Design of Unitless Normalized Measure of Nonlinearity for State Estimation

Ondřej Straka*, Jindřich Havlík*,

* Dept. of Cybernetics, & European Centre of Excellence NTIS
Faculty of Applied Sciences, Univ. of West Bohemia, Czech Republic
E-mails: {straka30,havlikj}@kky.zcu.cz

Abstract—The paper deals with measures of nonlinearity. In state estimation, they are utilized *i)* to select a suitable state estimation algorithm by assessing the nonlinearity of a system model, *ii)* to adapt the estimation algorithm structure or parameters, or *iii)* to indicate the possible effect of strong nonlinearity that leads to estimate credibility loss. This paper summarizes the state of the art of nonlinearity measures, focusing on the mean-square-error-based measure of nonlinearity. Its weak point is illustrated, and based on this, requirements for the new measure of nonlinearity are formulated. A new nonlinearity measure that is both unitless and normalized is designed. Its properties are demonstrated using numerical tracking experiments.

Index Terms—state estimation, nonlinearity measure, tracking

I. INTRODUCTION

Mathematical modeling of systems aims to describe the behavior of the system under study so that the behavior of individual components of the system can be better understood and predictions about future behavior can be made. The simplest models exhibit proportionality, i.e., they are linear. Linear problems can be solved efficiently and quickly. Unfortunately, linear behavior is idealized, and models of real systems are not purely linear. Moreover, the relationship between the variables often exhibits a character of uncertainty. It is not difficult to recognize that the model or its component behavior is not linear, but the question is: By how much? Is a linear approximation too simplifying? Do linear methods provide a viable approach? To answer these questions, one needs to be able to quantify the amount of deviation from linear behavior. Given the importance of these issues, it is not surprising that methods capable of quantifying nonlinearity are found in every field employing a mathematical approach to modeling. In the sequel, a method quantifying the nonlinearity will be called a measure of nonlinearity (MoN).

In 1960, Beale [1] was the first one who theoretically dealt with MoNs and proposed empirical and theoretical MoNs based on the difference between the nonlinear function and the Taylor expansion for the estimation of static non-random parameters. They were intended to indicate the accuracy of the approximate confidence intervals in regression analysis and to seek a linear approximation of the model. Many reactions investigating the properties and behavior of these MoNs appeared, such as [2], [3], which pointed out the inappropriate behavior of the MoNs for strongly nonlinear problems and proposed suitable modifications. Based on principles from

differential geometry, the so-called curvature[1] MoNs [4] were proposed for estimating a static parameter.

In nonlinear deterministic dynamic system control, MoNs are used to investigate whether the linear control theory can be used. For this purpose, MoNs based on functional norms of the difference between nonlinearity and the best linear approximation [5], [6], MoNs using vector space norms of the output space along the system trajectory [7], or gap metric [8] were designed. MoNs are also applied in other fields such as time series [9], [10], clinical biochemistry [11], or econometrics [12].

In the last few years, the MoNs have been the focus of intensive development in the state estimation of stochastic dynamical systems. However, the task of state estimation is challenging for MoNs for several reasons: *i)* In addition to a static measurement model, the system models are also composed of an equation of state dynamics. *ii)* The state is a random vector, and the nonlinearity has to be investigated in a specific region. *iii)* Noise is an inherent part of stochastic models and has cumulative effects.

Differential MoNs were used to answer the question of whether the state should be modeled in a Cartesian or a polar coordinate system with radar [13], [14] or video measurements [15]. MoNs were used to decide whether a simple or more complex nonlinear filtering algorithm should be used [16]. MoNs based on state density description by a mixture of Dirac functions were used to adapt nonlinear filter parameters to account for nonlinearity effects [17]. Several MoNs have been designed benefiting from the fact that the linear transformations preserve the Gaussian distribution of the transformed variable. These MoNs quantify the nonlinearity by the deviation of a variable density from the Gaussian distribution [18]. To this end, they use differential entropy, Kullback-Leibler divergence, various statistical tests, or higher-order moments. There, the MoNs are utilized to adapt the filter structure parameters or to provide a filter with self-assessment of estimate consistency. The paper [19] proposed a unique MoN based on minimizing the mean square error (MSE) between a nonlinear model influenced by a generally non-additive noise and its best linear approximation with an additive noise. Compared with other MoNs, the MSE-based MoN offers many favorable

---

[1]They assess deviation of the nonlinear function from a tangent (hyper)plane.

properties, such as normalization, simultaneous assessment of state dynamics and measurement relations, and consideration of generally non-additive noises.

The MSE-based MoN was shown in [20] to provide counterintuitive results for different nonlinear estimation problems, as the MoN value was inconsistent with experts' expectations. In addition, as will be shown in the motivation example, the MSE-based MoN behavior is highly sensitive to the choice of the variable units, which may be critical, especially for the MoN used in assessing the nonlinearity of different models. Therefore, the paper aims to identify the shortcomings of the MoN, revise the requirements for MoNs in relation to the task model nonlinearity assessment, and design a new MoN that does not suffer from these deficiencies.

The rest of the paper is organized as follows. Section II defines the MoN and describes the MSE-based MoN introduced in [19]. The requirements for the new MoN are particularized in Section III, and the new unitless MoN is designed in Section IV. The MoN is numerically illustrated in Section V, and concluding remarks are given in Section VI.

II. PROBLEM STATEMENT AND MSE-BASED MoN

Consider a nonlinear vector transformation $\mathbf{g} : \mathbb{R}^{n_u} \times \mathbb{R}^{n_v} \mapsto \mathbb{R}^{n_y}$ from the space of random variable $\mathbf{u} \in \mathbb{R}^{n_u}$ and noise $\mathbf{v} \in \mathbb{R}^{n_v}$ to the space of random variable $\mathbf{y} \in \mathbb{R}^{n_y}$

$$\mathbf{y} = \mathbf{g}(\mathbf{u}, \mathbf{v}) \tag{1}$$

with known probability density functions (PDFs) $p(\mathbf{u})$ and $p(\mathbf{v})$. The noise $\mathbf{v}$ is assumed without loss of generality zero-mean[2] and independent of $\mathbf{u}$. In discrete-time state space models

$$\mathbf{x}_{k+1} = \mathbf{f}_k(\mathbf{x}_k, \mathbf{w}_k) \tag{2a}$$

$$\mathbf{z}_k = \mathbf{h}_k(\mathbf{x}_k) + \mathbf{v}_k, \tag{2b}$$

the transformation (1) may represent the state $\mathbf{x}_k$ dynamics (2a), the relation (2b) between the state and the measurement $\mathbf{z}_k$, or both if they are stacked. Note that variables $\mathbf{w}_k$ and $\mathbf{v}_k$ represent the process and measurement noises, respectively.

The transformation (1) is the most general, with the noise $\mathbf{v}$ acting non-additively. Often, one is faced with a special case where the noise acts multiplicatively

$$\mathbf{y} = \mathbf{g}(\mathbf{u}, \mathbf{v}) = \mathbf{f}(\mathbf{u}) + \boldsymbol{\pi}(\mathbf{u})\boldsymbol{\gamma}(\mathbf{v}), \tag{3}$$

or additively

$$\mathbf{y} = \mathbf{g}(\mathbf{u}, \mathbf{v}) = \mathbf{f}(\mathbf{u}) + \mathbf{v}, \tag{4}$$

i.e., $n_v = n_y$ and $\mathbf{f} : \mathbb{R}^{n_u} \mapsto \mathbb{R}^{n_y}$. The term $\boldsymbol{\gamma}(\mathbf{v})$ in (3) is assumed zero mean without loss of generality, i.e., $\mathsf{E}[\boldsymbol{\gamma}(\mathbf{v})] = \mathbf{0}$. In the sequel, the following notation for covariance matrices is used: $\boldsymbol{\Sigma}_{\mathbf{gg}} := \mathrm{var}[\mathbf{g}(\mathbf{u}, \mathbf{v})]$, $\boldsymbol{\Sigma}_{\mathbf{uu}} := \mathrm{var}[\mathbf{u}]$, $\boldsymbol{\Sigma}_{\mathbf{gu}} := \mathrm{cov}[\mathbf{g}(\mathbf{u}, \mathbf{v}), \mathbf{u}]$, $\boldsymbol{\Sigma}_{\mathbf{ug}} = \boldsymbol{\Sigma}_{\mathbf{gu}}^{\mathsf{T}}$, $\boldsymbol{\Sigma}_{\mathbf{ff}} := \mathrm{var}[\mathbf{f}(\mathbf{u})]$, $\boldsymbol{\Sigma}_{\mathbf{fu}} := \mathrm{cov}[\mathbf{f}(\mathbf{u}), \mathbf{u}]$, $\boldsymbol{\Sigma}_{\mathbf{uf}} = \boldsymbol{\Sigma}_{\mathbf{fu}}^{\mathsf{T}}$, $\boldsymbol{\Sigma}_{\boldsymbol{\gamma\gamma}} := \mathrm{var}[\boldsymbol{\gamma}(\mathbf{u})]$.

---

[2]In the case of a nonzero-mean noise, the mean can be embedded into the function $\mathbf{g}$.

The MoNs measure the deviation of (1) (or (3) or (4)) from linear[3] behavior. The deviation can be assessed globally within the whole domain $\mathbb{R}^{n_u} \times \mathbb{R}^{n_v}$ of $\mathbf{g}$ or locally within a subset of the domain, where the subset is often determined by the PDFs $p(\mathbf{u})$ and $p(\mathbf{v})$.

Assume that the space of all mappings (1) is denoted by $\mathcal{G}$ and the space of the cumulative distribution functions (CDFs) corresponding to $p(\mathbf{u})$ and $p(\mathbf{v})$ is denoted by $\mathcal{D}$. Then, the MoN can be seen as a functional

$$M : \mathcal{G} \times \mathcal{D} \times \mathcal{D} \mapsto \mathbb{R}_0^+, \tag{5}$$

which for a linear function $\mathbf{g}$ the functional returns zero. Its value increases as $\mathbf{g}$ departs from linear behavior.

A. MSE-based MoN

The MoN based on the MSE was introduced in [21] to quantify the nonlinearity of stochastic models in the form $\mathbf{y} = \mathbf{f}(\mathbf{u})$. Later, it was generalized [19] to mappings in the form (1). The MoN was based on the MSE of a difference between the nonlinear transformation and its best linear approximation.

$$M = \sqrt{\min_{\mathbf{A}, \mathbf{b}, \psi_{\mathbf{v}, \mathbf{n}}} \mathsf{E}[\|\mathbf{g}(\mathbf{u}, \mathbf{v}) - \mathbf{L}(\mathbf{u}, \mathbf{n})\|_2^2]} \tag{6}$$

where the expectation is done over all involved random variables, i.e., $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{n}$ and the linear approximation $\mathbf{L}(\mathbf{u}, \mathbf{n})$ is of the form

$$\mathbf{L}(\mathbf{u}, \mathbf{n}) = \mathbf{A}\mathbf{u} + \mathbf{b} + \mathbf{n}, \tag{7}$$

with $\mathbf{A}$ and $\mathbf{b}$ being matrix and vector of corresponding dimensions and $\mathbf{n}$ being a zero-mean noise. The optimization in (6) is thus carried out over $\mathbf{A}$, $\mathbf{b}$, and the joint distribution $\psi_{\mathbf{v}, \mathbf{n}}$.

Also, a normalized version of (6) was introduced in [19] in the form

$$M^{\mathrm{norm}} = \frac{M}{\sqrt{\mathrm{tr}(\boldsymbol{\Sigma}_{\mathbf{gg}})}} \tag{8}$$

such that $M^{\mathrm{norm}} \in [0, 1]$. It was shown [21] that for (1) the optimal values of $\mathbf{A}$ and $\mathbf{b}$ are

$$\mathbf{A} = \boldsymbol{\Sigma}_{\mathbf{gu}}\boldsymbol{\Sigma}_{\mathbf{uu}}^{-1} \tag{9}$$

$$\mathbf{b} = \mathsf{E}[\mathbf{y}] - \mathbf{A}\mathsf{E}[\mathbf{u}]. \tag{10}$$

For the general noise case (1), the MoN cannot be simplified further and equals to

$$M = \sqrt{\begin{aligned}&\mathrm{tr}(\boldsymbol{\Sigma}_{\mathbf{gg}} - \boldsymbol{\Sigma}_{\mathbf{gu}}\boldsymbol{\Sigma}_{\mathbf{uu}}^{-1}\boldsymbol{\Sigma}_{\mathbf{ug}}) + \\ &+ \min_{\psi_{\mathbf{v}, \mathbf{n}}}(\mathsf{E}\|\mathbf{n}\|_2^2 - 2\mathsf{E}[\mathbf{g}(\mathbf{u}, \mathbf{v})^{\mathsf{T}}\mathbf{n}])\end{aligned}}. \tag{11}$$

For the multiplicative noise case (3), the MoN equals to

$$M = \sqrt{\mathrm{tr}(\boldsymbol{\Sigma}_{\mathbf{ff}} - \boldsymbol{\Sigma}_{\mathbf{fu}}\boldsymbol{\Sigma}_{\mathbf{uu}}^{-1}\boldsymbol{\Sigma}_{\mathbf{uf}}) + \mathrm{tr}(\mathsf{E}[\tilde{\boldsymbol{\pi}}^{\mathsf{T}}\tilde{\boldsymbol{\pi}}\boldsymbol{\Sigma}_{\boldsymbol{\gamma\gamma}}])} \tag{12}$$

and for the additive noise case (4), the MoN simplifies to

$$M = \sqrt{\mathrm{tr}(\boldsymbol{\Sigma}_{\mathbf{ff}} - \boldsymbol{\Sigma}_{\mathbf{fu}}\boldsymbol{\Sigma}_{\mathbf{uu}}^{-1}\boldsymbol{\Sigma}_{\mathbf{uf}})}. \tag{13}$$

---

[3]Mathematically strictly speaking, the MoNs often measure deviation from affine behavior rather than from linear behavior.
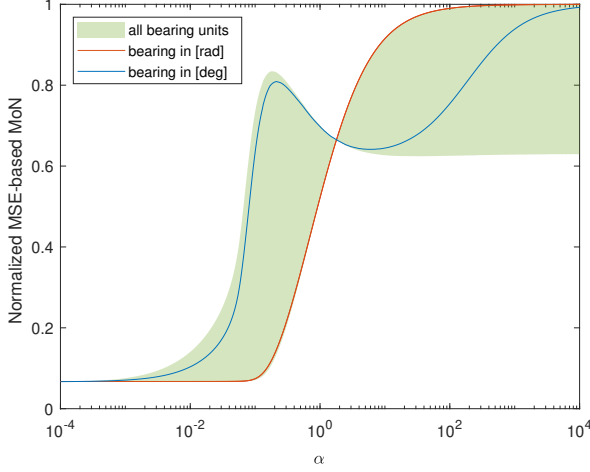
Figure 1. Effect of various bearing units on the MSE-based MoN.

B. Motivation Example

Consider the problem of transformation of a Gaussian distributed position from Cartesian to polar coordinates, i.e., $\mathbf{u} = [\mathsf{x}, \mathsf{y}]^\mathsf{T}$ given by

$$\begin{bmatrix} \mathsf{x} \\ \mathsf{y} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 1 \\ 10 \end{bmatrix} [\text{km}], \alpha \cdot \begin{bmatrix} 1 & 0 \\ 0 & 100 \end{bmatrix} [\text{km}^2] \right), \quad (14)$$

is transformed through either

$$\mathbf{y} = \mathbf{f}(\mathbf{u}) = \begin{bmatrix} \sqrt{\mathsf{x}^2 + \mathsf{y}^2} \\ \operatorname{atan2}(\mathsf{y}, \mathsf{x}) \frac{180}{\pi} \end{bmatrix} \begin{bmatrix} [\text{km}] \\ [^\circ] \end{bmatrix} \quad (15)$$

or

$$\mathbf{y} = \mathbf{f}(\mathbf{u}) = \begin{bmatrix} \sqrt{\mathsf{x}^2 + \mathsf{y}^2} \\ \operatorname{atan2}(\mathsf{y}, \mathsf{x}) \end{bmatrix} \begin{bmatrix} [\text{km}] \\ [\text{rad}] \end{bmatrix} \quad (16)$$

The transformations (15) and (16) differ only in the units for the bearing.

The values of α in (14) will be selected in the interval $\alpha \in [10^{-4}, 10^2]$. Changing the values of α facilitates the analysis of the MoN since the parameter affects the region over which the nonlinearity of $\mathbf{f}$ is studied. The region is small for small values of α, and the function will be close to linear in such a region. Thus, changing α indirectly changes the nonlinearity. Changing the nonlinearity of $\mathbf{f}$ directly is difficult.

The normalized MSE-based MoN [19]

$$M^{\text{norm}} = \frac{\sqrt{\operatorname{tr}(\boldsymbol{\Sigma}_{\mathbf{ff}} - \boldsymbol{\Sigma}_{\mathbf{fu}}\boldsymbol{\Sigma}_{\mathbf{uu}}^{-1}\boldsymbol{\Sigma}_{\mathbf{uf}})}}{\sqrt{\operatorname{tr}(\boldsymbol{\Sigma}_{\mathbf{ff}})}} \quad (17)$$

was calculated for both (15) and (16) and is shown in Figure 1. The lines for (16) and (15) indicate huge difference between the MoN values for some α (for $\alpha = 0.17$ the MoN for bearing in radians is $M^{\text{norm}} = 10\%$ while it is $M^{\text{norm}} = 80\%$ for bearing in degrees). In addition, Figure 1 also depicts the area of MoN values for arbitrary linear change of bearing units, (i.e., from units equivalent to $10\,\text{Grad}$ to units equivalent to $0.1\,\text{nrad}$). To summarize, the illustration example indicates a significant dependence of the MSE-based MoN on the units of $\mathbf{y}$, which motivates the design of a unitless MoN.

III. REQUIREMENTS FOR MoN

Motivated by the illustration and the discussion above, the designed MoN should possess the following properties:

P1: The MoN should be unitless, i.e., its value should be invariant to a change of units of $\mathbf{y}$ and $\mathbf{u}$.

P2: The MoN should be normalized, i.e., provide values in $[0, 1]$.

The property P1 can expressed through the invariance of MoN to a change of units. The change of units of quantities $\mathbf{u}$, $\mathbf{y}$, and $\mathbf{v}$ is assumed to be an affine transformation, i.e.,

$$\mathbf{u} = \mathbf{S_u}\bar{\mathbf{u}} + \mathbf{o_u} \quad (18\text{a})$$

$$\mathbf{y} = \mathbf{S_y}\bar{\mathbf{y}} + \mathbf{o_y}, \quad (18\text{b})$$

$$\mathbf{v} = \mathbf{S_v}\bar{\mathbf{v}}, \quad (18\text{c})$$

where $\bar{\mathbf{u}}$, $\bar{\mathbf{y}}$, and $\bar{\mathbf{v}}$ are the quantities $\mathbf{u}$, $\mathbf{y}$, and $\mathbf{v}$, respectively, expressed in some base units, $\mathbf{S_u} \in \mathbb{R}^{n_u \times n_u}$, $\mathbf{S_y} \in \mathbb{R}^{n_y \times n_y}$, and $\mathbf{S_v} \in \mathbb{R}^{n_v \times n_v}$ are positive definite *scale* matrices and $\mathbf{o_u} \in \mathbb{R}^{n_u}$ and $\mathbf{o_y} \in \mathbb{R}^{n_y}$ are offset vectors[4]. To regularize the problem, the base units will be chosen such that their elements have variances equal to one, i.e., $\operatorname{diag}(\boldsymbol{\Sigma}_{\bar{\mathbf{g}}\bar{\mathbf{g}}}) = \mathbf{1}_{n_y \times 1}$, $\operatorname{diag}(\boldsymbol{\Sigma}_{\bar{\mathbf{u}}\bar{\mathbf{u}}}) = \mathbf{1}_{n_u \times 1}$, and $\operatorname{diag}(\boldsymbol{\Sigma}_{\bar{\mathbf{v}}\bar{\mathbf{v}}}) = \mathbf{1}_{n_v \times 1}$, where $\mathbf{1}_{a \times b}$ represents an $a \times b$ matrix of ones. Analogously to (1), the base units are related through a mapping $\bar{\mathbf{g}}$ as

$$\bar{\mathbf{y}} = \bar{\mathbf{g}}(\bar{\mathbf{u}}, \bar{\mathbf{v}}) \quad (19)$$

and thus

$$\mathbf{y} = \mathbf{g}(\mathbf{u}, \mathbf{v}) = \mathbf{S_y}\bar{\mathbf{g}}\left(\mathbf{S_u}^{-1}(\mathbf{u} - \mathbf{o_u}), \mathbf{S_v}^{-1}\mathbf{v}\right) + \mathbf{o_y}. \quad (20)$$

The property P1 can then be written as

$$M(\mathbf{g}, \psi_\mathbf{u}, \psi_\mathbf{v}) = M(\bar{\mathbf{g}}, \psi_{\bar{\mathbf{u}}}, \psi_{\bar{\mathbf{v}}}), \quad (21)$$

where $\psi_\mathbf{u}$, $\psi_\mathbf{v}$, $\psi_{\bar{\mathbf{u}}}$, and $\psi_{\bar{\mathbf{v}}}$ are distributions of $\mathbf{u}$, $\mathbf{v}$, $\bar{\mathbf{u}}$, and $\bar{\mathbf{v}}$, respectively. Note that the MoN (6) was shown to be independent [19] of the affine transformation of $\mathbf{u}$ and $\mathbf{v}$.

The property P2 can be expressed simply as

$$M : \mathcal{G} \times \mathcal{D} \times \mathcal{D} \mapsto [0, 1]. \quad (22)$$

IV. UNITLESS MSE-BASED MoN

To achieve the unitless property of the MSE-based MoN (8), the L_2 norm $\|\cdot\|_2$ is replaced by a weighted norm $\|\cdot\|_\mathbf{W} = (\cdot)\mathbf{W}(\cdot)^\mathsf{T}$, where $\mathbf{W} \in \mathbb{R}^{n_y \times n_y}$ is a positive definite matrix.

$$M = \sqrt{\min_{\mathbf{A}, \mathbf{b}, \psi_{\mathbf{v}, \mathbf{n}}} \mathsf{E}[\|\mathbf{g}(\mathbf{u}, \mathbf{v}) - \mathbf{L}(\mathbf{u}, \mathbf{n})\|_\mathbf{W}]} \quad (23)$$

Note that using the weighted norm was mentioned in [21] as an option without further details.

---

[4]Note that the noise $\bar{\mathbf{v}}$ still has zero mean. No offset is needed to transform $\bar{\mathbf{v}}$ to $\mathbf{v}$.

A. Minimization

The assumptions used for the minimization in (23) are:

A1: Variable $\mathbf{u}$ is independent of $\mathbf{v}$ and $\mathbf{n}$, i.e. $\psi_{\mathbf{u},\mathbf{v},\mathbf{n}} = \psi_{\mathbf{u}}\psi_{\mathbf{v},\mathbf{n}}$.

A2: The minimization in (23) is done w.r.t. $\mathbf{A}$, $\mathbf{b}$ and all joint distributions $\psi_{\mathbf{n},\mathbf{v}}$ such that the marginal distribution $\psi_{\mathbf{v}}$ is fixed and $\mathbf{n}$ has zero mean.

From **A2**, it follows that the minimization is done w.r.t. conditional distribution $\psi_{\mathbf{n}|\mathbf{v}}$ as $\psi_{\mathbf{n},\mathbf{v}} = \psi_{\mathbf{n}|\mathbf{v}}\psi_{\mathbf{v}}$ and $\psi_{\mathbf{v}}$ is fixed. The minimization can thus be rewritten as

$$\min_{\mathbf{A},\mathbf{b},\psi_{\mathbf{n}|\mathbf{v}}} \mathsf{E}[\|\mathbf{g}(\mathbf{u},\mathbf{v}) - (\mathbf{A}\mathbf{u}+\mathbf{b}+\mathbf{n})\|_{\mathbf{W}}] \tag{24}$$

$$= \min_{\mathbf{A},\mathbf{b},\psi_{\mathbf{n}|\mathbf{v}}} \Big\{ \mathsf{E}[\|\mathbf{g}(\mathbf{u},\mathbf{v}) - (\mathbf{A}\mathbf{u}+\mathbf{b})\|_{\mathbf{W}}] + \mathsf{E}[\|\mathbf{n}\|_{\mathbf{W}}]$$
$$- 2\mathsf{E}[(\mathbf{g}(\mathbf{u},\mathbf{v}) - \mathbf{A}\mathbf{u} - \mathbf{b})\mathbf{W}\mathbf{n}^{\mathsf{T}}]\Big\}$$

$$= \min_{\mathbf{A},\mathbf{b}} \mathsf{E}[\|\mathbf{g}(\mathbf{u},\mathbf{v}) - (\mathbf{A}\mathbf{u}+\mathbf{b})\|_{\mathbf{W}}]$$
$$+ \min_{\psi_{\mathbf{n}|\mathbf{v}}} \Big\{ \mathsf{E}[\|\mathbf{n}\|_{\mathbf{W}}] - 2\mathsf{E}[(\mathbf{g}(\mathbf{u},\mathbf{v})\mathbf{W}\mathbf{n}^{\mathsf{T}}] \Big\}$$

$$= \underbrace{\min_{\mathbf{A},\mathbf{b}} \mathsf{E}[\|\mathbf{g}(\mathbf{u},\mathbf{v}) - (\mathbf{A}\mathbf{u}+\mathbf{b})\|_{\mathbf{W}}]}_{J_{\mathbf{A},\mathbf{b}}}$$
$$+ \underbrace{\min_{\psi_{\mathbf{n}|\mathbf{v}}} \mathsf{E}[\|\mathbf{g}(\mathbf{u},\mathbf{v}) - \mathbf{n}\|_{\mathbf{W}}]}_{J_{\mathbf{n}}} - \mathsf{E}[\|\mathbf{g}(\mathbf{u},\mathbf{v})\|_{\mathbf{W}}].$$

Using (24), the MoN (23) equals to

$$M = \sqrt{J_{\mathbf{A},\mathbf{b}} + J_{\mathbf{n}} - \mathsf{E}\left[\|\mathbf{g}(\mathbf{u},\mathbf{v})\|_{\mathbf{W}}\right]}, \tag{25}$$

which means that the minimization in (23) was split into the deterministic part $J_{\mathbf{A},\mathbf{b}}$ and the stochastic part $J_{\mathbf{n}}$. The deterministic minimization can easily be solved [21] and leads to

$$J_{\mathbf{A},\mathbf{b}} = \mathrm{tr}\left(\mathbf{W}(\boldsymbol{\Sigma}_{\mathbf{gg}} - \boldsymbol{\Sigma}_{\mathbf{gu}}\boldsymbol{\Sigma}_{\mathbf{uu}}^{-1}\boldsymbol{\Sigma}_{\mathbf{ug}})\right), \tag{26}$$

where $\mathbf{A}$ and $\mathbf{b}$ are given by (9) and (10), respectively.

On the other hand, the minimization $J_{\mathbf{n}}$ cannot be done analytically for the general transformation (1). For the multiplicative form (3), the minimization leads to

$$J_{\mathbf{n}} = \mathsf{E}[\|\mathbf{g}(\mathbf{u},\mathbf{v})\|_{\mathbf{W}}] - \mathrm{tr}(\mathbf{W}\bar{\boldsymbol{\pi}}\boldsymbol{\Sigma}_{\gamma\gamma}\bar{\boldsymbol{\pi}}^{\mathsf{T}}), \tag{27}$$

where $\bar{\boldsymbol{\pi}} = \mathsf{E}[\boldsymbol{\pi}(\mathbf{u})]$. For details, see Appendix A. Thus, for the multiplicative noise case (3), the MoN is

$$M = \sqrt{\mathrm{tr}\left(\mathbf{W}(\boldsymbol{\Sigma}_{\mathbf{gg}} - \boldsymbol{\Sigma}_{\mathbf{gu}}\boldsymbol{\Sigma}_{\mathbf{uu}}^{-1}\boldsymbol{\Sigma}_{\mathbf{ug}} - \bar{\boldsymbol{\pi}}\boldsymbol{\Sigma}_{\gamma\gamma}\bar{\boldsymbol{\pi}}^{\mathsf{T}})\right)} \tag{28}$$
$$= \sqrt{\mathrm{tr}\left(\mathbf{W}(\boldsymbol{\Sigma}_{\mathbf{ff}} - \boldsymbol{\Sigma}_{\mathbf{fu}}\boldsymbol{\Sigma}_{\mathbf{uu}}^{-1}\boldsymbol{\Sigma}_{\mathbf{uf}} + \mathsf{E}[\tilde{\boldsymbol{\pi}}(\mathbf{u})\boldsymbol{\Sigma}_{\gamma\gamma}\tilde{\boldsymbol{\pi}}(\mathbf{u})^{\mathsf{T}}])\right)},$$

where the relations $\boldsymbol{\Sigma}_{\mathbf{gu}} = \boldsymbol{\Sigma}_{\mathbf{fu}}$, $\boldsymbol{\Sigma}_{\mathbf{gg}} = \boldsymbol{\Sigma}_{\mathbf{ff}} + \mathsf{E}[\boldsymbol{\pi}(\mathbf{u})\boldsymbol{\Sigma}_{\gamma\gamma}\boldsymbol{\pi}(\mathbf{u})^{\mathsf{T}}]$, and $\tilde{\boldsymbol{\pi}}(\mathbf{u}) = \boldsymbol{\pi}(\mathbf{u}) - \bar{\boldsymbol{\pi}}$ were used. For the additive noise case (4), the MoN is

$$M = \sqrt{\mathrm{tr}\left(\mathbf{W}(\boldsymbol{\Sigma}_{\mathbf{ff}} - \boldsymbol{\Sigma}_{\mathbf{fu}}\boldsymbol{\Sigma}_{\mathbf{uu}}^{-1}\boldsymbol{\Sigma}_{\mathbf{uf}}\right)}. \tag{29}$$

B. Normalization of MoN

The normalization of the MoN (25) is based on finding an upper bound of $J_{\mathbf{A},\mathbf{b}}$ and $J_{\mathbf{n}}$. To find the upper bound of $J_{\mathbf{A},\mathbf{b}}$, no linearity in $\mathbf{g}$ is assumed, i.e., $\mathbf{A} = \mathbf{0}$ and the optimum value of the shift (10) is used to obtain a tight upper bound, i.e.,

$$J_{\mathbf{A},\mathbf{b}} = \min_{\mathbf{A},\mathbf{b}} \mathsf{E}[\|\mathbf{g}(\mathbf{u},\mathbf{v}) - (\mathbf{A}\mathbf{u}+\mathbf{b})\|_{\mathbf{W}}]$$
$$\leq \mathsf{E}\left[\|\mathbf{g}(\mathbf{u},\mathbf{v}) - \mathsf{E}[\mathbf{g}(\mathbf{u},\mathbf{v})]\|_{\mathbf{W}}\right] = \mathrm{tr}(\mathbf{W}\boldsymbol{\Sigma}_{\mathbf{gg}}). \tag{30}$$

The upper bound of the stochastic part $J_{\mathbf{n}}$ can be found by rewriting the norm in $J_{\mathbf{n}}$ as

$$\mathsf{E}[\|\mathbf{g}(\mathbf{u},\mathbf{v}) - \mathbf{n}\|_{\mathbf{W}}] = \mathsf{E}[\|\mathbf{g}(\mathbf{u},\mathbf{v})\|_{\mathbf{W}} + \|\mathbf{n}\|_{\mathbf{W}}$$
$$- 2\mathbf{g}(\mathbf{u},\mathbf{v})\mathbf{W}\mathbf{n}^{\mathsf{T}}]. \tag{31}$$

When $\mathbf{v}$ and $\mathbf{n}$ are independent, the expectation of the last term is zero as $\mathsf{E}[\mathbf{n}] = \mathbf{0}$. Also, a distribution $\psi_{\mathbf{n}}$ must exist such that $\mathsf{E}[\|\mathbf{n}\|_{\mathbf{W}}] = 0$. Then, the upper bound of $J_{\mathbf{n}}$ is

$$J_{\mathbf{n}} \leq \mathsf{E}[\|\mathbf{g}(\mathbf{u},\mathbf{v})\|_{\mathbf{W}}]. \tag{32}$$

Substituting (30) and (32) to (25) then results in

$$M = \sqrt{J_{\mathbf{A},\mathbf{b}} + J_{\mathbf{n}} - \mathsf{E}[\|\mathbf{g}(\mathbf{u},\mathbf{v})\|_{\mathbf{W}}]} \leq \sqrt{\mathrm{tr}(\mathbf{W}\boldsymbol{\Sigma}_{\mathbf{gg}})}. \tag{33}$$

Note that the upper bound (33) is achieved for zero correlation $\boldsymbol{\Sigma}_{\mathbf{fu}} = \mathbf{0}$ in the additive case (4) and for the zero correlation and $\mathsf{E}[\tilde{\boldsymbol{\pi}}(\mathbf{u})\boldsymbol{\Sigma}_{\gamma\gamma}\tilde{\boldsymbol{\pi}}(\mathbf{u})^{\mathsf{T}}] = \mathbf{0}$ in the multiplicative case (3).

Specific choices of the weighting matrix will be discussed in the following section.

C. Weighting Matrix Choice

The unitless property of the MoN (23) depends on the weight matrix $\mathbf{W}$, i.e., the following equality must hold

$$M(\mathbf{g},\psi_{\mathbf{u}},\psi_{\mathbf{v}};\mathbf{W}) = M(\bar{\mathbf{g}},\psi_{\bar{\mathbf{u}}},\psi_{\bar{\mathbf{v}}};\bar{\mathbf{W}}), \tag{34}$$

where the notation explicitly expresses the MoN as a functional parametrized by the corresponding weight matrix. Using the relations (18) it follows that

$$\bar{\mathbf{g}}(\bar{\mathbf{u}},\bar{\mathbf{v}}) = \mathbf{S}_{\mathbf{y}}^{-1}\mathbf{g}(\mathbf{S}_{\mathbf{u}}\bar{\mathbf{u}}+\mathbf{o}_{\mathbf{u}}, \mathbf{S}_{\mathbf{v}}\bar{\mathbf{v}}) - \mathbf{S}_{\mathbf{y}}^{-1}\mathbf{o}_{\mathbf{y}} \tag{35a}$$
$$\psi_{\bar{\mathbf{u}}}(\bar{\mathbf{u}}) = \psi_{\mathbf{u}}(\mathbf{S}_{\mathbf{u}}\bar{\mathbf{u}}+\mathbf{o}_{\mathbf{u}}) \tag{35b}$$
$$\psi_{\bar{\mathbf{v}}}(\bar{\mathbf{v}}) = \psi_{\mathbf{v}}(\mathbf{S}_{\mathbf{v}}\bar{\mathbf{v}}) \tag{35c}$$

Using the notation $\tilde{\mathbf{g}}(\mathbf{u},\mathbf{v}) = \mathbf{S}_{\mathbf{y}}^{-1}\mathbf{g}(\mathbf{u},\mathbf{v}) - \mathbf{S}_{\mathbf{y}}^{-1}\mathbf{o}_{\mathbf{y}}$, the MoN for $\bar{\mathbf{g}}$ can be rewritten as

$$M(\bar{\mathbf{g}},\psi_{\bar{\mathbf{u}}},\psi_{\bar{\mathbf{v}}};\bar{\mathbf{W}}) = M(\tilde{\mathbf{g}},\psi_{\mathbf{u}},\psi_{\mathbf{v}};\bar{\mathbf{W}}) \tag{36}$$

Using simple manipulations, it can be shown that

$$M(\tilde{\mathbf{g}},\psi_{\mathbf{u}},\psi_{\mathbf{v}};\bar{\mathbf{W}}) = M(\mathbf{g},\psi_{\mathbf{u}},\psi_{\mathbf{v}};\mathbf{S}_{\mathbf{y}}^{-1}\bar{\mathbf{W}}\mathbf{S}_{\mathbf{y}}^{-1}), \tag{37}$$

Combining (34), (36), and (37) it can be concluded that to achieve the property P1, the weights must be related as

$$\mathbf{W} = \mathbf{S}_{\mathbf{y}}^{-1}\bar{\mathbf{W}}\mathbf{S}_{\mathbf{y}}^{-1}. \tag{38}$$

The equality between the weight matrices (38) and the normalization condition $\mathrm{tr}(\mathbf{W}\boldsymbol{\Sigma}_{\mathbf{gg}}) = \mathrm{tr}(\bar{\mathbf{W}}\boldsymbol{\Sigma}_{\bar{\mathbf{g}}\bar{\mathbf{g}}}) = 1$ following

from (33) can be used to obtain matrix $\mathbf{W}$ leading to a unitless and normalized MoN.

Recalling $\mathrm{diag}(\boldsymbol{\Sigma}_{\bar{\mathbf{g}}\bar{\mathbf{g}}}) = \mathbf{I}$, simple choices of weight matrices $\bar{\mathbf{W}}$ for the base units following directly from the normalization condition $\mathrm{tr}(\bar{\mathbf{W}}\boldsymbol{\Sigma}_{\bar{\mathbf{g}}\bar{\mathbf{g}}}) = 1$ are

$$\bar{\mathbf{W}}^{\mathrm{diag}} = \frac{1}{n_y}\mathbf{I} \tag{39}$$

$$\bar{\mathbf{W}}^{\mathrm{full}} = \frac{1}{n_y}\boldsymbol{\Sigma}_{\bar{\mathbf{g}}\bar{\mathbf{g}}}^{-1}. \tag{40}$$

The complete set of weight matrices $\bar{\mathbf{W}}$ satisfying $\mathrm{tr}(\bar{\mathbf{W}}\boldsymbol{\Sigma}_{\bar{\mathbf{g}}\bar{\mathbf{g}}}) = 1$ is given[5] by

$$\bar{\mathbf{W}} \in \{\mathbf{V}\mathbf{Y}\mathbf{V}^{\mathsf{T}}; \mathbf{Y} \in \mathbf{S}_{++}^{n_y}, [\mathbf{Y}]_{ii} = \tfrac{\alpha_i}{\lambda_i}, i = 1 \ldots n_y\}, \tag{41}$$

where $\boldsymbol{\Sigma}_{\bar{\mathbf{g}}\bar{\mathbf{g}}} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^{\mathsf{T}}$, $\mathbf{V}$ is the matrix of eigenvectors of $\boldsymbol{\Sigma}_{\bar{\mathbf{g}}\bar{\mathbf{g}}}$, $\boldsymbol{\Lambda}$ is the matrix of eigenvalues, $\lambda_i = [\boldsymbol{\Lambda}]_{ii}$, $\alpha_i > 0$ are parameters satisfying $\sum_{i=1}^{n_y} \alpha_i = 1$ and $\mathbf{S}_{++}^{n_y} = \{\mathbf{X} \in \mathbf{S}_{++}^{n_y}; \mathbf{X} \succ \mathbf{0}\}$ is a set of n_y-dimensional positive definite matrices. For the explanation, see Appendix B.

The weights (41), for which (39) an (40) are special cases, ensure normalization of the MoN and achieve the unitless property. Note that both $\|\hat{\mathbf{g}}(\mathbf{u}, \mathbf{v}) - \mathbf{L}(\mathbf{u}, \mathbf{n})\|_{\frac{1}{n_y}\mathbf{I}}$ and $\|\hat{\mathbf{g}}(\mathbf{u}, \mathbf{v}) - \mathbf{L}(\mathbf{u}, \mathbf{n})\|_{\frac{1}{n_y}\boldsymbol{\Sigma}_{\bar{\mathbf{g}}\bar{\mathbf{g}}}^{-1}}$ are unitless because $\mathbf{I} = \mathrm{diag}(\mathrm{diag}(\boldsymbol{\Sigma}_{\bar{\mathbf{g}}\bar{\mathbf{g}}}))^{-1}$ is an inverse of a diagonal matrix with the diagonal[6] given by $\mathrm{diag}(\boldsymbol{\Sigma}_{\bar{\mathbf{g}}\bar{\mathbf{g}}})$.

The weight matrices $\mathbf{W}$ for the original units are then obtained by (38) with (39) or (40) in the form

$$\mathbf{W}^{\mathrm{diag}} = \frac{1}{n_y}\mathbf{S}_{\mathbf{y}}^{-1}\mathbf{S}_{\mathbf{y}}^{-1} = \frac{1}{n_y}\mathrm{diag}(\mathrm{diag}(\boldsymbol{\Sigma}_{\mathbf{g}\mathbf{g}}))^{-1} \tag{42}$$

$$\mathbf{W}^{\mathrm{full}} = \frac{1}{n_y}\mathbf{S}_{\mathbf{y}}^{-1}\boldsymbol{\Sigma}_{\bar{\mathbf{g}}\bar{\mathbf{g}}}^{-1}\mathbf{S}_{\mathbf{y}}^{-1} = \frac{1}{n_y}\boldsymbol{\Sigma}_{\mathbf{g}\mathbf{g}}^{-1}. \tag{43}$$

The MoN (23) computed by (28) for the multiplicative noise case and by (29) for the additive noise case with weight matrix $\mathbf{W} = \mathbf{S}_{\mathbf{y}}^{-1}\bar{\mathbf{W}}\mathbf{S}_{\mathbf{y}}^{-1}$, where $\bar{\mathbf{W}}$ is given by (41) constitutes the desired unitless normalized MoN. The individual weight matrices, such as (42) or (43), differ in the way they treat the correlations among individual elements of $\mathbf{y}$. For example, the MoN with the choice (42) ignores the off-diagonal elements of $(\boldsymbol{\Sigma}_{\mathbf{ff}} - \boldsymbol{\Sigma}_{\mathbf{fu}}\boldsymbol{\Sigma}_{\mathbf{uu}}^{-1}\boldsymbol{\Sigma}_{\mathbf{uf}})$ while the MoN with the choice (43) weighs them through $\boldsymbol{\Sigma}_{\mathbf{gg}}^{-1}$.

Remark 1: When considering the transformation (1) without the noise, i.e., $\mathbf{y} = \mathbf{g}(\mathbf{u})$, the MoN (23) with $\mathbf{W}^{\mathrm{full}}$ corresponds to square root of the multivariate generalization of coefficient of determination [22] which has many convenient features.

Remark 2: For the weight matrix $\mathbf{W} = \mathbf{I}$ the original MSE-based norm (6) is obtained and the weight matrix $\mathbf{W} = \frac{1}{\sqrt{\mathrm{tr}(\boldsymbol{\Sigma}_{\mathbf{gg}})}}\mathbf{I}$ leads to the normalized MoN (8). However, none of the choices satisfy the property P1, i.e., they are not unitless.

---

[5]The notation $[\cdot]_{ii}$ stands for the element at the i-th row and i-th column.
[6]Here, the MATLAB notation is used in the expression.

V. Numerical Illustration

To illustrate the proposed unitless MoN, measurement functions of three tracking problems are considered: bearing-only tracking (BOT) [23], [24], ground moving target indicator (GMTI) filtering [25], [26], and tracking with range and direction cosine (RDCOS) measurements [21]. Analysis of MoNs for these problems was provided in [20] with a focus on the normalized MSE-based MoN (8).

Assuming the Cartesian state of the target $\mathbf{x} = [\mathsf{x}, \mathsf{y}, \dot{\mathsf{x}}, \dot{\mathsf{y}}]^{\mathsf{T}}$ with units [m, m, m/s, m/s] distributed as

$$\mathbf{x} \sim \mathcal{N}\{[500, 500, 5, 8.7]^{\mathsf{T}}, \alpha\,\mathrm{diag}([10^3, 10^3, 1, 1])\} \tag{44}$$

For the BOT, GMTI, and RDCOS the sensors provide measurements in the form

$$\mathbf{z}^{\mathrm{BOT}} = \mathbf{h}^{\mathrm{BOT}}(\mathbf{x}) = \mathrm{atan2}(\mathsf{x}, \mathsf{y}), \tag{45}$$

$$\mathbf{z}^{\mathrm{GMTI}} = [h_r(\mathbf{x}, \mathbf{s}), h_b(\mathbf{x}, \mathbf{s}), h_{v_r}(\mathbf{x}, \mathbf{s})]^{\mathsf{T}}, \tag{46a}$$

$$h_r(\mathbf{x}, \mathbf{s}) = \sqrt{(\mathsf{x} - s_x)^2 + (\mathsf{y} - s_y)^2 + (s_z)^2}, \tag{46b}$$

$$h_b(\mathbf{x}, \mathbf{s}) = \mathrm{atan2}(\mathsf{x} - s_x, \mathsf{y} - s_y), \tag{46c}$$

$$h_{v_r}(\mathbf{x}, \mathbf{s}) = \frac{(\mathsf{x} - s_x)\dot{\mathsf{x}} + (\mathsf{y} - s_y)\dot{\mathsf{y}}}{h_r(\mathbf{x}, \mathbf{s})}, \tag{46d}$$

$$\mathbf{z}^{\mathrm{RDCOS}} = \mathbf{h}^{\mathrm{RDCOS}}(\mathbf{x}) = \begin{bmatrix} \sqrt{\mathsf{x}^2 + \mathsf{y}^2} \\ \mathsf{x}/\sqrt{\mathsf{x}^2 + \mathsf{y}^2} \end{bmatrix}, \tag{47}$$

where $\mathbf{s} = [s_x, s_y, s_z]^{\mathsf{T}} = [10^3, 10^3, 10^3]^{\mathsf{T}}$ is GMTI sensor position. The value of the scaling factor α was set to $\alpha \in [10^{-1}, 10^1]$. Three MoNs were calculated for the considered problems:

- unitless MoN (29) with the full weight matrix (43),
- unitless MoN (29) with the diagonal weight matrix (42),
- the original MoN (8) (denoted as orig).

The covariance matrices required to calculate the MoNs were obtained by the Monte Carlo method with 10^7 samples. The MoNs are depicted in Figure 2. The figure shows that for the BOT problem, all MoN values coincide. This is caused by the fact that $\mathbf{z}^{\mathrm{BOT}} \in \mathbb{R}$ is scalar. For such case $\mathbf{W}^{\mathrm{full}} = \mathbf{W}^{\mathrm{diag}}$. Also, the value of the original MoN is equal to the proposed unitless MoN since for the scalar case dividing by the trace of a scalar in (8) corresponds to multiplying by $\mathbf{W}^{\mathrm{full}} = \frac{1}{1}\boldsymbol{\Sigma}_{\mathbf{ff}}^{-1}$ in (29).

For the GMTI and RDCOS problems, the values of the original MoN and the proposed unitless MoN differ. To analyze sensitivity of the original MoN on the units, $\mathbf{z}^{\mathrm{GMTI}}$ was considered either in [m, rad, m/s]$^{\mathsf{T}}$ or [km, deg, km/h]$^{\mathsf{T}}$ and $\mathbf{z}^{\mathrm{RDCOS}}$ was considered in [m, $-$] or [km, $-$]. The figure shows that the original MoN may vary for the GMTI between $M = 0$ for one set of units and $M = 1$ for the other set of units (c.f. values of α close to one). For the RDCOS problem, the differences are not so prominent.
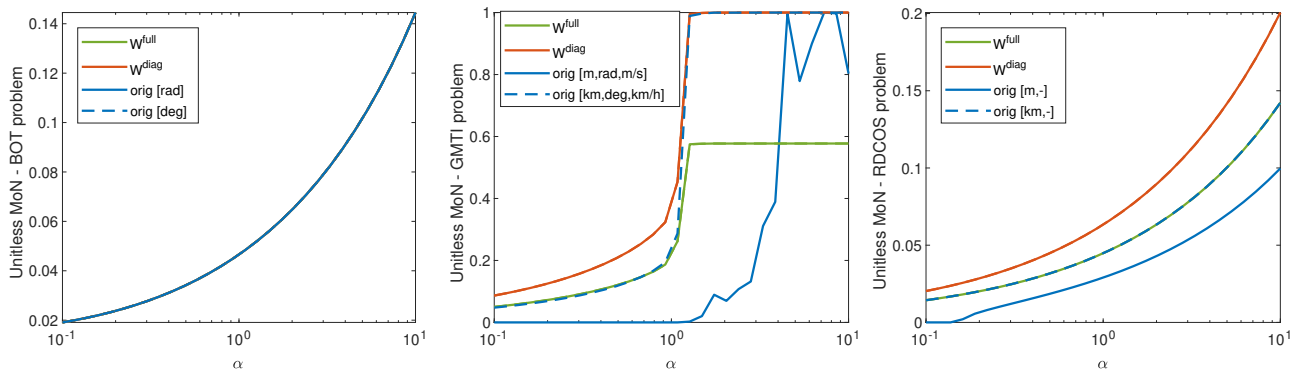
Figure 2. Comparison of MoNs for BOT, GMTI, and RDCOS problems: unitless MoN with $\mathbf{W}^{\text{full}}$ (green), unitless MoN with $\mathbf{W}^{\text{diag}}$ (red), original MSE-based MoN (blue and dashed blue).

The figure also shows that the unitless MoN's value depends on the weight matrix's choice $\mathbf{W}$. The weight matrix $\mathbf{W}^{\text{diag}}$ leads to higher values of MoN in the considered example because the difference $(\mathbf{\Sigma_{ff}} - \mathbf{\Sigma_{fu}}\mathbf{\Sigma_{uu}^{-1}}\mathbf{\Sigma_{uf}})$ is positively correlated. Generally, the unitless MoN with $\mathbf{W}^{\text{diag}}$ can be either lower or higher than the unitless MoN with $\mathbf{W}^{\text{full}}$.

Last, it shall be noticed that the MoN values for the BOT problem (interval $[2 \times 10^{-1}, 15 \times 10^{-1}]$) are slightly smaller than the MoN values for the RDCOS problem (interval $[0, 0.2]]$). The MoN values for the GMTI are significantly higher (interval $[0, 1]$), which indicates that compared to the BOT and RDCOS problems, the measurements of the GMTI exhibit strongly nonlinear behavior. This aligns with the findings in [13].

VI. Concluding Remarks

The paper focused on measures of nonlinearity for the state estimation problem. The measures of nonlinearity assess how far a transformation of random variables such as state and noise is from a linear behavior. The mean square error-based measure of nonlinearity was analyzed, and it was shown to suffer from sensitivity in the choice of the units. A new unitless nonlinearity measure was designed using a weight matrix to achieve the unitless property and normalization. A general set of weight matrices was specified, and two distinctive weight matrix choices were presented. The behavior of the new measure of nonlinearity was illustrated using the bearing-only tracking problem, ground target moving indicator filtering, and tracking with range and direction cosine.

VII. Acknowledgement

References

[1] E. M. L. Beale, "Confidence Regions in Non-Linear Estimation," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 22, no. 1, pp. 41–76, 12 1960.

[2] I. Guttman and D. A. Meeter, "On beale's measures of non-linearity," *Technometrics*, vol. 7, no. 4, pp. 623–637, 1965.

[3] H. N. Linssen, "Nonlinearity measures: a case study," *Statistica Neerlandica*, vol. 29, no. 3, pp. 93–99, 1975.

[4] D. M. Bates and D. G. Watts, *Nonlinear regression analysis and its applications*, ser. Wiley series in probability and mathematical statistics. New York: Wiley, 1988.

[5] C. Desoer and Y. T. Wang, "Foundations of feedback theory for nonlinear dynamical systems," *IEEE Transactions on Circuits and Systems*, vol. 27, no. 2, pp. 104–123, 1980.

[6] M. Nikolaou, "When is nonlinear dynamic modeling necessary?" in *1993 American Control Conference*, 1993, pp. 910–914.

[7] A. Helbig, W. Marquardt, and F. Allgöwer, "Nonlinearity measures: definition, computation and applications," *Journal of Process Control*, vol. 10, no. 2, pp. 113–123, 2000.

[8] A. El-Sakkary, "The gap metric: Robustness of stabilization of feedback systems," *IEEE Transactions on Automatic Control*, vol. 30, no. 3, pp. 240–247, 1985.

[9] T. Schreiber, "Interdisciplinary application of nonlinear time series methods," *Physics Reports*, vol. 308, no. 1, pp. 1 – 64, 1999.

[10] R. Haber, "Nonlinearity tests for dynamic processes," *IFAC Proceedings Volumes*, vol. 18, no. 5, pp. 409–414, 1985.

[11] K. Emancipator and M. H. Kroll, "A quantitative measure of nonlinearity," *Clinical Chemistry*, vol. 39, no. 5, pp. 766–772, 1993.

[12] R. Kotchoni, "Detecting and measuring nonlinearity," *Econometrics*, vol. 6, no. 3, 2018.

[13] M. Mallick, B. F. La Scala, and M. S. Arulampalam, "Differential geometry measures of nonlinearity for the bearing-only tracking problem," in *Signal Processing, Sensor Fusion, and Target Recognition XIV*, ser. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, I. Kadar, Ed., vol. 5809, May 2005, pp. 288–300.

[14] M. Mallick and B. Scala, "Differential geometry measures of nonlinearity for ground moving target indicator (gmti) filtering," in *2005 7th International Conference on Information Fusion*, vol. 1, 08 2005, pp. 219 – 226.

[15] M. M. and B. F. La Scala, "Differential geometry measures of nonlinearity for the video tracking problem," in *Signal Processing, Sensor Fusion, and Target Recognition XV*, I. Kadar, Ed., vol. 6235, International Society for Optics and Photonics. SPIE, 2006, p. 62350P.

[16] D. Pham, "Use of non-linearity as a characteristic in the selection of filtering algorithms in kinematic positioning," Ph.D. dissertation, 2020.

[17] O. Straka, J. Dunik, and M. Simandl, "Unscented kalman filter with controlled adaptation," in *Proceedings of 16th IFAC Symposium on System Identification (SYSID)*, vol. 16, no. 1, 2012, pp. 906–911.

[18] J. Duník, O. Straka, and A. García-Fernández, "Performance evaluation of nonlinearity and non-Gaussianity measures in state estimation," in *2017 20th International Conference on Information Fusion (Fusion)*, July 2017, pp. 1–10.

[19] Y. Liu and X. R. Li, "Measure of nonlinearity for estimation," *IEEE Transactions on Signal Processing*, vol. 63, no. 9, pp. 2377–2388, 2015.

[20] M. Mallick and B. Ristic, "Comparison of measures of nonlinearity for bearing-only and gmti filtering," in *2017 20th International Conference on Information Fusion (Fusion)*, 2017, pp. 1–8.

[21] X. R. Li, "Measure of nonlinearity for stochastic systems," in *Proceedings of the 15th International Conference on Information Fusion*, Singapore, 7 2012.

[22] A. C. Rencher and W. F. Christensen, *Methods of multivariate analysis*. Hoboken, New Jersey: Wiley, 2012.

[23] V. Aidala and S. Hammel, "Utilization of modified polar coordinates for bearings-only tracking," *IEEE Transactions on Automatic Control*, vol. 28, no. 3, pp. 283–294, 1983.

[24] S. Arulampalam and B. Ristic, "Comparison of the particle filter with range-parameterized and modified polar EKFs for angle-only tracking," in *Signal and Data Processing of Small Targets 2000*, ser. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, O. E. Drummond, Ed., vol. 4048, Jul. 2000, pp. 288–299.

[25] T. Kirubarajan, Y. Bar-Shalom, K. Pattipati, and I. Kadar, "Ground target tracking with variable structure imm estimator," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 36, no. 1, pp. 26–46, 2000.

[26] P. Wang, E. Blasch, X. R. Li, E. Jones, R. Hanak, W. H. Yin, A. Beach, and P. Brewer, "Degree of nonlinearity (don) measure for target tracking in videos," in *2016 19th International Conference on Information Fusion (FUSION)*, 2016, pp. 1390–1397.

APPENDIX A
MINIMIZATION OF $J_{\mathbf{n}}$

To calculate $J_{\mathbf{n}} = \min_{\psi_{\mathbf{n}|\mathbf{v}}} \mathsf{E}[\|\mathbf{g}(\mathbf{u}, \mathbf{v}) - \mathbf{n}\|_{\mathbf{W}}]$ for the multiplicative case $\mathbf{g}(\mathbf{u}, \mathbf{v}) = \mathbf{f}(\mathbf{u}) + \boldsymbol{\pi}(\mathbf{u})\boldsymbol{\gamma}(\mathbf{v})$, it should be reminded that $\mathsf{E}[\mathbf{n}] = \mathbf{0}$, $\mathsf{E}[\boldsymbol{\gamma}(\mathbf{v})] = \mathbf{0}$, and $\mathbf{u}$ and $\mathbf{v}$ are assumed independent. Then

$$J_{\mathbf{n}} = \min_{\psi_{\mathbf{n}|\mathbf{v}}} \mathsf{E}[\|\mathbf{f}(\mathbf{u}) + \boldsymbol{\pi}(\mathbf{u})\boldsymbol{\gamma}(\mathbf{v}) - \mathbf{n}\|_{\mathbf{W}}]$$
$$= \mathsf{E}\|\mathbf{f}(\mathbf{u})\|_{\mathbf{W}} + \min_{\psi_{\mathbf{n}|\mathbf{v}}} \mathsf{E}[\|\boldsymbol{\pi}(\mathbf{u})\boldsymbol{\gamma}(\mathbf{v}) - \mathbf{n}\|_{\mathbf{W}}]. \quad (48)$$

Now, by denoting $\bar{\boldsymbol{\pi}} := \mathsf{E}[\boldsymbol{\pi}(\mathbf{u})]$, the term $\boldsymbol{\pi}(\mathbf{u})$ can be written as $\boldsymbol{\pi}(\mathbf{u}) = \bar{\boldsymbol{\pi}} + \widetilde{\boldsymbol{\pi}}(\mathbf{u})$. It follows that

$$\min_{\psi_{\mathbf{n}|\mathbf{v}}} \mathsf{E}[\|\boldsymbol{\pi}(\mathbf{u})\boldsymbol{\gamma}(\mathbf{v}) - \mathbf{n}\|_{\mathbf{W}}]$$
$$= \min_{\psi_{\mathbf{n}|\mathbf{v}}} \{\mathsf{E}[\|\widetilde{\boldsymbol{\pi}}(\mathbf{u})\boldsymbol{\gamma}(\mathbf{v})\|_{\mathbf{W}}] + \mathsf{E}[\|\bar{\boldsymbol{\pi}}\boldsymbol{\gamma}(\mathbf{v}) - \mathbf{n}\|_{\mathbf{W}}]\}$$
$$= \mathsf{E}[\|\widetilde{\boldsymbol{\pi}}(\mathbf{u})\boldsymbol{\gamma}(\mathbf{v})\|_{\mathbf{W}}] + \min_{\psi_{\mathbf{n}|\mathbf{v}}} \mathsf{E}[\|\bar{\boldsymbol{\pi}}\boldsymbol{\gamma}(\mathbf{v}) - \mathbf{n}\|_{\mathbf{W}}]$$
$$= \operatorname{tr} \mathbf{W} \mathsf{E}[\widetilde{\boldsymbol{\pi}}(\mathbf{u})\boldsymbol{\Sigma}_{\boldsymbol{\gamma}\boldsymbol{\gamma}}\widetilde{\boldsymbol{\pi}}^{\mathsf{T}}], \quad (49)$$

where $\min_{\psi_{\mathbf{n}|\mathbf{v}}} \mathsf{E}[\|\bar{\boldsymbol{\pi}}\boldsymbol{\gamma}(\mathbf{v}) - \mathbf{n}\|_{\mathbf{W}}] = \mathbf{0}$ for a distribution $\psi_{\mathbf{n}|\mathbf{v}}$ of $\mathbf{n}$ such that $\mathbf{n} = \bar{\boldsymbol{\pi}}\boldsymbol{\gamma}(\mathbf{v})$ with probability one.

By expressing $\mathsf{E}[\|\mathbf{g}(\mathbf{u}, \mathbf{v})\|_{\mathbf{W}}]$ as

$$\mathsf{E}[\|\mathbf{g}(\mathbf{u}, \mathbf{v})\|_{\mathbf{W}}] = \mathsf{E}[\|\mathbf{f}(\mathbf{u}) + (\bar{\boldsymbol{\pi}} + \widetilde{\boldsymbol{\pi}}(\mathbf{u}))\boldsymbol{\gamma}(\mathbf{v})\|_{\mathbf{W}}]$$
$$= \mathsf{E}[\|\mathbf{f}(\mathbf{u})\|_{\mathbf{W}}] + \operatorname{tr} \mathbf{W} \mathsf{E}[\widetilde{\boldsymbol{\pi}}(\mathbf{u})\boldsymbol{\Sigma}_{\boldsymbol{\gamma}\boldsymbol{\gamma}}\widetilde{\boldsymbol{\pi}}(\mathbf{u})^{\mathsf{T}}] + \operatorname{tr} \mathbf{W}\bar{\boldsymbol{\pi}}\boldsymbol{\Sigma}_{\boldsymbol{\gamma}\boldsymbol{\gamma}}\bar{\boldsymbol{\pi}}^{\mathsf{T}}, \quad (50)$$

the stochastic part $J_{\mathbf{n}}$ is

$$J_{\mathbf{n}} = \min_{\psi_{\mathbf{n}|\mathbf{v}}} \mathsf{E}[\|\mathbf{g}(\mathbf{u}, \mathbf{v}) - \mathbf{n}\|_{\mathbf{W}}]$$
$$= \mathsf{E}[\|\mathbf{g}(\mathbf{u}, \mathbf{v})\|_{\mathbf{W}}] - \operatorname{tr} \mathbf{W}\bar{\boldsymbol{\pi}}\boldsymbol{\Sigma}_{\boldsymbol{\gamma}\boldsymbol{\gamma}}\bar{\boldsymbol{\pi}}^{\mathsf{T}}. \quad (51)$$

APPENDIX B
SET OF WEIGHTS

To find the set of matrices $\bar{\mathbf{W}}$ such that $\operatorname{tr}(\bar{\mathbf{W}}\boldsymbol{\Sigma}_{\bar{\mathbf{g}}\bar{\mathbf{g}}}) = 1$, where $\bar{\mathbf{W}} \succ \mathbf{0}$ and $\boldsymbol{\Sigma}_{\bar{\mathbf{g}}\bar{\mathbf{g}}} \succ \mathbf{0}$, the eigen decomposition of $\boldsymbol{\Sigma}_{\bar{\mathbf{g}}\bar{\mathbf{g}}} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^{\mathsf{T}}$ is computed first. Then,

$$\operatorname{tr}(\bar{\mathbf{W}}\boldsymbol{\Sigma}_{\bar{\mathbf{g}}\bar{\mathbf{g}}}) = \operatorname{tr}(\bar{\mathbf{W}}\mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^{\mathsf{T}}) = \operatorname{tr}(\mathbf{V}^{\mathsf{T}}\bar{\mathbf{W}}\mathbf{V}\boldsymbol{\Lambda}) = \operatorname{tr}(\mathbf{Y}\boldsymbol{\Lambda}),$$

where $\mathbf{Y} = \mathbf{V}^{\mathsf{T}}\bar{\mathbf{W}}\mathbf{V}$. Since $\boldsymbol{\Lambda} = \operatorname{diag}([\lambda_1, \dots \lambda_{n_y}])$ is a diagonal matrix, the trace will equal to one, $\operatorname{tr}(\mathbf{Y}\boldsymbol{\Lambda}) = 1$, if for the diagonal elements of $[\mathbf{Y}]_{ii}$ the following equality holds:

$$[\mathbf{Y}]_{ii} = \frac{\alpha_i}{\lambda_i}, \ i = 1, \dots, n_y,$$

where $\sum_{i=1}^{n_y} = 1$, $\alpha_i \geq 0$. The off-diagonal elements $\mathbf{Y}$ can be arbitrarily chosen such that $\mathbf{Y} \succ \mathbf{0}$. The weight matrix $\bar{\mathbf{W}}$ then is $\bar{\mathbf{W}} = \mathbf{V}\mathbf{Y}\mathbf{V}^{\mathsf{T}}$.