-1}$ and $\lambda_{j+1},\ldots,\lambda_k$. The notation does not imply any particular order of the coordinates which can be chosen from the full range $1,\ldots,k$. The following definition were introduced in [5]:

Definition 3.1 *Let $f : I_1 \times \cdots \times I_k \to \mathbf{R}$ be a real function of k variables defined on the product of k open intervals with continuous partial derivatives up to the second order. We define a data set Λ of order $(n_1,\ldots,n_k)$ for f to be an element $\Lambda \in I_1^{n_1} \times \cdots \times I_k^{n_k}$, and we usually write it in the form*

$$(*) \qquad \Lambda = \{\lambda_{m_i}(i)\}_{m_i=1,\ldots,n_i} \qquad i=1,\ldots,k.$$

To a given data set Λ we associate so-called generalized Hessian matrices. First we define to each tuple of natural numbers $(m_1,\ldots,m_k) \le (n_1,\ldots,n_k)$ and to any $s,u=1,\ldots,k$ a matrix denoted $H_{su}(m_1,\ldots,m_k)$ of order $n_u \times n_s$ in the following way:

1. *If $s \ne u$, then we set*

$$H_{su}(m_1,\ldots,m_k) =$$

$$\Big([\lambda_{m_1}(1)|\cdots|\lambda_{m_s}(s)\lambda_j(s)|\cdots|\lambda_p(u)\lambda_{m_u}(u)|\cdots|\lambda_{m_k}(k)]^{su}\Big)_{p=1,\ldots,n_u;\,j=1,\ldots,n_s}$$

2. *If $s=u$, then we set*

$$H_{ss}(m_1,\ldots,m_k) = 2\Big([\lambda_{m_1}(1)|\cdots|\lambda_{m_s}(s)\lambda_p(s)\lambda_j(s)|\cdots|\lambda_{m_k}(k)]^{s}\Big)_{p,j=1,\ldots,n_s}.$$

We then define the generalized Hessian matrix as the block matrix

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

which is quadratic and symmetric and of order $n_1+\cdots+n_k$.

If $n_i=1$ for $i=1,\ldots,k$ then the data set $(*)$ reduces to k numbers $\lambda(1),\ldots,\lambda(k)$ and there is only one (generalized) Hessian matrix H. The submatrix H_{su} is a 1×1 matrix with the partial derivative $f''_{su}(\lambda(1),\ldots,\lambda(k))$ as matrix element for $s,u=1,\ldots,k$. Therefore H can be identified with the usual Hessian matrix associated with a function of k variables.

4 Differential characterization of matrix convexity

The functional calculus $(x_1,\dots,x_k) \to f(x_1,\dots,x_k)$ for functions of several variables defines a mapping from (a subset of) the direct sum $B(H_1) \oplus \cdots \oplus B(H_k)$ to the tensor product $B(H_1) \otimes \cdots \otimes B(H_k)$. The mapping is twice Fréchet differentiable, if f has continuous partial derivatives of order $p > 2 + k/2$, cf. [5, Corollary 2.12]. For $k \leq 2$ there are sharper results by A.L. Brown and H.L. Vasudeva, and it may well be that $p = 2$ is a both necessary and sufficient condition for general k. The following result is of a classical nature and can be derived from [3].

Theorem 4.1 *Let the Hilbert spaces $H_1,\dots,H_k$ have finite dimensions $n_1,\dots,n_k$. If the functional calculus mapping is twice Fréchet differentiable, then f is matrix convex of order $(n_1,\dots,n_k)$ if and only if*

$$d^2 f(x_1,\dots,x_k)(h,h) \geq 0$$

for any tuple $h = (h^1,\dots,h^k)$ of selfadjoint matrices on $H_1,\dots,H_k$.

The above result is of great import in conjunction with the following structure theorem for the second Fréchet differential, cf. [5].

Theorem 4.2 *Let $f \in C^p(I_1 \times \cdots \times I_k)$ with $p > 2 + k/2$ where $I_1,\dots,I_k$ are open intervals and let $x = (x_1,\dots,x_k)$ be selfadjoint matrices of orders $(n_1,\dots,n_k)$ in the domain of f. The expectation value of the second Fréchet differential in a vector $\varphi \in H_1 \otimes \cdots \otimes H_k$ is given by*

$$(d^2 f(x)(h,h)\varphi \mid \varphi) = \sum_{m_1=1}^{n_1} \cdots \sum_{m_k=1}^{n_k} \Big(H(m_1,\dots,m_k)\Phi^h(m_1,\dots,m_k) \mid \Phi^h(m_1,\dots,m_k)\Big)$$

where $H(m_1,\dots,m_k)$ are the generalized Hessian matrices associated with f and the eigenvalues of $(x_1,\dots,x_k)$, while the vectors

$$\Phi^h(m_1,\dots,m_k) = \begin{pmatrix} \Phi^h_1(m_1,\dots,m_k) \\ \vdots \\ \Phi^h_k(m_1,\dots,m_k) \end{pmatrix} \qquad m_i = 1,\dots,n_i \quad \textit{for} \quad i = 1,\dots,k$$

are given by

$$\Phi^h_s(m_1,\dots,m_k)_{j_s} = h^s_{m_s j_s}\varphi(m_1,\dots,m_{s-1},j_s,m_{s+1},\dots,m_k)$$

for $j_s = 1,\dots,n_s$ and $s = 1,\dots,k$.

We immediately realize that even without calculating the vectors $\Phi^h(m_1,\dots,m_k)$, one can conclude that f is matrix convex of order $(n_1,\dots,n_k)$, provided 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. This can for example be done for the functions

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

where $\mu_1, \ldots, \mu_k \in [-1, 1]$. It is calculated in [5] that the generalized Hessian matrices for these functions are of the form

$$H(m_1, \ldots, m_k) = f(\lambda_{m_1}(1), \ldots, \lambda_{m_k}(k)) \begin{pmatrix} 2a(1)^t \cdot a(1) & a(1)^t \cdot a(2) & \cdots & a(1)^t \cdot a(k) \\ a(2)^t \cdot a(1) & 2a(2)^t \cdot a(2) & \cdots & a(2)^t \cdot a(k) \\ \vdots & \vdots & \ddots & \vdots \\ a(k)^t \cdot a(1) & a(k)^t \cdot a(2) & \cdots & 2a(k)^t \cdot a(k) \end{pmatrix}$$

where the vectors

$$a(i) = \mu_i\big(f_i(\lambda_1(i)), \ldots, f_i(\lambda_{n_i}(i))\big) \in \mathbf{R}^{n_i}$$

for $i = 1, \ldots, k$. The generalized Hessian matrices are bounded from below by

$$\begin{aligned} & f(\lambda_{m_1}(1), \ldots, \lambda_{m_k}(k)) \begin{pmatrix} a(1)^t \cdot a(1) & \cdots & a(1)^t \cdot a(k) \\ \vdots & \ddots & \vdots \\ a(k)^t \cdot a(1) & \cdots & a(k)^t \cdot a(k) \end{pmatrix} \\ = \; & f(\lambda_{m_1}(1), \ldots, \lambda_{m_k}(k)) \begin{pmatrix} a(1) & \cdots & a(k) \end{pmatrix}^t \begin{pmatrix} a(1) & \cdots & a(k) \end{pmatrix} \end{aligned}$$

which are positive semi-definite matrices.

Corollary 4.3 *Let ν be a non-negative Borel measure on the cube $[-1, 1]^k$ for $k \in \mathbf{N}$ and let $a_0, a_1, \ldots, a_k$ be real numbers. The function*

$$f(t_1, \ldots, t_k) = a_0 + a_1 t_1 + \cdots + a_k t_k + \int_{-1}^{1} \cdots \int_{-1}^{1} \prod_{i=1}^{k} \frac{1}{1 - \mu_i t_i} \, d\nu(\mu_1, \ldots, \mu_k)$$

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

References

[1] J.S. Aujla. Matrix convexity of functions of two variables. *Linear Algebra and Its Applications*, 194:149–160, 1993.

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

REFERENCES

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

[4] F. Hansen. Jensen's operator inequality for functions of two variables. *to appear in Proc. Amer. Math. Soc*, 1996.

[5] F. Hansen. Operator convex functions of several variables. *RIMS-1119*, 1996.

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

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