Abstract—Kalman filter is a frequently used tool for linear state estimation due to its simplicity and optimality. It can further be used for fusion of information obtained from multiple sensors. Kalman filtering is also often applied to nonlinear systems. As the direct application of bayesian functional recursion is computationally not feasible, approaches commonly taken use either a local approximation - Extended Kalman Filter based on linearization of the non-linear model - or the global one, as in the case of Particle Filters.

An approach to the local approximation is the so called Unscented Kalman Filter. It is based on a set of symmetrically distributed sample points used to parameterise the mean and the covariance. Such filter is computationally simple and no linearization step is required.

Another approach to selecting the set of sample points based on decorrelation of multivariable random variables and Hermite-Gauss Quadrature is introduced in this paper. This approach provides an additional justification of the Unscented Kalman Filter development and provides further options to improve the accuracy of the approximation, particularly for polynomial nonlinearities. A detailed comparison of the two approaches is presented in the paper.

I. INTRODUCTION

Whenever the state of the system needs to be estimated from noisy measurements, some kind of state estimator must be used. Minimum mean squared state error estimate for linear systems results in Kalman filter (KF) [1]. It is an excellent tool for information fusion - processing data obtained from different sensors. Kalman filter is the best tool for tracking and estimation due to its simplicity and optimality. However, its application to nonlinear systems is difficult.

The most common approach to estimating states of nonlinear systems is to utilize the Extended Kalman filter [2], [3], which is based on the linearization of the non-linear model [4]. It is known that some difficulties with utilization of this approach may occur. First, the linearization requires computing Jacobian matrices, which is nontrivial. Moreover, the resulting filters may be unstable.

Kalman filter operates on means and covariances of the probability distribution which may be non-Gaussian. The so called Unscented Kalman filter [5] was developed, based on the set of symmetrically distributed sample points (so called sigma points), which are used to parametrise the means and covariances. Such filter is simple and no linearization step is required.

In this paper, another approach to selecting the set of sample points for obtaining the mean and covariance of the distribution is proposed, which is based on Hermite-Gauss Quadrature [6], [7], [8]. This approach is simpler compared to that using sigma points and exact for polynomial nonlinearities. These two approaches to Unscented Filter are compared in the paper.

The paper is organized as follows: in Section II, basic properties of Hermite-Gauss Quadrature are shown; Section III shows how the mean and variance of nonlinear function can be computed using Hermite-Gauss Quadrature. In Section IV, sigma point transformation for Unscented filter is compared with the results obtained by Hermite-Gauss Quadrature. Finally, Section V describes how to use Hermite-Gauss Quadrature in Kalman filter.

II. HERMITE-GAUSS QUADRATURE

The objective of this section is to compute

$$\int_a^b v(x)f(x)dx$$

where $v(x)$ is a priori chosen weighting function and $f(x)$ is some nonlinear function. For our purposes (Hermite Quadrature) the weighting function is chosen as $v(x) = e^{-x^2}$, and the interval of integration equals to $(a, b) = (-\infty, \infty)$.

We want to approximate such integral by a quadrature formula of the form

$$\int_a^b v(x)f(x)dx = A_1f(a_1)+A_2f(a_2)+\ldots+A_{n_a}f(a_{n_a})$$

where $A_i$ are the weighting coefficients, $a_i < a, b >$ are the nodes and $n_a$ is the order of the quadrature formula.

Quadrature formula has the algebraic accuracy $m$, if polynomials up to order $m$ are integrated exactly. Quadrature formula with $n_a$ nodes has $2n_a$ parameters $A_i, a_i$ which leads to the
algebraic accuracy \( m = 2n_a - 1 \), because a polynomial of order \( m = 2n_a - 1 \) has \( 2n_a \) coefficients.

The solution of Gauss Quadrature is given in the following theorem [9].

**Theorem:** Quadratic formula is Gaussian if the product of roots factors

\[
\omega(x) = \prod_{i=1}^{n_a} (x - a_i)
\]

is an orthogonal polynomial with weight \( v(x) \) and when coefficients of quadrature formula equal

\[
A_i = \int_a^b v(x) l_i(x) \, dx, \quad i = 1, 2, \ldots n_a
\]

where \( l_i \) are elementary Lagrange interpolating polynomials.

**Elementary Lagrange interpolating polynomials** \( l_i(x) \) are equal to

\[
l_i(x) = \prod_{j=1, j \neq i}^{n_a} \frac{x - a_j}{a_i - a_j}, \quad i = 1, 2, \ldots n_a
\]

Hence,

\[
l_i(a_j) = \delta_{i,j} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}
\]

**Hermite orthogonal polynomials** on the interval \((-\infty, \infty)\) with respect to weighting function \( v(x) = e^{-x^2} \) are defined by the recurrent relation

\[
H_{n+1}(x) = 2xH_n(x) - 2nH_{n-1}(x),
\]

where \( H_0(x) = 1, \ H_1(x) = 2x, \ H_2(x) = 4x^2 - 2 \). These polynomials are called ”physical” Hermite orthogonal polynomials.

Nodes \( a_i \) are roots of Hermite polynomials. Thus, we are able to approximate the integral

\[
\int_{-\infty}^{\infty} e^{-x^2} f(x) \, dx = A_1 f(a_1) + \ldots + A_n f(a_n).
\]

If the function \( f(x) \) is polynomial of order \( m \leq 2n_a - 1 \), then the result is exact. The weights \( A_i \) can be computed by (4) or by the formula

\[
A_i = \frac{2^{n-1}n!\sqrt{\pi}}{n!_2[H_{n-1}(a_i)]^2}
\]

In the following table there are nodes \( a_i \) and coefficients \( A_i \) for different orders of approximation \( n_a \).

<table>
<thead>
<tr>
<th>( n_a )</th>
<th>( a_i )</th>
<th>( A_i )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>( \sqrt{\pi} )</td>
</tr>
<tr>
<td>2</td>
<td>( a_1 = \frac{1}{\sqrt{2}} )</td>
<td>( A_1 = \frac{\sqrt{\pi}}{2} )</td>
</tr>
<tr>
<td>&amp; ( a_2 = -\frac{1}{\sqrt{2}} )</td>
<td>( A_2 = \frac{\sqrt{\pi}}{2} )</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>( a_1 = 0 )</td>
<td>( A_1 = \frac{2}{3} \sqrt{\pi} )</td>
</tr>
<tr>
<td>&amp; ( a_2 = \sqrt{\frac{2}{3}} )</td>
<td>( A_2 = \frac{\sqrt{\pi}}{\sqrt{6}} )</td>
<td></td>
</tr>
<tr>
<td>&amp; ( a_3 = -\sqrt{\frac{2}{3}} )</td>
<td>( A_3 = \frac{\sqrt{\pi}}{\sqrt{6}} )</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>( a_1 = 1.6507 )</td>
<td>( A_1 = 0.0813 )</td>
</tr>
<tr>
<td>&amp; ( a_2 = -1.6507 )</td>
<td>( A_2 = 0.0813 )</td>
<td></td>
</tr>
<tr>
<td>&amp; ( a_3 = 0.5246 )</td>
<td>( A_3 = 0.8049 )</td>
<td></td>
</tr>
<tr>
<td>&amp; ( a_4 = -0.5246 )</td>
<td>( A_4 = 0.8049 )</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>( a_1 = 0 )</td>
<td>( A_1 = 0.945 )</td>
</tr>
<tr>
<td>&amp; ( a_2 = 2.0202 )</td>
<td>( A_2 = 0.01995 )</td>
<td></td>
</tr>
<tr>
<td>&amp; ( a_3 = -2.0202 )</td>
<td>( A_3 = 0.01995 )</td>
<td></td>
</tr>
<tr>
<td>&amp; ( a_4 = 0.9486 )</td>
<td>( A_4 = 0.3936 )</td>
<td></td>
</tr>
<tr>
<td>&amp; ( a_5 = -0.9486 )</td>
<td>( A_5 = 0.3936 )</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>( a_1 = 2.3506 )</td>
<td>( A_1 = 0.0045 )</td>
</tr>
<tr>
<td>&amp; ( a_2 = -2.3506 )</td>
<td>( A_2 = 0.0045 )</td>
<td></td>
</tr>
<tr>
<td>&amp; ( a_3 = 1.335849 )</td>
<td>( A_3 = 0.1571 )</td>
<td></td>
</tr>
<tr>
<td>&amp; ( a_4 = -1.335849 )</td>
<td>( A_4 = 0.1571 )</td>
<td></td>
</tr>
<tr>
<td>&amp; ( a_5 = 0.436077 )</td>
<td>( A_5 = 0.7246 )</td>
<td></td>
</tr>
<tr>
<td>&amp; ( a_6 = -0.436077 )</td>
<td>( A_6 = 0.7246 )</td>
<td></td>
</tr>
</tbody>
</table>

**Note:** Some references (e.g. [4]) use the weighting function \( v(x) = e^{-x^2} \); related Hermite orthogonal polynomials \( H_n^S(x) \) are called ”statistical”. Such polynomials are defined by the recurrent relation

\[ H_{n+1}^S(x) = 2xH_n^S(x) - 2nH_{n-1}^S(x), \]

where \( H_0^S(x) = 1, \ H_1^S(x) = x, \ H_2^S(x) = x^2 - 1 \).

The nodes \( a_i^S \) are the roots of Hermite polynomials \( H_n^S(x) \) and the weights \( A_i^S \) are equal to

\[
A_i^S = \frac{n!\sqrt{2\pi}}{n!(H_{n-1}^S(a_i^S))^2}
\]

Using ”statistical” Hermite orthogonal polynomials the quadrature formula has the form

\[
\int_{-\infty}^{\infty} f(x)e^{-x^2} \, dx = \sum_{i=1}^{n} A_i^S f(a_i^S)
\]

Simple transformation \( s = \frac{x}{\sqrt{2}} \) can be used to relate quadrature of ”physical” and ”statistical” Hermite polynomials

\[
\int_{-\infty}^{\infty} f(x)e^{-x^2} \, dx = \int_{-\infty}^{\infty} \sqrt{2} f(\sqrt{2}s)e^{-s^2} \, ds
\]

Similar results are obtained using Hermite-Gauss ”physical” Quadrature:

\[
\int_{-\infty}^{\infty} f(x)e^{-x^2} \, dx = \sum_{i=1}^{n} \tilde{A}_i f(\tilde{a}_i)
\]

where \( \tilde{A}_i = \sqrt{2}A_i \) and \( \tilde{a}_i = \sqrt{2}a_i \).
III. MEAN AND VARIANCE BY HERMITE-GAUSS QUADRATURE

In this section we use Hermite-Gauss quadrature to compute the mean and the covariance of function \( f(x) \) of a random variable \( x \). The mean \( \mu_f = \mathcal{E} \{ f(x) \} \), where \( x \) is a random variable with probability density function \( p(x) \), equals

\[
\mu_f = \int_{-\infty}^{\infty} f(x)p(x)dx.
\]

First, we will treat the one dimensional case; subsequently, the extension to multi-dimensional cases will be made.

A. One dimensional case

If a random variable \( x \) is of Gaussian distribution with mean \( \mu \) and variance \( \sigma^2 \), then \( \mu_f \) equals

\[
\mu_f = \int_{-\infty}^{\infty} f(x)\frac{1}{\sqrt{2\pi\sigma}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx
\]

If the substitution \( \frac{x-\mu}{\sqrt{\sigma^2}} = v \) is used, the formula for \( \mu_f \) equals

\[
\mu_f = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} f(\sqrt{2}\sigma v + \mu)e^{-v^2}dv
\]

Then Hermite-Gauss Quadrature can be used in the form

\[
\mu_f = \frac{1}{\sqrt{\pi}} \left\{ A_1 f(\bar{a}_1) + \ldots + A_n f(\bar{a}_n) \right\}
\]

where \( \bar{a}_i = \sqrt{2} \sigma a_i + \mu, i = 1, \ldots, n \), and \( A_i \) are coefficients of Hermite-Gauss Quadrature and \( a_i \) are nodes of Hermite orthogonal polynomials of order \( n \).

For the order of Hermite polynomial \( n = 2 \), the formula for \( \mu_f = \mathcal{E} \{ f(x) \} \) is simple:

\[
\mu_f = \frac{1}{2} \left\{ f(\mu + \sigma) + f(\mu - \sigma) \right\}
\]

For the order of Hermite polynomial \( n = 3 \), the formula for \( \mu_f = \mathcal{E} \{ f(x) \} \) becomes

\[
\mu_f = \frac{1}{6} \left\{ f(\mu) + f(\mu + \sqrt{3}\sigma) + f(\mu - \sqrt{3}\sigma) \right\}
\]

For variance \( \sigma_f^2 \) of function \( f(x) \), which equals \( \sigma_f^2 = \mathcal{E} \{ (f(x) - \mu_f)^2 \} \), a similar formula for Hermite-Gauss Quadrature can be obtained

\[
\sigma_f^2 = \frac{1}{\sqrt{\pi}} \left\{ A_1 \left[ f(\bar{a}_1) - \mu_f \right]^2 + \ldots + A_n \left[ f(\bar{a}_n) - \mu_f \right]^2 \right\}
\]

where the nodes \( \bar{a}_i \) and coefficients \( A_i \) are the same as in the previous case.

Example 1: For \( f(x) = x^2 \), formula (18) yields the mean of \( f(x) \) as

\[
\mu_f = \frac{1}{2} \left[ (\mu + \sigma)^2 + (\mu - \sigma)^2 \right] = \mu^2 + \sigma^2,
\]

which is a well-known formula. The same result is obtained if the order of Hermite polynomial equals \( n = 3 \).

For the variance of random function \( f(x) = x^2 \) it is necessary to use the order of Hermite polynomial \( n = 3 \). It is because the variance \( \sigma_f^2 = \mathcal{E} \{ (f(x) - \mu_f)^2 \} \); the order of the function for which the mean is computed equals \( 2n = 4 \) and \( 2n_a - 1 = 6 - 1 = 5 > 4 \). The variance \( \sigma_f^2 \) equals

\[
\sigma_f^2 = \frac{1}{\sqrt{\pi}} \sum_{i=1}^{3} A_i \left[ f(\sqrt{2} a_i + \mu) - \mu_f \right]^2.
\]

After the substitution of coefficients \( A_i \) and nodes \( a_i \), the formula for \( \sigma_f^2 = \mathcal{E} \{ (f(x) - \mu_f)^2 \} \) is

\[
\sigma_f^2 = \frac{1}{6} \left\{ (\mu + \sqrt{5}\sigma)^2 - \mu_f \right\}^2 + \left\{ (\mu - \sqrt{5}\sigma)^2 - \mu_f \right\}^2 = 2\sigma^4.
\]

Example 2: If \( f(x) = \sqrt{x} \) for \( x > 0 \) and \( f(x) = 0 \) for \( x < 0 \) then, for the order of Hermite polynomial \( n_a = 2 \), formula (18) for the mean yields

\[
\mu_f = \frac{1}{2} \left[ \sqrt{\mu + \sigma} + \sqrt{\mu - \sigma} \right]
\]

If the mean of random variable \( x \) equals \( \mu = 10 \) and variance \( \sigma = 3 \), the mean of the function \( f(x) \) equals

\[
\mu_f = \frac{1}{2} \left[ \sqrt{10 + 3} + \sqrt{10 - 3} \right] = 3.1257
\]

If \( n_a = 3 \), the mean of \( f(x) = \sqrt{x} \) under the same conditions equals

\[
\mu_f = \frac{1}{6} \left[ 4\sqrt{\mu + \sqrt{3}\sigma} + \sqrt{\mu - \sqrt{3}\sigma} \right] = 3.1232
\]

Sometimes approximate formula \( \mu_f = \bar{f}(\mu) = \sqrt{\bar{P}} \) is used, which in our case gives the result \( \mu_f = 3.1623 \). The same result is obtained by Hermite-Gauss Quadrature if the order of quadrature equals \( n_a = 1 \). The mean value obtained by Monte Carlo simulation (with \( 10^7 \) samples) gives the result \( \mu_f = 3.1499 \). The polynomial approximation is not exact in this case, but the results are satisfactory.

B. Extension to multidimensional cases

Let us assume \( n \) dimensional random vector \( x \) with mean \( \mu \) and covariance matrix \( P \). We are looking to compute first two moments of multidimensional function \( f(x) \) of random vector \( x \).

Vector mean \( \mu_f = \mathcal{E} \{ f(x) \} \) of function \( f(x) \) equals

\[
\mu_f = \int_{-\infty}^{\infty} f(x)\frac{1}{\sqrt{(2\pi)^n|P|}}e^{\frac{1}{2}(x-\mu)^T P^{-1}(x-\mu)}dx
\]

where \( |P| \) is the determinant of covariance matrix \( P \).

To simplify the exponent in the previous formula, let us make the substitution \( \frac{x-\mu}{\sqrt{\bar{P}}} = \sqrt{\bar{P}}v \), where the square root of covariance matrix \( P = \sqrt{\bar{P}} \). The square root of a covariance matrix can be obtained by Cholesky factorization. Realize that \( |P| = |(\sqrt{\bar{P}})^2| \) and according the substitution theorem \( dx = 2|\sqrt{\bar{P}}|dv \).
After the substitution, the formula for the mean of \( f(x) \) becomes

\[
\mu_f = \frac{\sqrt{2}}{\sqrt{(2\pi)^n}} \int_{-\infty}^{\infty} f(\sqrt{2} \sqrt{P} v + \mu) e^{-v^T v} dv
\]

Using the so called stochastic decorelation technique, vector \( \sqrt{P} v \) can be expressed as

\[
\sqrt{P} v = (\sqrt{P})_1 v_1 + (\sqrt{P})_2 v_2 + \ldots + (\sqrt{P})_n v_n
\]

where \((\sqrt{P})_i\) is the \(i\)th column of matrix \( \sqrt{P} \). According to the Fubinia theorem, a multidimensional integral can be changed to the product of one dimensional integrals of the form

\[
\frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} f(\sqrt{2} \sqrt{P} v_i + \mu) e^{-v_i^2} dv_i
\]

These integrals can be solved by Gauss-Hermite Quadrature and so the function mean \( \mu_f \) equals

\[
\mu_f = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[ A_1 f(\sqrt{2}(\sqrt{P})_i a_1 + \mu) + \ldots + A_n f(\sqrt{2}(\sqrt{P})_i a_n + \mu) \right]
\]

(25)

For the order of quadrature \( n_a = 2 \), the formula for function mean takes the simple form

\[
\mu_f = \sum_{i=1}^{n} \left[ f((\sqrt{P})_i + \mu) + f(-(\sqrt{P})_i + \mu) \right];
\]

(26)

the corresponding formula for \( n_a = 3 \) is given by

\[
\mu_f = \sum_{i=1}^{n} \frac{1}{6} \left[ 4 f(\mu) + f(\mu + \sqrt{3}(\sqrt{P})_i \xi_a + f(\mu - \sqrt{3}(\sqrt{P})_i \xi_a) \right]
\]

Covariance matrix \( P_f \) of vector function \( f(x) \) equals

\[
P_f = \mathcal{E} \{ (f(x) - \mu_f) (f(x) - \mu_f)^T \}
\]

and can be expressed by integral formula

\[
P_f = \int_{-\infty}^{\infty} (f(x) - \mu_f) (f(x) - \mu_f)^T e^{-v^T v} dv
\]

This relation is simplified using the substitution \( \frac{v - \mu}{\sqrt{2}} = \sqrt{P} v \), where \( \sqrt{P} \) is again obtained via Cholesky factorization. The resulting formula for covariance matrix \( P_f \) is obtained as

\[
P_f = \frac{\sqrt{2}}{\sqrt{(2\pi)^n}} \int_{-\infty}^{\infty} (f(\sqrt{2} \sqrt{P} v + \mu) - \mu_f) (f(\sqrt{2} \sqrt{P} v - \mu) - \mu_f)^T e^{-v^T v} dv
\]

This integral can be solved by Gauss-Hermite Quadrature

\[
P_f = \frac{1}{\sqrt{\pi}} \sum_{i=1}^{n} \left[ A_1 \left( f(\sqrt{2}(\sqrt{P})_i a_1 + \mu) - \mu_f \right) \right] \times
\]

\[
\left( f(\sqrt{2}(\sqrt{P})_i a_1 + \mu) - \mu_f \right)^T + \ldots + A_n \sum_{i=1}^{n} \left( f(\sqrt{2}(\sqrt{P})_i a_n + \mu) - \mu_f \right) \times
\]

\[
\left( f(\sqrt{2}(\sqrt{P})_i a_n + \mu) - \mu_f \right)^T
\]

(28)

For the order of quadrature \( n_a = 2 \) the formula for covariance matrix is obtained as

\[
P_f = \frac{1}{2} \sum_{i=1}^{n} s_i s_i^T + w_i w_i^T
\]

(29)

where \( s_i = (f((\sqrt{P})_i + \mu) - \mu_f) \) and \( w_i = (f(-((\sqrt{P})_i + \mu) - \mu_f) \).

For order of Hermite polynomial \( n_a = 3 \), the formula for covariance matrix takes the form

\[
P_f = \frac{1}{6} \sum_{i=1}^{n} 4 s_i s_i^T + w_i w_i^T + z_i z_i^T
\]

(30)

where \( s_i = (f(\mu) - \mu) \), \( w_i = (f(\sqrt{3}(\sqrt{P})_i + \mu) - \mu_f) \) and \( z_i = (f(-\sqrt{3}(\sqrt{P})_i + \mu) - \mu_f) \).

IV. SIGMA POINT TRANSFORMATION FOR UNSCENTED FILTER

Unscented Kalman filter based on the so called reduced sigma points is described in [5]. These points are mapped by a nonlinear transformation to obtain the mean and the covariance of nonlinear function. The algorithm will be described, using the notation introduced in [5].

A symmetrically distributed set of points which match the mean and covariance is obtained as

\[
\chi_0 = \hat{x} \\
\chi_i = \hat{x} + \sqrt{(n+k) P_i} \\
\chi_{i+n} = \hat{x} - \sqrt{(n+k) P_i}
\]

and a set of weights is chosen as

\[
W_0 = \frac{k}{n+k}, \quad W_i = W_{n+i} = \frac{1}{2(n+k)}
\]

There, \( k \in R \) is a tuning parameter and \( i = 1, \ldots, n \). Further, \( \hat{x} \) and \( P \) denote the mean and the covariance of \( x \), respectively; \( P_i \) is \( i \)th column of covariance matrix \( P \).

Let the random vectors \( x \) and \( y \) be related by a known nonlinear function \( y = f(x) \). The problem is to calculate the mean and the covariance matrix of vector \( y \) and the cross-covariance matrix \( P_{xy} \), i.e.,

\[
\hat{y} = \mathcal{E} \{ y \} = \mathcal{E} \{ f(x) \}
\]

\[
\begin{align*}
P_{yy} &= \mathcal{E} \{ (y - \hat{y})(y - \hat{y})^T \} \\
P_{xy} &= \mathcal{E} \{ (x - \hat{x})(y - \hat{y})^T \}
\end{align*}
\]
The solution of this problem by Unscented Transformation is based on the selection of symmetrically distributed set of points \( X_i, X_i \) and \( X_{i+n} \) and transformed points \( Y_i \) related to \( X_i \) as \( Y_i = f(X_i) \), \( i = 0, 1, \ldots, 2n \). The solution is given by

\[
\dot{y}^{UT} = \sum_{i=0}^{p} W_i \dot{Y}_i
\]

\[
P_{yy}^{UT} = \sum_{i=0}^{p} W_i (Y_i - \dot{y})(Y_i - \dot{y})^T
\]

\[
P_{xy}^{UT} = \sum_{i=0}^{p} W_i (X_i - \dot{x})(Y_i - \dot{y})^T
\]

where \( p = 2n \). The constant \( k \) is chosen so that \( (n + k) = 3 \).

If the notation introduced in the previous sections is used, the set of sigma point vectors is given by

\[
a_1 = \mu
\]

\[
a_{2,i} = \mu + \sqrt{(n + k)(\sqrt{P})i}
\]

\[
a_{3,i} = \mu - \sqrt{(n + k)(\sqrt{P})i}
\]

where \( (\sqrt{P})i \) is the \( i \)th column of square root of covariance matrix and the weights

\[
A_1 = \frac{k}{n + k}, \quad A_2 = A_3 = \frac{1}{2(n + k)}
\]

The mean of function \( f(x) \) using Unscented transformation equals

\[
\mu_f^{UT} = A_1 f(\mu) + \sum_{i=1}^{n} A_2 f(\mu + \sqrt{(n + k)(\sqrt{P})i}) + A_3 f(\mu - (n + k)(\sqrt{P})i)
\]

and after the substitution the formula simplifies to

\[
\mu_f^{UT} = \frac{k}{3} f(\mu) + \frac{1}{6} \sum_{i=1}^{n} f(\mu + \sqrt{3}(\sqrt{P})i) + f(\mu - \sqrt{3}(\sqrt{P})i)
\]

where \( n + k = 3 \) is used. The covariance matrix and cross-covariance matrix computed by sigma points result in

\[
P_{f}^{UT} = A_1 s s^T + \sum_{i=1}^{n} A_2 w_i u_i^T + A_3 z_i z_i^T
\]

\[
P_{xy}^{UT} = \sum_{i=1}^{n} A_2 (\sqrt{(n + k)(\sqrt{P})i}) u_i^T + A_3 (\sqrt{3}(\sqrt{P})i) z_i^T
\]

where \( s = (f(\mu) - \mu_f) \), \( w_i = (f(\mu + \sqrt{(n + k)(\sqrt{P})i}) - \mu_f) \) and \( z_i = (f(\mu - (n + k)(\sqrt{P})i) - \mu_f) \). Good performance of this algorithm is shown in [5].

**Example 3:** Let us compare the sigma point filter with Hermite-Gauss Quadrature (HGQ) algorithm for one dimensional case with function \( f(x) = x^2 \). The function mean using HGQ is solved in Example 1 obtaining \( \mu_f = \mu^2 + \sigma^2 \). Using sigma point transformation the function mean is obtained as

\[
\mu_f = \frac{k}{3} \mu^2 + \frac{1}{6} \left[ (\mu + \sqrt{3}\sigma)^2 + (\mu - \sqrt{3}\sigma)^2 \right]
\]

\[
= \left( \frac{k}{3} + 1 \right) \mu^2 + \sigma^2.
\]

For \( k = 2 \) the same result is obtained as for HGQ, equal to the exact value.

For variance \( \sigma^2_f \) of function \( f(x) = x^2 \), the result using HGQ was obtained in Example 1 as \( \sigma^2_f = 2\sigma^4 \). Using sigma point transformation, the variance approximated as

\[
\sigma^2_f = A_1 \left[ \mu^2 - \mu_f \right]^2 + A_2 \left[ (\mu + \sqrt{3}\sigma)^2 - \mu_f \right]^2 + A_3 \left[ (\mu - \sqrt{3}\sigma)^2 \right]^2
\]

and, after the substitution of the weight values, the same result \( \sigma^2_f = 2\sigma^4 \) are recovered (again for \( k = 2 \)).

For multidimensional cases, due to the stochastic decorrelation technique, the situation is similar. Unscented transformation based on sigma points gives similar results as Hermite-Gauss Quadrature for order of quadrature \( n_q = 3 \). For more complicated functions \( f(x) \), it is possible to use a higher order of quadrature resulting in more accurate results. So it can be stated that computation the mean and variance of function \( f(x) \) using Hermite-Gauss-Quadrature is more general.

**V. Utilization of Hermite-Gauss Quadrature in Kalman Filter**

State estimation problem for discrete time nonlinear stochastic system is solved. State equations of the system have the form

\[
x(t + 1) = f(x(t), u(t)) + v(t), \quad y(t) = g(x(t), u(t)) + e(t), \quad (31)
\]

where \( v(t) \) is zero mean white system noise sequence independent of the past and current states. Usually, normality of the noise is considered \( v(t) \sim \mathcal{N}(0, R_e(t)) \). Similarly \( e(t) \) is also zero mean white measurement noise sequence of known probability density function (p.d.f.), independent of past and present state and system noise. Normality of the measurement noise is also usually assumed, \( e(t) \sim \mathcal{N}(0, R_e(t)) \).

**A. Bayesian approach to state estimation**

Assume we observe the inputs \( u(\tau) \) and outputs \( y(\tau) \) for \( \tau = 1, \ldots, t-1 \) and our knowledge of the parameters and state of the process based on the data set

\[
\mathcal{D}^{t-1} = \{ u(1), y(1), \ldots, u(t-1), y(t-1) \}
\]

is described by a conditional probability density function (c.p.d.f.) \( p(x(t)|\mathcal{D}^{t-1}) \). The problem is how to update the knowledge described by c.p.d.f. \( p(x(t)|\mathcal{D}^{t-1}) \) to \( p(x(t+1)|\mathcal{D}^t) \) after new input-output data \( \{ u(t), y(t) \} \) have been measured. The output equation (31b), defining the c.p.d.f.
\( p(y(t)|x(t), u(t)) \) and state transition equations (31a) defining the c.p.d.f. \( p(x(t+1)|x(t), u(t), y(t)) \) are given. The solution can be given in the following steps:

1. C.p.d.f. \( p(x(t)|D^{t-1}) \) is given.
2. Using the output model \( p(y(t)|x(t), u(t)) \) determine the joint c.p.d.f.

\[
p(y(t), x(t)|D^{t-1}, u(t)) = p^2(y(t)|x(t)|D^{t-1}, u(t))
\]

(32)

\[
p(y(t)|x(t), u(t)) = \int p(y(t)|x(t)|D^{t-1}, u(t)) \, dx(t)
\]

(33)

Natural condition of control

\[
p(x(t)|D^{t-1}, u(t)) = p(x(t)|D^t)
\]

is used to complete this step. Natural condition of control expresses the fact that all information about state is in input-output data only and controller which generates the control \( u(t) \) has no extra information about the state.

3. Using the output measurement \( y(t) \), determine the c.p.d.f.

\[
p(x(t)|D^t) = \frac{p(y(t), x(t)|D^{t-1}, u(t))}{p(y(t)|D^{t-1}, u(t))}
\]

4. Using the state transition model (31a), for which

\[
p(x(t+1)|x(t), D^t) = p(x(t+1)|x(t), u(t), y(t))
\]

(35)

determine the predictive c.p.d.f.

\[
p(x(t+1)|D^t) = \int p(x(t+1)|x(t), D^t) \, p(x(t)|D^t) \, dx(t)
\]

(36)

**B. Kalman filter**

Kalman filter operates on the first two moments of random variable \( x(t) \). Our aim is to estimate the state mean of the system which can be denoted as \( \hat{x}(t, i) \) and state covariance matrix \( P(t, i) \). It is the state mean and covariance estimate in time \( t \) based on the data \( u(\tau), y(\tau) \) till time \( i \). Kalman filter consists on two steps:

**Data update step:**

Let us have state mean estimate \( \hat{x}(t, t-1) \) and state covariance matrix \( P(t, t-1) \) and new data \( y(t), u(t) \) are obtained. Data update step or prediction step of Kalman filter is the following: Predictive state mean equals

\[
\hat{x}(t, t) = \hat{x}(t, t-1) + P_{xy}(t, t-1)P_{yy}(t, t-1)^{-1} \times (y(t) - \hat{y}(t, t-1))
\]

(37)

where \( \hat{y}(t, t-1) = E \{g(x(t, t-1), u(t-1)) \} \) and state covariance matrix

\[
P(t, t) = P(t, t-1) - P_{xy}(t, t-1)P_{yy}(t, t-1)^{-1}P_{yx}(t, t-1)
\]

(38)

Covariances and cross-covariances are given by

\[
P_{yy}(t, t-1) = E \{(y(t) - \hat{y}(t, t-1)) \times (y(t) - \hat{y}(t, t-1))^T \} + R_e
\]

(39)

\[
P_{xy}(t, t-1) = E \{(x(t) - \hat{x}(t, t-1)) \times (y(t) - \hat{y}(t, t-1))^T \}
\]

(40)

**Time update step:**

Let us have the state mean estimate \( \hat{x}(t, t) \) and we can proceed further to obtain \( \hat{x}(t+1, t) \), which is the state mean estimate in time \( t+1 \), based on the same set of data. Such estimate is called time update or model update step, because such update is based only on the model of the system and no new data are given. So state time update mean equals

\[
\hat{x}(t+1, t) = E(f(x(t), u(t)) + v(t))
\]

(41)

and state covariance matrix

\[
P(t+1, t) = E \{(f(.) - \hat{x}(t, t))(f(.) - \hat{x}(t, t))^T\} + R_e.
\]

All means and covariances can be obtained by Hermite-Gauss Quadrature or Unscented transformation as it is shown in the next section.

**C. Kalman filter by Hermite-Gauss Quadrature**

Data update step:

State \( \hat{x}(t, t) \) is computed according to (37) where

\[
\hat{y}(t, t-1) = E \{g(x(t, t-1), u(t-1)) \} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} A_i g(\sqrt{\mathcal{P}} a_i + \hat{x}(t, t-1), u(t)) + \ldots + A_{na} g(\sqrt{\mathcal{P}} a_{na} + \hat{x}(t, t-1), u(t)),
\]

and the covariances

\[
P_{yy}(t, t-1) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} A_i s_{i,1} s_{i,1}^T + \ldots + A_{na} s_{i,na} s_{i,na}^T + R_e,
\]

(42)

where \( s_{i,j} = g(\sqrt{\mathcal{P}} a_j + \hat{x}(t, t-1), u(t)) - \hat{y}(t, t-1), \)

\[
P_{xy}(t, t-1) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} A_i w_{i,1} s_{i,1}^T + \ldots + A_{na} w_{i,na} s_{i,na}^T,
\]

where \( w_{i,j} = f(\sqrt{\mathcal{P}} a_j + \hat{x}(t, t-1), u(t)) - \hat{x}(t, t-1). \)

The solution for time update step by Hermite-Gauss Quadrature equals:

\[
\hat{x}(t+1, t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} A_i f(\sqrt{\mathcal{P}} a_i + \hat{x}(t, t), u(t)) + \ldots + A_{na} f(\sqrt{\mathcal{P}} a_{na} + \hat{x}(t, t), u(t)),
\]

and the state covariance

\[
P(t+1, t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} A_i z_{i,1} z_{i,1}^T + \ldots + A_{na} z_{i,na} z_{i,na}^T + R_e,
\]

where \( z_{i,j} = f(\sqrt{\mathcal{P}} a_j + \hat{x}(t, t), u(t)) - \hat{x}(t+1, t). \)

In data update step we set \( \sqrt{\mathcal{P}}_t = \sqrt{\mathcal{P}(t, t-1)_t} \), while in time update step it is \( \sqrt{\mathcal{P}}_t = \sqrt{\mathcal{P}(t, t-1)_t}. \)
D. Kalman filter by Cholesky factors of covariance matrices

From the previous formulas it is obvious that only Cholesky factors $\sqrt{P}$ of covariance matrices are needed. For the data update step according the formula (39) and (42) the covariance matrix $P_{yy}(t, t-1)$ can be expressed as

$$P_{yy}(t, t-1) = NN^T$$

where matrix $N$ is of the form

$$N = \begin{bmatrix} s_{1,1}, \ldots, s_{n,1}, \ldots, s_{n,n_a}, I_{m \times m} \end{bmatrix} \times \begin{bmatrix} \sqrt{A_1}/\sqrt{n} & \ddots & \sqrt{A_n}/\sqrt{n} \end{bmatrix}$$

It is clear from (37) that for Kalman gain $K = P_{zy}P_{yy}^{-1}$, (the arguments being omitted) there is $P_{zy} = KP_{yy}$. The mutual covariance matrix $P_c$ is given by

$$P_c = \begin{bmatrix} P_{yy} & P_{zx} \\ P_{xz} & P_{xx} \end{bmatrix} = \begin{bmatrix} NN^T & NN^TK^T \\ KN^T & MM^T \end{bmatrix}$$

where matrix $M$ is the Cholesky factor of the covariance matrix $P_{xx}$.

The mutual covariance matrix $P_c$ can be expressed in the form $P_c = QQ^T$ where $Q$ is given by

$$Q = \begin{bmatrix} \sqrt{A_1}/\sqrt{n} & \ddots & \sqrt{A_n}/\sqrt{n} & \sqrt{R_c} & 0 \\ 0 & \ddots & 0 & 0 & \sqrt{R_c} \end{bmatrix} \begin{bmatrix} s_{1,1}, \ldots, s_{n,1}, \ldots, s_{n,n_a}, I_{m \times m}, 0, I_{n \times n} \end{bmatrix}$$

Applying some orthogonal transformation, matrix $Q$ can be transformed to a lower triangular matrix such that

$$P_c = QQ^T = \begin{bmatrix} H & 0 & G^T \\ & H^T & F^T \end{bmatrix} = \begin{bmatrix} HH^T & H \sqrt{G} \sqrt{T} \\ & G^T + FF^T \end{bmatrix}$$

where matrices $H$ and $F$ are lower triangular matrices. It can be proved that according to formula (38) the state covariance matrix $P(t, t)$ equals

$$P(t, t) = FF^T$$

Comparing two previous matrices reveals that $GG^T + FF^T = MM^T$, and hence $FF^T = MM^T - GG^T$. The state covariance matrix $P(t, t)$ which is the result of data update step according to (38) equals

$$P(t, t) = MM^T - KNN^TK^T = MM^T - KHH^TK^T$$

and therefore, $G = KH$.

Updating state mean in the data update step is simple, using the formula

$$\hat{x}(t, t) = \hat{x}(t, t-1) + Gs$$

where the vector $s$ is obtained as the solution of linear algebraic equation

$$Hs = y(t) - \hat{y}(t, t-1).$$

where $H$ is a lower triangular matrix. Previous formulas follow from equation for Kalman gain $K = GH^{-1}$.

In the time update step, state $\hat{x}(t+1, t)$ is obtained in the standard way using the Hermite-Gauss Quadrature. The state covariance matrix is given by $P(t+1, t) = SS^T$, where the auxiliary matrix $S$ equals

$$S = \begin{bmatrix} z_{1,1}, \ldots, z_{n,1}, \ldots, z_{n,n_a}, I_{n \times n} \end{bmatrix} \times \begin{bmatrix} \sqrt{A_1}/\sqrt{n} & \ddots & \sqrt{A_n}/\sqrt{n} & \sqrt{R_c} \end{bmatrix}$$

Cholesky factorization

$$P(t+1, t) = SS^T = \begin{bmatrix} Q & 0 \\ & Q^T \end{bmatrix}$$

results in lower triangular matrix $Q$ – the desired Cholesky factor of state covariance matrix $QQ^T = P(t+1, t)$. This concludes the algorithm of Kalman filter operating entirely on Cholesky factors of covariance matrices.

VI. CONCLUSION

We described in this paper how to use Hermite-Gauss Quadrature for computing the mean and variance of function $f(x)$, where $x$ is a random variable of normal distribution whose mean and variance are known. Kalman filtering involving nonlinear systems results in a non-normal distribution of the random state $x$. Applying the proposed procedure thus yields an approximation of this random variable distribution by the first two moments.

Unscented filter based on sigma points selection was introduced in [5]. In this paper it was shown that this filter gives the same results as that using Hermite-Gauss Quadrature introduced here, if the order of quadrature equals $n_a = 3$. Hermite-Gauss-Quadrature can be used also for higher or lower order of quadrature and if the function $f(x)$ is a polynomial function of normal random variable $x$, the results obtained by Hermite-Gauss Quadrature of sufficient order are exact.

While attempts have been made to improve the accuracy of the Unscented filter by sigma point randomization [10], Hermite-Gauss Quadrature provides an alternative approach with a sound mathematical background.
ACKNOWLEDGMENT

This work was supported by the grant P103/11/1353 of the Czech Science Foundation.

REFERENCES