JOURNAL OF GUIDANCE, CONTROL, AND DYNAMICS
Vol. , No. ,

Engineering Notes

Covariance Estimation and Gaussianity Assessment for State and Measurement Noise

Jindřich Duník,* Oliver Kost,† and Ondřej Straka‡
University of West Bohemia, 306 14 Pilsen, Czech Republic
and
Erik Blasch§
Air Force Research Laboratory, Rome, New York 13441

https://doi.org/10.2514/1.G004348

1 ## I. Introduction

KNOWLEDGE of an appropriate system state-space model is a key prerequisite for an optimal design of many signal processing algorithms for applications such as global navigation satellite system (GNSS)–based routing and radar-measurement-based object tracking. The system state-space model includes the parameterized deterministic part and the distribution-based stochastic part. Although the deterministic model often arises from the first principles based on physical, kinematical, and mathematical laws governing the system behavior, the description of the stochastic part is often difficult to physically model and is identified from the measured data.

A tremendous research interest focuses on a design of methods estimating the properties of the stochastic part of the model, namely, on the estimation of the covariance matrices (CMs) of the state and measurement noise appearing in the state-space model [1–3]. A rather limited attention has been devoted to the assessment of whether the state and measurement noises are Gaussian or not, although such information is essential for the optimal design of Kalman filter (KF)–based navigation and tracking algorithms requiring consistent and integrity assured outputs.

In particular, two approaches for simultaneous noise characteristics estimation and probability density function (PDF) assessment can be found in the literature. The *first* approach is based on the estimation of *higher-order moments* (HOMs) (e.g., besides the noise CMs, estimating the noises skewness and kurtosis) and subsequent comparison with the expected HOMs computed under the assumption of a Gaussian distribution¶ [4]. The second approach estimates the noise CMs only and then analyses statistical properties of the *measurement prediction error*** (MPE) to decide about the Gaussianity of the noises. Compared with the HOM approach, the MPE approach can provide the decision about the noises Gaussianity with a required probability of false alarm (FA). The concept of the MPE approach has been recently outlined and validated in [5], which resulted in the *noise covariance matrices estimation with gaussianity assessment* (NEGA) method. However, the NEGA method was introduced with a strong focus on the linear time-varying (LTV) models and one particular Gaussianity goodness-of-fit test without

1) Consideration of the state noise shaping matrix
2) Discussion and illustration of design parameters selection
3) Design of the NEGA method for the linear time-invariant (LTI) models allowing efficient implementation

The objective of the paper is to derive the NEGA method for both LTV and LTI models *with* the state noise shaping matrix. Special attention is placed on selection of the design parameter, statistical test for the Gaussianity assessment, and on a computational complexity of the NEGA method especially for the LTI models. The exemplary MATLAB implementations of the NEGA method are submitted along with the paper to facilitate method understanding, implementation, and application.

The rest of the paper is organized as follows. In Sec. II, the state-space model is defined, the task of the noise CMs estimation is introduced, and the goal of the paper is drawn. The NEGA method is derived and discussed in Secs. III and IV. Numerical evaluation is provided in Sec. V and concluding remarks are given in Sec. VI.

II. System Definition, Problem Formulation, and Goal of the Paper

Let the state-space model describe an LTV discrete time stochastic dynamic system with additive noises [6–8]

$$\boldsymbol{x}_{k+1} = \boldsymbol{F}_k \boldsymbol{x}_k + \boldsymbol{u}_k + \boldsymbol{G}_k \boldsymbol{w}_k \tag{1}$$

$$\boldsymbol{z}_k = \boldsymbol{H}_k \boldsymbol{x}_k + \boldsymbol{v}_k \tag{2}$$

where $\boldsymbol{x}_k \in \mathbb{R}^{n_x}$, $\boldsymbol{u}_k \in \mathbb{R}^{n_x}$, and $\boldsymbol{z}_k \in \mathbb{R}^{n_z}$ represent the immeasurable state of the system, measurable input, and the known measurement at time instant k, respectively, with $k = 0, 1, \ldots, \tau$. The state matrix $\boldsymbol{F}_k \in \mathbb{R}^{n_x \times n_x}$, the measurement matrix $\boldsymbol{H}_k \in \mathbb{R}^{n_z \times n_x}$, and the state noise shaping matrix $\boldsymbol{G}_k \in \mathbb{R}^{n_x \times n_w}$ are known and bounded $\forall k$. The moments and the PDF of the initial state $\boldsymbol{x}_0$ are *not* assumed to be known. The variables $\boldsymbol{w}_k \in \mathbb{R}^{n_w}$ and $\boldsymbol{v}_k \in \mathbb{R}^{n_z}$ are the state and measurement *zero-mean* white noises with *unknown* noise CMs $\boldsymbol{Q} \in \mathbb{R}^{n_w \times n_w}$ and $\boldsymbol{R} \in \mathbb{R}^{n_z \times n_z}$, respectively, where $n_w \leq n_x$. Distribution type, that is, the form of the PDF $p(\boldsymbol{w}_k)$ and $p(\boldsymbol{v}_k)$, is *unknown* as well. The system state $\boldsymbol{x}_k$ is assumed to be *observable*, $\forall k$.

A. Noise Covariance Matrices Estimation

From the 1970s, an extensive number of the noise CMs estimation methods have been proposed. The methods, which estimate the unknown noise CMs $\boldsymbol{Q}$ and $\boldsymbol{R}$ using the available measurements $\boldsymbol{z}_k$, inputs $\boldsymbol{u}_k$, and the known system matrices $\boldsymbol{F}_k$, $\boldsymbol{G}_k$, and $\boldsymbol{H}_k$, can be divided into four groups according to their underlying approaches [1,2], namely, the *correlation* methods, *maximum likelihood* methods, *covariance matching* methods, and *Bayesian* methods. The state-of-the-art noise CMs estimation methods, surveyed in [2], have been designed for a wide range of the models (linear/nonlinear, time-invariant/time-varying), may provide unbiased and consistent estimates, and offer a tradeoff between estimation accuracy and computational complexity. Unfortunately, the vast majority of the methods do *not* provide any additional information regarding the PDF of the noises, mainly, whether the state and measurement noises are Gaussian or not. Without such

Received 8 February 2019; revision received 11 September 2019; accepted for publication 30 September 2019; published online XX epubMonth XXXX.

*Senior Researcher, Department of Cybernetics, Univerzitní 8.
†Ph.D. Candidate, Department of Cybernetics, Univerzitní 8.
‡Associate Professor, Department of Cybernetics, Univerzitní 8.
§Program Officer. Associate Fellow AIAA.

¶Gaussian distribution is fully parametrized by the first two moments; all higher-order moments can be uniquely computed on the basis of those and can considered as expected higher-order moments assuming the Gaussian PDF. By comparison of the expected and estimated (real) higher-order moments, the Gaussianity of the noises can be assessed.

**MPE can be shown to be a weighted sum of the state and measurement noise.

information, however, any optimal and reliable KF-based positioning, tracking, and navigation algorithms *cannot* be designed.

B. Goal of the Paper: Derivation of NEGA Method for LTI/LTV Models

The paper deals with statistical noise CMs estimation and noise Gaussianity assessment. The goal of the paper is to thoroughly derive the NEGA method for both LTV and time-invariant models of the forms (1) and (2) with specification of the rules for design parameter selection. Particular attention is also devoted to the selection of the statistical test for the Gaussianity assessment and on a computational complexity of the NEGA method especially for the LTI models.

The NEGA method is derived in the following two sections, where the noise CMs are estimated (Sec. III) and the noise Gaussianity assessed (Sec. IV). The paper builds upon the conference paper [5], where the basic idea and concept of the NEGA method have been proposed and illustrated. Compared with the conference paper, this paper substantially extends the NEGA method with:

1) Derivation of the NEGA method for a state-space model with the state noise shaping matrix in Eqs. (1) and (2)

2) Design of the NEGA method for the LTI models allowing more accurate and computational efficient implementation

3) Thorough discussion, analysis, and illustration of the NEGA method performance w.r.t the design parameters

III. Noise CMs Estimation Method

The first part of the NEGA method deals with the noise CMs estimation. The noise CMs estimation adopts the concept of the recently introduced correlation method for the noise CMs estimation [5,9], which provides *unbiased* and *consistent* estimates for the LTV models. The method for the noise CMs estimation consists of four steps; 1) design of a linear measurement predictor, 2) computation of the MPE, 3) statistical analysis of the MPE, and 4) sample-based estimate of the MPE CMs and of the noise CMs $\boldsymbol{Q}$ and $\boldsymbol{R}$. The method for the LTV models is proposed in Sec. III.A, and then in Sec. III.B, its computationally efficient version is developed for the LTI model.

A. Noise CMs Estimation for LTV Models

The method for the noise CMs estimation for the LTV models (1) and (2) with the noise shaping matrix is derived below.

1. Augmented Measurement Predictor

The *augmented measurement* $\boldsymbol{Z}_k^L$ and its *prediction* $\hat{\boldsymbol{Z}}_k^L$ can be expressed as

$$\boldsymbol{Z}_k^L = [z_k^T, z_{k+1}^T, \dots, z_{k+L-1}^T]^T, \qquad k = 0, 1, \dots, \tau - L + 1 \quad (3)$$

$$\hat{\boldsymbol{Z}}_k^L = \mathcal{O}_k^L(\boldsymbol{F}_{k-1}(\mathcal{O}_{k-1}^L)^\dagger(\boldsymbol{Z}_{k-1}^L - \boldsymbol{\Gamma}_{k-1}^L \boldsymbol{U}_{k-1}^L) + \boldsymbol{u}_{k-1}) + \boldsymbol{\Gamma}_k^L \boldsymbol{U}_k^L, \quad k = 1, \dots, \tau - L + 1 \quad (4)$$

where the notation $\boldsymbol{A}^T$ denotes the transpose of the matrix $\boldsymbol{A}$, and $\boldsymbol{U}_k^L \in \mathbb{R}^{Ln_x}$ and $\boldsymbol{\Gamma}_k^L \in \mathbb{R}^{Ln_z \times Ln_x}$ are defined as

$$\boldsymbol{U}_k^L = \begin{bmatrix} \boldsymbol{u}_k \\ \boldsymbol{u}_{k+1} \\ \vdots \\ \boldsymbol{u}_{k+L-1} \end{bmatrix},$$

$$\boldsymbol{\Gamma}_k^L = \begin{bmatrix} \boldsymbol{0}_{n_z\times n_x} & \boldsymbol{0}_{n_z\times n_x} & \cdots & \boldsymbol{0}_{n_z\times n_x} & \boldsymbol{0}_{n_z\times n_x} \\ \boldsymbol{H}_{k+1} & \boldsymbol{0}_{n_z\times n_x} & \cdots & \boldsymbol{0}_{n_z\times n_x} & \boldsymbol{0}_{n_z\times n_x} \\ \boldsymbol{H}_{k+2}\boldsymbol{F}_{k+1} & \boldsymbol{H}_{k+2} & \cdots & \boldsymbol{0}_{n_z\times n_x} & \boldsymbol{0}_{n_z\times n_x} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \boldsymbol{H}_{k+L-1}\mathcal{F}_{k+1}^{L-2} & \boldsymbol{H}_{k+L-1}\mathcal{F}_{k+2}^{L-3} & \cdots & \boldsymbol{H}_{k+L-1} & \boldsymbol{0}_{n_z\times n_x} \end{bmatrix} \quad (5)$$

where $\boldsymbol{0}_{m\times n}$ denotes a zero matrix of the indicated dimension. Parameter L is selected such that the observability matrix

$$\mathcal{O}_k^L = [(\boldsymbol{H}_k)^T,\ (\boldsymbol{H}_{k+1}\boldsymbol{F}_k)^T,\ \dots,\ (\boldsymbol{H}_{k+L-1}\mathcal{F}_k^{L-1})^T]^T \in \mathbb{R}^{Ln_z\times n_x} \quad (6)$$

is of full rank,[††] $\forall k$; that is, $L \geq n_x$, $(\mathcal{O}_{k-1}^L)^\dagger = ((\mathcal{O}_{k-1}^L)^T \mathcal{O}_{k-1}^L)^{-1}(\mathcal{O}_{k-1}^L)^T \in \mathbb{R}^{n_x\times Ln_z}$ is the pseudoinverse of the matrix $\mathcal{O}_{k-1}^L$, and $\mathcal{F}_k^M = \prod_{i=1}^M \boldsymbol{F}_{k+M-i} = \boldsymbol{F}_{k+M-1}\dots\boldsymbol{F}_{k+1}\boldsymbol{F}_k \in \mathbb{R}^{n_x\times n_x}$.

2. Augmented Measurement Prediction Error

The augmented measurement prediction error (AMPE) is defined as

$$\tilde{\boldsymbol{Z}}_k^L = \boldsymbol{Z}_k^L - \hat{\boldsymbol{Z}}_k^L, \qquad k = 1, \dots, \tau - L + 1 \quad (7)$$

where the augmented measurement vector can be, w.r.t. models (1) and (2), written as

$$\boldsymbol{Z}_k^L = \mathcal{O}_k^L \boldsymbol{x}_k + \boldsymbol{\Gamma}_k^L(\boldsymbol{U}_k^L + \tilde{\boldsymbol{W}}_k^L) + \boldsymbol{V}_k^L \quad (8)$$

and the augmented measurement vector prediction, w.r.t. Eqs. (8), (4), and (1), reads

$$\hat{\boldsymbol{Z}}_k^L = \mathcal{O}_k(\boldsymbol{x}_k - \tilde{\boldsymbol{w}}_{k-1}) + \mathcal{O}_k^L \boldsymbol{F}_{k-1}(\mathcal{O}_{k-1}^L)^\dagger(\boldsymbol{\Gamma}_{k-1}^L \tilde{\boldsymbol{W}}_{k-1}^L + \boldsymbol{V}_{k-1}^L) + \boldsymbol{\Gamma}_k^L \boldsymbol{U}_k^L \quad (9)$$

where the substitution $\tilde{\boldsymbol{w}}_k = \boldsymbol{G}_k \boldsymbol{w}_k \in \mathbb{R}^{n_x}$ was used, and the vectors and matrices $\boldsymbol{W}_k^L \in \mathbb{R}^{Ln_w}$, $\tilde{\boldsymbol{W}}_k^L \in \mathbb{R}^{Ln_x}$, $\boldsymbol{V}_k^L \in \mathbb{R}^{Ln_z}$, and $\boldsymbol{\Gamma}_k^L \in \mathbb{R}^{Ln_z\times Ln_x}$ are defined by

$$\tilde{\boldsymbol{W}}_k^L = \begin{bmatrix} \tilde{\boldsymbol{w}}_k \\ \tilde{\boldsymbol{w}}_{k+1} \\ \vdots \\ \tilde{\boldsymbol{w}}_{k+L-1} \end{bmatrix}, \quad \boldsymbol{W}_k^L = \begin{bmatrix} \boldsymbol{w}_k \\ \boldsymbol{w}_{k+1} \\ \vdots \\ \boldsymbol{w}_{k+L-1} \end{bmatrix}, \quad \boldsymbol{V}_k^L = \begin{bmatrix} \boldsymbol{v}_k \\ \boldsymbol{v}_{k+1} \\ \vdots \\ \boldsymbol{v}_{k+L-1} \end{bmatrix},$$

$$\boldsymbol{\Gamma}_k^L = \begin{bmatrix} \boldsymbol{0}_{n_z\times n_x} & \boldsymbol{0}_{n_z\times n_x} & \cdots & \boldsymbol{0}_{n_z\times n_x} & \boldsymbol{0}_{n_z\times n_x} \\ \boldsymbol{H}_{k+1} & \boldsymbol{0}_{n_z\times n_x} & \cdots & \boldsymbol{0}_{n_z\times n_x} & \boldsymbol{0}_{n_z\times n_x} \\ \boldsymbol{H}_{k+2}\boldsymbol{F}_{k+1} & \boldsymbol{H}_{k+2} & \cdots & \boldsymbol{0}_{n_z\times n_x} & \boldsymbol{0}_{n_z\times n_x} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \boldsymbol{H}_{k+L-1}\mathcal{F}_{k+1}^{L-2} & \boldsymbol{H}_{k+L-1}\mathcal{F}_{k+2}^{L-3} & \cdots & \boldsymbol{H}_{k+L-1} & \boldsymbol{0}_{n_z\times n_x} \end{bmatrix} \quad (10)$$

with $\tilde{\boldsymbol{W}}_{k-1}^L = [(\boldsymbol{G}_{k-1}\boldsymbol{w}_{k-1})^T, (\boldsymbol{G}_k \boldsymbol{w}_k)^T, \dots, (\boldsymbol{G}_{k+L-2}\boldsymbol{w}_{k+L-2})^T]^T = \mathcal{G}_{k-1}^L \boldsymbol{W}_{k-1}^L$ and

$$\mathcal{G}_{k-1}^L = \begin{bmatrix} \boldsymbol{G}_{k-1} & \boldsymbol{0}_{n_x\times n_w} & \cdots & \boldsymbol{0}_{n_x\times n_w} & \boldsymbol{0}_{n_x\times n_w} \\ \boldsymbol{0}_{n_x\times n_w} & \boldsymbol{G}_k & \cdots & \boldsymbol{0}_{n_x\times n_w} & \boldsymbol{0}_{n_x\times n_w} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \boldsymbol{0}_{n_x\times n_w} & \boldsymbol{0}_{n_x\times n_w} & \cdots & \boldsymbol{0}_{n_x\times n_w} & \boldsymbol{G}_{k+L-2} \end{bmatrix}$$

Then, the AMPE [Eq. (7)] can be written in a *compact* form

$$\tilde{\boldsymbol{Z}}_k^L = \tilde{\mathcal{A}}_k \tilde{\boldsymbol{\xi}}_k^{L^+} = \tilde{\mathcal{A}}_k \boldsymbol{\Pi}_k^{L^+} \boldsymbol{\xi}_k^{L^+} = \mathcal{A}_k \boldsymbol{\xi}_k^{L^+} \quad (11)$$

where $L^+ = L + 1$,

$$\boldsymbol{\xi}_k^{L^+} = [(\boldsymbol{W}_{k-1}^{L^+})^T,\ \ (\boldsymbol{V}_{k-1}^{L^+})^T]^T \in \mathbb{R}^{L^+(n_w+n_z)} \quad (12)$$

[††]The full rank matrix $\mathcal{O}_k^L$ always exists as the system state is supposed to be observable; that is, the state $\boldsymbol{x}_k$ is observable from the augmented measurement vector $\boldsymbol{Z}_k^L$.

$$\tilde{\xi}_k^{L^+} = [(\tilde{W}_{k-1}^{L^+})^T, \quad (V_{k-1}^{L^+})^T]^T \in \mathbb{R}^{L^+(n_x+n_z)} \tag{13}$$

$$\tilde{\mathcal{A}}_k = [\tilde{\mathcal{A}}_k^{(w)}, \quad \mathcal{A}_k^{(v)}] \in \mathbb{R}^{Ln_z \times L^+(n_x+n_z)} \tag{14}$$

with

$$\mathcal{A}_k = \tilde{\mathcal{A}}_k \Pi_k^{L^+} \tag{15}$$

$$\tilde{\xi}_k^{L^+} = \Pi_k^{L^+} \xi_k^{L^+} \tag{16}$$

$$\Pi_k = \begin{bmatrix} \mathcal{G}_{k-1}^{L^+} & \mathbf{0}_{L^+n_x \times L^+n_z} \\ \mathbf{0}_{L^+n_z \times L^+n_w} & I_{L^+n_z} \end{bmatrix}^T \tag{17}$$

$$\tilde{\mathcal{A}}_k^{(w)} = \begin{bmatrix} I_{Ln_z} \\ I_{Ln_z} \end{bmatrix}^T \begin{bmatrix} [\mathcal{O}_k^L, \quad \Gamma_k^L] \\ [-\mathcal{O}_k^L F_{k-1} (\mathcal{O}_{k-1}^L)^\dagger \Gamma_{k-1}^L, \mathbf{0}_{Ln_z \times n_x}] \end{bmatrix},$$
$$\mathcal{A}_k^{(v)} = \begin{bmatrix} I_{Ln_z} \\ I_{Ln_z} \end{bmatrix}^T \begin{bmatrix} [\mathbf{0}_{Ln_z \times n_z}, I_{Ln_z}] \\ [-\mathcal{O}_k^L F_{k-1} (\mathcal{O}_{k-1}^L)^\dagger, \mathbf{0}_{Ln_z \times n_z}] \end{bmatrix} \tag{18}$$

and the symbol I_n denotes the identity matrix of the indicated dimension.

Note 1: From Eq. (8), an *unbiased* but *not* a minimum mean square error state estimate can be computed [5].

Note 2: The matrix $\mathcal{A}_k$ in Eq. (11) is a function of the known model matrices F_k, H_k, and G_k, and thus it is known. The AMPE [Eq. (11)] is a *linear* function of the state and measurement noises stacked in the vector $\xi_k^{L^+}$ whose statistical properties are sought.

Note 3: Whereas the form of the AMPE [Eq. (7)] is suitable for the prediction error computation on the basis of measured data, the form (11) is suitable for the following AMPE statistical analysis.

3. Augmented Measurement Prediction Error Statistical Analysis

Because of the properties of the state and measurement noises forming the vector $\xi_k^{L^+}$, the AMPE $\tilde{Z}_k^L$ [Eq. (11)] is the *zero-mean* stochastic process with the covariance matrix (CM)‡‡ $C_k \in \mathbb{R}^{Ln_z \times Ln_z}$ defined by

$$C_k = \mathsf{E}[\tilde{Z}_k^L (\tilde{Z}_k^L)^T] = \mathcal{A}_k \mathsf{E}[\xi_k^{L^+} (\xi_k^{L^+})^T] \mathcal{A}_k^T \quad = \mathcal{A}_k \Xi \mathcal{A}_k^T \tag{19}$$

The matrix $\Xi \in \mathbb{R}^{n_\Xi \times n_\Xi}$ with $n_\Xi = L^+(n_w + n_z)$ is composed from the unknown and sought noise CMs Q and R as

$$\Xi = \mathsf{E}[\xi_k^{L^+} (\xi_k^{L^+})^T] = \begin{bmatrix} I_{L^+} \otimes Q & \mathbf{0}_{L^+n_w \times L^+n_z} \\ \mathbf{0}_{L^+n_z \times L^+n_w} & I_{L^+} \otimes R \end{bmatrix} \tag{20}$$

where the symbol $\otimes$ represents the Kronecker product [10].

The matrix C_k [Eq. (19)] can be, w.r.t. the matrix identity $ABC = (C^T \otimes A) B_S$ [10], rewritten into a convenient form

$$(C_k)_S = (\mathcal{A}_k \otimes \mathcal{A}_k) \Xi_S \tag{21}$$

where the notation $(A)_S$ stands for the columnwise stacking of a symmetric matrix $A \in \mathbb{R}^{n_A \times n_A}$ into a vector $(A)_S \in \mathbb{R}^{n_A^2}$.

It should be highlighted that the CM C_k is a *linear* function of the unknown noise CMs Q, R. Then, the CM in the vector form $(C_k)_S$ [Eq. (21)] can be written in a compact form as

$$\Lambda_k \theta = b_k \tag{22}$$

where $\Lambda_k \in \mathbb{R}^{n_b \times n_\theta}$, with $n_b = (Ln_z)^2$, is the matrix depending on the *known* model matrices defined as

$$\Lambda_k = (\mathcal{A}_k \otimes \mathcal{A}_k) \Psi \tag{23}$$

$b_k \in \mathbb{R}^{n_b}$ is the vector of elements of the CM given by

$$b_k = (C_k)_S \tag{24}$$

and $\theta \in \mathbb{R}^{n_\theta}$, with $n_\theta = [n_w(n_w + 1) + n_z(n_z + 1)/2]$, is the vector of *all* unknown *unique* elements of the noise CMs Q and R defined as

$$\theta = [(Q_{\mathrm{TS}})^T, \quad (R_{\mathrm{TS}})^T]^T \tag{25}$$

The notation A_{TS} stands for the columnwise stacking of only the *unique* $n_A(n_A + 1)/2$ elements of a symmetric matrix $A \in \mathbb{R}^{n_A \times n_A}$ by elimination of the supradiagonal elements. The matrix $\Psi \in \mathbb{R}^{n_\Xi^2 \times n_\theta}$ is a known duplication matrix [5,9] fulfilling the equality

$$\Xi_S = \Psi \theta \tag{26}$$

The AMPE CM Eq. (22) can be summarized for all time instants as

$$\Lambda \theta = b \tag{27}$$

with $\Lambda = [\Lambda_1^T, \Lambda_2^T, \Lambda_3^T, \ldots, \Lambda_{\tau-L+1}^T]^T$, $b = [b_1^T, b_2^T, b_3^T, \ldots, b_{\tau-L+1}^T]^T$.

4. Sample-Based Estimate of AMPE CM and Noise CMs

The matrix Λ in Eq. (27) is a function of the *known* model matrices F_k, G_k, and H_k and the *known* duplication matrix Ψ only. If the AMPE CM C_k were available $\forall k$ [i.e., available b in Eq. (27)], then the vector of the unknown noise CMs elements θ could be computed from Eq. (27) by the least-squares (LS) method. The CM C_k is, however, *not* available as it depends on the sought noise CMs Q and R [see the CM description Eq. (21)]. Nevertheless, the CM C_k can be estimated from a sequence of the measured and input data according to Eqs. (4) and (7) and similarly the vector b in Eq. (27).

Assuming available sequence of the measured and input data z_k and u_k, $\forall k$, the MPE sequence $\{\tilde{Z}_k^L\}_{k=1}^{\tau-L+1} = [\tilde{Z}_1^L, \tilde{Z}_2^L, \ldots, \tilde{Z}_{\tau-L+1}^L]$ can be computed according to Eq. (7) and the *sample-based* estimate of b_k in Eq. (22) is given by

$$\hat{b}_k = (\hat{C}_k)_S = (\tilde{Z}_k^L (\tilde{Z}_k^L)^T)_S = \tilde{Z}_k^L \otimes \tilde{Z}_k^L \tag{28}$$

and the *sample-based* estimate of b in Eq. (27) is then

$$\hat{b} = [\hat{b}_1^T, \hat{b}_2^T, \hat{b}_3^T, \ldots, \hat{b}_{\tau-L+1}^T]^T \tag{29}$$

Assuming that Λ is of full row rank, the *optimum* (LS) estimate of the vector of the noise CMs unknown elements is, due to Eq. (27), given by

$$\hat{\theta} = \Lambda^\dagger \hat{b} \tag{30}$$

Based on the vector $\hat{\theta}$ [Eq. (30)], the noise CMs estimates $\hat{Q}$ and $\hat{R}$ are recovered according to Eqs. (26) and (20).

Note 4: The parametric estimate $\hat{\theta}$ [Eq. (30)] (i.e., the noise CMs estimates $\hat{Q}$ and $\hat{R}$) are proven to be *unbiased* (i.e., $\mathsf{E}[\hat{\theta}] = \theta$), and *weakly consistent* (i.e., $\hat{\theta} \to \theta$ as $\tau \to \infty$). The full proofs can be found in [9].

Note 5: For a large set of data $\tau \to \infty$, the estimated noise CMs are positive-definite. For finite (and low) τ, the estimates may lose the positive-definiteness. In this case, the estimation procedure should be repeated with a higher number of data or an LS solution with an implicit constraint on the noise CMs positive-definiteness should be used as briefly discussed (e.g., in [11]). As the CMs are symmetric positive definite, the constraint can be formulated, for example, by a set of nonlinear inequality relations according to Sylvester's criterion. It should be also mentioned that consideration of the constrained LS method may affect properties of the estimate in terms of its unbiasedness and consistency.

‡‡The AMPE is a time-correlated stochastic process. Thus, besides the covariance matrix, the cross-covariance matrix can be used for noise CMs estimation as well. Further details on AMPE properties can be found in [9].

B. Noise CMs Estimation for LTI Models

Derivation of the method for the noise CMs estimation for the LTI model [i.e., models (1) and (2)] with $\boldsymbol{F}_k = \boldsymbol{F}, \boldsymbol{G}_k = \boldsymbol{G}$ and $\boldsymbol{H}_k = \boldsymbol{H}$, $\forall k$, is, in principle, the same as that for the LTV model, but taking into account the system time invariability the NEGA method derivation results in simpler relations allowing computationally efficient noise CMs estimation.

1. Augmented Measurement Predictor

Design of the one-step augmented measurement predictor starts from the definition of the augmented measurement

$$\boldsymbol{Z}_k^L = \mathcal{O}^L \boldsymbol{x}_k + \boldsymbol{\Gamma}^L(\boldsymbol{U}_k^L + \tilde{\boldsymbol{W}}_k^L) + \boldsymbol{V}_k^L \tag{31}$$

which is analogous to Eq. (8) with the exception of the *time-invariant* observability matrix $\mathcal{O}^L$ and matrix $\boldsymbol{\Gamma}^L$ defined as

$$\mathcal{O}^L = \begin{bmatrix} \boldsymbol{H} \\ \boldsymbol{HF} \\ \vdots \\ \boldsymbol{HF}^{L-1} \end{bmatrix}, \quad \boldsymbol{\Gamma}^L = \begin{bmatrix} \boldsymbol{0}_{n_z\times n_x} & \boldsymbol{0}_{n_z\times n_x} & \cdots & \boldsymbol{0}_{n_z\times n_x} & \boldsymbol{0}_{n_z\times n_x} \\ \boldsymbol{H} & \boldsymbol{0}_{n_z\times n_x} & \cdots & \boldsymbol{0}_{n_z\times n_x} & \boldsymbol{0}_{n_z\times n_x} \\ \boldsymbol{HF} & \boldsymbol{H} & \cdots & \boldsymbol{0}_{n_z\times n_x} & \boldsymbol{0}_{n_z\times n_x} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \boldsymbol{HF}^{L-2} & \boldsymbol{HF}^{L-3} & \cdots & \boldsymbol{H} & \boldsymbol{0}_{n_z\times n_x} \end{bmatrix} \tag{32}$$

The one-step augmented measurement prediction is then

$$\hat{\boldsymbol{Z}}_k^L = \mathcal{O}^L(\boldsymbol{F}(\mathcal{O}^L)^\dagger(\boldsymbol{Z}_{k-1}^L - \boldsymbol{\Gamma}^L\boldsymbol{U}_{k-1}^L) + \boldsymbol{u}_{k-1}) + \boldsymbol{\Gamma}^L\boldsymbol{U}_k^L \tag{33}$$

2. Augmented Measurement Prediction Error and Its Analysis

The AMPE $\tilde{\boldsymbol{Z}}_k^L$ [Eq. (7)] can be, with respect to Eqs. (31–33), written as

$$\tilde{\boldsymbol{Z}}_k^L = \mathcal{A}\boldsymbol{\xi}_k^{L^+} \tag{34}$$

where $\mathcal{A}$ is a known and *time-invariant* matrix defined as

$$\mathcal{A} = \tilde{\mathcal{A}}\boldsymbol{\Pi} \tag{35}$$

$$\boldsymbol{\Pi} = \begin{bmatrix} \mathcal{G}^{L^+} & \boldsymbol{0}_{L^+n_x\times L^+n_z} \\ \boldsymbol{0}_{L^+n_z\times L^+n_w} & \boldsymbol{I}_{L^+n_z} \end{bmatrix}^T \tag{36}$$

$$\mathcal{G}^{L^+} = \boldsymbol{I}_{L^+} \otimes \boldsymbol{G}, \qquad \tilde{\mathcal{A}} = [\tilde{\mathcal{A}}^{(w)}, \mathcal{A}^{(v)}] \tag{37}$$

$$\tilde{\mathcal{A}}^{(w)} = \begin{bmatrix} \boldsymbol{I}_{Ln_z} \\ \boldsymbol{I}_{Ln_z} \end{bmatrix}^T \begin{bmatrix} [\mathcal{O}^L, \boldsymbol{\Gamma}^L] \\ [-\mathcal{O}^L\boldsymbol{F}(\mathcal{O}^L)^\dagger\boldsymbol{\Gamma}^L, \boldsymbol{0}_{Ln_z\times n_x}] \end{bmatrix},$$
$$\mathcal{A}^{(v)} = \begin{bmatrix} \boldsymbol{I}_{Ln_z} \\ \boldsymbol{I}_{Ln_z} \end{bmatrix}^T \begin{bmatrix} [\boldsymbol{0}_{Ln_z\times n_z}, \boldsymbol{I}_{Ln_z}] \\ [-\mathcal{O}^L\boldsymbol{F}(\mathcal{O}^L), \boldsymbol{0}_{Ln_z\times n_z}] \end{bmatrix} \tag{38}$$

The AMPE $\tilde{\boldsymbol{Z}}_k^L$ [Eq. (34)] is again a *zero-mean* stochastic process with the *time-invariant* CM

$$\boldsymbol{C} = \mathsf{E}[\tilde{\boldsymbol{Z}}_k^L(\tilde{\boldsymbol{Z}}_k^L)^T] = \mathcal{A}\Xi\mathcal{A}^T \tag{39}$$

where Ξ is defined in Eq. (20). Using the Kronecker algebra, the AMPE CM $\boldsymbol{C}$ [Eq. (39)] can be written as a system of linear equations with respect to the unique elements of $\boldsymbol{Q}$ and $\boldsymbol{R}$, gathered in the vector $\boldsymbol{\theta}$, and the known duplication matrix $\boldsymbol{\Psi}$ as

$$\boldsymbol{\Delta\theta} = \boldsymbol{b} \tag{40}$$

where $\boldsymbol{\Delta} = (\mathcal{A} \otimes \mathcal{A})\boldsymbol{\Psi}$ and $\boldsymbol{b} = \boldsymbol{C}_S$.

3. Sample-Based Estimate of AMPE CM and Noise CMs

Similar to the LTV case, the system of linear Eq. (40) would be easily solvable if the vector $\boldsymbol{b}$ [and thus the AMPE CM $\boldsymbol{C}$, Eq. (39)] were known. The vector is, however, unknown, but it can be estimated from the available AMPE sequence $\{\tilde{\boldsymbol{Z}}_k^L\}_{k=1}^{\tau-L+1}$ by a sample AMPE CM as

$$\hat{\boldsymbol{C}} = \frac{1}{\tau - L + 1}\sum_{k=1}^{\tau-L+1} \tilde{\boldsymbol{Z}}_k^L(\tilde{\boldsymbol{Z}}_k^L)^T \tag{41}$$

$$\hat{\boldsymbol{b}} = \hat{\boldsymbol{C}}_S \tag{42}$$

Then, under the assumption of the full row rank of the matrix $\boldsymbol{\Delta}$, the optimum estimate of the vector of the noise CMs unknown elements is, due to Eq. (40), given by

$$\hat{\boldsymbol{\theta}} = \boldsymbol{\Delta}^\dagger\hat{\boldsymbol{b}} \tag{43}$$

Note 6: The dimension of the design matrix (i.e., matrix of regressors) $\boldsymbol{\Lambda}$ in Eq. (30) for the LTV models is $(\tau - L + 1)n_b \times n_\theta$, whereas the dimension of the design matrix $\boldsymbol{\Delta}$ in Eq. (43) for the LTI models is $n_b \times n_\theta$ only. Thus, the pseudoinverse of the design matrix is much simpler for the LTI models. The estimate $\hat{\boldsymbol{\theta}}$ [Eq. (30)] becomes $\hat{\boldsymbol{\theta}}$ [Eq. (43)] if the LTI model is considered.

C. Design Parameter Selection and Noise CMs Estimability

The NEGA method requires specification of one design parameter, namely, the parameter L. The parameter determines, in fact, the number of linear equations used for noise CMs estimation in Eqs. (30) and (43). The minimal choice $L_{\min}$, ensuring full rank observation matrix, results in the NEGA method with the minimal computational requirements. If L is selected to be greater than $L_{\min}$, then a higher number of linear equations are used for $\boldsymbol{Q}$ and $\boldsymbol{R}$ estimation, and consequently it results in higher computational complexity. The higher number of linear equations may (or may not as they are not optimally weighted [9]) result in higher quality estimates in terms of lower estimate error. Further discussion and analysis of selection of L can be found in a section devoted to the numerical illustration.

For the LTV models, if $L > L_{\min}$ and the model matrices $\boldsymbol{F}_k$, $\boldsymbol{G}_k$, and $\boldsymbol{H}_k$ are sufficiently varying, then the matrix $\boldsymbol{\Lambda}$ in Eq. (27) is of full column rank, and then the NEGA method provides estimates of all elements of the noise CMs. However, the matrix $\boldsymbol{\Lambda}$ rank cannot be assessed in general just from the properties of the model matrices $\boldsymbol{F}_k$, $\boldsymbol{G}_k$, and $\boldsymbol{H}_k$ as it is problem specific. Further discussion with examples is demonstrated in numerical illustrations.

As with correlation methods, the NEGA method for the LTI models may not generally allow estimation of all elements of the noise CMs (because of insufficient rank of the design matrix) independently of selected L. In particular, the complete measurement noise CMs $\boldsymbol{R}$ can be estimated, but only a subset of the state noise CMs $\boldsymbol{Q}$ elements can be estimated. The subset is guaranteed to have at least n_x elements. Thus, it is always possible to estimate all diagonal elements of $\boldsymbol{Q}$. Note, however, that with increasing n_z, the number of estimable elements of $\boldsymbol{Q}$ grows as well [1,9,11].

If some elements of the noise CMs $\boldsymbol{Q}$ and $\boldsymbol{R}$ are known, then the vector $\boldsymbol{\theta}$ [Eq. (25)] can be split into known and unknown parts, $\boldsymbol{\theta}_{\text{known}}$ and $\boldsymbol{\theta}_{\text{unknown}}$, respectively, and the unknown part is estimated by the LS method only. Namely, considering the noise CMs estimation for the LTI model, the linear Eq. (40) can be written as $\boldsymbol{\Delta}_{\text{known}}\boldsymbol{\theta}_{\text{known}} + \Delta_{\text{unknown}}\boldsymbol{\theta}_{\text{unknown}} = \boldsymbol{b}$, which allows estimation of the unknown noise CMs elements only according to $\hat{\boldsymbol{\theta}}_{\text{unknown}} = (\boldsymbol{\Delta}_{\text{unknown}})^\dagger \times (\hat{\boldsymbol{b}} - \boldsymbol{\Delta}_{\text{known}}\boldsymbol{\theta}_{\text{known}})$.

IV. Noises Gaussianity Assessment

Having the noise CMs estimates $\hat{\boldsymbol{Q}}$ and $\hat{\boldsymbol{R}}$, it remains to decide whether both the state and measurement noises $\boldsymbol{w}_k$ and $\boldsymbol{v}_k$ are Gaussian or not using a statistical hypothesis testing.

A *direct* hypothesis testing of noise Gaussianity would require availability of the state and measurements noise samples. These are, however, not available in practical situations. Therefore, an *indirect* statistical test of the noise Gaussianity was proposed [5], which tests, instead of the unavailable noise sequences, the *available* sequence of AMPE $\tilde{\boldsymbol{Z}}_k^L$ defined in Eq. (7). The AMPE $\tilde{\boldsymbol{Z}}_k^L$ is a "weighted" sum of the state and measurement noises that are stacked in $\boldsymbol{\xi}_k^{L^+}$ [Eq. (12)], where the weights are known and given by the matrix $\mathcal{A}_k$ in Eq. (11) for the LTV model and by the $\mathcal{A}$ in Eq. (34) for the LTI model. The *intuition* behind the indirect test thus lies in the fact that, if both noises are Gaussian, then the AMPE is Gaussian as well.

The indirect test as proposed in [5] is, in principle, a goodness-of-fit (GoF) test requiring available estimates of the noise CMs. Such a requirement is, however, an unnecessary limitation for the LTI models as the use of the noise CMs estimates instead of the true CMs may deteriorate the test performance. Also, the GoF test does not necessarily offer the best performance in terms of its statistical power. Therefore, in the following parts, the GoF test is briefly reviewed and several alternatives with better properties are introduced. Some of the introduced tests do not require known estimates $\hat{\boldsymbol{Q}}$ and $\hat{\boldsymbol{R}}$ for the LTI models.

A. Statistical Tests

In the statistical literature, many tests of Gaussianity have been proposed so far, where the *null* and *alternative* hypotheses are formulated as follows:

1) H_0: The AMPE $\tilde{\boldsymbol{Z}}_k^L$ [Eq. (7)] comes from a Gaussian distribution.

2) H_1: The AMPE $\tilde{\boldsymbol{Z}}_k^L$ [Eq. (7)] does not come from such a distribution.

The tests are designed to provide a decision whether the null hypothesis H_0 is accepted at the specified level of significance[§§] α or rejected. In the following section, several popular tests[¶¶] are reviewed.

1. Chi-Squared Goodness-of-Fit Test

The *chi-squared goodness-of-fit* test (Chi2GoF) [12] is based on the definition of a scalar normalized squared AMPE $\zeta_k^L = (\tilde{\boldsymbol{Z}}_k^L)^T(\mathcal{C}_k)^{-1}\tilde{\boldsymbol{Z}}_k^L \in \mathbb{R}$, where $\mathcal{C}_k = \boldsymbol{C}_k$ is the AMPE CM defined for the LTV model in Eq. (19) and $\mathcal{C}_k = \boldsymbol{C}$ defined for the LTI model in Eq. (39). If the state and measurement noises are Gaussian, then the AMPE $\tilde{\boldsymbol{Z}}_k^L$ [Eq. (7)] is Gaussian as well, and the normalized squared AMPE ζ_k^L is a chi-squared distributed variable with Ln_z degrees of freedom (DoF). If either of the state or the measurement noise is non-Gaussian, then the normalized squared AMPE ζ_k^L is generally not a chi-squared distributed variable. Unfortunately, the AMPE CMs $\boldsymbol{C}_k$ and $\boldsymbol{C}$ are not known (as they depend on the sought noise CMs). The CMs, however, can be estimated forms (19) and (39), respectively, using available estimates of the noise CMs $\hat{\boldsymbol{Q}}$ and $\hat{\boldsymbol{R}}$ (based on *all* data), as $\hat{\mathfrak{C}}_k = \mathcal{A}_k\hat{\Xi}\mathcal{A}_k^T$ and $\hat{\mathfrak{C}} = \mathcal{A}\hat{\Xi}\mathcal{A}^T$ for the LTV and LTI model, respectively, where $\hat{\Xi} = \begin{bmatrix} \boldsymbol{I}_{L^+} \otimes \hat{\boldsymbol{Q}} & \boldsymbol{0}_{L^+n_w \times L^+n_z} \\ \boldsymbol{0}_{L^+n_z \times L^+n_w} & \boldsymbol{I}_{L^+} \otimes \hat{\boldsymbol{R}} \end{bmatrix}$. Then, the resulting computable *test statistic* is defined as

$$\zeta_k^L = (\tilde{\boldsymbol{Z}}_k^L)^T(\hat{\mathcal{C}}_k)^{-1}\tilde{\boldsymbol{Z}}_k^L \tag{44}$$

where $\hat{\mathcal{C}}_k = \hat{\mathfrak{C}}_k$ or $\hat{\mathcal{C}}_k = \hat{\mathfrak{C}}$ and the test itself always depends on the quality of estimated noise CMs $\hat{\boldsymbol{Q}}$ and $\hat{\boldsymbol{R}}$.

[§§]The level of significance is the probability of rejection of the null hypothesis if it is true. The level may be denoted as the probability of false alert. Note also that, to be able to specify the probability of missed detection, it is necessary to particularize the alternative hypothesis H_1. It means that it is not enough to say that prediction error does not come from a Gaussian PDF, but it must be specified from which PDF the prediction error alternatively comes, which is quite challenging task partially solved in [4].

[¶¶]Most of the considered tests are available in a common statistical software such as MATLAB or R.

2. Anderson–Darling Test

The Anderson–Darling (AD) test [13] is able to directly decide whether the AMPE comes from a normal distribution. The test is based on the comparison of the empirical (sample-based) cumulative distribution function with its theoretical counterpart. The test requires *identically* distributed samples; that is, all the samples of the AMPE should follow the same distribution. Therefore, for the LTI models, the AMPE sequence $\tilde{\boldsymbol{Z}}_k^L$ can be *directly* used as a test statistic. The test statistic is *independent* of the estimated noise CMs and, thus, not affected by a possible estimation error. However, for the LTV models, the AMPE needs to be transformed[***] as

$$\varepsilon_k^L = (\hat{\boldsymbol{S}}_k)^{-1}\tilde{\boldsymbol{Z}}_k^L \tag{45}$$

where $\hat{\boldsymbol{S}}_k$ is a square-root factor of $\hat{\mathfrak{C}}_k$ fulfilling $\hat{\mathfrak{C}}_k = \hat{\boldsymbol{S}}_k(\hat{\boldsymbol{S}}_k)^T$. With the transformed AMPE, the quality of the noise CMs estimates directly influences the test statistic and, consequently, the performance of the statistical test, as shown in a section devoted to the numerical illustrations.

3. Jarque–Bera Test

The Jarque–Bera (JB) test statistic [14] compares the sample-based skewness and kurtosis with the ones expected for the Gaussian distribution. The test again requires identically distributed samples of the AMPE; thus for the LTI models, the AMPE can be directly used, and for the LTV models, the AMPE has to be stochastically decoupled.

4. Shapiro–Wilk Test

The Shapiro–Wilk (SW) test [14] compares transformed (and ordered) samples with samples generated from a standard Gaussian distribution. The test typically requires independent and identically distributed samples. Thus, the AMPE needs to be stochastically decoupled for the LTV models.

5. Lilliefors Test

The Lilliefors (LF) test [13] is, in some sense, similar to the AD test, with only a different criterion for evaluation of the sample-based, and expected cumulative distribution function is used. The test requires identically distributed samples; thus, for LTV models the AMPE has to be decoupled.

Note 7: The AMPE in Eqs. (11) and (34) can be interpreted as an addition of *two* sums of independent and identically distributed variables: one sum for the state noise sequence $\boldsymbol{W}_{k-1}^{L^+}$ and other for the measurement noise sequence $\boldsymbol{V}_{k-1}^{L^+}$. Each sum has L^+ addends. According to the central limit theorem, each sum will approach a normal distribution as L^+ goes to infinity. In the *limit* case, the AMPE would appear to be Gaussian, although the PDFs of the state and measurement noise $p(\boldsymbol{w}_k)$ and $p(\boldsymbol{v}_k)$, respectively, are non-Gaussian (e.g., heavy tailed or skewed), and consequently, the NEGA method would provide an incorrect assessment. However, in a typical NEGA method setting the parameter L^+ is kept *low* (see definition of $L_{\min}$) to minimize the method's computational complexity. In this case, non-Gaussian noises still result in the non-Gaussian AMPE and thus in the correct functionality of the (non-)Gaussian assessment.

B. Hypotheses Testing for LTV and LTI Models

For the LTV models, all introduced tests are based on the known noise CMs estimates $\hat{\boldsymbol{Q}}$ and $\hat{\boldsymbol{R}}$. Thus, the quality of the estimates affects the statistical test performance. For the LTI models, the noise CMs estimates are not required with the exception of the Chi2GoF test. As a consequence, the quality of the tests is not affected by the quality of the estimates. Nevertheless, the tests have different properties, for example, in terms of their power; that is, they have different probability of correct rejection of the null hypothesis H_0 and different sensitivity to the number of available measurements τ.

[***]The transformation is known as the stochastic decoupling, where the transformed variable has covariance matrix equal to the identity matrix.

Illustration is given in the section devoted to the numerical simulations.

Some of the statistical tests or their particular implementations may require independent samples (such as the SW test). In this case, the AMPE sequence should be "down-sampled"; that is, every L^+ element of the AMPE is selected and used as a test statistic. Then, the elements of the down-sampled sequence are certainly independent. The reason can be found in definition of the AMPE and especially in definition of the extended vector of the noises $\boldsymbol{\xi}_k^{L^+}$ in Eqs. (11) and (34), where the vectors $\boldsymbol{\xi}_k^{L^+}$ and $\boldsymbol{\xi}_{k+L^+}^{L^+}$ do not share any common state and measurement noise realization [see definition (12)].

V. Numerical Illustration

The performance of the proposed NEGA method is illustrated using following four examples:

1) Scalar LTI model illustrating an impact of the user defined parameter L and performance of the statistical tests.

2) Scalar LTV model for illustration of an impact of the noise CMs quality on statistical tests performance.

3) Multidimensional LTV model illustrating a relation between model time-variability and noise CMs estimability.

4) Multidimensional LTI model typically used in target tracking.

In the examples, the performance is assessed over $M = 10^4$ Monte Carlo (MC) simulations using the following criteria:

1) Estimated error variance $\widehat{\mathrm{var}}[\hat{Q}_{i,j}] = (1/M)\sum_{m=1}^{M} \times \left(Q_{i,j} - \hat{Q}_{i,j}^{(m)}\right)^2$, where $Q_{i,j}$ is element of the true CM $\boldsymbol{Q}$ at ith row and jth column and $\hat{Q}_{i,j}$ its estimate at mth MC simulation

2) Average noise CM estimate $\hat{\mathsf{E}}[\hat{Q}_{i,j}] = (1/M)\sum_{m=1}^{M} \hat{Q}_{i,j}^{(m)}$

3) Number of MC simulations with rejected null hypothesis H_0 over all MC simulations denoted as $M_{H0\text{-rejected}}$

The examples and evaluation criteria have been selected to illustrate all important properties of the NEGA method. However, even a broad set of simulations cannot cover all situations and setups appearing in applications. Therefore, the *source files* for scalar LTI and LTV models (i.e., first two examples) are *provided* along with the paper. The NEGA method was implemented with the stress on the readability of the code with a marginal focus on the computational complexity. Nevertheless, the source files may be modified up to certain level according to the user requirements. A design consideration included using the standard MATLAB functions and procedures without a need of a specialized toolbox. The source codes can be downloaded from https://idm.kky.zcu.cz/sw.

A. Scalar LTI Model

In the first example, scalar LTI models (1) and (2) are with $n_x = n_z = 1$, $F_k = 0.5$, $G_k = 1$, $H_k = 2$, and $u_k = 0$, $\forall k$. For the purposes of the *simulation*, the state noise w_k is assumed to have the Gaussian PDF with the mean $\mathsf{E}[w_k] = 0$ and variance $\mathrm{var}[w_k] = Q = 1$, that is, $p(w_k) = \mathcal{N}\{w_k; 0, Q\}$. The measurement noise v_k is assumed to be zero mean with variance $\mathrm{var}[v_k] = R = 2$ with the PDF defined in the following two *scenarios*:

a) Gaussian PDF $p(v_k) = \mathcal{N}\{v_k; 0, R\}$

b) Student's t PDF[†††] $p(v_k) = St\{v_k; \eta\}$, where the DoF $\eta = (2R/R - 1) = 4$

The considered number of measured data per MC simulation is $\tau = 10^3, 5 \times 10^3, 10^4$, and 10^5. The NEGA method is designed with the observability matrix length $L = 2n_x = 2$ (i.e., $L^+ = 3$) and with two levels of significance, $\alpha = 0.01$ and $\alpha = 0.002$. The results are given in Tables 1 and 2.

In Table 1, the quality of the noise CMs estimates provided by the NEGA method is assessed in terms of the average estimate value and estimate error variance. The results confirm that the NEGA method provides unbiased and consistent estimates as was theoretically proven in [9]. It can also be seen that estimate error variance $\widehat{\mathrm{var}}[\hat{R}]$ is strongly affected by the considered distribution of the measurement noise. For the Student's t–distributed measurement noise, the variance of the noise variance error is significantly greater. The reason can be found in the fact that the Student's t distribution resembles a more heavy-tailed distribution than the Gaussian distribution.

Table 1 Average noise CMs estimates and estimated error variance for LTI model

	$p(v_k)$	$\hat{\mathsf{E}}[\hat{Q}]$	$\widehat{\mathrm{var}}[\hat{Q}]$	$\hat{\mathsf{E}}[\hat{R}]$	$\widehat{\mathrm{var}}[\hat{R}]$
$\tau = 10^3$	Gaussian	1.0016	0.0153	1.9974	0.1838
	Student's	1.0012	0.0153	2.0009	0.2664
$\tau = 5 \times 10^3$	Gaussian	1.0007	0.0031	1.9968	0.0368
	Student's	1.0004	0.0030	1.9977	0.0502
$\tau = 10^4$	Gaussian	1.0002	0.0015	1.9984	0.0179
	Student's	1.0008	0.0015	1.9971	0.0271
$\tau = 10^5$	Gaussian	1.0000	0.0001	2.0001	0.0018
	Student's	1.0001	0.0002	1.9998	0.0029

In Table 2, the observed number of MC simulations, where the null hypothesis H_0 was rejected, is summarized for all five considered statistical tests. For normally distributed measurement noise, $M_{H0\text{-rejected,theor}}$ MC simulations with rejected null hypothesis (caused by the defined probability of FA α) should theoretically be observed, where

$$M_{H0\text{-rejected,theor}} = M \times \alpha \tag{46}$$

For the Student's t–distributed measurement noise, it would be ideal to observe M in MC simulations with rejected null hypothesis. The results show that for the Gaussian scenario all the tests reject the null hypothesis with the expected rate; that is, the number of rejected null hypothesis $M_{H0\text{-rejected}}$ is close to the theoretical one [Eq. (46)] and almost independent on the number of data τ. The Student's t scenario illustrates the power of the test, that is, the situation where the null hypothesis is *correctly* rejected. It can be seen that the particular tests have different power and also the power strongly depends on the number of measured data τ (with increasing number of data, the statistical power increases as well [12]). However, the tests do not significantly depend on the selected α. From all the considered tests, it seems that the JB and SW tests are the most suitable ones to be used within the NEGA method as they demonstrate the highest power of the test (rejected null hypothesis in Student's t scenario) independently of the length τ of the measurement sequence.

The last experiment within this example assesses the noise variances estimates in terms of the error variances $\widehat{\mathrm{var}}[\hat{Q}]$, $\widehat{\mathrm{var}}[\hat{R}]$ w.r.t. the choice of the design parameter L in scenario "a." As mentioned in Sec. V.B. the parameter needs to be selected as $L \geq n_x + 1 = 2$. The greater the value of L is selected, the higher the number of linear equations used for Q and R estimation in Eq. (43), and thus better estimates can be obtained (as the set of linear equations in Eq. (43) is not optimally weighted, the variance of the estimates need not necessarily decrease with the increasing number of equations [9]). This is illustrated in Fig. 1, where estimates variances $\widehat{\mathrm{var}}[\hat{Q}]$ and $\widehat{\mathrm{var}}[\hat{R}]$ are plotted against the parameter L. The plot indicates that the best performance w.r.t. minimal estimate variance can be reached for $L = 3$. Nevertheless, it must be pointed out that such a conclusion is valid for the considered setup only; for different models a different L may be optimal. Specification of the optimal number of equations in a correlation method is further discussed in [9].

B. Scalar LTV Model

In the second example, a scalar LTV model [Eqs. (1) and (2)] is considered, where $F_k = 0.5 + 0.4\sin(2\pi k/\tau)$, $H_k = 2 + \sin(10\pi k/\tau)$, $G_k = 1$, and $u_k = 0$, $\forall k$. The state and measurement noises are defined analogously to the LTI model defined in the previous section.

The simulation was performed for the number of data $\tau = 2 \times 10^3$ for both PDFs of the measurement noise in the MC simulations with $L = 2$ and $\alpha = 0.01$.

[†††]The Student's t-distributed noises have recently attracted significant attention, and a solution to the Bayesian recursive relations for models with Student's t-distributed noises was proposed [15].

Table 2 Observed number of MC with rejected null hypothesis for LTI model

		$\alpha = 0.01$					$\alpha = 0.002$				
	$p(v_k)$	Chi2GoF	AD	JB	SW	LF	Chi2GoF	AD	JB	SW	LF
$\tau = 10^3$	Gaussian	102	82	73	101	104	21	16	23	32	14
	Student's	554	1310	5401	4978	515	191	674	4261	3871	177
$\tau = 5 \times 10^3$	Gaussian	92	94	91	106	81	21	21	18	24	15
	Student's	5453	6604	9844	9790	2743	3823	4892	9744	9630	1198
$\tau = 10^4$	Gaussian	113	94	79	92	98	27	24	17	20	16
	Student's	9161	9592	10^4	9998	6263	8389	8994	9998	9996	3911
$\tau = 10^5$	Gaussian	92	94	77	101	108	25	19	19	24	20
	Student's	9163	10^4	10^4	10^4	10^4	9088	10^4	10^4	10^4	10^4

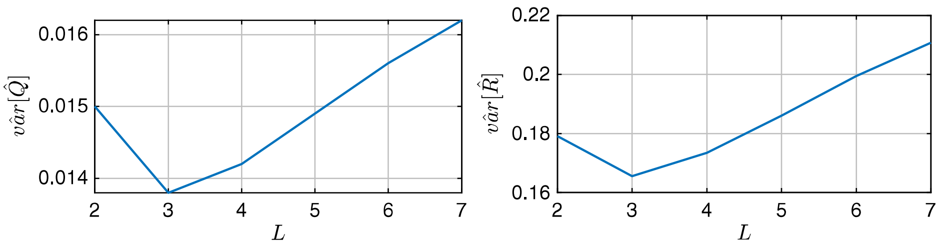

Fig. 1 Noise CMs estimate variance as a function of a design parameter.

In the previous LTI example, the AMPE was directly used as a test statistic (except the Chi2GoF test); thus, the tests were not influenced by the noise variances estimates quality. If an LTV model is considered, then the AMPE has to be stochastically decoupled according to Eq. (45) to ensure that the test statistic ε_k^L is identically distributed. Therefore, the test statistic ε_k^L depends on the quality of the noise CMs estimates through the matrix $\hat{S}_k$ in Eq. (45). The goal of the example is to illustrate the impact of the noise variances estimates $\hat{Q}$ and $\hat{R}$ on the performance of the JB and SW statistical tests, which exhibited the best performance in the previous example.

In each MC simulation, a sequence of measured data $z_k, k = 0, 1, \ldots, \tau = 2000$, is generated. Clearly, however, it is not possible to use all data for Q and R estimation and the same data for test statistic computation to prevent multiple uses of the same data. Therefore, it is necessary to split each MC data set into two parts:

1) Data for noise variances estimation
2) Data for Gaussianity assessment

In this section four different splitting ratios are considered, namely:

r1. Data $\{z_k\}_{k=0}^{400}$ for noise variances estimation and remaining data $\{z_k\}_{k=401}^{2000}$ for test statistic computation and Gaussianity assessment
r2. Data $\{z_k\}_{k=0}^{800}$ for estimation and data $\{z_k\}_{k=801}^{2000}$ for assessment
r3. Data $\{z_k\}_{k=0}^{1200}$ for estimation and data $\{z_k\}_{k=1201}^{2000}$ for assessment
3 r4. Data $\{z_k\}_{k=0}^{1600}$ for estimation and data $\{z_k\}_{k=1601}^{2000}$ for assessment

It means that the least precise and the most precise estimates of the noise variances Q and R can be expected for ratios $r1$ and $r4$, respectively. The question is how the variances estimate accuracy affects the performance and power of the statistical tests. The results are summarized in Table 3.

Table 3 confirms that, with the increasing number of measurements, the NEGA method provides estimates $\hat{Q}$ and $\hat{R}$ with a decreasing error variance. Such behavior is according to the theoretical expectations. However, more important are the observed simulations with rejected null hypothesis $M_{H0\text{-rejected}}$. It can be seen that for correct assessment of the noises' Gaussianity it is better to have rather a longer data set with possibly less accurate estimates $\hat{Q}$ and $\hat{R}$. Following the observations, if a limited set of data is available, then it is better to use smaller portion of the data for the noise CMs estimation and the remaining larger portion of data for the Gaussianity assessment. Note also that the same conclusions can be drawn if the later data are used for the estimation and earlier for the assessment.

C. Multidimensional LTV Model

The third example aims to illustrate the influence of model variability on the estimability of the noise CMs. Let two two-dimensional LTV models, (1) and (2), be considered with the following matrices:

$$M1.\ \boldsymbol{F}_k = \begin{bmatrix} f_k & 0 \\ 1 & 0.8 \end{bmatrix},\ \boldsymbol{G}_k = \boldsymbol{I}_2,\ \boldsymbol{H}_k = [0 \quad h_k]$$

$$M2.\ \boldsymbol{F}_k = \begin{bmatrix} 0.8 & 0 \\ 1 & f_k \end{bmatrix},\ \boldsymbol{G}_k = \boldsymbol{I}_2,\ \boldsymbol{H}_k = [0 \quad h_k]$$

where $n_x = 2$, $n_z = 1$, $f_k = 0.5 + 0.4\sin(2\pi k/\tau)$, $h_k = 2 + \sin(10\pi k/\tau)$, and $\boldsymbol{u}_k = \boldsymbol{0}_{2\times 1}$, $\forall k$. Two models, thus, differ only in the diagonal values of the dynamic matrix $\boldsymbol{F}_k$. Note also that the coefficients f_k and h_k are strictly positive $\forall k$.

The state noise CM $\boldsymbol{Q}$ is a 2×2 matrix with three unknown unique elements. The measurement noise variance R is an unknown scalar. Therefore, in total, the state and measurement noise CMs contain *four* unknowns, which are gathered in the vector $\boldsymbol{\theta}$ [Eq. (25)]. Noise CMs estimation for models (1) and (2) is based on a solution to the system of linear Eq. (30), in particular on the assumption of full column rank of the matrix $\boldsymbol{\Lambda}$. The matrix $\boldsymbol{\Lambda}$ is a known[‡‡‡] matrix computed from the model matrices $\boldsymbol{F}_k$ and $\boldsymbol{H}_k$, $\forall k$. The rank of the $\boldsymbol{\Lambda}$ computed for both models and several choices of the design parameter L is summarized in Table 4.

From Table 4 it can be seen that, although both models are similarly time varying, all elements of the noise CMs can be estimated only for model $M1$. For model $M2$, only three elements of the noise CMs can be found at best (out of which R is estimated and two elements of $\boldsymbol{Q}$). Note that, as the matrix $\boldsymbol{\Lambda}$ is known, determination of its rank (and thus of the number of estimable elements of $\boldsymbol{Q}$ and $\boldsymbol{R}$) does not impose any limitation on the proposed NEGA method.

D. Multidimensional LTI Model: Target Tracking

The last example illustrates performance of the NEGA method using the nearly constant velocity model with position measurements [6,7] with $n_x = 4$, $n_w = 2$, $n_z = 2$, and the following matrices

[‡‡‡]The particular values of the noise CMs are not important for this example; thus, they are not defined.

Table 3 Average noise CMs estimates, estimated error variance, and observed number of MC with rejected null hypothesis for LTV model

	$p(v_k)$	$\hat{\mathsf{E}}[\hat{Q}]$	$\widehat{\mathrm{var}}[\hat{Q}]$	$\hat{\mathsf{E}}[\hat{R}]$	$\widehat{\mathrm{var}}[\hat{R}]$	JB	SW
$r1$	Gaussian	0.9991	0.0355	2.0069	0.3025	115	123
	Student's	0.9983	0.0402	1.9936	0.4911	8613	8293
$r2$	Gaussian	1.0009	0.0194	1.9998	0.1553	71	94
	Student's	1.0001	0.0247	1.9986	0.3007	7593	7184
$r3$	Gaussian	1.0005	0.0131	1.9989	0.1097	60	78
	Student's	0.9998	0.0159	1.9968	0.2054	5984	5561
$r4$	Gaussian	0.9989	0.0096	2.0013	0.0873	63	89
	Student's	1.0006	0.0120	1.9888	0.1602	3585	3330

Table 4 Observability of noise CMs elements for multidimensional LTV model

	$M1$	$M2$
$L = n_x = 2$	4	2
$L = 3, L = 4$	4	3

Table 5 Average noise CMs estimates, estimated error variance, and observed number of MC with rejected null hypothesis for multidimensional LTI model

	$p(v_k)$	$\hat{\mathsf{E}}[\hat{q}]$	$\widehat{\mathrm{var}}[\hat{q}]$	$\hat{\mathsf{E}}[\hat{r}]$	$\widehat{\mathrm{var}}[\hat{r}]$	***JB***	***SW***
$\tau = 10^3$	Gaussian	0.1005	0.0043	1.0003	0.0038	96	98
	Gaussian mixture	0.1007	0.0043	1.0001	0.0050	6842	6352
$\tau = 10^4$	Gaussian	0.0996	0.0004	1.0002	0.0004	92	96
	Gaussian mixture	0.1003	0.0004	0.9997	0.005	10^4	10^4

$$\boldsymbol{F}_k = \begin{bmatrix} 1 & \Delta t & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & \Delta t \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad \boldsymbol{G}_k = \begin{bmatrix} (\Delta t)^2/2 & 0 \\ \Delta t & 0 \\ 0 & (\Delta t)^2/2 \\ 0 & \Delta t \end{bmatrix},$$

$$\boldsymbol{H}_k = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \tag{47}$$

where the sampling period is $\Delta t = 1$. The state noise is assumed to be Gaussian $p(\boldsymbol{w}_k) = \mathcal{N}\{\boldsymbol{w}_k; \boldsymbol{0}_{2\times 1}, q\boldsymbol{I}_2\}$ with $q = 0.1$ and the measurement noise to be either

a) Gaussian PDF $p(\boldsymbol{v}_k) = \mathcal{N}\{\boldsymbol{v}_k; \boldsymbol{0}_{2\times 1}, r\boldsymbol{I}_2\}$ with $r = 1$ or

b) Gaussian sum PDF $p(\boldsymbol{v}_k) = 0.1\mathcal{N}\{\boldsymbol{v}_k; [1, 1]^T, 2 \times \boldsymbol{I}_2\} + 0.9\mathcal{N}\{\boldsymbol{v}_k; [-0.11, -0.11]^T, 0.76 \times \boldsymbol{I}_2\}$.

Note that both measurement noise PDFs have the *same* first two moments, that is, with $\mathsf{E}[\boldsymbol{v}_k] = \boldsymbol{0}_{2\times 1}$ and $\mathrm{cov}[\boldsymbol{v}_k] = r\boldsymbol{I}_2$. The NEGA method is designed with the observability matrix length $L = 2$ and with the level of significance $\alpha = 0.01$. The results are given in Table 5 and confirm high-quality estimates of the noise variances q and r. Also, the number of the observed simulations with rejected null hypothesis corresponds with $M_{H0\text{-rejected,theor}}$ [Eq. (46)] for the Gaussian measurement noise. On the other hand, for the Gaussian mixture measurement noise, the number of the observed simulations with rejected null hypothesis is significantly higher and converge to M with the increasing number of data τ.

VI. Conclusions

The paper presents a complete analytical derivation of the NEGA method designed for estimation of the noise CMs and noises Gaussianity assessment for LTI and LTV models with the state noise shaping matrix. The method provides unbiased and consistent estimates of the noise CMs and a hypothesis testing based decision on whether the noises are Gaussian or not. Combination of these properties is unique in the state-of-the-art noise CMs estimation methods and is important for design of many optimal navigation and tracking algorithms requiring outputs with ensured integrity. The performance of the NEGA method has been illustrated in an extensive simulation study using four examples. The simulations confirmed all the theoretically derived properties of the method. The paper is accompanied with the exemplary MATLAB implementations of the method for the both LTI and LTV models.

Acknowledgments

This paper was based on the work supported by the Air Force Office of Scientific Research, Air Force Material Command, United States Air Force (USAF), under Award No. FA9550-16-1-0511. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the Air Force Office of Scientific Research, Air Force Materiel Command, USAF.

References

[1] Mehra, R. K., "On the Identification of Variances and Adaptive Filtering," *IEEE Transactions on Automatic Control*, Vol. 15, No. 2, 1970, pp. 175–184.
https://doi.org/10.1109/TAC.1970.1099422

[2] Duník, J., Straka, O., Kost, O., and Havlík, J., "Noise Covariance Matrices in State-Space Models: A Survey and Comparison—Part I," *International Journal of Adaptive Control and Signal Processing*, Vol. 31, No. 11, 2017, pp. 1505–1543.
https://doi.org/10.1002/acs.v31.11

[3] Olson, C. G., Russell, R. P., and Carpenter, J. R., "Precomputing Process Noise Covariance for Onboard Sequential Filters," *Journal of Guidance, Control, and Dynamics*, Vol. 40, No. 8, 2017, pp. 2062–2075.
https://doi.org/10.2514/1.G002144

[4] Kost, O., Duník, J., and Straka, O., "Noise Moment and Parameter Estimation of State-Space Model," *Proceedings of the 8th IFAC Symposium on System Identification*, Stockholm, 2018.
https://doi.org/10.1016/j.ifacol.2018.09.107

[5] Duník, J., Kost, O., Straka, O., and Blasch, E., "State and Measurement Noise in Positioning and Tracking: Covariance Matrices Estimation and Gaussianity Assessement," *Proceedings of the 2018 IEEE/ION Position Location and Navigation Symposium*, Monterey, CA, 2018.
https://doi.org/10.1109/PLANS.2018.8373523

[6] Bar-Shalom, Y., Li, X. R., and Kirubarajan, T., *Estimation with Applications to Tracking and Navigation: Theory Algorithms and Software*, Wiley, 2001.

[7] Graham, M. C., How, J. P., and Gustafson, D. E., "Robust State Estimation with Sparse Outliers," *Journal of Guidance, Control, and Dynamics*, Vol. 38, No. 7, 2015, pp. 1229–1240.
https://doi.org/10.2514/1.G000350

[8] Brink, K. M., "Partial-Update Schmidt-Kalman Fitler," *Journal of Guidance, Control, and Dynamics*, Vol. 40, No. 9, 2017, pp. 2214–2228.
https://doi.org/10.2514/1.G002808

[9] Duník, J., Kost, O., and Straka, O., "Design of Measurement Difference Autocovariance Method for Estimation of Process and Measurement Noise Covariances," *Automatica*, Vol. 90, April 2018, pp. 16–24.
https://doi.org/10.1016/j.automatica.2017.12.040

[10] Brewer, J. W., "Kronecker Products and Matrix Calculus in System Theory," *IEEE Transactions on Circuits and Systems*, Vol. 25, No. 9, 1978, pp. 772–781.
https://doi.org/10.1109/TCS.1978.1084534

[11] Odelson, B. J., Rajamani, M. R., and Rawlings, J. B., "A New Autocovariance Least-Squares Method for Estimating Noise Covariances," *Automatica*, Vol. 42, No. 2, 2006, pp. 303–308.
https://doi.org/10.1016/j.automatica.2005.09.006

[12] Papoulis, A., and Pillai, S. U., *Probability, Random Variables and Stochastic Processes*, 4th ed., McGraw–Hill, 2002.

[13] Razali, N. M., and Wah, Y. B., "Power Comparison of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors, and Anderson-Darling Tests," *Journal of Statistical Modeling and Analytics*, Vol. 2, No. 1, 2011, pp. 21–33.

[14] Thadewald, T., and Büning, H., "Jarque-Bera Test and Its Competitors for Testing Normality—A Power Comparison," *Journal of Applied Statistics*, Vol. 34, No. 1, 2007, pp. 87–105.
https://doi.org/10.1080/02664760600994539

[15] Roth, M., Özkan, E., and Gustafsson, F., "A Student's-t Filter for Heavy Tailed Process and Measurement Noise," *Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, Vancouver, 2013.
https://doi.org/10.1109/ICASSP.2013.6638770