Acta Math. (Szeged),
75(2009), 265–281

Gaussian Markov triplets approached by block matrices

Tsuyoshi Ando[2] and Dénes Petz[13]

[2]Shiroishi-ku, Hongo-dori 9, Sapporo, Japan
[3]Alfréd Rényi Institute of Mathematics, H-1364 Budapest, POB 127, Hungary

Abstract

Multivariate normal distributions are described by a positive definite matrix and if their joint distribution is Gaussian as well then it can be represented by a block matrix. The aim of this note is to study Markov triplets by using the block matrix technique. A Markov triplet is characterized by the form of its block covariance matrix and by the form of the inverse of this matrix. A strong subadditivity of entropy is proved for a triplet and equality corresponds to the Markov property. The results are applied to multivariate stationary homogeneous Gaussian Markov chains.

2000 *Mathematics Subject Classification.* Primary 54C70, 60J05; Secondary 40C05, 60G15.
Key words and phrases. Normal distributions, Markov property, entropy, Schur complement, Hida-Cramér representation, Markov chain.

1 Multivariate Gaussian distributions

This section is an introduction to multivariate Gaussian distributions [1, Chap. 2]. For the sake of completeness, the well-known results are proven and the block matrix techniques are emphasized.

Let $\mathbf{X} := (X_1, X_2, \dots, X_n)$ be an n-tuple of real or complex random variables. (The n-tuple is written as a row vector, but if an $n \times n$ matrix (of a linear transformation) is applied to the row vector, then it should be regarded as a column vector.)

[1]E-mail: petz@math.bme.hu. Partially supported by the Hungarian Research Grant OTKA T068258. The authors thank to Professor Takeyuki Hida and Ádám Nagy for useful discussions.

The (i,j) element of the $n \times n$ **covariance matrix** is

$$C_{ij} := \mathbb{E}(X_i\overline{X_j}) - \mathbb{E}(X_i)\mathbb{E}(\overline{X_j}), \tag{1}$$

where $\mathbb{E}$ denotes the expectation. The covariance matrix is positive semidefinite. The mean vector $\mathbf{m} := (m_1, m_2, \dots, m_n)$ consists of the expectations $m_i := \mathbb{E}(X_i)$ $(1 \le i \le n)$. We mostly assume that the random variables have 0 expectation.

Lemma 1 *Let $\mathbf{Y} := (Y_1, Y_2, \dots, Y_n)$ be an n-tuple of real or complex random variables with 0 expectation and let T be an $n \times n$ matrix. The covariance matrix of $\mathbf{X}$ determined by the equation $\mathbf{X} := T\mathbf{Y}$ is TCT^*, where C is covariance matrix of $\mathbf{Y}$.*

Proof: We simply compute the expectation of $X_i\overline{X_j} = \sum_k T_{ik}Y_k \sum_\ell \overline{Y_\ell T_{j\ell}}$. □

Let M be a positive definite $n \times n$ matrix and $\mathbf{m}$ be a vector. Then

$$f_{\mathbf{m},M}(\mathbf{x}) := \sqrt{\frac{\operatorname{Det} M}{(2\pi)^n}} \exp\Big(-\tfrac{1}{2}\langle \mathbf{x}-\mathbf{m}, M(\mathbf{x}-\mathbf{m})\rangle\Big) \tag{2}$$

is a multivariate Gaussian probability distribution with expectation $\mathbf{m}$, see, for example, III.6 in [4]. The distribution (2) is denoted by $N(\mathbf{m}, M^{-1})$. If $\mathbf{m} = 0$, then instead of $f_{\mathbf{m},M}(\mathbf{x})$ we write simply $f_M(\mathbf{x})$. More generally, if a function $g(\mathbf{x})$ is constant times $\exp\big(-\frac{1}{2}\langle \mathbf{x}, M\mathbf{x}\rangle\big)$, then M will be called its **quadratic matrix**.

If M is diagonal, then the probability distribution (2) is the product of functions of one-variable which means independence.

In probability theory the random variables are mostly real-valued, therefore in the rest we work in the real setting. However, in some application of the block matrix formulation complex matrices are used. (In the appendix, matrices can be real or complex.)

Lemma 2 *If the joint distribution of $\mathbf{X} := (X_1, X_2, \dots, X_n)$ is $f_M(\mathbf{x})$, then*

$$\int \langle \mathbf{x}, B\mathbf{x}\rangle f_M(\mathbf{x})\, d\mathbf{x} = \operatorname{Tr} BM^{-1}. \tag{3}$$

In particular, for $B = E_{ij}$ we have

$$\int x_i x_j f_M(\mathbf{x})\, d\mathbf{x} = \int \langle \mathbf{x}, E_{ij}\mathbf{x}\rangle f_M(\mathbf{x})\, d\mathbf{x} = \operatorname{Tr} E_{ij}M^{-1} = (M^{-1})_{ij}$$

and the covariance matrix is M^{-1}.

Proof: We first note that if (3) is true for a matrix M, then

$$\begin{aligned}\int \langle \mathbf{x}, B\mathbf{x}\rangle f_{U^*MU}(\mathbf{x})\, d\mathbf{x} &= \int \langle BU^*\mathbf{x}, U^*\mathbf{x}\rangle f_M(\mathbf{x})\, d\mathbf{x} \\ &= \operatorname{Tr}(UBU^*)M^{-1} = \operatorname{Tr} B(U^*MU)^{-1}\end{aligned}$$

for an orthogonal U, since the Lebesgue measure on $\mathbb{R}^n$ is invariant under orthogonal transformation. This means that (3) holds also for U^*MU. Therefore to check (3), we may assume that M is diagonal. Another reduction concerns B, we may assume that B is a matrix unit E_{ij}. Then the n variable integral reduces to integrals on $\mathbb{R}$ and the known formulas for the mean and variance of a Gaussian distribution can be used. □

Lemma 3 *Let the joint distribution of* $\mathbf{X} := (X_1, X_2, \dots, X_n)$ *be Gaussian. There exist an* n*-tuple* $\mathbf{Z} := (Z_1, Z_2, \dots, Z_n)$ *of independent Gaussian random variables and an orthogonal matrix* U *such that* $\mathbf{X} = U\mathbf{Z}$.

Proof: Let M be from (2). Since M is positive definite, we can write in the form $M = UDU^*$, where U is orthogonal and D is positive definite and diagonal. The covariance matrix is $M^{-1} = UD^{-1}U^*$, therefore $\mathbf{X}^t = U\mathbf{Z}^t$, where $\mathbf{Z}$ is determined by D. Since D is diagonal, the components of Z are independent. □

Lemma 4 *Let*

$$M = \begin{bmatrix} A & B \\ B^* & D \end{bmatrix}$$

be a positive definite $(m+k) \times (m+k)$ *matrix written in the form of a block matrix. Then the marginal of the Gaussian probability distribution*

$$f_M(\mathbf{x}_1, \mathbf{x}_2) := \sqrt{\frac{\mathrm{Det}\, M}{(2\pi)^{m+k}}} \exp\Big(-\tfrac{1}{2}\langle(\mathbf{x}_1, \mathbf{x}_2), M(\mathbf{x}_1, \mathbf{x}_2)\rangle\Big)$$

on $\mathbb{R}^m$ *is the distribution*

$$f_1(\mathbf{x}_1) = \sqrt{\frac{\mathrm{Det}\, M}{(2\pi)^m \mathrm{Det}\, D}} \exp\Big(-\tfrac{1}{2}\langle \mathbf{x}_1, (A - BD^{-1}B^*)\mathbf{x}_1\rangle\Big). \tag{4}$$

Proof: We have

$$\begin{aligned}
\langle(\mathbf{x}_1, \mathbf{x}_2), M(\mathbf{x}_1, \mathbf{x}_2)\rangle &= \langle A\mathbf{x}_1 + B\mathbf{x}_2, \mathbf{x}_1\rangle + \langle B^*\mathbf{x}_1 + D\mathbf{x}_2, \mathbf{x}_2\rangle \\
&= \langle A\mathbf{x}_1, \mathbf{x}_1\rangle + \langle B\mathbf{x}_2, \mathbf{x}_1\rangle + \langle B^*\mathbf{x}_1, \mathbf{x}_2\rangle + \langle D\mathbf{x}_2, \mathbf{x}_2\rangle \\
&= \langle A\mathbf{x}_1, \mathbf{x}_1\rangle + \langle D(\mathbf{x}_2 + W\mathbf{x}_1), (\mathbf{x}_2 + W\mathbf{x}_1)\rangle - \langle DW\mathbf{x}_1, W\mathbf{x}_1\rangle,
\end{aligned}$$

where $W = D^{-1}B^*$. We integrate on $\mathbb{R}^k$ as

$$\begin{aligned}
&\int \exp\Big(-\tfrac{1}{2}\langle(\mathbf{x}_1, \mathbf{x}_2), M(\mathbf{x}_1, \mathbf{x}_2)\rangle\Big)\, d\mathbf{x}_2 \\
&\quad = \exp\Big(-\tfrac{1}{2}(\langle A\mathbf{x}_1, \mathbf{x}_1\rangle - \langle DW\mathbf{x}_1, W\mathbf{x}_1\rangle)\Big) \\
&\qquad \times \int \exp\Big(-\tfrac{1}{2}\langle D(\mathbf{x}_2 + W\mathbf{x}_1), (\mathbf{x}_2 + W\mathbf{x}_1)\rangle\Big)\, d\mathbf{x}_2 \\
&\quad = \exp\Big(-\tfrac{1}{2}\langle(A - BD^{-1}B^*)\mathbf{x}_1, \mathbf{x}_1\rangle\Big) \sqrt{\frac{(2\pi)^k}{\mathrm{Det}\, D}}
\end{aligned}$$

and obtain the result. □

It is interesting that the matrix $A - BD^{-1}B^*$ appears also in matrix analysis. If

$$\mathbf{A} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix},$$

then $A_{11} - A_{12}A_{22}^{-1}A_{21}^*$ is called the **Schur complement** of A_{22} in $\mathbf{A}$ (when A_{22} is invertible) and it is denoted by $\mathbf{A}/A_{22}$ [7, Section 7.7]. The notation is justified by the **determinant formula**

$$\mathrm{Det}\,\mathbf{A} = \mathrm{Det}\,A_{22} \times \mathrm{Det}\,(\mathbf{A}/A_{22}). \tag{5}$$

When $\mathbf{A}$ is the above M, this formula comes also from Lemma 4.

For a positive semidefinite matrix, the Schur complement has an important meaning. The following lemma is from [2].

Lemma 5 *Let M be a positive operator and E be an orthogonal projection. The set*

$$\{N : 0 \le N \le M \text{ and } EN = N\}$$

always admits a maximal element, which is denoted by $[E]M$. When E and M are written in block matrix form such that

$$E = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix} \text{ and } M = \begin{bmatrix} A & B \\ B^* & D \end{bmatrix},$$

then

$$[E]M = \begin{bmatrix} A - BD^{-1}B^* & 0 \\ 0 & 0 \end{bmatrix}.$$

Another important formula is about the inverse:

$$\begin{bmatrix} A & B \\ B^* & D \end{bmatrix}^{-1} = \begin{bmatrix} W & -WBD^{-1} \\ -D^{-1}B^*W & D^{-1} + D^{-1}B^*WBD^{-1} \end{bmatrix}, \tag{6}$$

where $W = (A - BD^{-1}B^*)^{-1}$. This implies that the covariance matrix of the first m variables is $EM^{-1}E = (A - BD^{-1}B^*)^{-1}$, where the projection E is the same as in Lemma 5. The quadratic matrix is $A - BD^{-1}B^*$ which can be regarded as $[E]M$.

Now we move to conditional distributions. Given the random variables $\mathbf{X}_1$ and $\mathbf{X}_2$, the **conditional density**

$$f(\mathbf{x}_2|\mathbf{x}_1) := \frac{f_{(\mathbf{X}_1,\mathbf{X}_2)}(\mathbf{x}_1, \mathbf{x}_2)}{f_{\mathbf{X}_1}(\mathbf{x}_1)} \tag{7}$$

is a function of the variable $\mathbf{x}_2$ ($\mathbf{x}_1$ is fixed and gives the condition). So $f(\mathbf{x}_2|\mathbf{x}_1)$ is the quotient of the joint density and the density of $\mathbf{X}_1$.

Lemma 6 *Let* $(\mathbf{X}_1, \mathbf{X}_2)$ *be Gaussian with quadratic matrix*

$$\begin{bmatrix} A & B \\ B^* & D \end{bmatrix}$$

and expectation 0. Then the conditional distribution $f(\mathbf{x}_2|\mathbf{x}_1)$ *has the form*

$$c \exp\Big(-\frac{1}{2}\langle (\mathbf{x}_1, \mathbf{x}_2), R(\mathbf{x}_1, \mathbf{x}_2)\rangle\Big), \qquad R = \begin{bmatrix} BD^{-1}B^* & B \\ B^* & D \end{bmatrix} \tag{8}$$

and it is Gaussian $N(-D^{-1}B^*\mathbf{x}_1, D^{-1})$.

Proof: Lemma 4 tells us the quadratic matrix of $\mathbf{X}_1$. Hence the quadratic matrix of the quotient (7) is

$$\begin{bmatrix} A & B \\ B^* & D \end{bmatrix} - \begin{bmatrix} A - BD^{-1}B^* & 0 \\ 0 & 0 \end{bmatrix}. \tag{9}$$

The identity

$$\langle (\mathbf{x}_1, \mathbf{x}_2), R(\mathbf{x}_1, \mathbf{x}_2)\rangle = \langle (\mathbf{x}_2 + D^{-1}B^*\mathbf{x}_1), D(\mathbf{x}_2 + D^{-1}B^*\mathbf{x}_1)\rangle$$

implies the last statement. □

In the notation of Lemma 5, the property $[E]M \le M$ implies the positivity of (8).

2 Gaussian Markov triplets

Let $(\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3)$ be random variables with joint probability distribution $f(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$. The distribution of the appropriate marginals are $f(\mathbf{x}_1, \mathbf{x}_2)$, $f(\mathbf{x}_2, \mathbf{x}_3)$ and $f(\mathbf{x}_2)$.

$(\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3)$ will be called a **Markov triplet** if

$$f(\mathbf{x}_3|\mathbf{x}_1, \mathbf{x}_2) = f(\mathbf{x}_3|\mathbf{x}_2). \tag{10}$$

The notation $\mathbf{X}_1 \to \mathbf{X}_2 \to \mathbf{X}_3$ will mean that $(\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3)$ is a Markov triplet.

Let $(\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3)$ be Gaussian random variables with quadratic matrix

$$M = \begin{bmatrix} A_1 & A_2 & B_1 \\ A_2^* & A_3 & B_2 \\ B_1^* & B_2^* & D \end{bmatrix} \tag{11}$$

and with expectation 0. Then the conditional distribution

$$f(\mathbf{x}_3|\mathbf{x}_1, \mathbf{x}_2) \tag{12}$$

has quadratic matrix

$$\begin{bmatrix} B_1 D^{-1} B_1^* & B_1 D^{-1} B_2^* & B_1 \\ B_2 D^{-1} B_1^* & B_2 D^{-1} B_2^* & B_2 \\ B_1^* & B_2^* & D \end{bmatrix}, \tag{13}$$

see (9). If the corresponding quadratic form does not depend on $\mathbf{x}_1$, then $B_1 = 0$ and the matrix simplifies:

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & B_2 D^{-1} B_2^* & B_2 \\ 0 & B_2^* & D \end{bmatrix}. \tag{14}$$

The quadratic matrix of $(\mathbf{X}_2, \mathbf{X}_3)$ is

$$\begin{bmatrix} A_3 - A_2^* A_1^{-1} A_2 & B_2 - A_2^* A_1^{-1} B_1 \\ B_2^* - B_1^* A_1^{-1} A_2 & D - B_1^* A_1^{-1} B_1 \end{bmatrix}. \tag{15}$$

If $B_1 = 0$, then it simplifies:

$$\begin{bmatrix} A_3 - A_2^* A_1^{-1} A_2 & B_2 \\ B_2^* & D \end{bmatrix}. \tag{16}$$

Now we compute the quadratic matrix of the conditional distribution

$$\frac{f_{(\mathbf{X}_2,\mathbf{X}_3)}(\mathbf{x}_2, \mathbf{x}_3)}{f_{\mathbf{X}_2}(\mathbf{x}_2)} \tag{17}$$

and obtain

$$\begin{bmatrix} B_2^* D^{-1} B_2^* & B_2 \\ B_2^* & D \end{bmatrix}. \tag{18}$$

The above argument is part of the proof of the next theorem.

Theorem 1 *For the Gaussian triplet* $(\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3)$ *with quadratic matrix (11) and with expectation* 0*, the following conditions are equivalent.*

(a) $\mathbf{X}_1 \to \mathbf{X}_2 \to \mathbf{X}_3$.

(b) $B_1 = 0$.

(c) The conditional distribution $f(\mathbf{x}_3|\mathbf{x}_1, \mathbf{x}_2)$ *does not depend on* $\mathbf{x}_1$.

(d) The covariance matrix of $(\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3)$ *is of the form*

$$\begin{bmatrix} S_{11} & S_{12} & S_{12} S_{22}^{-1} S_{23} \\ S_{12}^* & S_{22} & S_{23} \\ S_{23}^* S_{22}^{-1} S_{12}^* & S_{23}^* & S_{33} \end{bmatrix}. \tag{19}$$

Proof: The equivalence of (a), (b) and (c) was obtained above. It remains to show the equivalence of (b) and (d). This fact is contained in Theorem 4 in Appendix. □

The (i,j)-block of the covariance matrix of $(\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3)$ can be called the correlation matrix of $\mathbf{X}_i$ and $\mathbf{X}_j$. Note that condition (d) of the Theorem says that the correlation matrix of $\mathbf{X}_1$ and $\mathbf{X}_3$ equals

$$S_{13} = S_{12}S_{22}^{-1}S_{23}, \tag{20}$$

where S_{ij} is the correlation matrix of $\mathbf{X}_i$ and $\mathbf{X}_j$. (S_{22} may be called the covariance matrix of X_2.) This condition appears in the standard literature, for example [4], in the single variable case.

It is interesting to compare (20) with the **Hida-Cramér representation** of a Markov triplet [5, 6].

Corollary 1 *For the Gaussian triplet* $(\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3)$ *with 0 expectation, the property* $\mathbf{X}_1 \to \mathbf{X}_2 \to \mathbf{X}_3$ *holds if and only if there is a representation*

$$\begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \mathbf{X}_3 \end{bmatrix} = \begin{bmatrix} A_1 & 0 & 0 \\ C_1 & A_2 & 0 \\ C_3 & C_2 & A_3 \end{bmatrix} \begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \\ \mathbf{Y}_3 \end{bmatrix},$$

where the covariance of the (standard) Gaussian triplet $(\mathbf{Y}_1, \mathbf{Y}_2, \mathbf{Y}_3)$ *is the identity matrix,* A_1, A_2, A_3 *are invertible matrices and* $C_3 = C_2A_2^{-1}C_1$.

Proof: Suppose that $(\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3)$ is a Gaussian Markov triplet with correlation matrix $\mathbf{S} := [S_{ij}]_{i,j=1}^3$. Markovianity means that $S_{13} = S_{12}S_{22}^{-1}S_{23}$.

Since $\mathbf{S}$ is positive definite, $S_{11} > 0$, $[S_{ij}]_{i,j=1}^2 > 0$ and $[S_{ij}]_{i,j=2}^3 > 0$. Therefore

$$S_{22} - S_{21}S_{11}^{-1}S_{12} > 0 \quad \text{and} \quad S_{33} - S_{32}S_{22}^{-1}S_{23} > 0$$

and we can define A_1, A_2, A_3, B_1, B_2 as follows:

$$A_1 = S_{11}^{1/2} > 0, \qquad A_2 = (S_{22} - S_{21}S_{11}^{-1}S_{12})^{1/2} > 0,$$

$$A_3 = (S_{33} - S_{32}S_{22}^{-1}S_{23})^{1/2} > 0$$

and

$$B_1 = S_{21}S_{11}^{-1} \quad \text{and} \quad B_2 = S_{32}S_{22}^{-1}.$$

Then we have

$$\begin{bmatrix} S_{11} & S_{12} & S_{13} \\ S_{12}^* & S_{22} & S_{23} \\ S_{13}^* & S_{23}^* & S_{33} \end{bmatrix} = \begin{bmatrix} A_1 & 0 & 0 \\ B_1A_1 & A_2 & 0 \\ B_2B_1A_1 & B_2A_2 & A_3 \end{bmatrix} \times \begin{bmatrix} A_1 & 0 & 0 \\ B_1A_1 & A_2 & 0 \\ B_2B_1A_1 & B_2A_2 & A_3 \end{bmatrix}^*.$$

Choosing $C_1 = B_1A_1, C_2 = B_2A_2$ and $C_3 = B_2B_1A_1$ we get the stated representation.

Now we prove the converse. Since

$$S_{12} = A_1 C_1^*, \quad S_{13} = A_1 C_3^*, \quad S_{22} = C_1 C_1^* + A_2 A_2^*, \quad S_{23} = C_1 C_3^* + A_2 C_2^*,$$

we have to prove

$$C_3^* = C_1^* (C_1 C_1^* + A_2 A_2^*)^{-1} (C_1 C_3^* + A_2 C_2^*).$$

This equality is equivalent to

$$C_3^* = D^* (DD^* + I)^{-1} (DC_3^* + C_2^*),$$

where $D := A_2^{-1} C_1$. We can write this in the form

$$(I - D^*(DD^* + I)^{-1} D) C_3^* = D^*(DD^* + I)^{-1} C_2^*$$

and from the relations

$$D^*(DD^* + I)^{-1} = (I + DD^* + I)^{-1} D^*, \quad I - D^*(DD^* + I)^{-1} D = (I + DD^* + I)^{-1}$$

we deduce that our statement is equivalent to

$$C_3^* = D^* C_2^* \quad \text{or} \quad C_3 = C_2 D.$$

Since this is our assumption, the proof is complete. □

The proof shows that if A_1, A_2, A_3 are assumed to be positive definite, then the representation is unique.

It is also interesting to note that $\mathbf{X}_{n+1}$ is a linear transformation of $\mathbf{X}_n$ plus an independent Gaussian random variable in the Hida-Cramér representation. For example,

$$X_3 = C_2 A_2^{-1} (C_1 Y_1 + A_2 Y_2) + A_3 Y_3 = C_2 A_2^{-1} X_2 + A_3 Y_3,$$

where $C_2 A_2^{-1} X_2$ and $A_3 Y_3$ are independent, since X_2 is a linear transformation of (Y_1, Y_2). In case of an infinite process such relation holds for all n.

3 Entropy

In this section the Markov property is characterized by entropy quantities. We refer to [3, Chap. 8] about the basic concepts and we adapt the notation.

Let $\mathbf{X} := (X_1, X_2, \dots, X_n)$ be an n-tuple of real random variables with density $f(x_1, x_2, \dots, x_n)$. The **Boltzmann-Gibbs entropy** (or differential entropy) of $\mathbf{X}$ is defined as

$$h(\mathbf{X}) := -\int f(\mathbf{x}) \log f(\mathbf{x}) \, d\mathbf{x}$$

(whenever this has a meaning). In particular, if $\mathbf{X}$ has a multivariate Gaussian distribution (2), then its entropy is equal to

$$\frac{n}{2}\log(2\pi e) - \frac{1}{2}\log \operatorname{Det} M\,, \tag{21}$$

where $\operatorname{Det} M$ denotes the determinant of M.

The **relative entropy** of the random variables $\mathbf{X}_1$ and $\mathbf{X}_2$ is defined as

$$D(\mathbf{X}_1\|\mathbf{X}_2) := \int f_1(\mathbf{x})(\log f_1(\mathbf{x}) - \log f_2(\mathbf{x}))\, d\,\mathbf{x},$$

where f_1 and f_2 are the probability densities. Up to a sign, the differential entropy is the relative entropy with respect to the Lebesgue measure. Since this is not a probability measure, we replace it with a Gaussian measure.

Let $\mathbf{X}$ and $\mathbf{Y}$ be n-tuples of real variables, assume that the distribution of $\mathbf{Y}$ is $N(0, C)$. Then

$$D(\mathbf{X}\|\mathbf{Y}) = -h(\mathbf{X}) + \tfrac{1}{2}\operatorname{Tr} C_{\mathbf{X}} M - \tfrac{1}{2}\log \operatorname{Det} M - \frac{n}{2}\log 2\pi, \tag{22}$$

where $M = C^{-1}$. The natural choice for C is I_n, then the expression simplifies. The formula makes sense if the differential entropy $h(\mathbf{X})$ and the covariance $C_{\mathbf{X}}$ are well-defined.

Theorem 2 *Let $\mathbf{X}_i$ be a random variable with values in $\mathbb{R}^{n_i}$, $i = 1, 2, 3$. Assume that the differential entropy and the covariance of $(\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3)$ are well-defined. Then the strong subadditivity for the differential entropy holds:*

$$h(\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3) + h(\mathbf{X}_2) \le h(\mathbf{X}_1, \mathbf{X}_2) + h(\mathbf{X}_2, \mathbf{X}_3). \tag{23}$$

The equality holds if and only if the Markov condition (10) is satisfied.

Proof: Let $\mathbf{Y}_i$ be independent Gaussian random variable of density $N(0, I_{n_i})$, $i = 1, 2, 3$. Then condition (23) is equivalent to

$$D(\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3\|\mathbf{Y}_1, \mathbf{Y}_2, \mathbf{Y}_3) + D(\mathbf{X}_2\|\mathbf{Y}_2) \ge D(\mathbf{X}_1, \mathbf{X}_2\|\mathbf{Y}_1, \mathbf{Y}_2) + D(\mathbf{X}_2, \mathbf{X}_3\|\mathbf{Y}_2, \mathbf{Y}_3) \tag{24}$$

or

$$D(\mathbf{X}_2, \mathbf{X}_3\|\mathbf{X}_2, \mathbf{Y}) \le D(\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3\|\mathbf{X}_1, \mathbf{X}_2, \mathbf{Y}). \tag{25}$$

This is true due to the monotonicity of the relative entropy [9].

Assume now the equality. Since $\mathbf{Y}$ is chosen to be independent, this condition can be reformulated for the Radon–Nikodym derivatives as follows:

$$\frac{f_{(\mathbf{X}_2,\mathbf{X}_3)}(\mathbf{x}_2, \mathbf{x}_3)}{f_{\mathbf{X}_2}(\mathbf{x}_2) f_{\mathbf{Y}}(\mathbf{x}_3)} = \frac{f_{(\mathbf{X}_1,\mathbf{X}_2,\mathbf{X}_3)}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)}{f_{(\mathbf{X}_1,\mathbf{X}_2)}(\mathbf{x}_1, \mathbf{x}_2) f_{\mathbf{Y}}(\mathbf{x}_3)},$$

see [8]. Since we can omit $f_{\mathbf{Y}}(\mathbf{x}_3)$, this is exactly the Markov property. $\square$

If $(\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3)$ is Gaussian with covariance

$$S = \begin{bmatrix} S_{11} & S_{12} & S_{13} \\ S_{12}^* & S_{22} & S_{23} \\ S_{13}^* & S_{23}^* & S_{33} \end{bmatrix},$$

then

$$h(\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3) + h(\mathbf{X}_2) = \frac{n_1 + n_2 + n_3}{2} \log(2\pi e) + \frac{1}{2} \log \operatorname{Det} S + \frac{n_2}{2} \log(2\pi e) + \frac{1}{2} \log \operatorname{Det} S_{22}$$

and inequality (23) in the Theorem becomes

$$\operatorname{Det} \begin{bmatrix} S_{11} & S_{12} & S_{13} \\ S_{12}^* & S_{22} & S_{23} \\ S_{13}^* & S_{23}^* & S_{33} \end{bmatrix} \times \operatorname{Det} S_{22} \le \operatorname{Det} \begin{bmatrix} S_{11} & S_{12} \\ S_{12}^* & S_{22} \end{bmatrix} \times \operatorname{Det} \begin{bmatrix} S_{22} & S_{23} \\ S_{23}^* & S_{33} \end{bmatrix}. \tag{26}$$

The condition for equality is $S_{13} = S_{12} S_{22}^{-1} S_{23}$. A different proof of this will be given in the Appendix with block matrix techniques.

4 Gaussian Markov chains

Let $\mathbf{X}_1, \mathbf{X}_2, \dots$ be a multivariate stationary homogeneous Gaussian Markov chain. Assume that $\mathbb{E}(\mathbf{X}_n) = 0$ and set $S := \mathbb{E}(\mathbf{X}_n \mathbf{X}_n^*)$, $C := \mathbb{E}(\mathbf{X}_n \mathbf{X}_{n+1}^*)$. Since for $u < v < w$, $(\mathbf{X}_u, \mathbf{X}_v, \mathbf{X}_w)$ is a Markovian triplet, application of Theorem 1 gives the covariance matrix $\mathbf{C}_n$ of $(\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n)$. The (i, j)-entry is

$$\mathbb{E}(\mathbf{X}_i \mathbf{X}_j^*) = S \left(S^{-1} C \right)^{j-i} \tag{27}$$

for $j \ge i$. (Note that a symmetric matrix is determined by its (i, j)-entries with $j \ge i$.)

Our aim is to find the inverse of $\mathbf{C}_n$ which is the quadratic matrix of $(\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n)$.

Lemma 7 *Let $\|K\| < 1$ and define an $n \times n$ symmetric block matrix $\mathbf{H}_n = [H_{ij}]_{i,j=1}^n$ as*

$$H_{ij} := K^{j-i} \quad \textit{for} \quad j \ge i.$$

Then the inverse $\mathbf{H}_n^{-1} := [G_{ij}]_{i,j=1}^n$ is symmetric tridiagonal matrix determined by the entries

$$G_{11} = I + KMK^*, \qquad G_{nn} = M$$

$$G_{ii} = KMK^* + M \quad \textit{for} \quad 2 \le i \le n-1,$$

$$G_{i,i+1} = -KM, \quad G_{i,j} = 0 \quad \textit{for} \quad j > i+1,$$

where $M = (I - K^* K)^{-1}$.

Proof: Assume that K acts on a Hilbert space $\mathcal{H}$ and identify the block matrices with elements of $B(\mathcal{H}) \otimes M_n$. We shall use a matrix $U \in M_n$ defined as

$$U_{ij} = \begin{cases} 1 & \text{if} \quad j = i+1, \\ 0 & \text{otherwise.} \end{cases}$$

Then we have

$$\begin{aligned}
\mathbf{H}_n &= I \otimes I + \sum_{k=1}^{n-1} (K \otimes U)^k + \sum_{k=1}^{n-1} (K^* \otimes U^*)^k \\
&= I \otimes I + \sum_{k=1}^{\infty} (K \otimes U)^k + \sum_{k=1}^{\infty} (K^* \otimes U^*)^k \\
&= -I \otimes I + (I \otimes I K \otimes U)^{-1} + (I \otimes I K^* \otimes U^*)^{-1} \\
&= (I \otimes I - K^* \otimes U^*)^{-1} \Big(I \otimes I - (K^* K) \otimes (U^* U) \Big) (I \otimes I - K \otimes U)^{-1}
\end{aligned}$$

based on the fact $U^n = 0$. Since we have product form, the inverse is computed now factor-by-factor. The middle factor is diagonal and its inverse is

$$\mathrm{Diag}\,(I, (I - K^*K)^{-1}, (I - K^*K)^{-1}, \ldots, (I - K^*K)^{-1}).$$

Now the statement is obtained by simple computation. □

Theorem 3 *Let $\mathbf{X}_1, \mathbf{X}_2, \ldots$ be a multivariate stationary homogeneous Gaussian Markov chain as it is described above. Then the quadratic matrix of $(\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n)$ is a tridiagonal symmetric matrix with diagonal*

$$L_1, L_1 + L_2 - S^{-1}, \ldots, L_1 + L_2 - S^{-1}, L_2$$

and with upper diagonal

$$-S^{-1} C L_2, -S^{-1} C L_2, \ldots, -S^{-1} C L_2,$$

where $L_1 = (S - C S^{-1} C^*)^{-1}$ and $L_2 = (S - C^* S^{-1} C)^{-1}$.

Proof: Apply the lemma to the operator

$$\mathbf{H}_n = \mathrm{Diag}\,(S^{-1/2}, S^{-1/2}, \ldots, S^{-1/2}) \mathbf{C}_n \mathrm{Diag}\,(S^{-1/2}, S^{-1/2}, \ldots, S^{-1/2})$$

and $K = S^{-1/2} C S^{-1/2}$. Then

$$\mathbf{C}_n^{-1} = \mathrm{Diag}\,(S^{-1/2}, S^{-1/2}, \ldots, S^{-1/2}) \mathbf{H}_n^{-1} \mathrm{Diag}\,(S^{-1/2}, S^{-1/2}, \ldots, S^{-1/2})$$

and we obtain the statement. □

Example 1 Assume that $X_1, X_2, \ldots$ is a stationary homogeneous Gaussian Markov chain such that $\mathbb{E}(X_n) = 0$, $\sigma^2 := \mathbb{E}(X_n^2)$ and $r := \mathbb{E}(X_1 X_2)$. Then the quadratic matrix of (X_1, X_2, X_3, X_4) is

$$\frac{1}{\sigma^2(1-q^2)} \begin{bmatrix} 1 & -q & 0 & 0 \\ -q & (1+q^2) & -q & 0 \\ 0 & -q & (1+q^2) & -q \\ 0 & 0 & -q & 1 \end{bmatrix},$$

where $q = r/\sigma^2$. □

Appendix

In this section we study the inverse of block matrices. We start with the (well-known) 2×2 case to warm up and then treat the 3×3 case which is related to the Markovian triplets and the result is used in Section 2. Since our result may have application in a non-probabilistic setting we emphasize that the matrices can be real or complex.

Consider an invertible numerical 2×2 matrix

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

with $a_{22} \neq 0$. A is invertible if and only if its determinant

$$\mathrm{Det}\, A = a_{11}a_{22} - a_{12}a_{21} = a_{22}(a_{11} - a_{12}a_{22}^{-1}a_{21}) \tag{28}$$

does not vanish, or equivalently $a_{11} - a_{12}a_{22}^{-1}a_{21}$ does not vanish. In this case

$$\begin{aligned} A^{-1} &= \frac{1}{\mathrm{Det}\, A} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix} \\ &= \begin{bmatrix} (a_{11} - a_{12}a_{22}^{-1}a_{21})^{-1} & -(a_{11} - a_{12}a_{22}^{-1}a_{21})^{-1}a_{12}a_{22}^{-1} \\ a_{22}^{-1}a_{21}(a_{11} - a_{12}a_{22}^{-1}a_{21})^{-1} & a_{22}^{-1} + a_{22}^{-1}a_{21}(a_{11} - a_{12}a_{22}^{-1}a_{21})^{-1}a_{12}a_{22}^{-1} \end{bmatrix}. \end{aligned}$$

Next let $\mathbf{A}$ be a 2×2 block matrix

$$\mathbf{A} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

such that A_{11} and A_{22} are square matrices, possibly of different order. If $\mathbf{A}$ and A_{22} are invertible, then

$$\mathbf{A}^{-1} = \begin{bmatrix} (A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1} & -(A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1}A_{12}A_{22}^{-1} \\ A_{22}^{-1}A_{21}(A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1} & A_{22}^{-1} + A_{22}^{-1}A_{21}(A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1}A_{12}A_{22}^{-1} \end{bmatrix}.$$

The analogue of (28) is **Schur's determinant formula**:

$$\mathrm{Det}\, \mathbf{A} = \mathrm{Det}\, A_{22} \times \mathrm{Det}\, (A_{11} - A_{12}A_{22}^{-1}A_{21}). \tag{29}$$

In the 3×3 case we intend to compute the $(1,3)$-entry of the inverse. Let $A = [a_{ij}]_{i,j=1}^{3}$ be an invertible numerical matrix with inverse $A^{-1} = [b_{ij}]_{i,j=1}^{3}$. Then

$$b_{13} = \frac{\mathrm{Det} \begin{bmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{bmatrix}}{\mathrm{Det}\, A}. \tag{30}$$

If $a_{22} \neq 0$, then

$$\mathrm{Det} \begin{bmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{bmatrix} = a_{22}(a_{33} - a_{32}a_{22}^{-1}a_{23}),$$

$$\mathrm{Det}\begin{bmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{bmatrix} = a_{22}(a_{12}a_{22}^{-1}a_{23} - a_{13}).$$

Moreover, if $[a_{ij}]_{i,j=2}^3$ is invertible, then we have

$$\mathrm{Det}\, A = \mathrm{Det}\left(a_{11} - [\, a_{12}, \quad a_{13}\,] \begin{bmatrix} a_{22}, & a_{23} \\ a_{32} & a_{33} \end{bmatrix}^{-1} \begin{bmatrix} a_{12} \\ a_{13} \end{bmatrix}\right) \times \mathrm{Det}\begin{bmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{bmatrix}$$

By substitution into (30) we conclude

$$b_{13} = \left(a_{11} - [\, a_{12}, \quad a_{13}\,] \begin{bmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{bmatrix}^{-1} \begin{bmatrix} a_{21} \\ a_{31} \end{bmatrix}\right)^{-1} (a_{12}a_{22}^{-1}a_{23} - a_{13})(a_{33} - a_{32}a_{22}^{-1}a_{23})^{-1}.$$

This formula admits a block matrix version whose proof is accomplished by repeated use the inverse formula of a 2×2 block matrix. This is the content of the next theorem.

Theorem 4 *Let $\mathbf{A} = [A_{ij}]_{i,j=1}^3$ be an invertible block matrix and assume that A_{22} and $[A_{ij}]_{i,j=2}^3$ are invertible. Then the $(1,3)$-entry of the inverse $\mathbf{A}^{-1} = [B_{ij}]_{i,j=1}^3$ is given by the following formula:*

$$\left(A_{11} - [\, A_{12}, \quad A_{13}\,] \begin{bmatrix} A_{22}, & A_{23} \\ A_{32} & A_{33} \end{bmatrix}^{-1} \begin{bmatrix} A_{12} \\ A_{13} \end{bmatrix}\right)^{-1} (A_{12}A_{22}^{-1}A_{23} - A_{13})(A_{33} - A_{32}A_{22}^{-1}A_{23})^{-1}.$$

Hence $B_{13} = 0$ if and only if $A_{13} = A_{12}A_{22}^{-1}A_{23}$.

The next theorem is the strong subadditivity of the entropy proven by block matrix techniques.

Theorem 5 *Let*

$$\mathbf{S} := \begin{bmatrix} S_{11} & S_{12} & S_{13} \\ S_{12}^* & S_{22} & S_{23} \\ S_{13}^* & S_{23}^* & S_{33} \end{bmatrix}$$

be a positive definite block matrix. Then

$$\mathrm{Det}\,\mathbf{S} \times \mathrm{Det}\, S_{22} \le \mathrm{Det}\begin{bmatrix} S_{11} & S_{12} \\ S_{12}^* & S_{22} \end{bmatrix} \times \mathrm{Det}\begin{bmatrix} S_{22} & S_{23} \\ S_{23}^* & S_{33} \end{bmatrix}$$

and the condition for equality is $S_{13} = S_{12}S_{22}^{-1}S_{23}$.

Proof: Let

$$P := \begin{bmatrix} I & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad \text{and} \quad Q := \begin{bmatrix} I & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

From Lemma 4, we have the matrix inequality

$$[P]\mathbf{S} \le [P]Q\mathbf{S}Q$$

which implies the determinant inequality

$$\mathrm{Det}\,[P]\mathbf{S} \le \mathrm{Det}\,[P]Q\mathbf{S}Q\,. \tag{31}$$

According to the Schur determinant formula, this is exactly the determinant inequality of the theorem. The equality in the determinant inequality implies $[P]\mathbf{S} = [P]Q\mathbf{S}Q$ which is

$$S_{11} - [\,S_{12}, \quad S_{13}\,] \begin{bmatrix} S_{22}, & S_{23} \\ S_{32} & S_{33} \end{bmatrix}^{-1} \begin{bmatrix} S_{21} \\ S_{31} \end{bmatrix} = S_{11} - S_{12}S_{22}^{-1}S_{21}.$$

This can be written as

$$[\,S_{12}, \quad S_{13}\,] \left(\begin{bmatrix} S_{22} & S_{23} \\ S_{32} & S_{33} \end{bmatrix}^{-1} - \begin{bmatrix} S_{22}^{-1}, & 0 \\ 0 & 0 \end{bmatrix} \right) \begin{bmatrix} S_{21} \\ S_{31} \end{bmatrix} = 0\,. \tag{32}$$

For a moment, let

$$\begin{bmatrix} S_{22} & S_{23} \\ S_{32} & S_{33} \end{bmatrix}^{-1} = \begin{bmatrix} C_{22}, & C_{23} \\ C_{32} & C_{33} \end{bmatrix}.$$

Then

$$\begin{aligned} \begin{bmatrix} S_{22} & S_{23} \\ S_{32} & S_{33} \end{bmatrix}^{-1} - \begin{bmatrix} S_{22}^{-1} & 0 \\ 0 & 0 \end{bmatrix} &= \begin{bmatrix} C_{23}C_{33}^{-1}C_{32}, & C_{23} \\ C_{32} & C_{33} \end{bmatrix} \\ &= \begin{bmatrix} C_{23}C_{33}^{-1/2} \\ C_{33}^{1/2} \end{bmatrix} [\,C_{33}^{-1/2}C_{32}, \quad C_{33}^{1/2}\,]\,. \end{aligned}$$

Comparing this with (32) we arrive at

$$[\,S_{12}, \quad S_{13}\,] \begin{bmatrix} C_{23}C_{33}^{-1/2} \\ C_{33}^{1/2} \end{bmatrix} = S_{12}C_{23}C_{33}^{-1/2} + S_{13}C_{33}^{1/2} = 0.$$

Equivalently,

$$S_{12}C_{23}C_{33}^{-1} + S_{13} = 0.$$

Since the concrete form of C_{23} and C_{33} is known, we can compute that $C_{23}C_{33}^{-1} = -S_{22}^{-1}S_{23}$ and this gives the condition stated in the theorem. □

References

[1] T.W. Anderson, *An introduction to multivariate statistical analysis*, Second edition. John Wiley & Sons, Inc., New York, 1984

[2] T. Ando, Generalized Schur complements, Linear Algebra Appl. **27**(1979), 173–186.

[3] T. M. Cover and J. A. Thomas, *Elements of information theory*, Wiley, 1991.

[4] W. Feller, *An introduction to probability theory and its applications*, vol. II., John Wiley & Sons, Inc., New York-London-Sydney, 1966.

[5] T. Hida, Canonical representations of Gaussian processes and their applications. Mem. Coll. Sci. Univ. Kyoto. Ser. A. Math. **33**1960/1961, 109–155.

[6] T. Hida and M. Hitsuda, *Gaussian processes.* Translations of Mathematical Monographs, **120**. American Mathematical Society, Providence, RI, 1993.

[7] R. Horn and Ch. Johnson, *Matrix Analysis*, Cambridge Univ. Press 1985.

[8] S. Kullback, *Information theory and statistics*, John Wiley and Sons, New York; Chapman and Hall, Ltd., London, 1959.

[9] D. Petz, *Quantum Information Theory and Quantum Statistics*, Springer, 2008.