Advances in Derivative-Free State Estimation for Nonlinear Systems

Magnus Nørgaard[⋆◇] Niels K. Poulsen[⋆] Ole Ravn[◇]
pmn@iau.dtu.dk nkp@imm.dtu.dk or@iau.dtu.dk

[⋆]Department of Mathematical Modelling
[◇]Department of Automation
Technical University of Denmark
2800 Lyngby, Denmark

Technical report: IMM-REP-1998-15

October 29, 2004

REVISED EDITION

Abstract

In this paper we show that it has considerable advantages to use polynomial approximations obtained with an interpolation formula for derivation of state estimators for nonlinear systems. The estimators become more accurate than estimators based on Taylor approximations; yet the implementation is significantly simpler as no derivatives are required. Thus, it is believed that estimators derived in this way can replace well-known filters, such as the extended Kalman filter (EKF) and its higher order relatives, in most practical applications. In addition to proposing a new set of state estimators, the paper also unifies recent developments in derivative-free state estimation.

1 Introduction

When it comes to state estimation for nonlinear systems there is not a single solution available that clearly outperforms all other strategies. A series of estimators have been proposed over time, which for the most part are nonlinear extensions of the celebrated Kalman filter. For each application one therefore has to pick the estimator which is found to best trade off various properties such as estimation accuracy, ease of implementation, numerical robustness, and computational burden. Up to now the extended Kalman filter (EKF) [GKN+74], [May82], [Lew86] has unquestionably been the dominating state estimation technique. The EKF is based on first-order Taylor approximations of state transition and observation equations about the estimated state trajectory. Application of the filter is therefore contingent upon the assumption that the required derivatives exist and can be obtained with a reasonable effort. The Taylor linearization provides an insufficiently accurate representation in many cases, and significant bias, or even convergence problems, are commonly encountered due to the overly crude approximation. Several estimation techniques are available that are more sophisticated than the EKF, e.g., re-iteration, higher order filters, and statistical linearization [GKN+74], [May82]. The more advanced techniques generally improve estimation accuracy, but it happens at the expense of a further complication in implementation and an increased computational burden.

In this paper we propose a new set of estimators, which are based on polynomial approximations of the nonlinear transformations obtained with particular multidimensional extension of Stirling's interpolation formula [Ste27], [Frö70]. Conceptually, the principle underlying the new filters resembles that of the EKF and its higher order relatives. The implementation is, however, quite different. In contrast to the Taylor approximation no derivatives are needed in the interpolation formula; only function evaluations. This accommodates easy implementation of the filters, and it enables state estimation even when there are singular points in which the derivatives are undefined. Although the implementation is less complicated than for filters based on Taylor approximations, the computational burden will often be comparable in size or only slightly bigger. Additionally, under certain assumptions on the distribution of the estimation errors, the new filters provide a similar or even superior performance.

Recently there has been interesting developments in derivative-free state estimation techniques [JU94], [JUDW95], [JU97], [Sch97]. It is shown in the paper that these filters occur as special cases of filters based on the interpolation formula. The filter described in [Sch97] corresponds to a suboptimal implementation of the filter derived using first-order approximations while the filter proposed in [JU94], [JUDW95] has the same *a priori* state estimate and a related (but less accurate) covariance estimate as the filter derived using second-order approximations. Due to these relationships we have found it natural to adopt some of the ideas on practical implementation suggested in [Sch97] and to analyze the performance of the filters by using the same approach as in [JU94].

The paper is organized as follows. First we introduce Stirling's interpolation formula and discuss under which circumstances it can provide more accurate approximations than Taylor's formula. A multidimensional extension of the interpolation formula is made, and it is discussed how it can be used for approximation of mean and covariance of stochastic variables generated by nonlinear transformation of stochastic variables with known mean and covariance. Based on the obtained results, two new filters are proposed. The DD1 filter is based on first-order approximations and the DD2 filter is based on second-order approximations. The performance of the new filters are demonstrated on a benchmark example. Readers only interested in the actual filter implementation may choose to skip Section 2 and Section 3.

2 Power Series Revisited

This section deals with polynomial approximations of arbitrary functions. In particular we will compare approximations obtained with Taylor's formula, which commonly underlies filters for nonlinear systems, with approximations obtained with an interpolation formula. Initially, functions of only one variable will be considered. Later the treatment is extended to multiple dimensions.

If the function f is analytic we can represent it by its Taylor series expanded about some point, $x = \bar{x}$

$$f(x) = f(\bar{x}) + f'(\bar{x})(x-\bar{x}) + \frac{f''(\bar{x})}{2!}(x-\bar{x})^2 + \frac{f^{(3)}(\bar{x})}{3!}(x-\bar{x})^3 + \dots \tag{1}$$

A commonly used approximation is obtained by truncating the series after a finite number of terms. As more terms are included, a locally better approximation is achieved since the remainder (the sum of high-order terms) converges as $O(|x-\bar{x}|^{n+1})$ (this holds even when f is not analytic). The principle of the Taylor series is that the approximation inherits still more characteristics of the true function in one particular point as the number of terms increases. Although the assumption that f is analytic implies that any desired accuracy can be achieved provided that a sufficient number of terms are retained, it is in general adviced to use a truncated series only in the proximity of the expansion point unless the remainder term has been properly analyzed.

Several *interpolation formulas* are available for deriving polynomial approximations that are to be used over an interval. Most of these do not require derivatives but are instead based on a finite number of evaluations of the function. Usually it is therefore much simpler to derive approximations with these formulas. Several textbooks are available that deal with interpolation, e.g., [DB74], [Ste27], [Frö70]. In the following we will consider one particular formula, namely *Stirling's* interpolation formula. Let the operators δ and μ perform the following operations (h denotes a selected *interval length*)

$$\delta f(x) = f(x+\frac{h}{2}) - f(x-\frac{h}{2}) \tag{2}$$

$$\mu f(x) = \frac{1}{2}\left(f(x+\frac{h}{2}) + f(x-\frac{h}{2})\right) . \tag{3}$$

With these operators Stirling's interpolation formula used around the point $x = \bar{x}$

can be expressed as [Frö70]

$$
\begin{aligned}
f(x) = f(\bar{x} + ph) \; = \; & f(\bar{x}) + p\mu\delta f(\bar{x}) + \tfrac{p^2}{2!}\delta^2 f(\bar{x}) + \begin{pmatrix} p+1 \\ 3 \end{pmatrix} \mu\delta^3 f(\bar{x}) \\
& + \tfrac{p^2(p^2-1)}{4!}\delta^4 f(\bar{x}) + \begin{pmatrix} p+2 \\ 5 \end{pmatrix} \mu\delta^5 f(\bar{x}) + \dots
\end{aligned}
\tag{4}
$$

Commonly, $-1 < p < 1$, but in our application we will allow occasional use outside this interval as we shall see in later sections.

In this paper the attention is restricted to first and second-order polynomial approximations. The formula (4) is in this case particularly simple

$$
f(x) \approx f(\bar{x}) + f'_{DD}(\bar{x})(x - \bar{x}) + \frac{f''_{DD}(\bar{x})}{2!}(x - \bar{x})^2 , \tag{5}
$$

where

$$
f'_{DD}(\bar{x}) = \frac{f(\bar{x} + h) - f(\bar{x} - h)}{2h} \qquad f''_{DD}(\bar{x}) = \frac{f(\bar{x} + h) + f(\bar{x} - h) - 2f(\bar{x})}{h^2} . \tag{6}
$$

One can be interpret (5) as a Taylor approximation with the derivatives replaced by central divided differences. To assess the accuracy of the approximation it is useful to insert the full Taylor series (1) in place of $f(\bar{x}+h)$ and $f(\bar{x}-h)$. We must assume that f is analytic to carry out this analysis

$$
\begin{aligned}
& f(\bar{x}) + f'_{DD}(\bar{x})(x - \bar{x}) + \frac{f''_{DD}(\bar{x})}{2!}(x - \bar{x})^2 = \\
& \quad f(\bar{x}) + f'(\bar{x})(x - \bar{x}) + \tfrac{f''(\bar{x})}{2!}(x - \bar{x})^2 \\
& \quad + \left(\tfrac{f^{(3)}(\bar{x})}{3!} h^2 + \tfrac{f^{(5)}(\bar{x})}{5!} h^4 + \dots \right)(x - \bar{x}) + \left(\tfrac{f^{(4)}(\bar{x})}{4!} h^2 + \tfrac{f^{(6)}(\bar{x})}{6!} h^4 + \dots \right)(x - \bar{x})^2 .
\end{aligned}
\tag{7}
$$

The first three terms on the right hand side of (7) are independent of the interval length, h, and are recognized as the first three terms of the Taylor series expansion of f. The "remainder" term given by the difference between (7) and the second-order Taylor approximation is controlled by h and will in general deviate from the higher order terms of the Taylor series expansion of f. As we shall see in the following section, the possibility of controlling the remainder term is what makes the interpolation formula more attractive than Taylor approximation in some applications. Certain interval lengths can ensure that the remainder term in some sense will be close to the higher order terms of the full Taylor series. Fig. 1 shows a typical example on the difference between a Taylor approximation and an approximation obtained with the interpolation formula.

We will now proceed with the multidimensional case. Let x be a vector, $x \in R^n$, and let $y = f(x)$ be a vector function. There are different ways in which the interpolation

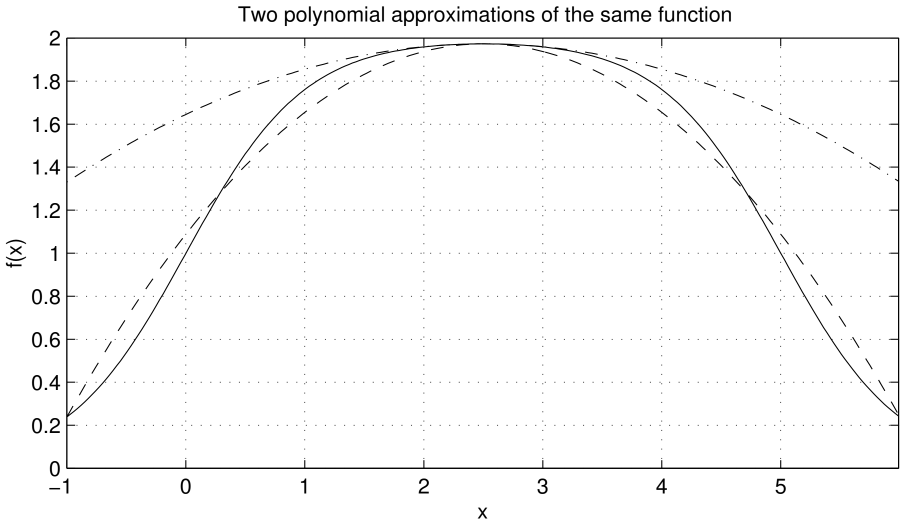

Figure 1. *Comparison of a second-order polynomial approximation obtained with Taylor's formula and one obtained with the interpolation formula. The expansion point is* $\bar{x} = 2.5$ *and for the interpolation formula the interval length was selected to* $h = 3.5$. *The solid line shows the true function, the dot-dashed line is the second-order Taylor approximation while the dashed line is the approximation obtained with the interpolation formula. Obviously, the Taylor polynomial is a better approximation near the expansion point while further away the error is much higher than for the approximation obtained with the interpolation formula.*

formula can be extended to multiple dimensions but before addressing this recall first that the multidimensional Taylor series expansion of f about $x = \bar{x}$ is given by

$$\begin{aligned} y = f(\bar{x} + \Delta x) &= \sum_{i=0}^{\infty} \frac{1}{i!} D^i_{\Delta x} f \\ &= f(\bar{x}) + D_{\Delta x} f + \tfrac{1}{2!} D^2_{\Delta x} f + \tfrac{1}{3!} D^3_{\Delta x} f + \ldots \end{aligned} \tag{8}$$

where the operator description employed by [JU94] has been adopted:

$$D^i_{\Delta x} f = \left(\Delta x_1 \frac{\partial}{\partial x_1} + \Delta x_2 \frac{\partial}{\partial x_2} + \cdots + \Delta x_n \frac{\partial}{\partial x_n} \right)^i f(x) \bigg|_{x=\bar{x}} . \tag{9}$$

The operators can also be written:

$$\begin{aligned} D_{\Delta x} f &= \left(\sum_{p=1}^{n} \Delta x_p \frac{\partial}{\partial x_p} \right) f(x) \bigg|_{x=\bar{x}} \\ D^2_{\Delta x} f &= \left(\sum_{p=1}^{n} \sum_{q=1}^{n} \Delta x_p \Delta x_q \frac{\partial^2}{\partial x_p \partial x_q} \right) f(x) \bigg|_{x=\bar{x}} \\ &\vdots \end{aligned} \tag{10}$$

By again restricting our attention to second-order polynomials we will write the multidimensional interpolation formula as

$$y \approx f(\bar{x}) + \tilde{D}_{\Delta x} f + \frac{1}{2!}\tilde{D}^2_{\Delta x} f \ . \tag{11}$$

As the divided difference operators, $\tilde{D}_{\Delta x}$, $\tilde{D}^2_{\Delta x}$, we will use

$$\tilde{D}_{\Delta x} f = \frac{1}{h}\left(\sum_{p=1}^{n} \Delta x_p \mu_p \delta_p\right) f(\bar{x}) \tag{12}$$

$$\tilde{D}^2_{\Delta x} f = \frac{1}{h^2}\left(\sum_{p=1}^{n} (\Delta x_p)^2 \delta_p^2 + \sum_{p=1}^{n} \sum_{q=1, q\neq p}^{n} \Delta x_p \Delta x_q (\mu_p \delta_p)(\mu_q \delta_q)\right) f(\bar{x}) \ , \tag{13}$$

where δ_p has been introduced as the "partial" difference operator

$$\delta_p f(\bar{x}) = f(\bar{x} + \frac{h}{2} e_p) - f(\bar{x} - \frac{h}{2} e_p) \ , \tag{14}$$

and e_p is the pth unit vector. A similar extension was made of the average operator μ.

The formula (11) is just one example of a multidimensional extension of the interpolation formula. To illustrate how others can be derived, the following linear transformation of x is introduced:

$$z = S^{-1} x \ , \tag{15}$$

and the function $\tilde{f}$ is defined by

$$\tilde{f}(z) \equiv f(Sz) = f(x) \ . \tag{16}$$

While the Taylor approximation of $\tilde{f}$ is identical to that of f, it is obviously not the case that the multidimensional interpolation formula (11) yields the same results for f and $\tilde{f}$. Since

$$2\mu_p \delta_p \tilde{f}(\bar{z}) = \tilde{f}(\bar{z} + h e_p) - \tilde{f}(\bar{z} - h e_p) = f(\bar{x} + h s_p) - f(\bar{x} - h s_p) \ , \tag{17}$$

where s_p denotes the pth column of S, $\tilde{D}_{\Delta x} f$ and $\tilde{D}^2_{\Delta x} f$ will clearly deviate from $\tilde{D}_{\Delta z} \tilde{f}$ and $\tilde{D}^2_{\Delta z} \tilde{f}$.

In the following section we are going to use the interpolation formula in a stochastic framework. In this case a particularly useful choice of transformation matrix (S) and interval length (h) exists.

3 Approximation of Mean and Covariance

Let x be a vector of stochastic variables for which the expectation and covariance are available

$$\bar{x} = E[x], \quad P_x = E\left[(x - \bar{x})(x - \bar{x})^T\right] \ . \tag{18}$$

We would now like to determine

$$\begin{aligned} \bar{y}_T &= E[f(x)] && (19)\\ (P_y)_T &= E\left[(f(x)-\bar{y}_T)(f(x)-\bar{y}_T)^T\right] && (20)\\ (P_{xy})_T &= E\left[(x-\bar{x})(f(x)-\bar{y}_T)^T\right] \ . && (21) \end{aligned}$$

As f is nonlinear we cannot rely on being able to calculate the exact expectations. Instead it is customary to insert a first or second-order polynomial approximation in place of f before taking the expectations. In this section we will focus on estimates of the expectations obtained using the interpolation formula in (11) for approximation of f. Additionally, we shall find it particularly useful to work with a linear transformation of x as described above. The transformation matrix is selected as a square Cholesky factor of the covariance matrix [Sch97]:

$$z = S_x^{-1}x, \qquad P_x = S_x S_x^T \ . \tag{22}$$

This transformation is sometimes said to perform a *stochastic decoupling* of the variables in x as the elements of z become mutually uncorrelated (and each with unity variance):

$$E\left[(z-E[z])(z-E[z])^T\right] = I \ . \tag{23}$$

We shall in the following use a rather wide interpretation of the so-called Cholesky factorization. For any symmetric matrix product $M = SS^T$ we will refer to S as a *Cholesky factor*. Thus, the Cholesky factor need not be square and triangular. However, most often a triangular Cholesky factor is considered as computationally efficient methods are available for performing such factorizations.

In the following subsections we shall work with $\tilde{f}(z)$ directly as this is most convenient. A few assumptions on $\tilde{f}$ (f) and z will be invoked. $\tilde{f}$ must in principle be defined for all $z \in R^n$ and the elements of $\Delta z = z - E[\Delta z]$ are assumed to belong to the same (zero mean) distribution. In Section 3.2 it is additionally assumed that Δz is Gaussian. For analysis purposes it is in Section 3.3 assumed that $\tilde{f}$ is analytic and that Δz is Gaussian. It should be stressed, however, that it is *not* necessary for $\tilde{f}$ to be analytic to apply the estimators.

3.1 A First-order Approximation

First estimates of mean and covariance will be derived by replacing the function $\tilde{f}$ by a first-order approximation

$$y = \tilde{f}(\bar{z}+\Delta z) \approx \tilde{f}(\bar{z}) + \tilde{D}_{\Delta z}\tilde{f} \ . \tag{24}$$

As the expectation $E[\Delta z] = 0$ by definition, the expectation of (24) is

$$\boxed{\bar{y} = E[\tilde{f}(\bar{z}) + \tilde{D}_{\Delta z}\tilde{f}] = \tilde{f}(\bar{z}) = f(\bar{x})} \tag{25}$$

An estimate of the covariance (20) is derived along the same lines. As before, the first-order moments can be neglected since Δz is zero mean. Moreover, the cross-terms evaluate to zero as z has been generated so that the cross-correlations $E[\Delta z_i \Delta z_j] = 0, i \neq j$.

$$\begin{aligned}
P_y &= E\left[\left(\tilde{f}(\bar{z}) + \tilde{D}_{\Delta z}\tilde{f} - \tilde{f}(\bar{z})\right)\left(\tilde{f}(\bar{z}) + \tilde{D}_{\Delta z}\tilde{f} - \tilde{f}(\bar{z})\right)^T\right] \\
&= E\left[\left(\tilde{D}_{\Delta z}\tilde{f}(z)\right)\left(\tilde{D}_{\Delta z}\tilde{f}(z)\right)^T\right] \\
&= E\left[\left(\sum_{p=1}^{n} \Delta z_p \mu_p \delta_p \tilde{f}(\bar{z})\right)\left(\sum_{p=1}^{n} \Delta z_p \mu_p \delta_p \tilde{f}(z)\right)^T\right] \\
&= \sigma_2 \sum_{p=1}^{n} \left(\mu_p \delta_p \tilde{f}(\bar{z})\right)\left(\mu_p \delta_p \tilde{f}(z)\right)^T \\
&= \frac{1}{4h^2} \sum_{p=1}^{n} \left(\tilde{f}(\bar{z} + he_p) - \tilde{f}(\bar{z} - he_p)\right)\left(\tilde{f}(\bar{z} + he_p) - \tilde{f}(\bar{z} - he_p)\right)^T . \qquad (26)
\end{aligned}$$

We shall denote the ith moment of an arbitrary element in Δz by σ_i. As all elements are assumed to be equally distributed their moments are obviously identical. As discussed above, $\sigma_2 = 1$. Higher moments depend on the distribution of Δz.

Recalling that $\tilde{f}(\bar{z} \pm he_p) = f(\bar{x} \pm hs_{x,p})$, where $s_{x,p}$ is the pth column of the square Cholesky factor of the covariance matrix S_x, (26) can also be written

$$\boxed{P_y = \frac{1}{4h^2} \sum_{p=1}^{n} \left(f(\bar{x} + hs_{x,p}) - f(\bar{x} - hs_{x,p})\right)\left(f(\bar{x} + hs_{x,p}) - f(\bar{x} - hs_{x,p})\right)^T} \qquad (27)$$

The estimate of the cross-covariance matrix can be derived a long the same lines

$$\begin{aligned}
P_{xy} &= E\left[(x - \bar{x})\left(\tilde{f}(\bar{z}) + \tilde{D}_{\Delta z}\tilde{f} - \tilde{f}(\bar{z})\right)^T\right] \\
&= E\left[(S_x \Delta z)\left(\tilde{D}_{\Delta z}\tilde{f}\right)^T\right] \\
&= E\left[\sum_{p=1}^{n} s_{x,p} \Delta z_p \left(\sum_{p=1}^{n} \Delta z_p \mu_p \delta_p \tilde{f}(z)\right)^T\right] \\
&= \sigma_2 \left[\sum_{p=1}^{n} s_{x,p} \left(\mu_p \delta_p \tilde{f}(z)\right)^T\right] \\
&= \frac{1}{2h} \sum_{p=1}^{n} s_{x,p} \left(\tilde{f}(\bar{z} + he_p) - \tilde{f}(\bar{z} - he_p)\right)^T , \qquad (28)
\end{aligned}$$

which we can also write

$$\boxed{P_{xy} = \frac{1}{2h}\sum_{p=1}^{n} s_{x,p}\left(f(\bar{x}+hs_{x,p}) - f(\bar{x}-hs_{x,p})\right)^T} \tag{29}$$

It is not clear from the derivations how the interval length, h, should be selected. The mean estimate is independent of the parameter while it has an obvious impact on the estimate of the covariance matrices. In Section 3.3 covering the analysis of the estimates it is shown that the optimal setting of h is dictated by the distribution of Δz. It turns out that h^2 should equal the kurtosis of the distribution, $h^2 = \sigma_4$.

3.2 A Second-order Approximation

More accurate estimates of mean and covariance of $\tilde{f}$ can be obtained with a limited extra effort by approximating the function with a second-order polynomial derived with the interpolation formula:

$$\begin{aligned} y &\approx \tilde{f}(\bar{z}) + \tilde{D}_{\Delta z}\tilde{f} + \frac{1}{2}\tilde{D}^2_{\Delta z}\tilde{f} \\ &= \tilde{f}(\bar{z}) + \frac{1}{h}\left(\sum_{p=1}^{n}\Delta z_p \mu_p \delta_p\right)\tilde{f}(\bar{z}) \\ &+ \frac{1}{2h^2}\left(\sum_{p=1}^{n}(\Delta z_p)^2\delta_p^2 + \sum_{p=1}^{n}\sum_{q=1,q\neq p}^{n}\Delta z_p \Delta z_q(\mu_p\delta_p)(\mu_q\delta_q)\right)\tilde{f}(\bar{z})\,. \end{aligned} \tag{30}$$

To obtain useful results the assumptions on Δz will now be slightly more restrictive as we demand that it is Gaussian. Since Δz is zero mean and the elements are uncorrelated, this new assumption implies that the elements are independent and the distribution is symmetric. The assumption is not needed for derivation of the mean estimate, but it is important when deriving the improved covariance estimate.

Utilizing that Δz is zero mean and its elements are uncorrelated, the expectation of $\tilde{f}$ can be estimated by

$$\begin{aligned} \bar{y} &= E\left[\tilde{f}(\bar{z}) + \frac{1}{2}\left(\sum_{p=1}^{n}(\Delta z_p)^2\delta_p^2\right)\tilde{f}(\bar{z})\right] \\ &= \tilde{f}(\bar{z}) + \frac{\sigma_2}{2}\sum_{p=1}^{n}\delta_p^2\tilde{f}(\bar{z}) \\ &= \tilde{f}(\bar{z}) + \frac{1}{2h^2}\sum_{p=1}^{n}\left(\tilde{f}(\bar{z}+he_p) + \tilde{f}(\bar{z}-he_p)\right) - \frac{n}{h^2}\tilde{f}(\bar{z}) \\ &= \frac{h^2-n}{h^2}\tilde{f}(\bar{z}) + \frac{1}{2h^2}\sum_{p=1}^{n}\left(\tilde{f}(\bar{z}+he_p) + \tilde{f}(\bar{z}-he_p)\right) \end{aligned} \tag{31}$$

$\Updownarrow$

$$\boxed{\bar{y} = \frac{h^2 - n}{h^2} f(\bar{x}) + \frac{1}{2h^2} \sum_{p=1}^{n} f(\bar{x} + h s_{x,p}) + f(\bar{x} - h s_{x,p})} \tag{32}$$

We will now proceed with a derivation of a covariance estimate. First we observe that

$$\begin{aligned} (P_y)_T &= E[(y - \bar{y})(y - \bar{y})^T] \\ &= E[(y - \tilde{f}(\bar{z}))(y - \tilde{f}(\bar{z}))^T] - E[y - \tilde{f}(\bar{z})]E[y - \tilde{f}(\bar{z})]^T . \end{aligned} \tag{33}$$

The estimate can therefore be written

$$\begin{aligned} P_y &= E\left[\left(\tilde{D}_{\Delta z}\tilde{f} + \frac{1}{2}\tilde{D}^2_{\Delta z}\tilde{f}\right)\left(\tilde{D}_{\Delta z}\tilde{f} + \frac{1}{2}\tilde{D}^2_{\Delta z}\tilde{f}\right)^T\right] \\ &- E\left[\tilde{D}_{\Delta z}\tilde{f} + \frac{1}{2}\tilde{D}^2_{\Delta z}\tilde{f}\right] E\left[\tilde{D}_{\Delta z}\tilde{f} + \frac{1}{2}\tilde{D}^2_{\Delta z}\tilde{f}\right]^T \\ &= E\left[\tilde{D}_{\Delta z}\tilde{f}\left(\tilde{D}_{\Delta z}\tilde{f}\right)^T\right] + \frac{1}{4}E\left[\tilde{D}^2_{\Delta z}\tilde{f}\left(\tilde{D}^2_{\Delta z}\tilde{f}\right)^T\right] \\ &- \frac{1}{4}E\left[\tilde{D}^2_{\Delta z}\tilde{f}\right]E\left[\tilde{D}^2_{\Delta z}\tilde{f}\right]^T . \end{aligned} \tag{34}$$

The second step was taken by using the fact that all odd order moments cancel as the elements of Δz are independent and the distribution symmetric. The first term in (34) is recognized as the covariance based on a first-order approximation of $\tilde{f}$ and has already been dealt with. Let us instead take a closer look at the two remaining terms:

$E\left[\tilde{D}^2_{\Delta z}\tilde{f}\left(\tilde{D}^2_{\Delta z}\tilde{f}\right)^T\right]$ is composed of 3 kinds of terms

$$E\left[\left((\Delta z_i)^2\delta_i^2\tilde{f}\right)\left((\Delta z_i)^2\delta_i^2\tilde{f}\right)^T\right] = \left(\delta_i^2\tilde{f}\right)\left(\delta_i^2\tilde{f}\right)^T\sigma_4 , \tag{35}$$

$$E\left[\left((\Delta z_i)^2\delta_i^2\tilde{f}\right)\left((\Delta z_j)^2\delta_j^2\tilde{f}\right)^T\right] = \left(\delta_i^2\tilde{f}\right)\left(\delta_j^2\tilde{f}\right)^T\sigma_2^2 , \tag{36}$$

$$E\left[\left(\Delta z_i\Delta z_j\mu_i\delta_i\mu_j\delta_j\tilde{f}\right)\left(\Delta z_i\Delta z_j\mu_i\delta_i\mu_j\delta_j\tilde{f}\right)^T\right] = \left(\mu_i\delta_i\mu_j\delta_j\tilde{f}\right)\left(\mu_i\delta_i\mu_j\delta_j\tilde{f}\right)^T\sigma_2^2 . \tag{37}$$

$E\left[\tilde{D}^2_{\Delta z}\tilde{f}\right]E\left[\tilde{D}^2_{\Delta z}\tilde{f}\right]^T$ is composed of 2 kinds of terms

$$E\left[(\Delta z_i)^2\delta_i\tilde{f}\right]E\left[(\Delta z_i)^2\delta_i\tilde{f}\right]^T = \left(\delta_i^2\tilde{f}\right)\left(\delta_i^2\tilde{f}\right)^T\sigma_2^2 , \tag{38}$$

$$E\left[(\Delta z_i)^2\delta_i\tilde{f}\right]E\left[(\Delta z_j)^2\delta_j\tilde{f}\right]^T = \left(\delta_i^2\tilde{f}\right)\left(\delta_j^2\tilde{f}\right)^T\sigma_2^2 . \tag{39}$$

All of the above terms appear for $\forall i, \forall j, i \neq j$.

The terms in (36) and (39) are identical and cancel when subtracted. Additionally, we will discard the terms containing cross-differences (37). This is done because their inclusion would lead to an excessive increase in the amount of computations as the number of such terms grows rapidly with the dimension of z. Moreover, the terms each require four additional evaluations of f for each dimension. The reason for not considering the extra effort worthwhile is that we are unable to capture all fourth moments anyway. This would require that f was approximated by a third-order polynomial (more details on this are given in Section 3.3).

Thus, we arrive at the following covariance estimate

$$\begin{aligned}
P_y &= \sigma_2 \sum_{p=1}^{n} \left(\mu_p \delta_p \tilde{f}(\bar{z})\right) \left(\mu_p \delta_p \tilde{f}(z)\right)^T + \frac{\sigma_4 - \sigma_2^2}{4} \sum_{p=1}^{n} \left(\delta_p^2 \tilde{f}(\bar{z})\right) \left(\delta_p^2 \tilde{f}(z)\right)^T \\
&= \frac{\sigma_2}{4h^2} \sum_{p=1}^{n} \left(\tilde{f}(\bar{z} + he_p) - \tilde{f}(\bar{z} - he_p)\right) \left(\tilde{f}(\bar{z} + he_p) - \tilde{f}(\bar{z} - he_p)\right)^T \\
&+ \frac{\sigma_4 - \sigma_2^2}{4h^4} \sum_{p=1}^{n} \left(\tilde{f}(\bar{z} + he_p) + \tilde{f}(\bar{z} - he_p) - 2\tilde{f}(\bar{z})\right) \times \\
&\qquad \left(\tilde{f}(\bar{z} + he_p) + \tilde{f}(\bar{z} - he_p) - 2\tilde{f}(\bar{z})\right)^T . \qquad (40)
\end{aligned}$$

Inserting that $\sigma_2 = 1$ and setting $h^2 = \sigma_4$ ($= 3$ for a Gaussian distribution) give

$$\boxed{\begin{aligned}
P_y &= \tfrac{1}{4h^2} \textstyle\sum_{p=1}^{n} \left[f(\bar{x} + hs_{x,p}) - f(\bar{x} - hs_{x,p})\right] \left[f(\bar{x} + hs_{x,p}) - f(\bar{x} - hs_{x,p})\right]^T \\
&+ \tfrac{h^2-1}{4h^4} \textstyle\sum_{p=1}^{n} \left[f(\bar{x} + hs_{x,p}) + f(\bar{x} - hs_{x,p}) - 2f(\bar{x})\right] \times \\
&\qquad \left[f(\bar{x} + hs_{x,p}) + f(\bar{x} - hs_{x,p}) - 2f(\bar{x})\right]^T
\end{aligned}} \qquad (41)$$

As

$$\sigma_4 - \sigma_2^2 = E[(\Delta z)^4] - E[(\Delta z)^2]^2 = Var[(\Delta z)^2] > 0 \; , \qquad (42)$$

$\sigma_4 \geq \sigma_2^2$ for all probability distributions. Therefore, we should always select $h^2 \geq 1$. Obviously, this implies that the covariance estimate will always be positive semidefinite.

The cross-covariance estimate, P_{xy}, turns out to be the same as when the first-order approximation is employed (29):

$$\begin{aligned}
P_{xy} &= E\left[(S_x \Delta z) \left(\tilde{D}_{\Delta z} \tilde{f} + \frac{1}{2} \tilde{D}_{\Delta z}^2 \tilde{f}\right)^T\right] \\
&= E\left[(S_x \Delta z) \left(\tilde{D}_{\Delta z} \tilde{f}\right)^T\right] \\
&= \frac{1}{2h} \sum_{p=1}^{n} s_{x,p} \left(f(\bar{x} + hs_{x,p}) - f(\bar{x} - hs_{x,p})\right)^T . \qquad (43)
\end{aligned}$$

3.3 Analysis of the Approximations

In this section the performance of the proposed mean and covariance estimators will be evaluated. The analysis proceeds according to the approach employed in [JU94]. That is, under the assumption that z is Gaussian and the function f is analytic, the Taylor series of the true mean and covariance are compared on a term-by-term basis with the Taylor series expansion of the estimators.

The derivative operator, $D^i_{\Delta z}$, has already been introduced in (9):

$$D^i_{\Delta z}\tilde{f} = \left(\sum_{p=1}^{n} \Delta z_p \frac{\partial}{\partial z_p} \right)^i \tilde{f}(z) \bigg|_{z=\bar{z}} . \tag{44}$$

Additionally, the following partial derivative operator will be useful during the analysis:

$$D^i_{he_p}\tilde{f} = h^i \nabla^i_p \tilde{f} = h^i \frac{\partial^i \tilde{f}(z)}{\partial z_p^i} \bigg|_{z=\bar{z}} . \tag{45}$$

It is not difficult to see that

$$\frac{1}{h^2} \sum_{p=1}^{n} D^i_{he_p}\tilde{f} \;=\; h^{i-2} \sum_{p=1}^{n} \nabla^i_p \tilde{f} \tag{46}$$

$$E\left[D^i_{\Delta z}\tilde{f} \right] \;=\; \sigma_i \sum_{p=1}^{n} \nabla^i_p \tilde{f} \quad + \quad [\text{cross-terms if } i \geq 4] . \tag{47}$$

It was mentioned previously that the Gaussian assumption implies that the elements of Δz are mutually independent and that the distribution of Δz is symmetric. Thus, all odd moments evaluate to zero in (47). The cross-terms are terms containing products of derivatives w.r.t. different variables and terms containing cross-derivatives. In a similar fashion we can evaluate the products:

$$\frac{1}{h^2} \sum_{p=1}^{n} D^i_{he_p}\tilde{f} \left(D^j_{he_p}\tilde{f} \right)^T \;=\; h^{i+j-2} \sum_{p=1}^{n} \left(\nabla^i_p \tilde{f} \right) \left(\nabla^j_p \tilde{f} \right)^T \tag{48}$$

$$E\left[D^i_{\Delta z}\tilde{f} \left(D^j_{\Delta z}\tilde{f} \right)^T \right] \;=\; \sigma^{i+j} \sum_{p=1}^{n} \left(\nabla^i_p \tilde{f} \right) \left(\nabla^j_p \tilde{f} \right)^T$$

$$+ \quad [\text{cross-terms if } i+j \geq 4] . \tag{49}$$

For the reasons called attention to above, (49) evaluate to zero for $i+j$ odd.

If, for a moment, we neglect the cross-terms in (47) and (49), the difference between the pair (46), (48) and the pair (47), (49) is for the even terms alone given by the discrepancy between h^{i+j} and σ_{i+j+2}. As Δz is Gaussian we have that [Pap84] $\sigma_{2i} = 1 \times 3 \times \cdots \times (2i-1)\sigma_2^i$. Thus, the moment grows factorially with i. As $\sigma_2 = 1$

we have $\sigma_{2i} = \{1, 3, 15, 105, ...\}$. In the second-order case (i.e., $i = 2$ or $i = j = 1$, respectively) the terms will agree regardless of the choice of h. If we select h^2 as the kurtosis, $h^2 = \sigma_4 = 3$, the terms will also agree in the fourth-order case (except for the cross-terms, which remain unmatched). In the higher order cases, (46) and (48) will underestimate (47) and (49), respectively, as $h^{2(i+j)}$ grows geometrically and therefore will be exceeded by $\sigma_{2i+2j+2}$ from the sixth order.

Series expansion of the true quantities

First the Taylor series expansion of the true expressions for mean and covariances (19), (20), (21) are determined. As the Taylor series of $\tilde{f}$ expanded around $z = \bar{z}$ is given by

$$y = \tilde{f}(z) + \sum_{i=1}^{\infty} \frac{D^{2i}_{\Delta z}\tilde{f}}{(2i)!} \tag{50}$$

we have for the true mean

$$\begin{aligned} \bar{y}_T = E[y] &= \tilde{f}(\bar{z}) + E\left[\sum_{i=1}^{\infty} \frac{D^{2i}_{\Delta z}\tilde{f}}{(2i)!}\right] \\ &= \tilde{f}(\bar{z}) + \sum_{i=1}^{\infty} \frac{\sigma_{2i}}{(2i)!} \sum_{p=1}^{n} \nabla_p^{2i}\tilde{f} + [\text{cross-terms if } i \geq 4]. \end{aligned} \tag{51}$$

For the true covariance we get

$$\begin{aligned} (P_y)_T &= E[(y - \tilde{f}(\bar{z}))(y - \tilde{f}(\bar{z}))^T] - E[y - \tilde{f}(\bar{z})]E[y - \tilde{f}(\bar{z})]^T \\ &= E\left[\sum_{i=1}^{\infty} \frac{D^{i}_{\Delta z}\tilde{f}}{i!} \sum_{j=1}^{\infty} \frac{(D^{j}_{\Delta z}\tilde{f})^T}{j!}\right] - E\left[\sum_{i=1}^{\infty} \frac{D^{2i}_{\Delta z}\tilde{f}}{(2i)!}\right] E\left[\sum_{i=1}^{\infty} \frac{D^{2i}_{\Delta z}\tilde{f}}{(2i)!}\right] \\ &= E\left[D_{\Delta z}\tilde{f}\left(D_{\Delta z}\tilde{f}\right)^T + \frac{D_{\Delta z}\tilde{f}(D^3_{\Delta z}\tilde{f})^T}{3!} + \frac{D^2_{\Delta z}\tilde{f}(D^2_{\Delta z}\tilde{f})^T}{2 \times 2!} + \frac{D^3_{\Delta z}\tilde{f}(D_{\Delta z}\tilde{f})^T}{3!}\right] \\ &- \quad E\left[\frac{D^2_{\Delta z}\tilde{f}}{2!}\right] E\left[\frac{D^2_{\Delta z}\tilde{f}}{2!}\right]^T \quad + \quad \cdots \\ &= \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \frac{\sigma_{2(i+j)+2}}{(2i+1)!(2j+1)!} \sum_{p=1}^{n} \left(\nabla_p^{2i+1}\tilde{f}\right)\left(\nabla_p^{2j+1}\tilde{f}\right)^T \\ &+ \quad \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} \frac{\sigma_{2(i+j)} - \sigma_{2i}\sigma_{2j}}{(2i)!(2j)!} \sum_{p=1}^{n} \left(\nabla_p^{2i}\tilde{f}\right)\left(\nabla_p^{2j}\tilde{f}\right)^T \\ &+ \quad [\text{cross-terms if } i + j \geq 4] \end{aligned} \tag{52}$$

while for the cross-covariance we have

$$(P_{xy})_T = E\left[(x - \bar{x})(\tilde{f}(z) - \bar{y}_T)^T\right]$$

$$
\begin{aligned}
&= E\left[(x-\bar{x})(\tilde{f}(z)-\tilde{f}(\bar{z}))^T\right] \\
&= E\left[S_x\Delta z\left(\sum_{i=0}^{\infty}\frac{(D_{\Delta z}^{2i+1}\tilde{f})^T}{(2i+1)!}\right)\right] \\
&= E\left[S_x\Delta z\left(D_{\Delta z}\tilde{f}\right)^T\right]+E\left[\frac{S_x\Delta z(D_{\Delta z}^3\tilde{f})^T}{(3)!}\right] \quad + \quad \cdots \\
&= \sum_{p=1}^{n}s_{x,p}\left(\sum_{i=0}^{\infty}\frac{\sigma_{2i+2}}{(2i+1)!}\nabla_p^{2i+1}\tilde{f}\right)^T+[\text{cross-terms if } i\geq 3]\,. \qquad (53)
\end{aligned}
$$

Series expansion of the mean estimates

The mean estimate based on the first-order approximation of $\tilde{f}$ (25) is simply the first term of the Taylor series:

$$\bar{y}=\tilde{f}(\bar{z}) \qquad (54)$$

while the Taylor series expansion of the mean estimate based on the second-order approximation in (31) is

$$
\begin{aligned}
\bar{y} &= \frac{h^2-n}{h^2}\tilde{f}(\bar{z})+\frac{1}{2h^2}\sum_{p=1}^{n}\tilde{f}(\bar{z}+he_p)+\tilde{f}(\bar{z}-hz_p) \\
&= \frac{h^2-n}{h^2}\tilde{f}(\bar{z})+\frac{1}{2h^2}\sum_{p=1}^{n}\left(2\tilde{f}(\bar{z})+\sum_{i=0}^{\infty}\frac{D_{he_p}^{i}\tilde{f}(\bar{z})}{i!}+\frac{(-D_{he_p})^i\tilde{f}(\bar{z})}{i!}\right) \\
&= \tilde{f}(\bar{z})+\frac{1}{h^2}\sum_{i=1}^{\infty}\sum_{p=1}^{n}\frac{D_{he_p}^{2i}\tilde{f}(\bar{z})}{(2i)!} \\
&= \tilde{f}(\bar{z})+\sum_{i=1}^{\infty}\frac{h^{2i-2}}{(2i)!}\sum_{p=1}^{n}\nabla_p^{2i}\tilde{f}\,. \qquad (55)
\end{aligned}
$$

The estimate based on a first-order approximation (54) is the same as if we had used an ordinary Taylor linearization of f. That is, the approximation error equals the second and higher order terms in the series expansion of the true mean (51).

For the estimate based on the second-order approximation we have the following approximation error for element k (obtained by subtracting (55) from (51)):

$$\bar{R}_2(k)=\sum_{i=3}^{\infty}\frac{\sigma_{2i}-h^{2i-2}}{(2i)!}\sum_{p=1}^{n}\nabla_p^{2i}\tilde{f}_k+\text{cross-terms.} \qquad (56)$$

Notice that the outer sum starts in $i=3$ as $h^2=\sigma_4$. Fourth-order derivatives are still present in the cross-terms, however. It is interesting to compare this approximation error to the error of a mean estimate obtained by employing a second-order Taylor approximation of $\tilde{f}$ as this is the traditional approach:

$$R_2(k) = \sum_{i=2}^{\infty} \frac{\sigma_{2i}}{(2i)!} \sum_{p=1}^{n} \nabla_p^{2i} \tilde{f}_k + \text{cross-terms}. \tag{57}$$

In the general case it is not possible to conclude that $|\bar{R}_2(k)|$ always will be smaller than $|R_2(k)|$ as the various derivatives can take any sign. However, one thing that can be said is that the magnitude of $R_2(k)$ will be bounded from above by

$$|R_2(k)| \leq M_2 = \sum_{i=2}^{\infty} \frac{\sigma_{2i}}{(2i)!} \left| \sum_{p=1}^{n} \nabla_p^{2i} \tilde{f}_k \right| + |\text{cross-terms}| \tag{58}$$

while $|\bar{R}_2(k)|$ will be bounded by

$$|\bar{R}_2(k)| \leq \bar{M}_2 = \sum_{i=3}^{\infty} \frac{\sigma_{2i} - h^{2i-2}}{(2i)!} \left| \sum_{p=1}^{n} \nabla_p^{2i} \tilde{f}_k \right| + |\text{cross-terms}| . \tag{59}$$

As $h^{2i-2} < \sigma_{2i}, \forall i \geq 3$ we have that $\bar{M}_2 \leq M_2$. The equality sign holds only when all the sums of derivatives in (58), (59) are 0. Thus, in general $|\bar{R}_2(k)|$ has a lower upper bound than $|R_2(k)|$.

To get an impression of the magnitude of the upper bound we observe that (recall that $\sigma_2 = 1$, $h^2 = \sigma_4 = 3$, $\sigma_{2i} = 1 \times 3 \times \cdots \times (2i-1)\sigma_2^{2i}$):

$$\left\{ \frac{\sigma_{2i}}{(2i)!} \right\}_1^{\infty} = \left\{ \frac{1}{2} \times \frac{1}{4} \times \frac{1}{6} \times \cdots \times \frac{1}{2i} \right\}_1^{\infty} = \{0.5, 0.125, 0.0208, 0.0026, \ldots\} \tag{60}$$

$$\left\{ \frac{h^{2i-2}}{(2i)!} \right\}_1^{\infty} = \left\{ \frac{1}{2}, \frac{1}{8}, \frac{1}{48}, \frac{1}{384}, \ldots \right\} = \{0.5, 0.125, 0.0125, 0.00067, \ldots\} . \tag{61}$$

Both fractions decay rapidly with i. Especially the fractions in (61) as the numerator in this case does not grow factorially. It is therefore reasonable to assume that also

$$\frac{\sigma_{2i}}{(2i)!} \left| \sum_{p=1}^{n} \nabla_p^{2i} \tilde{f}_k \right| \tag{62}$$

typically will decay rapidly with i and that the first few terms of the sum in (58) will dominate. If the upper bounds, $\bar{M}_2, M_2$, are not dominated by the cross-terms, $\bar{M}_2 \ll M_2$ as $\frac{\sigma_{2i} - h^{2i-2}}{(2i)!} \left| \sum_{p=1}^{n} \nabla_p^{2i} \tilde{f}_k \right|$ is 0 for $i = 2$ and less than half the size of (62) for $i = 3$. Recall that in the one-dimensional case there are no cross-terms. In this case errors are not introduced until the terms of order 6; i.e., a sixth-order Taylor approximation of $\tilde{f}$ would be necessary to achieve a better accuracy than what is offered by (55).

Series expansion of the covariance estimates

The same approach as above will now be used for assessing the accuracy of the covariance estimates. Note first that

$$\begin{aligned}
\frac{1}{2h}\left(\tilde{f}(\bar{z}+he_p)-\tilde{f}(\bar{z}-he_p)\right) &= \frac{1}{2h}\sum_{i=0}^{\infty}\frac{D^i_{he_p}\tilde{f}}{i!}-\frac{\left(-D_{he_p}\right)^i\tilde{f}}{i!} \\
&= \frac{1}{h}\sum_{i=0}^{\infty}\frac{D^{(2i+1)}_{he_p}\tilde{f}}{(2i+1)!} \qquad (63)\\
\frac{1}{2}\left(\tilde{f}(\bar{z}+he_p)+\tilde{f}(\bar{z}-he_p)-2\tilde{f}(\bar{z})\right) &= \frac{1}{2h}\sum_{i=0}^{\infty}\frac{D^i_{he_p}\tilde{f}}{i!}+\frac{\left(-D_{he_p}\right)^i\tilde{f}}{i!} \\
&= \sum_{i=1}^{\infty}\frac{D^{2i}_{he_p}\tilde{f}}{(2i)!}\,. \qquad (64)
\end{aligned}$$

Thus, when inserting the Taylor series in the estimate based on the first-order approximation (26) the following is obtained

$$\begin{aligned}
P_y &= \frac{1}{h^2}\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\sum_{p=1}^{n}\frac{D^{(2i+1)}_{he_p}\tilde{f}\left(D^{(2j+1)}_{he_p}\tilde{f}\right)^T}{(2i+1)!\,(2j+1)!} \\
&= \frac{1}{h^2}\sum_{p=1}^{n}D_{he_p}\tilde{f}\left(D_{he_p}\tilde{f}\right)^T \\
&\quad + \frac{1}{h^2}\sum_{p=1}^{n}\left(\frac{D_{he_p}\tilde{f}(D^3_{he_p}\tilde{f})^T}{3!}+\frac{D^3_{he_p}\tilde{f}(D_{he_p}\tilde{f})^T}{3!}\right) \\
&\quad + \cdots \\
&= \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\frac{h^{2(i+j)}}{(2i+1)!(2j+1)!}\sum_{p=1}^{n}\left(\nabla_p^{2i+1}\tilde{f}\right)\left(\nabla_p^{2j+1}\tilde{f}\right)^T\,. \qquad (65)
\end{aligned}$$

Similarly, we get for the estimate based on the second-order approximation (41):

$$\begin{aligned}
P_y &= \frac{1}{h^2}\sum_{p=1}^{n}\left(\sum_{i=0}^{\infty}\frac{D^{2i+1}_{he_p}\tilde{f}}{(2i+1)!}\right)\left(\sum_{i=0}^{\infty}\frac{D^{2i+1}_{he_p}\tilde{f}}{(2i+1)!}\right)^T \\
&\quad + \frac{1}{h^2}\sum_{p=1}^{n}\left(\sum_{i=1}^{\infty}\frac{D^{2i}_{he_p}\tilde{f}}{(2i)!}\right)\left(\sum_{i=1}^{\infty}\frac{D^{2i}_{he_p}\tilde{f}}{(2i)!}\right)^T \\
&\quad - \frac{1}{h^4}\sum_{p=1}^{n}\left(\sum_{i=1}^{\infty}\frac{D^{2i}_{he_p}\tilde{f}}{(2i)!}\right)\left(\sum_{i=1}^{\infty}\frac{D^{2i}_{he_p}\tilde{f}}{(2i)!}\right)^T \\
&= \frac{1}{h^2}\sum_{p=1}^{n}D_{he_p}\tilde{f}\left(D_{he_p}\tilde{f}\right)^T
\end{aligned}$$

$$
\begin{aligned}
&+ \quad \frac{1}{h^2}\sum_{p=1}^{n}\left(\frac{D_{he_p}\tilde{f}(D^3_{he_p}\tilde{f})^T}{3!} + \frac{D^2_{he_p}\tilde{f}(D^2_{he_p}\tilde{f})^T}{(2!)(2!)} + \frac{D^3_{he_p}\tilde{f}(D_{he_p}\tilde{f})^T}{3!}\right) \\
&- \quad \frac{1}{h^4}\sum_{p=1}^{n}\left(\frac{D^2_{he_p}\tilde{f}(D^2_{he_p}\tilde{f})^T}{(2!)(2!)}\right) \quad + \quad \cdots \\
= & \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\frac{h^{2(i+j)}}{(2i+1)!(2j+1)!}\sum_{p=1}^{n}\left(\nabla_p^{2i+1}\tilde{f}\right)\left(\nabla_p^{2j+1}\tilde{f}\right)^T \\
&+ \quad \sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\frac{h^{2(i+j)-2} - h^{2i-2}h^{2j-2}}{(2i)!(2j)!}\sum_{p=1}^{n}\nabla_p^{2i}\tilde{f}\left(\nabla_p^{2j}\tilde{f}\right)^T . \qquad (66)
\end{aligned}
$$

As before we will compare the new estimates with estimates obtained using Taylor approximations in place of $\tilde{f}$. For convenience we shall first look at the second-order approximation. The approximation error for element (k, l) in the covariance estimate obtained by employing a second-order Taylor approximation in place of $\tilde{f}$ is

$$
\begin{aligned}
Q_2(k,l) \quad = \quad & \sum_{\substack{i=0\\ j\neq 0}}^{\infty}\sum_{\substack{j=0\\ i\neq 0}}^{\infty}\frac{\sigma_{2(i+j)+2}}{(2i+1)!(2j+1)!}\sum_{p=1}^{n}\left(\nabla_p^{2i+1}\tilde{f}_k\right)\left(\nabla_p^{2j+1}\tilde{f}_l\right)^T \\
+ \quad & \sum_{\substack{i=1\\ j\neq 1}}^{\infty}\sum_{\substack{j=1\\ i\neq 1}}^{\infty}\frac{\sigma_{2(i+j)} - \sigma_{2i}\sigma_{2j}}{(2i)!(2j)!}\sum_{p=1}^{n}\left(\nabla_p^{2i}\tilde{f}_k\right)\left(\nabla_p^{2j}\tilde{f}_l\right)^T \\
+ \quad & [\text{cross-terms}] . \qquad (67)
\end{aligned}
$$

The subscripts on the first double sum mean that the case $i = j = 0$ is not included. Likewise, for the second double sum the case $i = j = 1$ is not included. To allow a comparison, the terms containing products of second-order cross-derivatives have been discarded as (37) was discarded for computational convenience (i.e., the terms are included in the "cross-terms"). It should be noticed that in the covariance estimate employed by the conventional second-order Gaussian filter these terms are usually calculated.

In a similar fashion as above, by subtracting (52) and (66), it is possible to write up the approximation error for the covariance estimate based on the new second-order approximation of $\tilde{f}$:

$$
\begin{aligned}
\bar{Q}_2(k,l) \quad = \quad & \sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\frac{\sigma_{2(i+j)+2} - h^{2(i+j)}}{(2i+1)!(2j+1)!}\sum_{p=1}^{n}\left(\nabla_p^{2i+1}\tilde{f}_k\right)\left(\nabla_p^{2j+1}\tilde{f}_l\right)^T \\
+ \quad & \sum_{\substack{i=1\\ j\neq 1}}^{\infty}\sum_{\substack{j=1\\ i\neq 1}}^{\infty}\frac{\sigma_{2(i+j)} - \sigma_{2i}\sigma_{2j} - h^{2(i+j)-4}(h^2-1)}{(2i)!(2j)!}\sum_{p=1}^{n}\left(\nabla_p^{2i}\tilde{f}_k\right)\left(\nabla_p^{2j}\tilde{f}_l\right)^T \\
+ \quad & [\text{cross-terms}] . \qquad (68)
\end{aligned}
$$

As

$$
\begin{array}{rcccl}
\sigma_{2(i+j)+2} & > & \sigma_{2(i+j)+2} - h^{2(i+j)} & > & 0 \\
\sigma_{2(i+j)} - \sigma_{2i}\sigma_{2j} & > & \sigma_{2(i+j)} - \sigma_{2i}\sigma_{2j} - h^{2(i+j)-4}(h^2-1) & > & 0 ,
\end{array} \qquad (69)
$$

we can use the same argumentation as was applied to evaluate the mean estimates and conclude that $|\bar{Q}_2(k,l)|$ has a lower upper bound than $|Q_2(k,l)|$. The new covariance estimate is therefore better than if we had inserted a second-order Taylor-approximation (without the cross-derivatives) of $\tilde{f}$. The missing fourth-order terms in (66) are the terms taking the form $\left(\frac{\partial \tilde{f}}{\partial z_p}\right)\left(\frac{\partial^3 \tilde{f}}{\partial z_p \partial z_q^2}\right)^T \sigma_2^2$ and $\left(\frac{\partial^2 \tilde{f}}{\partial z_p^2 \partial z_q^2}\right)\left(\frac{\partial^2 \tilde{f}}{\partial z_p^2 \partial z_q^2}\right)^T \sigma_2^2$. The last mentioned terms could have been present in the estimate had the cross-differences (37) not been discarded from the approximation of $\tilde{f}$.

Notice that for the one-dimensional case there are no cross-terms and all the sums are made over positive numbers. Thus, one can in this case skip the $|\cdot|$. Additionally, errors will obviously not appear until in the sixth-order terms for the estimate (66).

The approximation error for the covariance estimate based on the divided difference linearization of $\tilde{f}$ is

$$\begin{aligned} \bar{Q}_1(k,l) &= \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} \frac{\sigma_{2(i+j)+2} - h^{2(i+j)}}{(2i+1)!(2j+1)!} \sum_{p=1}^{n} \left(\nabla_p^{2i+1} \tilde{f}_k\right)\left(\nabla_p^{2j+1} \tilde{f}_l\right)^T \\ &+ \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} \frac{\sigma_{2(i+j)} - \sigma_{2i}\sigma_{2j}}{(2i)!(2j)!} \sum_{p=1}^{n} \left(\nabla_p^{2i} \tilde{f}_k\right)\left(\nabla_p^{2j} \tilde{f}_l\right)^T \\ &+ [\text{cross-terms}] \end{aligned} \tag{70}$$

The approximation error for the covariance estimate based on a Taylor linearization of $\tilde{f}$, $Q_1(k,l)$, is identical except that the quantity $h^{2(i+j)}$ is not subtracted. Obviously, $|\bar{Q}_1(k,l)|$ will therefore have a lower upper bound than $|Q_1(k,l)|$. The estimate will also have a lower upper bound than the estimate suggested in [Sch97] as in this paper $h = 1$.

For the estimate of the cross-covariance matrix, P_{xy}, given by (28) we have

$$P_{xy} = \frac{1}{h}\sum_{p=1}^{n} s_{x,p}\left(\sum_{i=0}^{\infty} \frac{D_{he_p}^{2i+1}\tilde{f}}{(2i+1)!}\right)^T = \frac{1}{h}\sum_{p=1}^{n} s_{x,p}\left(\sum_{i=0}^{\infty} \frac{h^{2i+1}}{(2i+1)!}\nabla_p^{2i+1}\tilde{f}\right)^T . \tag{71}$$

The conclusions above are valid for this estimate as well. The errors are again introduced on fourth-order terms in the series as the cross-derivative terms, $s_{x,p}\left(\frac{\partial^3 \tilde{f}}{\partial z_p \partial z_q^2}\right)^T \sigma_2^2$, $p \neq q$, do not appear in the series expansion of the estimate. However, unlike for the estimate based on a Taylor approximation, some of the fourth-order terms are matched with the new estimate.

4 State Estimation for Nonlinear Systems

We have now arrived at the central issue of this note, namely state estimation for nonlinear systems. Two new filters will be suggested that are based on the previously derived polynomial approximations. The filters are fundamentally different from filters based on Taylor approximations in that the polynomial approximations underlying the new filters take into account the uncertainty on the state estimate. The Taylor approximation underlying conventional filter designs for nonlinear systems, such as the EKF, depends only on the current state estimate and not on its variance. Nevertheless, the new filters can generally be implemented more easily as no derivatives are required.

The first filter we shall derive is based on a first-order polynomial approximation. This estimator is a generalized version of the filter presented in [Sch97]. Subsequently, a more accurate filter will be derived that also includes second-order terms. It turns out that this filter has certain similarities with the *unscented* filter described in [JU94], [JUDW95].

4.1 Review of State Estimation for Nonlinear Systems

Consider the following general nonlinear model of a dynamic system whose states are to be estimated

$$\begin{aligned} x_{k+1} &= f(x_k, u_k, v_k) && (72) \\ y_k &= g(x_k, w_k) \ . && (73) \end{aligned}$$

v_k and w_k are assumed i.i.d. and independent of current and past states, $v_k \sim (\bar{v}_k, Q(k)), w_k \sim (\bar{w}_k, R(k))$.

The commonly used state estimation principle for nonlinear systems is briefly outlined in the following. In-depth treatments of the topic can be found in [Lew86], [GKN+74], [May82]. Ideally, we would like to determine the *a priori* state and covariance estimates like in the Kalman filter. That is, as the conditional expectations

$$\begin{aligned} \bar{x}_k &= E[x_k | Y^{k-1}] && (74) \\ \bar{P}(k) &= E\left[(x_k - \bar{x}_k)(x_k - \bar{x}_k)^T | Y^{k-1}\right] \ , && (75) \end{aligned}$$

where Y^{k-1} is a matrix containing the past measurements

$$Y^{k-1} = \left[\begin{array}{cccc} y_0 & y_1 & \dots & y_{k-1} \end{array} \right]^T . \qquad (76)$$

For convenience, the measurement (*a posteriori*) update of the state estimate is usually restricted to be linear in the measurements. Selecting the update so that the (conditional) covariance of the estimation error is minimized, we obtain the following [Lew86]:

$$\begin{aligned} K_k &= P_{xy}(k) P_y^{-1}(k) && (77) \\ \hat{x}_k &= \bar{x}_k + K_k [y_k - \bar{y}_k] \ , && (78) \end{aligned}$$

where

$$\begin{aligned} \bar{y}_k &= E[y_k | Y^{k-1}] && (79) \\ P_{xy}(k) &= E\left[(x_k - \bar{x}_k)(y_k - \bar{y}_k)^T | Y^{k-1}\right] && (80) \\ P_y(k) &= E\left[(y_k - \bar{y}_k)(y_k - \bar{y}_k)^T | Y^{k-1}\right] \ . && (81) \end{aligned}$$

The corresponding update of the covariance matrix is

$$\hat{P}(k) = E\left[(x_k - \hat{x}_k)(x_k - \hat{x}_k)^T | Y^k\right] = \bar{P}(k) - K_k P_{yy}(k) K_k^T \; . \tag{82}$$

As the various expectations generally are intractable, some kind of approximation is commonly used; e.g., it is well-known that the extended Kalman filter is based on Taylor linearization of state transition and output equations (72), (73). The EKF equations are listed below to allow the reader to compare its complexity with that of the filters derived in the following. A treatment of the second-order filters may be found in [May82].

The state transition and observation equations are approximated by first-order polynomials

$$x_{k+1} \approx f(\hat{x}_k, u_k, \bar{v}_k) + F_x(k)(x_k - \hat{x}_k) + F_v(k)(v_k - \bar{v}_k) \tag{83}$$
$$y_k \approx g(\bar{x}_k, \bar{w}_k) + G_x(k)(x_k - \bar{x}_k) + G_w(k)(w_k - \bar{w}_k) \; , \tag{84}$$

where

$$\begin{aligned} F_x(k) &= \left.\frac{\partial f(x, u_k, \bar{v}_k)}{\partial x}\right|_{x=\hat{x}_k} \qquad F_v(k) = \left.\frac{\partial f(\hat{x}_k, u_k, v)}{\partial v}\right|_{v=\bar{v}_k} \\ G_x(k) &= \left.\frac{\partial g(x, \bar{w}_k)}{\partial x}\right|_{x=\bar{x}_k} \qquad G_w(k) = \left.\frac{\partial f(\bar{x}_k, w)}{\partial w}\right|_{w=\bar{w}_k} \; . \end{aligned} \tag{85}$$

When these approximations are inserted we arrive at [Lew86]:

A priori update:

$$\bar{x}_{k+1} = f(\hat{x}_k, u_k, v_k) \tag{86}$$
$$\bar{y}_k = g(\bar{x}_k, w_k) \tag{87}$$
$$\bar{P}(k+1) = F_x(k)\hat{P}(k)F_x(k)^T + F_v(k)Q(k)F_v(k)^T \tag{88}$$

A posteriori updates:

$$K_k = \bar{P}(k)G_x(k)^T \left[G_x(k)\bar{P}(k)G_x(k)^T + G_w(k)R(k)G_w(k)^T\right]^{-1} \tag{89}$$
$$\hat{x}_k = \bar{x}_k + K_k[y_k - \bar{y}_k] \tag{90}$$
$$\hat{P}(k) = \left[I - K_k G_w(k)\right]\bar{P}(k) \tag{91}$$

In the following subsections we will pursue the use of approximations obtained with the interpolation formula for derivation of state estimators for nonlinear systems.

4.2 The DD1 Filter

In this section a generalized version of the nonlinear state estimation scheme suggested in [Sch97] will be described. The filter is derived by employing the first-order approximation presented in Section 3.1. In principle this corresponds to the EKF except that the Jacobians (85) are replaced by divided differences. The state update is therefore the same as in the extended Kalman filter. The difference is alone found in the update of the various covariance matrices. Generally, they can be implemented more easily. We will use an approach much like the one suggested in [Sch97]. One of the particularly useful ideas provided in this paper is to update the Cholesky factors of the covariance matrices directly.

First we will introduce the following four square Cholesky factorizations

$$\begin{array}{ll} Q = S_v S_v^T & R = S_w S_w^T \\ \bar{P} = \bar{S}_x \bar{S}_x^T & \hat{P} = \hat{S}_x \hat{S}_x^T \; . \end{array} \tag{92}$$

Let the jth column of $\bar{S}_x$ be denoted $\bar{s}_{x,j}$ and vice versa for the other factors. Four matrices containing divided differences are now defined by

$$\begin{aligned} S_{x\hat{x}}^{(1)}(k) &= \left\{S_{x\hat{x}}^{(1)}(i,j)\right\} = \{(f_i(\hat{x}_k + h\hat{s}_{x,j}, u_k, \bar{v}_k) - f_i(\hat{x}_k - h\hat{s}_{x,j}, u_k, \bar{v}_k))/2h\} \\ S_{xv}^{(1)}(k) &= \left\{S_{xv}^{(1)}(i,j)\right\} = \{(f_i(\hat{x}_k, u_k, \bar{v}_k + hs_{v,j}) - f_i(\hat{x}_k, u_k, \bar{v}_k - hs_{v,j}))/2h\} \\ S_{y\bar{x}}^{(1)}(k) &= \left\{S_{y\bar{x}}^{(1)}(i,j)\right\} = \{(g_i(\bar{x}_k + h\bar{s}_{x,j}, \bar{w}_k) - g_i(\bar{x}_k - h\bar{s}_{x,j}, \bar{w}_k))/2h\} \\ S_{yw}^{(1)}(k) &= \left\{S_{yw}^{(1)}(i,j)\right\} = \{(g_i(\bar{x}_k, \bar{w}_k + hs_{w,j}) - g_i(\bar{x}_k, \bar{w}_k - hs_{w,j}))/2h\} \; . \end{aligned} \tag{93}$$

The a priori update

To understand how the results from Section 3.1 can be applied in a state estimation context it is useful to think off an *augmented state vector* consisting of state vector and process (or measurement) noise:

$$\breve{x} = \left[\; \breve{\bar{x}} + \Delta\breve{x} \;\right] = \begin{bmatrix} \hat{x} + \Delta x \\ \bar{v} + \Delta v \end{bmatrix} . \tag{94}$$

As the process noise is assumed to be independent of the state, the (conditional) covariance of $\Delta\breve{x}$ is

$$\hat{P}_{\breve{x}} = \begin{bmatrix} \hat{P} & 0 \\ 0 & Q \end{bmatrix} = \begin{bmatrix} \hat{S}_x & 0 \\ 0 & S_v \end{bmatrix} \begin{bmatrix} \hat{S}_x & 0 \\ 0 & S_v \end{bmatrix}^T = \hat{S}_{\breve{x}} \hat{S}_{\breve{x}}^T \; . \tag{95}$$

Introducing the vector z by stochastical decoupling of $\breve{x}$, $\breve{x} = S_{\breve{x}} z$, it is not difficult to see how the state estimation problem can be mapped into the treatment of the general vector function $\tilde{f}(z)$, which was presented in Section 3.1.

For the *a priori* update of the state estimate we will use (25):

$$\boxed{\bar{x}_{k+1} \approx \tilde{f}(\bar{z}_k) = f(\hat{x}_k, u_k, \bar{v}_k)} \tag{96}$$

which is the same as for the EKF.

As the basis of the covariance update we shall use (27). By application of the matrices defined in (93) the update can obviously be expressed in the following matrix notation

$$\begin{aligned}\bar{P}(k+1) &= \left[\begin{array}{cc} S_{x\hat{x}}^{(1)}(k) & S_{xv}^{(1)}(k)\end{array}\right]\left[\begin{array}{cc} S_{x\hat{x}}^{(1)}(k) & S_{xv}^{(1)}(k)\end{array}\right]^T \\ &= S_{x\hat{x}}^{(1)}(k)\left(S_{x\hat{x}}^{(1)}(k)\right)^T + S_{xv}^{(1)}(k)\left(S_{xv}^{(1)}(k)\right)^T . \end{aligned} \tag{97}$$

Due to the assumed independence between v_k and x_k, the update can be written as a sum of two matrix products.

It is well-known that a straightforward "text-book" implementation of the (extended) Kalman filter results in numerical problems after a number of iterations as the effect of round-off errors accumulates, thus making the covariance matrix asymmetric and non-positive definite. The usual remedy for this is to use a factored update. As the covariance update (97) is a sum of two quadratic terms, numerical problems of this kind should not occur with this update. Nevertheless, it is tempting to use a factored update anyway since the factor will be needed for the *a posteriori* update. Moreover, the (rectangular, nontriangular) Cholesky factor is immediately available as the following compound matrix:

$$\boxed{\bar{S}_x(k+1) = \left[\begin{array}{cc} S_{x\hat{x}}^{(1)}(k) & S_{xv}^{(1)}(k)\end{array}\right]} \tag{98}$$

This is a rectangular matrix and for later use it must be transformed to a square Cholesky factor. This can be achieved through Householder triangularization [GA93], [GL89].

The a posteriori update

The *a priori* estimate of output and covariance matrix for the *output* estimation error is derived in a similar fashion. The output estimate is given by

$$\bar{y}_k = g(\bar{x}_k, \bar{w}_k) , \tag{99}$$

and the compound matrix

$$S_y(k) = \left[\begin{array}{cc} S_{y\bar{x}}^{(1)}(k) & S_{yw}^{(1)}(k)\end{array}\right] \tag{100}$$

is a Cholesky factor of the covariance of the output estimation error,

$$P_y(k) = S_y(k)S_y(k)^T . \tag{101}$$

As for $\bar{S}_x$, $S_y(k)$ should be transformed to a quadratic matrix by Householder triangularization.

For approximation of the cross-covariance between state and output estimation error we will use the result in (29)

$$P_{xy}(k) = \bar{S}_x(k)\left(S_{y\bar{x}}^{(1)}(k)\right)^T . \tag{102}$$

The Kalman gain can now be calculated according to (77)

$$\boxed{K_k = P_{xy}(k)\left[S_y(k)S_y(k)^T\right]^{-1}} \tag{103}$$

and the state vector is updated according to to (78)

$$\boxed{\hat{x}_k = \bar{x}_k + K_k\left(y_k - \bar{y}_k\right)} \tag{104}$$

The factorization of P_y has deliberately been maintained in (103) because it is useful in the practical computation of the gain. Since S_y is triangular the equation $\left[S_y(k)S_y(k)^T\right]K_k = P_{xy}(k)$ is easily solved using only forward and back substitutions.

The *a posteriori* covariance can be updated according to (82). However, as suggested in [Sch97] one can also in this case update its Cholesky factor directly. As the following expressions are identical

$$\begin{aligned}
KP_yK^T &= \bar{S}_x\left(S_{yx}^{(1)}\right)^T K^T \\
&= KS_{yx}^{(1)}S_x^T \\
&= KS_{yx}^{(1)}\left(S_{yx}^{(1)}\right)^T K^T + KS_{yw}^{(1)}\left(S_{yw}^{(1)}\right)^T K^T ,
\end{aligned}$$

the *a posteriori* update can clearly be rewritten as

$$\begin{aligned}
\hat{P} &= \bar{P} - KP_yK^T \\
&= \bar{P} - KP_yK^T - KP_yK^T + KP_yK^T \\
&= \bar{S}_x\bar{S}_x^T - \bar{S}_x\left(S_{yx}^{(1)}\right)^T K^T - KS_{yx}^{(1)}S_x^T + KS_{yx}^{(1)}\left(S_{yx}^{(1)}\right)^T K^T + KS_{yw}^{(1)}\left(S_{yw}^{(1)}\right)^T K^T , \\
&= \left(\bar{S}_x - KS_{yx}^{(1)}\right)\left(\bar{S}_x - KS_{yx}^{(1)}\right)^T + KS_{yw}^{(1)}\left(KS_{yw}^{(1)}\right)^T
\end{aligned} \tag{105}$$

implying that a square Cholesky factor of the covariance matrix can be obtained by triangularization of the compound matrix

$$\boxed{\hat{S}(k) = \left[\begin{array}{cc} \bar{S}_x(k) - K_kS_{yx}^{(1)}(k) & K_kS_{yw}^{(1)}(k) \end{array}\right]} \tag{106}$$

4.3 The DD2 Filter

The DD2 filter is obtained by using the estimates of mean and covariance derived in Section 3.2. First we shall define four additional matrices containing divided differences

$$
\begin{aligned}
S^{(2)}_{x\hat{x}}(k) &= \left\{ \frac{\sqrt{h^2-1}}{2h^2}(f_i(\hat{x}_k + h\hat{s}_{x,j}, u_k, \bar{v}_k) + f_i(\hat{x}_k - h\hat{s}_{x,j}, u_k, \bar{v}_k) - 2f_i(\hat{x}_k, u_k, \bar{v}_k)) \right\} \\
S^{(2)}_{xv}(k) &= \left\{ \frac{\sqrt{h^2-1}}{2h^2}(f_i(\hat{x}_k, u_k, \bar{v}_k + hs_{v,j}) + f_i(\hat{x}_k, u_k, \bar{v}_k - hs_{v,j}) - 2f_i(\hat{x}_k, u_k, \bar{v}_k)) \right\} \\
S^{(2)}_{y\bar{x}}(k) &= \left\{ \frac{\sqrt{h^2-1}}{2h^2}(g_i(\bar{x}_k + h\bar{s}_{x,j}, \bar{w}_k) + g_i(\bar{x}_k - h\bar{s}_{x,j}, \bar{w}_k) - 2g_i(\bar{x}_k, \bar{w}_k)) \right\} \\
S^{(2)}_{yw}(k) &= \left\{ \frac{\sqrt{h^2-1}}{2h^2}(g_i(\bar{x}_k, \bar{w}_k + hs_{w,j}) + g_i(\bar{x}_k, \bar{w}_k - hs_{w,j}) - 2g_i(\bar{x}_k, \bar{w}_k)) \right\} .
\end{aligned}
$$

The a priori update

Proceeding as for the DD1 filter, we can obtain an improved state estimate by using (32):

$$
\boxed{
\begin{aligned}
\bar{x}_{k+1} = \; & \tfrac{h^2-n_x-n_v}{h^2} f(\hat{x}_k, u_k, \bar{v}_k) \\
& + \tfrac{1}{2h^2} \textstyle\sum_{p=1}^{n_x} f(\hat{x}_k + h\hat{s}_{x,p}, u_k, \bar{v}_k) + f(\hat{x}_k - h\hat{s}_{x,p}, u_k, \bar{v}_k) \\
& + \tfrac{1}{2h^2} \textstyle\sum_{p=1}^{n_v} f(\hat{x}_k, u_k, \bar{v}_k + hs_{v,p}) + f(\hat{x}_k, u_k, \bar{v}_k - hs_{v,p})
\end{aligned}
}
\tag{107}
$$

n_x denotes the dimension of the state vector and n_v denotes the dimension of process noise vector. It turns out that this estimate of the mean is identical to the one proposed in [JU94], [JUDW95]. This is interesting as the approach used in these papers is quite different from the one used here.

In agreement with the covariance estimate in (27), a triangular Cholesky factor of the *a priori* covariance is obtained by Householder transformation of the following compound matrix

$$
\boxed{\bar{S}_x(k+1) = \left[\begin{array}{cccc} S^{(1)}_{x\hat{x}}(k) & S^{(1)}_{xv}(k) & S^{(2)}_{x\hat{x}}(k) & S^{(2)}_{xv}(k) \end{array} \right]}
\tag{108}
$$

The covariance estimate $\bar{S}_x\bar{S}_x^T$ is *not* the same as the one derived in [JU94], [JUDW95], which was the case for the mean estimate. In Appendix A it is shown how the covariance estimate of [JU94] (which is less accurate than the one presented here) can be derived along the same lines as above.

The a posteriori update

The *a priori* estimate of the output and its covariance is calculated in a similar fashion as for the states

$$
\begin{aligned}
\bar{y}_k &= \tfrac{h^2-n_x-n_w}{h^2} g(\bar{x}_k, \bar{w}_k) \\
&+ \tfrac{1}{2h^2} \textstyle\sum_{p=1}^{n_x} g(\bar{x}_k + h\bar{s}_{x,p}, \bar{w}_k) + g(\bar{x}_k - h\bar{s}_{x,p}, \bar{w}_k) \\
&+ \tfrac{1}{2h^2} \textstyle\sum_{p=1}^{n_w} g(\bar{x}_k, \bar{w}_k + hs_{w,p}) + g(\bar{x}_k, \bar{w}_k - hs_{w,p})
\end{aligned}
\tag{109}
$$

and

$$
S_y(k) = \left[\begin{array}{cccc} S_{y\bar{x}}^{(1)}(k) & S_{yw}^{(1)}(k) & S_{y\bar{x}}^{(2)}(k) & S_{yw}^{(2)}(k) \end{array} \right] . \tag{110}
$$

n_w denotes the dimension of the measurement noise vector.

It follows from the discussion in Section 3.2 and (43) that the *a priori* cross-covariance matrix is the same as for the DD1 filter (102):

$$
P_{xy}(k) = \bar{S}_x(k) S_{y\bar{x}}(k)^T . \tag{111}
$$

Kalman gain and *a posteriori* update of the state is carried out exactly as for the DD1 filter:

Kalman gain:

$$
\boxed{K_k = P_{xy}(k) \left[S_y(k) S_y(k)^T \right]^{-1}} \tag{112}
$$

A posteriori update of state vector

$$
\boxed{\hat{x}_k = \bar{x}_k + K_k \left(y_k - \bar{y}_k \right)} \tag{113}
$$

The *a posteriori* update of the estimation error covariance has a few additional terms. Following the derivations in (105) we can write the covariance matrix

$$
\begin{aligned}
\hat{P} &= \left(\bar{S}_x - KS_{yx}^{(1)} \right) \left(\bar{S}_x - KS_{yx}^{(1)} \right)^T + KS_{yw}^{(1)} \left(KS_{yw}^{(1)} \right)^T \\
&+ KS_{yx}^{(2)} \left(KS_{yx}^{(2)} \right)^T + KS_{yw}^{(2)} \left(KS_{yw}^{(2)} \right)^T
\end{aligned} , \tag{114}
$$

which obviously has the Cholesky factor

$$
\boxed{\hat{S}_x(k) = \left[\begin{array}{cccc} \bar{S}_x(k) - K_k S_{yx}^{(1)}(k) & K_k S_{yw}^{(1)}(k) & K_k S_{yx}^{(2)}(k) & K_k S_{yw}^{(2)}(k) \end{array} \right]} \tag{115}
$$

4.4 The Complete Filter Algorithm

The following procedure outlines the implementation of the new filters. Recall that $h^2 = 3$ since $\sigma_4 = 3\sigma_2$ for a Gaussian distributed variable.

1. Initialize $\bar{x}_0$, $\bar{P}(0)$, $k = 0$.

a posteriori update

2. Compute $\bar{y}_k$, $S_{y\bar{x}}^{(1)}(k)$, $S_{yw}^{(1)}(k)$, $S_{y\bar{x}}^{(2)}(k)$, $S_{yw}^{(2)}(k)$

3. Compute P_{xy} according to (102) and determine $S_y(k)$ using Householder triangularization on (100) or (110).

4. Solve $K_k\left[\bar{S}_y(k)S_y(k)^T\right] = P_{xy}$ for the Kalman gain. Since S_y is square and triangular only forward and back-substitutions are needed: First solve for k': $k'S_y^T = P_{xy}$ and then solve for K_k: $K_kS_y = k'$.

5. *A posteriori* update of the state estimate $\hat{x}_k = \bar{x}_k + K_k\left(y_k - \bar{y}_k\right)$

6. *A posteriori* update of covariance matrix factor, $\hat{S}_x(k)$, is performed using Householder triangularization on (106) or (115).

a priori update

7. Determine $\bar{x}_{k+1}$, $S_{x\hat{x}}^{(1)}(k+1)$, $S_{xw}^{(1)}(k+1)$, $S_{x\hat{x}}^{(2)}(k+1)$, $S_{xw}^{(2)}(k+1)$.

8. Use Householder triangularization on (98) or (108) to compute $\bar{S}_x(k)$

9. $k = k+1$, go to step 2

Several textbooks provide details on how to perform the Householder triangularization, e.g., [PFTV88], [GL89], [GA93].

5 Example

To demonstrate the performance of the new filters they will in this section be evaluated on the often used vertically falling body example originating from [AWB68]. Several filter designs have been evaluated on this example [AWB68], [May82], [JU94]. The setup is briefly outlined below. The reader is referred to [AWB68] for a more detailed introduction to the problem.

We wish to estimate altitude (x_1), downward velocity (x_2), and a (constant) ballistic parameter (x_3) of a vertically falling body. The setup is depicted in Fig. 2.

The radar measures the range (r). The measurements, which appear with intervals of 1 second, are affected by additive, white Gaussian noise.

The model has the following form:

$$\begin{aligned}
\dot{x}_1(t) &= -x_2(t) && (116)\\
\dot{x}_2(t) &= -e^{-\gamma x_1(t)}x_2(t)^2x_3(t) && (117)\\
\dot{x}_3(t) &= 0 && (118)\\
y_k &= r_k + w_k = \sqrt{M^2 + (x_{1,k} - H)^2} + w_k\ . && (119)
\end{aligned}$$

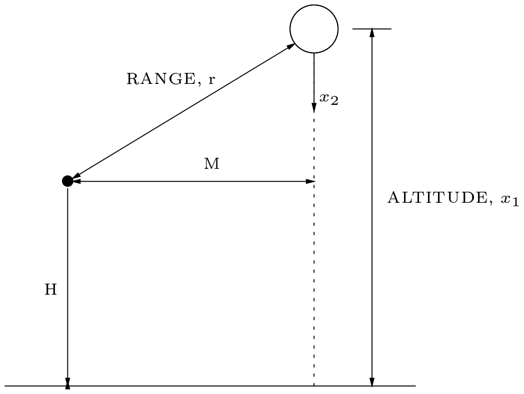

Figure 2. *Geometry of the vertically falling body problem.*

The model parameters are given by:

$$\begin{aligned} M &= 100,000 \text{ ft} \\ H &= 100,000 \text{ ft} \\ \gamma &= 5 \times 10^{-5} \\ E[w_k^2] &= 10^4 \text{ ft}^2 \end{aligned} \tag{120}$$

and the initial state of the system is

$$\begin{cases} x_{1,0} = 300,000 \text{ ft} \\ x_{2,0} = 20,000 \text{ ft/s} \\ x_{3,0} = 10^{-3} \end{cases} \tag{121}$$

We will compare the performances of the DD1 and DD2 filters with those of the EKF and the modified Gaussian second-order filter [AWB68]. The reader is referred to [JU94] for an evaluation of the unscented filter. Due to the nature of the problem it is common practice to employ a continuous-discrete filter implementation. The state equations (116)-(118) are integrated using a fourth-order Runge-Kutta method with 64 steps taken between each observation. It is straightforward to implement continuous-discrete versions of the DD1 and DD2 filters as there is no process noise. In [AWB68] it is described how to implement the EKF and the modified Gaussian second-order filter for the considered application.

In accordance with [AWB68] and [JU94] the following initialization of the state estimates is used

$$\begin{cases} \hat{x}_{1,0} = 300,000 \text{ ft} \\ \hat{x}_{2,0} = 20,000 \text{ ft/s} \\ \hat{x}_{3,0} = 3 \times 10^{-5} \end{cases} \tag{122}$$

and the covariance matrix is initialized to

$$\hat{P}(0) = \begin{bmatrix} 10^6 & 0 & 0 \\ 0 & 4 \times 10^6 & 0 \\ 0 & 0 & 10^{-4} \end{bmatrix}. \tag{123}$$

To enable a fair comparison of the estimates produced by each of the four filters, the estimates are averaged across a Monte Carlo simulation consisting of 50 runs. Each run is carried out with a different noise sample.

The results of the Monte Carlo simulation are shown in Figure 3–Figure 5.

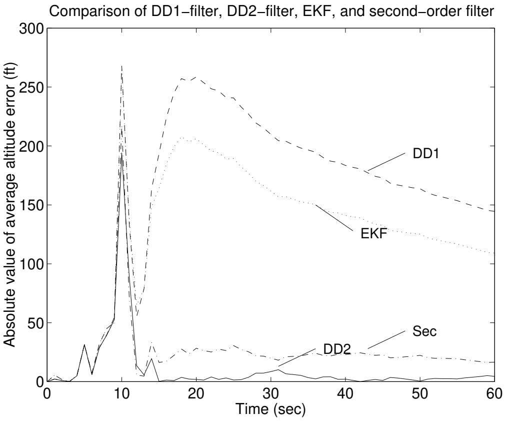

Figure 3. *Absolute error in position estimate (50 run average).*

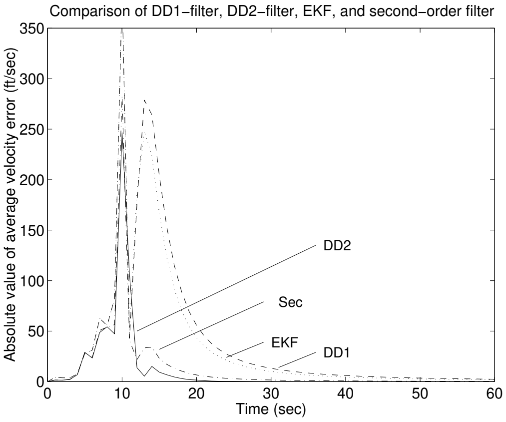

Figure 4. *Absolute error in velocity estimate (50 run average).*

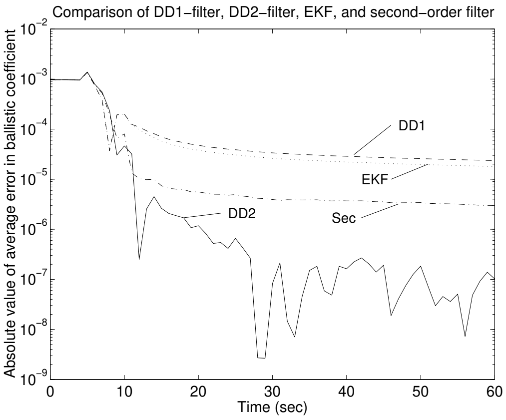

Figure 5. *Absolute error in estimate of ballistic coefficient (50 run average).*

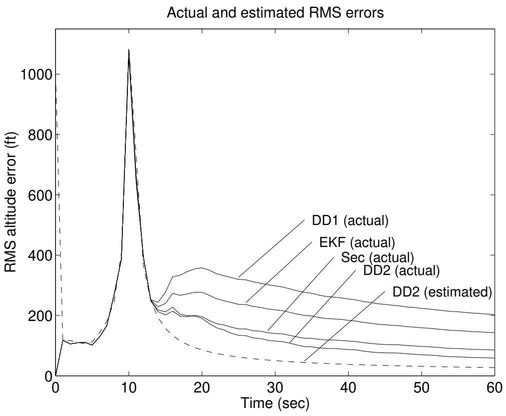

Figure 6. *"Actual" (50 run average) RMS altitude errors compared with the estimated RMS error, $\sqrt{\hat{P}_{11}(k)}$ for the DD2 filter.*

Not surprisingly, Figure 3–Figure 5 show that the DD2 filter exhibits a performance which is completely superior to the EKF and the DD1 filter. It is even better than the performance of the second-order filter. However, in contrast to what we would expect, the performance of the DD1 filter is slightly worse than that of the EKF. The difference is, however, marginal and must be contributed to the fact that the assumptions on which the accuracy of the DD1 filter was analyzed are partly violated. In particular, the assumption that the state estimate is unbiased is far from being satisfied here.

Comparison with the study of the unscented filter carried out in [JU94] shows that the performances of the unscented filter and the DD2 filter are similar. This agrees well with our expectations as the *a priori* state estimate is the same and the difference between the covariance updates are limited to fourth and higher order terms in their respective series expansions.

The RMS value of the altitude error is shown in Figure 6 for each of the four filters. For comparison, the estimated values $\sqrt{\hat{P}_{11}}$ have also been plotted for the DD2 filter. Note that the variations in the performance of the DD2 filter are seemingly smaller than for the EKF and DD1 filters. For all four filters, the actual estimation error variances exceed the variance estimates produced by the filters. However, the estimated variance is closer to the actual variance of the DD2 estimates than for the other three filters.

It should be noted that the simulation study also showed that there is little difference between the estimates of $\sqrt{\hat{P}_{11}}$ produced by the four filters. This is why only the estimates produced by the DD2 filter have been plotted in Figure 6. The marginal difference might lead to the suggestion that the (*a priori*) state estimate of the DD2 filter is used in conjunction with the covariance estimate of the DD1 filter in order to save computations.

6 Conclusions

In this paper we have proposed two new filters for nonlinear state estimation. Whereas filters for nonlinear systems commonly are based on polynomial approximations obtained with Taylor's formula, the approximations underlying the new filters are obtained with a multivariable extension of Stirling's interpolation formula. The filters are extremely simple to implement as no derivatives are needed, yet they provide an excellent accuracy. The DD1 filter is the simplest of the two filters. Essentially, it is similar to the filter proposed in [Sch97]. However, as it appears from Section 3.3, the (*a priori*) estimate of the covariance represents a more "faithful" approximation of the true covariance. The most important contribution of this note is the superior DD2 filter. This filter has the same *a priori* estimate as the "unscented" filter described in [JU94], [JUDW95], but a better covariance estimate.

The characteristics of the filters are briefly summarized below:

- Based on Gaussian assumptions, the accuracy of the DD1 filter will be comparable to the EKF in terms of expected error. The accuracy of the DD2 filter is comparable to the modified Gaussian second-order filter. As the employed polynomial approximations utilize knowledge about the covariance of the state estimates, we expect that the new filters will be superior to conventional (Taylor approximation based) filters for highly nonlinear systems, and systems with high noise levels.

- For "one-dimensional" systems (referring to the dimension of z) the accuracy of the DD2 filter is comparable to a fourth-order filter.
- The implementation is very simple as the filters do not require derivative information. Yet, the computational burden is relatively limited and will often be comparable to that of the EKF. As the user needs only provide models of dynamics and observation process, the filters are attractive for implementation of "generic" computer programs for nonlinear filtering.
- The filters are very useful for model calibration. It is straightforward to include a varying number of parameters in the state vector for joint state and parameter estimation. The user needs only initialize the parameter estimates and their variances and then run the filter again.
- The filters were derived based on considerations on how to estimate mean and covariance of arbitrary nonlinear transformations of variables with known mean and covariance. These results are not limited to state estimation; the approximations can easily be adopted by several other areas of statistics.
- Although the performance of the new filters was demonstrated based on the assumption that the nonlinear transformations are analytic, this is not a requirement for application of the filters. In fact, it is not even necessary to assume differentiability. The range of applications is therefore wider than for the EKF, which requires that the Jacobians exist.

7 Acknowledgements

This work was supported by the Danish Technical Research Council under contract no. 9502714.

References

[AWB68] M. Athans, R.P. Wishner, and A.B. Bertolini. Suboptimal state estimation for continuous-time nonlinear systems from discrete noisy measurements. *IEEE Transactions on Automatic Control*, AC-13(5):504–514, Oct. 1968.

[DB74] G. Dahlquist and Åke Björck. *Numerical Methods.* Prentice-Hall, 1974.

[Frö70] C.-E. Fröberg. *Introduction to Numerical Analysis.* Addison-Wesley, 1970.

[GA93] M.S. Grewal and A.P. Andrews. *Kalman Filtering: Theory and Practice.* Prentice Hall, Englewood Cliffs, New Jersey, 1993.

[GKN+74] A. Gelb, J.F. Kasper, R.A Nash, C.F. Price, and A.A. Sutherland. *Applied Optimal Estimation.* MIT Press, 1974.

[GL89] G.H. Golub and C.F. Van Loan. *Matrix Computations.* North Oxford Academic, London, 1989.

[JU94] S. J. Julier and J. K. Uhlmann. A general method for approximating nonlinear transformations of probability distributions. Technical report, Robotics Research Group, Department of Engineering Science, University of Oxford, 1994. (Internet publication: http://www.robots.ox.ac.uk/∼siju/index.html).

[JU97] S. J. Julier and J. K. Uhlmann. A new extension of the Kalman filter to nonlinear systems. In *Proceedings of AeroSense: The 11th International Symposium on Aerospace/Defense Sensing, Simulation and Controls, Orlando, Florida*, 1997.

[JUDW95] S. J. Julier, J. K. Uhlmann, and H. F. Durrant-Whyte. A new approach for filtering nonlinear systems. In *Proceedings of the 1995 American Control Conference, Seattle, Washington*, pages 1628–1632, 1995.

[Lew86] F.L. Lewis. *Optimal Estimation.* John Wiley & Sons, 1986.

[May82] P.S. Maybeck. *Stochastic Models, Estimation, and Control*, volume 2. Academic Press, 1982.

[Pap84] A. Papoulis. *Probability, Random Variables, and Stochastic Processess.* McGraw-Hill, Singapore, 1984.

[PFTV88] W.H. Press, B.P. Flannery, S.A. Tevkolsky, and W.T. Vetterling. *Numerical Recipes in C - The Art of Scientific Computing.* Cambridge University Press, 1988.

[Sch97] T.S. Schei. A finite-difference method for linearization in nonlinear estimation algorithms. *Automatica*, 33(11):2051–2058, 1997.

[Ste27] J. F. Steffensen. *Interpolation.* Williams & Wilkins, 1927.

A An Alternative Approximation of the Covariance

It was mentioned in Section 4.3 that the *a priori* state estimate of the DD2 filter is the one used in the *unscented* filter described in [JU94], [JUDW95]. In this appendix it is shown that also the covariance estimate of the unscented filter can be derived by following an approach similar to ours. This estimate is less accurate than the one presented previously in this paper. Moreover, it might occationally lead to an estimate which is non-positive semidefinite.

Recall from (34) that

$$\begin{aligned} P_y &= E\left[\tilde{D}_{\Delta z}\tilde{f}\left(\tilde{D}_{\Delta z}\tilde{f}\right)^T\right] + \frac{1}{4}E\left[\tilde{D}^2_{\Delta z}\tilde{f}\left(\tilde{D}^2_{\Delta z}\tilde{f}\right)^T\right] \\ &- \left(\bar{y} - \tilde{f}(\bar{z})\right)\left(\bar{y} - \tilde{f}(\bar{z})\right)^T . \end{aligned} \tag{124}$$

Maintaining from this expression the terms (35), (38), (39) we obtain (in (40) we did not include (39) as it cancels with (36))

$$\begin{aligned} P_y &= \sigma_2 \sum_{p=1}^{n} \left(\mu_p\delta_p\tilde{f}(\bar{z})\right)\left(\mu_p\delta_p\tilde{f}(z)\right)^T + \frac{\sigma_4}{4}\sum_{p=1}^{n}\left(\delta_p^2\tilde{f}(\bar{z})\right)\left(\delta_p^2\tilde{f}(z)\right)^T \\ &- \left(\bar{y} - \tilde{f}(\bar{z})\right)\left(\bar{y} - \tilde{f}(\bar{z})\right)^T \\ &= \frac{\sigma_2}{4h^2}\sum_{p=1}^{n}\left(\tilde{f}(\bar{z}+he_p) - \tilde{f}(\bar{z}-he_p)\right)\left(\tilde{f}(\bar{z}+he_p) - \tilde{f}(\bar{z}-he_p)\right)^T \\ &+ \frac{\sigma_4}{4h^4}\sum_{p=1}^{n}\left(\tilde{f}(\bar{z}+he_p) + \tilde{f}(\bar{z}-he_p) - 2\tilde{f}(\bar{z})\right) \times \\ &\qquad\qquad \left(\tilde{f}(\bar{z}+he_p) + \tilde{f}(\bar{z}-he_p) - 2\tilde{f}(\bar{z})\right)^T \\ &- \left(\bar{y} - \tilde{f}(\bar{z})\right)\left(\bar{y} - \tilde{f}(\bar{z})\right)^T \\ &= \frac{\sigma_2}{4h^2}\sum_{p=1}^{n}\Big[\tilde{f}(\bar{z}+he_p)f(\bar{z}+he_p)^T + f(\bar{z}-he_p)f(\bar{z}-he_p)^T \\ &\qquad - f(\bar{z}+he_p)f(\bar{z}-he_p)^T - f(\bar{z}-he_p)f(\bar{z}+he_p)^T\Big] \\ &+ \frac{\sigma_4}{4h^4}\sum_{p=1}^{n}\Big[f(\bar{z}+he_p)f(\bar{z}+he_p)^T + f(\bar{z}-he_p)f(\bar{z}-he_p)^T \\ &\qquad + f(\bar{z}+he_p)f(\bar{z}-he_p)^T + f(\bar{z}-he_p)f(\bar{z}+he_p)^T \\ &\qquad - 2f(\bar{z}+he_p)f(\bar{z})^T - 2f(\bar{z}-he_p)f(\bar{z})^T \\ &\qquad - 2f(\bar{z})f(\bar{z}+he_p)^T - 2f(\bar{z})f(\bar{z}-he_p)^T + 4f(\bar{z})f(\bar{z})^T\Big] \\ &- \left(\bar{y} - \tilde{f}(\bar{z})\right)\left(\bar{y} - \tilde{f}(\bar{z})\right)^T . \end{aligned} \tag{125}$$

Inserting that $\sigma_2 = 1$ and $h^2 = \sigma_4$, (125) can be greatly reduced.

$$
\begin{aligned}
P_y \quad = \quad & \frac{1}{2h^2} \sum_{p=1}^{n} \Big[\tilde{f}(\bar{z} + he_p) f(\bar{z} + he_p)^T + f(\bar{z} - he_p) f(\bar{z} - he_p)^T \\
& \qquad - f(\bar{z} + he_p) f(\bar{z})^T - f(\bar{z} - he_p) f(\bar{z})^T \\
& \qquad - f(\bar{z}) f(\bar{z} + he_p)^T - f(\bar{z}) f(\bar{z} - he_p)^T + 2 f(\bar{z}) f(\bar{z})^T \big] \\
- \quad & \left(\bar{y} - \tilde{f}(\bar{z}) \right) \left(\bar{y} - \tilde{f}(\bar{z}) \right)^T \\
= \quad & \frac{1}{2h^2} \sum_{p=1}^{n} [\tilde{f}(\bar{z} + he_p) - \tilde{f}(\bar{z})][\tilde{f}(\bar{z} + he_p) - \tilde{f}(\bar{z})]^T \\
+ \quad & \frac{1}{2h^2} \sum_{p=1}^{n} [\tilde{f}(\bar{z} - he_p) - \tilde{f}(\bar{z})][\tilde{f}(\bar{z} - he_p) - \tilde{f}(\bar{z})]^T \\
- \quad & \left(\bar{y} - \tilde{f}(\bar{z}) \right) \left(\bar{y} - \tilde{f}(\bar{z}) \right)^T \quad .
\end{aligned}
\tag{126}
$$

By straightforward vector manipulations and by using (32) it is easily shown that (126) can be rewritten as

$$
\begin{aligned}
P_y \quad = \quad & \frac{1}{2h^2} \sum_{p=1}^{n} [\tilde{f}(\bar{z} + he_p) - \bar{y}][\tilde{f}(\bar{z} + he_p) - \bar{y}]^T \\
+ \quad & \frac{1}{2h^2} \sum_{p=1}^{n} [\tilde{f}(\bar{z} - he_p) - \bar{y}][\tilde{f}(\bar{z} - he_p) - \bar{y}]^T \\
+ \quad & \frac{h^2 - n}{h^2} [\tilde{f}(\bar{z}) - \bar{y}][\tilde{f}(\bar{z}) - \bar{y}]^T \ .
\end{aligned}
\tag{127}
$$

If we use this result in a state estimation context, we arrive at the exact same covariance estimate as the one proposed in [JU94], [JUDW95]. The estimate has the drawback that when $h^2 < n$, the last term in (127) becomes negative semi-definite. A possible implication of this could be that the covariance estimate becomes non-positive definite. To remedy this, [JU94] recommends that the following, more conservative, estimate is used

$$
\begin{aligned}
P_y \quad = \quad & \tfrac{1}{2h^2} \textstyle\sum_{p=1}^{n} \Big[[\tilde{f}(\bar{z} + he_p) - \tilde{f}(\bar{z})][\tilde{f}(\bar{z} + he_p) - \tilde{f}(\bar{z})]^T \\
& \qquad + \quad [\tilde{f}(\bar{z} - he_p) - \tilde{f}(\bar{z})][\tilde{f}(\bar{z} - he_p) - \tilde{f}(\bar{z})]^T \Big] \ .
\end{aligned}
\tag{128}
$$

In our framework this expression is achieved by deriving the covariance estimate so that a second-order polynomial replaces y in the evaluation of $E[yy^T]$ in

$$
P_y = E[yy^T] - \bar{y}\bar{y} \tag{129}
$$

while only a first-order polynomial approximation is used for evaluating $\bar{y}$ (corresponding to $\bar{y} = \tilde{f}(\bar{z})$).

The interested reader is referred to [JU94] for a thorough analysis of the estimates.