ISSN 1064-2307, Journal of Computer and Systems Sciences International, 2019, Vol. 58, No. 6, pp. 869–882. © Pleiades Publishing, Ltd., 2019.
Russian Text © The Author(s), 2019, published in Izvestiya Akademii Nauk, Teoriya i Sistemy Upravleniya, 2019, No. 6, pp. 43–56.

CONTROL IN STOCHASTIC SYSTEMS AND UNDER UNCERTAINTY CONDITIONS

Two-Stage Algorithm for Estimation of Nonlinear Functions of State Vector in Linear Gaussian Continuous Dynamical Systems[1]

Won Choi[a],*, Il Young Song[b],, and Vladimir Shin[c],*****

[a]*Incheon National University, Incheon, 22012 Republic of Korea*
[b]*Hanwha Corporation, Daejeon, Republic of Korea*
[c]*Gyeongsang National University, Jinju, 660701 Republic of Korea*
**e-mail: choiwon@inu.ac.kr*
***e-mail: com21dud@hanwha.com*
****e-mail: vishin@gnu.ac.kr*

Received September 18, 2018; revised May 13, 2019; accepted July 22, 2019

Abstract—This paper focuses on the optimal minimum mean square error estimation of a nonlinear function of state (NFS) in linear Gaussian continuous-time stochastic systems. The NFS represents a multivariate function of state variables which carries useful information of a target system for control. The main idea of the proposed optimal estimation algorithm includes two stages: the optimal Kalman estimate of a state vector computed at the first stage is nonlinearly transformed at the second stage based on the NFS and the minimum mean square error (MMSE) criterion. Some challenging theoretical aspects of analytic calculation of the optimal MMSE estimate are solved by usage of the multivariate Gaussian integrals for the special NFS such as the Euclidean norm, maximum and absolute value. The polynomial functions are studied in detail. In this case the polynomial MMSE estimator has a simple closed form and it is easy to implement in practice. We derive effective matrix formulas for the true mean square error of the optimal and suboptimal quadratic estimators. The obtained results we demonstrate on theoretical and practical examples with different types of NFS. Comparison analysis of the optimal and suboptimal nonlinear estimators is presented. The subsequent application of the proposed estimators demonstrates their effectiveness.

DOI: 10.1134/S1064230719060169

INTRODUCTION

Estimation and filtering are powerful techniques for building models of complex control systems. The Kalman filtering and its variations are well-known state estimation techniques in wide use in a variety of applications such as navigation, target tracking, communications engineering, biomedical and chemical processing and other areas [1–6]. However, in many applications it is of interest to estimate not only a state vector $x_t \in \mathbb{R}^n$ but also a nonlinear function of the state vector (NFS), $z_t = f(x_t)$, which expresses practical and worthwhile information for control systems. The first motivating example of the NFS is the location of a target and radar. An angle (φ) and distance (d) from radar to target are shown in Fig. 1:

$$\varphi = f(x) = \arctan\left(\frac{x_3}{\sqrt{x_1^2 + x_2^2}}\right),$$
$$d = f(x) = \sqrt{x_1^2 + x_2^2 + x_3^2}\ .$$

The second example can be an arbitrary quadratic form, $f(x_t) = x_t^T \Omega x_t$, representing an energy-like function of an object [7], or the Euclidean distance (2-norm), $f(x_t) = \|x_t - x_t^n\|$, between the current x_t

[1] This work was supported by the Incheon National University Research Grant in 2015–2016.

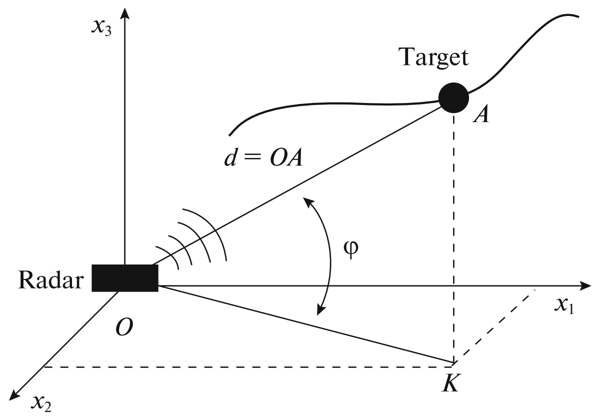

Fig. 1. Location of radar and target: $A(x_1, x_2, x_3)$, $d = OA$, $\tan\varphi = \dfrac{AK}{OK}$.

and nominal x_t^n states, respectively. So estimation and prediction of quantities represented by an NFS can be helpful in different applications, such as system control and target tracking.

The problem of estimation of nonlinear functions with unknown parameters and signals based on the minimax theory has been studied by many authors [8–11] and the references therein. Estimation of parameters of nonlinear functional model with known error covariance matrix is presented in [12]. Minimax quadratic estimate for integrated squared derivative of a periodic function is derived in [13]. In [14, 15], the optimal matrix of a quadratic function is searching based on cumulant criterions. Estimation of penalty considered as a quadratic cost functional for quantum harmonic oscillator is given in [16]. Estimators for integrals of nonlinear functions of a probability density are developed in [17, 18]. We also mention estimation of nonlinear functions of spectral density, periodogram of a stationary linear signals [19–21]. Some extension of these results obtained by [22]. In [23] an unknown distance between a target and radar is approximated by Taylor polynomial to subsequent estimation of its coefficients. For algorithms and theory for estimation information measures representing an nonlinear function of signals the reader is referred to [24, 25]. However, most authors have not focused on estimation of an NFS for vector signals defined by dynamical models, such as stochastic differential systems.

The aim of this paper is to develop an optimal two-stage minimum mean square error (MMSE) estimator for an arbitrary NFS in a linear Gaussian stochastic differential systems, and considerably study a special polynomial estimators for which one can obtain an important mean square estimation results. The main contributions of the paper are listed in the following:

1. This paper studies the estimation problem of an NFS within the continuous Kalman filtering framework. Using the mean square estimation approach, an optimal two-stage nonlinear estimator is proposed.

2. The optimal MMSE estimator for a polynomial functions (quadratic, cubic and quartic) is derived. We establish that the polynomial estimator represents a compact closed-form formula depends only on the Kalman estimate and error covariance.

3. Important class of quadratic estimators is comprehensively investigated, including derivation of a matrix equation for its true mean square error (MSE).

4. Performance of the proposed estimators for real NFS illustrates their theoretical and practical usefulness.

This paper is organized as follows. Section 1 presents a statement of the MMSE estimation problem for an NFS within the Kalman filtering framework. In Section 2, the general optimal MMSE estimator is proposed. Here we study the comparative analysis of the optimal and suboptimal estimators via several theoretical examples with a practical NFS. In Section 3, the importance of obtaining an optimal estimator in a closed form is studied. An optimal polynomial estimator represents a closed form expression in terms of the Kalman estimate and its error covariance. For an optimal and suboptimal quadratic estimators we derive matrix formulas for the true MSEs. The efficiency of the quadratic estimators is studied for a scalar random signal and on real model of the wind tunnel system.

1. PROBLEM STATEMENT

The Kalman framework involves estimation of the state of a continuous-time linear Gaussian dynamic system with additive white noise,

$$\begin{aligned} \dot{x}_t &= F_t x_t + G_t v_t, \quad t \geq 0, \\ y_t &= H_t x_t + w_t. \end{aligned} \tag{1.1}$$

Here, $x_t \in \mathbb{R}^n$ is a state vector, $y_t \in \mathbb{R}^m$ is an observation vector, $v_t \in \mathbb{R}^r$ and $w_t \in \mathbb{R}^m$ are zero-mean Gaussian white noises with intensities Q_t and R_t, respectively, i.e., $\mathbf{E}(v_t v_s^{\mathrm{T}}) = Q_t \delta_{t-s}$, $\mathbf{E}(w_t w_s^{\mathrm{T}}) = R_t \delta_{t-s}$, δ_t is the Dirac delta-function, $F_t \in \mathbb{R}^{n\times n}$, $G_t \in \mathbb{R}^{n\times r}$, $Q_t \in \mathbb{R}^{r\times r}$, $R_t \in \mathbb{R}^{m\times m}$ and $H_t \in \mathbb{R}^{m\times n}$, $\mathbf{E}(X)$ is the expectation of a random vector X.

We assume that the initial state $x_0 \sim \mathbb{N}(\bar{x}_0, P_0)$, and system and observation noises v_t, w_t are mutually uncorrelated.

A problem associated with such system (1.1) is that of estimation of the nonlinear function of state vector

$$z_t = f(x_t) : \mathbb{R}^n \to \mathbb{R} \tag{1.2}$$

from the overall noisy observations $y_0^t = \{y_s : 0 \leq s \leq t\}$.

There are a multitude of statistics-based methods to estimate an unknown function $z = f(x)$ from real sensor observations y_0^t. We focus on choosing the best estimate $\hat{z}$ minimizing MSE: $\min_{\hat{z}} \mathbf{E}[\|z - \hat{z}\|^2]$. In general the optimal MMSE solution (further "estimate" or "estimator") is given by the conditional mean, $\hat{z} = \mathbf{E}(z|y_0^t)$ [26, 27]. The most challenging problem is how to calculate the conditional mean. In this paper, we solve this problem for the NFS (1.2) within the Kalman filtering framework.

We propose optimal and suboptimal MMSE estimation algorithms for the NFS and their implementation in next section.

2. OPTIMAL TWO-STAGE MMSE ESTIMATOR FOR GENERAL NFS

Here the optimal two-stage estimator for the general NFS is derived. Also we propose a simple suboptimal estimator. The both estimators include two stages: the optimal Kalman estimate of the state vector $\hat{x}_t$, computed at the first stage is used at the second stage for estimation of the NFS (1.2).

2.1. Optimal Two-Stage Algorithm

First stage (calculation of Kalman estimate). The estimate $\hat{x}_t = \mathbf{E}(x_t \mid y_0^t)$ of the state x_t based on the observations y_0^t, and its error covariance $P_t = \mathbf{E}(e_t e_t^{\mathrm{T}})$, $e_t = x_t - \hat{x}_t$, are given by the continuous Kalman-Bucy filter (KBF) equations [4–6]:

$$\begin{aligned} \dot{\hat{x}}_t &= F_t \hat{x}_t + K_t (y_t - H_t \hat{x}_t), \quad \hat{x}_{t=0} = \bar{x}_0, \\ K_t &= P_t H_t^{\mathrm{T}} R_t^{-1}, \quad \tilde{G}_t = G_t Q_t G_t^{\mathrm{T}}, \\ \dot{P}_t &= F_t P_t + F_t^{\mathrm{T}} P_t - P_t H_t^{\mathrm{T}} R_t^{-1} H_t P_t + \tilde{G}_t, \quad P_{t=0} = P_0. \end{aligned} \tag{2.1}$$

Second stage (optimal estimator for NFS). The optimal estimate of the NFS $z_t = f(x_t)$ based on the observations y_0^t also represents a conditional mean, that is

$$\hat{z}_t = \mathbf{E}(z_t|y_0^t) = \int_{\mathbb{R}^n} f(x) p_t(x|y_0^t) dx, \tag{2.2}$$

where $p_t(x|y_0^t) = \mathbb{N}(\hat{x}_t, P_t)$ is a multivariate conditional Gaussian probability density function determining by the conditional mean $\hat{x}_t = \mathbf{E}(x_t \mid y_0^t)$ and covariance $P_t = \mathrm{cov}(x_t, x_t|y_0^t) = \mathbf{E}(e_t e_t^{\mathrm{T}})$.

Thus, the estimate in (2.2) represents the optimal MMSE estimator for the general NFS which depends on the Kalman estimate $\hat{x}_t$ and its error covariance P_t.

In practice, the nonlinear function (1.2) may depend not only on the state vector, but also on its estimate. Then the NSF takes the form

$$z_t = f(x_t, \hat{x}_t) : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}. \tag{2.3}$$

Taking into account that the state estimate $\hat{x}_t$ represents the known function of observations, $\hat{x}_t = \hat{x}_t(y_0^t)$, the MMSE estimate of an unknown function, $z_t = f(x_t, \hat{x}_t)$ is given by the formula similar to (2.2). Namely

$$\hat{z}_t = \mathbf{E}[f(x_t, \hat{x}_t(y_0^t))|y_0^t] = \int_{\mathbb{R}^n} f(x, \hat{x}_t) p_t(x|y_0^t)dx = \int_{\mathbb{R}^n} f(x, \hat{x}_t)\mathbb{N}(\hat{x}_t, P_t)dx. \tag{2.4}$$

Further, we consider several theoretical examples of application of the general nonlinear estimators (2.2).

2.2. Examples of Two-Stage Estimator

Let $x_t = [x_{1,t} \;\; x_{2,t}]^{\mathrm{T}}$ be a bivariate Gaussian vector; and $\hat{x}_t \in \mathbb{R}^2$ and $P_t \in \mathbb{R}^{2\times 2}$ are the Kalman estimate and error covariance, respectively.

Example 1 (expected value of the Euclidean norm). The overall estimation error is defined as

$$\|e_t\| = \sqrt{e_{1,t}^2 + e_{2,t}^2}, \quad e_t = x_t - \hat{x}_t, \quad e_t = [e_{1,t} \;\; e_{2,t}]^{\mathrm{T}}. \tag{2.5}$$

Taking into account that the Euclidean norm of the error depends on the difference between state vector and its estimate, $\|e_t\| = \sqrt{e_t^{\mathrm{T}} e_t} = \sqrt{(x_t - \hat{x}_t)^{\mathrm{T}}(x_t - \hat{x}_t)}$, one transform the formula (2.2). We have

$$\hat{z}_t = \mathbf{E}[\|e_t\||y_0^t] = \int (\sqrt{e^{\mathrm{T}}e}) p_t(e|y_0^t)de = \int (\sqrt{e^{\mathrm{T}}e})\mathbb{N}(0, P_t)\, de,$$

where $p_t(e|y_0^t) = \mathbb{N}(0, P_t)$ is the conditional probability density function of the error $e_t = x_t - \hat{x}_t$.

Next using [28] we have the analytical expression for the MMSE estimate of the Euclidean norm $z_t = \|e_t\|$,

$$\begin{gathered} \hat{z}_t = \mathbf{E}(\|e_t\||y_0^t) = \frac{2\lambda_{1,t}\lambda_{2,t}\sqrt{\pi}}{(\lambda_{1,t} + \lambda_{2,t})^{3/2}}\, {}_2F_1\left(\frac{3}{4}, \frac{5}{4}; 1; q_t^2\right), \\ q_t^2 = \frac{\lambda_{1,t} - \lambda_{2,t}}{\lambda_{1,t} + \lambda_{2,t}}, \end{gathered} \tag{2.6}$$

where ${}_2F_1(a, b; c; x)$ is the hypergeometric function, and $\lambda_{1,\,t}$ and $\lambda_{2,\,t}$ are the eigenvalues of P_t.

Example 2 (feedback control depends on maximum of state coordinate). In mechanical systems a piecewise feedback control law is given by

$$u_t = \begin{cases} 1, & \max(x_{1,t}, x_{2,t}) > D, \\ -1, & \max(x_{1,t}, x_{2,t}) \le D, \end{cases}$$

where D is a distance threshold. Then the MMSE estimate of the maximum $z_t = \max(x_{1,t}, x_{2,t})$ is obtained from [29],

$$\begin{aligned} \hat{z}_t = \mathbf{E}[\max(x_{1,t}, x_{2,t})|y_0^t] &= \hat{x}_{1,t}\Phi\left(\frac{\hat{x}_{1,t} - \hat{x}_{2,t}}{\theta_t}\right) \\ &+ \hat{x}_{2,t}\Phi\left(\frac{\hat{x}_{2,t} - \hat{x}_{1,t}}{\theta_t}\right) + \theta_t\phi\left(\frac{\hat{x}_{1,t} - \hat{x}_{2,t}}{\theta_t}\right), \\ \theta_t = \sqrt{P_{11,t} + P_{22,t} - 2P_{12,t}}, &\quad P_t = [P_{ij,t}], \end{aligned}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the standard conditional Gaussian density and cumulative distribution function, respectively.

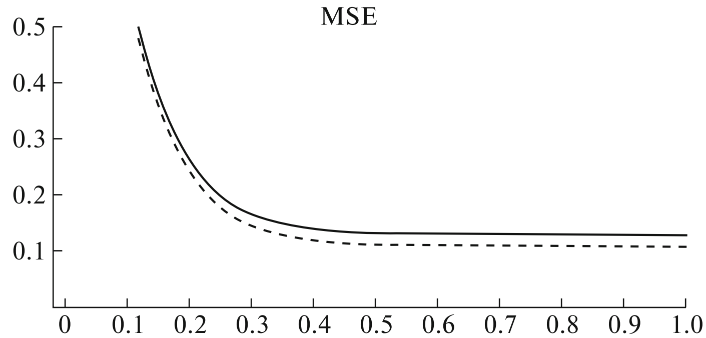

Fig. 2. Optimal MSE (dotted curve) and suboptimal MSE (solid curve) for $z = \sin\theta$.

Example 3 (estimation of sine function). Let θ be a random unknown angle which is measured in the presence of an additive white noise. Then

$$\begin{aligned} \dot{x}_t &= 0, \quad t \ge 0, \quad x_0 = \theta \sim \mathbb{N}(m_\theta, \sigma_\theta^2), \\ y_t &= x_t + w_t, \quad 0 \le t \le T, \end{aligned} \tag{2.7}$$

where $x_t = \theta$, and w_t is a zero-mean Gaussian white noise with intensity r.

The KBF equations (2.1) become

$$\begin{aligned} \dot{\hat{x}}_t &= P_t\,(y_t - \hat{x}_t)/r, \quad \hat{x}_0 = m_\theta, \\ \dot{P}_t &= -P_t^2/r, \quad P_0 = \sigma_\theta^2. \end{aligned} \tag{2.8}$$

Integrating the Riccati equation for $P_t = \mathbf{E}[(\theta - \hat{x}_t)^2]$, we obtain $P_t = \sigma_\theta^2/(1 + t\sigma_\theta^2/r)$. Further, we consider a sine function of an unknown angle θ. Then an NFS becomes $z = f(\theta) = \sin\theta$.

1. Optimal estimate. Using (2.2) the optimal estimate of the sine is

$$\hat{z}_t = \mathbf{E}[\sin\theta | y_0^t] = \int_{\mathbb{R}} \sin\theta \mathbb{N}(\hat{\theta}_t, P_t) d\theta = e^{-P_t/2}\sin\hat{\theta}_t, \tag{2.9}$$

where $\hat{\theta}_t = \hat{x}_t$ and P_t are determined by (2.8).

2. Suboptimal estimate. In parallel with the optimal mean square estimate (2.9) we consider a simple suboptimal estimate, $\tilde{z}_t = f(\hat{\theta}_t) = \sin\hat{\theta}_t$.

To compare estimation accuracy of two estimates, we derive the analytical formulas for their MSE: $P_{z,t}^{\text{opt}} = \mathbf{E}[(\sin\theta - \hat{z}_t)^2]$ and $P_{z,t}^{\text{sub}} = \mathbf{E}[(\sin\theta - \tilde{z}_t)^2]$, and demonstrate a comparative analysis. We obtain

$$P_{z,t}^{\text{opt}} = k_t^{(1)} - 2e^{-P_t/2}k_t^{(2)} + e^{-P_t}k_t^{(3)}, \quad P_{z,t}^{\text{sub}} = k_t^{(1)} - 2k_t^{(2)} + k_t^{(3)},$$

where

$$\begin{aligned} k_t^{(1)} &= \mathbf{E}[\sin^2\theta] = (1 - e^{-2\sigma_\theta^2})\cos(2m_\theta)/2, \\ k_t^{(2)} &= \mathbf{E}[\sin\theta\sin\hat{\theta}_t] = [e^{-P_t/2} - e^{-(4\sigma_\theta^2 - 3P_t)/2}\cos(2m_\theta)]/2, \\ k_t^{(3)} &= \mathbf{E}[\sin^2\hat{\theta}_t] = [1 - e^{-2(\sigma_\theta^2 - P_t)}\cos(2m_\theta)]/2. \end{aligned}$$

Figure 2 illustrates the exact MSEs $P_{z,t}^{\text{opt}}$ and $P_{z,t}^{\text{sub}}$ for the parameters $m_\theta = 0$; $\sigma_\theta^2 = 2$; $r = 1$ and Fig. 3 shows the relative error, $\Delta_t = |(P_{z,t}^{\text{opt}} - P_{z,t}^{\text{sub}})/P_{z,t}^{\text{opt}}| \cdot 100\%$. Not surprisingly, Figs. 2 and 3 illustrate that the optimal estimate is better than suboptimal one, i.e., $P_{z,t}^{\text{opt}} < P_{z,t}^{\text{sub}}$. We also observe that the difference between two estimates becomes negligible as the time increases.

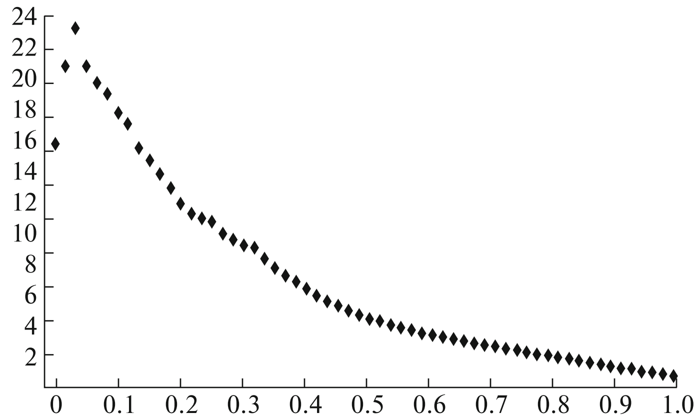

Fig. 3. Relative error Δ_t, % between optimal and suboptimal MSEs for $z = \sin\theta$.

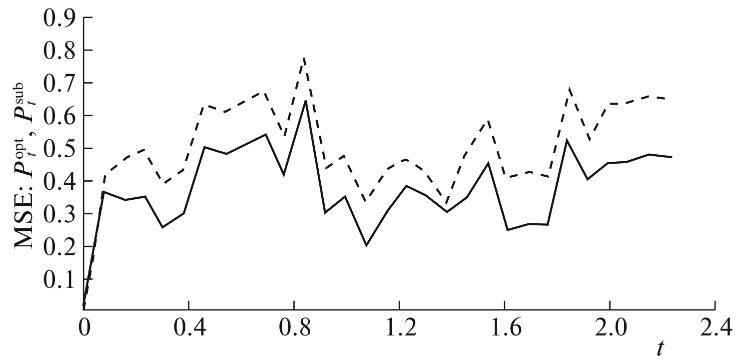

Fig. 4. MSEs for estimates of absolute value $z = |\theta|$: P_t^{opt} (solid curve), P_t^{sub} (dotted curve).

Example 4 (estimation of distance between random location θ and given point a). In this case an NFS becomes $z = |\theta - a|$. Under the model (2.7) for a random location $x_t = \theta$ we use the best estimate (2.2) for the distance $|\theta - a|$. Then

$$\hat{z}_t = \mathbf{E}[|\theta - a||y_0^t] = \int_{-\infty}^{\infty} |\theta - a| \mathbb{N}(\hat{\theta}_t, P_t) d\theta = \sqrt{\frac{2P_t}{\pi}} \exp\left[-\frac{(a-\hat{\theta}_t)^2}{2P_t}\right] + (a - \hat{\theta}_t)\left[2\Phi\left(\frac{a-\hat{\theta}_t}{\sqrt{P_t}}\right) - 1\right],$$

where $\hat{\theta}_t = \hat{x}_t$ and P_t are determined by (2.8).

In the particular case with $a = 0$ the MMSE estimate of an unknown modulus, $z = f(\theta) = |\theta|$, takes the form

$$\hat{z}_t = \sqrt{\frac{2P_t}{\pi}} \exp\left(-\frac{\hat{\theta}_t^2}{2P_t}\right) + \hat{\theta}_t\left[1 - 2\Phi\left(-\frac{\hat{\theta}_t}{\sqrt{P_t}}\right)\right].$$

In addition we can consider a simple suboptimal estimate $\tilde{z}_t = f(\hat{\theta}_t) = |\hat{\theta}_t|$.

To study the behavior of the MSEs, $P_t^{\text{opt}} = \mathbf{E}[(|\theta| - \hat{z}_t)^2]$ and $P_t^{\text{sub}} = \mathbf{E}[(|\theta| - \tilde{z}_t)^2]$ set $m_\theta = 1$; $\sigma_\theta^2 = 1$; $r = 0.2$. To compute the MSEs the Monte-Carlo simulation with 1000 runs was used. As shown in Fig. 4, the optimal estimate $\hat{z}_t$ has a great improvement over the suboptimal one $\tilde{z}_t$.

2.3. Alternative Idea of Suboptimal Estimation of NFS

In contrast to the proposed optimal MMSE solution (2.1) and (2.2) there is an alternative idea to estimate an NFS. In this case the NFS, $z_t = f(x_t)$, is considered as additional state variable z_t which is determined by the nonlinear stochastic equation $\dot{z}_t = a(x_t, z_t) + b(x_t, z_t)v_t$, in which the drift coefficient $a(x_t, z_t)$ and diffusion matrix $b(x_t, z_t)$ are determined by the Ito formula applied to the complex function $f(x_t)$, in which the argument is given by the equation $\dot{x}_t = F_t x_t + G_t v_t$ [30]. Including the variable z_t into the state of a system $x_t \in \mathbb{R}^n$, we obtain system with the augmented state $X_t = [x_t^{\mathrm{T}} \;\; z_t] \in \mathbb{R}^{n+1}$. Thus the problem of estimation of the NFS is reduced to the nonlinear filtering problem by replacing the original state x_t by the augmented one X_t. And approximate nonlinear filtering techniques can be used for simultaneously estimation of x_t and z_t. Many different approximate filters have been proposed [5, 6, 30–35], among which we distinguish the extended Kalman filter [5], the conditionally optimal Pugachev filter with given and optimal structures [30–32] and the unscented Kalman filter [33]. But computational complexity of the approximate nonlinear filters is considerably greater than complexity of the linear Kalman-Bucy filter. The proposed two-stage procedure (2.1) and (2.2) is more promising than estimation of the augmented state X_t.

2.4. Simple Suboptimal Estimator for NFS

In parallel to the optimal estimator $\hat{z}_t = \mathbf{E}(z_t|y_0^t)$ we propose a simple suboptimal estimate of the NFS, such as $\tilde{z}_t = f(\hat{x}_t)$ which depends only on the Kalman estimate of state $\hat{x}_t$ and does not require its error covariance P_t in contrast to the optimal one $\hat{z}_t$. The numerical results show that the suboptimal estimate $\tilde{z}_t$ may be either close to the optimal one (Example 3) or seriously worse (see Example 4 and further Example 6).

2.5. Real-Time Implementation of MMSE Estimator

The error covariance P_t can be pre-computed, because it does not depend on the sensor observations $y_0^t = \{y_s : 0 \le s \le t\}$, but only on the noise statistics Q_t, R_t, the system matrices F_t, G_t, H_t and the initial condition P_0, which are the part of system and observation model (1.1). Thus, once the observation schedule has been settled, the real-time implementation of the MMSE estimator, $\hat{z}_t = \hat{z}_t(\hat{x}_t, P_t)$, requires only the computation of the Kalman estimate $\hat{x}_t$.

2.6. Closed-Form MSE Estimator

For the general NFS, $z_t = f(x_t)$, calculation of the optimal MMSE estimate is reduced to calculation of the multivariate integral (2.2). Analytic calculation of the integral (closed-form MSE estimator) is possible only in special cases as in Examples 1–4.

Further, we consider a polynomial function of state (polynomial form) for which it is possible to obtain a simple closed-form MSE estimators that depend only on the Kalman statistics $(\hat{x}_t, P_t)$.

3. OPTIMAL CLOSED-FORM MMSE ESTIMATOR FOR POLYNOMIAL FUNCTIONS

Let consider a special NFS (1.2) that represents an arbitrary multivariate polynomial function (form) such as,

$$\begin{aligned} &\text{a) Linear form:} && f(x) = Ax; \\ &\text{b) Quadratic form:} && f(x) = x^{\mathrm{T}}Ax, \quad A = A^{\mathrm{T}}; \\ &\text{c) Cubic form:} && f(x) = Axx^{\mathrm{T}}Bx; \\ &\text{d) Quartic form:} && f(x) = x^{\mathrm{T}}Axx^{\mathrm{T}}Bx, \\ &&& A = A^{\mathrm{T}}, \quad B = B^{\mathrm{T}}, \end{aligned} \tag{3.1}$$

where $x \in \mathbb{R}^n$, and $A, B \in \mathbb{R}^{n \times n}$. For simplicity, we ignore the subscript t in this Section.

3.1. Optimal Polynomial Estimators

In case of the polynomial forms (3.1), the optimal estimate $\hat{z} = \mathbf{E}[f(x)|y_0^t]$ has a closed-form solution since the conditional expectation depends on high-order moments of a conditional Gaussian distribution $\mathbb{N}(\hat{x}_t, P_t)$, which can be explicitly calculated in terms of first- and second-order moments, namely, the Kalman estimate and its error covariance $(\hat{x}, P)$. The following theorem gives the best polynomial estimators.

Theorem 1. *The optimal MMSE estimators* $\hat{z} = \mathbf{E}[f(x)|y_0^t]$ *for the polynomial forms* (3.1) *are given by the following analytical formulas*:

$$\begin{aligned}
&\text{a) } \textit{linear estimator}: && \hat{z} = A\hat{x};\\
&\text{b) } \textit{quadratic estimator}: && \hat{z} = \mathrm{tr}(AP) + \hat{x}^{\mathrm{T}}A\hat{x};\\
&\text{c) } \textit{cubic estimator}: && \hat{z} = 2APB\hat{x} + A\hat{x}\hat{x}^{\mathrm{T}}B\hat{x} + A\hat{x}\mathrm{tr}(\mathrm{BP});\\
&\text{d) } \textit{quartic estimator}: && \hat{z} = 2\mathrm{tr}(APBP) + 4\hat{x}^{\mathrm{T}}APB\hat{x}\\
& && + [\mathrm{tr}(AP) + \hat{x}^{\mathrm{T}}A\hat{x}][\mathrm{tr}(BP) + \hat{x}^{\mathrm{T}}B\hat{x}].
\end{aligned} \tag{3.2}$$

The derivation of the estimators (3.2) is given in Appendix.

Example 5 (optimal and suboptimal estimation of the Euclidean norm). To determine an optimal relative location in wireless sensor networks we need to evaluate the cost function representing the Euclidean norm of an overall error,

$$z = \|e\| = \sqrt{e^{\mathrm{T}}e},$$

where e is n-dimensional random error, $e \sim \mathbb{N}(0, P)$

If $n = 2$ the optimal MMSE estimate $\hat{z}$ determined by formulas (2.5) and (2.6),

$$\hat{z} = \mathbf{E}(e_t|y_0^t) = \frac{2\lambda_1\lambda_2\sqrt{\pi}}{(\lambda_1+\lambda_2)^{3/2}}\,{}_2F_1\left(\frac{3}{4},\frac{5}{4};1;q^2\right), \quad q = \frac{\lambda_1-\lambda_2}{\lambda_1+\lambda_2}. \tag{3.3}$$

If $n > 2$ then it is difficult to derive an analytical formula for the optimal estimate $\hat{z}$ such as (3.3). In this case we propose a simple suboptimal estimate $\tilde{z}$ for the Euclidean norm. First, using the quadratic estimator (3.2) we calculate the MMSE estimate of the square norm $\|e\|^2 = e^{\mathrm{T}}e$, and then extract the square root. We have

$$\tilde{z} = \sqrt{\|\hat{e}^2\|} = \sqrt{\mathbf{E}(e^{\mathrm{T}}e|y_0^t)} = \sqrt{\mathrm{tr}(P)}, \tag{3.4}$$

where $\mathrm{tr}(P)$ denotes the trace of a matrix P.

To study behavior of the optimal and suboptimal estimates (3.3) and (3.4) at $n = 2$ we set

$$P = \begin{bmatrix} 1.5 & -0.5 \\ -0.5 & 1.5 \end{bmatrix}, \quad \lambda_1 = 1,\ \lambda_2 = 2,$$
$$q = -\frac{1}{3}, \quad {}_2F_1\left(\frac{3}{4},\frac{5}{4};1;q^2\right) = 1.1169.$$

Then $\hat{z} = 1.524$, $\tilde{z} = 1.732$. The relative error $\left|\frac{\tilde{z}-\hat{z}}{\tilde{z}}\right| \cdot 100\% = 13.6\%$ shows that the suboptimal estimate $\tilde{z}$ is more worse than the optimal one $\hat{z}$.

3.2. Exact Matrix Formulas for Quadratic Estimators

Consider an arbitrary quadratic form (QF)

$$z_t = f(x_t) = x_i^{\mathrm{T}}Ax_t. \tag{3.5}$$

Using Theorem 1, the optimal quadratic estimator (3.2) can be explicitly calculated in terms of the state estimate $\hat{x}_t$ and error covariance P_t,

$$\hat{z}_t = \mathrm{tr}(AP_t) + \hat{x}_t^{\mathrm{T}}A\hat{x}_t. \tag{3.6}$$

In parallel to the optimal estimate we propose a suboptimal estimate $\tilde{z}$ of the QF which depends only on the Kalman estimate $\hat{x}_t$ and does not require its error covariance P_t in contrast to the optimal one (3.6). The suboptimal estimate is obtained by direct calculation of the QF at the point $x_t = \hat{x}_t$ such as,

$$\tilde{z}_t = f(\hat{x}_t) = \hat{x}_t^{\mathrm{T}} A \hat{x}_t. \tag{3.7}$$

Let compare estimation accuracy of the optimal (3.6) and suboptimal (3.7) estimates. The following result completely defines the MSEs,

$$P_{z,t}^{\mathrm{opt}} = \mathbf{E}[(z_t - \hat{z}_t)^2],$$
$$P_{z,t}^{\mathrm{sub}} = \mathbf{E}[(z_t - \tilde{z}_t)^2].$$

Theorem 2. *The mean square errors $P_{z,t}^{\mathrm{opt}}$ and $P_{z,t}^{\mathrm{sub}}$ are given by*

$$P_{z,t}^{\mathrm{opt}} = 4\mathrm{tr}(AP_tAC_t) - 2\mathrm{tr}(AP_tAP_t) + 4\mu_t^{\mathrm{T}} AP_tA\mu_t, \tag{3.8}$$

and

$$P_{z,t}^{\mathrm{sub}} = 4\mathrm{tr}(AP_tAC_t) - 2\mathrm{tr}(AP_tAP_t) + \mathrm{tr}^2(AP_t) + 4\mu_t^{\mathrm{T}} AP_tA\mu_t \tag{3.9}$$

respectively. Here the unconditional mean $\mu_t = \mathbf{E}(x_t)$ and covariance $C_t = \mathrm{cov}(x_t, x_t)$ of the state vector x_t are determined by the equations of the method of moments [6],

$$\begin{aligned} \dot{\mu}_t &= F_t\mu_t, \quad t \ge 0, \quad \mu_0 = \bar{x}_0, \\ \dot{C}_t &= F_tC_t + C_tF_t^{\mathrm{T}} + G_tQ_tG_t^{\mathrm{T}}, \quad C_0 = P_0. \end{aligned} \tag{3.10}$$

The derivation of the MSEs (3.8) and (3.9) is given in Appendix.

Note that the difference between $P_{z,t}^{\mathrm{opt}}$ and $P_{z,t}^{\mathrm{sub}}$ is equal $P_{z,t}^{\mathrm{sub}} - P_{z,t}^{\mathrm{opt}} = \mathrm{tr}^2(AP_t)$.

Thus, equations (3.8)–(3.10) completely define the true MSEs of the optimal and suboptimal quadratic estimators, respectively.

3.3. Theoretical and Practical Examples of Application of Quadratic Estimators

Example 6 (theoretical example—estimation of power of scalar signal). Let x_t be a scalar random signal measured in additive white noise. Then the system model is

$$\dot{x}_t = ax_t + v_t, \quad t \ge 0, \quad a < 0,$$
$$y_t = x_t + w_t,$$

where v_t and w_t are the uncorrelated white Gaussian noises with intensities q and r, respectively, and $x_0 \sim \mathbb{N}(m_0, \sigma_0^2)$.

The KBF equations (2.1) give the following

$$\dot{\hat{x}}_t = a\hat{x}_t + P_t(y_t - \hat{x}_t)/r, \quad \hat{x}_0 = m_0,$$
$$\dot{P}_t = 2aP_t - P_t^2/r + q, \quad P_0 = \sigma_0^2.$$

Let consider power of the signal x_t, which is proportional to its square. Then $z_t = f(x_t) = x_t^2$

Using (3.6) and (3.7) we obtain the optimal and suboptimal estimates of power of the signal, respectively,

$$\hat{z}_t = \hat{x}_t^2 + P_t, \quad \tilde{z}_t = f(\hat{x}_t) = \hat{x}_t^2.$$

Compare accuracy of the estimates. Using Theorem 2, we obtain the true MSEs,

$$P_{z,t}^{\mathrm{opt}} = 4P_tC_t - 2P_t^2 + 4\mu_t^2 P_t,$$
$$P_{z,t}^{\mathrm{sub}} = 4P_tC_t - P_t^2 + 4\mu_t^2 P_t,$$

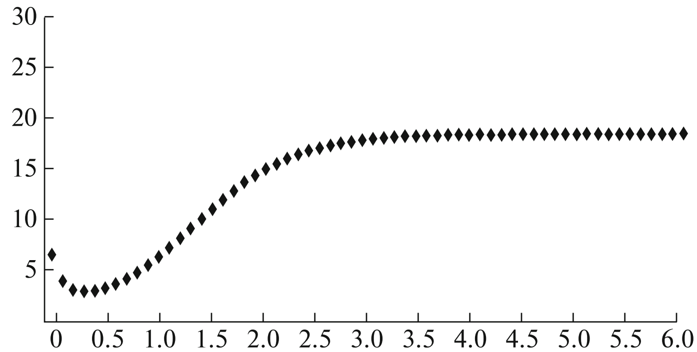

Fig. 5. Relative error Δ_t, % between optimal and suboptimal MSEs for quadratic estimators.

where the mean μ_t and covariance C_t of the signal x_t are determined by (3.10) as follows:

$$\dot{\mu}_t = a\mu_t, \quad \mu_0 = m_0,$$
$$\dot{C}_t = 2aC_t + q, \quad C_0 = \sigma_0^2.$$

Figure 5 shows the relative error $\Delta_t = |(P_{z,t}^{\text{opt}} - P_{z,t}^{\text{sub}})/P_{z,t}^{\text{opt}}| \cdot 100\%$ for the values $a = -1$, $q = 0.5$, $m_\theta = 0$, $\sigma_\theta^2 = 4$, $r = 0.1$ From Fig. 5 we observe that the relative error Δ_t varies from 3 to 6% within the time zone $t \in [0.1; 1.1]$, and then it increases. In steady-state regime $t > 4$ the relative error reaches the value $\Delta_\infty = 20.4\%$ and at the same time zone the absolute values of the MSEs are equal $P_\infty^{\text{opt}} = 0.1029$ and $P_\infty^{\text{sub}} = 0.1239$. Thus the numerical results show that the suboptimal estimate $\tilde{z}_t = \hat{x}_t^2$ may be seriously worse than the optimal one $\hat{z}_t = \hat{x}_t^2 + P_t$.

Example 7 (practical example—wind tunnel system). Here an experimental analysis of the quadratic estimators is considered on example of a total kinetic energy of the high-speed closed-air unit wind tunnel system [36]. The state vector $x_t \in \mathbb{R}^3$ consists of the state variables x_{1t}, x_{2t} and x_{3t}, representing derivatives from a chosen equilibrium point of the following quantities: x_1—Mach number, x_2—actuator position guyde vane angle in a driving fan, and x_3—actuator rate. Then the system model is given by

$$\dot{x}_t = \begin{bmatrix} 0.4032 & 0 & 0 \\ 0 & 0.1245 & 0.0701 \\ 0 & -2.5247 & -0.5488 \end{bmatrix} x_t + v_t,$$

where the initial conditions are $\bar{x}_0 = [3\ 28\ 10]^{\text{T}}$ and $P_0 = \text{diag}[1\ 1\ 0]$.

Two sensory measurement model is given

$$y_t = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} x_t + w_t.$$

The intensities of the white noises $v_t \in \mathbb{R}^3$ and $w_t \in \mathbb{R}^2$ are subjected to $Q = \text{diag}[0.01\ 0.01\ 0.01]$ and $R = \text{diag}[0.05\ 0.05]$, respectively.

The total kinetic energy of an actuator can be expressed as sum of the translational kinetic energy of the center of mass, $E^{tr} = mv_t^2/2$ and the rotational kinetic energy about the center of mass, $E^r = \mathbf{I}\omega_t^2/2$ where $\mathbf{I}$ is rotational inertia, $\omega_t = \dot{x}_{2t}$ is angular velocity, m is mass and $v_t = x_{3t}$ is linear velocity. Then the energy can be expressed in the following QF,

$$z_t = E^r + E^{tr} = \frac{1}{2}\mathbf{I}\dot{x}_{2,t}^2 + \frac{1}{2}mx_{3,t}^2 = X_t^{\text{T}} A X_t,$$
$$X_t = [x_{1,t}\ \ x_{2,t}\ \ x_{3,t}\ \ \dot{x}_{2,t}]^{\text{T}},$$
$$A = \text{diag}[0\ \ 0\ \ m/2\ \ \mathbf{I}/2],$$

where $X_t \in \mathbb{R}^4$ is the extended state vector, and $\mathbf{I} = 0.136$ kgm^2, $m = 7.39$ kg.

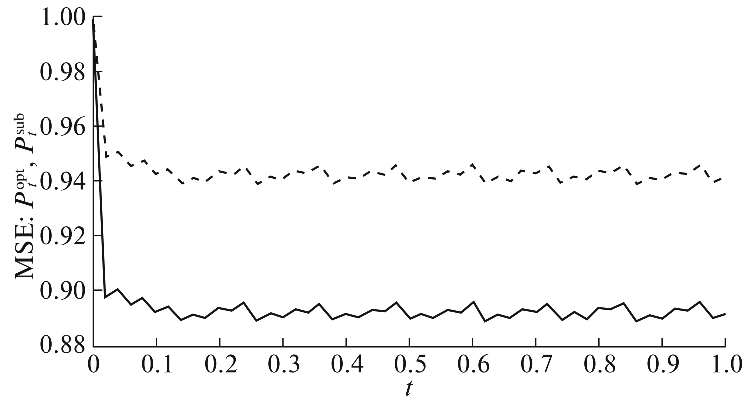

Fig. 6. Comparison of MSEs for total kinetic energy z_t with using optimal and suboptimal quadratic estimators: P_t^{opt} (solid curve), P_t^{sub} (dotted curve).

Using Theorem 1 the optimal and suboptimal quadratic estimators take the form

$$\hat{z}_t = \operatorname{tr}(AP_t) + \hat{X}_t^{\mathrm{T}} A\hat{X}_t, \;\; \tilde{z}_t = \hat{X}_t^{\mathrm{T}} A\hat{X}_t,$$

where the estimate of the state $\hat{X}_t \in \mathbb{R}^4$ and error covariance $P_t \in \mathbb{R}^{4\times4}$ are determined by (2.1).

Our point of interest is behavior of the MSEs, $P_{z,t}^{\text{opt}} = \mathbf{E}[(z_t - \hat{z}_t)^2]$ and $P_{z,t}^{\text{sub}} = \mathbf{E}[(z_t - \tilde{z}_t)^2]$ which can be calculated by using Theorem 2. We observe in Figure 6 that the optimal estimator has the best performance in contrast to the suboptimal one, i.e., $P_{z,t}^{\text{opt}} < P_{z,t}^{\text{sub}}$. The relative error Δ_t varies from 6.2 to 10% within the initial zone, $t \in [0.02; 0.07]$ of operation of the system, and then it decreases. In time zone, $t >$ 0.07 the values of the MSEs and relative error are equal $P_{z,t}^{\text{opt}} = 0.89$, $P_{z,t}^{\text{sub}} = 0.94$, and $\Delta_t = 5.6\%$, respectively.

As a result, we confirm that the proposed optimal quadratic estimator is more suitable for data processing in practice.

CONCLUSIONS

In some application problems, nonlinear function of state variables contains useful information of the target systems for control. In order to estimate an arbitrary NFS, an optimal two-stage MMSE estimation algorithm is proposed. At the first stage, a preliminary state estimate from the standard KBF method is calculated. And next the computed estimate is used at the second stage for the MMSE estimation of an NFS.

Particular attention is given for a polynomial functions of a state. In this case, it is possible to derive a closed-form polynomial estimator, which depends only on a parameters of the KBF. Interpretation of a quadratic functional as power or energy is considered in Examples 6 and 7.

In a view of importance of an NFS for practice, the proposed estimation algorithms are illustrated on theoretical and numerical examples for a real NFS. The examples show that the optimal MMSE estimator yields reasonably good estimation accuracy.

Using the MMSE method, an optimal two-stage nonlinear estimator is proposed. We establish that the polynomial estimators (quadratic, cubic and quartic) can be represented a compact closed-forms which depend only on the KBF characteristics (Theorem 1). An important class of quadratic functional is comprehensively investigated, including derivation of a matrix equations for a true MSE (Theorem 2). Performance of the proposed estimators for real NFS illustrates their theoretical and practical usefulness.

APPENDIX

Proof of Theorem 1. The derivation of the polynomial estimators (3.2) is based on the Lemma 1.

Lemma 1. *Let* $x \in \mathbb{R}^n$ *be a Gaussian random vector,* $x \sim \mathbb{N}(\mu, S)$ *and* $A, B \in \mathbb{R}^{n\times n}$ *be an arbitrary matrices. Then it holds that*

$$\begin{gathered}\mathbf{E}(Ax) = A\mu;\\ \mathbf{E}(x^{\mathrm{T}}Ax) = \mathrm{tr}(AS) + \mu^{\mathrm{T}}A\mu, \quad A = A^{\mathrm{T}};\\ \mathbf{E}(Axx^{\mathrm{T}}Bx) = 2ASB\mu + A\mu\mu^{\mathrm{T}}B\mu + A\mu\mathrm{tr}(BS);\\ E(x^TAxx^TBx) = 2tr(ASBS) + 4\mu^TASB\mu\\ + [\mathrm{tr}(AS) + \mu^TA\mu][\mathrm{tr}(BS) + \mu^TB\mu], \quad A = A^{\mathrm{T}}, \quad B = B^{\mathrm{T}}.\end{gathered} \tag{A.1}$$

The derivation of the formulas (A.1) is based on their scalar versions given in [37, 38], and standard transformations on random vectors.

This completes the proof of Lemma 1.

Next, replacing in (A.1) an unconditional expectations and covariance by their conditional versions, for example, $\mu \to \mathbf{E}(x|y_0^t) = \hat{x}$, $S \to \mathrm{cov}(\hat{x}, \hat{x}|y_0^t) = P$ we obtain (3.2).

This completes the proof of Theorem 1.

Proof of Theorem 2. The derivation of the MSEs is based on the Lemma 2.

Lemma 2. *Let* $X \in \mathbb{R}^{3n}$ *be a composite multivariate Gaussian vector,* $X^{\mathrm{T}} = [U^{\mathrm{T}}\ V^{\mathrm{T}}\ W^{\mathrm{T}}]$:

$$\begin{gathered}X \sim \mathbb{N}(\mu, S), \quad U, V, W \in \mathbb{R}^n,\\ \mu = \begin{bmatrix}\mu_u\\ \mu_v\\ \mu_w\end{bmatrix}, \quad S = \begin{bmatrix}S_{uu} & S_{uv} & S_{uw}\\ S_{vu} & S_{vv} & S_{vw}\\ S_{wu} & S_{wv} & S_{ww}\end{bmatrix}.\end{gathered} \tag{A.2}$$

Then the third- and fourth-order vector moments of the composite random vector are given by

$$\begin{gathered}\mathbf{E}(U^{\mathrm{T}}VW) = \mu_u^{\mathrm{T}}\mu_v\mu_w + \mathrm{tr}(S_{uv})\mu_w^{\mathrm{T}} + \mu_v^{\mathrm{T}}S_{uw} + \mu_u^{\mathrm{T}}S_{vw};\\ \mathbf{E}(U^{\mathrm{T}}UV^{\mathrm{T}}V) = \mu_u^{\mathrm{T}}\mu_u\mu_v^{\mathrm{T}}\mu_v + 2\mathrm{tr}(S_{uv}S_{vu}) + \mathrm{tr}(S_{uu})\mathrm{tr}(S_{vv})\\ + \mathrm{tr}(S_{uu})\mu_v^{\mathrm{T}}\mu_v + \mathrm{tr}(S_{vv})\mu_u^{\mathrm{T}}\mu_u + 4\mu_u^{\mathrm{T}}S_{uv}\mu_v;\\ \mathbf{E}(U^{\mathrm{T}}VV^{\mathrm{T}}U) = \mu_u^{\mathrm{T}}\mu_v\mu_v^{\mathrm{T}}\mu_u + \mathrm{tr}(S_{uu}S_{vv}) + \mathrm{tr}(S_{uv})\mathrm{tr}(S_{vu})\\ + \mathrm{tr}(S_{uv}^2) + \mu_v^{\mathrm{T}}S_{uu}\mu_v + \mu_u^{\mathrm{T}}S_{vv}\mu_u + \mu_u^{\mathrm{T}}S_{uv}\mu_u + \mu_u^{\mathrm{T}}S_{vu}\mu_v + 2\mathrm{tr}(S_{uv})\mu_u^{\mathrm{T}}\mu_v;\\ \mathbf{E}(U^{\mathrm{T}}VW^{\mathrm{T}}U) = \mu_u^{\mathrm{T}}\mu_v\mu_w^{\mathrm{T}}\mu_u + \mathrm{tr}(S_{uv})\mathrm{tr}(S_{uw}) + \mathrm{tr}(S_{uu}S_{wv})\\ + \mathrm{tr}(S_{uw}S_{uv}) + \mathrm{tr}(S_{uv})\mu_u^{\mathrm{T}}\mu_w + \mathrm{tr}(S_{uw})\mu_u^{\mathrm{T}}\mu_v\\ + \mu_v^{\mathrm{T}}S_{uu}\mu_w + \mu_v^{\mathrm{T}}S_{uw}\mu_u + \mu_u^{\mathrm{T}}S_{vu}\mu_w + \mu_u^{\mathrm{T}}S_{vw}\mu_u.\end{gathered} \tag{A.3}$$

The derivation of the vector formulas (A.3) is based on their scalar versions, and standard matrix manipulations,

$$\begin{gathered}\mathbf{E}\left(x_ix_jx_k\right) = \mu_i\mu_j\mu_k + \mu_iS_{jk} + \mu_jS_{ik} + \mu_kS_{ij};\\ \mathbf{E}\left(x_ix_jx_kx_\ell\right) = \mu_i\mu_j\mu_k\mu_\ell + S_{ij}S_{k\ell} + S_{ik}S_{\ell j} + S_{i\ell}S_{jk} + \mu_i\mu_jS_{k\ell} + \mu_i\mu_kS_{j\ell} + \mu_i\mu_\ell S_{jk}\\ + \mu_j\mu_kS_{i\ell} + \mu_j\mu_\ell S_{ik} + \mu_k\mu_\ell S_{ij},\end{gathered} \tag{A.4}$$

where $\mu_h = \mathbf{E}(x_h)$, $S_{pq} = \mathrm{cov}(x_p, x_q)$.

This completes the proof of Lemma 2.

Further, we derive the formula (3.8). Using (3.5) and (3.6), the error can be written as

$$\begin{gathered}e_z = z - \hat{z} = x^{\mathrm{T}}Ax - \mathrm{tr}\left(AP\right) - \hat{x}^{\mathrm{T}}A\hat{x} = \left(e + \hat{x}\right)^{\mathrm{T}}A\left(e + \hat{x}\right) - \hat{x}^{\mathrm{T}}A\hat{x} - \mathrm{tr}\left(AP\right)\\ = e^{\mathrm{T}}Ae + 2e^{\mathrm{T}}A\hat{x} - \mathrm{tr}\left(AP\right), \quad e = x - \hat{x}, \quad \hat{x}^{\mathrm{T}}Ae = e^{\mathrm{T}}A\hat{x}.\end{gathered} \tag{A.5}$$

Next, using the unbiased and orthogonality properties of the Kalman estimate $\mathbf{E}(e) = \mathbf{E}(e\hat{x}^{\mathrm{T}}) = 0$, we obtain

$$\begin{aligned} P_z^{\text{opt}} &= \mathbf{E}(e^{\mathrm{T}}Aee^{\mathrm{T}}Ae) + 2\mathbf{E}(e^{\mathrm{T}}A\hat{x}e^{\mathrm{T}}Ae) \\ &- \mathrm{tr}(AP)\mathbf{E}(e^{\mathrm{T}}Ae) + 2\mathbf{E}(e^{\mathrm{T}}Aee^{\mathrm{T}}A\hat{x}) \\ &+ 4\mathbf{E}(e^{\mathrm{T}}A\hat{x}e^{\mathrm{T}}A\hat{x}) - 2\mathrm{tr}(AP)\mathbf{E}(e^{\mathrm{T}}A\hat{x}) \\ &- \mathrm{tr}(AP)\mathbf{E}(e^{\mathrm{T}}Ae) - 2\mathrm{tr}(AP)\mathbf{E}(e^{\mathrm{T}}A\hat{x}) + \mathrm{tr}^2(AP). \end{aligned} \tag{A.6}$$

Using Lemma 2 we can calculate high-order moments in (A.6). We have

$$\begin{gathered} \mathbf{E}(e^{\mathrm{T}}Ae) = \mathrm{tr}(AP),\ \ \mathbf{E}(e^{\mathrm{T}}A\hat{x}) = 0,\quad \mathbf{E}(e^{\mathrm{T}}Aee^{\mathrm{T}}Ae) = \mathrm{tr}^2(AP) + 2\mathrm{tr}(APAP), \\ \mathbf{E}(e^{\mathrm{T}}A\hat{x}e^{\mathrm{T}}Ae) = \mathbf{E}(e^{\mathrm{T}}Aee^{\mathrm{T}}A\hat{x}) = 0,\quad \mathbf{E}(e^{\mathrm{T}}A\hat{x}e^{\mathrm{T}}A\hat{x}) = \mathbf{E}(e^{\mathrm{T}}A\hat{x}\hat{x}^{\mathrm{T}}Ae) = \mathrm{tr}(PAC_{\hat{x}\hat{x}}A) + \mu^{\mathrm{T}}AP\mu, \end{gathered} \tag{A.7}$$

where

$$\begin{gathered} \mu = \mathbf{E}(x) = \mathbf{E}(\hat{x}),\quad \mathrm{cov}(Ae, Ae) = APA^{\mathrm{T}}, \\ C_{\hat{x}\hat{x}} = C - P,\ \ P = \mathrm{cov}(e,e),\quad C = \mathrm{cov}(x,x). \end{gathered} \tag{A.8}$$

Substituting (A.7) in (A.6), and after some manipulations, we get the optimal MSE (3.8).

In the case of the suboptimal estimate $\tilde{z}$, the derivation of the MSE (3.9) is similar.

This completes the proof of Theorem 2.

REFERENCES

1. N. Davari and A. Gholami, "An asynchronous adaptive direct kalman filter algorithm to improve underwater navigation system performance," IEEE Sens. J. **17**, 1061–1068 (2017).
2. T. Rajaram, J. M. Reddy, and Y. Xu, "Kalman filter based detection and mitigation of subsynchronous resonance with SSSC," IEEE Trans. Power Syst. **32**, 1400–1409 (2017).
3. X. Deng and Z. Zhang, "Automatic multihorizons recognition for seismic data based on Kalman filter tracker," IEEE Geosci. Remote Sensing Lett. **14**, 319–323 (2017).
4. M. S. Grewal, A. P. Andrews, and C. G. Bartone, *Global Navigation Satellite Systems, Inertial Navigation, and Integration* (Wiley, NJ, 2013).
5. D. Simon, *Optimal State Estimation* (Wiley, NJ, 2006).
6. Y. Bar-Shalom, X. R. Li, and T. Kirubarajan, *Estimation with Applications to Tracking and Navigation* (Wiley, New York, 2001).
7. V. I. Arnold, *Mathematical Methods of Classical Mechanics* (Springer, New York, 1989).
8. T. T. Cai and M. G. Low, "Optimal adaptive estimation of a quadratic functional," Ann. Stat. **34**, 2298–2325 (2006).
9. J. Robins, L. Li, E. Tchetgen, and A. Vaart, "Higher order infuence functions and minimax estimation of nonlinear functionals," Prob. Stat. **2**, 335–421 (2008).
10. J. Jiao, K. Venkat, Y. Han, and T. Weissman, "Minimax estimation of functionals of discrete distributions," IEEE Trans. Inform. Theory **61**, 2835–2885 (2015).
11. J. Jiao, K. Venkat, Y. Han, and T. Weissman, "Maximum likelihood estimation of functionals of discrete distributions," IEEE Trans. Inform. Theory **63**, 6774–6798 (2017).
12. Y. Amemiya and W. A. Fuller, "Estimation for the nonlinear functional relationship," Ann. Stat. **16**, 147–160 (1988).
13. D. L. Donoho and M. Nussbaum, "Minimax quadratic estimation of a quadratic functional," J. Complexity **6**, 290–323 (1990).
14. D. S. Grebenkov, "Optimal and suboptimal quadratic forms for noncentered gaussian processes," Phys. Rev. **E88**, 032140 (2013).
15. B. Laurent and P. Massart, "Adaptive estimation of a quadratic functional by model selection," Ann. Stat. **28**, 1302–1338 (2000).
16. I. G. Vladimirov and I. R. Petersen, "Directly coupled observers for quantum harmonic oscillators with discounted mean square cost functionals and penalized back-action," in *Proceedings of the IEEE Conference on Norbert Wiener in the 21st Century, Melbourne, Australia, 2016,* pp. 78–83.
17. K. Sricharan, R. Raich, and A. O. Hero, "Estimation of nonlinear functionals of densities with confidence," IEEE Trans. Inform. Theory **58**, 4135–4159 (2012).

18. A. Wisler, V. Berisha, A. Spanias, and A. O. Hero, "Direct estimation of density functionals using a polynomial basis," IEEE Trans. Signal Process. **66**, 558–588 (2018).
19. M. Taniguchi, "On estimation of parameters of gaussian stationary processes," J. Appl. Prob. **16**, 575–591 (1979).
20. C. Zhao-Guo and E. J. Hanman, "The distribution of periodogram ordinates," J. Time Ser. Anal. **1**, 73–82 (1980).
21. D. Janas and R. Sachs, "Consistency for non-linear functions of the periodogram of tapered data," J. Time Ser. Anal. **16**, 585–606 (1995).
22. G. Fay, E. Moulines, and P. Soulier, "Nonlinear functionals of the periodogram," J. Time Ser. Anal. **23**, 523–553 (2002).
23. C. Noviello, G. Fornaro, P. Braca, and M. Martorella, "Fast and accurate ISAR focusing based on a doppler parameter estimation algorithm," IEEE Geosci. Remote Sens. Lett. **14**, 349–353 (2017).
24. Y. Wu and P. Yang, "Minimax rates of entropy estimation on large alphabets via best polynomial approximation," IEEE Trans. Inform. Theory **62**, 3702–3720 (2016).
25. Y. Wu and P. Yang, "Optimal entropy estimation on large alphabets via best polynomial approximation," in *Proceedings of the IEEE International Symposium on Information Theory, Hong Kong, 2015*, pp. 824–828.
26. S. O. Haykin, *Adaptive Filtering* (Prentice Hall, NJ, 2013).
27. T. K. Moon and W. C. Stirling, *Mathematical Methods and Algorithms for Signal Processing* (Prentice Hall, NJ, 2000).
28. A. Coluccia, "On the expected value and higher-order moments of the euclidean norm for elliptical normal variates," IEEE Commun. Lett. **17**, 2364–2367 (2013).
29. S. Nadarajah and S. Kotz, "Exact distribution of the max/Min of two gaussian random variables," IEEE Trans. Very Large Scale Integr. Syst. **16**, 210–212 (2008).
30. V. S. Pugachev and I. N. Sinitsyn, *Stochastic Differential Systems. Analysis and Filtering,* 2nd ed. (Nauka, Moscow, 1990) [in Russian].
31. V. S. Pugachev, "Assessment of the state and parameters of continuous nonlinear systems," Avtom. Telemekh., No. 6, 63–79 (1979).
32. E. A. Rudenko, "Optimal structure of continuous nonlinear reduced-order Pugachev filter," J. Comput. Syst. Sci. Int. **52**, 866 (2013).
33. S. J. Julier and J. K. Uhlmann, "Unscented filtering and nonlinear estimation," Proc. IEEE **92**, 401–422 (2004).
34. K. Ito and K. Xiong, "Gaussian filters for nonlinear filtering problems," IEEE Trans. Autom. Control **45**, 910–927 (2000).
35. A. Doucet, N. D. Freitas, and N. Gordon, *Sequential Monte Carlo Methods in Practice* (Springer, London, 2001).
36. E. S. Armstrong and J. S. Tripp, "An application of multivariable design techniques to the control of the national transonic facility," NASA Technical Paper, No. 1887 (NASA, Washington, DC, 1981), pp. 1–36.
37. R. Kan, "From moments of sum to moments of product," J. Multivar. Anal. **99**, 542–554 (2008).
38. B. Holmquist, "Expectations of products of quadratic forms in normal variables," Stoch. Anal. Appl. **14**, 149–164 (1996).