Approximate Conditional Least Squares Estimation of a Nonlinear State-Space Model via Unscented Kalman Filter

Kwang Woo AHN

Division of Biostatistics

Medical College of Wisconsin, Milwaukee, WI 53226

email: kwooahn@mcw.edu

Kung–Sik CHAN

Department of Statistics and Actuarial Science

The University of Iowa, Iowa City, Iowa 52242

email: kung-sik-chan@uiowa.edu

July 22, 2011

Abstract

Nonlinear state-space models driven by differential equations have been widely used in science. Their statistical inference generally requires computing the mean and co-variance 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 by Julier and Uhlmann (1997) to overcome these difficulties, but it lacks of 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). Simulation based methods such as sequential Monte Carlo methods have been popular for estimating such models, but it is computationally expensive (Doucet, deFreitas and Gordon 2001). In the engineering literature, the extended Kalman filter (EKF) has been proposed as a fast alternative, but it approximates the nonlinear system by its first order Taylor expansion, which is subject to rapidly increasing approximation errors with the degree of nonlinearity. To overcome these difficulties, Julier and Uhlmann (1997) proposed the unscented Kalman filter (UKF). 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). In contrast with Monte Carlo methods, the UT makes use of a small number of deterministic "sigma points" in estimating the mean and covariance matrix of some nonlinearly transformed random variable. Empirical works suggest that the UKF is a promising technique with satisfactory performance, see Julier and Uhlmann (1997), Julier and Uhlmann (2004), and Wan and van der Merwe (2000). However, theoretical justification of the UKF is lacking.

2

Since the UT plays a pivotal role in the UKF, it is essential to study the theoretical properties of the UT. As the UT is used for approximately computing the mean and covariance matrix of some nonlinear transformation of the state vector over a small time step, denoted by h, the main issue we address here concerns the error rate of the UT approximation as $h \to 0$. In Section 2, we elaborate the definition of the UT. The error rates of the UT are then derived in Section 3. We conclude briefly in Section 4.

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:

$$\hat{\mathbf{x}}^{(0)} = \mathrm{E}(\mathbf{x}), \ \hat{\mathbf{x}}^{(i)} = \mathrm{E}(\mathbf{x}) + \breve{\mathbf{x}}^{(i)}, i = 1, \ldots, 2c,$$

$$\breve{\mathbf{x}}^{(j)} = \left(\sqrt{(c+\lambda)\mathbf{P}}\right)^T_j, \ \breve{\mathbf{x}}^{(c+j)} = -\left(\sqrt{(c+\lambda)\mathbf{P}}\right)^T_j, j = 1, \ldots, c,$$

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 $\left(\sqrt{(c+\lambda)\mathbf{P}}\right)_j$ 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

$$\hat{\mathbf{y}} = \sum_{i=0}^{2c} W^{(i)} \hat{\mathbf{y}}^{(i)}, \ \hat{\mathbf{P}}_{\mathbf{y}} = \sum_{i=0}^{2c} W^{(i)} (\hat{\mathbf{y}}^{(i)} - \hat{\mathbf{y}})(\hat{\mathbf{y}}^{(i)} - \hat{\mathbf{y}})^T,$$

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

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:

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

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 Section 3.

3.   PROPERTIES OF UT

Define $\tilde{\mathbf{x}} = \mathbf{x} - E(\mathbf{x}) = (x_1 - E(x_1), \ldots, x_c - E(x_c))^T$. The derivative $\partial^i f(E(\mathbf{x}))/(\partial x_1^{k_1} \cdots \partial x_c^{k_c})$ is the derivative of f evaluated at $E(\mathbf{x})$ 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 of f around $E(\mathbf{x})$ is given as follows:

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

Define $D_{\tilde{\mathbf{x}}}^k f$ as

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

We assume that $D_{\tilde{\mathbf{x}}}^k f$ is integrable on Ω for any non-negative integer k. We also assume that there exists Y with finite absolute first moment such that $| \sum_{i=0}^{m} D_{\tilde{\mathbf{x}}}^i f/i! | \leq Y$ a.e. on Ω for all m. Since $\sum_{i=0}^{m} D_{\tilde{\mathbf{x}}}^i f/i! \to f$, we have $\lim_{m \to \infty} E(\sum_{i=0}^{m} D_{\tilde{\mathbf{x}}}^i f/i!) = \lim_{m \to \infty} \sum_{i=0}^{m} E(D_{\tilde{\mathbf{x}}}^i f/i!) =$

4

$E\{f(\mathbf{x})\}$ by the dominated convergence theorem. In practice, these conditions are satisfied if i) there exists a constant $Q > 0$ such that $E|\tilde{x}_1^{k_1} \cdots \tilde{x}_c^{k_c}| \leq Q^{\sum_{j=1}^c k_j}$, for any non-negative integers k_j; ii) for all j, $1 \leq j \leq q$, there exists some constant $R > 0$

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

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

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

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

$$+ \frac{1}{2(c+\lambda)} \sum_{i=1}^{2c} \sum_{j=2}^\infty \frac{1}{(2j)!} D_{\tilde{\mathbf{x}}^{(i)}}^{2j} f,$$

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 equation (3) provides an heuristic expansion of $E(\mathbf{y}) = E\{f(\mathbf{x}_h)\}$ with summands of orders h^j, $j = 0, 1, \cdots$. The order of the error rate of the UT estimator of $E(\mathbf{y})$ may then be studied by comparing Equations (3) and (4). If $\mathbf{x}$ has a symmetric distribution about its mean $E(\mathbf{x})$, we have

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

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

$$
\begin{aligned}
\mathbf{P_y} &= \mathbf{HPH}^T - \frac{1}{4}\mathrm{E}(D_{\tilde{\mathbf{x}}}^2 f)\mathrm{E}(D_{\tilde{\mathbf{x}}}^2 f)^T + \mathrm{E}\Big[\sum_{(i,j)\in\mathbb{A}} \frac{1}{i!j!}(D_{\tilde{\mathbf{x}}}^i f)(D_{\tilde{\mathbf{x}}}^j f)^T \Big] \\
&\quad - \Big[\sum_{(i,j)\in\mathbb{B}} \frac{1}{i!j!}\mathrm{E}(D_{\tilde{\mathbf{x}}}^i f)\mathrm{E}(D_{\tilde{\mathbf{x}}}^j f)^T \Big], \\
\hat{\mathbf{P}}_\mathbf{y} &= \mathbf{HPH}^T - \frac{1}{4}\mathrm{E}(D_{\tilde{\mathbf{x}}}^2 f)\mathrm{E}(D_{\tilde{\mathbf{x}}}^2 f)^T \\
&\quad + \frac{1}{2(c+\lambda)} \sum_{i=1}^{2c} \Big[\sum_{\substack{k+\ell=\text{ even} \\ (k,\ell)\in\mathbb{A}}} \frac{1}{k!\ell!}(D_{\tilde{\mathbf{x}}^{(i)}}^k f)(D_{\tilde{\mathbf{x}}^{(i)}}^\ell f)^T \Big] \\
&\quad - \sum_{(k,\ell)\in\mathbb{A}} \Big[\frac{1}{(2k!)(2\ell!)} \frac{1}{4(c+\lambda)^2} \sum_{i=1}^{2c}\sum_{j=1}^{2c}(D_{\tilde{\mathbf{x}}^{(i)}}^{2k} f)(D_{\tilde{\mathbf{x}}^{(j)}}^{2\ell} f)^T \Big].
\end{aligned}
\tag{6}
$$

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

$$
\begin{aligned}
\mathbf{P_y} &= \mathbf{HPH}^T - \frac{1}{4}\mathrm{E}(D_{\tilde{\mathbf{x}}}^2 f)\mathrm{E}(D_{\tilde{\mathbf{x}}}^2 f)^T + \mathrm{E}\Big[\sum_{\substack{i+j=\text{ even} \\ (i,j)\in\mathbb{A}}} \frac{1}{i!j!}(D_{\tilde{\mathbf{x}}}^i f)(D_{\tilde{\mathbf{x}}}^j f)^T \Big] \\
&\quad - \Big[\sum_{(i,j)\in\mathbb{A}} \frac{1}{(2i!)(2j!)}\mathrm{E}(D_{\tilde{\mathbf{x}}}^{2i} f)\mathrm{E}(D_{\tilde{\mathbf{x}}}^{2j} f)^T \Big].
\end{aligned}
\tag{7}
$$

In addition,

$$
\begin{aligned}
\mathbf{P_{xy}} &= \mathbf{PH}^T + \sum_{i=2}^{\infty} \frac{1}{i!}\mathrm{E}\Big[\tilde{\mathbf{x}}\Big(D_{\tilde{\mathbf{x}}}^i f\Big)^T \Big], \\
\hat{\mathbf{P}}_\mathbf{xy} &= \mathbf{PH}^T + \sum_{k=1}^{\infty} \frac{1}{(2k+1)!} \frac{1}{2(c+\lambda)} \sum_{i=1}^{2c} \tilde{\mathbf{x}}^{(i)}\Big(D_{\tilde{\mathbf{x}}^{(i)}}^{2k+1} f\Big)^T.
\end{aligned}
\tag{8}
$$

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

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

Detailed derivations of (3)-(9) can be found in Appendix. 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 $|\sum_{i=0}^m D_{\tilde{\mathbf{x}}}^i f / i!| \leq Y$ 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. $\mathbf{P_y} - \hat{\mathbf{P}}_\mathbf{y} = E\Big[\sum_{(i,j)\in\mathbb{A}} \frac{1}{i!j!}(D_{\tilde{\mathbf{x}}}^i f)(D_{\tilde{\mathbf{x}}}^j f)^T\Big] - \Big[\sum_{(i,j)\in\mathbb{B}} \frac{1}{i!j!} E(D_{\tilde{\mathbf{x}}}^i f)E(D_{\tilde{\mathbf{x}}}^j f)^T\Big]$
   $- \frac{1}{2(c+\lambda)} \sum_{i=1}^{2c} \Big[\sum_{\substack{k+\ell=\ even\\(k,\ell)\in\mathbb{A}}} \frac{1}{k!\ell!}(D_{\tilde{\mathbf{x}}^{(i)}}^k f)(D_{\tilde{\mathbf{x}}^{(i)}}^\ell f)^T\Big]$
   $+ \sum_{(k,\ell)\in\mathbb{A}} \Big[\frac{1}{(2k!)(2\ell!)} \frac{1}{4(c+\lambda)^2} \sum_{i=1}^{2c} \sum_{j=1}^{2c} (D_{\tilde{\mathbf{x}}^{(i)}}^{2k} f)(D_{\tilde{\mathbf{x}}^{(j)}}^{2\ell} f)^T\Big]$;

3. $\mathbf{P_{xy}} - \hat{\mathbf{P}}_\mathbf{xy} = \sum_{i=2}^{\infty} \frac{1}{i!} E\Big[\tilde{\mathbf{x}}\Big(D_{\tilde{\mathbf{x}}}^i f\Big)^T\Big] - \sum_{k=1}^{\infty} \frac{1}{(2k+1)!} \frac{1}{2(c+\lambda)} \sum_{i=1}^{2c} \tilde{\mathbf{x}}^{(i)}\Big(D_{\tilde{\mathbf{x}}^{(i)}}^{2k+1} f\Big)^T$.

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

3. $\mathbf{P_{xy}} - \hat{\mathbf{P}}_\mathbf{xy} = \sum_{i=1}^{\infty} \frac{1}{(2i+1)!} E\Big[\tilde{\mathbf{x}}\Big(D_{\tilde{\mathbf{x}}}^{2i+1} f\Big)^T\Big] - \sum_{k=1}^{\infty} \frac{1}{(2k+1)!} \frac{1}{2(c+\lambda)} \sum_{i=1}^{2c} \tilde{\mathbf{x}}^{(i)}\Big(D_{\tilde{\mathbf{x}}^{(i)}}^{2k+1} f\Big)^T$.

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. $\mathbf{P_y} - \hat{\mathbf{P}}_{\mathbf{y},L} = -\frac{1}{4} E(D_{\tilde{\mathbf{x}}}^2 f) E(D_{\tilde{\mathbf{x}}}^2 f)^T + E\Big[\sum_{(i,j)\in\mathbb{A}} \frac{1}{i!j!}(D_{\tilde{\mathbf{x}}}^i f)(D_{\tilde{\mathbf{x}}}^j f)^T\Big]$
   $- \Big[\sum_{(i,j)\in\mathbb{B}} \frac{1}{i!j!} E(D_{\tilde{\mathbf{x}}}^i f)E(D_{\tilde{\mathbf{x}}}^j f)^T\Big]$;

7

3. $\mathbf{P_{xy}} - \hat{\mathbf{P}}_{\mathbf{xy},L} = \sum_{i=2}^{\infty} \frac{1}{i!} E\Big[\tilde{\mathbf{x}} \Big(D_{\tilde{\mathbf{x}}}^i f \Big)^T \Big].$

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. $\mathbf{P_y} - \hat{\mathbf{P}}_{\mathbf{y},L} = -\frac{1}{4} E(D_{\tilde{\mathbf{x}}}^2 f) E(D_{\tilde{\mathbf{x}}}^2 f)^T + E\Big[\sum_{\substack{i+j=\ even \\ (i,j) \in \mathbb{A}}} \frac{1}{i!j!} (D_{\tilde{\mathbf{x}}}^i f)(D_{\tilde{\mathbf{x}}}^j f)^T \Big]$
   $- \Big[\sum_{(i,j) \in \mathbb{A}} \frac{1}{(2i!)(2j!)} E(D_{\tilde{\mathbf{x}}}^{2i} f) E(D_{\tilde{\mathbf{x}}}^{2j} f)^T \Big];$

3. $\mathbf{P_{xy}} - \hat{\mathbf{P}}_{\mathbf{xy},L} = \sum_{i=1}^{\infty} \frac{1}{(2i+1)!} E\Big[\tilde{\mathbf{x}} \Big(D_{\tilde{\mathbf{x}}}^{2i+1} f \Big)^T \Big].$

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 $f(E(\mathbf{x}))$ 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 $\mathbf{P_y}$ matches more terms than its linear approximation counterpart. For approximating $\mathbf{P_{xy}}$, 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 remaining higher-order terms. This compensation appears to be better when $\mathbf{x}$ has a symmetric distribution about its mean.*

Remark 2. *Lemma 1 shows that if λ is close to $-c$ or very large, 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, $E(x_1^{k_1} \cdots x_c^{k_c})$ can be expressed as a function of the components of $\mathbf{P}$, see Isserlis (1918), Holmquist (1988), and Triantafyllopoulos (2003). So we can calculate the mean squared error (MSE) in this case. From Lemma 1, the bias can be approximated as $E(\mathbf{y}) - \hat{\mathbf{y}} \approx \sum_{j=2}^{m} 1/(2j)! E[D_{\tilde{\mathbf{x}}}^{2j} f] - 1/(2c+2\lambda) \sum_{i=1}^{2c} \sum_{j=2}^{m} 1/(2j)! D_{\tilde{\mathbf{x}}^{(i)}}^{2j} f$ 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 3. *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), 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 $\mathbf{P}^*$. Thus, it is interesting to compare the error rates of the UT estimators of $E(\mathbf{y}) = E\{f(\mathbf{x})\}, \mathbf{P}_{\mathbf{y}}, \mathbf{P}_{\mathbf{xy}}$ and their counterparts from linear approximation when $\mathbf{P} = h\mathbf{P}^*$. We allow that f may depend on h. For simplicity, we assume that the distribution of $\mathbf{x}$ is symmetric about its mean. Then, we can obtain the following theorem:

Theorem 4. *Suppose $\mathbf{P} = h\mathbf{P}^*$ and $\mathbf{x}$ has a symmetric distribution about its mean. Assume*

1. there exists a h-free constant $M > 0$ such that $E|\tilde{x}_1^{k_1} \cdots \tilde{x}_c^{k_c}| \leq h^{i/2} M^{i/2}$, 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 $K > 0$,

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

9

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

Then,

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

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

Proof. See Appendix. □

Remark 5. *Condition 1 of Theorem 4 is motivated by the fact that $\tilde{x}_1^{k_1} \cdots \tilde{x}_c^{k_c} = O_p(h^{i/2})$ because $\tilde{x}_j = O_p(\sqrt{h})$. Condition 2 of Theorem 4 is satisfied if the function f is a polynomial.*

Remark 6. *Theorem 4 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, which is shown rigorously by Ahn and Chan (2011).*

4.   CONCLUSION

We have derived some theoretical properties of the UT and linear approximation. 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. Thus, the UT is faster than the Monte Carlo and more accurate than the method of linear approximation.

ACKNOWLEDGEMENTS

REFERENCES

Ahn, K. W., and Chan, K. S. (2011), "Nonlinear state-space modeling via the unscented Kalman filter," *Technical report, Medical College of Wisconsin,* .

Boyce, W. E., and DiPrima, R. C. (2004), *Elementary Differential Equations and Boundary Value Problems*, 8 edn, : Wiley.

Diekmann, O., and Heesterbeek, J. A. P. (2000), *Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation*, : John Wiley & Sons Ltd.

Doucet, A., deFreitas, J. F. G., and Gordon, N. J. (2001), *Sequential Monte Carlo Methods in Practice*, New York: Springer-Verlag.

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., and Uhlmann, J. K. (1997), "A new extension of the Kalman filter to nonlinear systems," *Proc. 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., and Uhlmann, J. K. (2004), "Unscented filtering and nonlinear estimation," *IEEE Review*, **92,**, 401–422.

Simon, D. (2006), *Optimal State Estimation*, New Jersey: John Wiley & Sons, Inc.

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

Wan, E. A., and van der Merwe, R. (2000), "The unscented Kalman filter for nonlinear estimation," *In Symposium 2000 on Adaptive Systems for Signal Processing*, .

APPENDIX A.   DERIVATION OF RESULTS FOR THEOREM 1

For simplicity, assume the pdf of the random vector $\mathbf{x}$ is symmetric about its mean with a known mean $E(\mathbf{x})$ and covariance matrix $\mathbf{P_x}$, and $\mathbf{y} = f(\mathbf{x})$. Assume f is an analytic

function. We also assume that $D_{\tilde{\mathbf{x}}}^k f$ is integrable on Ω for any non-negative integer k and there exists Y with finite absolute moment suth that $|\sum_{i=0}^m D_{\tilde{\mathbf{x}}}^i f / i!| \leq Y$ a.e. on Ω for all m. Replacing $\hat{\mathbf{y}}^{(i)}$ by its Taylor expansion around $\mathrm{E}(\mathbf{x})$, we get

$$\hat{\mathbf{y}} = W^{(0)}\hat{\mathbf{y}}^{(0)} + \sum_{i=1}^{2c} W^{(i)}\hat{\mathbf{y}}^{(i)}$$

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

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

Now notice that

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

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

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

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

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

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

$$\hat{\mathbf{y}} = f(\mathrm{E}(\mathbf{x})) + \frac{1}{2} \sum_{i=1}^{c} \sum_{j=1}^{c} P_{ij} \frac{\partial^2}{\partial x_i \partial x_j} f(\mathrm{E}(\mathbf{x}))$$

$$+ \frac{1}{2(c+\lambda)} \sum_{i=1}^{2c} \left(\frac{1}{4!} D_{\tilde{\mathbf{x}}^{(i)}}^4 f + \frac{1}{6!} D_{\tilde{\mathbf{x}}^{(i)}}^6 f + \cdots \right). \tag{A.2}$$

Similarly,

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

It can be easily shown

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

As a result, we have

$$
\mathrm{E}(\mathbf{y}) = f(\mathrm{E}(\mathbf{x})) + \frac{1}{2!}\Big[\sum_{i=1}^{c}\sum_{j=1}^{c} P_{ij}\frac{\partial^2}{\partial x_i \partial x_j}f(\mathrm{E}(\mathbf{x}))\Big]
$$

$$
+ \frac{1}{4!}\mathrm{E}[D_{\tilde{\mathbf{x}}}^4 f] + \cdots . \tag{A.3}
$$

Next, we turn our attention to the covariance structure:

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

Based on the results obtained so far, we have

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

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,

$$
\begin{aligned}
\mathbf{P_y} &= \mathrm{E}[(\mathbf{y} - \mathrm{E}(\mathbf{y}))(\mathbf{y} - \mathrm{E}(\mathbf{y}))^T] \\
&= \mathrm{E}\Big[\Big\{\Big(D_{\tilde{\mathbf{x}}}f + \frac{1}{2!}D_{\tilde{\mathbf{x}}}^2 f + \frac{1}{3!}D_{\tilde{\mathbf{x}}}^3 f + \cdots\Big) \\
&\quad - \Big(\frac{1}{2!}\mathrm{E}(D_{\tilde{\mathbf{x}}}^2 f) + \frac{1}{4!}\mathrm{E}(D_{\tilde{\mathbf{x}}}^4 f) + \frac{1}{6!}\mathrm{E}(D_{\tilde{\mathbf{x}}}^6 f) + \cdots\Big)\Big\} \\
&\quad \times \Big\{\Big(D_{\tilde{\mathbf{x}}}f + \frac{1}{2!}D_{\tilde{\mathbf{x}}}^2 f + \frac{1}{3!}D_{\tilde{\mathbf{x}}}^3 f + \cdots\Big) \\
&\quad - \Big(\frac{1}{2!}\mathrm{E}(D_{\tilde{\mathbf{x}}}^2 f) + \frac{1}{4!}\mathrm{E}(D_{\tilde{\mathbf{x}}}^4 f) + \frac{1}{6!}\mathrm{E}(D_{\tilde{\mathbf{x}}}^6 f) + \cdots\Big)\Big\}^T\Big] \\
&= \mathrm{E}[(D_{\tilde{\mathbf{x}}}f)(D_{\tilde{\mathbf{x}}}f)^T] - \frac{1}{4}\mathrm{E}(D_{\tilde{\mathbf{x}}}^2 f)\mathrm{E}(D_{\tilde{\mathbf{x}}}^2 f)^T \\
&\quad + \mathrm{E}\Big[\sum_{\substack{i+j=\text{ even} \\ (i,j)\in\mathbb{A}}} \frac{1}{i!j!}(D_{\tilde{\mathbf{x}}}^i f)(D_{\tilde{\mathbf{x}}}^j f)^T\Big] \\
&\quad - \Big[\sum_{(i,j)\in\mathbb{A}} \frac{1}{(2i!)(2j!)}\mathrm{E}(D_{\tilde{\mathbf{x}}}^{2i} f)\mathrm{E}(D_{\tilde{\mathbf{x}}}^{2j} f)^T\Big],
\end{aligned} \tag{A.4}
$$

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

$$
\begin{aligned}
\mathrm{E}[(D_{\tilde{x}}f)(D_{\tilde{\mathbf{x}}}f)^T] &= \mathrm{E}\Big[\Big(\sum_{i=1}^{c}\tilde{x}_i\frac{\partial f(\mathrm{E}(\mathbf{x}))}{\partial x_i}\Big)\Big(\sum_{i=1}^{c}\tilde{x}_i\frac{\partial f(\mathrm{E}(\mathbf{x}))}{\partial x_i}\Big)^T\Big] \\
&= \mathrm{E}\Big[\sum_{i=1}^{c}\sum_{j=1}^{c}\tilde{x}_i\frac{\partial f(\mathrm{E}(\mathbf{x}))}{\partial x_i}\frac{\partial f(\mathrm{E}(\mathbf{x}))^T}{\partial x_j}\tilde{x}_j\Big] \\
&= \sum_{i=1}^{c}\sum_{j=1}^{c}\mathbf{H}_i\mathrm{E}(\tilde{x}_i\tilde{x}_j)\mathbf{H}_j^T = \sum_{i=1}^{c}\sum_{j=1}^{c}\mathbf{H}_iP_{ij}\mathbf{H}_j^T = \mathbf{H}\mathbf{P}\mathbf{H}^T,
\end{aligned}
\tag{A.5}
$$

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

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

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

$$
\begin{aligned}
\hat{\mathbf{y}}^{(0)}-\hat{\mathbf{y}} &= f(\mathrm{E}(\mathbf{x})) - \Big\{f(\mathrm{E}(\mathbf{x})) + \frac{1}{2(c+\lambda)}\sum_{i=1}^{2c}\Big(\frac{1}{2!}D^2_{\tilde{\mathbf{x}}^{(i)}}f + \frac{1}{4!}D^4_{\tilde{\mathbf{x}}^{(i)}}f\cdots\Big)\Big\} \\
&= -\frac{1}{2(c+\lambda)}\sum_{i=1}^{2c}\Big(\frac{1}{2!}D^2_{\tilde{\mathbf{x}}^{(i)}}f + \frac{1}{4!}D^4_{\tilde{\mathbf{x}}^{(i)}}f\cdots\Big)
\end{aligned}
$$

Thus, we have

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

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 (A.1) and the fact that

$$\frac{1}{2(c+\lambda)}\sum_{i=1}^{2c}[(D_{\tilde{\mathbf{x}}^{(i)}}f)(D_{\tilde{\mathbf{x}}^{(i)}}f)^T] = \frac{1}{2(c+\lambda)}\sum_{i=1}^{2c}\sum_{k=1}^{2c}\sum_{\ell=1}^{2c}\left(\tilde{x}_k^{(i)}\frac{\partial f(\mathrm{E}(\mathbf{x}))}{\partial x_k}\right)\left(\tilde{x}_\ell^{(i)}\frac{\partial f(\mathrm{E}(\mathbf{x}))}{\partial x_\ell}\right)^T$$

$$= \frac{1}{c+\lambda}\sum_{i=1}^{c}\sum_{k=1}^{c}\sum_{\ell=1}^{c}\left(\tilde{x}_k^{(i)}\frac{\partial f(\mathrm{E}(\mathbf{x}))}{\partial x_k}\right)\left(\tilde{x}_\ell^{(i)}\frac{\partial f(\mathrm{E}(\mathbf{x}))}{\partial x_\ell}\right)^T$$

$$(\text{using } \tilde{x}_j^{(i)} = -\tilde{x}_j^{(c+i)})$$

$$= \sum_{k=1}^{c}\sum_{\ell=1}^{c}P_{k\ell}\left(\frac{\partial f(\mathrm{E}(\mathbf{x}))}{\partial x_k}\right)\left(\frac{\partial f(\mathrm{E}(\mathbf{x}))}{\partial x_\ell}\right)^T = \mathbf{HPH}^T$$

$$= \mathrm{E}[(D_{\tilde{x}}f)(D_{\tilde{\mathbf{x}}}f)^T] \quad (\text{By (A.5)}),$$

we can show

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

$$= \frac{1}{2(c+\lambda)}\sum_{i=1}^{2c}\left[\left\{\left(D_{\tilde{\mathbf{x}}^{(i)}}f + \frac{1}{2!}D_{\tilde{\mathbf{x}}^{(i)}}^2 f + \frac{1}{3!}D_{\tilde{\mathbf{x}}^{(i)}}^3 f + \cdots\right)\right.\right.$$

$$- \frac{1}{2(c+\lambda)}\sum_{j=1}^{2c}\left(\frac{1}{2!}D_{\tilde{\mathbf{x}}^{(j)}}^2 f + \frac{1}{4!}D_{\tilde{\mathbf{x}}^{(j)}}^4 f + \cdots\right)\right\}$$

$$\times \left\{\left(D_{\tilde{\mathbf{x}}^{(i)}}f + \frac{1}{2!}D_{\tilde{\mathbf{x}}^{(i)}}^2 f + \frac{1}{3!}D_{\tilde{\mathbf{x}}^{(i)}}^3 f + \cdots\right)\right.$$

$$\left.\left.- \frac{1}{2(c+\lambda)}\sum_{j=1}^{2c}\left(\frac{1}{2!}D_{\tilde{\mathbf{x}}^{(j)}}^2 f + \frac{1}{4!}D_{\tilde{\mathbf{x}}^{(j)}}^4 f + \cdots\right)\right\}^T\right] \tag{A.6}$$

$$= \mathrm{E}[(D_{\tilde{\mathbf{x}}}f)(D_{\tilde{\mathbf{x}}}f)^T] - \left(1 + \frac{\lambda}{c+\lambda}\right)\frac{1}{4}\mathrm{E}[D_{\tilde{\mathbf{x}}}^2 f]\mathrm{E}[D_{\tilde{\mathbf{x}}}^2 f]^T$$

$$+ \frac{1}{2(c+\lambda)}\sum_{i=1}^{2c}\left[\sum_{\substack{k+\ell=\text{ even}\\(k,\ell)\in\mathbb{A}}}\frac{1}{k!\ell!}(D_{\tilde{\mathbf{x}}^{(i)}}^k f)(D_{\tilde{\mathbf{x}}^{(i)}}^\ell f)^T\right]$$

$$- \sum_{(k,\ell)\in\mathbb{A}}\left[\frac{1}{(2k!)(2\ell!)}\frac{1}{4(c+\lambda)^2}\left(1 + \frac{\lambda}{c+\lambda}\right)\right.$$

$$\times \left.\sum_{i=1}^{2c}\sum_{j=1}^{2c}(D_{\tilde{\mathbf{x}}^{(i)}}^{2k} f)(D_{\tilde{\mathbf{x}}^{(j)}}^{2\ell} f)^T\right].$$

Thus, we have

$$
\begin{aligned}
\hat{\mathbf{P}}_{\mathbf{y}} &= \frac{\lambda}{c+\lambda}(\hat{\mathbf{y}}^{(0)} - \hat{\mathbf{y}})(\hat{\mathbf{y}}^{(0)} - \hat{\mathbf{y}})^T + \frac{1}{2(c+\lambda)}\sum_{i=1}^{2c}(\hat{\mathbf{y}}^{(i)} - \hat{\mathbf{y}})(\hat{\mathbf{y}}^{(i)} - \hat{\mathbf{y}})^T \\
&= \mathrm{E}[(D_{\tilde{\mathbf{x}}}f)(D_{\tilde{\mathbf{x}}}f)^T] - \frac{1}{4}\mathrm{E}[D_{\tilde{\mathbf{x}}}^2 f]\mathrm{E}[D_{\tilde{\mathbf{x}}}^2 f]^T \\
&\quad + \frac{1}{2(c+\lambda)}\sum_{i=1}^{2c}\Big[\sum_{\substack{k+\ell=\text{ even} \\ (k,\ell)\in\mathbb{A}}} \frac{1}{k!\ell!}(D_{\tilde{\mathbf{x}}^{(i)}}^k f)(D_{\tilde{\mathbf{x}}^{(i)}}^\ell f)^T\Big] \\
&\quad - \sum_{(k,\ell)\in\mathbb{A}}\Big[\frac{1}{(2k!)(2\ell!)}\frac{1}{4(c+\lambda)^2}\sum_{i=1}^{2c}\sum_{j=1}^{2c}(D_{\tilde{\mathbf{x}}^{(i)}}^{2k} f)(D_{\tilde{\mathbf{x}}^{(j)}}^{2\ell} f)^T\Big].
\end{aligned}
\tag{A.7}
$$

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

$$
\begin{aligned}
\mathbf{P}_{\mathbf{xy}} &= \mathrm{E}[(\mathbf{x} - \mathrm{E}(\mathbf{x}))(\mathbf{y} - \mathrm{E}(\mathbf{y}))^T] = \mathrm{E}[(\tilde{\mathbf{x}})(\mathbf{y} - \mathrm{E}(\mathbf{y}))^T] \\
&= \mathrm{E}\Big[\tilde{\mathbf{x}}\big(D_{\tilde{\mathbf{x}}}f\big)^T\Big] + \sum_{i=1}^{\infty}\frac{1}{(2i+1)!}\mathrm{E}\Big[\tilde{\mathbf{x}}\big(D_{\tilde{\mathbf{x}}}^{2i+1} f\big)^T\Big] \\
&= \mathbf{P}\mathbf{H}^T + \sum_{i=1}^{\infty}\frac{1}{(2i+1)!}\mathrm{E}\Big[\tilde{\mathbf{x}}\big(D_{\tilde{\mathbf{x}}}^{2i+1} f\big)^T\Big].
\end{aligned}
\tag{A.8}
$$

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

$$
\begin{aligned}
\hat{\mathbf{P}}_{\mathbf{xy}} &= \frac{1}{2(c+\lambda)}\sum_{i=0}^{2c}(\hat{\mathbf{x}}^{(i)} - \mathrm{E}(\mathbf{x}))(\hat{\mathbf{y}}^{(i)} - \hat{\mathbf{y}})^T \\
&= \frac{1}{2(c+\lambda)}\sum_{i=1}^{2c}\tilde{\mathbf{x}}^{(i)}\big(D_{\tilde{\mathbf{x}}^{(i)}}f\big)^T \\
&\quad + \sum_{k=1}^{\infty}\frac{1}{(2k+1)!}\frac{1}{2(c+\lambda)}\sum_{i=1}^{2c}\tilde{\mathbf{x}}^{(i)}\big(D_{\tilde{\mathbf{x}}^{(i)}}^{2k+1} f\big)^T \\
&= \mathbf{P}\mathbf{H}^T + \sum_{k=1}^{\infty}\frac{1}{(2k+1)!}\frac{1}{2(c+\lambda)}\sum_{i=1}^{2c}\tilde{\mathbf{x}}^{(i)}\big(D_{\tilde{\mathbf{x}}^{(i)}}^{2k+1} f\big)^T.
\end{aligned}
\tag{A.9}
$$

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). From the Taylor series in Section A of Appendix and Condtion 1 of Theorem 2, we have

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

$$- \left[f(\mathrm{E}(\mathbf{x})) + \frac{1}{2} \sum_{i=1}^{c} \sum_{j=1}^{c} P_{ij} \frac{\partial^2}{\partial x_i \partial x_j} f(\mathrm{E}(\mathbf{x})) \right.$$

$$\left. + \frac{1}{2(c+\lambda)} \sum_{j=1}^{2c} \left(\frac{1}{4!} D_{\tilde{\mathbf{x}}^{(j)}}^4 f + \frac{1}{6!} D_{\tilde{\mathbf{x}}^{(j)}}^6 f \cdots \right) \right]$$

$$= \sum_{i=2}^{\infty} \frac{1}{(2i)!} \mathrm{E}[D_{\tilde{\mathbf{x}}}^{2i} f] - \sum_{i=2}^{\infty} \frac{1}{2(c+\lambda)} \sum_{j=1}^{2c} \frac{1}{(2i)!} D_{\tilde{\mathbf{x}}^{(j)}}^{2i} f$$

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

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

$$\leq \frac{1}{(2i)!} \sum_{1 \leq k_1, \ldots, k_c \leq 2i} \left[\binom{2i}{k_1, \ldots, k_c} \left(h^i M^i \left| \frac{\partial^{2i} f_\ell(\mathrm{E}(\mathbf{x}))}{\partial x_1^{k_1} \cdots \partial x_c^{k_c}} \right| \right) \right] \qquad \text{(A.10)}$$

$$\equiv h^i a_{i,\ell}.$$

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

$$\frac{1}{2(c+\lambda)} \sum_{j=1}^{2c} \frac{1}{(2i)!} D_{\tilde{\mathbf{x}}^{(j)}}^{2i} f$$

$$= \frac{1}{(c+\lambda)} \sum_{j=1}^{c} \frac{1}{(2i)!} \left(\sum_{\ell=1}^{c} \sqrt{(c+\lambda)P}_{\ell j} \frac{\partial}{\partial x_\ell} \right)^{2i} f$$

$$= \frac{h^i}{(2i)!} \sum_{j=1}^{c} (c+\lambda)^{i-1} \left(\sum_{1 \leq k_1, \ldots, k_c \leq 2i} \binom{2i}{k_1, \ldots, k_c} \right. \qquad \text{(A.11)}$$

$$\left. \times (\sqrt{P^*}_{1j})^{k_1} \cdots (\sqrt{P^*}_{cj})^{k_c} \frac{\partial^{2i} f(\mathrm{E}(\mathbf{x}))}{\partial x_1^{k_1} \cdots \partial x_c^{k_n}} \right)$$

$$\equiv h^i \mathbf{b}_i,$$

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 (A.10) and (A.11) respectively are finite. Thus, we have

$$|\mathrm{E}(\mathbf{y}) - \hat{\mathbf{y}}| \leq \sum_{i=2}^{\infty} h^i(|\mathbf{a}_i| + |\mathbf{b}_i|) = h^2 \sum_{i=2}^{\infty} h^{i-2}(|\mathbf{a}_i| + |\mathbf{b}_i|) = O(h^2). \qquad \text{(A.12)}$$

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

$$\mathrm{E}(\mathbf{y}) - \hat{\mathbf{y}}_L = O(h). \tag{A.13}$$

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

$$\mathbf{P_y} - \hat{\mathbf{P}}_\mathbf{y} = \mathrm{E}\Big[\sum_{\substack{i+j=\text{ even} \\ (i,j)\in\mathbb{A}}} \frac{1}{i!j!}(D_{\tilde{\mathbf{x}}}^i f)(D_{\tilde{\mathbf{x}}}^j f)^T \Big] - \Big[\sum_{(i,j)\in\mathbb{A}} \frac{1}{(2i!)(2j!)}\mathrm{E}(D_{\tilde{\mathbf{x}}}^{2i} f)\mathrm{E}(D_{\tilde{\mathbf{x}}}^{2j} f)^T \Big]$$

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

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

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

$$\mathrm{E}\Big[\frac{1}{i!j!}(D_{\tilde{\mathbf{x}}}^i f)(D_{\tilde{\mathbf{x}}}^j f)^T\Big] = \mathrm{E}\Big[\Big(\frac{1}{i!} \sum_{1\le k_1,\dots,k_c\le i} \binom{i}{k_1,\dots,k_c} \tilde{x}_1^{k_1}\cdots\tilde{x}_c^{k_c} \frac{\partial^i f(\mathrm{E}(\mathbf{x}))}{\partial x_1^{k_1}\cdots\partial x_c^{k_c}} \Big)$$

$$\times \Big(\frac{1}{j!} \sum_{1\le \ell_1,\dots,\ell_c\le j} \binom{j}{k_1,\dots,k_c} \tilde{x}_1^{\ell_1}\cdots\tilde{x}_c^{\ell_c} \frac{\partial^j f(\mathrm{E}(\mathbf{x}))}{\partial x_1^{\ell_1}\cdots\partial x_c^{\ell_c}} \Big)^T.$$

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

$$
\begin{aligned}
&\left|\mathrm{E}\left[\frac{1}{i!j!}(D_{\tilde{\mathbf{x}}}^i f)(D_{\tilde{\mathbf{x}}}^j f)^T\right]_{ab}\right| \\
&= \left|\mathrm{E}\left[\left(\frac{1}{i!}\sum_{1\le k_1,\dots,k_c\le i}\binom{i}{k_1,\dots,k_c}\tilde{x}_1^{k_1}\cdots\tilde{x}_c^{k_c}\frac{\partial^i f_a(\mathrm{E}(\mathbf{x}))}{\partial x_1^{k_1}\cdots\partial x_c^{k_c}}\right)\right.\right. \\
&\qquad\left.\left.\times\left(\frac{1}{j!}\sum_{1\le \ell_1,\dots,\ell_c\le j}\binom{j}{\ell_1,\dots,\ell_c}\tilde{x}_1^{\ell_1}\cdots\tilde{x}_c^{\ell_c}\frac{\partial^j f_b(\mathrm{E}(\mathbf{x}))}{\partial x_1^{\ell_1}\cdots\partial x_c^{\ell_c}}\right)\right]\right| \\
&= \left|\mathrm{E}\left[\frac{1}{i!j!}\sum_{\substack{1\le k_1,\dots,k_c\le i\\1\le \ell_1,\dots,\ell_c\le j}}\binom{i}{k_1,\dots,k_c}\binom{j}{\ell_1,\dots,\ell_c}\tilde{x}_1^{k_1+\ell_1}\cdots\tilde{x}_c^{k_c+\ell_c}\right.\right. \\
&\qquad\left.\left.\times\frac{\partial^i f_a(\mathrm{E}(\mathbf{x}))}{\partial x_1^{k_1}\cdots\partial x_c^{k_c}}\frac{\partial^j f_b(\mathrm{E}(\mathbf{x}))}{\partial x_1^{\ell_1}\cdots\partial x_n^{\ell_c}}\right]\right| \\
&\le \frac{h^{(i+j)/2}M^{(i+j)/2}}{i!j!}\sum_{\substack{1\le k_1,\dots,k_c\le i\\1\le \ell_1,\dots,\ell_c\le j}}\binom{i}{k_1,\dots,k_c}\binom{j}{\ell_1,\dots,\ell_c} \\
&\qquad\times\left|\frac{\partial^i f_a(\mathrm{E}(\mathbf{x}))}{\partial x_1^{k_1}\cdots\partial x_c^{k_c}}\frac{\partial^j f_b(\mathrm{E}(\mathbf{x}))}{\partial x_1^{\ell_1}\cdots\partial x_c^{\ell_c}}\right| \\
&\equiv h^{(i+j)/2}R_{ab}^{ij}.
\end{aligned}
\tag{A.14}
$$

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

$$
\begin{aligned}
&\left|\mathrm{E}\left[\sum_{\substack{i+j=\text{ even}\\(i,j)\in\mathbb{A}}}\frac{1}{i!j!}(D_{\tilde{\mathbf{x}}}^i f)(D_{\tilde{\mathbf{x}}}^j f)^T\right]\right| \\
&\le \sum_{\substack{i+j=\text{ even}\\(i,j)\in\mathbb{A}}}h^{(i+j)/2}\mathbf{R}^{ij} = h^2\sum_{\substack{i+j=\text{ even}\\(i,j)\in\mathbb{A}}}h^{(i+j)/2-2}\mathbf{R}^{ij}.
\end{aligned}
\tag{A.15}
$$

By (A.10), the second term satisfies

$$
\begin{aligned}
&\sum_{(i,j)\in\mathbb{A}}\frac{1}{(2i!)(2j!)}\left|\mathrm{E}(D_{\tilde{\mathbf{x}}}^{2i} f)\mathrm{E}(D_{\tilde{\mathbf{x}}}^{2j} f)^T\right| \\
&\le \sum_{(i,j)\in\mathbb{A}}h^i\mathbf{a}_i(h^i\mathbf{a}_j)^T = h^2\sum_{(i,j)\in\mathbb{A}}h^{i+j-2}\mathbf{a}_i(\mathbf{a}_j)^T.
\end{aligned}
\tag{A.16}
$$

Next, let's consider the third therm $\frac{1}{2(c+\lambda)}\sum_{i=1}^{2c}\frac{1}{j!\ell!}(D_{\tilde{\mathbf{x}}^{(i)}}^j f)(D_{\tilde{\mathbf{x}}^{(i)}}^\ell f)^T$ where $j+\ell$ is even.

$$\frac{1}{2(c+\lambda)}\sum_{i=1}^{2c}\frac{1}{j!\ell!}(D_{\tilde{\mathbf{x}}^{(i)}}^j f)(D_{\tilde{\mathbf{x}}^{(i)}}^\ell f)^T = \frac{1}{c+\lambda}\sum_{i=1}^{c}\frac{1}{j!\ell!}(D_{\tilde{\mathbf{x}}^{(i)}}^j f)(D_{\tilde{\mathbf{x}}^{(i)}}^\ell f)^T$$

$$= \frac{1}{c+\lambda}\sum_{i=1}^{c}\left[\frac{1}{j!}\left(\sum_{r=1}^{c}\sqrt{(c+\lambda)P}_{ri}\frac{\partial}{\partial x_\ell}\right)^j f\right]\left[\frac{1}{\ell!}\left(\sum_{s=1}^{c}\sqrt{(c+\lambda)P}_{si}\frac{\partial}{\partial x_\ell}\right)^\ell f\right]^T$$

$$= \frac{(c+\lambda)^{(j+\ell)/2}}{c+\lambda}\sum_{i=1}^{c}\left[\frac{1}{j!}\sum_{1\le u_1,\ldots,u_c\le j}\binom{j}{u_1,\ldots,u_c}(\sqrt{P}_{1i})^{u_1}\cdots(\sqrt{P}_{ci})^{u_c}\right.$$

$$\times \left.\frac{\partial^j f(\mathrm{E}(\mathbf{x}))}{\partial x_1^{u_1}\cdots\partial x_c^{u_c}}\right]$$

$$\times \left[\frac{1}{\ell!}\sum_{1\le v_1,\ldots,v_c\le\ell}\binom{\ell}{v_1,\ldots,v_c}(\sqrt{P}_{1i})^{v_1}\cdots(\sqrt{P}_{ci})^{v_c}\frac{\partial^\ell f(\mathrm{E}(\mathbf{x}))}{\partial x_1^{v_1}\cdots\partial x_c^{v_c}}\right]^T.$$

Similarly with $\mathrm{E}\left[\frac{1}{i!j!}(D_{\tilde{\mathbf{x}}}^i f)(D_{\tilde{\mathbf{x}}}^j f)^T\right]_{ab}$, we have

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

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

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

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

Thus, we obtain

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

$$= \sum_{\substack{j+\ell=\text{ even}\\(j,\ell)\in\mathbb{A}}}h^{(i+j)/2}\mathbf{T}^{j\ell} = h^2\sum_{\substack{j+\ell=\text{ even}\\(j,\ell)\in\mathbb{A}}}h^{(i+j)/2-2}\mathbf{T}^{j\ell}. \tag{A.17}$$

For the last term of $\mathbf{P_y} - \hat{\mathbf{P}}_\mathbf{y}$, $\sum_{(j,\ell)\in\mathbb{A}}\left[\frac{1}{(2j!)(2\ell!)}\frac{1}{4(c+\lambda)^2}\sum_{i=1}^{2c}\sum_{m=1}^{2c}(D_{\tilde{\mathbf{x}}^{(i)}}^{2j}f)(D_{\tilde{\mathbf{x}}^{(m)}}^{2\ell}f)^T\right]$, by (A.11), we have

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

$$= \sum_{(j,\ell)\in\mathbb{A}}\left[\left(\sum_{i=1}^{2c}\frac{1}{(2j)!}\frac{1}{2(c+\lambda)}D_{\tilde{\mathbf{x}}^{(i)}}^{2j}f\right)\left(\sum_{m=1}^{2c}\frac{1}{(2\ell)!}\frac{1}{2(c+\lambda)}D_{\tilde{\mathbf{x}}^{(m)}}^{2\ell}f\right)^T\right] \quad (A.18)$$

$$= \sum_{(j,\ell)\in\mathbb{A}} h^j\mathbf{b}_j(h^\ell\mathbf{b}_\ell)^T = h^2\sum_{(j,\ell)\in\mathbb{A}} h^{j+\ell-2}\mathbf{b}_j(\mathbf{b}_\ell)^T.$$

Combining (A.15)-(A.18), we obtain

$$\mathbf{P_y} - \hat{\mathbf{P}}_\mathbf{y} = O(h^2), \quad (A.19)$$

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 (A.10), we have

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

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

$$\mathbf{P_y} - \hat{\mathbf{P}}_{\mathbf{y},L} = O(h^2). \quad (A.20)$$

Now, we know

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

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

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

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

As before, we can obtain

$$\frac{1}{(2i+1)!}\left|\mathrm{E}\left[\tilde{\mathbf{x}}\left(D_{\tilde{\mathbf{x}}}^{2i+1}f\right)^T\right]_{ab}\right|$$

$$=\frac{1}{(2i+1)!}\left|\mathrm{E}\left[\tilde{x}_a\sum_{1\le k_1,\ldots,k_c\le 2i+1}\binom{2i+1}{k_1,\ldots,k_c}\tilde{x}_1^{k_1}\cdots\tilde{x}_c^{k_c}\frac{\partial^{2i+1}f_b(\mathrm{E}(\mathbf{x}))}{\partial x_1^{k_1}\cdots\partial x_c^{k_c}}\right]\right|$$

$$=\frac{1}{(2i+1)!}\left|\mathrm{E}\left[\sum_{1\le k_1,\ldots,k_c\le 2i+1}\binom{2i+1}{k_1,\ldots,k_c}\tilde{x}_1^{k_1}\cdots\tilde{x}_a^{k_a+1}\cdots\tilde{x}_c^{k_c}\frac{\partial^{2i+1}f_b(\mathrm{E}(\mathbf{x}))}{\partial x_1^{k_1}\cdots\partial x_c^{k_c}}\right]\right|$$

$$\le\frac{h^{i+1}M^{i+1}}{(2i+1)!}\sum_{1\le k_1,\ldots,k_c\le 2i+1}\left[\binom{2i+1}{k_1,\ldots,k_c}\left|\frac{\partial^{2i+1}f_b(\mathrm{E}(\mathbf{x}))}{\partial x_1^{k_1}\cdots\partial x_c^{k_c}}\right|\right]$$

$$\equiv h^{i+1}U_{i,ab}.$$

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,

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

$$=\frac{h^{i+1}}{(2i+1)!}\frac{(c+\lambda)^{i+1}}{c+\lambda}\sum_{j=1}^c(\sqrt{P^*}_{aj})$$

$$\times\left(\sum_{1\le k_1,\ldots,k_c\le 2i+1}\binom{2i+1}{k_1,\ldots,k_c}(\sqrt{P^*}_{1j})^{k_1}\cdots(\sqrt{P^*}_{cj})^{k_c}\frac{\partial^{2i+1}f_b(\mathrm{E}(\mathbf{x}))}{\partial x_1^{k_1}\cdots\partial x_c^{k_c}}\right)$$

$$\equiv h^{i+1}V_{i,ab}.$$

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

$$|\mathbf{P_{xy}}-\hat{\mathbf{P}}_{\mathbf{xy}}|=\left|\sum_{i=1}^\infty\frac{1}{(2i+1)!}\mathrm{E}\left[\tilde{\mathbf{x}}\left(D_{\tilde{\mathbf{x}}}^{2i+1}f\right)^T\right]\right.$$

$$\left.-\sum_{i=1}^\infty\frac{1}{(2i+1)!}\frac{1}{2(c+\lambda)}\sum_{j=1}^{2c}\tilde{\mathbf{x}}^{(j)}\left(D_{\tilde{\mathbf{x}}^{(j)}}^{2i+1}f\right)^T\right| \tag{A.21}$$

$$\le h^2\sum_{i=1}^\infty h^{i-1}(|\mathbf{U}_i|+|\mathbf{V}_i|)=O(h^2),$$

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 (A.21) are finite. In linearization, $\hat{\mathbf{P}}_{\mathbf{xy},L} = \mathbf{P_x}\mathbf{H}^T$. Thus, it can be similarly shown that

$$|\mathbf{P_{xy}} - \hat{\mathbf{P}}_{\mathbf{xy},L}| \le h^2 \sum_{i=1}^{\infty} h^{i-1}|\mathbf{U}_i| = O(h^2). \tag{A.22}$$