

A Pessimistic Approximation for the Fisher Information Measure

Manuel S. Stein, *Student Member, IEEE*, and Josef A. Nossek, *Life Fellow, IEEE*

Abstract—The problem of determining the intrinsic quality of a signal processing system with respect to the inference of an unknown deterministic parameter θ is considered. While the Fisher information measure $F(\theta)$ forms a classical tool for such a problem, direct computation of the information measure can become difficult in various situations. For the estimation theoretic performance analysis of nonlinear measurement systems, the form of the likelihood function can make the calculation of the information measure $F(\theta)$ challenging. In situations where no closed-form expression of the statistical system model is available, the analytical derivation of $F(\theta)$ is not possible at all. Based on the Cauchy–Schwarz inequality, we derive an alternative information measure $S(\theta)$. It provides a lower bound on the Fisher information $F(\theta)$ and has the property of being evaluated with the mean, the variance, the skewness, and the kurtosis of the system model at hand. These entities usually exhibit good mathematical tractability or can be determined at low-complexity by real-world measurements in a calibrated setup. With various examples, we show that $S(\theta)$ provides a good conservative approximation for $F(\theta)$ and outline different estimation theoretic problems where the presented information bound turns out to be useful.

Index Terms—Cramér-Rao lower bound, estimation theory, Fisher information lower bound, smooth limiter, minimum Fisher information, nonlinear systems, squaring loss, worst-case noise.

I. INTRODUCTION

SUPPOSE we are given a parametric system, characterized by a probability density or mass function $q(\mathbf{y}; \theta)$, and face the problem of having to infer the deterministic but unknown system parameter $\theta \in \Theta$ from measurements at the system output Y . The output Y takes random values $\mathbf{y} \in \mathcal{Y}$, where \mathcal{Y} denotes the support of the random variable Y . In such a situation, estimation theory [1], [2] provides a variety of useful tools. On the one hand, we have guidelines for the design of *estimation algorithms* [3], and on the other hand, corresponding *performance bounds* [4]–[9]. While the latter have originally been derived to benchmark estimation algorithms, identify potential for further improvements, or to establish their efficiency, these error bounds have also become popular as a figure of merit for

Manuscript received September 15, 2015; revised May 8, 2016; accepted June 29, 2016. Date of publication October 13, 2016; date of current version November 17, 2016. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Tharmalingam Ratnarajah.

M. S. Stein is with the Digital Mathematics Group (DIMA), Mathematics Department (DWIS), Vrije Universiteit Brussel, Brussel B-1050, Belgium (e-mail: manuel.stein@vub.ac.be).

J. A. Nossek is with the Department of Teleinformatics Engineering, Universidade Federal do Ceará, Fortaleza 60020-181, Brasil, and also with the Department of Electrical and Computer Engineering, Technische Universität München, München 80333, Germany (e-mail: josef.a.nossek@tum.de).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TSP.2016.2617824

the design and optimization of the measurement system $q(\mathbf{y}; \theta)$. Such a problem frequently arises in the field of signal processing, where not only the efficient extraction of information from noisy data is in the interest of engineers, but also the design of the physical measurement system $q(\mathbf{y}; \theta)$ itself. Note that the layout of the measurement sensors can significantly influence technical properties such as computational complexity, power consumption, production cost, reliability, processing delay and system performance. Therefore, given the ability to modify the data gathering system $q(\mathbf{y}; \theta)$ to an alternative design $p(\mathbf{z}; \theta)$ with the altered output Z exhibiting realizations $\mathbf{z} \in \mathcal{Z}$, a rigorous method is required in order to draw a precise conclusion about the achievable performance of the two systems when operating with optimum estimation procedures $\hat{\theta}(\mathbf{y})$ or $\hat{\theta}(\mathbf{z})$. Note that here $\mathbf{y} \in \mathcal{Y}^N$ denotes a collection of N independent realizations of the system outputs Y , such that

$$q(\mathbf{y}; \theta) = \prod_{n=1}^N q(y_n; \theta), \quad \forall \mathbf{y} \in \mathcal{Y}^N. \quad (1)$$

A. Estimation Theory and the Fisher Information Measure

In order to motivate the use of an information measure, in the following, the performance of the estimator $\hat{\theta}(\mathbf{y})$ is analyzed. We restrict the discussion to the problem of performing unbiased estimation

$$\int_{\mathcal{Y}^N} \hat{\theta}(\mathbf{y}) q(\mathbf{y}; \theta) d\mathbf{y} = \theta. \quad (2)$$

Further, we assume that the system $q(\mathbf{y}; \theta)$ is differentiable in $\theta \in \Theta$ for every $\mathbf{y} \in \mathcal{Y}^N$, where the parameter set Θ is an open subset on the real line. All considered system models exhibit regularity, such that the statement

$$\frac{\partial}{\partial \theta} \int_{\mathcal{Y}^N} f(\mathbf{y}) q(\mathbf{y}; \theta) d\mathbf{y} = \int_{\mathcal{Y}^N} f(\mathbf{y}) \frac{\partial q(\mathbf{y}; \theta)}{\partial \theta} d\mathbf{y} \quad (3)$$

holds for any function $f(\cdot)$ which does not present θ as an argument. Applying (3) to (2) we can set

$$\int_{\mathcal{Y}^N} \hat{\theta}(\mathbf{y}) \frac{\partial q(\mathbf{y}; \theta)}{\partial \theta} d\mathbf{y} = 1. \quad (4)$$

With the requirement

$$\int_{\mathcal{Y}^N} q(\mathbf{y}; \theta) d\mathbf{y} = 1, \quad \forall \theta \in \Theta, \quad (5)$$

it follows that

$$\frac{\partial}{\partial \theta} \int_{\mathcal{Y}^N} q(\mathbf{y}; \theta) d\mathbf{y} = 0, \quad \forall \theta \in \Theta, \quad (6)$$

such that multiplying (6) by θ and expanding (4), we have

$$\int_{\mathcal{Y}^N} (\hat{\theta}(\mathbf{y}) - \theta) \frac{\partial q(\mathbf{y}; \theta)}{\partial \theta} d\mathbf{y} = 1. \quad (7)$$

Using the fact that

$$\frac{\partial \ln q(\mathbf{y}; \theta)}{\partial \theta} = \frac{1}{q(\mathbf{y}; \theta)} \frac{\partial q(\mathbf{y}; \theta)}{\partial \theta}, \quad (8)$$

equation (7) is manipulated, resulting in

$$\int_{\mathcal{Y}^N} (\hat{\theta}(\mathbf{y}) - \theta) \frac{\partial \ln q(\mathbf{y}; \theta)}{\partial \theta} q(\mathbf{y}; \theta) d\mathbf{y} = 1. \quad (9)$$

For two real-valued functions $f(\cdot)$, $g(\cdot)$ and a random variable $\mathbf{X} \in \mathbb{R}^N$, the Cauchy-Schwarz inequality [10] states that

$$\begin{aligned} & \int_{\mathcal{Y}^N} f^2(\mathbf{x}) p(\mathbf{x}) d\mathbf{x} \int_{\mathcal{Y}^N} g^2(\mathbf{x}) p(\mathbf{x}) d\mathbf{x} \\ & \geq \left(\int_{\mathcal{Y}^N} f(\mathbf{x}) g(\mathbf{x}) p(\mathbf{x}) d\mathbf{x} \right)^2, \end{aligned} \quad (10)$$

where $p(\cdot)$ is a probability distribution function. By setting

$$f(\cdot) = \hat{\theta}(\mathbf{y}) - \theta, \quad (11)$$

$$g(\cdot) = \frac{\partial \ln q(\mathbf{y}; \theta)}{\partial \theta} \quad (12)$$

and $p(\cdot) = q(\mathbf{y}; \theta)$, this allows us to derive the inequality

$$\begin{aligned} & \int_{\mathcal{Y}^N} (\hat{\theta}(\mathbf{y}) - \theta)^2 q(\mathbf{y}; \theta) d\mathbf{y} \\ & \geq \left(\int_{\mathcal{Y}^N} \left(\frac{\partial \ln q(\mathbf{y}; \theta)}{\partial \theta} \right)^2 q(\mathbf{y}; \theta) d\mathbf{y} \right)^{-1} \end{aligned} \quad (13)$$

from expression (9). As long as the observations are independent (1) and each element Y_n follows the identical statistical model

$$q(y_n; \theta) = q(y; \theta), \quad \forall n \in \{1, 2, \dots, N\}, \quad (14)$$

the right hand side of (13) simplifies to

$$\begin{aligned} & \int_{\mathcal{Y}^N} \left(\frac{\partial \ln q(\mathbf{y}; \theta)}{\partial \theta} \right)^2 q(\mathbf{y}; \theta) d\mathbf{y} \\ & = N \int_{\mathcal{Y}} \left(\frac{\partial \ln q(y; \theta)}{\partial \theta} \right)^2 q(y; \theta) dy. \end{aligned} \quad (15)$$

The left hand side of (13) is identified as the mean squared-error $\text{mse}_{\mathbf{Y}}(\theta)$ of the estimator $\hat{\theta}(\mathbf{y})$, such that the Cramér-Rao inequality [4], [5] for unbiased estimators

$$\text{mse}_{\mathbf{Y}}(\theta) = \text{var}_{\mathbf{Y}}(\theta) \geq \frac{1}{NF_{\mathbf{Y}}(\theta)} \quad (16)$$

is obtained. Note that an estimator $\hat{\theta}(\mathbf{y})$, which asymptotically in N attains equality with respect to (16), is called asymptotically efficient. Estimators designed along the principle of maximum-likelihood are known to exhibit efficiency in the asymptotic regime [3, App. 7B] under mild conditions. Consequently, when N is sufficiently large, the Fisher information

$$F_{\mathbf{Y}}(\theta) = \int_{\mathcal{Y}} \left(\frac{\partial \ln q(y; \theta)}{\partial \theta} \right)^2 q(y; \theta) dy \quad (17)$$

is a measure on the amount of intrinsic information about the unknown deterministic parameter θ contained in average within each observation of the random output Y . Note that the Fisher information measure also plays an important role for performance bounds in the Bayesian setting [11]–[15], where the parameter θ is considered to be a random variable. A comprehensive overview on this topic, which is out of the scope of this article, can be found in [16].

B. Relative Inference Capability

As the inequality (16) holds for all estimation procedures satisfying (2), the relative estimation theoretic quality of the modification $p(z; \theta)$ with respect to the reference system $q(y; \theta)$ can be assessed by the information ratio

$$\chi(\theta) = \frac{F_Z(\theta)}{F_{\mathbf{Y}}(\theta)}. \quad (18)$$

Note that $F_Z(\theta)$ is the Fisher information (17) evaluated on \mathcal{Z} with respect to the parametric probability function $p(z; \theta)$.

C. Fisher Information Lower Bound

Using the information ratio (18) for the design and the optimization of the measurement system $p(z; \theta)$ requires computing (17) for the benchmark experiment $q(y; \theta)$ and all modifications $p(z; \theta)$ which are of interest. If due to the alteration the probability distribution $p(z; \theta)$ takes a complicated form, this can become difficult. In a situation where the parametric probabilistic model $p(z; \theta)$ which defines the statistical behavior of the random output Z is unknown, a direct analytical formulation of the information measure (17) becomes impossible. However, if the mean

$$\mu_1(\theta) = \int_{\mathcal{Z}} z p_z(z; \theta) dz \quad (19)$$

of the system output Z and the variance

$$\mu_2(\theta) = \int_{\mathcal{Z}} (z - \mu_1(\theta))^2 p_z(z; \theta) dz \quad (20)$$

are known and are differentiable in θ , it can be shown that the Fisher information $F(\theta)$ is in general bounded from below [17], [18]

$$F(\theta) \geq \frac{1}{\mu_2(\theta)} \left(\frac{\partial \mu_1(\theta)}{\partial \theta} \right)^2. \quad (21)$$

While in [18] the example of a hard-limited Gaussian model was given where the information bound (21) holds with equality, a simple counter example is immediately constructed. To this end, consider the system output to follow the generic parametric Gaussian distribution

$$p(z; \theta) = \frac{1}{\sqrt{2\pi\mu_2(\theta)}} e^{-\frac{(z - \mu_1(\theta))^2}{2\mu_2(\theta)}}. \quad (22)$$

The exact Fisher information is [3, pp. 47]

$$F(\theta) = \frac{1}{\mu_2(\theta)} \left(\frac{\partial \mu_1(\theta)}{\partial \theta} \right)^2 + \frac{1}{2\mu_2^2(\theta)} \left(\frac{\partial \mu_2(\theta)}{\partial \theta} \right)^2 \quad (23)$$

and is equal to the right-hand side of (21) only for the special case where

$$\frac{\partial \mu_2(\theta)}{\partial \theta} = 0. \quad (24)$$

Obviously, the inequality (21) does not take into account the contribution of the variance $\mu_2(\theta)$ to the Fisher information measure $F(\theta)$.

D. Contribution and Outline

Motivated by the insight obtained in the preceding section, we aim to improve the lower bound (21). We achieve this by utilizing the Cauchy-Schwarz inequality (10) with a more general approach than in [18] and subsequently maximizing the resulting expression. This leads to an alternative information measure $S(\theta)$, which forms a pessimistic approximation for $F(\theta)$ and exclusively contains the mean, the variance, the skewness, and the kurtosis of the system output model in parametric form. A discussion on situations where the derivative of the variance vanishes (24) shows that the inequality (21) is contained in the presented result as a special case. Using various examples with continuous and discrete system outputs, we verify the quality of the alternative information measure $S(\theta)$. In order to demonstrate possible applications of the result and to provide further insights through $S(\theta)$, we approximately determine the estimation theoretic information loss when squaring a standard Gaussian input distribution and advance the discussion concerning minimum Fisher information [18]–[24]. Finally, we mimic a situation of practical relevance. Measuring the output moments of a smooth limiting device with standard Gaussian input, we demonstrate how to conservatively establish the intrinsic inference capability $F(\theta)$ of a nonlinear signal processing system through the information measure $S(\theta)$, when the analytic form of the parametric output model $p(z; \theta)$ is not available.

II. IMPROVED FISHER INFORMATION BOUND

For the discussion, we additionally require the central output moments

$$\mu_3(\theta) = \int_{\mathcal{Z}} (z - \mu_1(\theta))^3 p(z; \theta) dz, \quad (25)$$

$$\mu_4(\theta) = \int_{\mathcal{Z}} (z - \mu_1(\theta))^4 p(z; \theta) dz \quad (26)$$

and their normalized versions

$$\bar{\mu}_3(\theta) = \int_{\mathcal{Z}} \left(\frac{z - \mu_1(\theta)}{\sqrt{\mu_2(\theta)}} \right)^3 p_z(z; \theta) dz = \mu_3(\theta) \mu_2^{-\frac{3}{2}}(\theta), \quad (27)$$

$$\bar{\mu}_4(\theta) = \int_{\mathcal{Z}} \left(\frac{z - \mu_1(\theta)}{\sqrt{\mu_2(\theta)}} \right)^4 p_z(z; \theta) dz = \mu_4(\theta) \mu_2^{-2}(\theta). \quad (28)$$

Note that $\bar{\mu}_3(\theta)$ is referred to as the skewness, an indicator for the asymmetry of the output distribution $p(z; \theta)$, while $\bar{\mu}_4(\theta)$ is called the kurtosis, a characterization for the shape of the output distribution $p(z; \theta)$. Both moments stand in relation via Pearson's inequality [25]

$$\bar{\mu}_4(\theta) \geq \bar{\mu}_3^2(\theta) + 1, \quad (29)$$

for which a compact and elegant proof can be found in [26].

A. Derivation of the Information Bound

We apply the inequality (10) with

$$f(z; \theta) = \frac{\partial \ln p(z; \theta)}{\partial \theta} \quad (30)$$

and

$$g(z; \theta) = \left(\frac{z - \mu_1(\theta)}{\sqrt{\mu_2(\theta)}} \right) + \beta(\theta) \left(\frac{z - \mu_1(\theta)}{\sqrt{\mu_2(\theta)}} \right)^2 - \beta(\theta), \quad (31)$$

$\beta(\theta) \in \mathbb{R}$, in order to derive a lower bound on the Fisher information

$$F(\theta) = \int_{\mathcal{Z}} f^2(z; \theta) p(z; \theta) dz, \quad (32)$$

which takes into account the contribution of the variance to the Fisher information measure. With the manipulations

$$\begin{aligned} & \int_{\mathcal{Z}} \left(\frac{z - \mu_1(\theta)}{\sqrt{\mu_2(\theta)}} \right) \frac{\partial \ln p_z(z; \theta)}{\partial \theta} p_z(z; \theta) dz \\ &= \frac{1}{\sqrt{\mu_2(\theta)}} \left(\int_{\mathcal{Z}} z \frac{\partial p_z(z; \theta)}{\partial \theta} dz - \mu_1(\theta) \int_{\mathcal{Z}} \frac{\partial p_z(z; \theta)}{\partial \theta} dz \right) \\ &= \frac{1}{\sqrt{\mu_2(\theta)}} \left(\frac{\partial}{\partial \theta} \int_{\mathcal{Z}} z p_z(z; \theta) dz - \mu_1(\theta) \frac{\partial}{\partial \theta} \int_{\mathcal{Z}} p_z(z; \theta) dz \right) \\ &= \frac{1}{\sqrt{\mu_2(\theta)}} \frac{\partial \mu_1(\theta)}{\partial \theta} \end{aligned} \quad (33)$$

and

$$\begin{aligned} & \int_{\mathcal{Z}} \left(\frac{z - \mu_1(\theta)}{\sqrt{\mu_2(\theta)}} \right)^2 \frac{\partial \ln p_z(z; \theta)}{\partial \theta} p_z(z; \theta) dz \\ &= \frac{1}{\mu_2(\theta)} \left(\int_{\mathcal{Z}} z^2 \frac{\partial p_z(z; \theta)}{\partial \theta} dz - 2\mu_1(\theta) \int_{\mathcal{Z}} z \frac{\partial p_z(z; \theta)}{\partial \theta} dz \right. \\ & \quad \left. + \mu_1^2(\theta) \int_{\mathcal{Z}} \frac{\partial p_z(z; \theta)}{\partial \theta} dz \right) \\ &= \frac{1}{\mu_2(\theta)} \left(\frac{\partial}{\partial \theta} \int_{\mathcal{Z}} z^2 p_z(z; \theta) dz - 2\mu_1(\theta) \frac{\partial}{\partial \theta} \int_{\mathcal{Z}} z p_z(z; \theta) dz \right) \\ &= \frac{1}{\mu_2(\theta)} \left(\frac{\partial}{\partial \theta} (\mu_2(\theta) + \mu_1^2(\theta)) - 2\mu_1(\theta) \frac{\partial \mu_1(\theta)}{\partial \theta} \right) \\ &= \frac{1}{\mu_2(\theta)} \frac{\partial \mu_2(\theta)}{\partial \theta}, \end{aligned} \quad (34)$$

where we use the fact that

$$\int_{\mathcal{Z}} z^2 p_z(z; \theta) dz = \mu_2(\theta) + \mu_1^2(\theta), \quad (35)$$

the identity

$$\begin{aligned} & \int_{\mathcal{Z}} f(z; \theta) g(z; \theta) p(z; \theta) dz \\ &= \frac{1}{\sqrt{\mu_2(\theta)}} \frac{\partial \mu_1(\theta)}{\partial \theta} + \frac{\beta(\theta)}{\mu_2(\theta)} \frac{\partial \mu_2(\theta)}{\partial \theta}, \end{aligned} \quad (36)$$

is obtained. Note that

$$\begin{aligned} \int_{\mathcal{Z}} \beta(\theta) \frac{\partial \ln p(z; \theta)}{\partial \theta} p(z; \theta) dz &= \beta(\theta) \int_{\mathcal{Z}} \frac{\partial \ln p(z; \theta)}{\partial \theta} p(z; \theta) dz \\ &= \beta(\theta) \frac{\partial}{\partial \theta} \int_{\mathcal{Z}} p(z; \theta) dz = 0. \end{aligned} \quad (37)$$

Taking into account that

$$\int_{\mathcal{Z}} \left(\frac{z - \mu_1(\theta)}{\sqrt{\mu_2(\theta)}} \right) p(z; \theta) dz = 0, \quad (38)$$

$$\int_{\mathcal{Z}} \left(\frac{z - \mu_1(\theta)}{\sqrt{\mu_2(\theta)}} \right)^2 p(z; \theta) dz = 1, \quad (39)$$

we get

$$\begin{aligned} & \int_{\mathcal{Z}} g^2(z; \theta) p(z; \theta) dz \\ &= 1 + 2\beta(\theta) \bar{\mu}_3(\theta) + \beta^2(\theta) \bar{\mu}_4(\theta) - \beta^2(\theta). \end{aligned} \quad (40)$$

Therefore, from (10), (32), (36) and (40) it can be shown that the Fisher information can in general not fall below

$$\begin{aligned} F(\theta) &\geq \frac{\left(\int_{\mathcal{Z}} f(z; \theta) g(z; \theta) p(z; \theta) dz \right)^2}{\int_{\mathcal{Z}} g^2(z; \theta) p(z; \theta) dz} \\ &= \frac{1}{\mu_2(\theta) \left(1 + 2\beta(\theta) \bar{\mu}_3(\theta) + \beta^2(\theta) (\bar{\mu}_4(\theta) - 1) \right)} \end{aligned} \quad (41)$$

for any $\beta(\theta) \in \mathbb{R}$.

B. Optimization of the Information Bound

The factor $\beta(\theta)$ can be used to optimize the lower bound (41). For the trivial choice $\beta(\theta) = 0$, the expression (41) degenerates to

$$F(\theta) \geq \frac{1}{\mu_2(\theta)} \left(\frac{\partial \mu_1(\theta)}{\partial \theta} \right)^2, \quad (42)$$

which turns out to be the bound in (21). In order to improve this result, note that the problem

$$x^* = \arg \max_{x \in \mathbb{R}} h(x) \quad (43)$$

with

$$h(x) = \frac{(a + xb)^2}{1 + 2xc + x^2d}, \quad (44)$$

and $bc - ad \neq 0$ has a unique maximizing solution

$$x^* = \frac{ac - b}{bc - ad}. \quad (45)$$

Consequently, the tightest form of (41) is given by

$$F(\theta) \geq S(\theta), \quad (46)$$

where

$$S(\theta) = \frac{1}{\mu_2(\theta) \left(1 + 2\beta^*(\theta) \bar{\mu}_3(\theta) + \beta^{*2}(\theta) (\bar{\mu}_4(\theta) - 1) \right)} \quad (47)$$

with the optimization result

$$\begin{aligned} \beta^*(\theta) &= \frac{\frac{\partial \mu_1(\theta)}{\partial \theta} \bar{\mu}_3(\theta) - \frac{1}{\sqrt{\mu_2(\theta)}} \frac{\partial \mu_2(\theta)}{\partial \theta}}{\frac{1}{\sqrt{\mu_2(\theta)}} \frac{\partial \mu_2(\theta)}{\partial \theta} \bar{\mu}_3(\theta) - \frac{\partial \mu_1(\theta)}{\partial \theta} (\bar{\mu}_4(\theta) - 1)} \\ &= \frac{\frac{\partial \mu_1(\theta)}{\partial \theta} \sqrt{\mu_2(\theta)} \bar{\mu}_3(\theta) - \frac{\partial \mu_2(\theta)}{\partial \theta}}{\frac{\partial \mu_2(\theta)}{\partial \theta} \bar{\mu}_3(\theta) - \frac{\partial \mu_1(\theta)}{\partial \theta} \sqrt{\mu_2(\theta)} (\bar{\mu}_4(\theta) - 1)}. \end{aligned} \quad (48)$$

Note that for the case where it holds that

$$\frac{\partial \mu_1(\theta)}{\partial \theta} \sqrt{\mu_2(\theta)} \bar{\mu}_3(\theta) = \frac{\partial \mu_2(\theta)}{\partial \theta}, \quad (49)$$

the optimization of (47) results in

$$\beta^*(\theta) = 0 \quad (50)$$

and the approximation obtains the compact form (21). The inequality (46) states that the derived information measure $S(\theta)$ is always dominated by the Fisher information measure $F(\theta)$. Therefore, $S(\theta)$ gives a cautious approximation for $F(\theta)$. The Fisher information $F(\theta)$ requires integrating the squared score function

$$f^2(z; \theta) = \left(\frac{\partial \ln p(z; \theta)}{\partial \theta} \right)^2, \quad (51)$$

while in contrast, the alternative information measure $S(\theta)$ exclusively needs the mean $\mu_1(\theta)$, the variance $\mu_2(\theta)$, the skewness $\bar{\mu}_3(\theta)$, and the kurtosis $\bar{\mu}_4(\theta)$ in parametric form. These entities are usually analytically tractable, can be determined by simple measurements in a calibrated setup and are well studied for various probability laws. Note that based on raw moments and cumulants, an alternative to the bound (46) is found in [27].

C. Positiveness of the Information Bound

To ensure that the approximation (47) is always positive, it has to hold that

$$1 + 2\beta(\theta) \bar{\mu}_3(\theta) + \beta^2(\theta) (\bar{\mu}_4(\theta) - 1) \geq 0, \quad \forall \beta(\theta). \quad (52)$$

In order to demonstrate that this is the case, consider the fact that by construction

$$\begin{aligned} (1 + \beta(\theta) \sqrt{\bar{\mu}_4(\theta) - 1})^2 &= 1 + 2\beta(\theta) \sqrt{\bar{\mu}_4(\theta) - 1} \\ &+ \beta^2(\theta) (\bar{\mu}_4(\theta) - 1) \geq 0, \quad \forall \beta(\theta). \end{aligned} \quad (53)$$

With Pearson's inequality (29), we have

$$\sqrt{\bar{\mu}_4(\theta) - 1} \geq |\bar{\mu}_3(\theta)|, \quad \forall \theta, \quad (54)$$

such that with (53) the inequality

$$1 + 2\beta(\theta) |\bar{\mu}_3(\theta)| + \beta^2(\theta) (\bar{\mu}_4(\theta) - 1) \geq 0, \quad \forall \beta(\theta) \quad (55)$$

is obtained. As (55) holds irrespectively if $\beta(\theta)$ is positive or negative, we equivalently have

$$1 + 2\beta(\theta)\bar{\mu}_3(\theta) + \beta^2(\theta)(\bar{\mu}_4(\theta) - 1) \geq 0, \quad \forall \beta(\theta) \quad (56)$$

and the information bound $S(\theta)$ is always positive.

III. SPECIAL CASES OF THE INFORMATION BOUND

In order to derive simplified forms of the derived information measure (47), let us consider some special cases in the following.

A. Constant Mean

For the situation where the mean $\mu_1(\theta)$ does not vary with the system parameter θ , i.e.,

$$\frac{\partial \mu_1(\theta)}{\partial \theta} = 0, \quad \forall \theta \in \Theta, \quad (57)$$

we obtain

$$\beta^*(\theta) = -\frac{1}{\bar{\mu}_3(\theta)}, \quad (58)$$

such that a pessimistic approximation for $F(\theta)$ is

$$\begin{aligned} S(\theta) &= \frac{1}{\mu_2(\theta)} \frac{\left(-\frac{1}{\bar{\mu}_3(\theta)\sqrt{\mu_2(\theta)}} \frac{\partial \mu_2(\theta)}{\partial \theta}\right)^2}{1 - 2 + \frac{(\bar{\mu}_4(\theta) - 1)}{\bar{\mu}_3^2(\theta)}} \\ &= \frac{1}{\mu_2^2(\theta)} \frac{\left(\frac{\partial \mu_2(\theta)}{\partial \theta}\right)^2}{\bar{\mu}_4(\theta) - \bar{\mu}_3^2(\theta) - 1}. \end{aligned} \quad (59)$$

B. Constant Variance

When the variance $\mu_2(\theta)$ is constant with respect to θ , i.e.,

$$\frac{\partial \mu_2(\theta)}{\partial \theta} = 0, \quad \forall \theta \in \Theta, \quad (60)$$

it holds that

$$\beta^*(\theta) = -\frac{\bar{\mu}_3(\theta)}{(\bar{\mu}_4(\theta) - 1)}. \quad (61)$$

In this situation

$$\begin{aligned} S(\theta) &= \frac{1}{\mu_2(\theta)} \frac{\left(\frac{\partial \mu_1(\theta)}{\partial \theta}\right)^2}{1 - 2\frac{\bar{\mu}_3^2(\theta)}{(\bar{\mu}_4(\theta) - 1)} + \frac{\bar{\mu}_3^2(\theta)}{(\bar{\mu}_4(\theta) - 1)}} \\ &= \frac{1}{\mu_2(\theta)} \frac{\left(\frac{\partial \mu_1(\theta)}{\partial \theta}\right)^2}{1 - \frac{\bar{\mu}_3^2(\theta)}{(\bar{\mu}_4(\theta) - 1)}}. \end{aligned} \quad (62)$$

Note that (62) equals the expression in (21) whenever the skewness $\bar{\mu}_3(\theta)$ vanishes. In general, the relation (29) between skewness and kurtosis makes (62) larger than the unoptimized bound (21).

C. Symmetric Probability Distributions

For symmetric output distributions with zero skewness, i.e.,

$$\bar{\mu}_3(\theta) = 0, \quad (63)$$

we verify that the optimization of the information bound derived in (47) results in

$$\beta^*(\theta) = \frac{\frac{\partial \mu_2(\theta)}{\partial \theta}}{\frac{\partial \mu_1(\theta)}{\partial \theta} \sqrt{\mu_2(\theta)(\bar{\mu}_4(\theta) - 1)}}, \quad (64)$$

such that

$$\begin{aligned} S(\theta) &= \frac{1}{\mu_2(\theta)} \frac{\left(\frac{\partial \mu_1(\theta)}{\partial \theta} + \frac{\left(\frac{\partial \mu_2(\theta)}{\partial \theta}\right)^2}{\frac{\partial \mu_1(\theta)}{\partial \theta} \mu_2(\theta)(\bar{\mu}_4(\theta) - 1)}\right)^2}{1 + \left(\frac{\frac{\partial \mu_2(\theta)}{\partial \theta}}{\frac{\partial \mu_1(\theta)}{\partial \theta} \sqrt{\mu_2(\theta)(\bar{\mu}_4(\theta) - 1)}}\right)^2 (\bar{\mu}_4(\theta) - 1)} \\ &= \frac{\left(\frac{\partial \mu_1(\theta)}{\partial \theta}\right)^2 \mu_2(\theta)(\bar{\mu}_4(\theta) - 1) + \left(\frac{\partial \mu_2(\theta)}{\partial \theta}\right)^2}{\mu_2^2(\theta)(\bar{\mu}_4(\theta) - 1)} \\ &= \frac{1}{\mu_2(\theta)} \left(\frac{\partial \mu_1(\theta)}{\partial \theta}\right)^2 + \frac{1}{\mu_2^2(\theta)(\bar{\mu}_4(\theta) - 1)} \left(\frac{\partial \mu_2(\theta)}{\partial \theta}\right)^2. \end{aligned} \quad (65)$$

IV. APPROXIMATION QUALITY–CONTINUOUS OUTPUTS

In order to demonstrate the tightness of the derived information bound $S(\theta)$, we use different examples where $F(\theta)$ can be derived in a compact form. First, we discuss several well-studied distributions with continuous support \mathcal{Z} .

A. Gaussian System Model

Consider the system output Z which follows a generic Gaussian distribution of parametric form

$$p(z; \theta) = \frac{1}{\sqrt{2\pi\nu_2(\theta)}} e^{-\frac{(z - \nu_1(\theta))^2}{2\nu_2(\theta)}}. \quad (66)$$

The exact Fisher information measure is given by [3, p. 47]

$$F(\theta) = \frac{1}{\nu_2(\theta)} \left(\frac{\partial \nu_1(\theta)}{\partial \theta}\right)^2 + \frac{1}{2\nu_2^2(\theta)} \left(\frac{\partial \nu_2(\theta)}{\partial \theta}\right)^2. \quad (67)$$

As the mean, variance, skewness, and kurtosis of $p(z; \theta)$ are

$$\mu_1(\theta) = \nu_1(\theta), \quad (68)$$

$$\mu_2(\theta) = \nu_2(\theta), \quad (69)$$

$$\bar{\mu}_3(\theta) = 0, \quad (70)$$

$$\bar{\mu}_4(\theta) = 3, \quad (71)$$

with (65) we get the approximation

$$\begin{aligned} S(\theta) &= \frac{1}{\mu_2(\theta)} \left(\frac{\partial \mu_1(\theta)}{\partial \theta}\right)^2 + \frac{1}{\mu_2^2(\theta)(\bar{\mu}_4(\theta) - 1)} \left(\frac{\partial \mu_2(\theta)}{\partial \theta}\right)^2 \\ &= \frac{1}{\nu_2(\theta)} \left(\frac{\partial \nu_1(\theta)}{\partial \theta}\right)^2 + \frac{1}{2\nu_2^2(\theta)} \left(\frac{\partial \nu_2(\theta)}{\partial \theta}\right)^2. \end{aligned} \quad (72)$$

Comparing (72) to the original information measure in (67) it is obvious that here $S(\theta)$ forms a tight lower bound for $F(\theta)$.

B. Exponential System Model

As another example, we analyze the situation where samples from a parametric exponential distribution

$$p(z; \theta) = \nu(\theta) e^{-\nu(\theta)z}, \quad (73)$$

with $\nu(\theta) > 0$ and $z \geq 0$ can be collected at the random system output Z . The score function under this model is

$$\frac{\partial \ln p(z; \theta)}{\partial \theta} = \frac{1}{\nu(\theta)} \frac{\partial \nu(\theta)}{\partial \theta} - z \frac{\partial \nu(\theta)}{\partial \theta}, \quad (74)$$

such that the Fisher information is evaluated to be

$$\begin{aligned} F(\theta) &= \int_{\mathcal{Z}} \left(\frac{\partial \ln p(z; \theta)}{\partial \theta} \right)^2 p(z; \theta) dz \\ &= \frac{1}{\nu^2(\theta)} \left(\frac{\partial \nu(\theta)}{\partial \theta} \right)^2 + \left(\frac{\partial \nu(\theta)}{\partial \theta} \right)^2 \int_{\mathcal{Z}} z^2 p(z; \theta) dz \\ &\quad - \frac{2}{\nu(\theta)} \left(\frac{\partial \nu(\theta)}{\partial \theta} \right)^2 \int_{\mathcal{Z}} z p(z; \theta) dz \\ &= \frac{1}{\nu^2(\theta)} \left(\frac{\partial \nu(\theta)}{\partial \theta} \right)^2 \end{aligned} \quad (75)$$

by using

$$\int_{\mathcal{Z}} z p(z; \theta) dz = \frac{1}{\nu(\theta)}, \quad (76)$$

$$\int_{\mathcal{Z}} z^2 p(z; \theta) dz = \frac{2}{\nu^2(\theta)}. \quad (77)$$

For the approximation $S(\theta)$ the required mean, the variance, the skewness, and the kurtosis are

$$\mu_1(\theta) = \frac{1}{\nu(\theta)}, \quad (78)$$

$$\mu_2(\theta) = \frac{1}{\nu^2(\theta)}, \quad (79)$$

$$\bar{\mu}_3(\theta) = 2, \quad (80)$$

$$\bar{\mu}_4(\theta) = 3, \quad (81)$$

such that

$$\frac{\partial \mu_1(\theta)}{\partial \theta} \sqrt{\mu_2(\theta) \bar{\mu}_3(\theta)} = -\frac{2}{\nu^3(\theta)} \frac{\partial \nu(\theta)}{\partial \theta} = \frac{\partial \mu_2(\theta)}{\partial \theta}, \quad (82)$$

producing the optimization result $\beta^*(\theta) = 0$ as noted in (49). The approximation is therefore given by the simplified form

$$\begin{aligned} S(\theta) &= \frac{1}{\mu_2(\theta)} \left(\frac{\partial \mu_1(\theta)}{\partial \theta} \right)^2 = \nu^2(\theta) \left(-\frac{1}{\nu^2(\theta)} \frac{\partial \nu(\theta)}{\partial \theta} \right)^2 \\ &= \frac{1}{\nu^2(\theta)} \left(\frac{\partial \nu(\theta)}{\partial \theta} \right)^2, \end{aligned} \quad (83)$$

which matches the Fisher information $F(\theta)$ in (75) exactly.

C. Laplacian System Model

For a third example, we assume that the output Z follows a parametric Laplace distribution with zero mean, i.e.,

$$p(z; \theta) = \frac{1}{2\nu(\theta)} e^{-\frac{|z|}{\nu(\theta)}}, \quad (84)$$

with $\nu(\theta) > 0$. The score function is given by

$$\frac{\partial \ln p(z; \theta)}{\partial \theta} = -\frac{1}{\nu(\theta)} \frac{\partial \nu(\theta)}{\partial \theta} + \frac{|z|}{\nu^2(\theta)} \frac{\partial \nu(\theta)}{\partial \theta} \quad (85)$$

and the exact Fisher information is found to be

$$F(\theta) = \int_{\mathcal{Z}} \left(\frac{\partial \ln p(z; \theta)}{\partial \theta} \right)^2 p_z(z; \theta) dz = \frac{1}{\nu^2(\theta)} \left(\frac{\partial \nu(\theta)}{\partial \theta} \right)^2. \quad (86)$$

The first four moments of the output Z are

$$\mu_1(\theta) = 0, \quad (87)$$

$$\mu_2(\theta) = 2\nu^2(\theta), \quad (88)$$

$$\bar{\mu}_3(\theta) = 0, \quad (89)$$

$$\bar{\mu}_4(\theta) = 6. \quad (90)$$

As the first moment is constant with respect to the system parameter θ , the approximation takes the form (59)

$$\begin{aligned} S(\theta) &= \frac{1}{\mu_2^2(\theta)} \frac{\left(\frac{\partial \mu_2(\theta)}{\partial \theta} \right)^2}{(\bar{\mu}_4(\theta) - 1)} = \frac{1}{4\nu^4(\theta)} \frac{\left(4\nu(\theta) \frac{\partial \nu(\theta)}{\partial \theta} \right)^2}{5} \\ &= \frac{4}{5} \frac{1}{\nu^2(\theta)} \left(\frac{\partial \nu(\theta)}{\partial \theta} \right)^2. \end{aligned} \quad (91)$$

In contrast to the other examples, the information bound $S(\theta)$ is loose under the Laplacian system model.

V. APPROXIMATION QUALITY—DISCRETE OUTPUTS

In the following, we extend the discussion on the tightness of $S(\theta)$ to the case where the system output Z takes values from a discrete alphabet \mathcal{Z} .

A. Bernoulli System Model

As a first example for such kind of system outputs, observations from a parametric Bernoulli distribution with

$$p(z = 1; \theta) = 1 - p(z = 0; \theta) = \nu(\theta) \quad (92)$$

are considered, where $0 < \nu(\theta) < 1, \forall \theta \in \Theta$. The Fisher information measure under this model is

$$\begin{aligned} F(\theta) &= \int_{\mathcal{Z}} \left(\frac{\partial \ln p(z; \theta)}{\partial \theta} \right)^2 p(z; \theta) dz \\ &= \sum_{\mathcal{Z}} \left(\frac{\partial p(z; \theta)}{\partial \theta} \right)^2 \frac{1}{p(z; \theta)} = \frac{\left(\frac{\partial p(z=1; \theta)}{\partial \theta} \right)^2}{p(z=1; \theta)} \\ &\quad + \frac{\left(\frac{\partial p(z=0; \theta)}{\partial \theta} \right)^2}{p(z=0; \theta)} = \frac{1}{\nu(\theta)(1-\nu(\theta))} \left(\frac{\partial \nu(\theta)}{\partial \theta} \right)^2. \end{aligned} \quad (93)$$

The mean and the variance are

$$\mu_1(\theta) = \nu(\theta), \quad (94)$$

$$\mu_2(\theta) = \nu(\theta)(1 - \nu(\theta)), \quad (95)$$

with their derivatives

$$\frac{\partial \mu_1(\theta)}{\partial \theta} = \frac{\partial \nu(\theta)}{\partial \theta}, \quad (96)$$

$$\frac{\partial \mu_2(\theta)}{\partial \theta} = (1 - 2\nu(\theta)) \frac{\partial \nu(\theta)}{\partial \theta}. \quad (97)$$

The skewness is

$$\begin{aligned} \bar{\mu}_3(\theta) &= \sum_{\mathcal{Z}} \left(\frac{z - \mu_1(\theta)}{\sqrt{\mu_2(\theta)}} \right)^3 p(z; \theta) \\ &= \frac{1 - 2\nu(\theta)}{\sqrt{\nu(\theta)(1 - \nu(\theta))}} \end{aligned} \quad (98)$$

and the kurtosis

$$\begin{aligned} \bar{\mu}_4(\theta) &= \sum_{\mathcal{Z}} \left(\frac{z - \mu_1(\theta)}{\sqrt{\mu_2(\theta)}} \right)^4 p(z; \theta) \\ &= \frac{1}{\nu(\theta)(1 - \nu(\theta))} - 3. \end{aligned} \quad (99)$$

As

$$\begin{aligned} \frac{\partial \mu_1(\theta)}{\partial \theta} \sqrt{\mu_2(\theta)} \bar{\mu}_3(\theta) &= (1 - 2\nu(\theta)) \frac{\partial \nu(\theta)}{\partial \theta} \\ &= \frac{\partial \mu_2(\theta)}{\partial \theta} \end{aligned} \quad (100)$$

and consequently the optimization of the information bound results in $\beta^*(\theta) = 0$, the approximation takes its simplified form (21)

$$\begin{aligned} S(\theta) &= \frac{1}{\mu_2(\theta)} \left(\frac{\partial \mu_1(\theta)}{\partial \theta} \right)^2 \\ &= \frac{1}{\nu(\theta)(1 - \nu(\theta))} \left(\frac{\partial \nu(\theta)}{\partial \theta} \right)^2. \end{aligned} \quad (101)$$

It becomes clear that for a binary system output Z following a parametric Bernoulli distribution, the derived expression $S(\theta)$ is a tight approximation for the original inference capability $F(\theta)$ given in (93).

B. Poisson System Model

As a second example with discrete output, we consider the Poisson distribution. The samples z at the output Z are distributed according to the model

$$p(z; \theta) = \frac{\nu^z(\theta)}{z!} e^{-\nu(\theta)}, \quad (102)$$

with $\mathcal{Z} = \{0, 1, 2, \dots\}$ and $\nu(\theta) > 0, \forall \theta \in \Theta$. The derivative of the log-likelihood is given by

$$\frac{\partial \ln p(z; \theta)}{\partial \theta} = \frac{z}{\nu(\theta)} \frac{\partial \nu(\theta)}{\partial \theta} - \frac{\partial \nu(\theta)}{\partial \theta}, \quad (103)$$

such that we calculate

$$\begin{aligned} F(\theta) &= \int_{\mathcal{Z}} \left(\frac{\partial \ln p(z; \theta)}{\partial \theta} \right)^2 p(z; \theta) dz \\ &= \int_{\mathcal{Z}} \left(\frac{z}{\nu(\theta)} \frac{\partial \nu(\theta)}{\partial \theta} - \frac{\partial \nu(\theta)}{\partial \theta} \right)^2 p(z; \theta) dz \\ &= \frac{1}{\nu^2(\theta)} \left(\frac{\partial \nu(\theta)}{\partial \theta} \right)^2 \int_{\mathcal{Z}} z^2 p(z; \theta) dz \\ &\quad - \frac{2}{\nu(\theta)} \left(\frac{\partial \nu(\theta)}{\partial \theta} \right)^2 \int_{\mathcal{Z}} z p(z; \theta) dz + \left(\frac{\partial \nu(\theta)}{\partial \theta} \right)^2 \\ &= \frac{1}{\nu(\theta)} \left(\frac{\partial \nu(\theta)}{\partial \theta} \right)^2, \end{aligned} \quad (104)$$

where we have used

$$\int_{\mathcal{Z}} z p(z; \theta) dz = \sum_{z=0}^{\infty} z \frac{\nu^z(\theta)}{z!} e^{-\nu(\theta)} = \nu(\theta), \quad (105)$$

$$\int_{\mathcal{Z}} z^2 p(z; \theta) dz = \sum_{z=0}^{\infty} z^2 \frac{\nu^z(\theta)}{z!} e^{-\nu(\theta)} = \nu(\theta) + \nu^2(\theta). \quad (106)$$

In order to apply the approximation $S(\theta)$, we require the mean and the variance

$$\mu_1(\theta) = \mu_2(\theta) = \nu(\theta), \quad (107)$$

the skewness and the kurtosis

$$\bar{\mu}_3(\theta) = \frac{1}{\sqrt{\nu(\theta)}}, \quad (108)$$

$$\bar{\mu}_4(\theta) = \frac{1}{\nu(\theta)} + 3. \quad (109)$$

As these quantities exhibit the property

$$\frac{\partial \mu_1(\theta)}{\partial \theta} \sqrt{\mu_2(\theta)} \bar{\mu}_3(\theta) = \frac{\partial \nu(\theta)}{\partial \theta} = \frac{\partial \mu_2(\theta)}{\partial \theta}, \quad (110)$$

we obtain $\beta^*(\theta) = 0$ and the approximation for this example

$$S(\theta) = \frac{1}{\mu_2(\theta)} \left(\frac{\partial \mu_1(\theta)}{\partial \theta} \right)^2 = \frac{1}{\nu(\theta)} \left(\frac{\partial \nu(\theta)}{\partial \theta} \right)^2 \quad (111)$$

is tight when comparing it to (104).

VI. APPLICATIONS OF THE INFORMATION BOUND

Finally, we want to outline possible applications of the presented result and the opportunities provided by an information bound like (46). To this end, we present three problems for which $S(\theta)$ provides interesting and useful insights. The problems discussed cover theoretic as well as practical aspects in statistical signal processing.

A. Worst-Case Noise and Minimum Fisher Information

An important question in signal processing is to specify the worst-case noise distribution under the considered system model [28]. A common assumption in the field is that noise affects technical receive systems in an additive way. Therefore a model

of high practical relevance is

$$Z = x(\theta) + W, \quad (112)$$

where $x(\theta)$ is a deterministic pilot signal modulated by the unknown parameter θ (for example attenuation, time-delay, frequency-offset, etc.) and W is additive independent random noise with zero mean

$$\mathbb{E}[W] = 0. \quad (113)$$

If in addition the noise has the property

$$\mathbb{E}[W^2] = \nu, \quad (114)$$

i.e., the second central moment of Z is constant, it is well-understood that assuming the noise component W to follow the Gaussian probability density function

$$p(w) = \frac{1}{\sqrt{2\pi\nu}} e^{-\frac{w^2}{2\nu}}, \quad (115)$$

leads to minimum Fisher information $F(\theta)$ [19], [20]. Therefore, under an estimation theoretic perspective, Gaussian noise is the worst-case assumption for an additive system like (112) with constant second output moment [24]. The presented bounding approach (46) allows us to generalize these statements. If for any system $p(z; \theta)$ (including non-additive systems) the output Z exhibits the characteristic

$$\mu_1(\theta) = \mathbb{E}[Z] = x(\theta), \quad (116)$$

$$\mu_2(\theta) = \mathbb{E}[(Z - \mu_1(\theta))^2] = \nu, \quad (117)$$

the presented result (62) shows that $F(\theta)$ cannot violate

$$F(\theta) \geq \frac{1}{\mu_2(\theta)} \frac{\left(\frac{\partial \mu_1(\theta)}{\partial \theta}\right)^2}{1 - \frac{\bar{\mu}_3^2(\theta)}{(\bar{\mu}_4(\theta) - 1)}}. \quad (118)$$

This lower bound is minimized by a symmetric distribution, i.e., $\bar{\mu}_3(\theta) = 0$. The resulting expression

$$F(\theta) \geq \frac{1}{\mu_2(\theta)} \left(\frac{\partial \mu_1(\theta)}{\partial \theta}\right)^2, \quad (119)$$

reaches equality under an additive Gaussian system model

$$p(z; \theta) = \frac{1}{\sqrt{2\pi\nu}} e^{-\frac{(z-x(\theta))^2}{2\nu}}, \quad (120)$$

such that the worst-case model assumption with respect to Fisher information under the considered restrictions (116) and (117) is, in general, additive and Gaussian. In the more general setting, where also the output variance exhibits a dependency on the system parameter θ ,

$$\mu_1(\theta) = \mathbb{E}[Z] = x(\theta), \quad (121)$$

$$\mu_2(\theta) = \mathbb{E}[(Z - \mu_1(\theta))^2] = \nu(\theta) \quad (122)$$

and additionally the output distribution is symmetric, i.e.,

$$\bar{\mu}_3(\theta) = 0, \quad (123)$$

the presented result allows us to conclude that the Fisher information is in general bounded from below by

$$F(\theta) \geq \frac{1}{\nu(\theta)} \left(\frac{\partial x(\theta)}{\partial \theta}\right)^2 + \frac{1}{\nu^2(\theta)(\bar{\mu}_4(\theta) - 1)} \left(\frac{\partial \nu(\theta)}{\partial \theta}\right)^2. \quad (124)$$

As the system model

$$p(z; \theta) = \frac{1}{\sqrt{2\pi\nu(\theta)}} e^{-\frac{(z-x(\theta))^2}{2\nu(\theta)}} \quad (125)$$

exhibits the inference capability

$$F(\theta) = \frac{1}{\nu(\theta)} \left(\frac{\partial x(\theta)}{\partial \theta}\right)^2 + \frac{1}{2\nu^2(\theta)} \left(\frac{\partial \nu(\theta)}{\partial \theta}\right)^2, \quad (126)$$

by comparing (124) and (126) it can be concluded that for all cases where

$$\bar{\mu}_4(\theta) \leq 3, \quad (127)$$

a conservative system model $p(z; \theta)$ under an estimation theoretic perspective is the Gaussian model (125).

B. Information Loss of the Squaring Device

Another interesting problem in statistical signal processing is to characterize the estimation theoretic quality of nonlinear receive and measurement systems. The Fisher information measure $F(\theta)$ is a rigorous tool which allows us to draw precise conclusions. However, depending on the nature of the nonlinearity, the exact calculation of the information measure $F(\theta)$ can become complicated. As an example for such a scenario, consider the problem of analyzing the intrinsic capability of a system with a squaring output (power sensor)

$$Z = Y^2, \quad (128)$$

to infer the mean θ of a Gaussian input

$$p(y; \theta) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(y-\theta)^2}{2}} \quad (129)$$

with unit variance. In such a case, the system output Z follows a non-central chi-squared distribution with a single degree of freedom parameterized by θ . As the analytical description of the associated probability density function $p(z; \theta)$ includes a Bessel function, the characterization of the Fisher information $F(\theta)$ in compact analytical form is not trivial. We short-cut the derivation by using the approximation (47). The first two moments of the output (128) are found to be given by

$$\begin{aligned} \mathbb{E}[Z] &= \mathbb{E}[\theta^2 + 2\theta W + W^2] \\ &= \theta^2 + 1 = \mu_1(\theta), \end{aligned} \quad (130)$$

$$\begin{aligned} \mathbb{E}[(Z - \mu_1(\theta))^2] &= \mathbb{E}[(\theta^2 + 2\theta W + W^2 - \theta^2 - 1)^2] \\ &= 2(2\theta^2 + 1) = \mu_2(\theta), \end{aligned} \quad (131)$$

where we have introduced the auxiliary random variable

$$W = Y - \theta. \quad (132)$$

The third output moment is

$$\begin{aligned} E[(Z - \mu_1(\theta))^3] &= E[(\theta^2 + 2\theta W + W^2 - \theta^2 - 1)^3] \\ &= 8(3\theta^2 + 1) = \mu_3(\theta), \end{aligned} \quad (133)$$

while the fourth moment is

$$\begin{aligned} E[(Z - \mu_1(\theta))^4] &= E[(\theta^2 + 2\theta W + W^2 - \theta^2 - 1)^4] \\ &= 12((2\theta^2 + 1)^2 + 4(4\theta^2 + 1)) \\ &= \mu_4(\theta). \end{aligned} \quad (134)$$

Therefore, the skewness and the kurtosis are

$$\begin{aligned} \bar{\mu}_3(\theta) &= \mu_3(\theta)\mu_2^{-\frac{3}{2}}(\theta) \\ &= \frac{8(3\theta^2 + 1)}{2\sqrt{2}(2\theta^2 + 1)^{\frac{3}{2}}} = \frac{2\sqrt{2}(3\theta^2 + 1)}{(2\theta^2 + 1)^{\frac{3}{2}}}, \end{aligned} \quad (135)$$

$$\begin{aligned} \bar{\mu}_4(\theta) &= \mu_4(\theta)\mu_2^{-2}(\theta) = \frac{12((2\theta^2 + 1)^2 + 4(4\theta^2 + 1))}{4(2\theta^2 + 1)^2} \\ &= \frac{12(4\theta^2 + 1)}{(2\theta^2 + 1)^2} + 3. \end{aligned} \quad (136)$$

With the derivatives

$$\frac{\partial \mu_1(\theta)}{\partial \theta} = 2\theta, \quad (137)$$

$$\frac{\partial \mu_2(\theta)}{\partial \theta} = 8\theta, \quad (138)$$

we obtain

$$\begin{aligned} \beta^*(\theta) &= \frac{\frac{\partial \mu_1(\theta)}{\partial \theta} \sqrt{\mu_2(\theta)} \bar{\mu}_3(\theta) - \frac{\partial \mu_2(\theta)}{\partial \theta}}{\frac{\partial \mu_2(\theta)}{\partial \theta} \bar{\mu}_3(\theta) - \frac{\partial \mu_1(\theta)}{\partial \theta} \sqrt{\mu_2(\theta)} (\bar{\mu}_4(\theta) - 1)} \\ &= -\frac{\theta^2 \sqrt{2} \sqrt{(2\theta^2 + 1)}}{(4\theta^4 + 16\theta^2 + 3)} \end{aligned} \quad (139)$$

and the approximation is finally given by

$$\begin{aligned} S(\theta) &= \frac{1}{\mu_2(\theta)} \frac{\left(\frac{\partial \mu_1(\theta)}{\partial \theta} + \frac{\beta^*(\theta)}{\sqrt{\mu_2(\theta)}} \frac{\partial \mu_2(\theta)}{\partial \theta} \right)^2}{1 + 2\beta^*(\theta) \bar{\mu}_3(\theta) + \beta^{*2}(\theta) (\bar{\mu}_4(\theta) - 1)} \\ &= \frac{2\theta^2 (4\theta^4 + 12\theta^2 + 3)^2}{(4\theta^4 + 12\theta^2 + 3)(8\theta^6 + 24\theta^4 + 18\theta^2 + 3)} \\ &= \frac{2\theta^2 (4\theta^4 + 12\theta^2 + 3)}{(8\theta^6 + 24\theta^4 + 18\theta^2 + 3)}. \end{aligned} \quad (140)$$

Fig. 1 depicts the conservative approximation

$$\tilde{\chi}(\theta) = \frac{S_Z(\theta)}{F_Y(\theta)} \quad (141)$$

of the information loss (18) when squaring the random input variable Y . Note that Fig. 1 indicates that for small values of θ the squaring operation results in a strong degradation of the estimation capability. As a comparison, the corresponding loss for a symmetric hard-limiter [29]

$$Z = \text{sign}(Y) \quad (142)$$

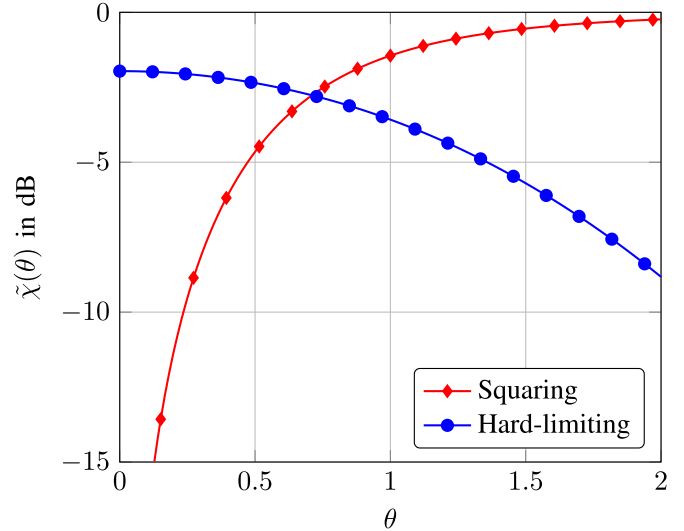


Fig. 1. Performance loss of two nonlinear systems.

is visualized in Fig. 1. Note that for hard-limiting the Gaussian model (129), it was shown in [18] that (21) forms a tight lower bound for the Fisher information measure. It can be observed that for small values of θ , the information about the algebraic sign (hard-limiting) of the system input Y conveys more information about the input mean θ than the amplitude (squaring). For $\theta \geq 0.75$, the situation changes as the statistics of the hard-limiter output vary slower with the parameter θ and therefore the squaring receiver outperforms the hard-limiter when it comes to estimating the mean θ of the system input Y from samples at the system output Z . Note that for the squaring device (128), Fig. 1 depicts a conservative approximation (141) of the exact squaring loss (18).

C. Measuring the Inference Capability With a Smooth Limiter

A situation that is often encountered in practice is that the analytical characterization of the system model $p(z; \theta)$ is difficult. If the appropriate system model $p(z; \theta)$ is unknown [30], the direct consultation of an analytical tool like the Fisher information measure $F(\theta)$ becomes impossible. However, in such a situation, an information bound like $S(\theta)$ allows us to numerically approximate the information measure $F(\theta)$ at low-complexity. To this end, the moments of the system output Z are measured in a calibrated setup, where the parameter θ can be controlled, or determined by Monte-Carlo simulations. We demonstrate this validation technique by using a smooth limiter model, i.e., the system input Y is transformed by

$$Z = \sqrt{\frac{2}{\pi\zeta^2}} \int_0^Y e^{-\frac{u^2}{2\zeta^2}} du = \text{erf} \left(\frac{Y}{\sqrt{2\zeta^2}} \right), \quad (143)$$

where $\zeta \in \mathbb{R}$ is a constant model parameter and

$$\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt \quad (144)$$

is the error function. This nonlinear model [31] can be used in order to characterize saturation effects in analog system com-

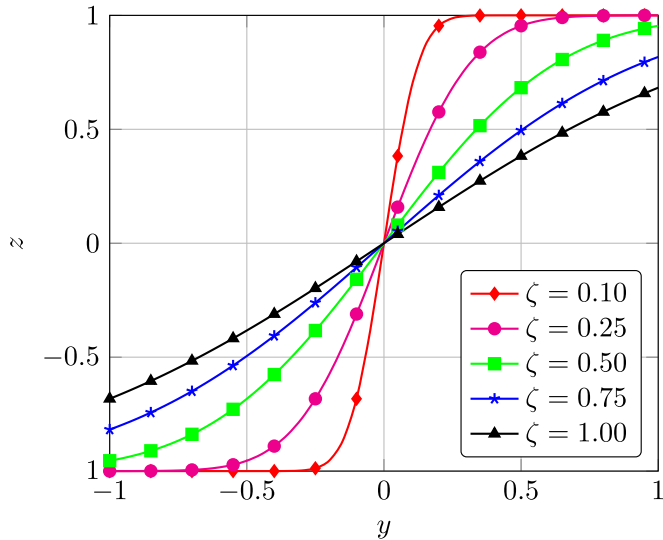


Fig. 2. Input-to-output relation of the smooth limiter.

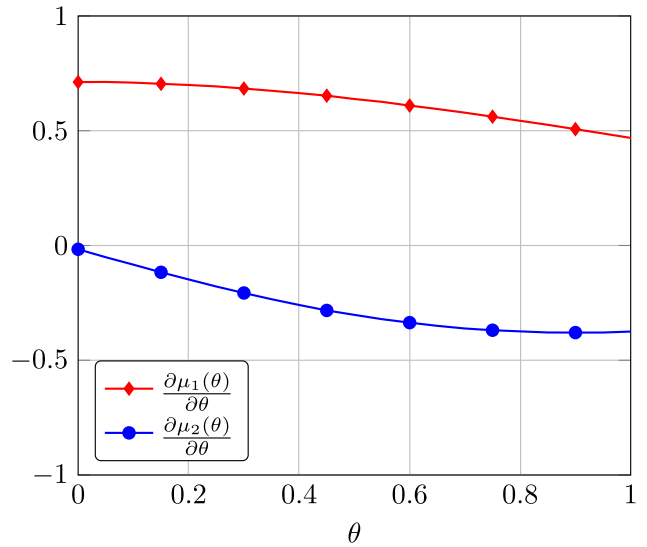
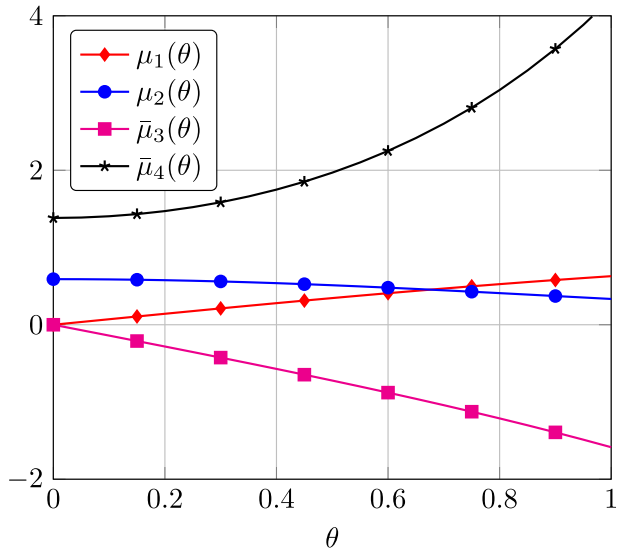
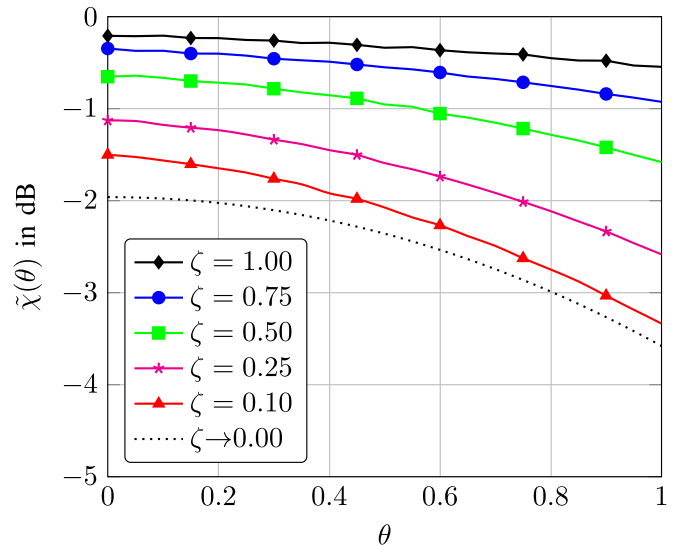
Fig. 4. Measured derivatives ($\zeta = 0.5$) of the smooth limiter.Fig. 3. Measured moments ($\zeta = 0.5$) of the smooth limiter.

Fig. 5. Measured information loss of the smooth limiter.

ponents like amplifiers [32], [33]. In Fig. 2, the input-to-output mapping of the model (143) is depicted for different setups ζ . As input, we consider a Gaussian distribution with unit variance like in (129). The output mean $\mu_1(\theta)$, variance $\mu_2(\theta)$, skewness $\bar{\mu}_3(\theta)$, and kurtosis $\bar{\mu}_4(\theta)$ are measured by 10^9 independent Monte-Carlo simulations of the nonlinear system output Z for each considered value of the input mean θ . The result is shown in Fig. 3. After numerically approximating the required derivatives $\frac{\partial \mu_1(\theta)}{\partial \theta}$, $\frac{\partial \mu_2(\theta)}{\partial \theta}$, which are depicted in Fig. 4, the approximation $S(\theta)$ is calculated.

In Fig. 5, the measured information loss $\tilde{\chi}(\theta)$ of the smooth limiter model is shown, where the dotted line indicates the exact information loss $\chi(\theta)$ with a hard-limiter (142) (as depicted in Fig. 1), which is equivalent to a smooth limiter with $\zeta \rightarrow 0$.

VII. CONCLUSION

We have established a generic and compact lower bound for the Fisher information measure. By various examples we have shown that the derived expression has the potential to provide a good approximation in a broad number of cases. This makes the presented information bound a versatile mathematical tool for a variety of problems encountered in the design and optimization of signal processing systems. Further, the pessimistic nature of the attained alternative information measure allows us to strengthen insights on worst-case noise and to generalize classical results on Gaussian system models which exhibit minimum Fisher information. Finally, we have outlined how to use the presented information bound in order to benchmark the estimation capability of physical measurement systems with output statistics of unknown analytical form.

REFERENCES

- [1] R. A. Fisher, "On the mathematical foundations of theoretical statistics," *Philosoph. Trans. Roy. Soc. London*, vol. 222, pp. 309–368, Jan. 1922.
- [2] R. A. Fisher, "Theory of statistical estimation," *Proc. Cambridge Philos. Soc.*, vol. 22, no. 5, pp. 700–725, Jul. 1925.
- [3] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ, USA: Prentice-Hall, 1993.
- [4] C. R. Rao, "Information and accuracy attainable in the estimation of statistical parameters," *Bull. Calcutta Math. Soc.*, vol. 37, no. 3, pp. 81–91, 1945.
- [5] H. Cramér, *Mathematical Methods of Statistics*. Princeton, NJ, USA: Princeton Univ. Press, 1946.
- [6] A. Bhattacharyya, "On some analogues of the amount of information and their use in statistical estimation," *Sankhya: Indian J. Statist.*, vol. 8, no. 1, pp. 1–14, 1946.
- [7] E. W. Barankin, "Locally best unbiased estimates," *Ann. Math. Statist.*, vol. 22, no. 4, pp. 477–501, 1949.
- [8] J. M. Hammersley, "On estimating restricted parameters," *J. Roy. Statist. Soc. vol. 12*, no. 2, pp. 192–240, 1950.
- [9] D. G. Chapman and H. Robbins, "Minimum variance estimation without regularity assumptions," *Ann. Math. Stat.*, vol. 22, no. 4, pp. 581–586, 1951.
- [10] H. Kobayashi, B. L. Mark, and W. Turin, *Probability, Random Processes, and Statistical Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 2011.
- [11] H. L. Van Trees, *Detection, Estimation, and Modulation Theory, Part I*, New York, NY, USA: Wiley, 1968.
- [12] J. Ziv and M. Zakai, "Some lower bounds on signal parameter estimation," *IEEE Trans. Inf. Theory*, vol. 15, no. 3, pp. 386–391, May 1969.
- [13] B. Bobrovsky and M. Zakai, "A lower bound on the estimation error for Markov processes," *IEEE Trans. Autom. Control*, vol. 20, no. 6, pp. 785–788, Dec. 1975.
- [14] A. J. Weiss and E. Weinstein, "A lower bound on the mean-square error in random parameter estimation," *IEEE Trans. Inf. Theory*, vol. 31, no. 5, pp. 680–682, Sep. 1985.
- [15] P. Tichavský, C. H. Muravchik, and A. Nehorai, "Posterior Cramer-Rao bounds for discrete-time nonlinear filtering," *IEEE Trans. Signal Process.*, vol. 46, no. 5, pp. 1386–1396, May 1998.
- [16] H. L. Van Trees and K. L. Bell, *Bayesian Bounds for Parameter Estimation and Nonlinear Filtering/Tracking*, 1st ed. Piscataway, NJ, USA: Wiley, 2007.
- [17] M. Sankaran, "On an analogue of Bhattacharya bound," *Biometrika*, vol. 51, no. 1/2, pp. 268–270, 1964.
- [18] M. Stein, A. Mezghani, and J. A. Nossek, "A lower bound for the Fisher information measure," *IEEE Signal Process. Lett.*, vol. 21, no. 7, pp. 796–799, Jul. 2014.
- [19] A. J. Stam, "Some mathematical properties of quantities of information," Ph.D. dissertation, Delft Univ. Technol., Delft, Netherlands, 1959.
- [20] D. E. Boeke, "A generalization of the Fisher information measure," Ph.D. dissertation, Delft Univ. Technol., Delft, Netherlands, 1977.
- [21] E. Uhrmann-Klingen, "Minimal Fisher information distributions with compact-supports," *Sankhyā: Indian J. Statist.*, vol. 57, no. 3, pp. 360–374, Oct. 1995.
- [22] V. Živojnović, "Minimum Fisher information of moment-constrained distributions with application to robust blind identification," *Elsevier Signal Process.*, vol. 65, no. 2, pp. 297–313, Oct. 1998.
- [23] J. F. Bercher and C. Vignat, "On minimum Fisher information distributions with restricted support and fixed variance," *Inf. Sci.*, vol. 179, no. 22, pp. 3832–3842, Nov. 2009.
- [24] P. Stoica and P. Babu, "The Gaussian data assumption leads to the largest Cramèr-Rao bound," *IEEE Signal Process. Mag.*, vol. 28, no. 3, pp. 132–133, May 2011.
- [25] K. Pearson, "Mathematical contributions to the theory of evolution. XIX. Second supplement to a Memoir on Skew variation," *Philos. Trans. Roy. Soc. (A)*, vol. 216, pp. 429–457, 1916.
- [26] J. E. Wilkins, "A note on skewness and kurtosis," *Ann. Math. Statist.*, vol. 15, no. 3, pp. 333–335, 1944.
- [27] R. G. Jarrett, "Bounds and expansions for Fisher information when the moments are known," *Biometrika* vol. 71, no. 1, pp. 101–113, 1984.
- [28] P. J. Huber, "Robust statistics: A review," *Ann. Math. Statist.*, vol. 43, no. 4, pp. 1041–1067, 1972.
- [29] J. H. Van Vleck and D. Middleton, "The spectrum of clipped noise," *Proc. IEEE*, vol. 54, no. 1, pp. 2–19, Jan. 1966.
- [30] V. Berisha and A. O. Hero, "Empirical non-parametric estimation of the Fisher information," *IEEE Signal Process. Lett.*, vol. 22, no. 7, pp. 988–992, Jul. 2015.
- [31] R. Baum, "The correlation function of smoothly limited Gaussian noise," *IRE Trans. Inf. Theory*, vol. 3, no. 3, pp. 193–197, Sep. 1957.
- [32] R. F. Pawula, T. S. Fong, and M. R. O'Sullivan, "Intermodulation distortion in frequency multiplexed satellite repeaters," *Proc. IEEE*, vol. 59, no. 2, pp. 213–218, Feb. 1971.
- [33] A. J. Cann, "Nonlinearity model with variable knee sharpness," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 16, no. 6, pp. 874–877, Nov. 1980.



Manuel S. Stein (S'12) received the B.Sc., Dipl.-Ing., and Dr.-Ing. degrees in electrical engineering and information technology from the Technische Universität München (TUM), Germany, in 2009, 2010, and 2016, respectively. From 2011 to 2016, he has been a Research Associate in the Institute of Circuit Theory and Signal Processing (NWS), TUM, while serving as a Teaching Assistant in the Department of Electrical and Computer Engineering, TUM. In 2016, he was with the Signal and System Theory Group, Universität Paderborn, Germany, before joining the

Digital Mathematics Group (DIMA), Vrije Universiteit Brussel, Belgium, as a Postdoctoral Researcher. He is the first recipient of the DAAD/MSCA P.R.I.M.E. fellowship in the electrical engineering sciences. His research focuses on the fundamentals of statistical signal processing with low-complexity system design and its application in wireless synchronization, array processing, communication, spectrum monitoring and radar.



Josef A. Nossek (S'72–M'74–SM'81–F'93–LF'13) received the Dipl.-Ing. and Dr. techn. degrees in electrical engineering from Vienna University of Technology, Austria, in 1974 and 1980, respectively. He joined Siemens AG, Germany, in 1974, where he was engaged in filter design for communication systems. From 1987 to 1989, he was the Head of the Radio Systems Design Department, where he was instrumental in introducing high-speed VLSI signal processing into digital microwave radio. From 1989 to 2016, he has been the Head of the Institute of Circuit

Theory and Signal Processing (NWS) at the Technische Universität München, Germany. In 2016, he joined the Universidade Federal do Ceará, Fortaleza, Brasil as a Full Professor. He was the President Elect, President, and Past President of the IEEE Circuits and Systems Society in 2001, 2002, and 2003, respectively. He was the President of Verband der Elektrotechnik, Elektronik, und Informationstechnik (VDE) in 2007 and the President of the Convention of National Associations of Electrical Engineers of Europe (EUREL) in 2013. He received the ITG Best Paper Award in 1988, the Mannesmann Mobilfunk (currently Vodafone) Innovations Award in 1998, and the Award for Excellence in Teaching from the Bavarian Ministry for Science, Research and Art in 1998. From the IEEE Circuits and Systems Society, he received the Golden Jubilee Medal for Outstanding Contributions to the Society in 1999 and the Education Award in 2008. He received the Order of Merit of the Federal Republic of Germany (Bundesverdienstkreuz am Bande) in 2008. In 2011, he received the IEEE Guillemin-Cauer Best Paper Award and in 2013 the honorary doctorate (Dr. h. c.) from the Peter Pazmany Catholic University, Hungary. He received the VDE Ring of Honor in 2014 and the TUM Emeritus of Excellence in 2016.