A Lower Bound for the Fisher Information Measure

Manuel Stein, Amine Mezghani, and Josef A. Nossek

Abstract—The problem how to approximately determine the value of the Fisher information measure for a general parametric probabilistic system is considered. Having available the first and second moment of the system output in a parametric form, it is shown that the information measure can be bounded from below through a replacement of the original system by a Gaussian system with equivalent moments. The presented technique is applied to a system of practical importance and the potential quality of the bound is demonstrated.

Index Terms—Estimation theory, minimum Fisher information, non-linear systems.

I. INTRODUCTION

ASSESSING the quality of a given parametric probabilistic system with respect to the inference of it's unknown parameters plays a key role in the analysis and design of signal processing systems. Through the Fisher information measure [1][2], estimation theory provides a method to characterize the capability of such systems with respect to this problem in a compact way. As the information measure is inversely proportional to the minimum achievable variance with unbiased estimators and exhibits desirable properties like convexity [3] and [4], it can be used for performance characterization or as a figure of merit when aiming at the optimum system design. However, the application of such rigorous tools to practical problems requires a probabilistic model which accurately captures all underlying effects within the system of interest. Therefore, a crucial point during the conceptual transition from real-world physical systems to theoretical mathematical models is to find a correct analytical representation of their probabilistic nature, i.e. to determine how uncertainty impacts these entities. As superposition characterizes in an appropriate way how noise enters into different practical systems, additive models have become a widely-used assumption among engineers and researchers. If the noise is caused by a sufficiently large number of sources, the central limit theorem can be applied in order to justify an independent additive noise model with Gaussian distribution. Otherwise, using the fact that among all additive noise distributions with fixed variance, the Gaussian one minimizes Fisher information [5]–[7], allows to assess the quality of any probabilistic system with independent additive noise in a conservative way.

A. Motivation

While mathematical tractability makes models with independent and additive noise attractive, such an assumption must be questioned if one takes into account that in general probabilistic systems follow non-additive laws. This becomes relevant in practice when considering that a variety of technical systems exhibit severe non-linear characteristics behind their dominant noise sources. Therefore, additive independent noise does not provide an accurate and generic mathematical model for the input-output relation of physical systems. In order to contribute to the scientific understanding of this general class of systems, here we show how to approximately validate the Fisher information measure for an arbitrary parametric probabilistic model in a simple way.

B. Related Works

An early discussion about the least favorable distributions with respect to location and scale estimation problems is found in [8]. For the generalized form of the Fisher information measure of order s, report [5, pp. 73 ff.] shows that for models with location parameter, the exponential power distribution with fixed s-th moment attains minimum Fisher information and cites [3] for a proof of the standard case $s = 2$. The work [9] focuses on minimum Fisher information under fixed higher order moments. The articles [10] and [11] analyze the problem under a restriction on the support of the system output. A recent lecture note concerned with minimum Fisher information under a univariate model with independent and additive noise is provided by [6], while [7] generalizes the result to multivariate problems, i.e. models with independent and additive correlated noise.

C. Contribution

While the discussion has basically focused on determining the additive noise distribution minimizing the Fisher information under a fixed second moment, to the best of our knowledge, none of the previous works provide a bounding approach for the Fisher information under general models including non-additive systems. The strength of the method presented here lies in it's generality, providing a bound on the Fisher information measure under any probabilistic model and it's simplicity, being based exclusively on the dependency between the first two moments of the system output and the model parameters. This can be useful when analyzing the effects of non-linearities inherent in signal processing systems. The presented bounding technique can be interpreted as a replacement of the original system by an appropriate Gaussian model with equivalent moments and therefore allows to use well-known Gaussian expressions in the estimation theoretic analysis of non-additive parametric probabilistic systems.

II. SYSTEM MODEL

For the discussion, we assume access to the random output $y \in \mathcal{Y}$ of a parametric probabilistic system $p_y(y; \theta)$ and it's deterministic parameter $\theta \in \Theta$, where $\mathcal{Y} \subseteq \mathbb{R}$ denotes the support of the output y and $\Theta \subseteq \mathbb{R}$ is the parameter set. Further, the first and the second central output moment are available in the form

$$\mu(\theta) = \int_{\mathcal{Y}} y p_y(y; \theta) \mathrm{d}y, \tag{1}$$

$$\sigma^2(\theta) = \int_{\mathcal{Y}} (y - \mu(\theta))^2 p_y(y; \theta) \mathrm{d}y. \tag{2}$$

III. FISHER INFORMATION BOUND - UNIVARIATE CASE

Given the possibility to observe the system output y, following the parameterized probability density function $p_y(y; \theta)$, the Fisher information measure [12]

$$F(\theta) = \int_{\mathcal{Y}} p_y(y; \theta) \left(\frac{\partial \log p_y(y; \theta)}{\partial \theta} \right)^2 \mathrm{d}y \tag{3}$$

is associated with the information carried by realizations of the output y for inference of the unknown parameter θ since the minimum estimation error which is achievable with an unbiased estimate $\hat{\theta}(y)$ is restricted to

$$E_y \left[(\hat{\theta}(y) - \theta)^2 \right] \geq \frac{1}{F(\theta)}. \tag{4}$$

The goal here is to show that if the derivative $\frac{\partial \mu(\theta)}{\partial \theta}$ exists, the Fisher information measure can be lower bounded by

$$F(\theta) \geq \frac{1}{\sigma^2(\theta)} \left(\frac{\partial \mu(\theta)}{\partial \theta} \right)^2. \tag{5}$$

A. Proof

Starting from the definition of the information measure (3) and using the Cauchy–Schwarz inequality (45),

$$F(\theta) = \int_{\mathcal{Y}} \left(\frac{\partial \log p_y(y; \theta)}{\partial \theta} \right)^2 p_y(y; \theta) \mathrm{d}y$$

$$\geq \frac{\left(\int_{\mathcal{Y}} (y - \mu(\theta)) \frac{\partial \log p_y(y; \theta)}{\partial \theta} p_y(y; \theta) dy \right)^2}{\int_{\mathcal{Y}} (y - \mu(\theta))^2 p_y(y; \theta) \mathrm{d}y}$$

$$= \frac{\left(\int_{\mathcal{Y}} (y - \mu(\theta)) \frac{\partial p_y(y; \theta)}{\partial \theta} dy \right)^2}{\sigma^2(\theta)}$$

$$= \frac{\left(\int_{\mathcal{Y}} y \frac{\partial p_y(y; \theta)}{\partial \theta} \mathrm{d}y - \int_{\mathcal{Y}} \mu(\theta) \frac{\partial p_y(y; \theta)}{\partial \theta} \mathrm{d}y \right)^2}{\sigma^2(\theta)}$$

$$= \frac{\left(\frac{\partial}{\partial \theta} \int_{\mathcal{Y}} y p_y(y; \theta) \mathrm{d}y - \mu(\theta) \frac{\partial}{\partial \theta} \int_{\mathcal{Y}} p_y(y; \theta) \mathrm{d}y \right)^2}{\sigma^2(\theta)} \tag{6}$$

$$= \frac{1}{\sigma^2(\theta)} \left(\frac{\partial \mu(\theta)}{\partial \theta} \right)^2, \tag{7}$$

allows to bound (3) in general[1] from below.

[1] Note that step (6) necessitates $p_y(y; \theta)$ to fulfill certain regularity conditions [5]. This is the case for most systems of practical relevance.

IV. FISHER INFORMATION BOUND - MULTIVARIATE CASE

For the general case of a multivariate system output $\boldsymbol{y}$ with $\mathcal{Y} \subseteq \mathbb{R}^N$ and multiple parameters $\boldsymbol{\theta}$ with $\Theta \subseteq \mathbb{R}^K$, the parametric probabilistic system is written $p_y(\boldsymbol{y}; \boldsymbol{\theta})$. The required first and second output moments are given by

$$\boldsymbol{\mu}(\boldsymbol{\theta}) = \int_{\mathcal{Y}} \boldsymbol{y} p_y(\boldsymbol{y}; \boldsymbol{\theta}) \mathrm{d}\boldsymbol{y} \tag{8}$$

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \int_{\mathcal{Y}} (\boldsymbol{y} - \boldsymbol{\mu}(\boldsymbol{\theta}))(\boldsymbol{y} - \boldsymbol{\mu}(\boldsymbol{\theta}))^{\mathrm{T}} p_y(\boldsymbol{y}; \boldsymbol{\theta}) \mathrm{d}\boldsymbol{y}, \tag{9}$$

and the Fisher information measure is defined in matrix form

$$\boldsymbol{F}(\boldsymbol{\theta}) = \int_{\mathcal{Y}} p_y(\boldsymbol{y}; \boldsymbol{\theta}) \left(\frac{\partial \log p_y(\boldsymbol{y}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^{\mathrm{T}} \frac{\partial \log p_y(\boldsymbol{y}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \mathrm{d}\boldsymbol{y}. \tag{10}$$

The goal now is to show that if the derivative $\frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$ exists,

$$\boldsymbol{F}(\boldsymbol{\theta}) \succeq \left(\frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^{\mathrm{T}} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) \left(\frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right) \tag{11}$$

holds, where with $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{K \times K}$, the matrix inequality

$$\boldsymbol{A} \succeq \boldsymbol{B}, \tag{12}$$

stands for the fact that

$$\boldsymbol{x}^{\mathrm{T}} (\boldsymbol{A} - \boldsymbol{B}) \boldsymbol{x} \geq 0 \quad \forall \boldsymbol{x} \in \mathbb{R}^K. \tag{13}$$

A. Proof

Inequality (41) from the appendix is used, such that

$$\boldsymbol{F}(\boldsymbol{\theta}) = \int_{\mathcal{Y}} \left(\frac{\partial \log p_y(\boldsymbol{y}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^{\mathrm{T}} \frac{\partial \log p_y(\boldsymbol{y}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} p_y(\boldsymbol{y}; \boldsymbol{\theta}) \mathrm{d}\boldsymbol{y}$$

$$\succeq \int_{\mathcal{Y}} \left(\frac{\partial \log p_y(\boldsymbol{y}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^{\mathrm{T}} (\boldsymbol{y} - \boldsymbol{\mu}(\boldsymbol{\theta}))^{\mathrm{T}} p_y(\boldsymbol{y}; \boldsymbol{\theta}) \mathrm{d}\boldsymbol{y}$$

$$\cdot \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) \int_{\mathcal{Y}} (\boldsymbol{y} - \boldsymbol{\mu}(\boldsymbol{\theta})) \frac{\partial \log p_y(\boldsymbol{y}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} p_y(\boldsymbol{y}; \boldsymbol{\theta}) \mathrm{d}\boldsymbol{y}$$

$$= \int_{\mathcal{Y}} \left(\frac{\partial p_y(\boldsymbol{y}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^{\mathrm{T}} (\boldsymbol{y} - \boldsymbol{\mu}(\boldsymbol{\theta}))^{\mathrm{T}} \mathrm{d}\boldsymbol{y}$$

$$\cdot \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) \int_{\mathcal{Y}} (\boldsymbol{y} - \boldsymbol{\mu}(\boldsymbol{\theta})) \frac{\partial p_y(\boldsymbol{y}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \mathrm{d}\boldsymbol{y}$$

$$= \left(\frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^{\mathrm{T}} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) \left(\frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right). \tag{14}$$

V. INTERPRETATION - EQUIVALENT PESSIMISTIC SYSTEM

For an interpretation of the presented bounding technique, we consider instead of the model $p_y(y; \theta)$, the Gaussian system

$$q_y(y; \theta) = \frac{1}{\sqrt{2\pi\sigma^2(\theta)}} \mathrm{e}^{-\frac{(y - \mu(\theta))^2}{2\sigma^2(\theta)}}, \tag{15}$$

with equivalent first and second moments. If the dependency between $\sigma^2(\theta)$ and the parameter θ is ignored and

$$\frac{\partial \sigma^2(\theta)}{\partial \theta} = 0 \tag{16}$$

is postulated, the Fisher information measure

$$\tilde{F}(\theta) = \int_{\mathcal{Y}} q_y(y; \theta) \left(\frac{\partial \log q_y(y; \theta)}{\partial \theta} \right)^2 \mathrm{d}y \tag{17}$$

attains the value

$$\tilde{F}(\theta) = \frac{1}{\sigma^2(\theta)} \left(\frac{\partial \mu(\theta)}{\partial \theta} \right)^2. \tag{18}$$

For the multivariate case, the Gaussian system

$$q_y(\boldsymbol{y}; \boldsymbol{\theta}) = \frac{e^{-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{\mu}(\boldsymbol{\theta}))^{\mathrm{T}} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})(\boldsymbol{y}-\boldsymbol{\mu}(\boldsymbol{\theta}))}}{(2\pi)^{\frac{N}{2}} (\det \boldsymbol{\Sigma}(\boldsymbol{\theta}))^{\frac{1}{2}}} \tag{19}$$

exhibits the Fisher information

$$\begin{aligned}
\tilde{\mathbf{F}}(\boldsymbol{\theta}) &= \int_{\mathcal{Y}} q_y(\boldsymbol{y}; \boldsymbol{\theta}) \left(\frac{\partial \log q_y(\boldsymbol{y}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^{\mathrm{T}} \frac{\partial \log q_y(\boldsymbol{y}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \mathrm{d}\boldsymbol{y} \\
&= \left(\frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^{\mathrm{T}} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) \left(\frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right),
\end{aligned} \tag{20}$$

when claiming that

$$\frac{\partial \boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \theta_k} = \mathbf{0} \tag{21}$$

for $k = 1, \ldots, K$. This shows that the bounding approach can be interpreted as a replacement of the system $p_y(\boldsymbol{y}; \boldsymbol{\theta})$ by an equivalent pessimistic counterpart $q_y(\boldsymbol{y}; \boldsymbol{\theta})$, for which the inequality

$$\boldsymbol{F}(\boldsymbol{\theta}) \succeq \tilde{\mathbf{F}}(\boldsymbol{\theta}) \tag{22}$$

holds. So, the Fisher information $\boldsymbol{F}(\boldsymbol{\theta})$ of the original system $p_y(\boldsymbol{y}; \boldsymbol{\theta})$ always dominates the information measure $\tilde{\boldsymbol{F}}(\boldsymbol{\theta})$ calculated for the equivalent Gaussian system $q_y(\boldsymbol{y}; \boldsymbol{\theta})$.

VI. MINIMUM FISHER INFORMATION - A SPECIAL CASE

A question which has received attention in the field of statistical signal processing is to specify the worst-case independent additive noise model under a fixed variance [6][7]. In order to show that our approach includes this relevant special case, consider the multivariate additive system model

$$\boldsymbol{y} = \boldsymbol{s}(\theta) + \boldsymbol{\eta}, \tag{23}$$

where the location parameter $\theta \in \mathbb{R}$ modulates the signal $\boldsymbol{s}(\theta) \in \mathbb{R}^N$ and $\boldsymbol{\eta} \in \mathbb{R}^N$ is a zero-mean independent random process with constant second moment $\boldsymbol{R} \in \mathbb{R}^{N \times N}$. As the first two central moments are

$$\begin{aligned}
\boldsymbol{\mu}(\theta) &= \int_{\mathcal{Y}} \boldsymbol{y} p_y(\boldsymbol{y}; \theta) \mathrm{d}\boldsymbol{y} \\
&= \boldsymbol{s}(\theta) \tag{24} \\
\boldsymbol{\Sigma} &= \int_{\mathcal{Y}} (\boldsymbol{y} - \boldsymbol{\mu}(\theta))(\boldsymbol{y} - \boldsymbol{\mu}(\theta))^{\mathrm{T}} p_y(\boldsymbol{y}; \theta) \mathrm{d}\boldsymbol{y} \\
&= \boldsymbol{R}, \tag{25}
\end{aligned}$$

it follows from (11), that assuming $\boldsymbol{\eta}$ to be normally distributed with constant covariance $\boldsymbol{R}$ is the worst-case assumption from an estimation theoretic perspective. However, note that the presented result allows a pessimistic statement about the Fisher information even for the more general case with parametric covariance $\boldsymbol{R}(\theta)$. Therefore, it provides a tool for the analysis of a brighter class of systems than those considered in [6][7].

VII. THEORETICAL AND PRACTICAL APPLICABILITY

For a given system model $p_y(\boldsymbol{y}; \theta)$ with $\mathcal{Y} \subseteq \mathbb{R}^N$, the calculation of the Fisher information measure can be difficult, due to the fact that

$$F(\theta) = \int_{\mathcal{Y}} p_y(\boldsymbol{y}; \theta) \left(\frac{\partial \log p_y(\boldsymbol{y}; \theta)}{\partial \theta} \right)^2 \mathrm{d}\boldsymbol{y} \tag{26}$$

generally requires the evaluation of an N-fold integral. The presented bound $\tilde{F}(\theta)$ has the advantage that the calculation of the two required moments

$$\boldsymbol{\mu}(\theta) = \int_{\mathcal{Y}} \boldsymbol{y} p_y(\boldsymbol{y}; \theta) \mathrm{d}\boldsymbol{y} \tag{27}$$

$$\boldsymbol{\Sigma}(\theta) = \int_{\mathcal{Y}} (\boldsymbol{y} - \boldsymbol{\mu}(\theta))(\boldsymbol{y} - \boldsymbol{\mu}(\theta))^{\mathrm{T}} p_y(\boldsymbol{y}; \theta) \mathrm{d}\boldsymbol{y}, \tag{28}$$

can be performed element-wise, which after marginalization, only requires computing a single and a two-fold integral, respectively. Further, for the case where a mathematical model $p_y(\boldsymbol{y}; \theta)$ is not available, which is a situation frequently encountered in practice, our method allows to estimate the Fisher information by determining the required parametric moments in a calibrated measurement or simulation environment.

VIII. BOUNDING QUALITY - HARD-LIMITER

In order to finally show, that the presented procedure has the potential to bound the Fisher information measure in an accurate way, we consider, as an example, a Gaussian signal with mean $\theta \in \mathbb{R}$, which is processed by a hard-limiting device

$$y = \mathrm{sign}_{\alpha}(\theta + \eta), \tag{29}$$

where the quantization operation with binary output reads as

$$\mathrm{sign}_{\alpha}(x) = \begin{cases} +1 & \text{if } x \geq \alpha, \\ -1 & \text{if } x < \alpha \end{cases} \tag{30}$$

and the additive independent noise term $\eta \in \mathbb{R}$ follows the normal probability density with unit variance

$$p_\eta(\eta) = \frac{e^{-\frac{\eta^2}{2}}}{\sqrt{2\pi}}. \tag{31}$$

In this case

$$\begin{aligned}
p(y = +1; \theta) &= \int_{-\theta+\alpha}^{\infty} p_\eta(\eta) \mathrm{d}\eta \\
&= \frac{1}{2} \left(1 + \mathrm{erf} \left(\frac{\theta - \alpha}{\sqrt{2}} \right) \right), \tag{32} \\
p(y = -1; \theta) &= \int_{-\infty}^{-\theta+\alpha} p_\eta(\eta) \mathrm{d}\eta \\
&= \frac{1}{2} \left(1 - \mathrm{erf} \left(\frac{\theta - \alpha}{\sqrt{2}} \right) \right) \tag{33}
\end{aligned}$$

and

$$\frac{\partial p(y = +1; \theta)}{\partial \theta} = \frac{e^{-\left(\frac{\theta - \alpha}{\sqrt{2}} \right)^2}}{\sqrt{2\pi}}, \tag{34}$$

$$\frac{\partial p(y = -1; \theta)}{\partial \theta} = -\frac{e^{-\left(\frac{\theta - \alpha}{\sqrt{2}} \right)^2}}{\sqrt{2\pi}}. \tag{35}$$

Therefore, the exact Fisher information of the system (29) is

$$
\begin{aligned}
F(\theta) &= \int_{\mathcal{Y}} \frac{1}{p_y(y;\theta)} \left(\frac{\partial p_y(y;\theta)}{\partial \theta} \right)^2 \mathrm{d}y \\
&= \left(\frac{\mathrm{e}^{-\left(\frac{\theta-\alpha}{\sqrt{2}} \right)^2}}{\sqrt{2\pi}} \right)^2 \left(\frac{2}{1+\mathrm{erf}\left(\frac{\theta-\alpha}{\sqrt{2}} \right)} + \frac{2}{1-\mathrm{erf}\left(\frac{\theta-\alpha}{\sqrt{2}} \right)} \right) \\
&= \frac{2}{\pi} \frac{\mathrm{e}^{-(\theta-\alpha)^2}}{1-\mathrm{erf}^2\left(\frac{\theta-\alpha}{\sqrt{2}} \right)}.
\end{aligned}
\tag{36}
$$

In order to apply the information bound (5), the first moment of the output (29) is found to be

$$
\begin{aligned}
\mu(\theta) &= \int_{\mathcal{Y}} y\, p_y(y;\theta)\mathrm{d}y \\
&= \mathrm{erf}\left(\frac{\theta-\alpha}{\sqrt{2}} \right),
\end{aligned}
\tag{37}
$$

while the second central moment is

$$
\begin{aligned}
\sigma^2(\theta) &= \int_{\mathcal{Y}} \left(y - \mathrm{erf}\left(\frac{\theta-\alpha}{\sqrt{2}} \right) \right)^2 p_y(y;\theta)\mathrm{d}y \\
&= 1 - \mathrm{erf}^2\left(\frac{\theta-\alpha}{\sqrt{2}} \right).
\end{aligned}
\tag{38}
$$

With the derivative of the first moment

$$
\frac{\partial \mu(\theta)}{\partial \theta} = \sqrt{\frac{2}{\pi}}\mathrm{e}^{-\left(\frac{\theta-\alpha}{\sqrt{2}} \right)^2},
\tag{39}
$$

the lower bound for the Fisher information measure

$$
\begin{aligned}
\tilde{F}(\theta) &= \frac{1}{\sigma^2(\theta)} \left(\frac{\partial \mu(\theta)}{\partial \theta} \right)^2 \\
&= \frac{2}{\pi} \frac{\mathrm{e}^{-(\theta-\alpha)^2}}{1-\mathrm{erf}^2\left(\frac{\theta-\alpha}{\sqrt{2}} \right)}
\end{aligned}
\tag{40}
$$

interestingly matches the exact result (36).

IX. Conclusion

We have established a lower bound for the Fisher information measure under an arbitrary parametric probabilistic system. Being able to characterize the first and the second central moment of the system output as functions of the system parameters, consultation of an equivalent Gaussian system leads to a pessimistic description of the information measure. Using the presented approach for a non-linear system of practical relevance, we have demonstrated that the presented bound has the potential to be tight to the exact information measure. In particular, in situations where the analytical characterization of the information measure with the model of interest is difficult, but where the first two central moments are available, the presented technique allows to analyze the system with respect to the achievable estimation performance.

Appendix A

Proposition: Given multivariate random variables $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^N$ which follow the joint probability distribution $p_{xy}(\boldsymbol{x}, \boldsymbol{y})$,

$$
\mathrm{E}_y\left[\boldsymbol{y}\boldsymbol{y}^{\mathrm{T}} \right] \succeq \mathrm{E}_{xy}\left[\boldsymbol{y}\boldsymbol{x}^{\mathrm{T}} \right] \mathrm{E}_x\left[\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}} \right]^{-1} \mathrm{E}_{xy}\left[\boldsymbol{x}\boldsymbol{y}^{\mathrm{T}} \right].
\tag{41}
$$

Proof: Given $\boldsymbol{x}$, construct the auxiliary random variable

$$
\hat{\boldsymbol{y}}(\boldsymbol{x}) = \mathrm{E}_{xy}\left[\boldsymbol{y}\boldsymbol{x}^{\mathrm{T}} \right] \mathrm{E}_x\left[\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}} \right]^{-1} \boldsymbol{x}.
\tag{42}
$$

Observing that by construction,

$$
\mathrm{E}_{xy}\left[(\boldsymbol{y} - \hat{\boldsymbol{y}}(\boldsymbol{x}))(\boldsymbol{y} - \hat{\boldsymbol{y}}(\boldsymbol{x}))^{\mathrm{T}} \right] \succeq \boldsymbol{0}
\tag{43}
$$

proves that

$$
\mathrm{E}_y\left[\boldsymbol{y}\boldsymbol{y}^{\mathrm{T}} \right] - \mathrm{E}_{xy}\left[\boldsymbol{y}\boldsymbol{x}^{\mathrm{T}} \right] \mathrm{E}_x\left[\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}} \right]^{-1} \mathrm{E}_{xy}\left[\boldsymbol{x}\boldsymbol{y}^{\mathrm{T}} \right] \succeq \boldsymbol{0}.
\tag{44}
$$

∎

Corollary: Given scalar random variables $x, y \in \mathbb{R}$, which follow the joint probability distribution $p_{xy}(x, y)$, it holds that

$$
\int_{\mathcal{Y}} y^2 p_y(y)\mathrm{d}y \int_{\mathcal{X}} x^2 p_x(x)\mathrm{d}x \geq \left(\int_{\mathcal{X}} \int_{\mathcal{Y}} xy\, p_{xy}(x,y)\mathrm{d}x\mathrm{d}y \right)^2.
\tag{45}
$$

Proof: With $N = 1$, inequality (41) simplifies to

$$
\int_{\mathcal{Y}} y^2 p_y(y)\mathrm{d}y \geq \frac{\left(\int_{\mathcal{X}} \int_{\mathcal{Y}} xy\, p_{xy}(x,y)\mathrm{d}x\mathrm{d}y \right)^2}{\int_{\mathcal{X}} x^2 p_x(x)\mathrm{d}x},
\tag{46}
$$

such that (45) follows directly. ∎

References

[1] R. A. Fisher, "On the mathematical foundations of theoretical statistics," *Philos. Trans. Roy. Soc. London*, vol. 222, pp. 309–368, Jan. 1922.

[2] R. A. Fisher, "Theory of statistical estimation," in *Proc. Cambridge Philos. Soc*, Jul. 1925, vol. 22, no. 5, pp. 700–725.

[3] A. J. Stam, "Some mathematical properties of quantities of information," Ph.D. thesis, TU Delft, Delft, The Netherlands, 1959.

[4] M. Cohen, "The Fisher information and convexity," *IEEE Trans. Inf. Theory*, vol. 14, pp. 591–592, Jul. 1968.

[5] D. E. Boekee, "A generalization of the Fisher information measure," Ph.D. thesis, TU Delft, Delft, The Netherlands, 1977.

[6] P. Stoica and P. Babu, "The Gaussian data assumption leads to the largest Cramèr-Rao bound," *IEEE Signal Process. Mag.*, vol. 28, pp. 132–133, May 2011.

[7] S. Park, E. Serpedin, and K. Qaraqe, "Gaussian assumption: The least favorable but the most useful," *IEEE Signal Process. Mag.*, vol. 30, pp. 183–186, Apr. 2013.

[8] P. J. Huber, "Robust statistics: A review," *Ann. Math. Statist.*, vol. 43, no. 4, pp. 1041–1067, 1972.

[9] V. Živojnović, "Minimum Fisher information of moment-constrained distributions with application to robust blind identification," *Signal Process.*, vol. 65, no. 2, pp. 297–313, Oct. 1998.

[10] E. Uhrmann-Klingen, "Minimal Fisher information distributions with compact-supports," *Sankhyā: Indian J. Stat.*, vol. 57, no. 3, pp. 360–374, Oct. 1995.

[11] J. F. Bercher and C. Vignat, "On minimum Fisher information distributions with restricted support and fixed variance," *Inf. Sci.*, vol. 179, no. 22, Nov. 2009.

[12] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddler River, NJ, USA: Prentice-Hall, 1993.