

# Equivalent Linearization Kalman Filter with Application to Cubic Sensor Problems

Tohru Katayama

Guest Professor, Faculty of Science and Engineering, Ritsumeikan University  
 1-1-1 Noji-Higashi, Kusatsu, Shiga 525-8577, Japan  
*e-mail: tohru\_katayama(at)nifty.com*

**Abstract**— We revisit the equivalent linearization technique to clarify a relationship between the extended Kalman filter (EKF) and equivalent linearization Kalman filter (EqKF). By deriving the equivalent gain for a static nonlinearity, we show that the equivalent linearization for the EqKF is a global method, though the first-order linearization for the EKF is a local one. Then, we consider discrete-time cubic sensor problems and analyze the Kalman gains and filtered covariances of respective filters, showing that the EqKF is quite close to the Gaussian filter (GF). Moreover, numerical results are included to compare the performances of the EqKF, EKF and GF.

## I. INTRODUCTION

The equivalent linearization, or statistical linearization, was developed by Booton [1], Kazakov [2], [3], Sawaragi *et al.* [4], [5], Caughey [6] and others over 50 years ago to analyze various nonlinear stochastic dynamic systems [7], [8]. The method has also been used for deriving a nonlinear filter called the equivalent linearization Kalman filter (EqKF) as an alternative to the extended Kalman filter (EKF) [10]; see e.g. Katayama [11], Sunahara [12] and Gelb [13]. Moreover, the EqKF has been applied to trajectory estimation by Austin and Leondes [14], and to Gaussian sum filtering for a GPS signal processing by Sugimoto *et al.* [15].

In this paper, we revisit the equivalent linearization technique for static nonlinearities to clarify the difference and similarity between the EKF and EqKF in the framework of Gaussian filter [16]. We show that, while the first-order linearization is a local method, the equivalent linearization is a global one. Then, we consider discrete-time cubic sensor problems [9] and analyze the Kalman gains and filtered covariances of EqKF, EKF and GF, showing that the EqKF algorithm is close to the GF algorithm, but is quite different from the EKF algorithm. Simulation studies show that the EqKF gives good performance like the GF, though the EKF often fails to track the state.

## II. NONLINEAR FILTERING

Consider a discrete-time nonlinear dynamical system

$$x_{t+1} = f_t(x_t) + w_t \quad (1a)$$

$$y_t = h_t(x_t) + v_t \quad (1b)$$

where  $x_t \in \mathbb{R}^n$  is the state,  $y_t \in \mathbb{R}^p$  is the output observation, and  $f_t : \mathbb{R}^n \rightarrow \mathbb{R}^n$ ,  $h_t : \mathbb{R}^n \rightarrow \mathbb{R}^p$  are nonlinear functions. Also,  $w_t$  and  $v_t$  are mutually independent Gaussian white

noises with mean zeros and covariance matrices  $Q_t$  and  $R_t$ , respectively. Let  $Y^t = \{y_0, y_1, \dots, y_t\}$  be the collection of observations up to time  $t$ , and define the conditional mean estimates as

$$\hat{x}_{t+l/t} = E\{x_{t+l} | Y^t\}, \quad l = 0, 1$$

and the conditional covariance matrices as

$$P_{t+l/t} = \text{cov}(x_{t+l} | Y^t), \quad l = 0, 1$$

Although many nonlinear filtering algorithms have been developed in the past [13], [18], a general procedure that propagates the conditional means and covariance matrices, under the assumption that the conditional probability density functions (pdfs) are Gaussian, is described by the following algorithm [16], [19].

*Algorithm 1:* (Gaussian filter algorithm)

- 1) Set initial values  $\hat{x}_{0/-1} = \bar{x}_0$ ,  $P_{0/-1} = P_0$  and  $t = 0$ .
- 2) The measurement update

$$\hat{y}_{t/t-1} = E\{y_t | Y^{t-1}\} \quad (2)$$

$$U_{t/t-1} = \text{cov}(x_t, y_t | Y^{t-1}) \quad (3)$$

$$V_{t/t-1} = \text{cov}(y_t | Y^{t-1}) \quad (4)$$

$$K_t = U_{t/t-1} V_{t/t-1}^{-1} \quad (5)$$

$$\hat{x}_{t/t} = \hat{x}_{t/t-1} + K_t [y_t - \hat{y}_{t/t-1}] \quad (6)$$

$$P_{t/t} = P_{t/t-1} - U_{t/t-1} V_{t/t-1}^{-1} U_{t/t-1}^T \quad (7)$$

- 3) The time update

$$\hat{x}_{t+1/t} = E\{x_{t+1} | Y^t\} \quad (8)$$

$$P_{t+1/t} = \text{cov}(x_{t+1} | Y^t) \quad (9)$$

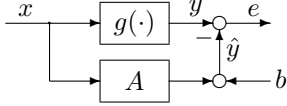
- 4) Set  $t := t + 1$ , and go back to Step 2.  $\square$

If we find a new approximation method of evaluating the conditional expectations in Algorithm 1, then a new nonlinear filtering algorithm can be derived. In fact, the EKF is derived by linearizing nonlinearities  $f_t(x_t)$  and  $h_t(x_t)$  at the filtered and predicted estimates, respectively, before taking the conditional expectations [10]. The EqKF is derived by replacing two nonlinearities by the respective equivalent linear systems under the Gaussian assumption [10], [13]. Also, to evaluate the conditional means and covariance matrices, Monte Carlo samplings are employed in the ensemble Kalman filter (EnKF) [20], and the deterministic sampling

based on  $\sigma$ -points is used to develop the UKF [17]. The UKF algorithm is analyzed from the point of view of statistical linearization [21]. Further, numerical issues of Gaussian filters are discussed in [16], [19], [22].

### III. EQUIVALENT LINEARIZATION OF A STATIC NONLINEARITY

In this section, we briefly review a statistical equivalent linearization of a static nonlinearity. Let  $x \in \mathbb{R}^n$  be a Gaussian random vector with mean  $m_x \in \mathbb{R}^n$  and covariance matrix  $\Sigma_{xx} \in \mathbb{R}^{n \times n}$ . Let a nonlinear function be  $y = g(x)$  with  $g : \mathbb{R}^n \rightarrow \mathbb{R}^p$ . We derive a linear unbiased minimum variance estimate  $\hat{y}$  of  $y$  as shown in the diagram below:



Let  $\hat{y} = Ax + b$ , where  $A \in \mathbb{R}^{p \times n}$ ,  $b \in \mathbb{R}^p$ . Then, the unbiasedness of  $\hat{y}$  implies that  $E\{y - Ax - b\} = 0$ , so that  $m_y = Am_x + b$ , where  $m_y = E\{y\} = E\{g(x)\}$ . Thus, from  $b = m_y - Am_x$ , the mean square error is given by

$$J = E\{(y - Ax - b)^T(y - Ax - b)\} \\ = \text{trace}\left(\Sigma_{yy} - A\Sigma_{yx} - \Sigma_{yx}A^T + A\Sigma_{xx}A^T\right)$$

Suppose that  $\Sigma_{xx}$  is positive definite. Then, the minimum value of  $J$  is attained by  $A^* = \Sigma_{yx}\Sigma_{xx}^{-1}$ , where  $\Sigma_{yx} = E\{g(x)[x - m_x]^T\}$ , so that  $b^* = m_y - \Sigma_{yx}\Sigma_{xx}^{-1}m_x$ . Hence the optimum linear approximation is obtained as

$$\hat{y} = m_y + G^e(x - m_x)$$

where  $G^e = \Sigma_{yx}\Sigma_{xx}^{-1} \in \mathbb{R}^{p \times n}$  is called the equivalent gain matrix in the literature [7]. The covariance matrix of the output  $y$  is therefore approximated as

$$\text{cov}(y) \simeq \text{cov}(\hat{y}) = G^e \Sigma_{xx} (G^e)^T$$

Moreover, since  $J(A^*) \geq 0$ , the actual output covariance matrix is bounded below, i.e.  $\text{cov}(y) \geq G^e \Sigma_{xx} (G^e)^T$ .

*Proposition 1:* Suppose that  $x \in \mathbb{R}^n$  is a Gaussian random vector with  $N(m, P)$ . Consider a nonlinear transformation  $y = g(x)$  with  $g : \mathbb{R}^n \rightarrow \mathbb{R}^p$  differentiable, and assume that the following conditions are satisfied:

$$E\{|g_i(x)|\} < \infty, \quad E\{|\partial g_i(x)/\partial x_j|\} < \infty \quad (10)$$

where  $i = 1, \dots, p$ ,  $j = 1, \dots, n$ . Then, the equivalent gain matrix  $G^e \in \mathbb{R}^{p \times n}$  of  $g(x)$  is expressed as

$$G^e = E\left\{\frac{\partial}{\partial x} g(x)\right\} = \frac{\partial}{\partial m} E\{g(x)\} \quad (11)$$

*Proof:* See Appendix.  $\square$

The formula (11) implies that the equivalent gain matrix can be obtained by either the conditional expectation of Jacobian matrix of the nonlinearity or the Jacobian of conditional expectation of it. The first result is found in [7], and the second one in [3].

*Proposition 2:* We have two linear approximations for  $y = g(x)$ , i.e. a local approximation:

$$g(x) \simeq g(m) + \left[\frac{\partial g(x)}{\partial x}\right]_{x=m} (x - m) \quad (12)$$

and a "global" approximation:

$$g(x) \simeq E\{g(x)\} + E\left\{\frac{\partial g(x)}{\partial x}\right\} (x - m) \quad (13a)$$

$$= E\{g(x)\} + \frac{\partial}{\partial m} E\{g(x)\} (x - m) \quad (13b)$$

It is well known that the EKF is derived by using a local approximation of (12), while the EqKF is derived by using a global approximation of (13).  $\square$

Once we computed the conditional expectation  $\hat{y}_{t/t-1} = E\{y_t | Y^{t-1}\}$ , it is convenient to compute the equivalent gain matrix as

$$H_t^e = \frac{\partial \hat{y}_{t/t-1}}{\partial \hat{x}_{t/t-1}} = \frac{\partial}{\partial \hat{x}_{t/t-1}} E\{h_t(x_t) | Y^{t-1}\} \quad (14)$$

where  $x_t \sim N(\hat{x}_{t/t-1}, P_{t/t-1})$ . Similarly, the equivalent gain matrix of  $f_t(x_t)$  is given by

$$F_t^e = \frac{\partial \hat{x}_{t+1/t}}{\partial \hat{x}_{t/t}} = \frac{\partial}{\partial \hat{x}_{t/t}} E\{f_t(x_t) | Y^t\}$$

where  $x_t \sim N(\hat{x}_{t/t}, P_{t/t})$ .

*Example 1:* Consider a cubic function  $y = x^3$ . Let  $x \sim N(m, P)$ . Then, the output mean  $m_y$ , cross-covariance  $\Sigma_{xy}$  and covariance  $\Sigma_{yy}$  are given by the second line (Exact) of Table I, in which the output statistics computed by the UT [17], the equivalent linearization (13) and the local approximation (12) are also included. We see that the coefficients of  $\Sigma_{yy}$  by UT with  $\kappa = 2$  lie between those of Exact and (13), so that  $\Sigma_{yy}^{\text{Exact}} \geq \Sigma_{yy}^{\text{UT}} \geq \Sigma_{yy}^{(13)} \geq \Sigma_{yy}^{(12)}$ .  $\square$

TABLE I  
OUTPUT STATISTICS FOR  $y = x^3$ .

	$m_y$	$\Sigma_{xy}$	$\Sigma_{yy}$
Exact	$m^3 + 3mP$	$3m^2P + 3P^2$	$(9m^4 + 36m^2P + 15P^2)P$
UT*	$m^3 + 3mP$	$3m^2P + 3P^2$	$(9m^4 + 36m^2P + 9P^2)P$
(13)	$m^3 + 3mP$	$3m^2P + 3P^2$	$(9m^4 + 18m^2P + 9P^2)P$
(12)	$m^3$	$3m^2P$	$9m^4P$

\* UT:=unscented transformation with  $\kappa = 2$  [17].

### IV. CUBIC SENSOR PROBLEM 1

Consider a scalar linear discrete-time system with a cubic observation equation

$$x_{t+1} = ax_t + bu_t + w_t \quad (15a)$$

$$y_t = \beta x_t^3 + v_t \quad (15b)$$

where  $a, b$  and  $\beta > 0$  are constants,  $u_t = \sin(2\pi t/50)$  is a probing input, and  $w_t$  and  $v_t$  are white Gaussian noises with means zero and variances  $q$  and  $r$ , respectively.

The above model of (15) with  $b = 0$  has been treated by Bucy [9]; developed is a grid method of solving recursive equations satisfied by conditional pdfs.

To apply Algorithm 1 to the present problem, we assume that  $p(x_t | Y^{t-1}) \sim N(\hat{x}_{t/t-1}, P_{t/t-1})$  and  $p(x_t | Y^t) \sim N(\hat{x}_{t/t}, P_{t/t})$ . Then, according to Table I, the GF for the cubic sensor problem is given by the following.

*Algorithm 2: (GF algorithm)*

1) The measurement update

$$\hat{y}_{t/t-1} = \beta(\hat{x}_{t/t-1}^3 + 3\hat{x}_{t/t-1}P_{t/t-1}) \quad (16)$$

$$U_{t/t-1} = 3\beta(\hat{x}_{t/t-1}^2 + P_{t/t-1})P_{t/t-1} \quad (17)$$

$$V_{t/t-1} = \beta^2[9\hat{x}_{t/t-1}^4 + 36\hat{x}_{t/t-1}^2P_{t/t-1} + 15P_{t/t-1}^2]P_{t/t-1} + r \quad (18)$$

$$K_t = U_{t/t-1}V_{t/t-1}^{-1} \quad (19)$$

$$\hat{x}_{t/t} = \hat{x}_{t/t-1} + K_t[y_t - \hat{y}_{t/t-1}] \quad (20)$$

$$P_{t/t} = P_{t/t-1} - U_{t/t-1}V_{t/t-1}^{-1}U_{t/t-1}^T \quad (21)$$

2) The time update

$$\hat{x}_{t+1/t} = a\hat{x}_{t/t} + bu_t \quad (22)$$

$$P_{t+1/t} = a^2P_{t/t} + q \quad (23)$$

3) The initial conditions  $\hat{x}_{0/-1} = \bar{x}_0, P_{0/-1} = P_0$ .  $\square$

We now derive the Kalman gains and filtered covariance equations for EqKF, GF and EKF to analyze these filters.

For the EqKF, from (14) and (16), the equivalent gain is given by  $H_t^e = 3\beta(\hat{x}_{t/t-1}^2 + P_{t/t-1})$ , so that

$$K_t^{EqKF} = \frac{3\beta(\hat{x}_{t/t-1}^2 + P_{t/t-1})P_{t/t-1}}{9\beta^2(\hat{x}_{t/t-1}^2 + P_{t/t-1})^2P_{t/t-1} + r} \quad (24)$$

$$P_{t/t} = \frac{rP_{t/t-1}}{9\beta^2(\hat{x}_{t/t-1}^2 + P_{t/t-1})^2P_{t/t-1} + r} \quad (25)$$

For the GF, we see from (17), (18) and (21) that the Kalman gain and filtered covariance are

$$K_t^{GF} = \frac{3\beta(\hat{x}_{t/t-1}^2 + P_{t/t-1})P_{t/t-1}}{9\beta^2(\hat{x}_{t/t-1}^2 + P_{t/t-1})^2P_{t/t-1} + r + Z_t} \quad (26)$$

$$P_{t/t} = \frac{(r + Z_t)P_{t/t-1}}{9\beta^2(\hat{x}_{t/t-1}^2 + P_{t/t-1})^2P_{t/t-1} + r + Z_t} \quad (27)$$

where  $Z_t := 6\beta^2(3\hat{x}_{t/t-1}^2 + P_{t/t-1})P_{t/t-1}^2$  is introduced to make it easy to compare the GF and EqKF.

For the EKF, it follows that  $\hat{y}_{t/t-1} = \beta\hat{x}_{t/t-1}^3$ , so that the Jacobian evaluated at the predicted estimate is  $3\beta\hat{x}_{t/t-1}^2$ . Hence, the Kalman gain and filtered covariance become

$$K_t^{EKF} = \frac{3\beta(\hat{x}_{t/t-1}^2)P_{t/t-1}}{9\beta^2(\hat{x}_{t/t-1}^2)^2P_{t/t-1} + r} \quad (28)$$

$$P_{t/t} = \frac{rP_{t/t-1}}{9\beta^2(\hat{x}_{t/t-1}^2)^2P_{t/t-1} + r} \quad (29)$$

*Proposition 3:* If  $P_{t/t-1}^{GF} = P_{t/t-1}^{EqKF}$  and if  $\hat{x}_{t/t-1}^{GF} = \hat{x}_{t/t-1}^{EqKF}$ , then we have

$$K_t^{GF} \leq K_t^{EqKF}, \quad P_{t/t}^{EqKF} \leq P_{t/t}^{GF} \quad (30)$$

*Proof:* The first result is clear from (24) and (26), and the second one is proved by using (25), (27) and  $Z_t \geq 0$ .  $\square$

It follows from (30) that, given the same information, the GF is more cautious than the EqKF, because the variance of observation noise in the GF is regarded as  $r + Z_t$  from the viewpoint of the EqKF.

Also, under the assumption that  $P_{t/t-1}^{EKF} = P_{t/t-1}^{EqKF}$  and  $\hat{x}_{t/t-1}^{EKF} = \hat{x}_{t/t-1}^{EqKF}$ , we have  $K_t^{EqKF} \leq K_t^{EKF}$ , if

$$r \leq \frac{\hat{x}_{t/t-1}^2}{\hat{x}_{t/t-1}^2 + P_{t/t-1}}(H_t^e)^2P_{t/t-1}$$

This implies that if  $\hat{x}_{t/t-1}^2$  is not very small, then  $K_t^{EKF}$  is more optimistic than  $K_t^{EqKF}$ . But, if  $\hat{x}_{t/t-1}^2 \simeq 0$ , then  $K_t^{EKF}$  reduces to zero, so that the EKF ceases to track the state [9]. Since the Jacobian of  $h_t$  at  $\hat{x}_{t/t-1}$  does not include the covariance  $P_{t/t-1}$ , we see that  $K_t^{EKF}$  is quite different from  $K_t^{EqKF}$  and  $K_t^{GF}$ ; see (24), (26) and (28).

It may be noted that there does not exist clear relation between the GF and EKF.

We show numerical results for the cubic sensor problem, comparing EKF, EqKF and GF algorithms. For simulations, we assume that  $a = 1, b = 0.01, \beta = 1/10, q = r = 1$  in (15), and  $x_0 = 0$ . The initial values are assumed to be  $\hat{x}_{0/-1} = 5$  and  $P_{0/-1} = 1$  for all three filters.

Figs. 1 and 2 show sample plots of the true state and observation process and the estimation results by the EKF and EqKF, respectively. In Fig. 2(left), there are time intervals (e.g.  $t = 125, \dots, 150; t = 170, \dots, 190$ ), where the state estimation is interrupted ( $\hat{x}_{t/t} \sim 10^{-9}$ ); whereas the estimation result by EqKF shows a smooth tracking as in Fig. 2(right). Note that tracking behavior of the GF (not shown here) is similar to that of the EqKF.

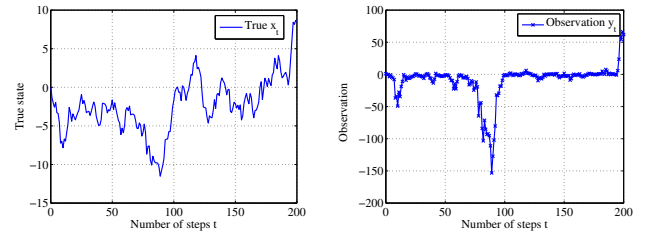


Fig. 1. A true state trajectory (left), and observations (right).

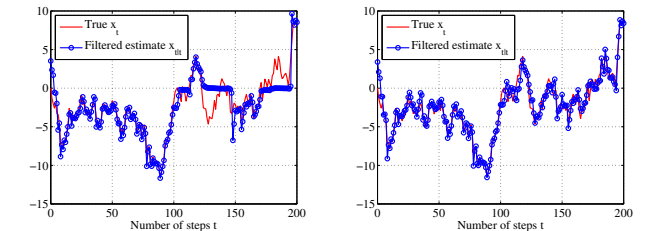


Fig. 2. Estimation by EKF (left) and by EqKF (right).

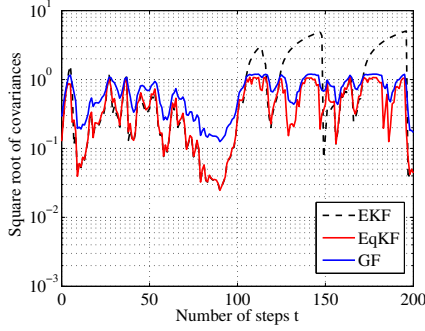


Fig. 3. Plot of  $\sqrt{P_{t/t}}$  by EKF (black), EqKF (red) and GF (blue).

Fig. 3 depicts a sample plot of the standard deviation  $\sqrt{P_{t/t}}$  of the EKF (dash-black), EqKF (red) and GF (blue). We observe an increase of the filtered covariance of the EKF in the intervals where the state estimation is interrupted; in fact, if  $\hat{x}_{t/t-1} \simeq 0$ , then from (28) and (29), we have  $K_t^{EKF} \simeq 0$  and  $P_{t+1/t} \simeq a^2 P_{t/t-1} + q$ .

However, we see that the two covariances of EKF and EqKF almost coincide in the intervals where good estimates (hence small covariances) are obtained. In fact, if  $P_{t/t-1}$  in the equivalent gain  $H_t^e$  is much smaller than  $\hat{x}_{t/t-1}^2$ , then the equivalent gain approaches to the Jacobian of EKF, so that the covariance equation of (25) reduces to (29). Moreover, we observe that the covariance of the EqKF of (25) is bounded above by the covariance of the GF of (27)<sup>1</sup>.

To compare the performance, we define the root mean square error (RMSE) as

$$E_N^{(j)} = \sqrt{\frac{1}{N} \sum_{t=0}^N (x_t^{(j)} - \hat{x}_{t/t}^{(j)})^2}, \quad j = 1, \dots, M$$

where the superscript denotes the  $j$ th sample. Table II displays the average RMSE

$$\bar{E} = \frac{1}{M} \sum_{j=1}^M E_N^{(j)}, \quad M = 50, \quad N = 200$$

where the numbers in parentheses denote standard deviation of  $E_N^{(j)}$ . We see from Table II that the performances of EqKF and GF are much better than that of the EKF. Also, note that if there is no probing input ( $b = 0$ ), then the performance of the EKF considerably degrades, whereas we do not see any degradation in the performances of the EqKF and GF.

TABLE II  
PERFORMANCE OF EKF, EQKF AND GF.

	EKF	EqKF	GF
$b = 0$	4.3275(6.5931)	0.6692(0.1990)	0.6962(0.1851)
$b = 0.01$	1.7726(2.1468)	0.6704(0.1995)	0.6969(0.1853)

<sup>1</sup>Simulation studies so far show that the filtered covariance of GF at any time instant is larger than that of the EqKF. This fact may be partially supported by (30), but is not proved analytically.

## V. CUBIC SENSOR PROBLEM 2

In this section, we consider again the scalar linear system of (15) with  $b = 0$ , where  $a$  is now an unknown parameter to be estimated,  $\beta > 0$  is a known constant, and  $w_t$  and  $v_t$  are white Gaussian noises with means zero and variances  $q$  and  $r$ , respectively.

Define  $x_{1,t} := x_t$  and  $x_{2,t} := a$ . Then, the extended state space model is expressed as

$$\begin{bmatrix} x_{1,t+1} \\ x_{2,t+1} \end{bmatrix} = \begin{bmatrix} x_{1,t} & x_{2,t} \\ & x_{2,t} \end{bmatrix} + \begin{bmatrix} w_t \\ 0 \end{bmatrix} \quad (31a)$$

$$y_t = \beta x_{1,t}^3 + v_t \quad (31b)$$

Let the conditional pdf of the state vector  $x_t = [x_{1,t} \ x_{2,t}]^T$  based on  $Y^{t-1}$  be given by  $N(\hat{x}_{t/t-1}, P_{t/t-1})$ , i.e.

$$\begin{bmatrix} x_{1,t} \\ x_{2,t} \end{bmatrix} \sim N \left( \begin{bmatrix} \hat{x}_{1,t/t-1} \\ \hat{x}_{2,t/t-1} \end{bmatrix}, \begin{bmatrix} P_{t/t-1}^{(11)} & P_{t/t-1}^{(12)} \\ P_{t/t-1}^{(21)} & P_{t/t-1}^{(22)} \end{bmatrix} \right)$$

Since the output nonlinearity is  $h_t = \beta x_{1,t}^3$ , the output prediction is expressed as

$$\hat{y}_{t/t-1} = \beta(\hat{x}_{1,t/t-1}^3 + 3\hat{x}_{1,t/t-1}P_{t/t-1}^{(11)}) \quad (32)$$

so that from (14), the equivalent gain for  $h_t$  is given by

$$H_t^e = [3\beta(\hat{x}_{1,t/t-1}^2 + P_{t/t-1}^{(11)}) \quad 0] \quad (33)$$

Also, the prediction of  $x_{t+1}$  based on  $Y^t$  is

$$\hat{x}_{t+1/t} = [\hat{x}_{1,t/t} \hat{x}_{2,t/t} + P_{t/t}^{(12)} \quad \hat{x}_{2,t/t}]^T$$

Hence, the equivalent gain matrix for  $f_t$  becomes

$$F_t^e = \begin{bmatrix} \hat{x}_{2,t/t} & \hat{x}_{1,t/t} \\ 0 & 1 \end{bmatrix}$$

Define  $Q = \text{diag}(q, 0)$ . Then, the EqKF algorithm can be summarized as follows.

*Algorithm 3:* (EqKF algorithm)

1) The measurement update

$$K_t^e = P_{t/t-1} (H_t^e)^T [H_t^e P_{t/t-1} (H_t^e)^T + r]^{-1}$$

$$\hat{x}_{t/t} = \hat{x}_{t/t-1} + K_t^e [y_t - \hat{y}_{t/t-1}]$$

$$P_{t/t} = P_{t/t-1} - K_t^e H_t^e P_{t/t-1}$$

where  $\hat{y}_{t/t-1}$  is given by (32).

2) The time update

$$\hat{x}_{t+1/t} = \begin{bmatrix} \hat{x}_{1,t/t} \hat{x}_{2,t/t} + P_{t/t}^{(12)} \\ \hat{x}_{2,t/t} \end{bmatrix} \quad (34)$$

$$P_{t+1/t} = F_t^e P_{t/t} (F_t^e)^T + Q \quad (35)$$

3) The initial values  $\hat{x}_{0/-1} = \bar{x}_0$ ,  $P_{0/-1} = P_0$ .  $\square$

It may be noted that the EKF algorithm is obtained by deleting  $P_{t/t-1}^{(11)}$  in (32), (33) and  $P_{t/t-1}^{(12)}$  in (34).

For the GF, it follows from (3) that

$$U_{t/t-1} = \begin{bmatrix} 3\beta P_{t/t-1}^{(11)} (\hat{x}_{1,t/t-1}^2 + P_{t/t-1}^{(11)}) \\ 3\beta P_{t/t-1}^{(21)} (\hat{x}_{1,t/t-1}^2 + P_{t/t-1}^{(11)}) \end{bmatrix} \quad (36)$$

Also, from (4),

$$V_{t/t-1} = \beta^2 [9\hat{x}_{1,t/t-1}^4 + 36\hat{x}_{1,t/t-1}^2 P_{t/t-1}^{(11)} + 15(P_{t/t-1}^{(11)})^2] P_{t/t-1}^{(11)} + r \quad (37)$$

The measurement update equations are determined by  $\hat{y}_{t/t-1}$  of (32) and  $K_t = U_{t/t-1} V_{t/t-1}^{-1}$  using (36) and (37).

For the time update, the prediction of  $x_{t+1}$  based on  $Y^t$  is given by (34), so that we see from (9) that

$$\begin{aligned} P_{t+1/t}^{(11)} &= \hat{x}_{1,t/t-1}^2 P_{t/t}^{(22)} + 2\hat{x}_{1,t/t-1} \hat{x}_{2,t/t-1} P_{t/t}^{(12)} \\ &\quad + \hat{x}_{2,t/t-1}^2 P_{t/t}^{(11)} + P_{t/t}^{(11)} P_{t/t}^{(22)} + (P_{t/t}^{(12)})^2 + q \\ P_{t+1/t}^{(12)} &= \hat{x}_{1,t/t-1} P_{t/t}^{(22)} + \hat{x}_{2,t/t-1} P_{t/t}^{(12)} \\ P_{t+1/t}^{(21)} &= \hat{x}_{1,t/t-1} P_{t/t}^{(22)} + \hat{x}_{2,t/t-1} P_{t/t}^{(12)} \\ P_{t+1/t}^{(22)} &= P_{t/t}^{(22)} \end{aligned} \quad (38)$$

Note that if we delete  $P_{t/t}^{(11)} P_{t/t}^{(22)} + (P_{t/t}^{(12)})^2$  in the second line of (38), then it reduces to the time update equation (35) of EqKF.

We show simulation results by the EqKF and GF, where we assume that  $a = 0.96$ , but other parameters are the same as in Section V, i.e.  $\beta = 1/10$ ,  $q = r = 1$  and  $x_{1,0} = 0$ . The initial values are  $\hat{x}_{1,0/-1} = 5$ ,  $\hat{x}_{2,0/-1} = 0.01$  and  $P_{0/-1} = \text{diag}(2, 2)$ .

There is not much difference in the sample estimates  $\hat{x}_{2,t/t}$  of the parameter  $a$  by the EqKF and GF. Table III displays the sample average of the estimates  $\hat{x}_{2,N/N}$  with  $M = 100$  and  $N = 250$ , where the numbers in parentheses denote standard deviation. This implies that the performance of the GF is somewhat superior to that of the EqKF.

TABLE III  
PARAMETER ESTIMATION BY EQKF AND GF.

	True	EqKF	GF
$a$	0.96	0.9316 (0.0340)	0.9496 (0.0261)

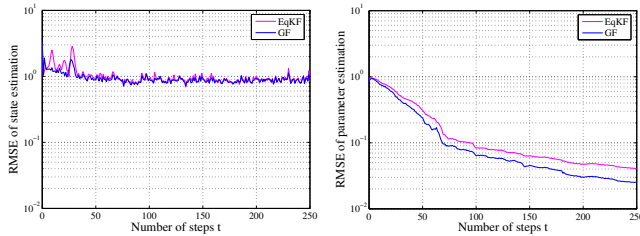


Fig. 4. Plot of RMSEs  $E_{1,t}$  (left) and  $E_{2,t}$  (right).

Moreover, Fig. 4 depicts RMSEs  $E_{i,t}$ ,  $i = 1, 2$  of state and parameter estimation for  $M = 100$  runs, where

$$E_{i,t} = \sqrt{\frac{1}{M} \sum_{j=1}^M (x_{i,t}^{(j)} - \hat{x}_{i,t/t}^{(j)})^2}, \quad i = 1, 2$$

It may be noted that  $i = 1, 2$  denote the state and parameter, respectively.

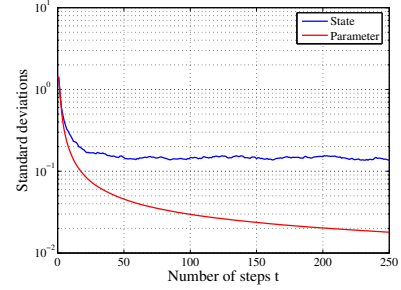


Fig. 5. Square roots of Cramér-Rao lower bounds for state and parameter estimation.

Fig. 5 displays the square roots of Cramér-Rao lower bounds for state and parameter estimation, that are computed by using 2000 particles; see [23] for computational algorithm. The curves of Fig. 4(left) are not very close to the square root of Cramér-Rao lower bound in Fig. 5. The curves of Fig. 4(right) are, however, closer to the square root of Cramér-Rao lower bound in Fig. 5. We see that as far as the parameter estimation is concerned, both the performances of the EqKF and GF are close to the optimal performance evaluated by the Cramér-Rao lower bound, though the GF shows somewhat better performance than the EqKF.

## VI. CONCLUSIONS

In this paper, we have revisited the statistical equivalent linearization technique for static nonlinearities to better understand the difference and similarity between the EKF and EqKF in the framework of the GF. Results of this paper can be summarized as follows.

- We have examined the relationship between the first-order linearization and the equivalent linearization for differentiable nonlinearities, showing that the equivalent linearization is a "global" linearization method. Also, two methods of computing the equivalent gain are derived for a differentiable nonlinearity.
- For cubic sensor problems, we have shown that the Kalman gain and covariance equation of the EqKF are close to those of the GF, but are quite different from those of the EKF. This is because the Jacobian of the EKF includes the mean information only and not the covariance information.
- Simulation studies show that though the EKF often degenerates for the cubic sensor problems, the EqKF exhibits a good estimation and tracking performance like the GF.

A disadvantage of the EqKF is that its applicability is limited to stochastic systems with polynomial or product type nonlinearities. Thus, for general nonlinear systems, we need some numerical procedures such as Gauss-Hermite or cubature integration rule [16], [19], [22] to approximately evaluate the following conditional expectations at each step:

$$E\{h_t(x_t) | Y^{t-1}\}, \quad E\left\{\frac{\partial h_t(x_t)}{\partial x_t} \Big| Y^{t-1}\right\}$$

and

$$E\{f_t(x_t) | Y^t\}, \quad E\left\{\frac{\partial f_t(x_t)}{\partial x_t} \Big| Y^t\right\}$$

#### APPENDIX: PROOF OF PROPOSITION 1

Let  $\xi = x - m$ . Then,  $\xi \sim N(0, P)$ , so that

$$\begin{aligned} G^e &= E\{g(x)(x - m)^T\}P^{-1} = E\{g(\xi + m)\xi^T\}P^{-1} \\ &= \frac{1}{\sqrt{(2\pi)^n |P|}} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(\xi + m)\xi^T P^{-1} \\ &\quad \times e^{-\frac{1}{2}\xi^T P^{-1}\xi} d\xi_1 \cdots d\xi_n \end{aligned}$$

Hence, the  $(i, j)$  component of  $G^e$  becomes

$$G_{ij}^e = \frac{1}{\sqrt{(2\pi)^n |P|}} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g_i(\xi + m)(\xi^T P^{-1})_j \times e^{-\frac{1}{2}\xi^T P^{-1}\xi} d\xi_1 \cdots d\xi_n$$

Since

$$\frac{d}{d\xi_j} e^{-\frac{1}{2}\xi^T P^{-1}\xi} = -(\xi^T P^{-1})_j e^{-\frac{1}{2}\xi^T P^{-1}\xi}$$

we see that

$$G_{ij}^e = \frac{1}{\sqrt{(2\pi)^n |P|}} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g_i(\xi + m) \times \left(-\frac{d}{d\xi_j} e^{-\frac{1}{2}\xi^T P^{-1}\xi}\right) d\xi_1 \cdots d\xi_n$$

Integration by parts with respect to  $\xi_j$  yields

$$\begin{aligned} G_{ij}^e &= \frac{1}{\sqrt{(2\pi)^n |P|}} \underbrace{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty}}_{n-1} \\ &\quad \times \left[-g_i(\xi + m)e^{-\frac{1}{2}\xi^T P^{-1}\xi}\right]_{\xi_j=-\infty}^{\infty} d\xi_1 \cdots \hat{d\xi_j} \cdots d\xi_n \\ &\quad + \frac{1}{\sqrt{(2\pi)^n |P|}} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \frac{\partial}{\partial \xi_j} g_i(\xi + m) \\ &\quad \times e^{-\frac{1}{2}\xi^T P^{-1}\xi} d\xi_1 \cdots d\xi_n \end{aligned} \quad (39)$$

where  $\hat{d\xi_j}$  denotes that  $d\xi_j$  is removed. It follows from the assumption (10) that

$$\lim_{\xi_j \rightarrow \pm\infty} g_i(\xi + m)e^{-\frac{1}{2}\xi^T P^{-1}\xi} = 0$$

so that the first term in the right-hand side of (39) vanishes. Hence, we get

$$\begin{aligned} G_{ij}^e &= \frac{1}{\sqrt{(2\pi)^n |P|}} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \frac{\partial}{\partial \xi_j} g_i(\xi + m) \\ &\quad \times e^{-\frac{1}{2}\xi^T P^{-1}\xi} d\xi_1 \cdots d\xi_n \\ &= E\left\{\frac{\partial}{\partial \xi_j} g_i(\xi + m)\right\} \end{aligned} \quad (40)$$

The first equality in (11) is now proved. Moreover,

$$\frac{\partial}{\partial \xi_j} g_i(\xi + m) = \frac{\partial}{\partial m_j} g_i(\xi + m)$$

so that the right-hand side of (40) equals

$$E\left\{\frac{\partial}{\partial m_j} g_i(\xi + m)\right\} = \frac{\partial}{\partial m_j} E\{g_i(\xi + m)\}$$

This completes a proof of the second equality.  $\square$

#### REFERENCES

- [1] B. C. Booton, The analysis of nonlinear control systems with random inputs, *IRE Trans. Circuit Theory*, Vol. CT-1, No. 1, pp. 32–34, 1954.
- [2] I. E. Kazakov, Approximate probabilistic analysis of the accuracy of operation of essentially nonlinear systems, *Automation & Remote Control*, Vol. 17, No. 5, pp. 423–450, 1956.
- [3] I. E. Kazakov, Generalization of the method of statistical linearization to multidimensional systems, *Automation and Remote Control*, Vol. 26, pp. 1201–1206, 1965.
- [4] Y. Sawaragi and S. Takahashi, Response of control systems containing zero-memory nonlinearity to sinusoidal and Gaussian inputs, *Proc. Int. Congress of Automatic Control*, Heidelberg, 1956, pp. 271–274.
- [5] Y. Sawaragi, N. Sugai and Y. Sunahara, *Statistical Studies on Nonlinear Control Systems*, Nippon Printing & Publishing Co., Osaka, 1962.
- [6] T. K. Caughey, Equivalent linearization technique, *J. Acoustical Soc. America*, Vol. 35, No. 11, pp. 1706–1711, 1963.
- [7] J. B. Roberts and P. D. Spanos, *Random Vibration and Statistical Linearization*, Wiley 1990 (Dover edition 2003).
- [8] S. H. Crandall, On using non-Gaussian distribution to perform statistical linearization, *Int. J. Non-Linear Mechanics*, Vol. 39, No. 11, pp. 1395–1406, 2004.
- [9] R. S. Bucy, Bayes theorem and digital realizations for nonlinear filters, *J. Astronautical Sci.*, Vol. 17, No. 2, pp. 80–94, 1969.
- [10] A. H. Jazwinski, *Stochastic Processes and Filtering Theory*, Academic, 1970.
- [11] T. Katayama, Studies on state and parameter estimation for discrete-time stochastic systems, PhD Thesis, Department of Applied Mathematics and Physics, Kyoto University, March 1969.
- [12] Y. Sunahara, An approximate method of state estimation for nonlinear dynamical systems, *Trans. ASME, J. Basic Eng.*, Vol. 92D, No. 2, pp. 385–393, 1970.
- [13] A. Gelb (ed.), *Applied Optimal Estimation*, MIT Press, 1974.
- [14] J. W. Austin and C. T. Leondes, Statistically linearized estimation of reentry trajectories, *Trans. Aerospace & Electronic Systems*, Vol. AES-17, No. 1, 54–61, 1981.
- [15] S. Sugimoto, Y. Kubo and N. Munetomo, A quasi-linear filter with conditional Gaussian sum distributions for nonlinear dynamical systems, *Preprints of the 18th IFAC World Congress*, Milano, 2011, pp. 7797–7802.
- [16] K. Ito and K. Xiong, Gaussian filters for nonlinear filtering problems, *IEEE Trans. Automat. Control*, Vol. 45, No. 5, pp. 910–927, 2000.
- [17] S. J. Julier, J. K. Uhlmann and H. F. Durrant-Whyte, A new method for the nonlinear transformation of means and covariances in filters and estimators, *IEEE Trans. Automat. Control*, Vol. 45, No. 3, pp. 477–482, 2000.
- [18] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*, Prentice-Hall, 1979.
- [19] Y. Wu, D. Hu, M. Wu and X. Hu, A numerical-integration perspective on Gaussian filters, *IEEE Trans. Signal Processing*, Vol. 54, No. 8, pp. 2910–2921, 2006.
- [20] G. Evensen, The ensemble Kalman filter for combined state and parameter estimation, *IEEE Control Systems Magazine*, Vol. 29, No. 3, pp. 83–104, 2009.
- [21] T. Lefebvre, H. Bruyninckx and J. De Schutter, Comment on “A new method for the nonlinear transformation of means and covariances in filters and estimators”, *IEEE Trans. Automat. Control*, Vol. 47, No. 8, pp. 1406–1408, 2002.
- [22] I. Arasaratnam and S. Haykin, Cubature Kalman filters, *IEEE Trans. Automat. Control*, Vol. 54, No. 6, pp. 1254–1269, 2009.
- [23] M. Šimandl, J. Kravovec and P. Tichavský, Filtering, predictive and smoothing Cramér-Rao bounds for discrete-time nonlinear dynamic systems, *Automatica*, Vol. 37, No. 11, pp. 1703–1716, 2001.