Means of positive matrices: Geometry and a conjecture*

Dénes Petz

Alfréd Rényi Institute of Mathematics
Hungarian Academy of Sciences
H–1053 Budapest, Reáltanoda u. 13–15, Hungary
e-mail: petz@math.bme.hu

Abstract

Means of positive numbers are well-know but the theory of matrix means due to Kubo and Ando is less known. The lecture gives a short introduction to means, the emphasis is on matrices. It is shown that any two-variable-mean of matrices can be extended to more variables. The n-variable-mean $\mathbf{M}_n(A_1, A_2, \ldots, A_n)$ is defined by a symmetrization procedure when the n-tuple $(A_1, A_2, \ldots, A_n)$ is ordered, it is continuous and monotone in each variable. The geometric mean of matrices has a nice interpretation in terms of an information geometry and the ordering of the n-tuple is not necessary for the definition. It is conjectured that this strong condition might be weakened for some other means, too.

Key Words: operator means, information geometry, logarithmic mean, geometric mean, positive matrices.

AMS Classification Number: 47A64 (15A48, 47A63)

1. Introduction

The **geometric mean** $\sqrt{xy}$, the **arithmetic mean** $(x+y)/2$ and the inequality between them go back to the ancient Greeks. That time the means of the positive numbers x and y were treated as geometric proportions.

The arithmetic mean can be extended to more variables as

$$\mathbf{A}(x_1, x_2, \ldots, x_n) := \frac{1}{n}(x_1 + x_2 + \cdots + x_n)$$

and the formula makes sense not only for positive numbers but in any vectorspace. The quantity appears under various names depending on the context of the application, for example, average, barycenter, or the center of mass.

Suppose we have a device which can compute the mean of two variables. How to compute the mean of three? Assume that we aim to obtain the mean of x, y and z. We can make a new device

$$W : (a, b, c) \mapsto (\mathbf{A}(a, b), \mathbf{A}(a, c), \mathbf{A}(b, c)) \tag{1.1}$$

which applied to (x, y, z) many times gives the mean of x, y and z. More mathematically,

$$W^n(x, y, z) \to \mathbf{A}(x, y, z) \quad \text{as} \quad n \to \infty. \tag{1.2}$$

To show the relation (1.2), we have various possibilities. The simplest might be to observe that $W^n(x, y, z)$ is a convex combination of x, y and z,

$$W^n(x, y, z) = \lambda_1^{(n)} x + \lambda_2^{(n)} y + \lambda_3^{(n)} z.$$

One can compute the coefficients $\lambda_i^{(n)}$ explicitly and show that $\lambda_i^{(n)} \to 1/3$.

Another possibility is to compute the eigenvalues of the linear transformation W. It turns out that 1 is the only eigenvalue and the only peripheral eigenvalue, so W^n converges to the corresponding eigenprojection according to ergodic theory. In other words,

$$W^n(x, y, z) \to \frac{1}{3}(x + y + z).$$

Assume that x, y and z are linearly independent vectors in a vectorspace. Their convex hull is a triangle Δ_0. Let the convex hull of the three vectors $W^n(x, y, z)$ be the triangle Δ_n. Since

$$\bigcap_{n=0}^{\infty} \Delta_n = \left\{ \frac{x + y + z}{3} \right\}$$

we can visualize the convergence (1.2) and we can observe its exponential speed, since the diameter of Δ_{n+1} is the half of that of Δ_n.
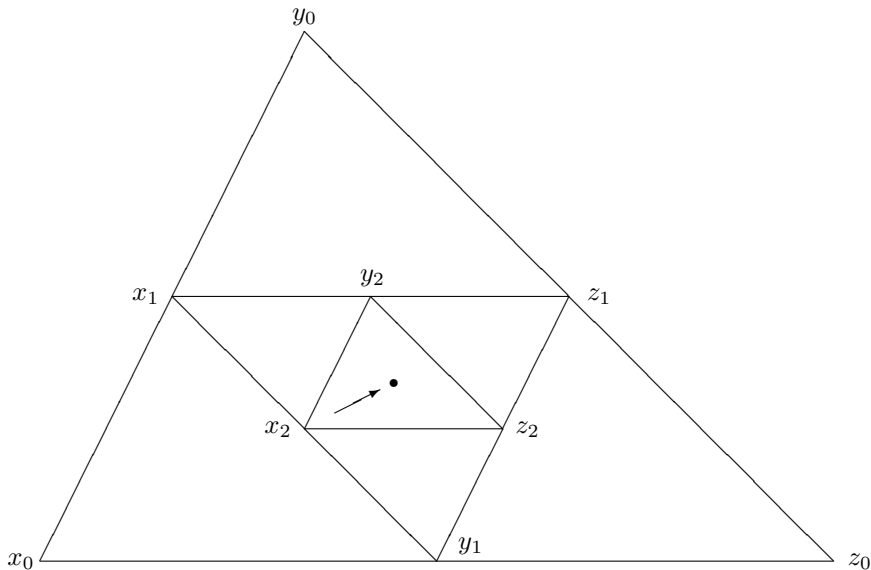
Of course, the above approach to the three-variable arithmetic mean is a possibility. If $x_1, x_2, \ldots, x_n$ are numbers, then one can minimize the functional

$$z \mapsto \sum_{i=1}^{n} (x_i - z)^2. \tag{1.3}$$

The minimizer is

$$z = \mathbf{A}(x_1, x_2, \ldots, x_n).$$

The aim of this lecture is to study the symmetrization procedure (1.1) not only for the arithmetic mean but for several other means and not for numbers but mostly for matrices.

Figure 1: The triangles Δ_0, Δ_1 and Δ_2.

2. Means for three variables

The geometric mean, the arithmetic mean and the harmonic mean are extended to three (and more) variables as

$$\sqrt{xyz}, \quad (x+y+z)/3, \quad 3/(x^{-1}+y^{-1}+z^{-1}).$$

Our target is to extend an arbitrary mean $M_2(x,y)$ of two variables to three variables. In order to do this we need to specify what we mean by a mean. A function $\mathbf{M} : \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}^+$ may be called a *mean* of positive numbers if

  (i) $\mathbf{M}(x,x) = x$ for every $x \in \mathbb{R}^+$.

  (ii) $\mathbf{M}(x,y) = \mathbf{M}(y,x)$ for every $x,y \in \mathbb{R}^+$.

  (iii) If $x < y$, then $x < \mathbf{M}(x,y) < y$.

  (iv) If $x < x'$ and $y < y'$, then $\mathbf{M}(x,y) < \mathbf{M}(x',y')$.

  (v) $\mathbf{M}(x,y)$ is continuous.

  (vi) $\mathbf{M}(tx,ty) = t\mathbf{M}(x,y)$ $(t,x,y \in \mathbb{R}^+)$.

A two-variable function $\mathbf{M}(x,y)$ satisfying condition (vi) can be reduced to a one-variable function $f(x) := \mathbf{M}(1,x)$. Namely, $\mathbf{M}(x,y)$ is recovered from f as

$$\mathbf{M}(x,y) = xf\left(\frac{y}{x}\right). \tag{2.1}$$

What are the properties of f which imply conditions (i)–(v)? They are as follows.

(i)′ $f(1) = 1$

(ii)′ $tf(t^{-1}) = f(t)$

(iii)′ $f(t) > 1$ if $t > 1$ and $f(t) < 1$ if $0 < t < 1$.

(iv)′ f is monotone increasing.

(v)′ f is continuous.

Then a homogeneous and continuous mean is uniquely described by a function f satisfying the properties (i)′–(v)′.

Assume that $0 < x \leqslant y \leqslant z$ and define a recursion

$$x_1 = x, \qquad y_1 = y, \qquad z_1 = z \tag{2.2}$$

$$x_{n+1} = M_2(x_n, y_n), \qquad y_{n+1} = M_2(x_n, z_n), \qquad z_{n+1} = M_2(y_n, z_n). \tag{2.3}$$

Below we refer to this recursion as **symmetrization procedure**.

One can show by induction that $x_n \leqslant y_n \leqslant z_n$, moreover the sequence (x_n) is increasing and (z_n) is decreasing. Therefore, the limits

$$L := \lim_{n \to \infty} x_n \qquad \text{and} \qquad U = \lim_{n \to \infty} z_n \tag{2.4}$$

exist. It is not difficult to show that $L = U$ must hold [12].

Given a mean $\mathbf{M}_2(x, y)$ of two variables, we define $\mathbf{M}_3(x, y, z)$ as the limit $\lim_n x_n = \lim_n y_n = \lim_n z_n$ in the above recursion (2.2) and (2.3) for any $x, y, z \in \mathbb{R}^+$. Note that the existence of the limit does not require the symmetry of the $\mathbf{M}_2(x, y)$.

Example 2.1. Let
$$\mathbf{M}_f(x, y) = f^{-1} \left(\frac{f(x) + f(y)}{2} \right)$$

be a **quasi-arithmetic mean** defined by a strictly monotone function f [5]. For these means
$$\mathbf{M}_3(x, y, z) = f^{-1} \left(\frac{f(x) + f(y) + f(z)}{3} \right).$$

Note that arithmetic, geometric and harmonic means belong to this class.   □

When our paper [12] was written, we were not aware of the paper [6], in which the following definition was given. Assume that m is a mean of two variables. A mean $\mathbf{M}$ of three variables is said to be of type 1 invariant mean with respect to m if

$$\mathbf{M}(m(a, c), m(a, b), m(b, c)) = \mathbf{M}(a, b, c).$$

It is obtained in [6] that to each m there exists a unique $\mathbf{M}$ which is type 1 invariant with respect to m. The proof is exactly the above symmetrization procedure.

A theory of means of positive operators was developed by Kubo and Ando [7]. The key point of the theory that for positive matrices A and B formula (2.1) should be modified as

$$\mathbf{M}(A, B) = A^{1/2} f\big(A^{-1/2} B A^{-1/2}\big) A^{1/2} \tag{2.5}$$

and the function f is required to be operator monotone. This property implies that f' is decreasing and $f'(1) = 1/2$. Since the matrix case is in the center of our interest, we assume these properties below.

Assume that $x < y$.

$$\frac{\mathbf{M}(x, y) - x}{y - x} = \frac{x(f(y/x) - 1)}{y - x} = \frac{x(y/x - 1)}{y - x} f'(t) = f''(t),$$

where $1 < t < y/x$. When f' is decreasing, we have

$$f'(y/x) \le \frac{\mathbf{M}(x, y) - x}{y - x} \le f'(1). \tag{2.6}$$

Lemma 2.2. *Let $0 < x \leqslant y \leqslant z$ and assume that f' is decreasing and $f'(1) = 1/2$. Then*

$$\frac{\mathbf{M}(y, z) - \mathbf{M}(x, y)}{z - x} \le 1 - f'(y/x).$$

Proof. From (2.6) we have

$$\frac{y - \mathbf{M}(x, y)}{y - x} \le 1 - f'(y/x) \qquad \text{and} \qquad \frac{\mathbf{M}(y, z) - y}{z - y} \le f'(1),$$

moreover $\max\big(f'(1), 1 - f'(y/x)\big) = 1 - f'(y/x)$. $\qquad\qquad\square$

It is an important consequence of the lemma that the limit (2.4) is exponential. Namely,

$$z_n - x_n \le (z - x)\big(1 - f'(z/x)\big)^n \tag{2.7}$$

holds. When n is large, then z_n/x_n is close to 1 and $(z_{n+1} - x_{n+1})/(z_n - x_n)$ has an upper bound close to $1/2$.

A sort of mean of three positive matrices can be obtained by a symmetrization procedure from the two-variable-means, at least under some restriction.

Theorem 2.3. *Let $A, B, C \in \mathbf{M}_n(\mathbb{C})$ be positive definite matrices and let $\mathbf{M}_2$ be an operator mean. Assume that $A \leqslant B \leqslant C$. Set a recursion as*

$$A_1 = A, \qquad B_1 = B, \qquad C_1 = C, \tag{2.8}$$

$$A_{n+1} = \mathbf{M}_2(A_n, B_n), \qquad B_{n+1} = \mathbf{M}_2(A_n, C_n), \qquad C_{n+1} = \mathbf{M}_2(B_n, C_n). \tag{2.9}$$

Then the limits

$$\mathbf{M}_3(A, B, C) := \lim_n A_n = \lim_n B_n = \lim_n C_n \tag{2.10}$$

exist.

Proof. By the monotonicity of $\mathbf{M}_2$ and mathematical induction, we see that $A_n \leqslant B_n \leqslant C_n$. It follows that the sequence (A_n) is increasing and (C_n) is decreasing. Therefore, the limits

$$L := \lim_{n \to \infty} A_n \qquad \text{and} \qquad U = \lim_{n \to \infty} C_n$$

exist. We claim that $L = U$.

Assume that $L \neq U$. By continuity, $B_n \to \mathbf{M}_2(L, U) =: M$, where $L < M < U$. Since

$$\mathbf{M}_2(B_n, C_n) = C_{n+1},$$

the limit $n \to \infty$ gives $\mathbf{M}_2(M, U) = U$, which contradicts $M < U$. $\qquad \square$

By the symmetrization procedure any operator mean of two variables has an extension to those triplets (A, B, C) which can be ordered. In a few cases, the latter restriction can be skipped. The arithmetic and the harmonic means belong to this class.

The three-variable matrix means defined by symmetrization are continuous and monotone in each of the variables. This facts follow straightforwardly from the construction.

Example 2.4. The **logarithmic mean** of the positive numbers x and y is

$$\frac{x - y}{\log x - \log y} . \tag{2.11}$$

The corresponding function

$$f(x) = \frac{x - 1}{\log x}$$

is operator monotone and so the mean makes sense for positive matrices as well. If A and B are positive matrices, then

$$\mathbf{G}(A, B) \leq \mathbf{L}(A, B) \leq \mathbf{A}(A, B)$$

holds for the geometric, logarithmic and arithmetic means. When $A \leq B \leq C$ are positive matrices, then the symmetrization procedure defines the three-variable means $\mathbf{G}_3, \mathbf{L}_3$ and $\mathbf{A}_3$. (Of course, $\mathbf{A}_3$ is clear without symmetrization.) It follows from the procedure, that

$$\mathbf{G}_3(A, B, C) \leq \mathbf{L}_3(A, B, C) \leq \mathbf{A}_3(A, B, C).$$

Note that the paper aimed to discuss the three-variable mean $\mathbf{L}_3$ for numbers, but only some inequalities were obtained.

3. Means of matrices and information geometry

An important non-trivial example of operator means is the **geometric mean**:

$$A\#B = A^{1/2}\big(A^{-1/2}BA^{-1/2}\big)^{1/2}A^{1/2} \tag{3.1}$$

which has the special property

$$(\lambda A)\#(\mu B) = \sqrt{\lambda\mu}(A\#B) \tag{3.2}$$

for positive numbers λ and μ. The geometric mean was found before the general theory of matrix means, see [13], in a very different context.

Theorem 3.1. *Let $A, B, C \in M_n(\mathbb{C})$ be positive definite matrices. Set a recursion as*

$$A_1 = A, \qquad B_1 = B, \qquad C_1 = C, \tag{3.3}$$

$$A_{n+1} = A_n\#B_n, \qquad B_{n+1} = A_n\#C_n, \qquad C_{n+1} = B_n\#C_n. \tag{3.4}$$

Then the limit

$$G(A, B, C) := \lim_n A_n = \lim_n B_n = \lim_n C_n \tag{3.5}$$

exists.

Proof. Choose positive numbers λ and μ such that

$$A' := A < B' := \lambda B < C' := \mu C.$$

Start the recursion with these matrices. By Theorem 2.3 the limits

$$G(A', B', C') := \lim_n A'_n = \lim_n B'_n = \lim_n C'_n$$

exist. For the numbers

$$a := 1, \qquad b := \lambda \qquad \text{and} \qquad c := \mu$$

the recursion provides a convergent sequence (a_n, b_n, c_n) of triplets.

$$(\lambda\mu)^{1/3} = \lim_n a_n = \lim_n b_n = \lim_n c_n.$$

Since

$$A_n = A'_n/a_n, \qquad B_n = B'_n/b_n \qquad \text{and} \qquad C_n = C'_n/c_n$$

due to property (3.2) of the geometric mean, the limits stated in the theorem must exist and equal $G(A', B', C')/(\lambda\mu)^{1/3}$. $\qquad\square$

The result of Theorem 3.1 was obtained in [2] but our proof is different and completely elementary. A different approach is based on Riemannian geometry [3, 9].

The positive definite matrices might be considered as the variance of multivariate normal distributions and the information geometry of Gaussians yields a natural Riemannian metric. The simplest way to construct an information geometry is to start with an **information potential function** and to introduce the Riemannian metric by the Hessian of the potential. We want a geometry on the family of non-degenerate multivariate Gaussian distributions with zero mean vector. Those distributions are given by a positive definite real matrix A in the form

$$f_A(x) := \frac{1}{\sqrt{(2\pi)^n \det A}} \exp\left(-\langle A^{-1}x, x\rangle/2\right) \qquad (x \in \mathbb{R}^n). \qquad (3.6)$$

We identify the Gaussian (3.6) with the matrix A, and we can say that the Riemannian geometry is constructed on the space of positive definite real matrices. There are many reasons (originated from statistical mechanics, information theory and mathematical statistics) that the **Boltzmann entropy**

$$S(f_A) := C + \frac{1}{2} \log \det A \qquad (C \text{ is a constant}) \qquad (3.7)$$

is a candidate for being an information potential.

The $n \times n$ real symmetric matrices can be identified with the Euclidean space of dimension $n(n+1)/2$ and the positive definite matrices form an open set. Therefore the set of Gaussians has a simple and natural manifold structure. The tangent space at each foot point is the set of symmetric matrices. The Riemannian metric is defined as

$$g_A(H_1, H_2) := \frac{\partial^2}{\partial s \partial t} S(f_{A+tH_1+sH_2})\Big|_{t=s=0}, \qquad (3.8)$$

where H_1 and H_2 are tangents at A. The differentiation easily gives

$$g_A(H_1, H_2) = \operatorname{Tr} A^{-1} H_1 A^{-1} H_2. \qquad (3.9)$$

The corresponding information geometry of the Gaussians was discussed in [10] in details. We note here that this geometry has many symmetries, each similarity transformation of the matrices becomes a symmetry. In the statistical model of multivariate distributions (3.9) plays the role of the **Fisher-Rao metric**.

(3.9) determines a Riemannian metric on the set $\mathcal{P}$ of all positive definite complex matrices as well and below we prefer to consider the complex case. The geodesic connecting $A, B \in \mathcal{P}$ is

$$\gamma(t) = A^{1/2}(A^{-1/2}BA^{-1/2})^t A^{1/2} \qquad (0 \le t \le 1)$$

and we observe that the midpoint $\gamma(1/2)$ is just the geometric mean $A\#B$. The geodesic distance is

$$\delta(A, B) = \|\log(A^{-1/2}BA^{-1/2})\|_2,$$

where $\|\cdot\|_2$ stands for the Hilbert–Schmidt norm. (It was computed in [1] that the scalar curvature of the space $\mathcal{P}$ is constant.) These observations show that the information Riemannian geometry is adequate to treat the geometric mean of positive definite matrices [3, 9].

Let A, B and C be positive definite matrices. The mean $C' := A\#B$ is the middle point of the geodesic connecting A with B, $B' := C\#A$ and $C' := A\#B$ have similar geometric description. Since

$$\delta(A\#C, B\#C) \leq \tfrac{1}{2}\delta(A, B) \tag{3.10}$$

(see Prop. 6 in [3]), the diameter of the triangle $A'B'C'$ is at most the half of the diameter of ABC.

When A_n, B_n, C_n are defined by the symmetrization procedure, the sequences $(A_n), (B_n)$ and (C_n) form Cauchy sequences with respect to the geodesic distance δ. The space is complete with respect to this metric and the three sequences have a common limit point.
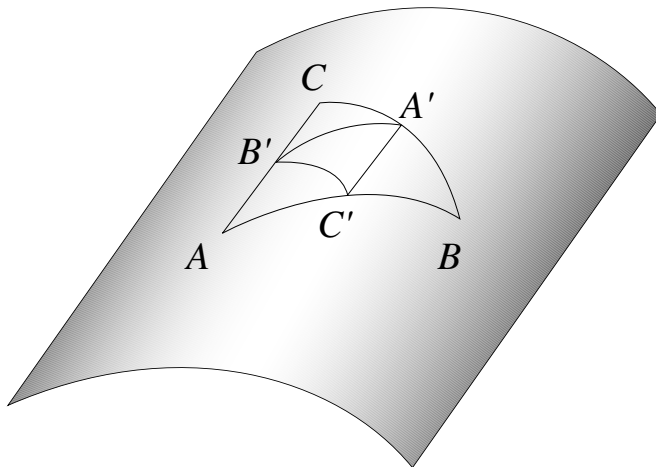

Figure 2: Geometric view of the symmetrization

The geometric view of the symmetrization procedure concerning the geometric mean in the Riemannian space of positive definite matrices resembles very much the procedure concerning the arithmetic mean in the flat space.

The arithmetic mean of matrices A_1, A_2 and A_3 is the minimizer of the functional

$$Z \mapsto \|Z - A_1\|^2 + \|Z - A_2\|^2 + \|Z - A_3\|^2,$$

where the norm is the Hilbert–Schmidt norm. Following this example, one may define the geometric mean of the positive matrices A_1, A_2 and A_3 as the minimizer

of the functional

$$Z \mapsto \delta(Z, A_1)^2 + \delta(Z, A_2)^2 + \delta(Z, A_3)^2.$$

This approach is discussed in several papers [8, 9, 3]. The minimizer is unique, the mean is well-defined but it is different from the three-variable geometric mean coming out from the symmetrization procedure.

Note that there are several natural Riemannian structures on the cone of positive definite matrices. When such a matrix is considered as a quantum statistical operator (without the normalization constraint), the information geometries correspond to different Riemannian metrics, see [11].

4. Discussion

It seems that there are more 3-variable-means without the ordering constraint than the known arithmetic, geometric and harmonic means. (Computer simulation has been carried out for some other means as well.)

An operator monotone function f associated with a matrix mean has the property $f(1) = 1$ and $f'(1) = 1/2$. The latter formula follows from (ii)$'$. From the power series expansion around 1, one can deduce that there are some positive numbers ε and δ such that

$$\|A^{1/2}f(A^{-1/2}B_1A^{-1/2})A^{1/2} - A^{1/2}f(A^{-1/2}B_2A^{-1/2})A^{1/2}\| \le (1-\delta)\|B_1 - B_2\|$$

whenever $I - \varepsilon \le A, B_1, B_2 \le I + \varepsilon$. This estimate equivalently means

$$\|\mathbf{M}_f(A, B_1) - \mathbf{M}_f(A, B_2)\| \le (1-\delta)\|B_1 - B_2\|, \qquad (4.1)$$

where $\|\cdot\|$ denotes the operator norm.

If the **diameter** of a triplet (A, B, C) is defined as

$$D(A, B, C) := \max\{\|A - B\|, \|B - C\|, \|A - C\|\},$$

then we have

$$D(A_{n+1}, B_{n+1}, C_{n+1}) \le (1-\delta)D(A_n, B_n, C_n),$$

provided that the condition $I - \varepsilon \le A_1, B_1, C_1 \le I + \varepsilon$ holds. Therefore in a small neighborhood of the identity the symmetrization procedure converges exponentially fast for any triplet of matrices and for any matrix mean. We conjecture that this holds in general.

References

[1] ANDAI, A., Information geometry in quantum mechanics, Ph.D. dissertation (in Hungarian), BUTE, 2004.

[2] ANDO, T., LI, C-K., MATHIAS, R., Geometric means, *Linear Algebra Appl.*, Vol. 385 (2004), 305–334.

[3] BHATIA, R., HOLBROOK, J., Geometry and means, preprint, 2005.

[4] CARLSON, B. C., The logarithmic mean, *Amer. Math. Monthly*, Vol. 79 (1972), 615–618.

[5] DARÓCZY, Z., PÁLES, ZS., Gauss-composition of means and the solution of the Matkowski-Sutő problem, *Publ. Math. Debrecen*, **61**(2002), 157–218.

[6] HORWITZ, A., Invariant means, *J. Math. Anal. Appl.*, Vol. 270 (2002), 499–518.

[7] KUBO, F., ANDO, T., Means of positive linear operators, *Math. Ann.*, Vol. 246 (1980), 205–224.

[8] LIM, Y., Geometric means of symmetric connes, *Arch. Math.*, Vol. 75 (2000), 39–45.

[9] MOAKHER, M., A differential geometric approach to the geometric mean of symmetric positive definite matrices, *SIAM J. Matrix Anal. Appl.*, Vol. 26 (2005), 735–747.

[10] OHARA, A., SUDA, N., AMARI, S., Dualistic differential geometry of positive definite matrices and its applications to related problems, *Linear Algebra Appl.*, Vol. 247 (1996), 31–53.

[11] PETZ, D., Monotone metrics on matrix spaces, *Linear Algebra Appl.*, Vol. 244 (1996), 81–96.

[12] PETZ, D., Temesi, R., Means of positive numbers and matrices, *SIAM J. Matrix Anal. Appl.*, to appear.

[13] PUSZ, W., Woronowicz, S. L., Functional calculus for sesquilinear forms and the purification map, *Rep. Math. Phys.*, Vol. 8 (1975), 159–170.

[14] SKOVGAARD, L. T., A Riemannian geometry of the multivariate normal model, *Scand. J. Statistics*, Vol. 11 (1984), 211–223.

Dénes Petz
Alfréd Rényi Institute of Mathematics
Hungarian Academy of Sciences
H–1053 Budapest, Reáltanoda u. 13–15, Hungary