JOURNAL OF GUIDANCE, CONTROL, AND DYNAMICS
Vol. 34, No. 2, March–April 2011

Parameter Estimation of Aircraft Dynamics via Unscented Smoother with Expectation-Maximization Algorithm

Nobuhiro Yokoyama*
National Defense Academy, Yokosuka 239-8686, Japan

DOI: 10.2514/1.51696

This paper proposes a new method for parameter estimation of aircraft dynamics modeled in state space. The developed method employs the smoother, which estimates the state and unknown parameters, combined with expectation-maximization algorithm to estimate unknown statistics in the problem, i.e., the mean and covariance of an initial state, and the noise covariances. To approximate the expectation values in the expectation-maximization with a reasonable computational cost, an unscented transform based on the estimates obtained by the unscented Rauch–Tung–Striebel smoother is employed. Moreover, nonconvex numerical optimization algorithms are not necessary in the maximization step of the expectation-maximization, because the optimums of the unknown statistics are given in analytical forms. Thus, the developed method achieves low computational cost and high robustness. Its effectiveness is demonstrated through two problems of estimating aircraft aerodynamic parameters.

I. Introduction

PARAMETER estimation of aircraft through flight test data is an important technique for analyzing flight mechanics and designing flight control systems [1–3]. This paper proposes a new method for parameter estimation of aircraft dynamics modeled in state space.

Many estimation methods have been developed so far. The equation error method [4,5] and the output error method [6,7] have been widely used for parameter estimation of aircraft due to their simplicity. However, practical flight test data is often significantly affected by both the process and measurement noise [3]. Thus, the filter error method (FEM) [8,9] and the filtering approach (FA) [10], which cover both the process and measurement noise, have been regarded as reliable estimation methods [3]. In the FEM, the maximum likelihood problem based on the estimated output obtained by an optimal filter is solved. This method has been frequently applied to linear dynamic models [8] using the Kalman filter, but it has also been extended to nonlinear dynamic models using the extended Kalman filter (EKF) [9]. In the FA, on the other hand, a parameter estimation problem is transformed into an augmented state estimation problem by regarding the unknown parameters as additional state variables. The Kalman filter offers optimal estimation of the state variables for linear dynamic models, while the EKF and the unscented Kalman filter (UKF) [11,12], which estimates the states and outputs via an unscented transform, are frequently used for nonlinear dynamic models. Both the FEM and FA are based on the optimal filtering, i.e., the optimal estimation using the measurement data at a certain time is iterated from an initial time to a final time. In contrast, the smoothing approach computes optimal estimates of unknowns based on the entire measurement history. Thus, it has a potential to offer more accurate estimations than the method based on the filtering.

Methods using the nonlinear programming [13,14] are the candidates for effective smoothing. However, they can be computationally expensive and not easy to implement, because they include nonconvex numerical optimization algorithm such as the interior point method [15] and the sequential quadratic programming [16,17]. In this paper, on the other hand, we adopt the unscented Rauch–Tung–Striebel smoother (URTSS) [18], which is computationally inexpensive and relatively easy to implement. The URTSS uses an optimal backward smoothing, which recursively computes corrections to the forward estimation by the UKF. By using the unscented transform, the URTSS approximate the probability density functions (PDFs) more accurately than the smoother based on the first-order approximation, i.e., the extended Kalman smoother (EKS). Actually, it has been reported [18] that the URTSS achieved more accurate estimation than the UKF, EKF, and EKS. Nevertheless, similarly to the UKF, EKF, and EKS, one of the major limitations of the URTSS is that it needs another approach to estimate statistics in the problem, i.e., the mean and covariance of initial state (or initial augmented state) and the noise covariances.

Therefore, this paper proposes a new algorithm which combines the URTSS and the expectation-maximization (EM) algorithm [19]. The EM is used to estimate the unknown statistics in the problem. Similar methods which combine some smoothers with the EM are given in [20,21]. The application area of the method in [20] is limited to linear systems. On the other hand, the method in [21], which adopts the particle smoother [22], is applicable to nonlinear systems with unknown parameters and statistics. Nevertheless, it is necessary to numerically solve an optimization problem of a log-likelihood function, which is generally nonconvex and often numerically ill-conditioned. Thus, the convergence of the algorithm is affected by the robustness of the numerical optimization algorithm used therein. Moreover, the particle smoother may be computationally expensive because it needs many sample points to achieve accurate estimation. In sharp contrast to this, the method proposed in this paper does not require nonconvex optimization algorithm, because the optimums of the unknown statistics at each step of the EM are derived in analytical forms. Moreover, the computational cost of the developed method is relatively low, because an unscented transform based on the estimation results by the URTSS, which attains high accuracy with a reasonable number of sample points (i.e., sigma-points), is adopted.

The effectiveness of the developed method is demonstrated through two examples. The one is a problem of estimating lateral-directional aerodynamic parameters of linearized aircraft dynamics. This is a relatively simple benchmark problem and the flight test data used therein is artificially generated. The other is a problem of estimating longitudinal aerodynamic parameters of nonlinear dynamics of a research aircraft. A real flight test data is used therein and the problem is of more practical importance.

Received 20 July 2010; revision received 20 September 2010; accepted for publication 21 September 2010.

*Research Associate, Department of Aerospace Engineering, 1-10-20 Hashirimizu. Senior Member AIAA.

II. Developed Method

A. EM Algorithm

Let $\boldsymbol{\theta}$, $\boldsymbol{X}$, and $\boldsymbol{Y}$ be a vector of unknown statistics, a set of missing data, and a set of measured data, respectively. The elements of $\boldsymbol{X}$ and $\boldsymbol{Y}$ are assumed as probabilistic variables. Thus, due to the definition of conditional PDFs, the following equation holds:

$$\log p_{\boldsymbol{\theta}}(\boldsymbol{Y}) = \log p_{\boldsymbol{\theta}}(\boldsymbol{X}, \boldsymbol{Y}) - \log p_{\boldsymbol{\theta}}(\boldsymbol{X}|\boldsymbol{Y}) \tag{1}$$

where $p_{\boldsymbol{\theta}}(\cdot)$ denotes a PDF that depends on $\boldsymbol{\theta}$. The objective of the maximum likelihood estimation is to find $\boldsymbol{\theta}$ which maximizes the $p_{\boldsymbol{\theta}}(\boldsymbol{Y})$, i.e., the likelihood function. Let $\hat{\boldsymbol{\theta}}$ be a current estimate of $\boldsymbol{\theta}$. Then, expectations of the both sides of Eq. (1) conditioned on $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{Y}$ are

$$\begin{aligned} \log p_{\boldsymbol{\theta}}(\boldsymbol{Y}) &= \int p_{\hat{\boldsymbol{\theta}}}(\boldsymbol{X}|\boldsymbol{Y}) \log p_{\boldsymbol{\theta}}(\boldsymbol{Y})\, \mathrm{d}\boldsymbol{X} \\ &= \int p_{\hat{\boldsymbol{\theta}}}(\boldsymbol{X}|\boldsymbol{Y}) \log p_{\boldsymbol{\theta}}(\boldsymbol{X}, \boldsymbol{Y})\, \mathrm{d}\boldsymbol{X} \\ &\quad - \int p_{\hat{\boldsymbol{\theta}}}(\boldsymbol{X}|\boldsymbol{Y}) \log p_{\boldsymbol{\theta}}(\boldsymbol{X}|\boldsymbol{Y})\, \mathrm{d}\boldsymbol{X} \end{aligned} \tag{2}$$

From (2), the following equation holds:

$$\begin{aligned} \log p_{\boldsymbol{\theta}}(\boldsymbol{Y}) - \log p_{\hat{\boldsymbol{\theta}}}(\boldsymbol{Y}) &= L(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}) - L(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\theta}}) \\ &\quad + \int p_{\hat{\boldsymbol{\theta}}}(\boldsymbol{X}|\boldsymbol{Y}) \log \frac{p_{\hat{\boldsymbol{\theta}}}(\boldsymbol{X}|\boldsymbol{Y})}{p_{\boldsymbol{\theta}}(\boldsymbol{X}|\boldsymbol{Y})}\, \mathrm{d}\boldsymbol{X} \end{aligned} \tag{3}$$

where

$$L(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}) \triangleq \int p_{\hat{\boldsymbol{\theta}}}(\boldsymbol{X}|\boldsymbol{Y}) \log p_{\boldsymbol{\theta}}(\boldsymbol{X}, \boldsymbol{Y})\, \mathrm{d}\boldsymbol{X} \tag{4}$$

The third term of Eq. (3) is referred to as the Kullback–Leibler information distance [23] from $p_{\boldsymbol{\theta}}(\boldsymbol{X}|\boldsymbol{Y})$ to $p_{\hat{\boldsymbol{\theta}}}(\boldsymbol{X}|\boldsymbol{Y})$, and it has been proved to be nonnegative for arbitrary $\boldsymbol{\theta}$ and $\hat{\boldsymbol{\theta}}$. Therefore, from Eq. (3), if $L(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}) > L(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\theta}})$, then $p_{\boldsymbol{\theta}}(\boldsymbol{Y}) > p_{\hat{\boldsymbol{\theta}}}(\boldsymbol{Y})$. On the basis of this viewpoint, the EM calculates the optimum of $\boldsymbol{\theta}$, which maximizes the likelihood $p_{\boldsymbol{\theta}}(\boldsymbol{Y})$, by alternately repeating the following steps:

1) E (expectation) step:

$$\text{Calculate}: \ L(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}_t) \tag{5}$$

2) M (maximization) step:

$$\text{Find}: \ \hat{\boldsymbol{\theta}}_{t+1} = \arg\max_{\boldsymbol{\theta}} L(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}_t) \tag{6}$$

where $\hat{\boldsymbol{\theta}}_t$ denotes the estimate of $\boldsymbol{\theta}$ at $t(\geq 1)$-th iteration. An initial estimate of $\boldsymbol{\theta}$, i.e., $\hat{\boldsymbol{\theta}}_0$, should be given in advance.

B. Application of EM to Parameter Estimation of Nonlinear Systems

The following nonlinear discrete system is covered in this paper:

$$\tilde{\boldsymbol{x}}_{k+1} = \boldsymbol{f}(\tilde{\boldsymbol{x}}_k, \boldsymbol{u}_k, \boldsymbol{\rho}) + \boldsymbol{w}_k \tag{7}$$

$$\boldsymbol{y}_k = \boldsymbol{h}(\tilde{\boldsymbol{x}}_k, \boldsymbol{u}_k, \boldsymbol{\rho}) + \boldsymbol{v}_k \tag{8}$$

where $k(=0, 1, \cdots, N)$ is the index of time step, $\tilde{\boldsymbol{x}}_k \in \mathbb{R}^q$ is the state, $\boldsymbol{u}_k \in \mathbb{R}^m$ is the measured input, $\boldsymbol{\rho} \in \mathbb{R}^l$ is the unknown parameter, $\boldsymbol{y}_k \in \mathbb{R}^p$ is the measured output, $\boldsymbol{w}_k \in \mathbb{R}^q$ is the process noise, and $\boldsymbol{v}_k \in \mathbb{R}^p$ is the measurement noise. It is assumed that the measured inputs $\boldsymbol{u}_0, \cdots, \boldsymbol{u}_N$ are deterministic variables, whereas the measured outputs $\boldsymbol{y}_0, \cdots, \boldsymbol{y}_N$ are probabilistic variables. In addition, the process noise $\boldsymbol{w}_k$ and the measurement noise $\boldsymbol{v}_k$ are assumed to yield the Gaussians with zero means, i.e.,

$$\boldsymbol{w}_k \sim N(\boldsymbol{0}, \boldsymbol{Q}) \tag{9}$$

$$\boldsymbol{v}_k \sim N(\boldsymbol{0}, \boldsymbol{R}) \tag{10}$$

where the covariances $\boldsymbol{Q} \in \mathbb{R}^{q\times q}$ and $\boldsymbol{R} \in \mathbb{R}^{p\times p}$ are positive definite matrices. Although Eq. (7) appears to only cover the additive process noise, it is possible to convert state equations with nonadditive process noise in the form of Eq. (7), as described in Appendix A.

Let us introduce an augmented state $\boldsymbol{x}_k \triangleq [\tilde{\boldsymbol{x}}_k^T \quad \boldsymbol{\rho}_k^T]^T \in \mathbb{R}^n$ ($n \triangleq q + l$) and the following augmented state equation:

$$\boldsymbol{x}_{k+1} = \begin{bmatrix} \tilde{\boldsymbol{x}}_{k+1} \\ \boldsymbol{\rho}_{k+1} \end{bmatrix} = \begin{bmatrix} \boldsymbol{f}(\tilde{\boldsymbol{x}}_k, \boldsymbol{u}_k, \boldsymbol{\rho}_k) + \boldsymbol{w}_k \\ \boldsymbol{\rho}_k \end{bmatrix} \tag{11}$$

Moreover, assume that the initial augmented state $\boldsymbol{x}_0$ yields a Gaussian as follows:

$$\boldsymbol{x}_0 \sim N(\hat{\boldsymbol{x}}_0, \boldsymbol{P}_0) \tag{12}$$

The statistics $\hat{\boldsymbol{x}}_0, \boldsymbol{P}_0, \boldsymbol{Q}$, and $\boldsymbol{R}$ are also unknown in general. Thus, the problem considered in this paper is estimating these unknown statistics as well as the unknown system parameter $\boldsymbol{\rho}$ based on the measured data sets of inputs $\boldsymbol{U}_N \triangleq \{\boldsymbol{u}_0, \cdots, \boldsymbol{u}_N\}$ and outputs $\boldsymbol{Y}_N \triangleq \{\boldsymbol{y}_0, \cdots, \boldsymbol{y}_N\}$. The unknown statistics are estimated in the EM by collecting their elements into $\boldsymbol{\theta}$. On the other hand, the unknown parameter $\boldsymbol{\rho}$ is estimated by applying the URTSS to Eqs. (8) and (11) based on the current estimate of $\boldsymbol{\theta}$ given by the EM. This means that a set of augmented states $\boldsymbol{X}_N \triangleq \{\boldsymbol{x}_0, \cdots, \boldsymbol{x}_N\}$ and $\boldsymbol{Y}_N$ are, respectively, taken as the missing data and the measured data in the EM.

From Eqs. (8) and (11), $\log p_{\boldsymbol{\theta}}(\boldsymbol{X}_N, \boldsymbol{Y}_N)$ can be written as follows:

$$\begin{aligned} \log p_{\boldsymbol{\theta}}(\boldsymbol{X}_N, \boldsymbol{Y}_N) &= \log p_{\boldsymbol{\theta}}(\boldsymbol{x}_0) + \sum_{k=0}^{N-1} \log p_{\boldsymbol{\theta}}(\boldsymbol{x}_{k+1}|\boldsymbol{x}_k) \\ &\quad + \sum_{k=0}^{N} \log p_{\boldsymbol{\theta}}(\boldsymbol{y}_k|\boldsymbol{x}_k) \end{aligned} \tag{13}$$

Now let $J(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}) = -2L(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$ be the cost function to be minimized, i.e.,

$$J(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}) = -2L(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}) = -2\int p_{\hat{\boldsymbol{\theta}}}(\boldsymbol{X}_N|\boldsymbol{Y}_N) \log p_{\boldsymbol{\theta}}(\boldsymbol{X}_N, \boldsymbol{Y}_N)\, \mathrm{d}\boldsymbol{X}_N \tag{14}$$

Substituting Eq. (13) into Eq. (14), $J(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$ can be rewritten as follows:

$$J(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}) = J_1 + J_2 + J_3 \tag{15}$$

where

$$J_1 \triangleq -2\int p_{\hat{\boldsymbol{\theta}}}(\boldsymbol{x}_0|\boldsymbol{Y}_N) \log p_{\boldsymbol{\theta}}(\boldsymbol{x}_0)\, \mathrm{d}\boldsymbol{x}_0 \tag{16}$$

$$J_2 \triangleq -2\sum_{k=0}^{N-1} \int p_{\hat{\boldsymbol{\theta}}}(\boldsymbol{x}_k, \boldsymbol{x}_{k+1}|\boldsymbol{Y}_N) \log p_{\boldsymbol{\theta}}(\boldsymbol{x}_{k+1}|\boldsymbol{x}_k)\, \mathrm{d}\boldsymbol{x}_{k+1}\, \mathrm{d}\boldsymbol{x}_k \tag{17}$$

$$J_3 \triangleq -2\sum_{k=0}^{N} \int p_{\hat{\boldsymbol{\theta}}}(\boldsymbol{x}_k|\boldsymbol{Y}_N) \log p_{\boldsymbol{\theta}}(\boldsymbol{y}_k|\boldsymbol{x}_k)\, \mathrm{d}\boldsymbol{x}_k \tag{18}$$

In the developed method, the expectation values in Eqs. (16–18) are calculated by using an unscented transform based on the smoothing. For this purpose, the UKF and URTSS are adopted. Thus, we give brief overviews of the UKF and URTSS in Secs. II.C and II.D before deriving the method to calculate Eqs. (16–18).

C. Overview of UKF

Similar to the conventional Kalman filters, the UKF is composed of the time updates and the measurement updates. In the time update, sigma-points $\chi^{(0)}_{k|k}, \chi^{(1)}_{k|k}, \cdots, \chi^{(2n)}_{k|k}$, and their associated weights defined next are adopted:

$$\chi^{(i)}_{k|k} = \begin{bmatrix} \chi^{x(i)}_{k|k} \\ \chi^{\rho(i)}_{k|k} \end{bmatrix} \triangleq \begin{cases} \hat{x}_{k|k} & : i = 0 \\ \hat{x}_{k|k} + \alpha_f \sqrt{n+\kappa_f} (P^{xx}_{k|k})^{1/2}_i & : 1 \le i \le n \\ \hat{x}_{k|k} - \alpha_f \sqrt{n+\kappa_f} (P^{xx}_{k|k})^{1/2}_{i-n} & : n+1 \le i \le 2n \end{cases} \tag{19}$$

$$W^{(i)}_{mf} \triangleq \begin{cases} 1 - n/[\alpha_f^2(n+\kappa_f)] & : i = 0, \\ 1/[2\alpha_f^2(n+\kappa_f)] & : 1 \le i \le 2n \end{cases} \tag{20}$$

$$W^{(i)}_{cf} \triangleq \begin{cases} W^{(i)}_{mf} + 1 - \alpha_f^2 + \beta_f & : i = 0, \\ W^{(i)}_{mf} & : 1 \le i \le 2n \end{cases} \tag{21}$$

where $\chi^{x(i)}_{k|k} \in \mathbb{R}^q$ and $\chi^{\rho(i)}_{k|k} \in \mathbb{R}^l$ correspond to the state and the unknown parameter, respectively. $\hat{x}_{k|k}$ and $P^{xx}_{k|k}$ denotes the mean and covariance of $p_{\hat{\theta}}(x_k|Y_k)$. In addition, $(P^{xx}_{k|k})^{1/2}_i$ denotes the i-th column of $(P^{xx}_{k|k})^{1/2}$. Furthermore, α_f, β_f, and κ_f are tuning parameters, whose values are typically $\alpha_f \in [10^{-3}, 1)$, $\beta_f = 2$, and $\kappa_f = 3 - n$ or 0. Then, the means and covariances of $p_{\hat{\theta}}(x_{k+1}|Y_k)$ and $p_{\hat{\theta}}(y_{k+1}|Y_k)$ are approximated by the following unscented transform:

$$\hat{x}_{k+1|k} \cong \sum_{i=0}^{2n} W^{(i)}_{mf} \chi^{(i)}_{k+1|k} \tag{22}$$

$$P^{xx}_{k+1|k} \cong \sum_{i=0}^{2n} W^{(i)}_{cf} (\chi^{(i)}_{k+1|k} - \hat{x}_{k+1|k})(\chi^{(i)}_{k+1|k} - \hat{x}_{k+1|k})^T + \mathrm{diag}[Q, 0_{l\times l}] \tag{23}$$

$$\hat{y}_{k+1|k} \cong \sum_{i=0}^{2n} W^{(i)}_{mf} \eta^{(i)}_{k+1|k} \tag{24}$$

$$P^{yy}_{k+1|k} \cong \sum_{i=0}^{2n} W^{(i)}_{cf} (\eta^{(i)}_{k+1|k} - \hat{y}_{k+1|k})(\eta^{(i)}_{k+1|k} - \hat{y}_{k+1|k})^T + R \tag{25}$$

$$P^{xy}_{k+1|k} \cong \sum_{i=0}^{2n} W^{(i)}_{cf} (\chi^{(i)}_{k+1|k} - \hat{x}_{k+1|k})(\eta^{(i)}_{k+1|k} - \hat{y}_{k+1|k})^T \tag{26}$$

where

$$\chi^{(i)}_{k+1|k} = \begin{bmatrix} \chi^{x(i)}_{k+1|k} \\ \chi^{\rho(i)}_{k+1|k} \end{bmatrix} = \begin{bmatrix} f(\chi^{x(i)}_{k|k}, u_k, \chi^{\rho(i)}_{k|k}) \\ \chi^{\rho(i)}_{k|k} \end{bmatrix} \quad (i = 0, \cdots, 2n) \tag{27}$$

$$\eta^{(i)}_{k+1|k} = h(\chi^{x(i)}_{k+1|k}, u_{k+1}, \chi^{\rho(i)}_{k+1|k}) \quad (i = 0, \cdots, 2n) \tag{28}$$

On the other hand, in the measurement update, the mean and covariance of the augmented state are calculated by the following equations:

$$\hat{x}_{k+1|k+1} = \hat{x}_{k+1|k} + L_{k+1}(y_{k+1} - \hat{y}_{k+1|k}) \tag{29}$$

$$P^{xx}_{k+1|k+1} = P^{xx}_{k+1|k} - L_{k+1} P^{yy}_{k+1|k} L^T_{k+1} \tag{30}$$

where L_{k+1} is the optimal gain given as follows:

$$L_{k+1} = P^{xy}_{k+1|k} (P^{yy}_{k+1|k})^{-1} \tag{31}$$

The algorithm of the UKF can be described as follows:

1) Step F-1: set $k = 0$. Let the mean and covariance of the initial augmented state as $\hat{x}_{0|0} = \hat{x}_0$ and $P^{xx}_{0|0} = P_0$, respectively.

2) Step F-2: calculate the means and covariances of $p_{\hat{\theta}}(x_{k+1}|Y_k)$ and $p_{\hat{\theta}}(y_{k+1}|Y_k)$ by Eqs. (22–26) using the sigma-points $\chi^{(i)}_{k|k}$, $\chi^{(i)}_{k+1|k}$, and $\eta^{(i)}_{k+1|k}(i = 0, \cdots, 2n)$ defined by Eqs. (19), (27), and (28).

3) Step F-3: calculate the optimal gain L_{k+1} by Eq. (31). Then calculate the mean $\hat{x}_{k+1|k+1}$ and covariance $P^{xx}_{k+1|k+1}$ of the augmented state by Eqs. (29) and (30).

4) Step F-4: increment k by one. If $k = N$, calculate $\chi^{(i)}_{k|N}(i = 0, \cdots, 2n)$ by Eq. (19) and exit. Otherwise, go back to step F-2.

D. Overview of URTSS

For convenience, let us define $z_k \triangleq [x_k^T \quad x_{k+1}^T]^T$. Because the sigma-points $\chi^{(i)}_{k+1|k}(i = 0, \cdots, 2n)$ given by Eq. (27) yield a joint PDF $p_{\hat{\theta}}(z_k|Y_k) = p_{\hat{\theta}}(x_{k+1}|x_k) p_{\hat{\theta}}(x_k|Y_k)$, the following equation holds:

$$p_{\hat{\theta}}(z_k|Y_k) = N\left(\begin{bmatrix} \hat{x}_{k|k} \\ \hat{x}_{k+1|k} \end{bmatrix}, \begin{bmatrix} P^{xx}_{k|k} & C_{k+1} \\ C^T_{k+1} & P^{xx}_{k+1|k} \end{bmatrix} \right) \tag{32}$$

where

$$C_{k+1} = \sum_{i=0}^{2n} W^{(i)}_{cf} (\chi^{(i)}_{k|k} - \hat{x}_{k|k})(\chi^{(i)}_{k+1|k} - \hat{x}_{k+1|k})^T \tag{33}$$

On the other hand, the PDF $p_{\hat{\theta}}(x_k|x_{k+1}, Y_N)$ can be rewritten as follows:

$$p_{\hat{\theta}}(x_k|x_{k+1}, Y_N) = p_{\hat{\theta}}(x_k|x_{k+1}, Y_k) = \frac{p_{\hat{\theta}}(z_k|Y_k)}{p_{\hat{\theta}}(x_{k+1}|Y_k)} \tag{34}$$

Let us assume $p_{\hat{\theta}}(x_{k+1}|Y_k)$ and $p_{\hat{\theta}}(x_{k+1}|Y_N)$ yield Gaussian distributions, i.e.,

$$p_{\hat{\theta}}(x_{k+1}|Y_k) = N(\hat{x}_{k+1|k}, P^{xx}_{k+1|k}) \tag{35}$$

$$p_{\hat{\theta}}(x_{k+1}|Y_N) = N(\hat{x}_{k+1|N}, P^{xx}_{k+1|N}) \tag{36}$$

Substituting Eqs. (32) and (35) into Eq. (34) yields

$$p_{\hat{\theta}}(x_k|x_{k+1}, Y_N) = N(\hat{x}_{k|k} + D_k(\hat{x}_{k+1|N} - \hat{x}_{k+1|k}), P^{xx}_{k|k} - D_k P^{xx}_{k+1|k} D^T_k) \tag{37}$$

where

$$D_k = C_{k+1}(P^{xx}_{k+1|k})^{-1} \tag{38}$$

In addition, by substituting Eqs. (36) and (37) and into the equation $p_{\hat{\theta}}(z_k|Y_N) = p_{\hat{\theta}}(x_k|x_{k+1}, Y_N) p_{\hat{\theta}}(x_{k+1}|Y_N)$, one obtains

$$p_{\hat{\theta}}(z_k|Y_k) = N(\hat{z}_{k|N}, S_{k|N}) \tag{39}$$

where

$$\hat{z}_{k|N} = \begin{bmatrix} \hat{x}_{k|k} + D_k(\hat{x}_{k+1|N} - \hat{x}_{k+1|k}) \\ \hat{x}_{k+1|N} \end{bmatrix} \tag{40}$$

$$S_{k|N} = \begin{bmatrix} P^{xx}_{k|k} + D_k(P^{xx}_{k+1|N} - P^{xx}_{k+1|k})D^T_k & D_k P^{xx}_{k+1|N} \\ P^{xx}_{k+1|N} D^T_k & P^{xx}_{k+1|N} \end{bmatrix} \tag{41}$$

Marginalizing Eq. (39) with respect to $p_{\hat{\theta}}(\boldsymbol{x}_{k+1}|\boldsymbol{Y}_N)$ then yields

$$p_{\hat{\theta}}(\boldsymbol{x}_k|\boldsymbol{Y}_N) = N(\hat{\boldsymbol{x}}_{k|N}, \boldsymbol{P}^{xx}_{k|N}) \tag{42}$$

where

$$\hat{\boldsymbol{x}}_{k|N} = \hat{\boldsymbol{x}}_{k|k} + \boldsymbol{D}_k(\hat{\boldsymbol{x}}_{k+1|N} - \hat{\boldsymbol{x}}_{k+1|k}) \tag{43}$$

$$\boldsymbol{P}^{xx}_{k|N} = \boldsymbol{P}^{xx}_{k|k} + \boldsymbol{D}_k(\boldsymbol{P}^{xx}_{k+1|N} - \boldsymbol{P}^{xx}_{k+1|k})\boldsymbol{D}_k^T \tag{44}$$

In summary, the procedure of the URTSS can be described as follows:

1) Step S-1: set $k = N - 1$. Set the mean $\hat{\boldsymbol{x}}_{N|N}$ and covariance $\boldsymbol{P}^{xx}_{N|N}$ as those obtained by the UKF.

2) Step S-2: calculate $\boldsymbol{C}_{k+1}$, $\boldsymbol{D}_k$, $\hat{\boldsymbol{x}}_{k|N}$, and $\boldsymbol{P}^{xx}_{k|N}$ by Eqs. (33), (38), (43), and (44).

3) Step S-3: exit if $k = 0$. Otherwise, decrement k by one and go back to step S-2.

E. Calculation of Cost Function

In this section, we derive a method to calculate the expectations appear in the constituent terms of the cost function given by Eqs. (16–18).

For convenience, let $\boldsymbol{f}_k$ and $\boldsymbol{h}_k$ denote $\boldsymbol{f}(\tilde{\boldsymbol{x}}_k, \boldsymbol{u}_k, \boldsymbol{\rho}_k)$ and $\boldsymbol{h}(\tilde{\boldsymbol{x}}_k, \boldsymbol{u}_k, \boldsymbol{\rho}_k)$, respectively. From $p_{\hat{\theta}}(\tilde{\boldsymbol{x}}_{k+1}|\boldsymbol{x}_k) = N(\boldsymbol{f}_k, \boldsymbol{Q})$, $\boldsymbol{\rho}_{k+1} = \boldsymbol{\rho}_k$, $p_{\hat{\theta}}(\boldsymbol{y}_k|\boldsymbol{x}_k) = N(\boldsymbol{h}_k, \boldsymbol{R})$ and Eq. (12), the following equations hold:

$$\log p_{\theta}(\boldsymbol{x}_0) = -\frac{1}{2}[\|\boldsymbol{x}_0 - \hat{\boldsymbol{x}}_0\|^2_{\boldsymbol{P}_0^{-1}} + \log\det \boldsymbol{P}_0 + n\log(2\pi)] \tag{45}$$

$$\log p_{\theta}(\boldsymbol{x}_{k+1}|\boldsymbol{x}_k) = -\frac{1}{2}[\|\tilde{\boldsymbol{x}}_{k+1} - \boldsymbol{f}_k\|^2_{\boldsymbol{Q}^{-1}} + \log\det \boldsymbol{Q} + q\log(2\pi)] \tag{46}$$

$$\log p_{\theta}(\boldsymbol{y}_k|\boldsymbol{x}_k) = -\frac{1}{2}[\|\boldsymbol{y}_k - \boldsymbol{h}_k\|^2_{\boldsymbol{R}^{-1}} + \log\det \boldsymbol{R} + p\log(2\pi)] \tag{47}$$

where $\|\boldsymbol{x}\|^2_A \triangleq \boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}$. By substituting Eqs. (45–47) into Eqs. (16–18) and manipulating them, the constituent expectation terms of the cost function, i.e., J_1, J_2, and J_3, can be expressed as follows:

$$\begin{aligned} J_1 &= \int[\|\boldsymbol{x}_0 - \hat{\boldsymbol{x}}_0\|^2_{\boldsymbol{P}_0^{-1}} + \log\det \boldsymbol{P}_0] p_{\hat{\theta}}(\boldsymbol{x}_0|\boldsymbol{Y}_N)\,\mathrm{d}\boldsymbol{x}_0 + \text{const} \\ &= \int \|\boldsymbol{x}_0 - \hat{\boldsymbol{x}}_{0|N} + \hat{\boldsymbol{x}}_{0|N} - \hat{\boldsymbol{x}}_0\|^2_{\boldsymbol{P}_0^{-1}} p_{\hat{\theta}}(\boldsymbol{x}_0|\boldsymbol{Y}_N)\,\mathrm{d}\boldsymbol{x}_0 + \log\det \boldsymbol{P}_0 \\ &\quad + \text{const} = Tr[\boldsymbol{P}_0^{-1}\{\boldsymbol{P}^{xx}_{0|N} + (\hat{\boldsymbol{x}}_{0|N} - \hat{\boldsymbol{x}}_0)(\hat{\boldsymbol{x}}_{0|N} - \hat{\boldsymbol{x}}_0)^T\}] \\ &\quad + \log\det \boldsymbol{P}_0 + \text{const} \end{aligned} \tag{48}$$

$$\begin{aligned} J_2 &= \sum_{k=0}^{N-1}\int(\|\tilde{\boldsymbol{x}}_{k+1} - \boldsymbol{f}_k\|^2_{\boldsymbol{Q}^{-1}} + \log\det \boldsymbol{Q}) p_{\hat{\theta}}(z_k|\boldsymbol{Y}_N)\,\mathrm{d}z_k + \text{const} \\ &= \sum_{k=0}^{N-1}\int \|\tilde{\boldsymbol{x}}_{k+1} - \hat{\boldsymbol{x}}^x_{k+1|N} - \boldsymbol{f}_k + \hat{\boldsymbol{f}}_{k|N} + \hat{\boldsymbol{x}}^x_{k+1|N} - \hat{\boldsymbol{f}}_{k|N}\|^2_{\boldsymbol{Q}^{-1}} \\ &\quad \times p_{\hat{\theta}}(z_k|\boldsymbol{Y}_N)\,\mathrm{d}z_k + N\log\det \boldsymbol{Q} + \text{const} \\ &= Tr\Bigg[\boldsymbol{Q}^{-1}\sum_{k=0}^{N-1}\Big\{(\boldsymbol{P}^{xx}_{k+1|N})^x + \boldsymbol{P}^{ff}_{k|N} - \boldsymbol{P}^{xf}_{k+1|N} - \boldsymbol{P}^{xfT}_{k+1|N} \\ &\quad + \boldsymbol{U}_{k+1}\Big\}\Bigg] + N\log\det \boldsymbol{Q} + \text{const} \end{aligned} \tag{49}$$

$$\begin{aligned} J_3 &= \sum_{k=0}^{N}\int(\|\boldsymbol{y}_k - \boldsymbol{h}_k\|^2_{\boldsymbol{R}^{-1}} + \log\det \boldsymbol{R}) p_{\hat{\theta}}(\boldsymbol{x}_k|\boldsymbol{Y}_N)\,\mathrm{d}\boldsymbol{x}_k + \text{const} \\ &= \sum_{k=0}^{N}\int \|\boldsymbol{y}_k - \hat{\boldsymbol{h}}_{k|N} + \hat{\boldsymbol{h}}_{k|N} - \boldsymbol{h}_k\|^2_{\boldsymbol{R}^{-1}} p_{\hat{\theta}}(\boldsymbol{x}_k|\boldsymbol{Y}_N)\,\mathrm{d}\boldsymbol{x}_k + (N \\ &\quad + 1)\log\det \boldsymbol{R} + \text{const} = Tr\Bigg[\boldsymbol{R}^{-1}\sum_{k=0}^{N}\{(\boldsymbol{y}_k - \hat{\boldsymbol{h}}_{k|N})(\boldsymbol{y}_k \\ &\quad - \hat{\boldsymbol{h}}_{k|N})^T + \boldsymbol{P}^{hh}_{k|N}\}\Bigg] + (N+1)\log\det \boldsymbol{R} + \text{const} \end{aligned} \tag{50}$$

where

$$\hat{\boldsymbol{f}}_{k|N} \triangleq \int \boldsymbol{f}_k p_{\hat{\theta}}(\boldsymbol{x}_k|\boldsymbol{Y}_N)\,\mathrm{d}\boldsymbol{x}_k \tag{51}$$

$$\boldsymbol{P}^{ff}_{k|N} \triangleq \int(\boldsymbol{f}_k - \hat{\boldsymbol{f}}_{k|N})(\boldsymbol{f}_k - \hat{\boldsymbol{f}}_{k|N})^T p_{\hat{\theta}}(z_k|\boldsymbol{Y}_N)\,\mathrm{d}z_k \tag{52}$$

$$\boldsymbol{P}^{xf}_{k+1|N} \triangleq \int(\tilde{\boldsymbol{x}}_{k+1} - \hat{\boldsymbol{x}}^x_{k+1|N})(\boldsymbol{f}_k - \hat{\boldsymbol{f}}_{k|N})^T p_{\hat{\theta}}(z_k|\boldsymbol{Y}_N)\,\mathrm{d}z_k \tag{53}$$

$$\hat{\boldsymbol{h}}_{k|N} \triangleq \int \boldsymbol{h}_k p_{\hat{\theta}}(\boldsymbol{x}_k|\boldsymbol{Y}_N)\,\mathrm{d}\boldsymbol{x}_k \tag{54}$$

$$\boldsymbol{P}^{hh}_{k|N} \triangleq \int(\boldsymbol{h}_k - \hat{\boldsymbol{h}}_{k|N})(\boldsymbol{h}_k - \hat{\boldsymbol{h}}_{k|N})^T p_{\hat{\theta}}(\boldsymbol{x}_k|\boldsymbol{Y}_N)\,\mathrm{d}\boldsymbol{x}_k \tag{55}$$

$$\boldsymbol{U}_{k+1} \triangleq (\hat{\boldsymbol{x}}^x_{k+1|N} - \hat{\boldsymbol{f}}_{k|N})(\hat{\boldsymbol{x}}^x_{k+1|N} - \hat{\boldsymbol{f}}_{k|N})^T \tag{56}$$

In addition, $\hat{\boldsymbol{x}}^x_{k+1|N} \in \mathbb{R}^q$ and $(\boldsymbol{P}^{xx}_{k+1|N})^x \in \mathbb{R}^{q\times q}$ denote the part of the state (i.e., the nonaugmented state) in the mean and covariance of the augmented state $\hat{\boldsymbol{x}}_{k+1|N}$, respectively. To calculate the statistics given by Eqs. (51–55), an unscented transform based on $p_{\hat{\theta}}(z_k|\boldsymbol{Y}_N)$ given by Eq. (39) is used. That is, sigma-points $\boldsymbol{\zeta}^{(i)}_{k|N} \in \mathbb{R}^n$, $\boldsymbol{\xi}^{(i)}_{k+1|N} \in \mathbb{R}^n$ $(i = 0, \cdots, 4n)$ and their associated weights $W^{(i)}_{ms}$, $W^{(i)}_{cs}$, $(i = 0, \cdots, 4n)$ defined next are adopted:

$$\begin{bmatrix} \boldsymbol{\zeta}^{(i)}_{k|N} \\ \boldsymbol{\xi}^{(i)}_{k+1|N} \end{bmatrix} \triangleq \begin{cases} \hat{z}_{k|N} & : i = 0 \\ \hat{z}_{k|N} + \alpha_s\sqrt{2n+\kappa_s}(\boldsymbol{S}_{k|N})^{1/2}_i & : 1 \le i \le 2n \\ \hat{z}_{k|N} - \alpha_s\sqrt{2n+\kappa_s}(\boldsymbol{S}_{k|N})^{1/2}_{i-2n} & : 2n+1 \le i \le 4n \end{cases} \tag{57}$$

$$W^{(i)}_{ms} \triangleq \begin{cases} 1 - 2n/[\alpha_s^2(2n+\kappa_s)] & : i = 0, \\ 1/[2\alpha_s^2(2n+\kappa_s)] & : 1 \le i \le 4n \end{cases} \tag{58}$$

$$W^{(i)}_{cs} \triangleq \begin{cases} W^{(i)}_{ms} + 1 - \alpha_s^2 + \beta_s & : i = 0, \\ W^{(i)}_{ms} & : 1 \le i \le 4n \end{cases} \tag{59}$$

where α_s, β_s, and κ_s are tuning parameters to be set similarly as α_f, β_f, and κ_f. It should be noted that $\boldsymbol{\zeta}^{(i)}_{k|N}(i = 0, \cdots, 4n)$ yields not only $p_{\hat{\theta}}(z_k|\boldsymbol{Y}_N)$ but also $p_{\hat{\theta}}(\boldsymbol{x}_k|\boldsymbol{Y}_N)$ given by Eq. (42). Furthermore, by dividing the sigma-points in Eq. (57) as $\boldsymbol{\zeta}^{(i)}_{k|N} = [\boldsymbol{\zeta}^{x(i)}_{k|N} \quad \boldsymbol{\zeta}^{\rho(i)}_{k|N}]^T$ $(\boldsymbol{\zeta}^{x(i)}_{k|N} \in \mathbb{R}^q \quad \boldsymbol{\zeta}^{\rho(i)}_{k|N} \in \mathbb{R}^l)$ and $\boldsymbol{\xi}^{(i)}_{k+1|N} = [\boldsymbol{\xi}^{x(i)}_{k+1|N} \quad \boldsymbol{\xi}^{\rho(i)}_{k+1|N}]^T$ $(\boldsymbol{\xi}^{x(i)}_{k+1|N} \in \mathbb{R}^q, \quad \boldsymbol{\xi}^{\rho(i)}_{k+1|N} \in \mathbb{R}^l)$), and applying the unscented transform to Eqs. (51–55), the statistics can be approximated as follows:

$$\hat{\boldsymbol{f}}_{k|N} \cong \sum_{i=0}^{4n} W^{(i)}_{ms}\boldsymbol{f}(\boldsymbol{\zeta}^{x(i)}_{k|N}, \boldsymbol{u}_k, \boldsymbol{\zeta}^{\rho(i)}_{k|N}) \tag{60}$$

$$P^{ff}_{k|N} \cong \sum_{i=0}^{4n} W^{(i)}_{cs}[f(\zeta^{x(i)}_{k|N}, u_k, \zeta^{\rho(i)}_{k|N}) - \hat{f}_{k|N}][f(\zeta^{x(i)}_{k|N}, u_k, \zeta^{\rho(i)}_{k|N}) - \hat{f}_{k|N}]^T \quad (61)$$

$$P^{xf}_{k+1|N} \cong \sum_{i=0}^{4n} W^{(i)}_{cs}(\xi^{x(i)}_{k+1|N} - \hat{x}^x_{k+1|N})[f(\zeta^{x(i)}_{k|N}, u_k, \zeta^{\rho(i)}_{k|N}) - \hat{f}_{k|N}]^T \quad (62)$$

$$\hat{h}_{k|N} \cong \sum_{i=0}^{4n} W^{(i)}_{ms} h(\zeta^{x(i)}_{k|N}, u_k, \zeta^{\rho(i)}_{k|N}) \quad (63)$$

$$P^{hh}_{k|N} \cong \sum_{i=0}^{4n} W^{(i)}_{cs}[h(\zeta^{x(i)}_{k|N}, u_k, \zeta^{\rho(i)}_{k|N}) - \hat{h}_{k|N}][h(\zeta^{x(i)}_{k|N}, u_k, \zeta^{\rho(i)}_{k|N}) - \hat{h}_{k|N}]^T \quad (64)$$

It should be noted that the weights $W^{(i)}_{ms}(i = 0, \cdots, 4n)$ are used to calculate the means $\hat{f}_{k|N}$, $\hat{h}_{k|N}$, and the weights $W^{(i)}_{cs}(i = 0, \cdots, 4n)$ are used to calculate the covariances $P^{ff}_{k|N}$, $P^{xf}_{k+1|N}$, and $P^{hh}_{k|N}$. Unlike the particle methods [21,22], which requires many sample points to attain accurate approximation of the expectations, the unscented transform is able to accurately approximate the expectations with a reasonable number (i.e., $4n + 1$) of sigma-points.

F. Minimization of the Cost Function

Now let us consider the optimal θ (i.e., the collection of the elements of $\hat{x}_0, P_0, Q$, and R) minimizing the cost function $J(\theta, \hat{\theta}_t)$ at each M-step in the EM. To derive it, we use the following theorem:

Theorem 1: let us define the following function:

$$\phi(Z) = Tr(Z^{-1}A) + \log\det Z \quad (65)$$

where Z and A are positive definite symmetric matrices. The minimizer of the preceding function is given as follows:

$$Z = A \quad (66)$$

The proof is given in Appendix B.

$\hat{x}_0$ and P_0 appears only in the term of J_1 among the constituent terms of the cost function. From Eq. (48), the optimal $\hat{x}_0$ is trivially given as

$$\hat{x}_0 = \hat{x}_{0|N} \quad (67)$$

By substituting Eq. (67) to Eq. (48), the term J_1 can be expressed as follows:

$$J_1 = Tr(P_0^{-1}P^{xx}_{0|N}) + \log\det P_0 + \text{const} \quad (68)$$

Then, by applying Theorem 1, the optimal P_0 is derived as follows:

$$P_0 = P^{xx}_{0|N} \quad (69)$$

Furthermore, Q and R appear only in J_2 and J_3, respectively. By applying Theorem 1 again to Eqs. (49) and (50) and, the optimal Q and R are derived as follows:

$$Q = \frac{1}{N}\sum_{k=0}^{N-1}[(P^{xx}_{k+1|N})^x + P^{ff}_{k|N} - P^{xf}_{k+1|N} - P^{xf}_{k+1|N} + U_{k+1}] \quad (70)$$

$$R = \frac{1}{N+1}\sum_{k=0}^{N}[(y_k - \hat{h}_{k|N})(y_k - \hat{h}_{k|N})^T + P^{hh}_{k|N}] \quad (71)$$

Consequently, in the M-step of the EM, the statistics θ is updated by Eqs. (67) and (69–71). In the sense that the optimal θ at each M-step can readily be calculated in the analytical form as in Eqs. (67) and (69–71), the developed method is advantageous to the other smoothing approaches [13,14,21], which require numerical optimization algorithms for the resulting nonconvex optimization problem.

G. Summary of the Developed Method

The algorithm of the developed method can be summarized as follows:

1) Step 1: give initial estimates of $\hat{x}_0$, P_0, Q, and R.
2) Step 2: perform the UKF algorithm (step F-1 to step F-4).
3) Step 3: perform the URTSS algorithm (step S-1 to step S-3) with an insertion of the following step between step S-2 and step S-3.
4) Step S-2': calculate $\hat{z}_{k|N}$ and $S_{k|N}$ from Eqs. (40) and (41). Subsequently calculate the sigma-points $\zeta^{(i)}_{k|N}$, $\xi^{(i)}_{k+1|N}$ $(i = 0, \cdots, 4n)$ from Eq. (57). Then calculate $\hat{f}_{k|N}$, $P^{ff}_{k|N}$, $\hat{h}_{k|N}$, $P^{hh}_{k|N}$, and U_{k+1} by Eqs. (56) and (60–64).
5) Step 4: update $\hat{x}_0$, P_0, Q, and R by Eqs. (67) and (69–71).
6) Step 5: calculate the estimate of the parameter $\hat{\rho}$ by

$$\hat{\rho} = \left(\sum_{k=0}^{N} x_{\{k|N\}^{\rho}}\right)/(N+1) \quad (72)$$

7) Step 6: if either the number of iterations reaches the specified value or the changes of the estimates of $\hat{x}_0$, P_0, Q, and R are all small, then terminate the algorithm. Otherwise, go back to step 2.

III. Numerical Examples

A. Estimation of Aerodynamic Parameters of Linearized Aircraft Dynamics

Let us first consider a problem of estimating lateral-directional aerodynamic parameters of a linearized dynamics model of an aircraft. The model is postulated as follows:

1) State equations:

$$p_{k+1} = p_k + \int_{t_k}^{t_{k+1}}(L_p p + L_r r + L_{\delta a}\delta_a + L_{\delta r}\delta_r + L_\beta \beta + L_0)\,dt + w_{k(1)} \quad (73)$$

$$r_{k+1} = r_k + \int_{t_k}^{t_{k+1}}(N_p p + N_r r + N_{\delta a}\delta_a + N_{\delta r}\delta_r + N_\beta \beta + N_0)\,dt + w_{k(2)} \quad (74)$$

where t is the time (s), p is the roll rate (rad/s), r is the yaw rate (rad/s), δ_a is the aileron deflection angle (rad), δ_r is the rudder deflection angle (rad), and β is the sideslip angle (rad), and $w_{k(i)}(i = 1, 2)$ is the process noise.

2) Observation equations:

$$p_{mk} = p_k + v_{k(1)} \quad (75)$$

$$r_{mk} = r_k + v_{k(2)} \quad (76)$$

where the subscript m denotes the measured output, and $v_{k(i)}(i = 1, 2)$ is the measurement noise.

The sideslip angle β is treated as one of the measured inputs, and hence its dynamics is not considered as a state equation. This simplified model is introduced in [3]. The state variable vector is $\tilde{x}_k = [\,p_k \quad q_k\,]^T$, the measured input vector is $u_k = [\,\delta_{ak} \quad \delta_{rk} \quad \beta_k\,]^T$, the measured output vector is $y_k = [\,p_{mk} \quad r_{mk}\,]^T$, and the unknown parameter vector is $\rho = [\,L_p \quad L_r \quad L_{\delta a} \quad L_{\delta r} \quad L_\beta L_0 N_p N_r N_{\delta a} N_{\delta r} N_\beta N_0\,]^T$. The integral terms in Eqs. (73) and (74) were numerically calculated by the fourth-order Runge–Kutta scheme with a time step of 0.02 s. It should be noted that the state equations are nonlinear with respect to the augmented state $x_k = [\,\tilde{x}_k^T \quad \rho_k^T\,]^T$. We artificially generated a flight test data via numerical simulation using the time histories of the inputs and reference values of the aerodynamic parameters taken from [3]. The true initial state and covariances were defined as follows:

$$p_0 = 9.661 \times 10^{-3}\ [\mathrm{rad/s}], \qquad r_0 = 1.434 \times 10^{-4}\ [\mathrm{rad/s}] \quad (77)$$

$$\boldsymbol{P}_0 = \boldsymbol{0}_{14\times 14}, \qquad \boldsymbol{Q} = 10^{-8}\boldsymbol{I}_2, \qquad \boldsymbol{R} = 10^{-6}\boldsymbol{I}_2 \quad (78)$$

Figure 1 shows the time histories of the inputs and the true state variables generated by the simulation.

The developed method was applied with initial estimates of $\hat{\boldsymbol{x}}_0 = [\,1.153\times 10^{-2}\ 3.049\times 10^{-2}\quad \boldsymbol{0}_{1\times 12}\,]^T$, $\boldsymbol{P}_0 = \boldsymbol{I}_{14}$, $\boldsymbol{Q} = \boldsymbol{I}_2$, and $\boldsymbol{R} = \boldsymbol{I}_2$. The first two elements of the initial estimate of $\hat{\boldsymbol{x}}_0$ are the same as the measured outputs at the initial time. The tuning parameters of the unscented transform were specified as $\alpha_f = \alpha_s = 10^{-2}$, $\beta_f = \beta_s = 2$, $\kappa_f = 1$, and $\kappa_s = -1$. The maximum number of iteration was specified as 10^4, and for simplicity reaching it was the only condition to terminate the algorithm. For comparison, the methods of [20] (the linear smoother with EM) and [21] (the particle smoother with EM) were also applied to this problem. Since the method of [20] estimates the system matrices of discrete linear systems, i.e., matrices $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$, and $\boldsymbol{D}$ in $\tilde{\boldsymbol{x}}_{k+1} = \boldsymbol{A}\tilde{\boldsymbol{x}}_k + \boldsymbol{B}\boldsymbol{u}_k$ and $\boldsymbol{y}_k = \boldsymbol{C}\tilde{x}_k + \boldsymbol{D}\boldsymbol{u}_k$, we modified it such that the matrices $\boldsymbol{C}$ and $\boldsymbol{D}$ were mandatorily updated to identity matrix and zero matrix, respectively, for the uniqueness of the parameters to be estimated. In addition, once the estimated matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ were obtained, they were transformed into the matrices for continuous system by the bilinear approximation. Moreover, we modified the method of [21] such that the Powell's direction set method [24] was used as the optimization algorithm, because, in our experience, it was more robust than the original gradient-based algorithm. The number of particles was specified as 100. In each method, the maximum number of iterations and the initial estimates were the same as the developed method.

Table 1 shows the true values and the estimates of unknown parameters obtained by the three methods. As can be seen, the estimates obtained by the developed method agreed best with the true values. The reason for the relatively inaccurate estimates by the method of [20] may be the modification of the mandatory update of the matrices $\boldsymbol{C}$ and $\boldsymbol{D}$, because the combination of these mandatorily updated matrices and the matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ produced by each M-step does not necessarily maximize the expectation. Nevertheless, without this modification, we obtained less accurate estimates. On the other hand, the method of [21] stopped after 14 iterations and got trapped in a poor local optimum. The reason for this may be the insufficiency of the number of particles as well as the poor initial estimates. However, the specified number of particles was larger than that of the sigma-points used in the developed algorithm. In addition, the method of [20] and the developed method started from the same

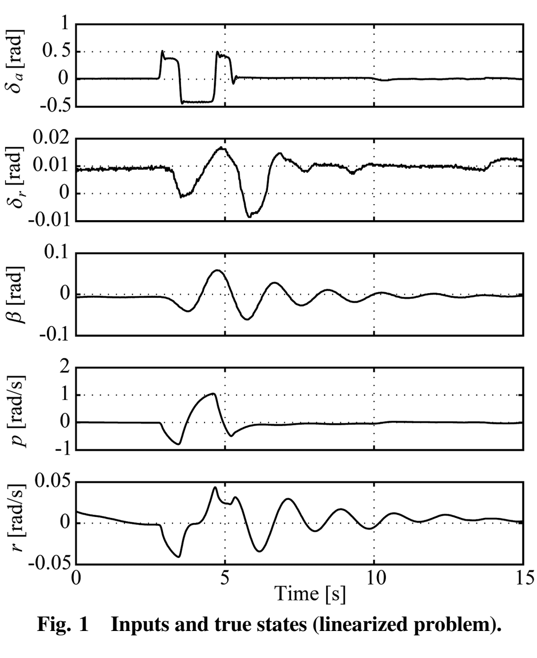

Fig. 1 Inputs and true states (linearized problem).

Table 1 Estimated parameters (linearized problem)

Parameter	True value	Developed method	Method of [20]	Method of [21]
L_p, 1/s	−2.072	−2.075	−2.080	−1.648
L_r, 1/s	1.019	1.045	1.074	1.755
$L_{\delta a}$, 1/s^2	−6.334	−6.336	−6.491	−5.722
$L_{\delta r}$, 1/s^2	1.195	1.160	1.560	2.750
L_β, 1/s^2	−3.800	−3.779	−3.694	−7.385
L_0, rad/s^2	0.010	0.011	0.0090	1.4×10^{-4}
N_p, 1/s	−0.170	−0.168	−0.149	−2.836
N_r, 1/s	−0.425	−0.419	−1.081	−79.628
$N_{\delta a}$, 1/s^2	−0.378	−0.376	−0.360	−2.929
$N_{\delta r}$, 1/s^2	−1.719	−1.608	−0.0763	1.629
N_β, 1/s^2	3.031	2.984	2.852	0.594
N_0, rad/s^2	0.034	0.033	0.021	−0.995

initial estimates. Using a computer equipped with Intel Core 2 2.4 GHz and 4.0 GB RAM, the computational times were 1.071×10^3 s for the developed method, 76.44 s for the method of [20], and 4.454×10^4 [s] for the method of [21], respectively. Although the computational time for the developed method is longer than that of the method of [20] because it employs the unscented transform to cover nonlinear functions, it is much shorter than that for the method of [21]. The long computational time for the method of [21] was because the calculations of the expectation were repeated many times in an iteration of the optimization algorithm.

Figures 2–4 show the estimation profiles of the parameters, the statistics of the initial state, and the noise covariances as functions of the iteration number. As can be seen, the estimated parameters produced by the first iteration were very poor. This was presumably because of the large uncertainty in the estimates of the statistics of the initial state and noise covariances. As these statistics were updated to more accurate values by the iterations, the agreement between the estimated parameters and the true values became better. The parameters and the statistics except for the covariance of the initial state almost converged after the first 100 iterations. The noise covariances produced by the final iteration ($=10^4$-th iteration) were as follows:

$$\boldsymbol{Q} = \begin{bmatrix} 6.651\times 10^{-9} & 1.830\times 10^{-9} \\ 1.830\times 10^{-9} & 6.415\times 10^{-9} \end{bmatrix}, \quad \boldsymbol{R} = \begin{bmatrix} 9.285\times 10^{-7} & -3.356\times 10^{-8} \\ -3.356\times 10^{-8} & 1.224\times 10^{-6} \end{bmatrix} \quad (79)$$

It can be confirmed from Figs. 3 and 4 and Eq. (79) that the converged statistics agreed well with their true values as a whole. In particular, an acceptable agreement was observed in the process noise covariance $\boldsymbol{Q}$, which has often been difficult to estimate relative to the other statistics [3].

The errors between the estimated and true state variables are shown in Fig. 5, wherein $\hat{\boldsymbol{x}}_{k|N}(k = 0, \cdots, N)$ obtained at the final iteration are treated as the estimated state variables (the subscript e in this figure denotes the error). It can be confirmed that these errors are sufficiently small relative to the absolute values of the true state, i.e., the estimated state agreed well with the true state.

B. Estimation of Longitudinal Aerodynamic Parameters of Nonlinear Aircraft Dynamics

As a practical application, let us consider a problem of estimating longitudinal aerodynamic parameters of a nonlinear dynamic model of the research aircraft HFB-320. An actual flight test data was published in [3]. Figure 6 shows the time history of the elevator deflection angle δ_e (rad) and the thrust T (N) measured in the flight test. The longitudinal motion was excited through multistep elevator input, resulting in short-period motion, and a pulse input leading to phugoid motion.

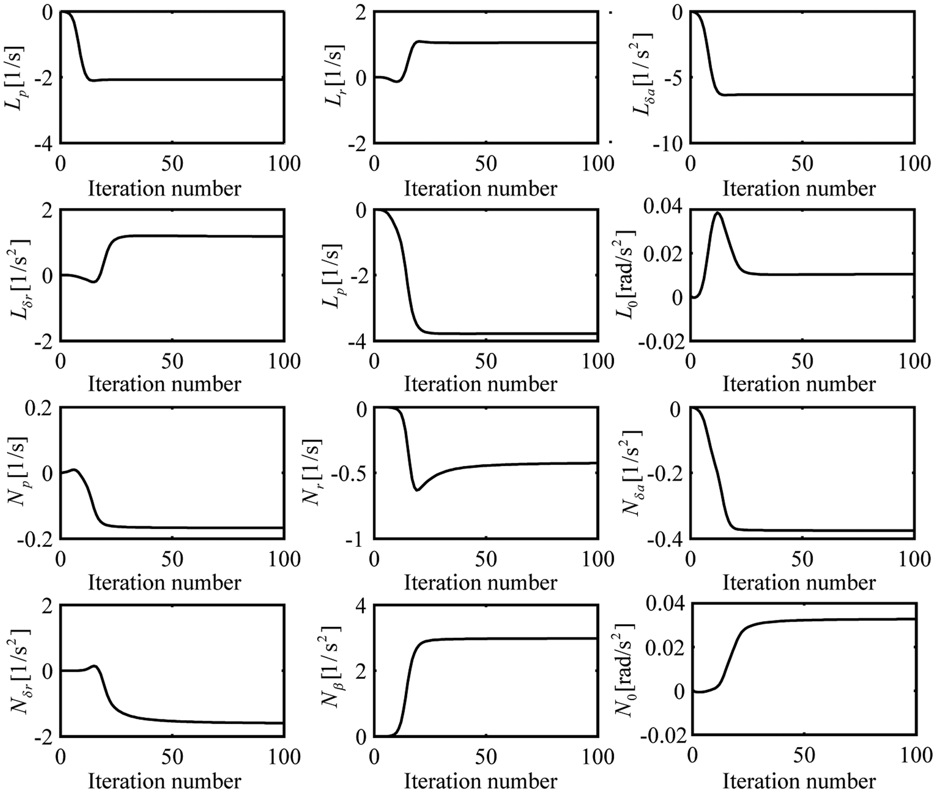

Fig. 2 Parameters as functions of iteration number (linearized problem).

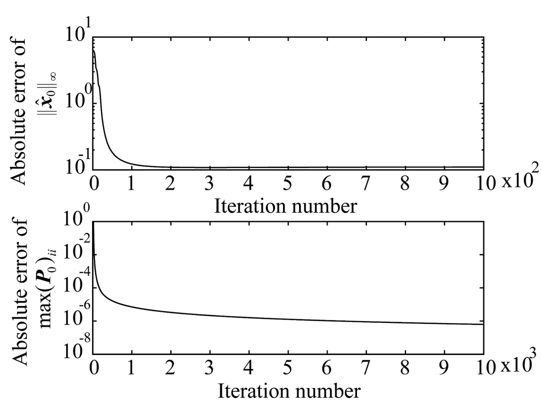

Fig. 3 Statistics of initial states as functions of iteration number (linearized problem).

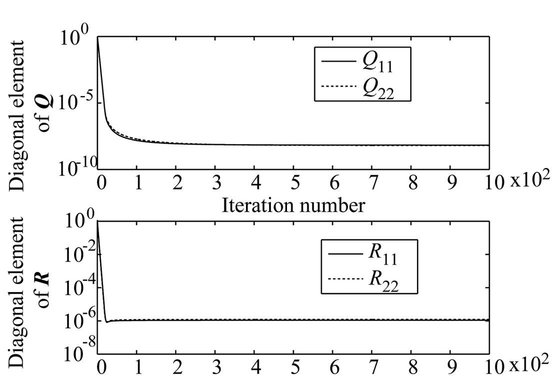

Fig. 4 Noise covariances as functions of iteration number (linearized problem).

The following model has been postulated:

1) State equations:

$$V_{k+1} = V_k + \int_{t_k}^{t_{k+1}} [\{T\cos(\alpha + \sigma_T) - D\}/m + g\sin(\alpha - \theta)]\,\mathrm{d}t + w_{k(1)} \tag{80}$$

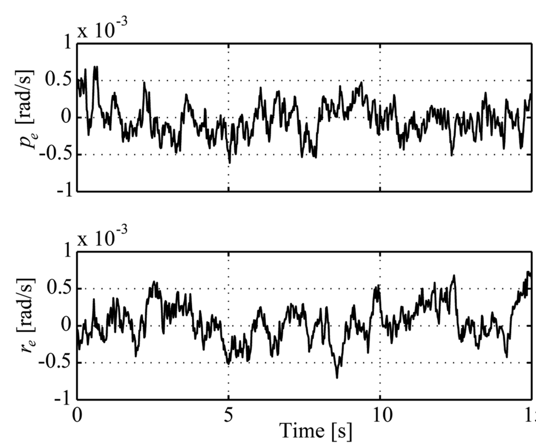

Fig. 5 Errors between the estimated and true states (linearized problem).

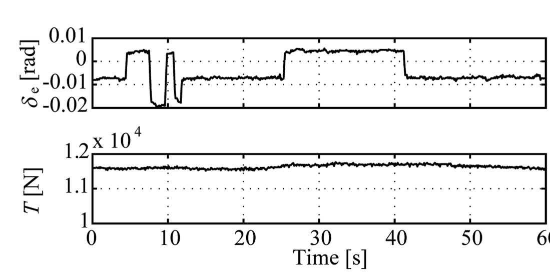

Fig. 6 Measured inputs (nonlinear problem).

Table 2 Constant parameters in the nonlinear problem

Parameter	Value
Air density $\rho_{\rm air}$	$0.7920\ \mathrm{kg/m^3}$
Gravity acceleration g	$9.80665\ \mathrm{m/s^2}$
Mass m	7472.0 kg
Moment of inertia I_{yy}	$2.2847 \times 10^4\ \mathrm{kg\,m^2}$
Reference area S	$30.097\ \mathrm{m^2}$
Mean aerodynamic chord $\bar{c}$	0.6075 m
Reference velocity $V_{\rm ref}$	104.67 m/s
Thrust inclination angle σ_T	0.0525 rad
Moment arm of the thrust $l_{tx}\sin\sigma_T + l_{tz}\sin\sigma_T$	−0.1603 m

$$\alpha_{k+1} = \alpha_k + \int_{t_k}^{t_{k+1}} [q - \{L + T\sin(\alpha + \sigma_T)\}/(mV) + g\cos(\alpha - \theta)/V]\,\mathrm{d}t + w_{k(2)} \tag{81}$$

$$\theta_{k+1} = \theta_k + \int_{t_k}^{t_{k+1}} q\,\mathrm{d}t + w_{k(3)} \tag{82}$$

$$q_{k+1} = q_k + \int_{t_k}^{t_{k+1}} [M + T(l_{tx}\sin\sigma_T + l_{tz}\sin\sigma_T)]/I_{yy}\,\mathrm{d}t + w_{k(4)} \tag{83}$$

where t is the time (s), V is the velocity (m/s), α is the angle of attack (rad), θ is the pitch angle (rad), q is the pitch rate (rad/s), D is the drag (N), L is the lift (N), M is the pitching moment $(\mathrm{N}\cdot\mathrm{m})$, and $w_{k(i)} (i = 1, \cdots, 4)$ is the process noise. Table 2 summarizes the constant parameters. In addition, D, L, and M are expressed as

$$D = q_d S[C_{D0} + C_{DV}(V/V_{\rm ref}) + C_{D\alpha}\alpha] \tag{84}$$

$$L = q_d S[C_{L0} + C_{LV}(V/V_{\rm ref}) + C_{L\alpha}\alpha] \tag{85}$$

$$M = q_d S\bar{c}[C_{m0} + C_{mV}(V/V_{\rm ref}) + C_{m\alpha}\alpha + C_{mq}(q\bar{c})/(2V_{\rm ref}) + C_{m\delta e}\delta_e] \tag{86}$$

where $q_d \triangleq \rho_{air}V^2/2$.

2) Observation equations:

$$V_{mk} = V_k + v_{k(1)} \tag{87}$$

$$\alpha_{mk} = \alpha_k + v_{k(2)} \tag{88}$$

$$\theta_{mk} = \theta_k + v_{k(3)} \tag{89}$$

$$q_{mk} = q_k + v_{k(4)} \tag{90}$$

$$\dot{q}_{mk} = [M_k + T_k(l_{tx}\sin\sigma_T + l_{tz}\cos\sigma_T)]/I_{yy} + v_{k(5)} \tag{91}$$

$$a_{xmk} = (L_k\sin\alpha_k - D_k\cos\alpha_k + T_k\cos\sigma_T)/m + v_{k(6)} \tag{92}$$

$$a_{zmk} = (-L_k\cos\alpha_k - D_k\sin\alpha_k + T_k\sin\sigma_T)/m + v_{k(7)} \tag{93}$$

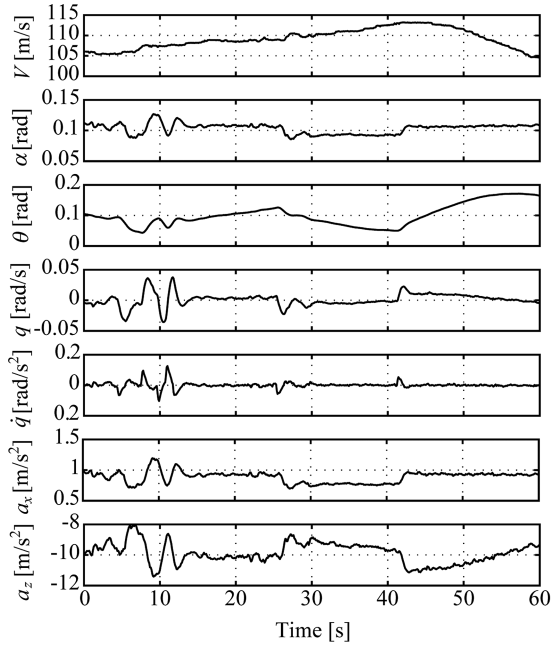

Fig. 7 Estimated outputs (nonlinear problem).

where the subscript m denotes the measured output, a_x is the acceleration along the X-axis $(\mathrm{m/s^2})$, a_z is the acceleration along the Z-axis $(\mathrm{m/s^2})$, and $v_{k(i)} (i = 1, \cdots, 7)$ is the measurement noise.

The state variable vector is $\tilde{\boldsymbol{x}}_k = [V_k \quad \alpha_k \quad \theta_k \quad q_k]^T$, the measured input vector is $\boldsymbol{u}_k = [\delta_{ek} \quad T_k]^T$, the measured output vector is $\boldsymbol{y}_k = [V_{mk} \quad \alpha_{mk} \quad \theta_{mk} \quad q_{mk} \quad \dot{q}_{mk} \quad a_{xm(k)} \quad a_{zm(k)}]^T$, and the unknown parameter vector is $\boldsymbol{\rho} = [C_{D0} \quad C_{DV} \quad C_{D\alpha} \quad C_{L0} \quad C_{LV} \quad C_{L\alpha} \quad C_{m0} \quad C_{mV} \quad C_{m\alpha} \quad C_{mq} C_{m\delta e}]^T$. The measurements were performed over 60 s at a frequency of 10 Hz. The integral terms in Eqs. (80–83) were numerically calculated by the fourth-order Runge–Kutta scheme with a time step of 0.1 s.

The developed method was applied with initial estimates of $\hat{\boldsymbol{x}}_0 = [106.3 \quad 0.1117 \quad 0.1054 \quad -3.440 \times 10^{-3} \quad \boldsymbol{0}_{11\times 1}]^T$, $\boldsymbol{P}_0 = \boldsymbol{I}_{15}$, $\boldsymbol{Q} = \boldsymbol{I}_4$, and $\boldsymbol{R} = \boldsymbol{I}_7$. The first four elements of the estimate of

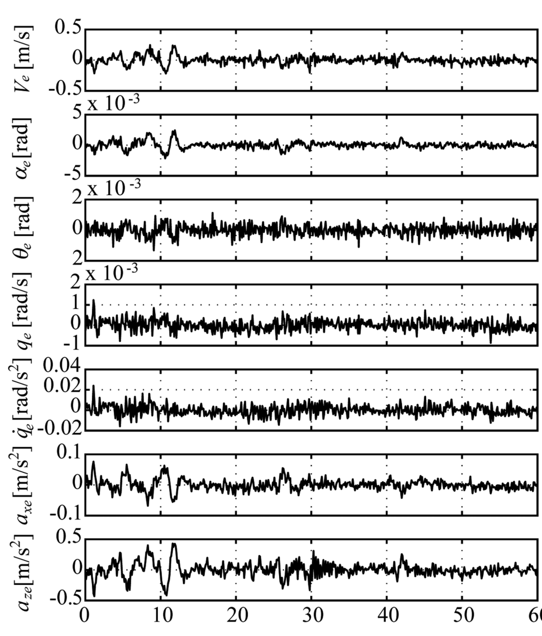

Fig. 8 Deviation of the estimated outputs from the measured outputs (nonlinear problem).

Table 3 Estimated parameters (nonlinear problem)

Parameter	Reference	Estimate by developed method
C_{D0}	0.123	0.123
C_{DV}	−0.0645	−0.0651
$C_{D\alpha}$	0.320	0.321
C_{L0}	−0.0929	−0.0911
C_{LV}	0.149	0.149
$C_{L\alpha}$	4.328	4.308
C_{m0}	0.112	0.107
C_{mV}	0.0039	0.0116
$C_{m\alpha}$	−0.968	−1.000
C_{mq}	−34.710	−38.634
$C_{m\delta e}$	−1.529	−1.616

$\hat{\boldsymbol{x}}_0$ are the same as the measured outputs at the initial time. The tuning parameters of the unscented transform were specified as $\alpha_f = \alpha_s = 10^{-2}$, $\beta_f = \beta_s = 2$, $\kappa_f = -1$, and $\kappa_s = -5$. As is the case with the previous example, the maximum number of iterations was specified as 10^4, and reaching it was the only condition to terminate the algorithm.

The estimated outputs $\hat{\boldsymbol{y}}_{k|N} (k = 0, \cdots, N)$ and the deviations of them from the measured outputs are shown in Figs. 7 and 8, respectively. As can be seen in these figures, the deviations are sufficiently small relative to the absolute values of the estimated outputs, i.e., the estimated outputs agreed well with the measured outputs.

Table 3 shows the estimated values of the unknown parameters. The values in the column labeled "Reference" in this table are the successful estimation results by the FEM presented in [3]. It should be noted that in the calculation of the FEM, the initial values of the unknown parameters were specified at the vicinity of the estimated optimum. On the other hand, arbitrarily determined initial estimates were used for the developed method, but the final estimates produced by it show good agreement with the references obtained successfully by the FEM. We tried to perform the estimation by the FEM and the method of [21] with 100 particles starting from the same initial estimates as those in the developed method, but they failed. These results indicate the convergence robustness of the developed method.

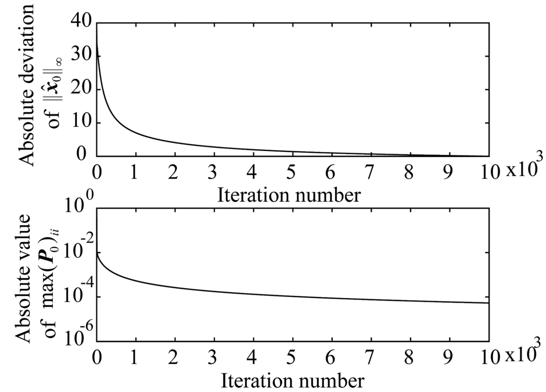

Fig. 10 Statistics of initial states as functions of iteration number (nonlinear problem).

Figures 9–11 show the estimation profiles of the parameters, the statistics of the initial state, and the noise covariances as functions of the iteration number. "Absolute deviation of $\|\hat{\boldsymbol{x}}_0\|_\infty$" in Fig. 10 means the absolute deviation between $\|\hat{\boldsymbol{x}}_0\|_\infty$ at the current iteration and that at the final iteration. The parameters produced by the first iteration were far from the reference values presumably due to the large uncertainty in the estimates of the statistics. As these statistics converged to certain values by the iterations, the agreement between the estimated parameters and the reference values became better. This result indicates that the statistics significantly affected the accuracy of the estimated parameters also in this problem, and the developed method achieved accurate estimation of the parameters by combining the estimation of the statistics via the EM.

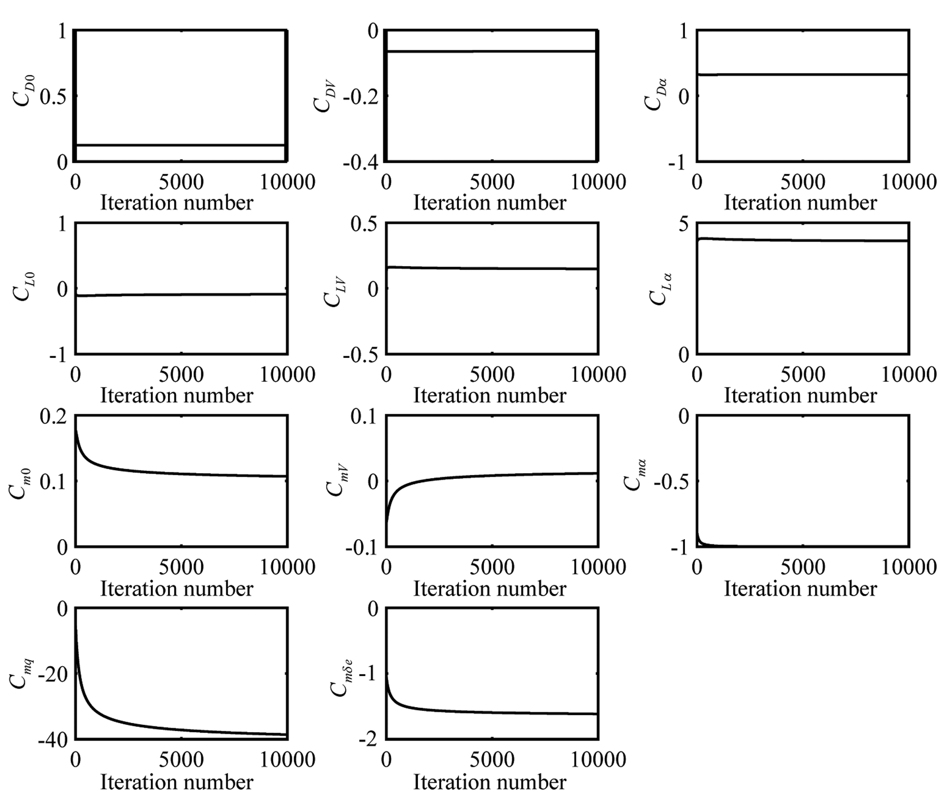

Fig. 9 Parameters as functions of iteration number (nonlinear problem).

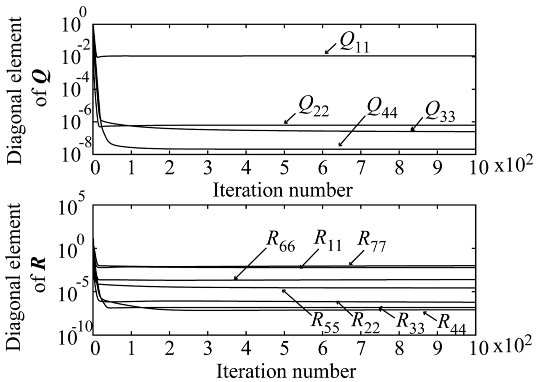

Fig. 11 Noise covariances as functions of iteration number (nonlinear problem).

IV. Conclusions

This paper proposed a new method for parameter estimation of aircraft's dynamics in state space using the URTSS combined with the EM algorithm. The developed method calculates the expectation in the EM by an unscented transform, and analytically finds the maximizer of the expectation. These properties enable the developed method to achieve low computational cost and high robustness. By applying the developed method to the aerodynamic parameters estimation problems and comparing with the existing methods, the accuracy, robustness, and efficiency of the developed method were confirmed.

Possible future work includes extending the algorithm to be applicable to the multiple flight test data.

Appendix A: Treatment of Nonadditive Process Noise

Let us consider the state equation in the following form:

$$\tilde{\boldsymbol{x}}_{k+1} = \boldsymbol{f}(\tilde{\boldsymbol{x}}_k, \boldsymbol{\omega}_k, \boldsymbol{u}_k, \boldsymbol{\rho}) + \boldsymbol{w}_k \tag{A1}$$

$$\begin{bmatrix} \boldsymbol{w}_k \\ \boldsymbol{\omega}_k \end{bmatrix} \sim N(\boldsymbol{0}, \boldsymbol{Q}) \tag{A2}$$

where $\boldsymbol{w}_k \in \mathbb{R}^q$ and $\boldsymbol{\omega}_k \in \mathbb{R}^r$ are the additive process noise and the nonadditive process noise, respectively. Although the state equation defined preceding is more general than those defined by Eq. (7), by defining a new state as $\tilde{\boldsymbol{x}}_k^a \triangleq [\tilde{\boldsymbol{x}}_k^T \quad \boldsymbol{\omega}_k^T]^T$, Eq. (A1) can be rewritten in the form of Eq. (7) as follows:

$$\tilde{\boldsymbol{x}}_{k+1}^a = \begin{bmatrix} \tilde{\boldsymbol{x}}_{k+1} \\ \boldsymbol{\omega}_{k+1} \end{bmatrix} = \boldsymbol{f}^a(\tilde{\boldsymbol{x}}_k^a, \boldsymbol{u}_k, \boldsymbol{\rho}) + \boldsymbol{w}_k^a \tag{A3}$$

where

$$\boldsymbol{f}^a(\tilde{\boldsymbol{x}}_k^a, \boldsymbol{u}_k, \boldsymbol{\rho}) \triangleq \begin{bmatrix} \boldsymbol{f}(\tilde{\boldsymbol{x}}_k, \boldsymbol{\omega}_k, \boldsymbol{u}_k, \boldsymbol{\rho}) \\ 0 \end{bmatrix}, \qquad \boldsymbol{w}_k^a \sim N(\boldsymbol{0}, \boldsymbol{Q}) \tag{A4}$$

Because the initial nonadditive noise $\boldsymbol{\omega}_0$ is a part of $\boldsymbol{x}_0^a$, its PDF is included in that of the augmented state $\boldsymbol{x}_0 = [(\boldsymbol{x}_0^a)^T \quad \boldsymbol{\rho}_0^T]^T$, i.e., Eq. (12).

Appendix B: Proof of Theorem 1

This appendix gives a proof of Theorem 1. Let us assume $\boldsymbol{Z}$ is a positive definite symmetric matrix in $\mathbb{R}^{a\times a}$, and define S_+^a as a cone of positive definite symmetric matrices in $\mathbb{R}^{a\times a}$. Then S_+^a includes a matrix $\boldsymbol{S} \triangleq \boldsymbol{Z}^{-1}$. The problem of finding $\boldsymbol{Z}$ which minimizes $\phi(\boldsymbol{Z})$ is equivalent to finding $\boldsymbol{S}$ which minimizes the following equation:

$$\psi(\boldsymbol{S}) = Tr(\boldsymbol{SA}) - \log\det \boldsymbol{S} \tag{B1}$$

Differentiation of Eq. (B1) with respect to $\boldsymbol{S}$ is

$$\frac{\partial \psi(\boldsymbol{S})}{\partial \boldsymbol{S}} = \boldsymbol{A} - \boldsymbol{S}^{-1} \tag{B2}$$

Therefore, $\boldsymbol{S} = \boldsymbol{A}^{-1}$ is the extremum of $\psi(\boldsymbol{S})$. Moreover, it is known that $\psi(\boldsymbol{S})$ is a convex function in S_+^a [15]. Let us prove this fact next. Without loss of generality, any element of S_+^a can be expressed as $\boldsymbol{A}^{-1} + \theta\boldsymbol{V}$, where $\boldsymbol{V} \in S_+^a$ and $\theta \in \mathbb{R}$. Then $\psi(\boldsymbol{S})$ is rewritten as follows:

$$\begin{aligned}\psi(\boldsymbol{S}) &= \psi(\boldsymbol{A}^{-1} + \theta\boldsymbol{V}) = Tr[(\boldsymbol{A}^{-1} + \theta\boldsymbol{V})\boldsymbol{A}] - \log\det(\boldsymbol{A}^{-1} + \theta\boldsymbol{V}) \\ &= Tr(\boldsymbol{I}_a + \theta\boldsymbol{VA}) - \log\det(\boldsymbol{I} + \theta\boldsymbol{W}) + \log\det\boldsymbol{A} = a \\ &+ \theta Tr(\boldsymbol{VA}) - \sum_{i=1}^{a}\log(1 + \theta\lambda_i) + \log\det\boldsymbol{A}\end{aligned} \tag{B3}$$

where $\boldsymbol{W} \triangleq \boldsymbol{A}^{1/2}\boldsymbol{V}\boldsymbol{A}^{1/2}$, and $\lambda_1, \cdots \lambda_a$ are eigenvalues of $\boldsymbol{W}$. Then, the second-order derivative of $\psi(\boldsymbol{A}^{-1} + \theta\boldsymbol{V})$ with respect to θ is given as

$$\frac{\partial^2}{\partial\theta^2}\psi(\boldsymbol{A}^{-1} + \theta\boldsymbol{V}) = \sum_{i=1}^{a}\frac{\lambda_i^2}{(1 + \theta\lambda_i)^2} > 0 \tag{B4}$$

This means $\psi(\boldsymbol{S})$ is a convex function of arbitrary $\boldsymbol{S} \in S_+^a$, and hence $\boldsymbol{S} = \boldsymbol{A}^{-1}$ minimizes $\psi(\boldsymbol{S})$. Therefore, $\boldsymbol{Z}(=\boldsymbol{S}^{-1}) = \boldsymbol{A}$ minimizes the function $\phi(\boldsymbol{Z})$.

References

[1] Hamel, P. G., and Jategaonkar, R. V., "Evolution of Flight Vehicle System Identification," *Journal of Aircraft*, Vol. 33, No. 1, 1996, pp. 9–28.
doi:10.2514/3.46898

[2] Morelli, E. A., and Klein, V., "Application of System Identification to Aircraft at NASA Langley Research Center," *Journal of Aircraft*, Vol. 42, No. 1, 2005, pp. 12–25.
doi:10.2514/1.3648

[3] Jategaonkar, R. V., *Flight Vehicle System Identification: A Time Domain Methodology*, AIAA, Reston, VA, 2006.

[4] Klein, V., "Estimation of Aircraft Aerodynamic Parameters from Flight Data," *Progress in Aerospace Sciences*, Vol. 26, No. 1, 1989, pp. 1–77.
doi:10.1016/0376-0421(89)90002-X, Pergamon, Oxford, UK.

[5] Paris, A. C., and Alaverdi, O., "Nonlinear Aerodynamic Model Extraction from Flight-Test Data for the S-3B Viking," *Journal of Aircraft*, Vol. 42, No. 1, 2005, pp. 26–32.
doi:10.2514/1.3172

[6] Murray, J. E., and Maine, R. E., "pEst Version 2.1 User's Manual," NASA TM-88280, 1987.

[7] Rohlfs, M., "Identification of Non-Linear Derivative Models from Bo 105 Flight Test Data," *The Aeronautical Journal*, Vol. 102, No. 1011, 1998, pp. 1–8.

[8] Maine, R. E., and Iliff, K. W., "Formulation and Implementation of a Practical Algorithm for Parameter Estimation with Process and Measurement Noise," *SIAM Journal on Applied Mathematics*, Vol. 41, No. 3, 1981, pp. 558–579.
doi:10.1137/0141045

[9] Jategaonkar, R. V., and Plaetschke, E., "Algorithms for Aircraft Parameter Estimation Accounting for Process and Measurement Noise," *Journal of Aircraft*, Vol. 26, No. 4, 1989, pp. 360–372.
doi:10.2514/3.45769

[10] Chowdhary, G., and Jategaonkar, R., "Aerodynamic Parameter Estimation from Flight Data Applying Extended and Unscented Kalman Filter," AIAA Paper 2006-6146, 2006.

[11] Julier, S., Uhlmann, J., and Durrant-Whyte, H. F., "A New Method for the Nonlinear Transformation of Means and Covariances in Filters and Estimators," *IEEE Transactions on Automatic Control*, Vol. 45, No. 3, 2000, pp. 477–482.
doi:10.1109/9.847726

[12] Julier, S., and Uhlmann, J. K., "Unscented Filtering for Nonlinear Estimation," *Proceedings of the IEEE*, Vol. 92, No. 3, 2004, pp. 401–422.
doi:10.1109/JPROC.2003.823141

[13] Yokoyama, N., "System Identification of Flight Vehicle Using Nonlinear Smoothing Combined with Unscented Transform," *Proceedings of the 18th IFAC Symposium on Automatic Control in Aerospace*, 2010.

[14] Bell, B. M., Burke, J. V., and Pillonetto, G., "An Inequality Constrained Nonlinear Kalman–Bucy Smoother by Interior Point Likelihood Maximization," *Automatica*, Vol. 45, No. 1, 2009, pp. 25–33. doi:10.1016/j.automatica.2008.05.029

[15] Boyd, S., and Vandenberghe, L., *Convex Optimization*, Cambridge Univ. Press, Cambridge, England, U.K., 2004, pp. 355–357.

[16] Betts, J. T., *Practical Methods for Optimal Control Using Nonlinear Programming*, SIAM, Philadelphia, PA, 2001.

[17] Yokoyama, N., Suzuki, S., and Tsuchiya, T., "Convergence Acceleration of Direct Trajectory Optimization Using Novel Hessian Calculation Methods," *Journal of Optimization Theory and Applications*, Vol. 136, No. 3, 2008, pp. 297–319. doi:10.1007/s10957-008-9351-0

[18] Särkkä, S., "Unscented Rauch–Tung–Striebel Smoother," *IEEE Transactions on Automatic Control*, Vol. 53, No. 3, 2008, pp. 845–849. doi:10.1109/TAC.2008.919531

[19] Dempster, A. P., Laird, N. M., and Rubin, D. B., "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society Series B (Methodological)*, Vol. 39, No. 1, 1977, pp. 1–38.

[20] Gibson, S., and Ninness, B., "Robust Maximum-Likelihood Estimation of Multivariable Dynamic Systems," *Automatica*, Vol. 41, No. 10, 2005, pp. 1667–1682. doi:10.1016/j.automatica.2005.05.008

[21] Wills, A., Schön, T. B., and Ninness, B., "Parameter Estimation for Discrete-Time Nonlinear Systems Using EM," *Proceedings of the 17th IFAC World Congress*, 2008.

[22] Christophe, A., Doucet, A., Singh, S. S., and Tadić V. B., "Particle Methods for Change Detection, System Identification, and Control," *Proceedings of the IEEE*, Vol. 92, No. 3, 2004, pp. 423–438. doi:10.1109/JPROC.2003.823142

[23] Kullback, S., and Leibler, R. A., "On Information and Sufficiency," *Annals of Mathematical Statistics*, Vol. 22, No. 1, 1951, pp. 79–86. doi:10.1214/aoms/1177729694

[24] Brent, R. P., *Algorithms for Minimization Without Derivatives*, Prentice–Hall, Englewood Cliffs, NJ, 1973, Chap. 7.