Matrix Fisher inequalities for non-invertible linear systems

C. Vignat[1] and J.-F. Bercher[1,2]

(1)
Laboratoire Systèmes de Communications
Université de Marne la Vallée
93 166 Noisy-le-Grand, FRANCE
vignat@univ-mlv.fr

(2)
Laboratoire Signaux et Télécoms,
Groupe ESIEE
93 162 Noisy-le-Grand, FRANCE
bercherj@esiee.fr

1 Introduction

The Fisher informatin matrix I_X of a random vector X appears as useful information theoretic tool to describe the proppagation 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 variables (vectors) submitted to linear transformations. The first results about information transformation were given in the 60's by Blachman [1] and Stam [2]. Later, Papathanasiou [3] derived an important serie of Fisher Information Inequalities (FII) with applications to characterization of normality. In a fascinating contribution, Zamir [4] extended the FII to the case of non-invertible linear systems. He also pointed that such inequalities may have interesting applications in areas such as (blind) deconvolution and sources separation. However, the proofs given in his paper, completed in the technical report [5], involve complicated derivations, especially for the characterization of the cases of equality.

The main contributions of the paper are twofold: first we give two alternate derivations of Zamir's FII inequalitiesand show that how they can be related to Papathanasiou's results. Second, we examine the case of equality and give an interpretation that highlights the concept of extractible component of the input vector of a linear system, together with its relationship with the concepts of pseudoinverse and gaussianity.

2 A new derivation of Zamir's Fisher Information Inequalities

2.1 Notations and definitions

In this paper, 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, 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 suposed to satisfy the three regularity conditions (cf. [3])

1. $f_X(x)$ is continuous and has continuous first and second order partial derivatives,
2. $f_X(x)$ is defined on $\mathbb{R}$ 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

$$[J_X]_{i,j} = E_X\left[\frac{\partial \ln f_X(x)}{\partial x_i}\frac{\partial \ln f_X(x)}{\partial x_j}\right],$$

and is supposed non singular.

Finally, we denote $\phi_X(.)$ and $\phi_Y(.)$ the score (log derivative) functions associated with f_X and f_Y.

2.2 A first result

We derive here a first theorem that extends [5, Lemma 1].

Theorem 1 *Under the hypotheses expressed in 2.1, the best estimate (in the minimum mean square error sense) of $\phi_X(x)$ from observations Y is*

$$\hat{\phi}_X(X) = A^T \phi_Y(Y) \tag{1}$$

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

Lemma 2 *If X and Y are two random vectors such that $Y = AX$ where A is a $m \times n$ full row-rank matrix then for all scalar valued function g*

$$E_X g(AX) = E_Y g(Y) \tag{2}$$

Proof. *see [6] vol. 2 p.133* ■

Lemma 3 *Let X and Y be two random vectors such that $Y = AX$ where A is a $m \times n$ full row-rank matrix. Let us denote J_X and J_Y the Fisher information matrix of X and Y respectively. Then for any scalar valued function g and any vector valued function h,*

$$E_X \phi_X(X) g(X) = -E_X \nabla_X g(X) \tag{rule 1}$$

$$E_X \phi_X(X) h^T(X) = -E_X \nabla_X h^T(X) \tag{rule 2}$$

$$\nabla_X h^T(AX) = A^T \nabla_Y h^T(Y) \tag{rule 3}$$

$$E_X \nabla_X \phi_Y^T(Y) = -A^T J_Y \tag{rule 4}$$

The proofs of these algebraic rules are given in Appendix A. They allow to prove the theorem via the following lemma (see proof in Appendix B):

Lemma 4 *For all multivariate function $h : \mathbb{R}^m \to \mathbb{R}^n$,*

$$E_X \left(\phi_X(X) - \hat{\phi}_X(Y) \right)^T h(Y) = 0 \tag{3}$$

where $\hat{\phi}_X(Y) = A^T \phi_Y(Y)$.

Finally, the proof of theorem follows directly from Lemma 4 and from the classical projection result of Minimum Mean Square Error (MMSE) estimation that expresses the best (MSE) estimator of $\phi_X(X)$ as orthogonal to any arbitrary function $h(Y)$ of the observations.

Observe that the theorem 1 extends the corresponding lemma in Zamir's paper where the components of X are supposed independent.

3 Application to Zamir's FII

As was shown by Zamir [4], the result of theorem 1 may be used to derive a very interesting Fisher Information Inequality. We exhibit here an extension and two alternate proofs of these results: the first proof relies on a classical atrix inequality combined with the algebraic properties of the score functions as expressed by rules (rule1) to (rule4). The second proof is deduced as a particular case of results expressed by Papathanasiou [3]. The theorem is the following.

Theorem 5 *Under the assumptions of theorem 1,*

$$J_X \geq A^T J_Y A \tag{4}$$

and

$$J_Y \leq \left(A J_X^{-1} A^T \right)^{-1} \tag{5}$$

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

Lemma 6 *If $U = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$ is a block symmetric non-negative matrix such that D^{-1} exists, then*

$$A - BD^{-1}C \geq 0, \tag{6}$$

(in the sense that $A - BD^{-1}C$ is non-negative definite), with equality if and only if $rank(U) = \dim(D)$

Proof. (of the inequality). Consider the block LΔM factorization [7] of matrix U :

$$U = \underbrace{\begin{bmatrix} I & BD^{-1} \\ 0 & I \end{bmatrix}}_{L} \underbrace{\begin{bmatrix} A - BD^{-1}C & 0 \\ 0 & D \end{bmatrix}}_{\Delta} \underbrace{\begin{bmatrix} I & 0 \\ D^{-1}C & I \end{bmatrix}}_{M^T} \tag{7}$$

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

$$\Delta = L^{-1} U L^{-T} \tag{8}$$

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

$$A - BD^{-1}C \geq 0 \tag{9}$$

■ Using this matrix inequality, we can complete the proof of theorem 5 by considering the two following matrices

$$U_1 = E \begin{bmatrix} \phi_X(X) \\ \phi_Y(Y) \end{bmatrix} \begin{bmatrix} \phi_X^T(X) & \phi_Y^T(Y) \end{bmatrix} \tag{10}$$

$$U_2 = E \begin{bmatrix} \phi_Y(Y) \\ \phi_X(X) \end{bmatrix} \begin{bmatrix} \phi_Y^T(Y) & \phi_X^T(X) \end{bmatrix} \tag{11}$$

Applying the result of lemma 6 and of the rules of 3, to matrices U_1 and U_2 yields straigthforwardly inequalities (4) and (5).

Second proof of (5)

A second proof of inequality 5 is now exhibited, as a consequence of a general result derived by Papathanasiou [3]. This result states as follows.

Theorem 7 *(Papathanasiou [3]) If $g(X)$ is a function $\mathbb{R}^n \rightarrow \mathbb{R}^m$ such that, $\forall 1 \leq i \leq m$, $g_i(x)$ is differentiable and var[$g_i(X)$]$\leq \infty$. The covariance matrix cov[$g(X)$] of $g(X)$ verifies:*

$$\mathrm{cov}[g(X)] \geq E[\nabla^t g(X)] J_X^{-1} E[\nabla g(X)].$$

Now, inequality 5 simply results from the choice $g(X) = \phi_Y(AX)$, since in this case $\mathrm{cov}[g(X)] = J_Y$ and $E[\nabla^t g(X)] = -J_Y A$.

4 Case of equality in Zamir's FII

4.1 Introduction

We now explicit the cases of equality in both inequalities (4) and (5). Case of equality in 5 was already chereacterized in [5] and introduces the notion of 'extractible' components of vector X. Our alternate proof also make use of this notion and establish a link with the pseudo inverse of operator A.

4.2 Case of equality in inequality (4)

The case of equality in inequality (4) is characterized by the following theorem.

Theorem 8 *Suppose that the components X_i of X are mutually independent. Then equality holds in (4) if and only of matrix A possesses $(n-m)$ null columns or, equivalently, if A writes, up to a permutation of its column vectors*

$$A = [A_0|0],$$

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

Proof. According to the first proof of theorem 5 and the case of equality in lemma 6, equality holds in (4) if there exists a matrix B and a constant vector c such that

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

where B and c are a constant matrix and a constant vector respectively. However, as the random variables $\phi_X(X)$ and $\phi_Y(Y)$ are zero-mean, then necessarily $c = 0$. Moreover, applying rules 2 and 4

$$E\phi_X(X)\,\phi_Y(Y)^T = A^T J_Y$$

on one side, and

$$E\phi_X(X)\,\phi_Y(Y)^T = B J_Y$$

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

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

Now, since A has rank m, it can be written under the following form

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

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

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

so that

$$\begin{aligned} Y &= AX && (12) \\ &= A_0 X_0 + A_0 M X_1 && (13) \\ &= A_0 \tilde{X} && (14) \end{aligned}$$

with $\tilde{X} = X_0 + MX_1$. It is easy to check that

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

so that

$$\begin{aligned} \phi_X &= A^T \phi_Y(Y) && (15) \\ &= A^T A_0^{-T} \phi_{\tilde{X}}\left(\tilde{X}\right) && (16) \\ &= \begin{bmatrix} A_0^T \\ M^T A_0^T \end{bmatrix} A_0^{-T} \phi_{\tilde{X}}\left(\tilde{X}\right) && (17) \\ &= \begin{bmatrix} I \\ M^T \end{bmatrix} \phi_{\tilde{X}}\left(\tilde{X}\right) && (18) \\ &= \begin{bmatrix} \phi_{\tilde{X}}\left(\tilde{X}\right) \\ M^T \phi_{\tilde{X}}\left(\tilde{X}\right) \end{bmatrix} && (19) \end{aligned}$$

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 finaly

$$\left[\begin{array}{c}\phi_{X_0}(X_0)\\ \phi_{X_1}(X_1)\end{array}\right]=\left[\begin{array}{c}\phi_{\tilde{X}}\left(\tilde{X}\right)\\ M^T\phi_{\tilde{X}}\left(\tilde{X}\right)\end{array}\right]$$

from what 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 $M=0$, what is the equality condition expressed in theorem 8.

Reciprocally, if these conditions are met, then obviously, equality is reached in inequality (4). ■

4.3 Case of equality in inequality (5)

The case of equality in (5), in the case of independent components, is characterized as follows:

Theorem 9 *Equality holds in (5) if and only if each component X_i of X belongs to one of the three groups*

a *X_i is gaussian*

b *X_i can be recovered from the observation of $Y=AX$, i.e. X_i is 'extractible'*

c *X_i correspond to a null column of A.*

Proof. According to the (first) proof of (5), equality holds, as previously, if and only if there exists a matrix C such that

$$\phi_Y(Y)=C\phi_X(X) \qquad (20)$$

Remark that then $J_Y=CJ_XC^t$, so that, as $J_Y^{-1}=AJ_X^{-1}A^t$, and $C=J_YAJ_X^{-1}$ is such a matrix. Then denoting $\tilde{\phi_X}(X)=J_X^{-1}\phi_X(X)$ and $\tilde{\phi_Y}(Y)=J_Y^{-1}\phi_Y(Y)$, equality (20) writes

$$\tilde{\phi_Y}(Y)=A\tilde{\phi_X}(X). \qquad (21)$$

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

- if X is Gaussian then equality holds in (5),
- if A is a non singular square matrix, equality holds in (5) irrespectively of X.

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

$$A^{\#}A=\left[\begin{array}{ccc}I&0&0\\0&M&0\\0&0&0\end{array}\right]$$

where I is the $n_i\times n_i$ identity, M is a $n_{ni}\times n_{ni}$ matrix and 0 is a $n_z\times n_z$ matrix with $n_z=n-n_i-n_{ni}$ (i for invertible, ni for not invertible and z for zero). Remark that n_z is exactly the number of null columns of A. Following [4, 5], n_i is the number of 'extractable' components, that is the number of components of X that can be deduced from the observation $Y=AX$. This number can be characterized as follows: there always exists a vector Z such that

$$X=A^{\#}Y+(1-A^{\#}A)Z=X_0+(1-A^{\#}A)Z.$$

Here X_0 is the minimum norm solution of the linear system $Y=AX$. Thus, n_i is exactly the number of components shared by X and X_0. [1]

The expression of $A^{\#}A$ allows to decompose $\mathbb{R}^n$ as the direct sum $\mathbb{R}^n=\mathbb{R}^i+\mathbb{R}^{ni}+\mathbb{R}^z$, and to decompose accordingly X as $X=\left[X_i^T,X_{ni}^T,X_z^T\right]^T$. Then equality in (5) can be studied separately in the three subspaces as follows:

1. in R^i, A is an invertible operator, and thus equality holds without condition
2. in R^{ni}, equality (21) writes $M\tilde{\phi}(X_{ni})=\tilde{\phi}(MX_{ni})$ that means that necessarily all components of X_{ni} are gaussian
3. in R^z, equality holds without condition.

■

[1] On the invertibility. Remark that, although A is supposed full rank, $n_i\le\mathrm{rank}A$. For instance, consider matrix

$$A=\left[\begin{array}{ccc}1&0&0\\0&1&1\end{array}\right]$$

for which $n_i=1$ and $n_ni=2$. This example shows that the notion of extractability as defined by Zamir should not be confused with the invertibility restricted to a subspace. In the previous example, A is clearly invertible in the subspace $x_3=0$. However, such subspace is irrelevant here since we deal with random input vectors, excluding the case of a deterministic component. We may also note that any X in span A^t (the subspace generated by the columns of A^t) is such that $X=A^{\#}Y$, and thus 'invertible' ; but in this case the components of X in this subspace are not necessarily independent, that is an hypothesis in the theorem.

References

[1] N. M. Blachman, “The convolution inequality for entropy powers”, IEEE trans. on Information Theory, IT 11, pp. 267-271, 1965.

[2] A. J. Stam, “Some inequalities satisfied by the quantities of information of Fisher and Shannon”, Inform. Control, 2, pp. 101-112, june 1959.

[3] V. Papathanasiou, “Some Characteristic Properties of the Fisher Information Matrix via Cacoullos-type Inequalities”, Journal of Multivariate analysis, 14, pp. 256-265, 1993.

[4] R. Zamir, “A Proof of the Fisher Information Matrix via a Data Processing Argument”, IEEE trans. on Information Theory, IT 44, 3, pp. 1246-1250, 1998.

[5] 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 .

[6] W. Feller, “Introduction to probability theory and its applications”, New York, Wiley.

[7] G. Golub, ”Matrix Computations”, Johns Hopkins University Press, 1996

5 Appendix A – Proof of Lemma 3

Proof. we prove only rules 3 and 4. Consider component k of vector h, namely h_k and remark that

$$\frac{\partial}{\partial x_j} h_k(AX) = \sum_{i=1}^{m} \frac{\partial h_k(AX)}{\partial y_i} \frac{\partial y_i}{\partial x_j} \tag{22}$$

$$= \sum_{i=1}^{m} A_{ij} \left(\nabla_Y h_k(AX)\right)_i \tag{23}$$

$$= \left(A^T \nabla_Y h_k(AX)\right)_j \tag{24}$$

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

Rule 4 can be deduced as follows:

$$E_X \nabla_X \phi_Y^T(Y) \tag{25}$$

$$\overset{\text{rule 3}}{=} A^T E_X \nabla_Y \phi_Y^T(Y) \tag{26}$$

$$\overset{\text{lemma 2}}{=} A^T E_Y \nabla_Y \phi_Y^T(Y) \tag{27}$$

$$\overset{\text{rule 2}}{=} -A^T E_Y \phi_Y(Y) \phi_Y(Y)^T \tag{28}$$

■

6 Appendix B – Proof of Lemma 4

Proof. develop into two terms and compute first term as

$$E_X \phi_X(X)^T h(Y) \tag{29}$$

$$= tr\ E_X \phi_X(X) h^T(Y) \tag{30}$$

$$\overset{\text{rule 2}}{=} -tr\ E_X \nabla_X h^T(Y) \tag{31}$$

$$\overset{\text{rule 3}}{=} -tr\ A^T E_X \nabla_Y h^T(Y) \tag{32}$$

Second term writes

$$E_X \hat{\phi}_X(Y)^T h(Y) \tag{33}$$

$$= tr\ E_X \hat{\phi}_X(Y) h^T(Y) \tag{34}$$

$$= tr\ E_X A^T \phi_Y(Y) h^T(Y) \tag{35}$$

$$= tr\ A^T E_Y \phi_Y(Y) h^T(Y) \tag{36}$$

$$\overset{\text{rule 2}}{=} -tr\ A^T E_Y \nabla_Y h^T(Y) \tag{37}$$

$$\overset{\text{lemma 2}}{=} -tr\ A^T E_X \nabla_Y h^T(Y) \tag{38}$$

thus both terms are equal. ■

7 Appendix C – First proof of theorem 5

Applying the result of lemma 6 and of the rules of 3, to matrices U_1 and U_2 yields straigthforwardly inequalities (4) and (5).

For matrix U_1, we thus have

$$E\phi_X(X)\phi_X^T(X) \geq E\phi_X(X)\phi_Y^T(Y) \tag{39}$$

$$\left(E\phi_Y(Y)\phi_Y^T(Y)\right)^{-1} E\phi_Y(Y)\phi_X^T(X) \tag{40}$$

and recognizing $E\phi_X(X)\phi_X^T(X) = J_X$, $E\phi_Y(Y)\phi_Y^T Y = J_Y$ and

$$E\phi_X(X)\phi_Y^T(Y) = -E\nabla\phi_Y^T(Y) = A^T J_Y \tag{41}$$

$$E\phi_Y(Y)\phi_X^T(X) = \left(A^T J_Y\right)^T = J_Y A \tag{42}$$

Replacing these expressions in (39), we deduce the first inequality (4).

Applying the result of the lemma to matrix U_2 yields similarly

$$J_Y \geq J_Y^T A J_X^{-1} A^T J_Y \tag{43}$$

Multiplying both on left and right by $J_Y^{-1} = \left(J_Y^{-1}\right)^T$ yields inequality (5).