ON NONLINEAR TRANSFORMATIONS OF GAUSSIAN DISTRIBUTIONS

Gustaf Hendeby and Fredrik Gustafsson

Department of Electrical Engineering
Linköping University
SE-581 83 Linköping, Sweden
{hendeby, fredrik}@isy.liu.se

ABSTRACT

The *unscented Kalman filter* (UKF) relies on the *unscented transformation* (UT) that fits a Gaussian distribution to nonlinearly transformed so called *sigma points*. This contribution firstly gives the exact first and second order moments of the nonlinear transformation as a function of the rest term in a second order Taylor expansion, given that the gradient and Hessian are computable, and secondly compares these with what is obtained using the UT. A simple quadratic example illustrates the differences in the compared transformations.

***Index Terms*—** unscented transform, nonlinear transformation, extended Kalman filter

1. INTRODUCTION

This contribution considers the problem to compute a Gaussian approximation of $z = g(x)$ of a Gaussian distributed x. This transformation approximation is a fundamental step in nonlinear filtering, with the *extended Kalman filter* (EKF) [1–3] and the *unscented Kalman filter* (UKF) [4–6] as particular examples. Four methods are considered in this paper:

TT1 Gauss' approximation formula that is based on a first order Taylor expansion.

TT2 which compensates with the mean and covariance of the rest term of a second order Taylor expansion. [1, 3, 7]

UT the unscented transformation. [4]

MC the Monte Carlo approach.

TT1 (Gauss' approximation formula) is a simple and often used approximation, for instance in the EKF.

The exact form of TT2 is not often seen in literature, but is the exact way to include the correct first and second order moments of the transformation. TT2 relies on the gradient and Hessian of the nonlinear transformation to be computed, which is not always the case. For some reason, the resulting *Kalman filter* is not often seen in literature.

The UT only requires function evaluations, and often achieves a fairly good approximation of TT2. Since its introduction, the *unscented Kalman filter* (UKF) has met a tremendous interest in the literature.

It is often claimed in literature, that the UT is exact for transformations of Gaussian distributions for at least the two first moments (see *e.g.*, [4]). However, it will demonstrated using a simple example that this is not always the case. By letting the sigma points approach the mean, it will be shown that the UT gets a bias in the approximated variance.

However, the transformed sigma points can be used to recover important parts of the second order term. This paper shows the inner workings of the UT in terms of approximated derivatives, that is then compared to the TT2 method.

The authors gratefully acknowledge fundings from SSF (Swedish Foundation for Strategic Research) Strategic Research Center MOVIII, and VR (the Swedish Research Council) project Sensor Informatics.

2. NONLINEAR TRANSFORMATIONS REVISITED

2.1. Taylor Expansion

Consider a general nonlinear transformation and its second order Taylor expansion

$$z = g(x) = g(\mu_x) + g'(\mu_x)(x - \mu_x) + \underbrace{\left[\tfrac{1}{2}(x-\mu_x)^T g_i''(\xi)(x-\mu_x)\right]_i}_{r\left(x;\mu_x,g''(\xi)\right)}, \quad (1)$$

where n_x is the dimension of the vector x, $x \in \mathbb{R}^{n_x}$, and $z \in \mathbb{R}^{n_z}$. The notation $[v_i]_i$ is used to denote a vector in which element i is v_i. Analogously, the notation $[m_{ij}]_{ij}$ will be used to denote the matrix where the (i, j) element is m_{ij}.

Basically, nonlinear filters in literature are based on the following assumptions:

- EKF is based on the constant and linear terms in (1). This works fine as long as the rest term is small. Small here relates both to the state estimation error and the degree of nonlinearity of g. As a rule of thumb, the rest term is negligible if either
  - the model is almost linear,
  - the SNR is high, in which case the estimation error can be considered sufficiently small.
- The second order compensated EKF approximates the rest term $r(x; \mu_x, g''(\xi))$ with $r\big(x; \mu_x, g''(\mu_x)\big)$, and compensates for the mean and variance of this term. This works well if g'' varies little over the principal support of x.
- UKF estimates first moments of the nonlinear transformation in (1), without explicitly computing, or even assuming existence, of any derivatives of g.

There are several links and interpretations between UKF and EKF as will be pointed out later on.

2.2. Summary of Approximate Transformations

To summarize, the following options are available (the derivation of these will be elaborated on more in the full length paper):

TT1 First order Taylor approximation:

$$x \sim \mathcal{N}\big(\mu_x, P\big) \to z \sim \mathcal{N}\big(g(\mu_x), g'(\mu_x)P\big(g'(\mu_x)\big)^T\big). \tag{2}$$

TT2 Second order Taylor approximation:

$$x \sim \mathcal{N}\big(\mu_x, P\big) \to z \sim \mathcal{N}\Big(g(\mu_x) + [\mathrm{tr}(g_i''(\mu_x)P)]_i,$$
$$g'(\mu_x)P\big(g'(\mu_x)\big)^T + \tfrac{1}{2}\Big[\mathrm{tr}(Pg_i''(\mu_x)Pg_j''(\mu_x))\Big]_{ij}\Big). \tag{3}$$

UT Unscented transform approximation:

$$x \sim \mathcal{N}\big(\mu_x, P\big) \to z \sim \mathcal{N}\big(\mu_z, P_z\big), \tag{4}$$

where μ_z and P_z are defined as follows.

First define, u_i and σ_i from the *singular value decomposition* (SVD) of the covariance matrix P,

$$P = U\Sigma U^T = \sum_{i=1}^{n_x} \sigma_i^2 u_i u_i^T,$$

and then let

$$x^{(0)} = \mu_x, \qquad x^{(\pm i)} = \mu_x \pm \sqrt{n_x + \lambda}\sigma_i u_i,$$
$$\omega^{(0)} = \frac{\lambda}{n_x + \lambda}, \qquad \omega^{(\pm i)} = \frac{1}{2(n_x + \lambda)},$$

where $i = 1, \ldots, n_x$. Let $z^{(i)} = g(x^{(i)})$, and apply

$$\mu_z = \sum_{i=-n_x}^{n_x} \omega^{(i)} z^{(i)},$$
$$P_z = \sum_{i=-n_x}^{n_x} \omega^{(i)} (z^{(i)} - \mu_z)(z^{(i)} - \mu_z)^T$$
$$+ (1 - \alpha^2 + \beta)(z^{(0)} - \mu_z)(z^{(0)} - \mu_z)^T,$$

where $\omega^{(0)} + (1 - \alpha^2 + \beta)$ is often denoted $w_c^{(0)}$ and used to make the notation more compact for the covariance matrix expression.

MC Monte Carlo approximation:

$$x \sim \mathcal{N}\big(\mu_x, P\big) \to z \sim \frac{1}{N}\sum_{i=1}^{N} \delta_{z^{(i)}}, \tag{7}$$

where

$$x^{(i)} \sim \mathcal{N}\big(\mu_x, P\big), \quad i = 1, \ldots, N,$$
$$z^{(i)} = g(x^{(i)}),$$
$$\mu_z = \frac{1}{N}\sum_{i=1}^{N} z^{(i)},$$
$$P_z = \frac{1}{N-1}\sum_{i=1}^{N} \big(z^{(i)} - \mu_z\big)\big(z^{(i)} - \mu_z\big)^T.$$

The design parameters of UT have here the same notation as in UKF literature (*e.g.*, [6]):

- λ is defined by $\lambda = \alpha^2(n_x + \kappa) - n_x$.
- α controls the spread of the sigma points and is suggested to be approximately 10^{-3}.
- β compensates for the distribution, and should be chosen to $\beta = 2$ for Gaussian distributions.
- κ is usually chosen to zero.

Note that $n_x + \lambda = \alpha^2 n_x$ when $\kappa = 0$, and that for $n_x + \lambda \to 0^+$ the central weight $\omega^{(0)} \to -\infty$. Furthermore, $\sum_i \omega^{(i)} = 1$.

Recall from the introduction that TT1 is a computationally cheap approximation, TT2 recovers the first two moments if the gradient and Hessian in μ_x are available (for quadratic functions TT2 is completely correct, otherwise it is often a good approximation), the MC approach is always asymptotically correct, and that the UT is a fairly good compromise between TT2 and MC, that improves computational complexity to MC and the need for prior knowledge to TT2.

3. ANALYSIS OF THE UNSCENTED TRANSFORM

In this section the UT will be analyzed and expressions for the resulting mean and variance are given and interpreted in the limit as the sigma points approach the center point.

First, consider the mean of the transformed variable, for simplicity $\omega = \omega^{(0)}$ and remaining express the remaining weights in terms of it,

$$\mu_z = \sum_{i=-n_x}^{n_x} \omega^{(i)} z^{(i)} = z^{(0)} + \tfrac{1-\omega}{2n_x}\sum_{i=1}^{n_x}(z^{(i)} - 2z^{(0)} + z^{(-i)}). \tag{8a}$$

Continuing to compute the variance yields, using $(a)(\cdot)^T$ to denote aa^T to reduce the size of the expressions:

$$\begin{aligned}
P_z &= (1 - \alpha^2 + \beta + \omega)(z^{(0)} - \mu_z)(\cdot)^T \\
&\quad + \sum_{i\neq 0} \tfrac{1-\omega}{2n_x}(z^{(i)} - \mu_z)(\cdot)^T \\
&= (1 - \alpha^2 + \beta)\tfrac{(1-\omega)^2}{4n_x^2}\Big(\sum_{i>0} z^{(i)} - 2z^{(0)} + z^{(-i)}\Big)\Big(\,\cdot\,\Big)^T \\
&\quad + \tfrac{\omega(1-\omega)^2}{4n_x^2}\Big(\sum_{i>0} z^{(i)} - 2z^{(0)} + z^{(-i)}\Big)\Big(\,\cdot\,\Big)^T \\
&\quad + \tfrac{(1-\omega)(\omega^2-1)}{4n_x^2}\Big(\sum_{i>0} z^{(i)} - 2z^{(0)} + z^{(-i)}\Big)\Big(\,\cdot\,\Big)^T \\
&\quad + \tfrac{1-\omega}{2n_x}\sum_{i\neq 0}(z^{(i)} - z^{(0)})(\cdot)^T \\
&= (1 - \alpha^2 + \beta)\tfrac{(1-\omega)^2}{4n_x^2}\Big(\sum_{i>0} z^{(i)} - 2z^{(0)} + z^{(-i)}\Big)\Big(\,\cdot\,\Big)^T \\
&\quad - \tfrac{(1-\omega)^2}{4n_x^2}\Big(\sum_{i>0} z^{(i)} - 2z^{(0)} + z^{(-i)}\Big)\Big(\,\cdot\,\Big)^T \\
&\quad + \tfrac{1-\omega}{n_x}\sum_{i>0}(z^{(i)} - z^{(0)})(\cdot)^T.
\end{aligned} \tag{8b}$$

Observe that with the original formulation of the UT where the same weights are used for both mean and covariance computation the covariance matrix estimate may very well end up being indefinite or even negative definite, in contradiction to the definition of a covariance matrix.

With sigma points defined using the SVD, differences can be constructed that in the limit as $n_x + \lambda \to 0^+$, *i.e.*, $\alpha \to 0^+$ with $\kappa = 0$, yields the derivatives:

$$\frac{z^{(i)} - z^{(0)}}{\sigma_i \sqrt{n_x + \lambda}} \to g'(\mu_x) u_i \tag{9}$$

$$\frac{z^{(i)} - 2z^{(0)} + z^{(-i)}}{\sigma_i^2 (n_x + \lambda)} \to \big[u_i^T g_k'(\mu_x) u_i\big]_k \tag{10}$$

Note that $n_x + \lambda = (\omega - 1)/n_x$.

Using this, the limit case of (8) can be evaluated,

$$\mu_z \to g(\mu_x) + \tfrac{1}{2}\big[\operatorname{tr}(g_i'' P)\big]_i \tag{11a}$$

and

$$P_z \to g'(\mu_x) P \big(g'(\mu_x)\big)^T + \tfrac{(\beta-\alpha^2)}{4}\big[\operatorname{tr}\big(P g_i''(\mu_x)\big)\operatorname{tr}\big(P g_j''(\mu_x)\big)\big]_{ij}. \tag{11b}$$

These expressions for the approximation the UT can now be used to better compare it to the other derivative based algorithms.

4. COMPARING THE DERIVATIVE BASED METHODS

4.1. Scalar Case

First looking at the completely scalar case $g : \mathbb{R} \mapsto \mathbb{R}$, where TT2 and UT yield the same results ($\beta - \alpha^2 \approx 2$, for suggested parameter values), and TT1 lacks the second order term in both the mean and the variance approximation

$$\mu_z = g(\mu_x) + \tfrac{1}{2} P g''(\mu_x) \tag{12a}$$

$$P_z = \big(g'(\mu_x)\big)^2 P + \tfrac{1}{2}\big(g''(\mu_x)P\big)^2. \tag{12b}$$

4.2. Multidimensional Case

For the multidimensional case, TT2 and UT differs in their terms compensating for the quadratic effects. They both include the same first order terms as TT1. The correct second order compensation, given by TT2, is

$$\tfrac{1}{2}\Big[\operatorname{tr}(P g_i''(\mu_x) P g_j''(\mu_x))\Big]_{ij} \tag{13}$$

and the corresponding UT compensation is

$$\tfrac{(\beta-\alpha^2)}{4}\big[\operatorname{tr}\big(P g_i''(\mu_x)\big)\operatorname{tr}\big(P g_j''(\mu_x)\big)\big]_{ij}, \tag{14}$$

where $\beta - \alpha^2 \approx 2$ is recommended. The terms are similar, but not identical. The reason for the difference is that the UT cannot express the mixed second order derivatives needed for the TT2 compensation term without increasing the number of sigma points. The result of this approximation depends on the transformation and must analyzed for the case at hand.

5. EXAMPLE: SUM OF TWO SQUARED GAUSSIANS

This example demonstrates the effects of not having the cross-derivatives for the Hessian available in the UT, it also demonstrates the poor performance of TT1 even for quite well-behaved functions. To do this, study the transformation

$$z = g(x) = x^T x, \tag{15}$$

Table 1. Results of the different transformation methods on (15). The analytic result is a $\chi^2(2)$ distribution. The UT is for the recommended $\alpha = 10^{-3}$, $\beta = 2$, and $\kappa = 0$, the MC method uses 100 000 samples.

Method	**Mean**	**Variance**
Analytic	2	4
TT1	0	0
TT2	2	4
UT	2	8 ($= 4\beta + 2\alpha^2\kappa$)
MC	2.0	4.0

with $x \sim \mathcal{N}(\mathbf{0}_{2\times 1}, I_{2\times 2})$. The transformed variable z is now the sum of two squares of Gaussians, and hence $z \sim \chi^2(2)$ distributed with $\mathsf{E}(z) = 2$ and $\mathsf{var}(x) = 4$.

The results obtained when using the methods discussed in this paper can be found in Table 1. Some observations:

- TT1 totally fails since it does not at all consider the second order effects in the transformation.
- TT2 gives the correct answer, as expected for quadratic transformations.
- The approximation made in the UT due to the lack of mixed derivatives does in this case produce a too large variance, but the mean is correct. Note, there is no guarantee that the UT will overestimate the covariance and hence make the UKF more robust than a EKF based on TT2.
- The MC is correct, but the exact result will vary with different instances of random numbers.

6. REFERENCES

[1] Brian D. O. Anderson and John B. Moore, *Optimal Filtering*, Prentice-Hall, Inc, Englewood Cliffs, NJ, 1979.

[2] Andrew H. Jazwinski, *Stochastic Processes and Filtering Theory*, vol. 64 of *Mathematics in Science and Engineering*, Academic Press, Inc, 1970.

[3] Yaakvo Bar-Shalom and Thomas E. Fortmann, *Tracking and Data Association*, Academic Press, Inc, 1988.

[4] Simon J. Julier and Jeffrey K. Uhlmann, "Unscented filtering and nonlinear estimation," *Proc. IEEE*, vol. 92, no. 3, pp. 401–422, Mar. 2004.

[5] Simon J. Julier, Jeffrey K. Uhlmann, and Hugh F. Durrant-Whyte, "A new approach for filtering nonlinear systems," in *Proc. American Contr. Conf*, Seattle, WA, USA, June 1995, pp. 1628–1632.

[6] Eric A. Wan and Rudolph van der Merwe, "The uncented Kalman filter for nonlinear estimation," in *Proc. IEEE Adapt. Syst. for Signal Processing*, Lake Louise, AB, Canada, Oct. 2000, pp. 153–158.

[7] Fredrik Gustafsson, *Adaptive Filtering and Change Detection*, John Wiley & Sons, Ltd, 2000, 2 reprint.