A NOTE ON EXTREMAL DECOMPOSITIONS OF COVARIANCES

ZOLTÁN LÉKA

ABSTRACT. We shall present an elementary approach to extremal decompositions of (quantum) covariance matrices determined by densities. We give a new proof on former results and provide a sharp estimate of the ranks of the densities that appear in the decomposition theorem.

1. Introduction. Let $D \in M_n(\mathbb{C})$ denote an $n \times n$ (complex) density matrix (i.e., $D \geq 0$ and $\operatorname{Tr} D = 1$), and let X_i , $1 \leq i \leq k$, stand for self-adjoint matrices in $M_n(\mathbb{C})$. Then the non-commutative covariance matrix is defined by

$$Var_D(\mathbf{X})_{ij} := Tr DX_i X_j - (Tr DX_i)(Tr DX_j)$$

$$1 \le i, \quad j \le k,$$

where **X** stands for the tuple (X_1, \ldots, X_k) , see [7, page 13]. We note that there are more general versions of variances and covariance matrices. For instance, in [1, 2] Bhatia and Davis introduced them by means of completely positive maps and applied the concept for improving non-commutative Schwarz inequalities.

Covariances naturally appear in quantum information theory as well and it seems that there is a recent interest in order to understand their extremal properties [8, 9]. More precisely, in [8], Petz and Tóth proved that any density matrix D can be written as the convex combination of projections $\{P_l\}$, i.e.,

$$D = \sum_{l} \lambda_l P_l,$$

²⁰¹⁰ AMS $Mathematics\ subject\ classification.$ Primary 62J10, 81Q10, Secondary 15B48, 15B57.

 $Keywords\ and\ phrases.$ Decomposition, density, covariance, correlation, extreme points.

Received by the editors on October 4, 2013, and in revised form on June 17, 2014

DOI:10.1216/RMJ-2016-46-2-571

such that

$$\operatorname{Var}_D(X) = \sum_l \lambda_l \operatorname{Var}_{P_l}(X)$$

holds, where X denotes a fixed Hermitian. It is worth mentioning here that, quite recently, Yu pointed out some extremal aspects of the variances which yield a descriptions of the quantum Fisher information in terms of variances (for the details, see [11]).

In this short note, we study analogous questions in the multivariable case. Actually, we are interested in the following problem: let us find densities $D_l \in M_n(\mathbb{C})$ such that

$$D = \sum_{l} \lambda_{l} D_{l}$$
 and $\operatorname{Var}_{D}(\mathbf{X}) = \sum_{l} \lambda_{l} \operatorname{Var}_{D_{l}}(\mathbf{X}),$

where

$$\sum_{l} \lambda_l = 1 \quad \text{and} \quad 0 < \lambda_l < 1.$$

Let us call a density D extreme with respect to $\mathbf{X} = (X_1, \dots, X_k)$ if it admits only the trivial decomposition, i.e., $D_l = D$ for every l. It was proved in the cases k = 1 and k = 2 that the extreme densities are rank-one projections [5, 8]. Furthermore, the number of projections used, i.e., the length of the decomposition, is polynomial in rank D (see [5]).

The aim of this note is to present a simple approach to the extremal problem above and to look at the question from the theory of extreme correlation matrices (see [3, 4, 6]). In this context, we shall give a new proof to the decomposition theorems appearing in [5, 8, 9], and we present a sharp rank-estimate of the extreme densities.

2. Results and examples. First, we collect some basic properties of the covariance matrix $\operatorname{Var}_D(\mathbf{X})$. We note that the matrix does not change by (real) scalar perturbations of the tuple (X_1, \ldots, X_k) . In fact, an elementary calculation on the entries gives that

(2.1)
$$\operatorname{Var}_{D}(\mathbf{X}) = \operatorname{Var}_{D}(X_{1} - \lambda_{1}I, \dots, X_{k} - \lambda_{k}I),$$

where $\lambda_i \in \mathbb{R}$ for every i. Moreover, one can readily check that $\operatorname{Var}_D(\mathbf{X})$ is positive. For the sake of completeness, here is a simple proof.

Lemma 2.1. $Var_D(\mathbf{X}) \geq 0$.

Proof. By equation (2.1), without loss of generality, one can assume that $\operatorname{Tr} DX_i = 0$ holds for every $1 \leq i \leq k$. The density D defines a semi-inner product $\langle A, B \rangle_D := \operatorname{Tr} DA^*B$ on $M_n(\mathbb{C})$. Since $\operatorname{Var}_D(\mathbf{X})_{ij} = \langle X_i, X_j \rangle_D$, for any $y = (y_1, \dots, y_k) \in \mathbb{C}^k$, we get that

$$y \operatorname{Var}_D(\mathbf{X}) y^* = \left\langle \sum_i y_i X_i, \sum_i y_i X_i \right\rangle_D \ge 0,$$

and the proof is done.

Next, we show that the covariance is a concave function on the set of the density matrices.

Lemma 2.2. Let

$$D = \sum_{l} \lambda_l D_l$$

be a finite sum of densities $D_l \in M_n(\mathbb{C})$ such that

$$\sum_{l} \lambda_{l} = 1$$

and $0 \le \lambda_l \le 1$. Then

$$\operatorname{Var}_D(\mathbf{X}) \ge \sum_l \lambda_l \operatorname{Var}_{D_l}(\mathbf{X}).$$

Proof. Choose $0 < \lambda < 1$. If $D = \lambda D_1 + (1 - \lambda)D_2$, a straightforward calculation gives that

$$\operatorname{Var}_{D}(\mathbf{X}) - (\lambda \operatorname{Var}_{D_{1}}(\mathbf{X}) + (1 - \lambda) \operatorname{Var}_{D_{2}}(\mathbf{X})) = \lambda (1 - \lambda) [x_{ij}]_{1 \le i, j \le k},$$

where $x_{ij} = \text{Tr}(D_1 - D_2)X_i\text{Tr}(D_1 - D_2)X_j$. Therefore, $[x_{ij}]_{1 \leq i,j \leq k} = XX^* \geq 0$ holds with

$$X = \begin{bmatrix} \operatorname{Tr} (D_1 - D_2) X_1 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ \operatorname{Tr} (D_1 - D_2) X_k & 0 & \cdots & 0 \end{bmatrix} \in M_k(\mathbb{C}),$$

and the lemma readily follows.

The scalar perturbation property $\operatorname{Var}_D(\mathbf{X}) = \operatorname{Var}_D(\mathbf{X} - \lambda)$ guarantees that it is enough to solve the extremal problem when $\operatorname{Tr} DX_i = 0$ comes for every $1 \leq i \leq k$. Then the nonlinear part of the covariance vanishes; thus, we can simply transform our problem into a geometrical one: let $X_i \in M_n(\mathbb{C})$, $1 \leq i \leq k$, be self-adjoints, and define the set

$$\mathcal{D}(\mathbf{X}) := \{D : D \in M_n(\mathbb{C}) \text{ is density and}$$

$$\operatorname{Tr} DX_i = 0 \text{ for every } 1 < i < k\}.$$

Clearly, $\mathcal{D}(\mathbf{X})$ is a convex, compact set. From the Krein-Milman theorem, $\mathcal{D}(\mathbf{X})$ is the convex hull of its extreme points. Precisely, these extreme points are the extreme densities we are looking for in the decomposition of $\operatorname{Var}_{\mathcal{D}}(\mathbf{X})$.

Notice that there is no restriction if we assume that X_1, \ldots, X_k are linearly independent over \mathbb{R} . Hence, from here on, we shall use this assumption on X_i 's.

When $k \geq 3$, one can see that it is no longer true that the extreme points of $\mathcal{D}(\mathbf{X})$ are rank-one projections. In fact, look at the following simple example in $M_2(\mathbb{C})$ with k=3.

Example 2.3. Recall that the Pauli matrices are given by

$$\sigma_x = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \qquad \sigma_y = \begin{bmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{bmatrix}, \qquad \sigma_z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

Any 2×2 Hermitian Z with Tr Z = 1 can be expressed in the form

$$Z = \frac{1}{2}(I_2 + x\sigma_x + y\sigma_y + z\sigma_z),$$

where x, y and $z \in \mathbb{R}$. Then the points of the Bloch sphere, i.e., $x^2 + y^2 + z^2 = 1$, correspond to the rank-one projections. It is standard that the self-adjoints of trace 1, which are orthogonal to a fixed Z, form an affine two-dimensional subspace of \mathbb{R}^3 . Hence, one can find X_1, X_2 and X_3 so that the only density D that satisfies $\operatorname{Tr} DX_i = 0$, $1 \le i \le 3$, is inside the Bloch ball. Then $\mathcal{D}(\mathbf{X}) = \{D\}$ and D is a density of rank 2.

We shall present a simple characterization of extreme densities or the extreme points of $\mathcal{D}(\mathbf{X})$. We recall that, for any positive operators D and C, $D-\varepsilon C$ is positive for some $\varepsilon>0$ if and only if ran $C\leq \operatorname{ran} D$ holds. Then we can prove:

Lemma 2.4. The following statements are equivalent:

- (i) D is an extreme point of $\mathcal{D}(\mathbf{X})$.
- (ii) If $C \in \mathcal{D}(\mathbf{X})$ such that ran $C \leq \operatorname{ran} D$, then C = D.

Proof. Let us assume that ran $C \leq \operatorname{ran} D$ and $D \neq C \in \mathcal{D}(\mathbf{X})$. Then

$$(1 - \varepsilon) \left(\frac{1}{1 - \varepsilon} (D - \varepsilon C) \right) + \varepsilon C = D,$$

where $0 < \varepsilon < 1$; hence, D cannot be an extreme point of $\mathcal{D}(\mathbf{X})$.

Conversely, if D is not extreme, then $D = 1/2D_1 + 1/2D_2$ which implies that ran $(D-1/2D_1) \le \operatorname{ran} D$, since $D-1/2D_1$ is positive. \square

To produce a description of ext $\mathcal{D}(\mathbf{X})$ which is more effective for our purposes, we need some basic facts about correlation matrices. We recall that a positive semidefinite matrix is a correlation matrix if its diagonal entries are ones. Correlation matrices form a convex, compact set in $M_n(\mathbb{C})$. Its extreme points, or extreme correlation matrices, were described by several authors, see e.g., [4, 6]. It is well known that an $n \times n$ extreme correlation matrix has rank at most \sqrt{n} (see, e.g., [3]). Later, we shall present an estimate of the rank of extreme density matrices (with respect to tuples).

The perturbation method used by Li and Tam is relevant for us. Let us say that a nonzero Hermitian $S \in M_n(\mathbb{C})$ is a perturbation of D if there exists an $\varepsilon > 0$ such that $D \pm \varepsilon S$ are density matrices as well. Then D is an extreme density with respect to X_1, \ldots, X_k if and only if there does not exist a perturbation S of D such that $\operatorname{Tr} S = 0$ and $\operatorname{Tr} S X_i = 0$ for every $1 \le i \le k$. In fact, if D is not extreme, one can find D_1 and D_2 densities such that $D = 1/2D_1 + 1/2D_2$ and $\operatorname{Tr} D_j X_i = 0$. It follows that $S = D_1 - D_2$ is a perturbation of D. The converse statement is trivial.

From here on, let $H_n(\mathbb{C})$ denote the real Hilbert space of $n \times n$ complex Hermitian matrices with the usual inner product $\langle A, B \rangle = \text{Tr } AB$. One can easily conclude that an extreme density D (with respect to \mathbf{X}) must be singular if $n^2 > k + 1$. Actually, the last

inequality guarantees the existence of a Hermitian perturbation S which satisfies the orthogonality constraints, i.e., S is orthogonal to I and X_i 's. Moreover, the continuity of the spectra here gives that any small perturbation $D \pm \varepsilon S$ is positive if D is invertible.

Let $\sigma(A)$ denote the spectrum of any $A \in M_n(\mathbb{C})$. Suppose that the matrix D is of rank r. Then there do exist a $Y \in M_{n \times r}(\mathbb{C})$ and $R \in H_r(\mathbb{C})$ such that $D = YRY^*$.

Now one can prove the following lemma which is analogous to [6, Theorem 1 (a)].

Lemma 2.5. Let $D = YRY^* \in \mathcal{D}(\mathbf{X})$ be a density of rank r. Then S is a perturbation of D if and only if $\operatorname{Tr} S = 0$ and $S = YQY^*$ where $Q \in H_r(\mathbb{C})$.

Proof. First, assume that $S = YQY^*$. Then S is nonzero if and only if $Q \neq 0$. Indeed, we have rank $S = \operatorname{rank} Q$ because Y has full column rank r. Since $D = YRY^*$ is positive, we obtain that R is positive and invertible. From $0 \notin \sigma(R)$, there does exist an $\varepsilon > 0$ such that $D \pm \varepsilon S = Y(R \pm \varepsilon Q)Y^*$ are positive. Obviously, we get that S is a perturbation.

Conversely, let us assume that S is a perturbation of D. Clearly, $\operatorname{Tr} S = 0$ must hold. Expand Y with a matrix $Z \in M_{n \times (n-r)}(\mathbb{C})$ such that V = (Y|Z) is invertible and $V(R \oplus 0_{n-r})V^* = D$ holds. Next, let us write $V^{-1}S(V^*)^{-1}$ into blocks that correspond to the block form of $R \oplus 0_{n-r}$. Since $V^{-1}(D \pm \varepsilon S)(V^{-1})^*$ are positive for some $\varepsilon > 0$, it follows that $S = V(Q \oplus 0_{n-r})V^*$ must hold for some $Q \in H_r(\mathbb{C})$. \square

Here we present our main result which reflects some similarity with the characterization theorem of extreme correlations, see [6, Theorem 1].

Theorem 2.6. Let $X_i \in H_n(\mathbb{C})$, $1 \le i \le k$, and $D = YRY^* \in \mathcal{D}(\mathbf{X})$ be a density of rank r, where $Y \in M_{n \times r}(\mathbb{C})$. The following are equivalent:

- (i) D is an extreme point of $\mathcal{D}(\mathbf{X})$,
- (ii) span $\{Y^*X_1Y, \dots, Y^*X_kY, Y^*Y\} = H_r(\mathbb{C}),$
- (iii) $\{DX_1D, \dots, DX_kD, D^2\}$ has (real) rank r^2 .

Moreover, if $D = YY^*$, then the above statements are equivalent to:

(iv) $r^{-1}I_r$ is an extreme density with respect to Y^*XY , that is,

$$\mathcal{D}(Y^*\mathbf{X}Y) = \{r^{-1}I_r\}.$$

Proof.

- (i) \Leftrightarrow (ii). From Lemma 2.5, D is extreme if and only if there does not exist $0 \neq YQY^*$ such that $\operatorname{Tr} YQY^*X_i = \operatorname{Tr} Q(Y^*X_iY) = 0$ and $\operatorname{Tr} YQY^* = \operatorname{Tr} Q(Y^*Y) = 0$. We notice that Q = 0 if and only if the linear span of Y^*X_1Y, \ldots, Y^*X_kY and Y^*Y is the full space $H_r(\mathbb{C})$.
- (iii) \Leftrightarrow (ii). Let us choose the decomposition $D = YY^*$, that is, $R = I_r$. Note that the self-adjoint $Y^*Y \in M_r(\mathbb{C})$ is invertible. In fact, $\sigma(YY^*) \cup \{0\} = \sigma(Y^*Y) \cup \{0\}$ holds; thus, $\sigma(Y^*Y)$ equals the set of positive eigenvalues of D (with multiplicities). This implies that

$$\sum_{i=0}^{k} \alpha_i Y^* X_i Y = 0$$

if and only if

$$\sum_{i=0}^{k} \alpha_i Y Y^* X_i Y Y^* = 0$$
$$(\alpha_i \in \mathbb{R}, \quad X_0 = I_n),$$

so the systems $\{Y^*X_1Y,\ldots,Y^*X_kY,Y^*Y\}$ and $\{DX_1D,\ldots,DX_kD,D^2\}$ have the same rank.

- (i) \Rightarrow (iv). Since D is an extreme point, we get from (ii) that $\{Y^*X_1Y,\ldots,Y^*X_kY\}$ has rank at least r^2-1 . However, I_r is not in the linear span of the above system because it is orthogonal to every matrix Y^*X_iY . Adjusting $r^{-1}I_r$ to $Y^*\mathbf{X}Y$, we get a full rank system of $H_r(\mathbb{C})$. Hence, by (iii), we conclude that $r^{-1}I_r$ is an extreme point of $\mathcal{D}(Y^*\mathbf{X}Y)$.
- (iv) \Rightarrow (i). If $r^{-1}I_r$ is an extreme point, it has no perturbation S which is orthogonal to every Y^*X_iY . Thus, it follows that $I_r, Y^*X_1Y, \ldots, Y^*X_kY$ must span $H_r(\mathbb{C})$, that is, $\mathcal{D}(Y^*\mathbf{X}Y) = \{r^{-1}I_r\}$. Note that $Y^*Y, Y^*X_1Y, \ldots, Y^*X_kY$ span $H_r(\mathbb{C})$ as well because $\operatorname{Tr} Y^*Y = \operatorname{Tr} D = 1$ and Y^*X_iY s are traceless. Thus, (ii) implies that D is an extreme point.

The theorem gives a straightforward estimate of the rank of extreme densities.

Corollary 2.7. Let $D \in M_n(\mathbb{C})$ be an extreme density with respect to $X_1, \ldots, X_k \in H_n(\mathbb{C})$. Then

$$\operatorname{rank} D \le \sqrt{k+1}.$$

The Krein-Milman theorem implies that $Var_D(\mathbf{X})$ can be written as the convex sum of covariances determined by densities of rank at most $\sqrt{k+1}$. Moreover, one can easily deduce the following result which first appeared in [5], [8, Theorem] and [9].

Corollary 2.8. Let $D \in M_n(\mathbb{C})$ denote a density matrix. In the case of k = 1 and k = 2, there exist rank-one projections P_1, \ldots, P_m such that

$$D = \sum_{l=1}^{m} \lambda_l P_l \quad and \quad \text{Var}_D(\mathbf{X}) = \sum_{l=1}^{m} \lambda_l \text{Var}_{P_l}(\mathbf{X})$$

hold, where

$$\sum_{l=1}^{m} \lambda_l = 1$$

and $0 < \lambda_l < 1$.

In the case of $k \geq 3$, one might expect that the covariance matrix still can be decomposed by means of projections if n is large enough. However, this is not necessarily true. The next example shows that the estimate of Corollary 2.7 is sharp if n is large enough.

Example 2.9. Let $n = \lfloor \sqrt{k+1} \rfloor$. The special unitary group SU(n) has dimension $n^2 - 1$, so let λ_i , $1 \le i \le n^2 - 1$, denote a collection of its traceless, Hermitian infinitesimal generators. One can also assume that $\operatorname{Tr} \lambda_i \lambda_j = 0$ holds for every $i \ne j$ (for the generalized Gell-Mann matrices, see e.g., [10]). Then the matrices $\{I_n, \lambda_1, \ldots, \lambda_{n^2-1}\}$ span the real vector space $H_n(\mathbb{C})$. Thus, it follows that

$$\mathcal{D}(\lambda_1, \dots, \lambda_{n^2 - 1}) = \left\{ \frac{I_n}{n} \right\}$$

is a singleton; hence, $(1/n)I_n$ is an extreme density of rank n. If $n^2 < k+1$, let us choose arbitrary $\lambda_{n^2}, \ldots, \lambda_k \in M_m(\mathbb{C})$ Hermitians which are linearly independent where m is large enough. From Theorem 1 (iii), $(1/n)I_n \oplus 0_m$ remains extremal with respect to $\lambda = (\lambda_1 \oplus 0_m, \ldots, \lambda_{n^2-1} \oplus 0_m, 0_n \oplus \lambda_{n^2}, \ldots, 0_n \oplus \lambda_k)$; hence, $\mathrm{Var}_{(1/n)I_n \oplus 0_m}(\lambda)$ is not decomposable.

Applying direct sums as above, for every large n, one can construct $n \times n$ extreme densities of arbitrary rank between 1 and $\sqrt{k+1}$.

The method we used is very similar to that of describing extreme correlations. However, the next example shows that $Var_D(\mathbf{X})$ is not necessarily extreme even if it is a correlation matrix and D is an extreme density (with respect to some tuple).

Example 2.10. Let D be the projection diag $(1, 0, ..., 0) \in \mathbb{R}^{n+1}$. We define the Hermitians in $H_{n+1}(\mathbb{C})$ as

$$X_1 := \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \oplus 0_{n-1},$$

$$X_2 := \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \oplus 0_{n-2}, \dots, X_n := \begin{bmatrix} 0 & \cdots & 0 & 1 \\ \vdots & \vdots & \vdots & 0 \\ 0 & \vdots & \vdots & \vdots \\ 1 & 0 & \cdots & 0 \end{bmatrix}.$$

Then a simple calculation gives that $Var_D(\mathbf{X}) = I_n$, which is obviously not an extreme correlation matrix.

Finally, for the converse, we give an example that $\operatorname{Var}_D(\mathbf{X})$ can be an extreme correlation matrix while D is not necessarily extremal (with respect to \mathbf{X}).

Example 2.11. Consider $D = (1/n)I_n \oplus 0_n \in H_{2n}(\mathbb{C}), n > 2$. Let us choose reals x_1, \ldots, x_n such that

$$\sum_{i=1}^{n} x_i = 0$$
 and $\sum_{i=1}^{n} nx_i^2 = 1$

hold. For any $\widetilde{X}_i \in H_n(\mathbb{C})$, $1 \leq i \leq n$, we set

$$X_i = \operatorname{diag}(x_1, \dots, x_n) \oplus \widetilde{X}_i \in H_{2n}(\mathbb{C}), \quad 1 \le i \le n.$$

Then we get that $\operatorname{Var}_D(\mathbf{X})$ is the $n \times n$ matrix which consists only of ones, that is, it is a rank-one extreme correlation matrix. From Corollary 2.7, D cannot be extreme with respect to \mathbf{X} .

REFERENCES

- R. Bhatia, Positive definite matrices, Princeton University Press, Oxford, 2007.
- 2. R. Bhatia and C. Davis, More operator versions of the Schwarz inequality, Comm. Math. Phys. 215 (2000), 239–244.
- **3.** J.P.R. Christensen and J. Vesterstrøm, A note on extreme positive definite matrices, Math. Ann. **244** (1979), 65–68.
- 4. R. Grone, S. Pierce and W. Watkins, *Extremal correlation matrices*, Lin. Alg. Appl. **134** (1990), 63–70.
- Z. Léka and D. Petz, Some decompositions of matrix variances, Prob. Math. Stat. 33 (2013), 191–199.
- C-K. Li and B-S. Tam, A note on extreme correlation matrices, SIAM J. Matrix Anal. Appl. 15 (1994), 903–908.
- 7. K.R. Parthasarathy, An introduction to quantum stochastic calculus, Birkhäuser Verlag, Basel, 1992.
- D. Petz and G. Tóth, Matrix variances with projections, Acta Sci. Math. 78 (2012), 683–688.
- 9. _____, Extremal properties of the variance and the quantum Fisher information, Phys. Rev. 87 (2013), 032324.
- 10. H-J. Sommers and K. Zyckowski, Hilbert-Schmidt volume of the set of mixed quantum states, J. Phys. 36 (2003), 10115–10130.
- S. Yu, Quantum Fisher information as the convex roof of variance, preprint, arXiv:1302.5311.

ALFRÉD RÉNYI INSTITUTE OF MATHEMATICS, BUDAPEST, REÁLTANODA U. 13-15, 1053 HUNGARY

Email address: leka.zoltan@renyi.mta.hu