Ann Inst Stat Math (2013) 65:889–912
DOI 10.1007/s10463-013-0397-x

On the convergence rate of the unscented transformation

Kwang Woo Ahn · Kung-Sik Chan

Received: 9 August 2011 / Revised: 11 October 2012 / Published online: 7 April 2013

Abstract Nonlinear state-space models driven by differential equations have been widely used in science. Their statistical inference generally requires computing the mean and covariance matrix of some nonlinear function of the state variables, which can be done in several ways. For example, such computations may be approximately done by Monte Carlo, which is rather computationally expensive. Linear approximation by the first-order Taylor expansion is a fast alternative. However, the approximation error becomes non-negligible with strongly nonlinear functions. Unscented transformation was proposed to overcome these difficulties, but it lacks theoretical justification. In this paper, we derive some theoretical properties of the unscented transformation and contrast it with the method of linear approximation. Particularly, we derive the convergence rate of the unscented transformation.

Keywords Unscented transformation · Nonlinear transformation · Monte Carlo · Linear approximation

1 Introduction

Many scientific studies employ nonlinear state-space models for describing the dynamics of a continuous-time state process driven by a system of an ordinary differential equation (Diekmann and Heesterbeek 2000; Simon 2006). The unscented Kalman

K. W. Ahn (✉)
Division of Biostatistics, Medical College of Wisconsin, Milwaukee, WI 53226, USA
e-mail: kwooahn@mcw.edu

K.-S. Chan
Department of Statistics and Actuarial Science, The University of Iowa,
Iowa City, IA 52242, USA
e-mail: kung-sik-chan@uiowa.edu

filter (UKF) was proposed by Julier and Uhlmann (1997) to overcome difficulties such as high computational cost from simulation-based methods and inaccuracy of the extended Kalman filter (EKF), which approximates the nonlinear system by its first-order Taylor expansion. See Simon (2006) for details. Empirical works suggest that the UKF is a promising technique with satisfactory performance, see Julier and Uhlmann (1997, 2004); Wan and van der Merwe (2000). The UKF mimics the updating scheme of the Kalman filter, with each updating step requiring the computation of the mean and covariance matrix of some nonlinear function of the state vector, which is done via the unscented transformation (UT). Specifically, the latter problem concerns computing the mean and covariance matrix of some nonlinear transformation $\mathbf{y} = f(\mathbf{x})$, as well as the covariance matrix between $\mathbf{y}$ and $\mathbf{x}$, where the mean and covariance matrix of $\mathbf{x}$ are known. In contrast with Monte Carlo methods, the UT makes use of a small number of deterministic "sigma points" in estimating these population characteristics. Julier and Uhlmann (1997, 2004) studied the error rates of the UT estimators. However, their derivation of the error rates was heuristic as they did not provide exact error terms and conditions under which such error rates can be obtained. In addition, they did not investigate the error rate of the UT estimator of the covariance matrix between $\mathbf{x}$ and $\mathbf{y}$. Since the UKF has been widely applied to inference with a nonlinear differential equation model that is discretized with a small time step, denoted by $h > 0$, it is essential to study the theoretical properties of the UT where the nonlinear transformation is indexed by a parameter $h > 0$. The main issue we address in this paper concerns a rigorous derivation of the error rates of the UT as $h \to 0$. (We do not pursue the study of the error rate of the UKF in this paper.) Note that for an antisymmetric transformation, i.e., $f(\cdot) = -f(-\cdot)$, if the distribution of $\mathbf{x}$ is symmetric, i.e., the distribution of $\mathbf{x}$ is same as that of $-\mathbf{x}$, then the distribution of $\mathbf{y} = f(\mathbf{x})$ is symmetric. Thus, it is also of interest to study the error rates of the UT for the case of symmetric distributions. In Sect. 2, we elaborate the definition of the UT. The error rates of the UT are then derived in Sect. 3. Section 4 includes a simulation study to compare UT, Monte Carlo, and linear approximation, which shows that the UT is a relatively fast method that generally outperforms the linear approximation method. We conclude briefly in Sect. 5.

2 Unscented transformation

The UT is an approximate scheme for computing the mean and covariance matrix of $\mathbf{y} = f(\mathbf{x})$, where $\mathbf{x}$ is a a $c \times 1$ random vector with known mean $\mathrm{E}(\mathbf{x})$ and covariance matrix $\mathbf{P}$, and $f : \Omega \to \mathbb{R}^q$ is a $q \times 1$ vector function, i.e., $f = (f_1, \ldots, f_q)$, where $\mathbb{R}$ is a set of real numbers and $\Omega \subseteq \mathbb{R}^c$ is the sample space of $\mathbf{x}$, i.e., $P(\mathbf{x} \in \Omega) = 1$. Let $\mathbf{P_y}$ and $\mathbf{P_{xy}}$ be the covariance matrix of $\mathbf{y}$ and the covariance matrix between $\mathbf{x}$ and $\mathbf{y}$, respectively. For a constant $\lambda > -c$, the sigma points $\hat{\mathbf{x}}^{(0)}, \ldots, \hat{\mathbf{x}}^{(2c)}$ are defined as follows:

$$\hat{\mathbf{x}}^{(0)} = \mathrm{E}(\mathbf{x}), \quad \hat{\mathbf{x}}^{(i)} = \mathrm{E}(\mathbf{x}) + \check{\mathbf{x}}^{(i)}, \quad i = 1, \ldots, 2c,$$
$$\check{\mathbf{x}}^{(j)} = \left(\sqrt{(c+\lambda)\mathbf{P}}\right)_j^T, \check{\mathbf{x}}^{(c+j)} = -\left(\sqrt{(c+\lambda)\mathbf{P}}\right)_j^T, \quad j = 1, \ldots, c,$$

where $\sqrt{(c+\lambda)\mathbf{P}}$ is the matrix square root of $(c+\lambda)\mathbf{P}$ such that

$$\left(\sqrt{(c+\lambda)\mathbf{P}}\right)^T\left(\sqrt{(c+\lambda)\mathbf{P}}\right)=(c+\lambda)\mathbf{P},$$

and $(\sqrt{(c+\lambda)\mathbf{P}})_j$ is the jth row of $\sqrt{(c+\lambda)\mathbf{P}}$. Here, $\sqrt{(c+\lambda)\mathbf{P}}$ can be obtained by the Cholesky decomposition or singular value decomposition. The constant λ controls the distance between the sigma points and $\mathrm{E}(\mathbf{x})$. If $\lambda \to -c$, the sigma points tend to be closer to $\mathrm{E}(\mathbf{x})$. If $\lambda \to \infty$, the sigma points tend to be further away from $\mathrm{E}(\mathbf{x})$. Hence, λ is a tuning parameter that controls the error between the true mean or covariance matrices and their UT approximations to be defined below. Let $\hat{\mathbf{y}}^{(i)} = f(\hat{\mathbf{x}}^{(i)})$, $i = 0, \ldots, 2c$. The UT formulas for the approximate mean $\hat{\mathbf{y}}$, covariance matrix $\hat{\mathbf{P}}_{\mathbf{y}}$ of $\mathbf{y}$, and covariance matrix $\hat{\mathbf{P}}_{\mathbf{xy}}$ between $\mathbf{x}$ and $\mathbf{y}$ are

$$\begin{aligned}\hat{\mathbf{y}} &= \sum_{i=0}^{2c} W^{(i)}\hat{\mathbf{y}}^{(i)}, \quad \hat{\mathbf{P}}_{\mathbf{y}} = \sum_{i=0}^{2c} W^{(i)}(\hat{\mathbf{y}}^{(i)}-\hat{\mathbf{y}})(\hat{\mathbf{y}}^{(i)}-\hat{\mathbf{y}})^T,\\ \hat{\mathbf{P}}_{\mathbf{xy}} &= \sum_{i=0}^{2c} W^{(i)}(\hat{\mathbf{x}}^{(i)}-E(\mathbf{x}))(\hat{\mathbf{y}}^{(i)}-\hat{\mathbf{y}})^T,\end{aligned} \tag{1}$$

where $W^{(0)} = \lambda/(c+\lambda)$, $W^{(i)} = 1/(2c+2\lambda)$, $i = 1, \ldots, 2c$. Hence, the UT estimates are weighted sample analogues based on the sigma points, see Simon (2006). On the other hand, the linear approximation scheme used by the EKF approximates the mean and covariance matrices via the first-order Taylor expansion, resulting in the following formulas:

$$\hat{\mathbf{y}}_L = f(\mathrm{E}(\mathbf{x})), \hat{\mathbf{P}}_{\mathbf{y},L} = \mathbf{HPH}^T, \hat{\mathbf{P}}_{\mathbf{xy},L} = \mathbf{PH}^T \tag{2}$$

for estimating $E(\mathbf{y})$, $\mathbf{P}_{\mathbf{y}}$, and $\mathbf{P}_{\mathbf{xy}}$, respectively, where $\mathbf{H}$ is the Jacobian matrix of f evaluated at $E(\mathbf{x})$. Clearly, the preceding UT method is computationally more efficient than the Monte Carlo simulation. In addition, the UT does not require calculating the Jacobian matrix. Below, we show that the UT method provides more accurate approximation than linear approximation, see Sect. 3.

3 Convergence rates of the UT

Define $\tilde{\mathbf{x}} = \mathbf{x} - \mathrm{E}(\mathbf{x}) = (x_1 - \mathrm{E}(x_1) \ldots x_c - \mathrm{E}(x_c))^T$. The derivative $\partial^i f(\mathrm{E}(\mathbf{x}))/(\partial x_1^{k_1} \ldots \partial x_c^{k_c})$ is the derivative of f evaluated at $\mathrm{E}(\mathbf{x})$ where $i = \sum_{j=1}^{c} k_j$, and k_j's are non-negative integers. Assume f is an analytic function over Ω. Then, the Taylor series expansion of f around $\mathrm{E}(\mathbf{x})$ is given as follows:

$$f(\mathbf{x}) = f(\mathrm{E}(\mathbf{x})) + \sum_{i=1}^{\infty}\left(\tilde{x}_1\frac{\partial}{\partial x_1} + \cdots + \tilde{x}_c\frac{\partial}{\partial x_c}\right)^i \frac{f(\mathrm{E}(\mathbf{x}))}{i!}.$$

Define $D^k_{\tilde{\mathbf{x}}} f$ as

$$D^k_{\tilde{\mathbf{x}}} f = \left(\sum_{i=1}^{c} \tilde{x}_i \frac{\partial}{\partial x_i} \right)^k f(\mathrm{E}(\mathbf{x})).$$

We assume that $D^k_{\tilde{\mathbf{x}}} f$ is integrable on Ω for any non-negative integer k. We also assume that there exists Y with finite absolute first moment such that $|\sum_{i=0}^m D^i_{\tilde{\mathbf{x}}} f/i!| \le Y$ a.e. on Ω for all m. Since $\sum_{i=0}^m D^i_{\tilde{\mathbf{x}}} f/i! \to f$, we have $\lim_{m\to\infty} E(\sum_{i=0}^m D^i_{\tilde{\mathbf{x}}} f/i!) = \lim_{m\to\infty} \sum_{i=0}^m E(D^i_{\tilde{\mathbf{x}}} f/i!) = E\{f(\mathbf{x})\}$ by the dominated convergence theorem. In practice, these conditions are satisfied if i) there exists a constant $Q > 0$ such that $E|\tilde{x}_1^{k_1} \ldots \tilde{x}_c^{k_c}| \le Q^{\sum_{j=1}^c k_j}$, for any non-negative integers k_j; ii) for all j, $1 \le j \le q$, there exists some constant $R > 0$ such that

$$\left| \frac{\partial^{\sum_{i=1}^c k_i} f_j}{\partial x_1^{k_1} \ldots \partial x_c^{k_c}} \right| \le R^{\sum_{i=1}^c k_i},$$

for any non-negative integers k_j, $1 \le j \le c$. Thus, we can derive the following results:

$$\mathrm{E}(\mathbf{y}) = f(\mathrm{E}(\mathbf{x})) + \frac{1}{2!} \sum_{i=1}^{c} \sum_{j=1}^{c} P_{ij} \frac{\partial^2}{\partial x_i \partial x_j} f(\mathrm{E}(\mathbf{x})) + \sum_{j=3}^{\infty} \frac{1}{j!} \mathrm{E}\left[D^j_{\tilde{\mathbf{x}}} f \right], \tag{3}$$

$$\hat{\mathbf{y}} = f(\mathrm{E}(\mathbf{x})) + \frac{1}{2} \sum_{i=1}^{c} \sum_{j=1}^{c} P_{ij} \frac{\partial^2}{\partial x_i \partial x_j} f(\mathrm{E}(\mathbf{x})) + \frac{1}{2(c+\lambda)} \sum_{i=1}^{2c} \sum_{j=2}^{\infty} \frac{1}{(2j)!} D^{2j}_{\tilde{\mathbf{x}}^{(i)}} f, \tag{4}$$

where P_{ij} is the (i, j)th entry of $\mathbf{P}$. We consider the case that the random variable $\mathbf{x} = \mathbf{x}_h$ is indexed by a positive number h such that $\mathbf{x} = E(\mathbf{x}) + o_p(h)$ where $E(\mathbf{x})$ is independent of h, in which case Eq. (3) provides an heuristic expansion of $E(\mathbf{y}) = E\{f(\mathbf{x}_h)\}$ with summands of orders h^j, $j = 0, 1 \ldots$. The order of the error rate of the UT estimator of $E(\mathbf{y})$ may then be studied by comparing Eqs. (3) and (4). If $\mathbf{x}$ has a symmetric distribution about its mean $E(\mathbf{x})$, we have

$$\mathrm{E}(\mathbf{y}) = f(\mathrm{E}(\mathbf{x})) + \frac{1}{2!} \sum_{i=1}^{c} \sum_{j=1}^{c} P_{ij} \frac{\partial^2}{\partial x_i \partial x_j} f(\mathrm{E}(\mathbf{x})) + \sum_{j=2}^{\infty} \frac{1}{(2j)!} \mathrm{E}\left[D^{2j}_{\tilde{\mathbf{x}}} f \right]. \tag{5}$$

Define $\mathbb{A} = \{(a, b)|a, b \in \mathbb{N}$ but $(a, b) \ne (1, 1)\}$ and $\mathbb{B} = \{(a, b)|a, b \in \mathbb{N}$, but $a \ne 1, b \ne 1, (a, b) \ne (2, 2)\}$ where $\mathbb{N}$ is the set of natural numbers. Then, we have

$$\mathbf{P_y} = \mathbf{HPH}^T - \frac{1}{4}\mathrm{E}\left(D^2_{\tilde{\mathbf{x}}} f\right)\mathrm{E}\left(D^2_{\tilde{\mathbf{x}}} f\right)^T + \mathrm{E}\left[\sum_{(i,j)\in\mathbb{A}} \frac{1}{i!j!}\left(D^i_{\tilde{\mathbf{x}}} f\right)\left(D^j_{\tilde{\mathbf{x}}} f\right)^T\right]$$
$$-\left[\sum_{(i,j)\in\mathbb{B}} \frac{1}{i!j!}\mathrm{E}\left(D^i_{\tilde{\mathbf{x}}} f\right)\mathrm{E}\left(D^j_{\tilde{\mathbf{x}}} f\right)^T\right],$$
$$\hat{\mathbf{P}}_\mathbf{y} = \mathbf{HPH}^T - \frac{1}{4}\mathrm{E}\left(D^2_{\tilde{\mathbf{x}}} f\right)\mathrm{E}\left(D^2_{\tilde{\mathbf{x}}} f\right)^T \tag{6}$$
$$+\frac{1}{2(c+\lambda)}\sum_{i=1}^{2c}\left[\sum_{\substack{k+\ell=\text{ even}\\(k,\ell)\in\mathbb{A}}} \frac{1}{k!\ell!}\left(D^k_{\tilde{\mathbf{x}}^{(i)}} f\right)\left(D^\ell_{\tilde{\mathbf{x}}^{(i)}} f\right)^T\right]$$
$$-\sum_{(k,\ell)\in\mathbb{A}}\left[\frac{1}{(2k!)(2\ell!)}\frac{1}{4(c+\lambda)^2}\sum_{i=1}^{2c}\sum_{j=1}^{2c}\left(D^{2k}_{\tilde{\mathbf{x}}^{(i)}} f\right)\left(D^{2\ell}_{\tilde{\mathbf{x}}^{(j)}} f\right)^T\right].$$

If the distribution of $\mathbf{x}$ is symmetric about its mean,

$$\mathbf{P_y} = \mathbf{HPH}^T - \frac{1}{4}\mathrm{E}\left(D^2_{\tilde{\mathbf{x}}} f\right)\mathrm{E}\left(D^2_{\tilde{\mathbf{x}}} f\right)^T + \mathrm{E}\left[\sum_{\substack{i+j=\text{ even}\\(i,j)\in\mathbb{A}}} \frac{1}{i!j!}\left(D^i_{\tilde{\mathbf{x}}} f\right)\left(D^j_{\tilde{\mathbf{x}}} f\right)^T\right]$$
$$-\left[\sum_{(i,j)\in\mathbb{A}} \frac{1}{(2i!)(2j!)}\mathrm{E}\left(D^{2i}_{\tilde{\mathbf{x}}} f\right)\mathrm{E}\left(D^{2j}_{\tilde{\mathbf{x}}} f\right)^T\right]. \tag{7}$$

In addition,

$$\mathbf{P_{xy}} = \mathbf{PH}^T + \sum_{i=2}^{\infty}\frac{1}{i!}\mathrm{E}\left[\tilde{\mathbf{x}}\left(D^i_{\tilde{\mathbf{x}}} f\right)^T\right],$$
$$\hat{\mathbf{P}}_\mathbf{xy} = \mathbf{PH}^T + \sum_{k=1}^{\infty}\frac{1}{(2k+1)!}\frac{1}{2(c+\lambda)}\sum_{i=1}^{2c}\tilde{\mathbf{x}}^{(i)}\left(D^{2k+1}_{\tilde{\mathbf{x}}^{(i)}} f\right)^T. \tag{8}$$

When the distribution of $\mathbf{x}$ is symmetric about the mean $E(\mathbf{x})$,

$$\mathbf{P_{xy}} = \mathbf{PH}^T + \sum_{i=1}^{\infty}\frac{1}{(2i+1)!}\mathrm{E}\left[\tilde{\mathbf{x}}\left(D^{2i+1}_{\tilde{\mathbf{x}}} f\right)^T\right]. \tag{9}$$

Detailed derivations of (3)–(9) can be found in Appendix A. From (3)–(9), we have the following lemma:

Lemma 1 *Assume that*

1. f *is an analytic function;*
2. $D_{\tilde{\mathbf{x}}}^k f$ *is integrable on* Ω *for any non-negative integer* k*;*
3. *there exists* Y *with finite absolute first moment such that* $|\sum_{i=0}^m D_{\tilde{\mathbf{x}}}^i f/i!| \le Y$ *a.e. on* Ω *for all* m.

Then,

1. $E(\mathbf{y}) - \hat{\mathbf{y}} = \sum_{j=3}^{\infty} \frac{1}{j!} E[D_{\tilde{\mathbf{x}}}^j f] - \frac{1}{2(c+\lambda)} \sum_{i=1}^{2c} \sum_{j=2}^{\infty} \frac{1}{(2j)!} D_{\tilde{\mathbf{x}}^{(i)}}^{2j} f;$
2. $\mathbf{P_y} - \hat{\mathbf{P}}_{\mathbf{y}} = E\left[\sum_{(i,j)\in\mathbb{A}} \frac{1}{i!j!} (D_{\tilde{\mathbf{x}}}^i f)(D_{\tilde{\mathbf{x}}}^j f)^T\right] - \left[\sum_{(i,j)\in\mathbb{B}} \frac{1}{i!j!} E(D_{\tilde{\mathbf{x}}}^i f) E(D_{\tilde{\mathbf{x}}}^j f)^T\right]$
$-\frac{1}{2(c+\lambda)} \sum_{i=1}^{2c} \left[\sum_{\substack{k+\ell=\,even\\(k,\ell)\in\mathbb{A}}} \frac{1}{k!\ell!} (D_{\tilde{\mathbf{x}}^{(i)}}^k f)(D_{\tilde{\mathbf{x}}^{(i)}}^\ell f)^T\right]$
$+\sum_{(k,\ell)\in\mathbb{A}} \left[\frac{1}{(2k!)(2\ell!)} \frac{1}{4(c+\lambda)^2} \sum_{i=1}^{2c} \sum_{j=1}^{2c} (D_{\tilde{\mathbf{x}}^{(i)}}^{2k} f)(D_{\tilde{\mathbf{x}}^{(j)}}^{2\ell} f)^T\right];$
3. $\mathbf{P_{xy}} - \hat{\mathbf{P}}_{\mathbf{xy}} = \sum_{i=2}^{\infty} \frac{1}{i!} E\left[\tilde{\mathbf{x}} \left(D_{\tilde{\mathbf{x}}}^i f\right)^T\right] - \sum_{k=1}^{\infty} \frac{1}{(2k+1)!} \frac{1}{2(c+\lambda)} \sum_{i=1}^{2c} \tilde{\mathbf{x}}^{(i)} \left(D_{\tilde{\mathbf{x}}^{(i)}}^{2k+1} f\right)^T.$

Furthermore, if the distribution of $\mathbf{x}$ *is symmetric about the mean* $E(\mathbf{x})$*, then*

1. $E(\mathbf{y}) - \hat{\mathbf{y}} = \sum_{j=2}^{\infty} \frac{1}{(2j)!} E[D_{\tilde{\mathbf{x}}}^{2j} f] - \frac{1}{2(c+\lambda)} \sum_{i=1}^{2c} \sum_{j=2}^{\infty} \frac{1}{(2j)!} D_{\tilde{\mathbf{x}}^{(i)}}^{2j} f;$
2. $\mathbf{P_y} - \hat{\mathbf{P}}_{\mathbf{y}} = E\left[\sum_{\substack{i+j=\,even\\(i,j)\in\mathbb{A}}} \frac{1}{i!j!} (D_{\tilde{\mathbf{x}}}^i f)(D_{\tilde{\mathbf{x}}}^j f)^T\right]$
$-\left[\sum_{(i,j)\in\mathbb{A}} \frac{1}{(2i!)(2j!)} E(D_{\tilde{\mathbf{x}}}^{2i} f) E(D_{\tilde{\mathbf{x}}}^{2j} f)^T\right]$
$-\frac{1}{2(c+\lambda)} \sum_{i=1}^{2c} \left[\sum_{\substack{k+\ell=\,even\\(k,\ell)\in\mathbb{A}}} \frac{1}{k!\ell!} (D_{\tilde{\mathbf{x}}^{(i)}}^k f)(D_{\tilde{\mathbf{x}}^{(i)}}^\ell f)^T\right]$
$+\sum_{(k,\ell)\in\mathbb{A}} \left[\frac{1}{(2k!)(2\ell!)} \frac{1}{4(c+\lambda)^2} \sum_{i=1}^{2c} \sum_{j=1}^{2c} (D_{\tilde{\mathbf{x}}^{(i)}}^{2k} f)(D_{\tilde{\mathbf{x}}^{(j)}}^{2\ell} f)^T\right];$
3. $\mathbf{P_{xy}} - \hat{\mathbf{P}}_{\mathbf{xy}} = \sum_{i=1}^{\infty} \frac{1}{(2i+1)!} E\left[\tilde{\mathbf{x}} \left(D_{\tilde{\mathbf{x}}}^{2i+1} f\right)^T\right]$
$-\sum_{k=1}^{\infty} \frac{1}{(2k+1)!} \frac{1}{2(c+\lambda)} \sum_{i=1}^{2c} \tilde{\mathbf{x}}^{(i)} \left(D_{\tilde{\mathbf{x}}^{(i)}}^{2k+1} f\right)^T.$

For the linear approximation scheme defined by (2), we have the following parallel results:

Lemma 2 *Under the same conditions of Lemma 1, the linear approximation based on the first-order Taylor expansion enjoys the following properties:*

1. $E(\mathbf{y}) - \hat{\mathbf{y}}_L = \frac{1}{2!} \sum_{i=1}^{c} \sum_{j=1}^{c} P_{ij} \frac{\partial^2}{\partial x_i \partial x_j} f(E(\mathbf{x})) + \sum_{j=3}^{\infty} \frac{1}{j!} E[D_{\tilde{\mathbf{x}}}^j f];$
2. $\mathbf{P_y} - \hat{\mathbf{P}}_{\mathbf{y},L} = -\frac{1}{4} E(D_{\tilde{\mathbf{x}}}^2 f) E(D_{\tilde{\mathbf{x}}}^2 f)^T + E\left[\sum_{(i,j)\in\mathbb{A}} \frac{1}{i!j!} (D_{\tilde{\mathbf{x}}}^i f)(D_{\tilde{\mathbf{x}}}^j f)^T\right]$
$-\left[\sum_{(i,j)\in\mathbb{B}} \frac{1}{i!j!} E(D_{\tilde{\mathbf{x}}}^i f) E(D_{\tilde{\mathbf{x}}}^j f)^T\right];$
3. $\mathbf{P_{xy}} - \hat{\mathbf{P}}_{\mathbf{xy},L} = \sum_{i=2}^{\infty} \frac{1}{i!} E\left[\tilde{\mathbf{x}} \left(D_{\tilde{\mathbf{x}}}^i f\right)^T\right].$

If the distribution of $\mathbf{x}$ *is, furthermore, symmetric about its mean, then*

1. $E(\mathbf{y}) - \hat{\mathbf{y}}_L = \frac{1}{2!}\sum_{i=1}^{c}\sum_{j=1}^{c} P_{ij}\frac{\partial^2}{\partial x_i \partial x_j} f(E(\mathbf{x})) + \sum_{j=2}^{\infty}\frac{1}{(2j)!}E[D_{\tilde{\mathbf{x}}}^{2j} f]$;
2. $\mathbf{P}_\mathbf{y} - \hat{\mathbf{P}}_{\mathbf{y},L} = -\frac{1}{4}E(D_{\tilde{\mathbf{x}}}^2 f)E(D_{\tilde{\mathbf{x}}}^2 f)^T + E\left[\sum_{i+j=\,even(i,j)\in\mathbb{A}}\frac{1}{i!j!}(D_{\tilde{\mathbf{x}}}^i f)(D_{\tilde{\mathbf{x}}}^j f)^T\right]$ $-\left[\sum_{(i,j)\in\mathbb{A}}\frac{1}{(2i!)(2j!)}E(D_{\tilde{\mathbf{x}}}^{2i} f)E(D_{\tilde{\mathbf{x}}}^{2j} f)^T\right]$;
3. $\mathbf{P}_{\mathbf{xy}} - \hat{\mathbf{P}}_{\mathbf{xy},L} = \sum_{i=1}^{\infty}\frac{1}{(2i+1)!}E\left[\tilde{\mathbf{x}}\left(D_{\tilde{\mathbf{x}}}^{2i+1} f\right)^T\right]$.

Remark 1 These results show that the UT estimator of the mean $E(\mathbf{y}) = E\{f(\mathbf{x})\}$ matches the true value correctly up to the second-order terms in the expansion (3) whereas the estimator from linearization of $f(\mathrm{E}(\mathbf{x}))$ only matches the true mean up to the first-order term. If the distribution of $\mathbf{x}$ is symmetric about its mean, the UT estimator of $E(\mathbf{y})$ matches the true value correctly up to the third-order terms in the expansion. The UT estimator of the covariance matrix $\mathbf{P}_\mathbf{y}$ matches more terms than its linear approximation counterpart. For approximating $\mathbf{P}_{\mathbf{xy}}$, the UT estimators and its linear approximation counterpart provide the same order of approximation. Nevertheless, in all cases, the UT estimator has other terms to compensate for the higher-order remainder terms. This compensation appears to be better when $\mathbf{x}$ has a symmetric distribution about its mean. The matching patterns between $\hat{\mathbf{y}}$ ($\hat{\mathbf{P}}_\mathbf{y}$) and their population counterparts for the case that $\mathbf{x}$ has a symmetric distribution was earlier noted by Julier and Uhlmann (2004) although they did not provide the exact error terms and the conditions under which the error rates obtain.

Remark 2 Lemma 1 shows that if λ is close to $-c$ or very large, the bias may be severe unless the nonlinearity of f is minimal. Choosing an optimal λ is a difficult problem because it is generally infeasible to calculate the expectations stated in Lemma 1 for a nonlinear f. However, they can be calculated for certain distributions, for example, normal distributions. When $\mathbf{x}$ is normal, $E(x_1^{k_1}\ldots x_c^{k_c})$ can be expressed as a function of the components of $\mathbf{P}$, see Isserlis (1918), Holmquist (1988), Triantafyllopoulos (2003). So we can calculate the mean squared error (MSE) in this case. From Lemma 1, the bias can be approximated as $E(\mathbf{y}) - \hat{\mathbf{y}} \approx \sum_{j=2}^{m} 1/(2j)!\mathrm{E}[D_{\tilde{\mathbf{x}}}^{2j} f] - 1/(2c + 2\lambda)\sum_{i=1}^{2c}\sum_{j=2}^{m} 1/(2j)!D_{\tilde{\mathbf{x}}(i)}^{2j} f$ for some m, where the first term can be expressed as a function of the components of $\mathbf{P}$. The corresponding variance can be derived based on (1). Then, we can minimize the MSE to obtain an optimal λ.

Next, define D_k as follows:

Definition 1

$$D_k = \left\{f : \Omega \subset \mathbb{R}^c \to \mathbb{R}^q \,|\, f \text{ is a polynomial of degree at most } k\right\}.$$

Theorem 1 1. *If $f \in D_2$, then $\hat{\mathbf{y}} = E(\mathbf{y})$. In addition, if the distribution of $\mathbf{x}$ is symmetric about its mean and $f \in D_3$, then, $\hat{\mathbf{y}}$=$E(\mathbf{y})$.*
2. *If $f \in D_1$, i.e., f is linear, then $\hat{\mathbf{P}}_\mathbf{y} = \mathbf{P}_\mathbf{y}$ and $\hat{\mathbf{P}}_{\mathbf{xy}} = \mathbf{P}_{\mathbf{xy}}$.*

Proof This follows from the Taylor series expansions in (3)–(9). □

The UKF has been applied for estimating a nonlinear state-space model where the state equation is an ordinary differential equation or a stochastic differential equation with observations at time $t_1 \ldots t_n$, where $t_{i+1} - t_i = h$ for $i = 1 \ldots n-1$, where, with no loss of generality, $h \leq 1$. To solve these differential equations, one may discretize the system equation using the Euler method or Runge-Kutta method (Boyce and DiPrima 2004; Milstein and Tretyakov 2004), which results in the case that conditionally $\mathbf{x}$ has a covariance matrix $\mathbf{P} = h\mathbf{P}^*$ for some h-free positive definite matrix $\mathbf{P}^*$, see Ahn and Chan (2011). Thus, it is interesting to compare the error rates of the UT estimators of $E(\mathbf{y}) = E\{f(\mathbf{x})\}$, $\mathbf{P_y}$, $\mathbf{P_{xy}}$ and their counterparts from linear approximation when $\mathbf{P} = h\mathbf{P}^*$. We allow that f may depend on h. Then, we can obtain the following theorem:

Theorem 2 *Suppose* $\mathbf{P} = h\mathbf{P}^*$. *Assume*

1. *there exists a h-free constant* $M > 0$ *such that* $E|\tilde{x}_1^{k_1} \ldots \tilde{x}_c^{k_c}| \leq h^{i/2} M^{i/2}$, *for any non-negative integers* k_j *where* $\sum_{j=1}^{c} k_j = i$;
2. *for all* j, $1 \leq j \leq q$, *and for all* $0 < h \leq 1$, *there exists some h-free constant* $K > 0$,

$$\left| \frac{\partial^i f_j(E(\mathbf{x}))}{\partial x_1^{k_1} \ldots \partial x_c^{k_c}} \right| \leq K^i,$$

for any non-negative integers k_j, $1 \leq j \leq c$, *where* $i = \sum_{j=1}^{c} k_j$.

Then,

$$E(\mathbf{y}) - \hat{\mathbf{y}} = O(h^{3/2}),\ \mathbf{P_y} - \hat{\mathbf{P}}_\mathbf{y} = O(h^{3/2}),\ \mathbf{P_{xy}} - \hat{\mathbf{P}}_\mathbf{xy} = O(h^{3/2}),$$
$$E(\mathbf{y}) - \hat{\mathbf{y}}_L = O(h),\ \mathbf{P_y} - \hat{\mathbf{P}}_{\mathbf{y},L} = O(h^{3/2}),\ \mathbf{P_{xy}} - \hat{\mathbf{P}}_{\mathbf{xy},L} = O(h^{3/2}).$$

If $\mathbf{x}$ *has a symmetric distribution about its mean, then*

$$E(\mathbf{y}) - \hat{\mathbf{y}} = O(h^2),\ \mathbf{P_y} - \hat{\mathbf{P}}_\mathbf{y} = O(h^2),\ \mathbf{P_{xy}} - \hat{\mathbf{P}}_\mathbf{xy} = O(h^2),$$
$$E(\mathbf{y}) - \hat{\mathbf{y}}_L = O(h),\ \mathbf{P_y} - \hat{\mathbf{P}}_{\mathbf{y},L} = O(h^2),\ \mathbf{P_{xy}} - \hat{\mathbf{P}}_{\mathbf{xy},L} = O(h^2).$$

Proof See Appendix B. □

Remark 3 Condition 1 of Theorem 2 is motivated by the fact that $\tilde{x}_1^{k_1} \ldots \tilde{x}_c^{k_c} = O_p(h^{i/2})$ because $\tilde{x}_j = O_p(\sqrt{h})$. Condition 2 of Theorem 2 is satisfied if the function f is a polynomial.

Remark 4 Theorem 2 shows that the UT estimator $E(\mathbf{y})$ has a smaller error rate than the estimator from linear approximation although the estimators of the covariance matrices $\mathbf{P_y}$ and $\mathbf{P_{xy}}$ have the same error rate. This suggests that discretization methods based on the UKF would be more accurate than the EKF. Using Theorem 2, Ahn and Chan (2011) showed that discretization methods based on the UKF are more accurate than those based on the EKF.

4 Simulation study

A simulation study was conducted to compare the three methods, namely, UT, Monte Carlo, and linear approximation, for estimating the mean and standard deviation of the random variable

$$f(x, y) = \sqrt{x + 1}\cos(y),$$

where x and y are independent of each other, $x \sim \text{Gamma}(2, 2/\sqrt{h})$ with mean $\sqrt{h}$ and variance $h/2$, and $y \sim \text{Unif}(0, \pi\sqrt{h})$. Thus, the covariance matrix of $(x, y)^T$ equals

$$h\begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{\pi^2}{12} \end{pmatrix}.$$

We conducted 240 scenarios by considering 240 h values: $h = 0.1, 0.11, 0.12 \ldots 2.5$. For each h, we computed the estimates of the mean and standard deviation of $f(x, y)$, via (i) UT with $\lambda = -0.5, 0$, and 0.5, (ii) Monte Carlo with two sample sizes, $n = 100$ and $n = 1{,}000$, and (iii) linearization. Table 1 shows the results for $h = 0.1, 0.5, 1.5, 2, 2.5$. The true means and standard deviations were obtained based

Table 1 Comparison of the moment estimates of $f(x, y)$ for various methods

h	True values	UT $\lambda = -0.5$	UT $\lambda = 0$	UT $\lambda = 0.5$	MC ($n = 100$)	MC ($n = 1{,}000$)	Linearization
0.1							
M	0.964	0.964	0.964	0.964	1.003	0.959	1.009
SD	0.176	0.178	0.179	0.181	0.169	0.173	0.179
0.5							
M	0.464	0.460	0.462	0.464	0.339	0.455	0.580
SD	0.673	0.685	0.666	0.648	0.716	0.682	0.756
1							
M	0.0004	0.000	0.000	0.000	−0.059	−0.053	0.000
SD	1.000	1.035	0.959	0.886	1.002	0.983	1.283
1.5							
M	−0.247	−0.233	−0.247	−0.260	−0.090	−0.223	−0.516
SD	1.094	1.136	1.025	0.923	1.064	1.101	1.558
2							
M	−0.332	−0.292	−0.334	−0.375	−0.475	−0.351	−0.941
SD	1.080	1.109	1.038	0.976	1.005	1.081	1.598
2.5							
M	−0.305	−0.235	−0.321	−0.401	−0.293	−0.324	−1.271
SD	1.064	1.069	1.118	1.141	1.144	1.097	1.435

The rows with heading “M” and “SD” are the estimated mean and standard deviation, respectively

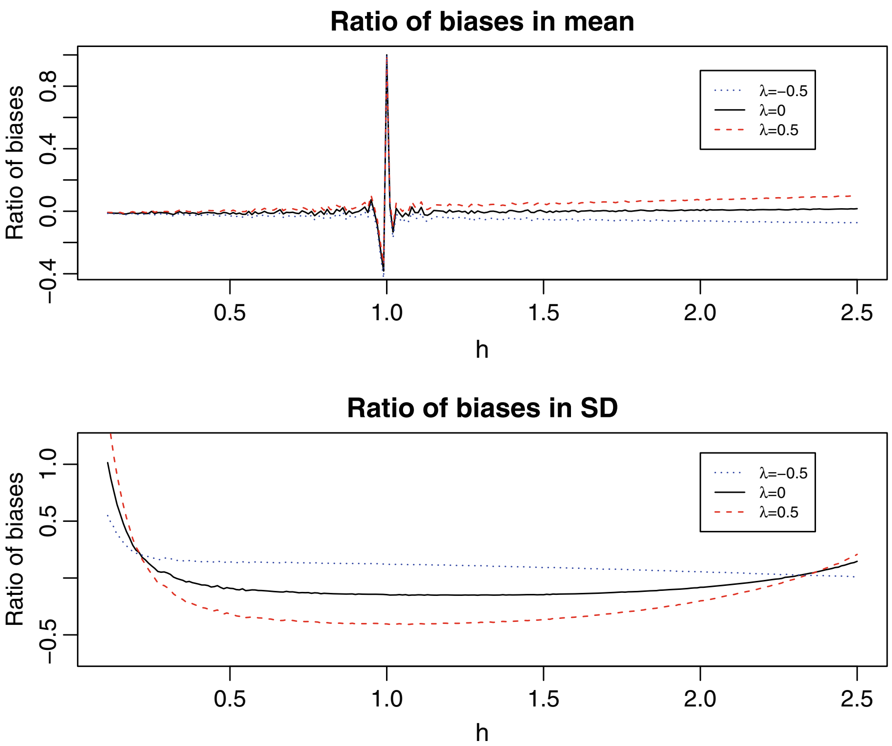

Fig. 1 Scatter diagram of the ratio of the bias in the UT estimate of the mean (*upper diagram*) and SD (*lower diagram*) of $f(x, y)$ to their linear approximation counterparts versus h. Note that, except for $h = 0.98, 0.99, 1$, the ratios of the bias in the mean estimates fall on a straight line approximately, for fixed α. In addition, except for $h \leq 0.13$, the ratios of the bias in SD are well below 1

on Monte Carlo with $n = 1{,}000{,}000$. In general, the UT estimates are close to the true values and the UT method outperformed the Monte Carlo method with $n = 100$ and the linear approximation method. The UT estimates of the mean are robust to the three λ values for $h \leq 1.5$, while those of the standard deviation are relatively more robust to λ, for all h values examined. In general, the UT estimates of the mean and standard deviation, with λ equal to zero, are generally closer to their true values than the UT estimates with other λ values, for all h values examined. The upper diagram of Fig. 1 shows the ratios of the bias in the UT estimate of the mean of $f(x, y)$ to that of the linear approximation method. The blue dotted, black solid, and red dashed lines indicate the ratios for $\lambda = -0.5, 0, 0.5$, respectively. Except for $h = 0.98, 0.99, 1$, these ratios for each λ are approximately a linear function, i.e., of the form $c_0 \times h$ for some constant c_0, which is consistent with the results in Theorem 2. For $h = 1$, it is readily checked that the distribution of $f(x, y)$ is symmetric about 0, in which case UT and linear approximation estimates coincide with 0. (Note that, for $h = 1$, the "true" value obtained from Monte Carlo with $n = 1{,}000{,}000$ is 0.0004.) Theorem 2 implies that the ratio of the biases in mean is asymptotically a linear function of h, except at $h = 0$ where the ratio jumps discontinuously to 1. In a simulation study with

Table 2 Comparison for user CPU time (in seconds)

	UT	Monte Carlo ($n = 100$)	Monte Carlo ($n = 1{,}000$)	Linearization
User CPU time	0.04	0.05	0.18	0.03

finite simulation size, the ratio of the biases is a continuous function of h since the simulation is initialized with the same random seed for each h. The approximation may then be expected to admit large fluctuation around the point of discontinuity, which is analogous to the Gibbs phenomenon: a finite partial Fourier series approximation to a piecewise continuous function has large oscillation around any point of discontinuity of the limit function, regardless of the number of terms included in the approximation. The Gibbs phenomenon holds in more general situations, see Foster and Richards (1991), Shim and Park (2005), and the references therein.

The lower diagram of Fig. 1 plots the ratios of the UT estimate of the standard deviation of $f(x, y)$ to that of the linear approximation method, which shows that in general the bias of the UT estimate is a fraction of that of the linear approximation method, in terms of magnitude, which, again, is consistent with the results in Theorem 2. Table 2 shows the total user CPU time for each method to compute the estimates for all the 240 scenarios using a computer with Intel Core(TM)2 Duo CPU 2.13 GHz and 6GB RAM. The UT method is faster than the Monte Carlo method, but slightly slower than the linear approximation method. Altogether, the UT method gives reasonable estimates and is more accurate than the linear approximation method and the Monte Carlo method with $n = 100$, but faster than the Monte Carlo method with $n = 1000$.

Ahn and Chan (2011) used the UT to estimate the mean and covariance matrix of some random vector. They observed that in general the UT induced more bias in estimating the covariance matrix than estimating the mean. In addition, when h is large, the UT estimates were more sensitive to the choice of λ in the multivariate case than the scalar case.

5 Conclusion

We have derived some theoretical properties of the UT and linear approximation, and conducted a simulation study. The UT makes use of some deterministic sigma points in contrast with the Monte Carlo method. In addition, it does not require calculating the Jacobian matrix unlike the method of linear approximation. Derivations based on Taylor expansions show that the estimated mean and covariance matrix from the UT match their true values to more higher-order terms than the method of linear approximation. The simulation study confirms that the UT is fast and outperforms the linear approximation in terms of accuracy.

Appendix A: Derivation of results for (3)–(9)

For simplicity, assume the pdf of the random vector $\mathbf{x}$ is symmetric about its mean with a known mean $\mathrm{E}(\mathbf{x})$ and covariance matrix $\mathbf{P_x}$, and $\mathbf{y} = f(\mathbf{x})$. Assume f is an analytic function. We also assume that $D^k_{\tilde{\mathbf{x}}} f$ is integrable on Ω for any non-negative integer k and there exists Y with finite absolute moment such that $|\sum_{i=0}^{m} D^i_{\tilde{\mathbf{x}}} f/i!| \le Y$ a.e. on Ω for all m. Replacing $\hat{\mathbf{y}}^{(i)}$ by its Taylor expansion around $\mathrm{E}(\mathbf{x})$, we get

$$\begin{aligned}\hat{\mathbf{y}} &= W^{(0)}\hat{\mathbf{y}}^{(0)} + \sum_{i=1}^{2c} W^{(i)}\hat{\mathbf{y}}^{(i)} \\ &= \frac{\lambda}{c+\lambda} f(\mathrm{E}(\mathbf{x})) + \frac{1}{2(c+\lambda)} \sum_{i=1}^{2c} \left(f(\mathrm{E}(\mathbf{x})) + D_{\tilde{\mathbf{x}}^{(i)}} f + \frac{1}{2!} D^2_{\tilde{\mathbf{x}}^{(i)}} f + \cdots \right) \\ &= f(\mathrm{E}(\mathbf{x})) + \frac{1}{2(c+\lambda)} \sum_{i=1}^{2c} \left(D_{\tilde{\mathbf{x}}^{(i)}} f + \frac{1}{2!} D^2_{\tilde{\mathbf{x}}^{(i)}} f + \cdots \right).\end{aligned}$$

Now notice that

$$\sum_{j=1}^{2c} D^{2k+1}_{\tilde{\mathbf{x}}^{(j)}} f = \mathbf{0},$$

because $\tilde{\mathbf{x}}^{(j)} = -\tilde{\mathbf{x}}^{(c+j)},\ j = 1, \ldots, c$. See Simon (2006) for the details. Therefore,

$$\hat{\mathbf{y}} = f(\mathrm{E}(\mathbf{x})) + \frac{1}{2(c+\lambda)} \sum_{i=1}^{2c} \frac{1}{2!} D^2_{\tilde{\mathbf{x}}^{(i)}} f + \frac{1}{2(c+\lambda)} \sum_{i=1}^{2c} \left(\frac{1}{4!} D^4_{\tilde{\mathbf{x}}^{(i)}} f + \frac{1}{6!} D^6_{\tilde{\mathbf{x}}^{(i)}} f \cdots \right).$$

In addition, similar to Simon (2006), we can obtain

$$\frac{1}{2(c+\lambda)} \sum_{i=1}^{2c} \frac{1}{2!} D^2_{\tilde{\mathbf{x}}^{(i)}} f = \frac{1}{2} \sum_{k=1}^{c} \sum_{\ell=1}^{c} P_{k\ell} \frac{\partial^2}{\partial x_k \partial x_\ell} f(\mathrm{E}(\mathbf{x})), \tag{10}$$

because $\tilde{\mathbf{x}}^{(i)} = -\tilde{\mathbf{x}}^{(c+i)}$. Thus,

$$\begin{aligned}\hat{\mathbf{y}} = f(\mathrm{E}(\mathbf{x})) + \frac{1}{2} \sum_{i=1}^{c} \sum_{j=1}^{c} P_{ij} \frac{\partial^2}{\partial x_i \partial x_j} f(\mathrm{E}(\mathbf{x})) \\ + \frac{1}{2(c+\lambda)} \sum_{i=1}^{2c} \left(\frac{1}{4!} D^4_{\tilde{\mathbf{x}}^{(i)}} f + \frac{1}{6!} D^6_{\tilde{\mathbf{x}}^{(i)}} f + \cdots \right).\end{aligned} \tag{11}$$

Similarly,

$$\mathrm{E}(\mathbf{y}) = f(\mathrm{E}(\mathbf{x})) + \frac{1}{2!} \mathrm{E}\left[D^2_{\tilde{\mathbf{x}}} f \right] + \frac{1}{4!} \mathrm{E}\left[D^4_{\tilde{\mathbf{x}}} f \right] + \cdots.$$

It can be easily shown

$$\frac{1}{2!}\mathrm{E}\left[D_{\tilde{\mathbf{x}}}^2 f\right] = \frac{1}{2!}\sum_{k=1}^{c}\sum_{\ell=1}^{c} P_{k\ell}\frac{\partial^2}{\partial x_k \partial x_\ell} f(\mathrm{E}(\mathbf{x})).$$

As a result, we have

$$\mathrm{E}(\mathbf{y}) = f(\mathrm{E}(\mathbf{x})) + \frac{1}{2!}\left[\sum_{i=1}^{c}\sum_{j=1}^{c} P_{ij}\frac{\partial^2}{\partial x_i \partial x_j} f(\mathrm{E}(\mathbf{x}))\right] + \frac{1}{4!}\mathrm{E}\left[D_{\tilde{\mathbf{x}}}^4 f\right] + \cdots . \tag{12}$$

Next, we turn our attention to the covariance structure:

$$\mathbf{P}_{\mathbf{y}} = \mathrm{E}[(\mathbf{y} - \mathrm{E}(\mathbf{y}))(\mathbf{y} - \mathrm{E}(\mathbf{y}))^T].$$

Based on the results obtained so far, we have

$$\begin{aligned}\hat{\mathbf{y}} - \mathrm{E}(\mathbf{y}) &= \left[f(\mathrm{E}(\mathbf{x})) + D_{\tilde{\mathbf{x}}} f + \frac{1}{2!}D_{\tilde{\mathbf{x}}}^2 f + \cdots\right] \\ &\quad - \left[f(\mathrm{E}(\mathbf{x})) + \frac{1}{2!}\mathrm{E}(D_{\tilde{\mathbf{x}}}^2 f) + \frac{1}{4!}\mathrm{E}(D_{\tilde{\mathbf{x}}}^4 f) + \cdots\right] \\ &= \left[D_{\tilde{\mathbf{x}}} f + \frac{1}{2!}D_{\tilde{\mathbf{x}}}^2 f + \cdots\right] - \left[\frac{1}{2!}\mathrm{E}(D_{\tilde{\mathbf{x}}}^2 f) + \frac{1}{4!}\mathrm{E}(D_{\tilde{\mathbf{x}}}^4 f) + \cdots\right].\end{aligned}$$

Then, using the definition of $\mathbb{A} = \{(a, b)|a, b \in \mathbb{N}\} - \{(1, 1)\}$ where $\mathbb{N}$ is the set of natural numbers,

$$\begin{aligned}\mathbf{P}_{\mathbf{y}} &= \mathrm{E}[(\mathbf{y} - \mathrm{E}(\mathbf{y}))(\mathbf{y} - \mathrm{E}(\mathbf{y}))^T] \\ &= \mathrm{E}\Bigg[\Bigg\{\left(D_{\tilde{\mathbf{x}}} f + \frac{1}{2!}D_{\tilde{\mathbf{x}}}^2 f + \frac{1}{3!}D_{\tilde{\mathbf{x}}}^3 f + \cdots\right) \\ &\quad - \left(\frac{1}{2!}\mathrm{E}\left(D_{\tilde{\mathbf{x}}}^2 f\right) + \frac{1}{4!}\mathrm{E}\left(D_{\tilde{\mathbf{x}}}^4 f\right) + \frac{1}{6!}\mathrm{E}\left(D_{\tilde{\mathbf{x}}}^6 f\right) + \cdots\right)\Bigg\} \\ &\quad \times \Bigg\{\left(D_{\tilde{\mathbf{x}}} f + \frac{1}{2!}D_{\tilde{\mathbf{x}}}^2 f + \frac{1}{3!}D_{\tilde{\mathbf{x}}}^3 f + \cdots\right) \\ &\quad - \left(\frac{1}{2!}\mathrm{E}\left(D_{\tilde{\mathbf{x}}}^2 f\right) + \frac{1}{4!}\mathrm{E}\left(D_{\tilde{\mathbf{x}}}^4 f\right) + \frac{1}{6!}\mathrm{E}\left(D_{\tilde{\mathbf{x}}}^6 f\right) + \cdots\right)\Bigg\}^T\Bigg] \\ &= \mathrm{E}[(D_{\tilde{\mathbf{x}}} f)\,(D_{\tilde{\mathbf{x}}} f)^T] - \frac{1}{4}\mathrm{E}\left(D_{\tilde{\mathbf{x}}}^2 f\right)\mathrm{E}\left(D_{\tilde{\mathbf{x}}}^2 f\right)^T\end{aligned}$$

$$+\mathrm{E}\left[\sum_{\substack{i+j=\text{even}\\(i,j)\in\mathbb{A}}}\frac{1}{i!j!}\left(D_{\tilde{\mathbf{x}}}^{i}f\right)\left(D_{\tilde{\mathbf{x}}}^{j}f\right)^{T}\right]$$
$$-\left[\sum_{(i,j)\in\mathbb{A}}\frac{1}{(2i!)(2j!)}\mathrm{E}\left(D_{\tilde{\mathbf{x}}}^{2i}f\right)\mathrm{E}\left(D_{\tilde{\mathbf{x}}}^{2j}f\right)^{T}\right], \tag{13}$$

where all odd-powered terms in the expected value are zero. Here,

$$\mathrm{E}\left[(D_{\tilde{x}}f)(D_{\tilde{\mathbf{x}}}f)^{T}\right]=\mathrm{E}\left[\left(\sum_{i=1}^{c}\tilde{x}_i\frac{\partial f(\mathrm{E}(\mathbf{x}))}{\partial x_i}\right)\left(\sum_{i=1}^{c}\tilde{x}_i\frac{\partial f(\mathrm{E}(\mathbf{x}))}{\partial x_i}\right)^{T}\right]$$
$$=\mathrm{E}\left[\sum_{i=1}^{c}\sum_{j=1}^{c}\tilde{x}_i\frac{\partial f(\mathrm{E}(\mathbf{x}))}{\partial x_i}\frac{\partial f(\mathrm{E}(\mathbf{x}))^{T}}{\partial x_j}\tilde{x}_j\right]$$
$$=\sum_{i=1}^{c}\sum_{j=1}^{c}\mathbf{H}_i\mathrm{E}(\tilde{x}_i\tilde{x}_j)\mathbf{H}_j^{T}=\sum_{i=1}^{c}\sum_{j=1}^{c}\mathbf{H}_i P_{ij}\mathbf{H}_j^{T}=\mathbf{HPH}^{T}, \tag{14}$$

where $\mathbf{H}$ is the Jacobian matrix of $f(\mathbf{x})$ and $\mathbf{H}_i$ is the ith column of $\mathbf{H}$. Now we consider the approximate covariance matrix, $\hat{\mathbf{P}}_{\mathbf{y}}$, defined as

$$\hat{\mathbf{P}}_{\mathbf{y}}=\sum_{i=0}^{2c}W^{(i)}(\hat{\mathbf{y}}^{(i)}-\hat{\mathbf{y}})(\hat{\mathbf{y}}^{(i)}-\hat{\mathbf{y}})^{T}$$
$$=\frac{\lambda}{c+\lambda}(\hat{\mathbf{y}}^{(0)}-\hat{\mathbf{y}})(\hat{\mathbf{y}}^{(0)}-\hat{\mathbf{y}})^{T}+\frac{1}{2(c+\lambda)}\sum_{i=1}^{2c}(\hat{\mathbf{y}}^{(i)}-\hat{\mathbf{y}})(\hat{\mathbf{y}}^{(i)}-\hat{\mathbf{y}})^{T}.$$

Consider $\frac{\lambda}{c+\lambda}(\mathbf{y}^{(0)}-\hat{\mathbf{y}})(\mathbf{y}^{(0)}-\hat{\mathbf{y}})^{T}$ first. By the Taylor expansion, we get

$$\hat{\mathbf{y}}^{(0)}-\hat{\mathbf{y}}=f(\mathrm{E}(\mathbf{x}))-\left\{f(\mathrm{E}(\mathbf{x}))+\frac{1}{2(c+\lambda)}\sum_{i=1}^{2c}\left(\frac{1}{2!}D_{\tilde{\mathbf{x}}^{(i)}}^{2}f+\frac{1}{4!}D_{\tilde{\mathbf{x}}^{(i)}}^{4}f\cdots\right)\right\}$$
$$=-\frac{1}{2(c+\lambda)}\sum_{i=1}^{2c}\left(\frac{1}{2!}D_{\tilde{\mathbf{x}}^{(i)}}^{2}f+\frac{1}{4!}D_{\tilde{\mathbf{x}}^{(i)}}^{4}f\cdots\right)$$

Thus, we have

$$
\begin{aligned}
&\frac{\lambda}{c+\lambda}(\hat{\mathbf{y}}^{(0)}-\hat{\mathbf{y}})(\hat{\mathbf{y}}^{(0)}-\hat{\mathbf{y}})^T\\
&=\frac{\lambda}{c+\lambda}\frac{1}{4(c+\lambda)^2}\sum_{i=1}^{2c}\left(\frac{1}{2!}D^2_{\tilde{\mathbf{x}}^{(i)}}f\right)\sum_{i=1}^{2c}\left(\frac{1}{2!}D^2_{\tilde{\mathbf{x}}^{(i)}}f\right)^T\\
&\quad+\frac{\lambda}{c+\lambda}\sum_{(k,\ell)\in\mathbb{A}}\left[\frac{1}{(2k!)(2\ell!)}\frac{1}{4(c+\lambda)^2}\sum_{i=1}^{2c}\sum_{j=1}^{2c}(D^{2k}_{\tilde{\mathbf{x}}^{(i)}}f)(D^{2\ell}_{\tilde{\mathbf{x}}^{(j)}}f)^T\right]\\
&=\frac{\lambda}{c+\lambda}\frac{1}{4}\mathrm{E}[D^2_{\tilde{\mathbf{x}}}f]\mathrm{E}[D^2_{\tilde{\mathbf{x}}}f]^T\\
&\quad+\frac{\lambda}{c+\lambda}\sum_{(k,\ell)\in\mathbb{A}}\left[\frac{1}{(2k!)(2\ell!)}\frac{1}{4(c+\lambda)^2}\sum_{i=1}^{2c}\sum_{j=1}^{2c}(D^{2k}_{\tilde{\mathbf{x}}^{(i)}}f)(D^{2\ell}_{\tilde{\mathbf{x}}^{(j)}}f)^T\right].
\end{aligned}
$$

Next, consider $\frac{1}{2(c+\lambda)}\sum_{i=1}^{2c}(\mathbf{y}^{(i)}-\hat{\mathbf{y}})(\mathbf{y}^{(i)}-\hat{\mathbf{y}})^T$. Using (10) and the fact that

$$
\begin{aligned}
&\frac{1}{2(c+\lambda)}\sum_{i=1}^{2c}\left[(D_{\tilde{\mathbf{x}}^{(i)}}f)(D_{\tilde{\mathbf{x}}^{(i)}}f)^T\right]\\
&=\frac{1}{2(c+\lambda)}\sum_{i=1}^{2c}\sum_{k=1}^{2c}\sum_{\ell=1}^{2c}\left(\tilde{x}_k^{(i)}\frac{\partial f(\mathrm{E}(\mathbf{x}))}{\partial x_k}\right)\left(\tilde{x}_\ell^{(i)}\frac{\partial f(\mathrm{E}(\mathbf{x}))}{\partial x_\ell}\right)^T\\
&=\frac{1}{c+\lambda}\sum_{i=1}^{c}\sum_{k=1}^{c}\sum_{\ell=1}^{c}\left(\tilde{x}_k^{(i)}\frac{\partial f(\mathrm{E}(\mathbf{x}))}{\partial x_k}\right)\left(\tilde{x}_\ell^{(i)}\frac{\partial f(\mathrm{E}(\mathbf{x}))}{\partial x_\ell}\right)^T\\
&\quad\left(\text{using } \tilde{x}_j^{(i)}=-\tilde{x}_j^{(c+i)}\right)\\
&=\sum_{k=1}^{c}\sum_{\ell=1}^{c}P_{k\ell}\left(\frac{\partial f(\mathrm{E}(\mathbf{x}))}{\partial x_k}\right)\left(\frac{\partial f(\mathrm{E}(\mathbf{x}))}{\partial x_\ell}\right)^T=\mathbf{HPH}^T\\
&=\mathrm{E}\left[(D_{\tilde{x}}f)(D_{\tilde{\mathbf{x}}}f)^T\right]\quad\text{(By (14))},
\end{aligned}
$$

we can show

$$
\begin{aligned}
&\frac{1}{2(c+\lambda)}\sum_{i=1}^{2c}(\hat{\mathbf{y}}^{(i)}-\hat{\mathbf{y}})(\hat{\mathbf{y}}^{(i)}-\hat{\mathbf{y}})^T\\
&=\frac{1}{2(c+\lambda)}\sum_{i=1}^{2c}\Bigg[\Bigg\{\left(D_{\tilde{\mathbf{x}}^{(i)}}f+\frac{1}{2!}D^2_{\tilde{\mathbf{x}}^{(i)}}f+\frac{1}{3!}D^3_{\tilde{\mathbf{x}}^{(i)}}f+\cdots\right)\\
&\qquad-\frac{1}{2(c+\lambda)}\sum_{j=1}^{2c}\left(\frac{1}{2!}D^2_{\tilde{\mathbf{x}}^{(j)}}f+\frac{1}{4!}D^4_{\tilde{\mathbf{x}}^{(j)}}f+\cdots\right)\Bigg\}
\end{aligned}
$$

$$
\begin{aligned}
&\times\Bigg\{\left(D_{\tilde{\mathbf{x}}^{(i)}}f+\frac{1}{2!}D^2_{\tilde{\mathbf{x}}^{(i)}}f+\frac{1}{3!}D^3_{\tilde{\mathbf{x}}^{(i)}}f+\cdots\right)\\
&-\frac{1}{2(c+\lambda)}\sum_{j=1}^{2c}\left(\frac{1}{2!}D^2_{\tilde{\mathbf{x}}^{(j)}}f+\frac{1}{4!}D^4_{\tilde{\mathbf{x}}^{(j)}}f+\cdots\right)\Bigg\}^T\Bigg]\\
&=\mathrm{E}[(D_{\tilde{\mathbf{x}}}f)(D_{\tilde{\mathbf{x}}}f)^T]-\left(1+\frac{\lambda}{c+\lambda}\right)\frac{1}{4}\mathrm{E}[D^2_{\tilde{\mathbf{x}}}f]\mathrm{E}[D^2_{\tilde{\mathbf{x}}}f]^T\\
&+\frac{1}{2(c+\lambda)}\sum_{i=1}^{2c}\left[\sum_{\substack{k+\ell=\text{ even}\\(k,\ell)\in\mathbb{A}}}\frac{1}{k!\ell!}(D^k_{\tilde{\mathbf{x}}^{(i)}}f)(D^\ell_{\tilde{\mathbf{x}}^{(i)}}f)^T\right]\\
&-\sum_{(k,\ell)\in\mathbb{A}}\Bigg[\frac{1}{(2k!)(2\ell!)}\frac{1}{4(c+\lambda)^2}\left(1+\frac{\lambda}{c+\lambda}\right)\\
&\times\sum_{i=1}^{2c}\sum_{j=1}^{2c}(D^{2k}_{\tilde{\mathbf{x}}^{(i)}}f)(D^{2\ell}_{\tilde{\mathbf{x}}^{(j)}}f)^T\Bigg].
\end{aligned}\tag{15}
$$

Thus, we have

$$
\begin{aligned}
\hat{\mathbf{P}}_{\mathbf{y}}&=\frac{\lambda}{c+\lambda}(\hat{\mathbf{y}}^{(0)}-\hat{\mathbf{y}})(\hat{\mathbf{y}}^{(0)}-\hat{\mathbf{y}})^T+\frac{1}{2(c+\lambda)}\sum_{i=1}^{2c}(\hat{\mathbf{y}}^{(i)}-\hat{\mathbf{y}})(\hat{\mathbf{y}}^{(i)}-\hat{\mathbf{y}})^T\\
&=\mathrm{E}[(D_{\tilde{\mathbf{x}}}f)(D_{\tilde{\mathbf{x}}}f)^T]-\frac{1}{4}\mathrm{E}[D^2_{\tilde{\mathbf{x}}}f]\mathrm{E}[D^2_{\tilde{\mathbf{x}}}f]^T\\
&+\frac{1}{2(c+\lambda)}\sum_{i=1}^{2c}\left[\sum_{\substack{k+\ell=\text{ even}\\(k,\ell)\in\mathbb{A}}}\frac{1}{k!\ell!}(D^k_{\tilde{\mathbf{x}}^{(i)}}f)(D^\ell_{\tilde{\mathbf{x}}^{(i)}}f)^T\right]\\
&-\sum_{(k,\ell)\in\mathbb{A}}\left[\frac{1}{(2k!)(2\ell!)}\frac{1}{4(c+\lambda)^2}\sum_{i=1}^{2c}\sum_{j=1}^{2c}(D^{2k}_{\tilde{\mathbf{x}}^{(i)}}f)(D^{2\ell}_{\tilde{\mathbf{x}}^{(j)}}f)^T\right].
\end{aligned}\tag{16}
$$

Now we consider the covariance matrix $\mathbf{P}_{\mathbf{xy}}$. Then,

$$
\begin{aligned}
\mathbf{P}_{\mathbf{xy}}&=\mathrm{E}[(\mathbf{x}-\mathrm{E}(\mathbf{x}))(\mathbf{y}-\mathrm{E}(\mathbf{y}))^T]=\mathrm{E}[(\tilde{\mathbf{x}})(\mathbf{y}-\mathrm{E}(\mathbf{y}))^T]\\
&=\mathrm{E}\left[\tilde{\mathbf{x}}\,(D_{\tilde{\mathbf{x}}}f)^T\right]+\sum_{i=1}^{\infty}\frac{1}{(2i+1)!}\mathrm{E}\left[\tilde{\mathbf{x}}\left(D^{2i+1}_{\tilde{\mathbf{x}}}f\right)^T\right]\\
&=\mathbf{PH}^T+\sum_{i=1}^{\infty}\frac{1}{(2i+1)!}\mathrm{E}\left[\tilde{\mathbf{x}}\left(D^{2i+1}_{\tilde{\mathbf{x}}}f\right)^T\right].
\end{aligned}\tag{17}
$$

Using $\mathbf{x}^{(0)} = \mathrm{E}(\mathbf{x})$, the approximate covariance matrix, $\hat{\mathbf{P}}_{\mathbf{xy}}$, equals

$$\begin{aligned}\hat{\mathbf{P}}_{\mathbf{xy}} &= \frac{1}{2(c+\lambda)}\sum_{i=0}^{2c}(\hat{\mathbf{x}}^{(i)} - \mathrm{E}(\mathbf{x}))(\hat{\mathbf{y}}^{(i)} - \hat{\mathbf{y}})^T \\ &= \frac{1}{2(c+\lambda)}\sum_{i=1}^{2c}\tilde{\mathbf{x}}^{(i)}\left(D_{\tilde{\mathbf{x}}^{(i)}} f\right)^T \\ &\quad + \sum_{k=1}^{\infty}\frac{1}{(2k+1)!}\frac{1}{2(c+\lambda)}\sum_{i=1}^{2c}\tilde{\mathbf{x}}^{(i)}\left(D_{\tilde{\mathbf{x}}^{(i)}}^{2k+1} f\right)^T \\ &= \mathbf{P}\mathbf{H}^T + \sum_{k=1}^{\infty}\frac{1}{(2k+1)!}\frac{1}{2(c+\lambda)}\sum_{i=1}^{2c}\tilde{\mathbf{x}}^{(i)}\left(D_{\tilde{\mathbf{x}}^{(i)}}^{2k+1} f\right)^T. \qquad (18)\end{aligned}$$

Appendix B: Derivation of results for Theorem 2

In this section, $|A|$ indicates the absolute value of A, where A may be a scalar, a vector, or a matrix. When A is a vector (matrix), $|A|$ is a vector (matrix) consisting of the absolute values of A's components. In addition, when "$\leq$" is used for a vector (matrix), it implies that every component of the left vector (matrix) is less than or equal to the corresponding component of the right vector (matrix). For simplicity, we assume that the distribution of $\mathbf{x}$ is symmetric about its mean. A general case, which does not assume the symmetric distribution of $\mathbf{x}$ about its mean, can be shown similarly. From the Taylor series in Section A of Appendix and Condition 1 of Theorem 2, we have

$$\begin{aligned}\mathrm{E}(\mathbf{y}) - \hat{\mathbf{y}} &= \left[f(\mathrm{E}(\mathbf{x})) + \frac{1}{2!}\sum_{i=1}^{c}\sum_{j=1}^{c}P_{ij}\frac{\partial^2}{\partial x_i \partial x_j}f(\mathrm{E}(\mathbf{x})) + \frac{1}{4!}\mathrm{E}[D_{\tilde{\mathbf{x}}}^4 f] + \cdots\right] \\ &\quad - \left[f(\mathrm{E}(\mathbf{x})) + \frac{1}{2}\sum_{i=1}^{c}\sum_{j=1}^{c}P_{ij}\frac{\partial^2}{\partial x_i \partial x_j}f(\mathrm{E}(\mathbf{x}))\right. \\ &\quad \left. + \frac{1}{2(c+\lambda)}\sum_{j=1}^{2c}\left(\frac{1}{4!}D_{\tilde{\mathbf{x}}^{(j)}}^4 f + \frac{1}{6!}D_{\tilde{\mathbf{x}}^{(j)}}^6 f \cdots\right)\right] \\ &= \sum_{i=2}^{\infty}\frac{1}{(2i)!}\mathrm{E}[D_{\tilde{\mathbf{x}}}^{2i} f] - \sum_{i=2}^{\infty}\frac{1}{2(c+\lambda)}\sum_{j=1}^{2c}\frac{1}{(2i)!}D_{\tilde{\mathbf{x}}^{(j)}}^{2i} f\end{aligned}$$

For $1 \leq \ell \leq q$, we have

$$\left|\frac{1}{(2i)!}\mathrm{E}[D_{\tilde{\mathbf{x}}}^{2i} f_\ell]\right| = \frac{1}{(2i)!}\left|\mathrm{E}\left(\sum_{j=1}^{c}\tilde{x}_j\frac{\partial}{\partial x_j}\right)^{2i} f_\ell\right|$$

$$\leq \frac{1}{(2i)!} \sum_{1\leq k_1,\ldots,k_c\leq 2i} \left[\binom{2i}{k_1,\ldots,k_c}\left(h^i M^i \left|\frac{\partial^{2i} f_\ell(\mathrm{E}(\mathbf{x}))}{\partial x_1^{k_1}\ldots\partial x_c^{k_c}}\right|\right)\right]$$
$$\equiv h^i a_{i,\ell}. \tag{19}$$

Let $\mathbf{a}_i = (a_{i,1}\ldots a_{i,q})$. Similarly,

$$\frac{1}{2(c+\lambda)}\sum_{j=1}^{2c}\frac{1}{(2i)!}D^{2i}_{\tilde{\mathbf{x}}^{(j)}} f$$
$$= \frac{1}{(c+\lambda)}\sum_{j=1}^{c}\frac{1}{(2i)!}\left(\sum_{\ell=1}^{c}\sqrt{(c+\lambda)P}_{\ell j}\frac{\partial}{\partial x_\ell}\right)^{2i} f$$
$$= \frac{h^i}{(2i)!}\sum_{j=1}^{c}(c+\lambda)^{i-1}\left(\sum_{1\leq k_1,\ldots,k_c\leq 2i}\binom{2i}{k_1,\ldots,k_c}\right.$$
$$\left.\times(\sqrt{P^*}_{1j})^{k_1}\ldots(\sqrt{P^*}_{cj})^{k_c}\frac{\partial^{2i} f(\mathrm{E}(\mathbf{x}))}{\partial x_1^{k_1}\ldots\partial x_c^{k_n}}\right)$$
$$\equiv h^i \mathbf{b}_i, \tag{20}$$

where P^*_{ij} is the (i,j) component of $\mathbf{P}^*$. Since $\sum_{i=1}^{\infty} K^i/i!$ is finite, Condition 2 of Theorem 2 implies that all components of $\sum_{i=1}^{\infty} h^i \mathbf{a}_i$ and $\sum_{i=1}^{\infty} h^i \mathbf{b}_i$ defined in (19) and (20) respectively are finite. Thus, we have

$$|\mathrm{E}(\mathbf{y}) - \hat{\mathbf{y}}| \leq \sum_{i=2}^{\infty} h^i(|\mathbf{a}_i| + |\mathbf{b}_i|) = h^2\sum_{i=2}^{\infty} h^{i-2}(|\mathbf{a}_i| + |\mathbf{b}_i|) = O(h^2). \tag{21}$$

On the other hand, linearization uses $\hat{\mathbf{y}}_L = f(\mathrm{E}(\mathbf{x}))$. It can be similarly shown that

$$\mathrm{E}(\mathbf{y}) - \hat{\mathbf{y}}_L = O(h). \tag{22}$$

Now, we turn our attention to $\mathbf{P}_\mathbf{y} - \hat{\mathbf{P}}_\mathbf{y}$. From the results so far, we get

$$\mathbf{P}_\mathbf{y} - \hat{\mathbf{P}}_\mathbf{y}$$
$$= \mathrm{E}\left[\sum_{\substack{i+j=\text{ even}\\(i,j)\in\mathbb{A}}}\frac{1}{i!j!}(D^i_{\tilde{\mathbf{x}}} f)(D^j_{\tilde{\mathbf{x}}} f)^T\right] - \left[\sum_{(i,j)\in\mathbb{A}}\frac{1}{(2i!)(2j!)}\mathrm{E}(D^{2i}_{\tilde{\mathbf{x}}} f)\mathrm{E}(D^{2j}_{\tilde{\mathbf{x}}} f)^T\right]$$
$$-\frac{1}{2(c+\lambda)}\sum_{i=1}^{2c}\left[\sum_{\substack{j+\ell=\text{ even}\\(j,\ell)\in\mathbb{A}}}\frac{1}{j!\ell!}\left(D^j_{\tilde{\mathbf{x}}^{(i)}} f\right)\left(D^\ell_{\tilde{\mathbf{x}}^{(i)}} f\right)^T\right]$$

$$+ \sum_{(j,\ell)\in\mathbb{A}} \left[\frac{1}{(2j!)(2\ell!)} \frac{1}{4(c+\lambda)^2} \sum_{i=1}^{2c} \sum_{m=1}^{2c} \left(D^{2j}_{\tilde{\mathbf{x}}^{(i)}} f \right) \left(D^{2\ell}_{\tilde{\mathbf{x}}^{(m)}} f \right)^T \right].$$

First of all, we consider the first term $\mathrm{E}\left[\frac{1}{i!j!}(D^i_{\tilde{\mathbf{x}}} f)(D^j_{\tilde{\mathbf{x}}} f)^T\right]$ where $i + j$ is even. Then, we have

$$\begin{aligned}
&\mathrm{E}\left[\frac{1}{i!j!}(D^i_{\tilde{\mathbf{x}}} f)(D^j_{\tilde{\mathbf{x}}} f)^T\right] \\
&\quad = \mathrm{E}\left[\left(\frac{1}{i!} \sum_{1\le k_1,\ldots,k_c\le i} \binom{i}{k_1,\ldots,k_c} \tilde{x}_1^{k_1}\ldots\tilde{x}_c^{k_c} \frac{\partial^i f(\mathrm{E}(\mathbf{x}))}{\partial x_1^{k_1}\ldots\partial x_c^{k_c}}\right)\right. \\
&\qquad \left. \times \left(\frac{1}{j!} \sum_{1\le \ell_1,\ldots\ell_c\le j} \binom{j}{k_1,\ldots,k_c} \tilde{x}_1^{\ell_1}\ldots\tilde{x}_c^{\ell_c} \frac{\partial^j f(\mathrm{E}(\mathbf{x}))}{\partial x_1^{\ell_1}\ldots\partial x_c^{\ell_c}}\right)^T\right].
\end{aligned}$$

Let the ith component of a function f be f_i. Then, the (a, b) component of $\mathrm{E}\left[\frac{1}{i!j!}(D^i_{\tilde{x}} f)(D^j_{\tilde{x}} f)^T\right]$, $\mathrm{E}\left[\frac{1}{i!j!}(D^i_{\tilde{x}} f)(D^j_{\tilde{x}} f)^T\right]_{ab}$ satisfies

$$\begin{aligned}
&\left|\mathrm{E}\left[\frac{1}{i!j!}(D^i_{\tilde{\mathbf{x}}} f)(D^j_{\tilde{\mathbf{x}}} f)^T\right]_{ab}\right| \\
&= \left|\mathrm{E}\left[\left(\frac{1}{i!} \sum_{1\le k_1,\ldots,k_c\le i} \binom{i}{k_1,\ldots,k_c} \tilde{x}_1^{k_1}\ldots\tilde{x}_c^{k_c} \frac{\partial^i f_a(\mathrm{E}(\mathbf{x}))}{\partial x_1^{k_1}\ldots\partial x_c^{k_c}}\right)\right.\right. \\
&\quad \left.\left. \times \left(\frac{1}{j!} \sum_{1\le \ell_1,\ldots,\ell_c\le j} \binom{j}{\ell_1,\ldots,\ell_c} \tilde{x}_1^{\ell_1}\ldots\tilde{x}_c^{\ell_c} \frac{\partial^j f_b(\mathrm{E}(\mathbf{x}))}{\partial x_1^{\ell_1}\ldots\partial x_c^{\ell_c}}\right)\right]\right| \\
&= \left|\mathrm{E}\left[\frac{1}{i!j!} \sum_{\substack{1\le k_1,\ldots,k_c\le i \\ 1\le \ell_1,\ldots,\ell_c\le j}} \binom{i}{k_1,\ldots,k_c}\binom{j}{\ell_1,\ldots,\ell_c} \tilde{x}_1^{k_1+\ell_1}\ldots\tilde{x}_c^{k_c+\ell_c}\right.\right. \\
&\quad \left.\left. \times \frac{\partial^i f_a(\mathrm{E}(\mathbf{x}))}{\partial x_1^{k_1}\ldots\partial x_c^{k_c}} \frac{\partial^j f_b(\mathrm{E}(\mathbf{x}))}{\partial x_1^{\ell_1}\ldots\partial x_n^{\ell_c}}\right]\right| \\
&\le \frac{h^{(i+j)/2} M^{(i+j)/2}}{i!j!} \sum_{\substack{1\le k_1,\ldots,k_c\le i \\ 1\le \ell_1,\ldots,\ell_c\le j}} \binom{i}{k_1,\ldots,k_c}\binom{j}{\ell_1,\ldots,\ell_c} \\
&\quad \times \left|\frac{\partial^i f_a(\mathrm{E}(\mathbf{x}))}{\partial x_1^{k_1}\ldots\partial x_c^{k_c}} \frac{\partial^j f_b(\mathrm{E}(\mathbf{x}))}{\partial x_1^{\ell_1}\ldots\partial x_c^{\ell_c}}\right| \\
&\equiv h^{(i+j)/2} R^{ij}_{ab}. \qquad (23)
\end{aligned}$$

Thus, $\left|\mathrm{E}\left[\frac{1}{i!j!}(D^i_{\tilde{\mathbf{x}}} f)(D^j_{\tilde{\mathbf{x}}} f)^T\right]\right| \leq h^{(i+j)/2}\mathbf{R}^{ij}$. Then, we obtain

$$\left|\mathrm{E}\left[\sum_{\substack{i+j=\text{ even}\\(i,j)\in\mathbb{A}}} \frac{1}{i!j!}\left(D^i_{\tilde{\mathbf{x}}} f\right)\left(D^j_{\tilde{\mathbf{x}}} f\right)^T\right]\right| \leq \sum_{\substack{i+j=\text{ even}\\(i,j)\in\mathbb{A}}} h^{(i+j)/2}\mathbf{R}^{ij} = h^2 \sum_{\substack{i+j=\text{ even}\\(i,j)\in\mathbb{A}}} h^{(i+j)/2-2}\mathbf{R}^{ij}. \tag{24}$$

By (19), the second term satisfies

$$\sum_{(i,j)\in\mathbb{A}} \frac{1}{(2i!)(2j!)}\left|\mathrm{E}\left(D^{2i}_{\tilde{\mathbf{x}}} f\right)\mathrm{E}\left(D^{2j}_{\tilde{\mathbf{x}}} f\right)^T\right| \leq \sum_{(i,j)\in\mathbb{A}} h^i\mathbf{a}_i(h^i\mathbf{a}_j)^T = h^2\sum_{(i,j)\in\mathbb{A}} h^{i+j-2}\mathbf{a}_i(\mathbf{a}_j)^T. \tag{25}$$

Next, let us consider the third term $\frac{1}{2(c+\lambda)}\sum_{i=1}^{2c}\frac{1}{j!\ell!}(D^j_{\tilde{\mathbf{x}}^{(i)}} f)(D^\ell_{\tilde{\mathbf{x}}^{(i)}} f)^T$ where $j+\ell$ is even.

$$\begin{aligned}
&\frac{1}{2(c+\lambda)}\sum_{i=1}^{2c}\frac{1}{j!\ell!}(D^j_{\tilde{\mathbf{x}}^{(i)}} f)(D^\ell_{\tilde{\mathbf{x}}^{(i)}} f)^T = \frac{1}{c+\lambda}\sum_{i=1}^{c}\frac{1}{j!\ell!}(D^j_{\tilde{\mathbf{x}}^{(i)}} f)(D^\ell_{\tilde{\mathbf{x}}^{(i)}} f)^T\\
&=\frac{1}{c+\lambda}\sum_{i=1}^{c}\left[\frac{1}{j!}\left(\sum_{r=1}^{c}\sqrt{(c+\lambda)P}_{ri}\frac{\partial}{\partial x_\ell}\right)^j f\right]\left[\frac{1}{\ell!}\left(\sum_{s=1}^{c}\sqrt{(c+\lambda)P}_{si}\frac{\partial}{\partial x_\ell}\right)^\ell f\right]^T\\
&=\frac{(c+\lambda)^{(j+\ell)/2}}{c+\lambda}\sum_{i=1}^{c}\left[\frac{1}{j!}\sum_{1\leq u_1,\ldots,u_c\leq j}\binom{j}{u_1,\ldots,u_c}(\sqrt{P}_{1i})^{u_1}\ldots(\sqrt{P}_{ci})^{u_c}\right.\\
&\quad\left.\times\frac{\partial^j f(\mathrm{E}(\mathbf{x}))}{\partial x_1^{u_1}\ldots\partial x_c^{u_c}}\right]\\
&\quad\times\left[\frac{1}{\ell!}\sum_{1\leq v_1,\ldots,v_c\leq\ell}\binom{\ell}{v_1,\ldots,v_c}(\sqrt{P}_{1i})^{v_1}\ldots(\sqrt{P}_{ci})^{v_c}\frac{\partial^\ell f(\mathrm{E}(\mathbf{x}))}{\partial x_1^{v_1}\ldots\partial x_c^{v_c}}\right]^T.
\end{aligned}$$

Similar to $\mathrm{E}\left[\frac{1}{i!j!}(D^i_{\tilde{\mathbf{x}}} f)(D^j_{\tilde{\mathbf{x}}} f)^T\right]_{ab}$, we have

$$\left[\frac{1}{2(c+\lambda)}\sum_{i=1}^{2c}\frac{1}{j!\ell!}(D^j_{\tilde{\mathbf{x}}^{(i)}} f)(D^\ell_{\tilde{\mathbf{x}}^{(i)}} f)^T\right]_{ab}$$

$$= h^{(j+\ell)/2} \sum_{i=1}^{c} \Bigg[\frac{(c+\lambda)^{(j+\ell)/2-1}}{j!\ell!} \sum_{\substack{1 \le u_1,\ldots,u_c \le j \\ 1 \le v_1,\ldots,v_c \le \ell}} \binom{j}{u_1,\ldots,u_c} \binom{\ell}{v_1,\ldots,v_c} \cdot$$

$$\times (\sqrt{P^*}_{1i})^{u_1+v_1} \ldots (\sqrt{P^*}_{ni})^{u_c+v_c} \frac{\partial^j f_a(\mathrm{E}(\mathbf{x}))}{\partial x_1^{u_1} \ldots \partial x_c^{u_c}} \frac{\partial^\ell f_b(\mathrm{E}(\mathbf{x}))}{\partial x_1^{v_1} \ldots \partial x_c^{v_c}} \Bigg]$$

$$\equiv h^{(j+\ell)/2} T_{ab}^{j\ell}.$$

Thus, we obtain

$$\frac{1}{2(c+\lambda)} \sum_{i=1}^{2c} \Bigg[\sum_{\substack{j+\ell=\text{ even} \\ (j,\ell)\in\mathbb{A}}} \frac{1}{j!\ell!} (D_{\tilde{\mathbf{x}}^{(i)}}^{j} f)(D_{\tilde{\mathbf{x}}^{(i)}}^{\ell} f)^T \Bigg]$$
$$= \sum_{\substack{j+\ell=\text{ even} \\ (j,\ell)\in\mathbb{A}}} h^{(i+j)/2} \mathbf{T}^{j\ell} = h^2 \sum_{\substack{j+\ell=\text{ even} \\ (j,\ell)\in\mathbb{A}}} h^{(i+j)/2-2} \mathbf{T}^{j\ell}. \tag{26}$$

For the last term of $\mathbf{P_y} - \hat{\mathbf{P}}_\mathbf{y}$, $\sum_{(j,\ell)\in\mathbb{A}} \big[\frac{1}{(2j!)(2\ell!)} \frac{1}{4(c+\lambda)^2} \sum_{i=1}^{2c} \sum_{m=1}^{2c} (D_{\tilde{\mathbf{x}}^{(i)}}^{2j} f)$ $(D_{\tilde{\mathbf{x}}^{(m)}}^{2\ell} f)^T\big]$, by (20), we have

$$\sum_{(j,\ell)\in\mathbb{A}} \Bigg[\frac{1}{(2j!)(2\ell!)} \frac{1}{4(c+\lambda)^2} \sum_{i=1}^{2c} \sum_{m=1}^{2c} (D_{\tilde{\mathbf{x}}^{(i)}}^{2j} f)(D_{\tilde{\mathbf{x}}^{(m)}}^{2\ell} f)^T \Bigg]$$
$$= \sum_{(j,\ell)\in\mathbb{A}} \Bigg[\Bigg(\sum_{i=1}^{2c} \frac{1}{(2j)!} \frac{1}{2(c+\lambda)} D_{\tilde{\mathbf{x}}^{(i)}}^{2j} f \Bigg) \Bigg(\sum_{m=1}^{2c} \frac{1}{(2\ell)!} \frac{1}{2(c+\lambda)} D_{\tilde{\mathbf{x}}^{(m)}}^{2\ell} f \Bigg)^T \Bigg]$$
$$= \sum_{(j,\ell)\in\mathbb{A}} h^j \mathbf{b}_j (h^\ell \mathbf{b}_\ell)^T = h^2 \sum_{(j,\ell)\in\mathbb{A}} h^{j+\ell-2} \mathbf{b}_j (\mathbf{b}_\ell)^T. \tag{27}$$

Combining (24)–(27), we obtain

$$\mathbf{P_y} - \hat{\mathbf{P}}_\mathbf{y} = O(h^2), \tag{28}$$

where Condition 2 of Theorem 2 implies that all components of $\sum_{\substack{i+j=\text{ even} \\ (i,j)\in\mathbb{A}}} h^{(i+j)/2}$ $\mathbf{R}^{ij}$, $\sum_{(i,j)\in\mathbb{A}} h^{i+j} \mathbf{a}_i (\mathbf{a}_j)^T$, $\sum_{\substack{j+\ell=\text{ even} \\ (j,\ell)\in\mathbb{A}}} h^{(j+\ell)/2} \mathbf{T}^{j\ell}$, and $\sum_{(j,\ell)\in\mathbb{A}} h^{j+\ell-2} \mathbf{b}_j (\mathbf{b}_\ell)^T$ are finite. By (19), we have

$$\frac{1}{4} \mathrm{E}\left[D_{\tilde{\mathbf{x}}}^2 f\right] \mathrm{E}\left[D_{\tilde{\mathbf{x}}}^2 f\right]^T = h^2 \mathbf{a}_1 (\mathbf{a}_1)^T.$$

In linearization, $\hat{\mathbf{P}}_{\mathbf{y},L} = \mathbf{H}\mathbf{P}_{\mathbf{x}}\mathbf{H}^T$. Thus, it can be similarly shown that

$$\mathbf{P}_{\mathbf{y}} - \hat{\mathbf{P}}_{\mathbf{y},L} = O(h^2). \tag{29}$$

Now, we know

$$\begin{aligned}\mathbf{P}_{\mathbf{xy}} - \hat{\mathbf{P}}_{\mathbf{xy}} &= \sum_{i=1}^{\infty} \frac{1}{(2i+1)!} \mathrm{E}\left[\tilde{\mathbf{x}} \left(D_{\tilde{\mathbf{x}}}^{2i+1} f\right)^T\right] \\ &\quad - \sum_{i=1}^{\infty} \frac{1}{(2i+1)!} \frac{1}{2(c+\lambda)} \sum_{j=1}^{2c} \tilde{\mathbf{x}}^{(j)} \left(D_{\tilde{\mathbf{x}}^{(j)}}^{2i+1} f\right)^T.\end{aligned}$$

We consider $\frac{1}{(2i+1)!}\mathrm{E}\left[\tilde{\mathbf{x}}\left(D_{\tilde{\mathbf{x}}}^{2i+1} f\right)^T\right]$ first.

$$\begin{aligned}&\frac{1}{(2i+1)!}\mathrm{E}\left[\tilde{\mathbf{x}}\left(D_{\tilde{\mathbf{x}}}^{2i+1} f\right)^T\right] \\ &= \frac{1}{(2i+1)!}\mathrm{E}\left[\tilde{\mathbf{x}}\left(\sum_{1\le k_1,\ldots,k_c\le 2i+1}\binom{2i+1}{k_1,\ldots,k_c}\tilde{x}_1^{k_1}\ldots\tilde{x}_c^{k_c}\frac{\partial^{2i+1} f(\mathrm{E}(\mathbf{x}))}{\partial x_1^{k_1}\ldots\partial x_c^{k_c}}\right)^T\right].\end{aligned}$$

As before, we can obtain

$$\begin{aligned}&\frac{1}{(2i+1)!}\left|\mathrm{E}\left[\tilde{\mathbf{x}}\left(D_{\tilde{\mathbf{x}}}^{2i+1} f\right)^T\right]_{ab}\right| \\ &= \frac{1}{(2i+1)!}\left|\mathrm{E}\left[\tilde{x}_a\sum_{1\le k_1,\ldots,k_c\le 2i+1}\binom{2i+1}{k_1,\ldots,k_c}\tilde{x}_1^{k_1}\ldots\tilde{x}_c^{k_c}\frac{\partial^{2i+1} f_b(\mathrm{E}(\mathbf{x}))}{\partial x_1^{k_1}\ldots\partial x_c^{k_c}}\right]\right| \\ &= \frac{1}{(2i+1)!}\left|\mathrm{E}\left[\sum_{1\le k_1,\ldots,k_c\le 2i+1}\binom{2i+1}{k_1,\ldots,k_c}\tilde{x}_1^{k_1}\ldots\tilde{x}_a^{k_a+1}\ldots\tilde{x}_c^{k_c}\frac{\partial^{2i+1} f_b(\mathrm{E}(\mathbf{x}))}{\partial x_1^{k_1}\ldots\partial x_c^{k_c}}\right]\right| \\ &\le \frac{h^{i+1}M^{i+1}}{(2i+1)!}\sum_{1\le k_1,\ldots,k_c\le 2i+1}\left[\binom{2i+1}{k_1,\ldots,k_c}\left|\frac{\partial^{2i+1} f_b(\mathrm{E}(\mathbf{x}))}{\partial x_1^{k_1}\ldots\partial x_c^{k_c}}\right|\right] \\ &\equiv h^{i+1}U_{i,ab}.\end{aligned}$$

Thus, we get

$$\sum_{i=1}^{\infty}\frac{1}{(2i+1)!}\left|\mathrm{E}\left[\tilde{\mathbf{x}}\left(D_{\tilde{\mathbf{x}}}^{2i+1} f\right)^T\right]\right| \le \sum_{i=1}^{\infty} h^{i+1}\mathbf{U}_i = h^2\sum_{i=1}^{\infty} h^{i-1}\mathbf{U}_i.$$

Furthermore,

$$\left[\frac{1}{(2i+1)!}\frac{1}{2(c+\lambda)}\sum_{j=1}^{2c}\tilde{\mathbf{x}}^{(j)}\left(D_{\tilde{\mathbf{x}}^{(j)}}^{2i+1}f\right)^T\right]_{ab}$$
$$= \frac{h^{i+1}}{(2i+1)!}\frac{(c+\lambda)^{i+1}}{c+\lambda}\sum_{j=1}^{c}(\sqrt{P^*}_{aj})$$
$$\times\left(\sum_{1\le k_1,\ldots,k_c\le 2i+1}\binom{2i+1}{k_1,\ldots,k_c}(\sqrt{P^*}_{1j})^{k_1}\ldots(\sqrt{P^*}_{cj})^{k_c}\frac{\partial^{2i+1}f_b(\mathrm{E}(\mathbf{x}))}{\partial x_1^{k_1}\ldots\partial x_c^{k_c}}\right)$$
$$\equiv h^{i+1}V_{i,ab}.$$

Thus, we have

$$\sum_{i=1}^{\infty}\frac{1}{(2i+1)!}\frac{1}{2(c+\lambda)}\sum_{j=1}^{2c}\tilde{\mathbf{x}}^{(j)}\left(D_{\tilde{\mathbf{x}}^{(j)}}^{2i+1}f\right)^T = \sum_{i=1}^{\infty}h^{i+1}\mathbf{V}_i = h^2\sum_{i=1}^{\infty}h^{i-1}\mathbf{V}_i.$$

Therefore, we obtain

$$\begin{aligned}|\mathbf{P_{xy}}-\hat{\mathbf{P}}_{\mathbf{xy}}| &= \left|\sum_{i=1}^{\infty}\frac{1}{(2i+1)!}\mathrm{E}\left[\tilde{\mathbf{x}}\left(D_{\tilde{\mathbf{x}}}^{2i+1}f\right)^T\right]\right.\\ &\qquad\left.-\sum_{i=1}^{\infty}\frac{1}{(2i+1)!}\frac{1}{2(c+\lambda)}\sum_{j=1}^{2c}\tilde{\mathbf{x}}^{(j)}\left(D_{\tilde{\mathbf{x}}^{(j)}}^{2i+1}f\right)^T\right|\\ &\le h^2\sum_{i=1}^{\infty}h^{i-1}(|\mathbf{U}_i|+|\mathbf{V}_i|) = O(h^2),\end{aligned} \tag{30}$$

where Condition 2 of Theorem 2 implies that all components of $\sum_{i=1}^{\infty}h^{i+1}\mathbf{U}_i$ and $\sum_{i=1}^{\infty}h^{i+1}\mathbf{V}_i$ defined in (30) are finite. In linearization, $\hat{\mathbf{P}}_{\mathbf{xy},L}=\mathbf{P_x}\mathbf{H}^T$. Thus, it can be similarly shown that

$$|\mathbf{P_{xy}}-\hat{\mathbf{P}}_{\mathbf{xy},L}| \le h^2\sum_{i=1}^{\infty}h^{i-1}|\mathbf{U}_i| = O(h^2). \tag{31}$$

Acknowledgments This work was partially supported by the U.S. National Science Foundation (DMS-1021292). The authors would like to thank an anonymous reviewer for valuable comments which helped improve the manuscript.

References

Ahn, K. W., Chan, K. S. (2011). *Approximate conditional least squares estimation of a nonlinear state-space model via unscented Kalman filter.* Technical report# 405 Department of Statistics and Actuarial Science, University of Iowa.

Boyce, W. E., DiPrima, R. C. (2004). *Elementary differential equations and boundary value problems* (8th ed.). New Jersey: Wiley.

Diekmann, O., Heesterbeek, J. A. P. (2000). *Mathematical epidemiology of infectious diseases: Model building, analysis and interpretation*. West Sussex: Wiley

Foster, J., Richards, F. B. (1991). The Gibbs phenomenon for piecewise-linear approximation. *The American Mathematical Monthly*, *98*, 47–49.

Holmquist, B. (1988). Moments and cumulants of the multivariate normal distribution. *Stochastic Analysis and Applications*, *6*, 273–278.

Isserlis, L. (1918). On a formula for the product-moment coefficient of any order of a noraml frequency distribution in any number of variables. *Biometrika*, *12*, 134–139.

Julier, S. J., Uhlmann, J. K. (1997). A new extension of the Kalman filter to nonlinear systems. *Proceedings of AeroSense: The 11th International symposium on aerospace/defense sensing, simulation and controls. Orlando, Florida, Vol Multi Sensor Fusion, Tracking and Resource Management II*.

Julier, S. J., Uhlmann, J. K. (2004). Unscented filtering and nonlinear estimation. *IEEE Review*, *92*, 401–422.

Milstein, G. N., Tretyakov, M. V. (2004). *Stochastic numerics for mathematical physics*. Berlin, Heidelberg: Springer.

Shim, H. T., Park, C. H. (2005). The survey of Gibbs phenomenon from Fourier series to hybrid sampling series. *Journal of Applied Mathematics and Computing*, *17*, 719–736.

Simon, D. (2006). *Optimal state estimation*. New Jersey: Wiley.

Triantafyllopoulos, K. (2003). On the central moments of the multidimensional gaussian distribution. *The Mathematical Scientist*, *28*, 125–128.

Wan, E. A., van der Merwe, R. (2000). The unscented Kalman filter for nonlinear estimation. *Adaptive systems for signal processing, communications, and control symposium (AS-SPCC) The IEEE*, Oct 2000 Lake Louise, Alberta, pp. 153–158.