Journal of Inequalities in Pure and Applied Mathematics

http://jipam.vu.edu.au/

Volume 4, Issue 4, Article 71, 2003

ON FISHER INFORMATION INEQUALITIES AND SCORE FUNCTIONS IN NON-INVERTIBLE LINEAR SYSTEMS

C. VIGNAT AND J.-F. BERCHER

E.E.C.S. UNIVERSITY OF MICHIGAN
1301 N. BEAL AVENUE
ANN ARBOR MI 48109, USA.
vignat@univ-mlv.fr
URL: http://www-syscom.univ-mlv.fr/ vignat/

ÉQUIPE SIGNAL ET INFORMATION
ESIEE AND UMLV
93 162 NOISY-LE-GRAND, FRANCE.
jf.bercher@esiee.fr

Received 04 June, 2003; accepted 09 September, 2003
Communicated by F. Hansen

ABSTRACT. In this note, we review score functions properties and discuss inequalities on the Fisher Information Matrix of a random vector subjected to linear non-invertible transformations. We give alternate derivations of results previously published in [6] and provide new interpretations of the cases of equality.

Key words and phrases: Fisher information, Non-invertible linear systems.

2000 *Mathematics Subject Classification.* 62B10, 93C05, 94A17.

1. INTRODUCTION

The Fisher information matrix J_X of a random vector X appears as a useful theoretic tool to describe the propagation of information through systems. For instance, it is directly involved in the derivation of the Entropy Power Inequality (EPI), that describes the evolution of the entropy of random vectors submitted to linear transformations. The first results about information transformation were given in the 60's by Blachman [1] and Stam [5]. Later, Papathanasiou [4] derived an important series of Fisher Information Inequalities (FII) with applications to characterization of normality. In [6], Zamir extended the FII to the case of non-invertible linear systems. However, the proofs given in his paper, completed in the technical report [7], involve complicated derivations, especially for the characterization of the cases of equality.

The main contributions of this note are threefold. First, we review some properties of score functions and characterize the estimation of a score function under linear constraint. Second,

ISSN (electronic): 1443-5756

Part of this work was presented at International Symposium on Information Theory, Lausanne, Switzerland, 2002.
077-03

we give two alternate derivations of Zamir's FII inequalities and show how they can be related to Papathanasiou's results. Third, we examine the cases of equality and give an interpretation that highlights the concept of extractable component of the input vector of a linear system, and its relationship with the concepts of pseudoinverse and gaussianity.

2. NOTATIONS AND DEFINITIONS

In this note, we consider a linear system with a $(n \times 1)$ random vector input X and a $(m \times 1)$ random vector output Y, represented by a $m \times n$ matrix A, with $m \leq n$ as

$$Y = AX.$$

Matrix A is assumed to have full row rank (rank $A = m$).

Let f_X and f_Y denote the probability densities of X and Y. The probability density f_X is supposed to satisfy the three regularity conditions (cf. [4])

(1) $f_X(x)$ is continuous and has continuous first and second order partial derivatives,
(2) $f_X(x)$ is defined on $\mathbb{R}^n$ and $\lim_{||x|| \to \infty} f_X(x) = 0$,
(3) the Fisher information matrix J_X (with respect to a translation parameter) is defined as

$$[J_X]_{i,j} = \int_{\mathbb{R}^n} \left[\frac{\partial \ln f_X(x)}{\partial x_i} \frac{\partial \ln f_X(x)}{\partial x_j} \right] f_X(x) \mathrm{d}x,$$

and is supposed nonsingular.

We also define the score functions $\phi_X(\cdot) : \mathbb{R}^n \to \mathbb{R}^n$ associated with f_X according to:

$$\phi_X(x) = \frac{\partial \ln f_X(x)}{\partial x}.$$

The statistical expectation operator E_X is

$$\mathrm{E}_X \left[h(X) \right] = \int_{\mathbb{R}^n} h(x) f_X(x) \mathrm{d}x.$$

$\mathrm{E}_{X,Y}$ and $\mathrm{E}_{X|Y}$ will denote the mutual and conditional expectations, computed with the mutual and conditional probability density functions $f_{X,Y}$ and $f_{X|Y}$ respectively.

The covariance matrix of a random vector $g(X)$ is defined by

$$\mathrm{cov}[g(X)] = \mathrm{E}_X \left[(g(X) - \mathrm{E}_X \left[g(X) \right])(g(X) - \mathrm{E}_X \left[g(X) \right])^T \right].$$

The gradient operator ∇_X is defined by

$$\nabla_X h(X) = \left[\frac{\partial h(X)}{\partial x_1}, \ \ldots, \ \frac{\partial h(X)}{\partial x_n} \right]^T.$$

Finally, in what follows, a matrix inequality such as $A \geq B$ means that matrix $(A - B)$ is nonnegative definite.

3. PRELIMINARY RESULTS

We derive here a first theorem that extends Lemma 1 of [7]. The problem addressed is to find an estimator $\widehat{\phi_X(X)}$ of the score function $\phi_X(X)$ in terms of the observations $Y = AX$. Obviously, this estimator depends of Y, but this dependence is omitted here for notational convenience.

Theorem 3.1. *Under the hypotheses expressed in Section 2, the solution to the minimum mean square estimation problem*

$$\widehat{\phi_X(X)} = \arg\min_w \mathrm{E}_{X,Y}\left[||\phi_X(X) - w(Y)||^2\right] \text{ subject to } Y = AX, \tag{3.1}$$

is

$$\widehat{\phi_X(X)} = A^T \phi_Y(Y). \tag{3.2}$$

The proof we propose here relies on elementary algebraic manipulations according to the rules expressed in the following lemma.

Lemma 3.2. *If X and Y are two random vectors such that $Y = AX$, where A is a full row-rank matrix then for any smooth functions $g_1 : \mathbb{R}^m \to \mathbb{R}$, $g_2 : \mathbb{R}^n \to \mathbb{R}$, $h_1 : \mathbb{R}^n \to \mathbb{R}^n$, $h_2 : \mathbb{R}^m \to \mathbb{R}^m$,*

Rule 0
$$\mathrm{E}_X[g_1(AX)] = \mathrm{E}_Y[g_1(Y)]$$

Rule 1
$$\mathrm{E}_X[\phi_X(X)\, g_2(X)] = -\mathrm{E}_X[\nabla_X g_2(X)]$$

Rule 2
$$\mathrm{E}_X\left[\phi_X(X)\, h_1^T(X)\right] = -\mathrm{E}_X\left[\nabla_X h_1^T(X)\right]$$

Rule 3
$$\nabla_X h_2^T(AX) = A^T \nabla_Y h_2^T(Y)$$

Rule 4
$$\mathrm{E}_X\left[\nabla_X \phi_Y^T(Y)\right] = -A^T J_Y.$$

Proof. Rule 0 is proved in [2, vol. 2, p.133]. Rule 1 and Rule 2 are easily proved using integration by parts. For Rule 3, denote by h_k the k^{th} component of vector $h = h_2$, and remark that

$$\begin{aligned}
\frac{\partial}{\partial x_j} h_k(AX) &= \sum_{i=1}^{m} \frac{\partial h_k(AX)}{\partial y_i} \frac{\partial y_i}{\partial x_j} \\
&= \sum_{i=1}^{m} A_{ij} \left[\nabla_Y h_k(Y)\right]_i \\
&= \left[A^T \nabla_Y h_k(Y)\right]_j .
\end{aligned}$$

Now $h^T(Y) = \left[h_1^T(Y), \ldots, h_n^T(Y)\right]$ so that $\nabla_X h^T(Y) = \left[\nabla_X h_1^T(AX), \ldots, \nabla_X h_n^T(AX)\right] = A^T \nabla_Y h^T(Y)$.

Rule 4 can be deduced as follows:

$$\begin{aligned}
&\mathrm{E}_X\left[\nabla_X \phi_Y^T(Y)\right] \\
&\quad \overset{\text{Rule 3}}{=} A^T \mathrm{E}_X\left[\nabla_Y \phi_Y^T(Y)\right] \\
&\quad \overset{\text{Rule 0}}{=} A^T \mathrm{E}_Y\left[\nabla_Y \phi_Y^T(Y)\right] \\
&\quad \overset{\text{Rule 2}}{=} -A^T \mathrm{E}_Y\left[\phi_Y(Y)\, \phi_Y(Y)^T\right].
\end{aligned}$$

□

For the proof of Theorem 3.1, we will also need the following orthogonality result.

Lemma 3.3. *For all multivariate functions* $h:\mathbb{R}^m \to \mathbb{R}^n$, $\widehat{\phi_X(X)} = A^T\phi_Y(Y)$ *satisfies*

$$\mathrm{E}_{X,Y}\left(\phi_X(X) - \widehat{\phi_X(X)}\right)^T h(Y) = 0. \tag{3.3}$$

Proof. Expand into two terms and compute first term using the trace operator $tr(\cdot)$

$$\begin{aligned}\mathrm{E}_{X,Y}\left[\phi_X(X)^T h(Y)\right] &= tr\,\mathrm{E}_{X,Y}\left[\phi_X(X)\,h^T(Y)\right]\\ &\overset{\text{Rule 2, Rule 0}}{=} -tr\,\mathrm{E}_Y\left[\nabla_X h^T(Y)\right]\\ &\overset{\text{Rule 3}}{=} -tr\ A^T\mathrm{E}_Y\left[\nabla_Y h^T(Y)\right].\end{aligned}$$

Second term writes

$$\begin{aligned}\mathrm{E}_{X,Y}\left[\widehat{\phi_X(X)}^T h(Y)\right] &= tr\,\mathrm{E}_{X,Y}\left[\widehat{\phi_X(X)}h^T(Y)\right]\\ &= tr\,\mathrm{E}_Y\left[A^T\phi_Y(Y)\,h^T(Y)\right]\\ &= tr\,A^T\mathrm{E}_Y\left[\phi_Y(Y)\,h^T(Y)\right]\\ &\overset{\text{Rule 2}}{=} -tr\ A^T\mathrm{E}_Y\left[\nabla_Y h^T(Y)\right]\end{aligned}$$

thus the terms are equal. □

Using Lemma 3.2 and Lemma 3.3 we are now in a position to prove Theorem 3.1.

Proof of Theorem 3.1. From Lemma 3.3, we have

$$\begin{aligned}\mathrm{E}_{X,Y}\left[\left(\phi_X(X) - \widehat{\phi_X(X)}\right)h(Y)\right] &= \mathrm{E}_{X,Y}\left[\left(\phi_X(X) - A^T\phi_Y(Y)\right)h(Y)\right]\\ &= \mathrm{E}_Y\left[\mathrm{E}_{X|Y}\left[\left(\phi_X(X) - A^T\phi_Y(Y)\right)h(Y)\right]\right] = 0.\end{aligned}$$

Since this is true for all h, it means the inner expectation is null, so that

$$\mathrm{E}_{X|Y}\left[\phi_X(X)\right] = A^T\phi_Y(Y).$$

Hence, we deduce that the estimator $\widehat{\phi_X(X)} = A^T\phi_Y(Y)$ is nothing else but the conditional expectation of $\phi_X(X)$ given Y. Since it is well known (see [8] for instance) that the conditional expectation is the solution of the Minimum Mean Square Error (MMSE) estimation problem addressed in Theorem 3.1, the result follows. □

Theorem 3.1 not only restates Zamir's result in terms of an estimation problem, but also extends its conditions of application since our proof does not require, as in [7], the independence of the components of X.

4. FISHER INFORMATION MATRIX INEQUALITIES

As was shown by Zamir [6], the result of Theorem 3.1 may be used to derive the pair of Fisher Information Inequalities stated in the following theorem:

Theorem 4.1. *Under the assumptions of Theorem 3.1,*

$$J_X \geq A^T J_Y A \tag{4.1}$$

and

$$J_Y \leq \left(AJ_X^{-1}A^T\right)^{-1}. \tag{4.2}$$

We exhibit here an extension and two alternate proofs of these results, that do not even rely on Theorem 3.1. The first proof relies on a classical matrix inequality combined with the algebraic properties of score functions as expressed by Rule 1 to Rule 4. The second (partial) proof is deduced as a particular case of results expressed by Papathanasiou [4].

The first proof we propose is based on the well-known result expressed in the following lemma.

Lemma 4.2. *If* $U = \left[\begin{smallmatrix} A & B \\ C & D \end{smallmatrix}\right]$ *is a block symmetric non-negative matrix such that* D^{-1} *exists, then*

$$A - BD^{-1}C \geq 0,$$

with equality if and only if $rank(U) = \dim(D)$.

Proof. Consider the block $L\Delta M$ factorization [3] of matrix U :

$$U = \underbrace{\left[\begin{array}{cc} I & BD^{-1} \\ 0 & I \end{array}\right]}_{L} \underbrace{\left[\begin{array}{cc} A - BD^{-1}C & 0 \\ 0 & D \end{array}\right]}_{\Delta} \underbrace{\left[\begin{array}{cc} I & 0 \\ D^{-1}C & I \end{array}\right]}_{M^T}.$$

We remark that the symmetry of U implies that $L = M$ and thus

$$\Delta = L^{-1} U L^{-T}$$

so that Δ is a symmetric nonnegative definite matrix. Hence, all its principal minors are non-negative, and

$$A - BD^{-1}C \geq 0.$$

□

Using this matrix inequality, we can complete the proof of Theorem 4.1 by considering the two following $(m+n) \times (m+n)$ matrices

$$U_1 = \mathrm{E}\left[\begin{array}{c} \phi_X(X) \\ \phi_Y(Y) \end{array}\right] \left[\begin{array}{cc} \phi_X^T(X) & \phi_Y^T(Y) \end{array}\right],$$

$$U_2 = \mathrm{E}\left[\begin{array}{c} \phi_Y(Y) \\ \phi_X(X) \end{array}\right] \left[\begin{array}{cc} \phi_Y^T(Y) & \phi_X^T(X) \end{array}\right].$$

For matrix U_1, we have, from Lemma 4.2

$$\begin{aligned} &\mathrm{E}_X\left[\phi_X(X)\,\phi_X^T(X)\right] \\ &\geq \mathrm{E}_{X,Y}\left[\phi_X(X)\,\phi_Y^T(Y)\right]\left(\mathrm{E}_Y\left[\phi_Y(Y)\,\phi_Y^T(Y)\right]\right)^{-1}\mathrm{E}_{X,Y}\left[\phi_Y(Y)\,\phi_X^T(X)\right]. \end{aligned} \tag{4.3}$$

Then, using the rules of Lemma 3.2, we can recognize that

$$\begin{aligned} \mathrm{E}_X\left[\phi_X(X)\,\phi_X^T(X)\right] &= J_X, \\ \mathrm{E}_Y\left[\phi_Y(Y)\,\phi_Y^T(Y)\right] &= J_Y, \\ \mathrm{E}_{X,Y}\left[\phi_X(X)\,\phi_Y^T(Y)\right] &= -\mathrm{E}_Y\left[\nabla\phi_Y^T(Y)\right] = A^T J_Y, \\ \mathrm{E}_{X,Y}\left[\phi_Y(Y)\,\phi_X^T(X)\right] &= \left(A^T J_Y\right)^T = J_Y A. \end{aligned}$$

Replacing these expressions in inequality (4.3), we deduce the first inequality (4.1).

Applying the result of Lemma 4.2 to matrix U_2 yields similarly

$$J_Y \geq J_Y^T A J_X^{-1} A^T J_Y.$$

Multiplying both on left and right by $J_Y^{-1} = \left(J_Y^{-1}\right)^T$ yields inequality (4.2).

Another proof of inequality (4.2) is now exhibited, as a consequence of a general result derived by Papathanasiou [4]. This result states as follows.

Theorem 4.3. *(Papathanasiou* [4]*) If $g(X)$ is a function $\mathbb{R}^n \to \mathbb{R}^m$ such that, $\forall i \in [1, m]$, $g_i(x)$ is differentiable and $var[g_i(X)] \leq \infty$, the covariance matrix $cov[g(X)]$ of $g(X)$ verifies:*

$$\operatorname{cov}[g(X)] \geq \mathrm{E}_X\left[\nabla^T g(X)\right] J_X^{-1} \mathrm{E}_X\left[\nabla g(X)\right].$$

Now, inequality (4.2) simply results from the choice $g(X) = \phi_Y(AX)$, since in this case $\operatorname{cov}[g(X)] = J_Y$ and $\mathrm{E}_X\left[\nabla^T g(X)\right] = -J_Y A$. Note that Papathanasiou's theorem does not allow us to retrieve inequality (4.1).

5. Case of Equality in Matrix FII

We now explicit the cases of equality in both inequalities (4.1) and (4.2). Case of equality in inequality (4.2) was already characterized in [7] and introduces the notion of 'extractable components' of vector X. Our alternate proof also makes use of this notion and establishes a link with the pseudoinverse of matrix A.

Case of equality in inequality (4.1). The case of equality in inequality (4.1) is characterized by the following theorem.

Theorem 5.1. *Suppose that components X_i of X are mutually independent. Then equality holds in (4.1) if and only if matrix A possesses $(n-m)$ null columns or, equivalently, if A writes, up to a permutation of its column vectors*

$$A = [A_0 \mid 0_{m\times(n-m)}],$$

where A_0 is a $m \times m$ non-singular matrix.

Proof. According to the first proof of Theorem 4.1 and the case of equality in Lemma 4.2, equality holds in (4.1) if there exists a non-random matrix B and a non-random vector c such that

$$\phi_X(X) = B\phi_Y(Y) + c.$$

However, as random variables $\phi_X(X)$ and $\phi_Y(Y)$ have zero-mean, $\mathrm{E}_X[\phi(X)] = 0$, $\mathrm{E}_Y[\phi(Y)] = 0$, then necessarily $c = 0$. Moreover, applying Rule 2 and Rule 4 yields

$$\mathrm{E}_{X,Y}\left[\phi_X(X)\,\phi_Y(Y)^T\right] = A^T J_Y$$

on one side, and

$$\mathrm{E}_{X,Y}\left[\phi_X(X)\,\phi_Y(Y)^T\right] = BJ_Y$$

on the other side, so that finally $B = A^T$ and

$$\phi_X(X) = A^T\phi_Y(Y).$$

Now, since A has rank m, it can be written, up to a permutation of its columns, under the form

$$A = [A_0 \mid A_0 M],$$

where A_0 is an invertible $m \times m$ matrix, and M is an $m \times (n-m)$ matrix. Suppose $M \neq 0$ and express equivalently X as

$$X = \begin{bmatrix} X_0 \\ X_1 \end{bmatrix} \begin{matrix} \}m \\ \}n-m \end{matrix}$$

so that

$$\begin{aligned} Y &= AX \\ &= A_0X_0 + A_0MX_1 \\ &= A_0\tilde{X}, \end{aligned}$$

with $\tilde{X} = X_0 + MX_1$. Since A_0 is square and invertible, it follows that

$$\phi_Y(Y) = A_0^{-T}\phi_{\tilde{X}}\left(\tilde{X}\right)$$

so that

$$\begin{aligned}
\phi_X &= A^T\phi_Y(Y) \\
&= A^T A_0^{-T}\phi_{\tilde{X}}\left(\tilde{X}\right) \\
&= \begin{bmatrix} A_0^T \\ M^T A_0^T \end{bmatrix} A_0^{-T}\phi_{\tilde{X}}\left(\tilde{X}\right) \\
&= \begin{bmatrix} I \\ M^T \end{bmatrix} \phi_{\tilde{X}}\left(\tilde{X}\right) \\
&= \begin{bmatrix} \phi_{\tilde{X}}\left(\tilde{X}\right) \\ M^T\phi_{\tilde{X}}\left(\tilde{X}\right) \end{bmatrix}.
\end{aligned}$$

As X has independent components, ϕ_X can be decomposed as

$$\phi_X = \begin{bmatrix} \phi_{X_0}(X_0) \\ \phi_{X_1}(X_1) \end{bmatrix}$$

so that finally

$$\begin{bmatrix} \phi_{X_0}(X_0) \\ \phi_{X_1}(X_1) \end{bmatrix} = \begin{bmatrix} \phi_{\tilde{X}}\left(\tilde{X}\right) \\ M^T\phi_{\tilde{X}}\left(\tilde{X}\right) \end{bmatrix},$$

from which we deduce that

$$\phi_{X_1}(X_1) = M^T\phi_{X_0}(X_0).$$

As X_0 and X_1 are independent, this is not possible unless $M = 0$, which is the equality condition expressed in Theorem 5.1.

Reciprocally, if these conditions are met, then obviously, equality is reached in inequality (4.1). □

Case of equality in inequality (4.2). Assuming that components of X are mutually independent, the case of equality in inequality (4.2) is characterized as follows:

Theorem 5.2. *Equality holds in inequality (4.2) if and only if each component X_i of X verifies at least one of the following conditions*

a) *X_i is Gaussian,*
b) *X_i can be recovered from the observation of $Y = AX$, i.e. X_i is 'extractable',*
c) *X_i corresponds to a null column of A.*

Proof. According to the (first) proof of inequality (4.2), equality holds, as previously, if and only if there exists a matrix C such that

$$\phi_Y(Y) = C\phi_X(X), \tag{5.1}$$

which implies that $J_Y = CJ_XC^t$. Then, as by assumption $J_Y^{-1} = AJ_X^{-1}A^t$, $C = J_YAJ_X^{-1}$ is such a matrix. Denoting $\tilde{\phi}_X(X) = J_X^{-1}\phi_X(X)$ and $\tilde{\phi}_Y(Y) = J_Y^{-1}\phi_Y(Y)$, equality (5.1) writes

$$\tilde{\phi}_Y(Y) = A\tilde{\phi}_X(X). \tag{5.2}$$

The rest of the proof relies on the following two well-known results:

- if X is Gaussian then equality holds in inequality (4.2),

- if A is a non singular square matrix, equality holds in inequality (4.2) irrespectively of X.

We thus need to isolate the 'invertible part' of matrix A. In this aim, we consider the pseudoinverse $A^{\#}$ of A and form the product $A^{\#}A$. This matrix writes, up to a permutation of rows and columns

$$A^{\#}A = \begin{bmatrix} I & 0 & 0 \\ 0 & M & 0 \\ 0 & 0 & \mathbf{0} \end{bmatrix},$$

where I is the $n_i \times n_i$ identity, M is a $n_{ni} \times n_{ni}$ matrix and $\mathbf{0}$ is a $n_z \times n_z$ matrix with $n_z = n - n_i - n_{ni}$ (i stands for invertible, n_i for not invertible and z for zero). Remark that n_z is exactly the number of null columns of A. Following [6, 7], n_i is the number of 'extractable' components, that is the number of components of X that can be deduced from the observation $Y = AX$. We provide here an alternate characterization of n_i as follows: the set of solutions of $Y = AX$ is an affine set

$$X = A^{\#}Y + (I - A^{\#}A)Z = X_0 + (I - A^{\#}A)Z,$$

where X_0 is the minimum norm solution of the linear system $Y = AX$ and Z is any vector. Thus, n_i is exactly the number of components shared by X and X_0.

The expression of $A^{\#}A$ allows us to express $\mathbb{R}^n$ as the direct sum $\mathbb{R}^n = \mathbb{R}^i \oplus \mathbb{R}^{ni} \oplus \mathbb{R}^z$, and to express accordingly X as $X = \left[X_i^T, X_{ni}^T, X_z^T\right]^T$. Then equality in inequality (4.2) can be studied separately in the three subspaces as follows:

(1) restricted to subspace $\mathbb{R}^i$, A is an invertible operator, and thus equality holds without condition,
(2) restricted to subspace $\mathbb{R}^{ni}$, equality (5.2) writes $M\tilde{\phi}(X_{ni}) = \tilde{\phi}(MX_{ni})$ that means that necessarily all components of X_{ni} are gaussian,
(3) restricted to subspace $\mathbb{R}^z$, equality holds without condition.

□

As a final note, remark that, although A is supposed full rank, $n_i \leq \operatorname{rank} A$. For instance, consider matrix

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix}$$

for which $n_i = 1$ and $n_{ni} = 2$. This example shows that the notion of 'extractability' should not be confused with the invertibility restricted to a subspace. A is clearly invertible in the subspace $x_3 = 0$. However, such a subspace is irrelevant here since, as we deal with continuous random input vectors, X has a null probability to belong to this subspace.

ACKNOWLEDGMENT

The authors wish to acknowledge Pr. Alfred O. Hero III at EECS for useful discussions and suggestions, particularly regarding Theorem 3.1. The authors also wish to thank the referee for his careful reading of the manuscript that helped to improve its quality.

REFERENCES

[1] N.M. BLACHMAN, The convolution inequality for entropy powers, *IEEE Trans. on Information Theory, IT* , **11** (1965), 267–271.

[2] W. FELLER, *Introduction to Probability Theory and its Applications*, New York, John Wiley & Sons, 1971.

[3] G. GOLUB, *Matrix Computations*, Johns Hopkins University Press, 1996.

[4] V. PAPATHANASIOU, Some characteristic properties of the Fisher information matrix via Cacoullos-type inequalities, *J. Multivariate Analysis*, **14** (1993), 256–265.

[5] A.J. STAM, Some inequalities satisfied by the quantities of information of Fisher and Shannon, *Inform. Control*, **2** (1959), 101–112.

[6] R. ZAMIR, A proof of the Fisher information matrix inequality via a data processing argument, *IEEE Trans. on Information Theory, IT,* **44**(3) (1998), 1246–1250.

[7] R. ZAMIR, A necessary and sufficient condition for equality in the Matrix Fisher Information inequality, *Technical report, Tel Aviv University, Dept. Elec. Eng. Syst.*, 1997. Available online `http://www.eng.tau.ac.il/~zamir/techreport/crb.ps.gz`

[8] H.L. van TREES, *Detection, Estimation, and Modulation Theory,* Part I, New York London, John Wiley and Sons, 1968.