A new set of sigma/Quadrature points for a Gaussian PDF that can capture all 4th and 6th moments "exactly"

Nagavenkat Adurthi

August 22, 2011

From the basics

Consider a discrete dynamic system with noise

$$x_{k+1} = f(x_k, k) + \nu_k$$

The PDF of this dynamic system propagates according to the **Chapman Kolmogorov Equation**.

$$P(x_{k+1}) = \int P(x_{k+1}|x_k)P(x_k)dx_k$$

- $P(x_k)$ in the CKE need not be gaussian at all times even though the initial condition was gaussian.

- It would be gaussian at all time only when system is linear and the noise is also gaussian. In this case it is very easy to **solve this equation analytically**.

Incase the the system in nonlinear and noise is gaussian, the CKE would be

$$P(x_{k+1}) = \int \mathsf{N}(x_{k+1}|x_k)P(x_k)dx_k$$

Extended Kalman Filter

$P(x_k)$ is not always gaussian and there is no analytical solution to this equation. Hence the EKF emerged that can provide an **analytical solution** to this CKE by considering the following **two approximations**

- $P(x_k)$ is replaced by an equivalent gaussian PDF that has the same first two moments as the original PDF $P(x_k)$.

- $f(x_k, k)$ is linearized in $N(x_{k+1}|x_k)$.

- The EKF works well for systems in which the linearized dynamics is a good approximation to the nonlinear system.- **Higher order terms in Taylor series are negligible**.

- If the nonlinearity is too strong the EKF would diverge.

- Above that during the linearization process the computation of the jacobian is **computationally expensive**.

- The EKF **disregards the actual state PDF** and propagates only the first two moments of the state PDF. The linearized dynamics are used in the propagation of the first two moments.

Propagation of mean

$$\mu_{k+1} = f(\mu_k)$$

Propagation of Covariance

$$P_{k+1} = AP_kA^T + Q_k$$

Where A is the jacobian of the system.

$$A = \frac{\partial f}{\partial x}|_{\mu_k}$$

Linear Regression Kalman Filter - LRKF

• The LRKF also uses linearized dynamics but the linearization is not done using the taylor series (hence evaluating jacobian) but by a method called **statistical linearization**.

• Firstly sample points are chosen about the current mean at time k such that the mean of the samples and covariance of the samples **match** the current mean and current covariance.

• Each point is propagated using the **nonlinear dynamics** of the system to time step $k + 1$.

- Now between the current sample points at time k and the corresponding propagated points at time $k+1$ a **linear model is fit** which gives rise to the linearized dynamics of the system at time k.

- Now this linearized dynamics is used to compute the mean and covariance at time step $k+1$.

The sample points at time k are

$$\mu_k = \frac{1}{n} \sum_{i=1}^{N} X^i$$

$$P_k = \frac{1}{n} \sum_{i=1}^{N} (X^i - \mu_k)(X^i - \mu_k)^T$$

And now each sample point is individually propagated using nonlinear dynamics

$$Y^i = f(X^i)$$

Now trying to fit a linear model between the points (Y^i, X^i)

$$Y = AX + B$$

This is standard linear fit procedure by minimizing the least square error

$$e_i = Y^i - AX^i - B$$
$$E = (e_i)^T (e_i)$$

The mean and covariance at time $k+1$ can be found from the Kalman filter propagation equations

$$\mu_{k+1} = A\mu_k$$
$$P_{k+1} = AP_k A^T + Q_k$$

Unscented Kalman Filter- Unscented Transform

•The UKF works in the same ways as the LRKF but the samples are chosen in a **determined** way such that they always match the mean and covariance or higher moments at the current step.

• The points are propagated using the **nonlinear dynamics**. The mean and covariance at time step $k+1$ are calculated from these propagated points.

• The **first approximation** the UKF does to the CKE is that it replaces the current state PDF with a gaussian PDF with same first two moments.

• There is **no linearizations involved**, instead the integral is itself evaluated approximately using quadrature points of the gaussian weighting function-**second approximation**.

• The UKF like the EKF only keeps track of the first two moments of the state PDF.

• We can still improve the **second approximation** by better evaluating the integrals.

One could look at the UKF in the following perspective, starting from CKE

$$P(x_{k+1}) = \int \mathsf{N}(x_{k+1}|x_k)P(x_k)dx_k$$

Replacing the state PDF with gaussian pdf with equivalent mean μ_k and covariance P_k.

$$P(x_{k+1}) = \int \mathsf{N}(x_{k+1}|x_k)\mathsf{N}(x_k)dx_k$$

Calculating the mean of $P(x_{k+1}$ by integrating on both side wrt x_{k+1}.

$$\begin{aligned}\mu_{k+1} &= \int \int \mathsf{N}(x_{k+1} - f(x_k)|Q_k)dx_{k+1}\mathsf{N}(x_k)dx_k \\ &= \int f(x_k)\mathsf{N}(x_k)dx_k\end{aligned}$$

The covariance can be calculated from the second raw moment

$$E[x_{k+1}x_{k+1}^T] = \int\int x_{k+1}x_{k+1}^T \mathsf{N}(x_{k+1} - f(x_k)|Q_k)dx_{k+1}\mathsf{N}(x_k)dx_k$$
$$= \int (f(x_k)f(x_k)^T + Q_k)\mathsf{N}(x_k)dx_k$$

By parallel axis theorem for moments the covariace is calculated as

$$P_k = E[x_{k+1}x_{k+1}^T] - \mu_{k+1}\mu_{k+1}^T$$

•Thus by evaluating two integrals we get the mean and covariance.

• The second raw moment and the mean need to be evaluated accurate enough or else the parallel axis theorem for moments might render the **covariance to be positive semi definite**.

For example in a 1D case, the integrals would be

$$
\begin{aligned}
\mu_{k+1} &= \int f(x_k)\mathsf{N}(x_k)dx_k \\
&= \sum_{i=1}^{N} w^i f(x_k^i) \\
E[x_{k+1}^2] &= \int (f(x_k)^2 + Q_k)\mathsf{N}(x_k)dx_k \\
&= \sum_{i=1}^{N} w^i f(x_k^i)^2 + Q_k \sum_{i=1}^{N} w^i \\
&= \sum_{i=1}^{N} w^i f(x_k^i)^2 + Q_k
\end{aligned}
$$

The covariance is

$$
\begin{aligned}
P_k &= E[x_{k+1}^2] - \mu_{k+1}^2 \\
&= \sum_{i=1}^{N} w^i (f(x_k^i) - \mu_{k+1})^2
\end{aligned}
$$

Thus the UKF boils down to just evaluating the integrals involving gaussian kernel using the quadrature points.

Unscented Transform-Integration using Quadrature points for Gaussian Kernel

- This section shows the derivation of a new set of quadrature points that can exactly match all the first four and even six moments of the gaussian weighting function for any dimension.

- The main **draw back** of the $2n+1$, $4n+1$ or even the $6n+1$ sigma points is that they cannot capture the 4th moment or any higher order even moment exactly. It is illustrated that **we need to search in additional directions** to satisfy higher order moments.

- An example is presented that evidently shows that these methods cannot capture the 4th moment of the gaussian weigting function.

An Example illustrating the fact that the $2n + 1$, $4n+1$ and $6n+1$ set of sigma points cannot satisfy the 4th moment

• All these set of sigma points tend to pick symmetric points on the principle axis of the σ contours.

• Their distances from the mean and the corresponding weights are chosen in a heuristic manner such that they satisfy the covariances.

•The main motive is that we have a gaussian pdf and we want to discretize it such that the discrete pdf can capture the moments of the continuous gaussian pdf.

For example consider a 2D system.The square root of the covariance matrix P is defined as

$$\sqrt{P} = AA^T$$

where A is the matrix square root. Let

$$P = \begin{bmatrix} P_{11} & P_{12} \\ P_{12} & P_{22} \end{bmatrix}$$

if A is considered to be the principle square root of the symmetric matrix P then A is also symmetric.

$$A = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix}$$

$$
\begin{aligned}
P_{11} &= \sigma_{11}^2 + \sigma_{12}^2 \\
P_{12} &= \sigma_{11}\sigma_{12} + \sigma_{12}\sigma_{22} \\
P_{22} &= \sigma_{12}^2 + \sigma_{22}^2
\end{aligned}
$$

The priciple axis of the 1σ contour are columns of A i.e.

$$\sigma_1 = \begin{bmatrix} \sigma_{11} \\ \sigma_{12} \end{bmatrix}$$

and

$$\sigma_2 = \begin{bmatrix} \sigma_{12} \\ \sigma_{22} \end{bmatrix}$$

- Now for example the $4n + 1$ set of sigma points for the 2D case has 1 point on the mean, 4 points on the principle axis of $r_1\sigma$ contour and the rest 4 points on the principle axis of another $r_2\sigma$ contour.

- The mean is considered as the origin as once the points are found they can be translated by the given mean vector.

- The points that lie on the same contour have the same weights. So now there are only 3 weights: w_0 for the mean, w_1 for the 4 $r_1\sigma$ contour points and w_2 for the 4 $r_2\sigma$ contour points.

The points are

$$(0, 0)$$
$$(r_1\sigma_{11}, r_1\sigma_{12})$$
$$(-r_1\sigma_{11}, -r_1\sigma_{12})$$
$$(r_1\sigma_{12}, r_1\sigma_{22})$$
$$(-r_1\sigma_{12}, -r_1\sigma_{22})$$
$$(r_2\sigma_{11}, r_2\sigma_{12})$$
$$(-r_2\sigma_{11}, -r_2\sigma_{12})$$
$$(r_2\sigma_{12}, r_2\sigma_{22})$$
$$(-r_2\sigma_{12}, -r_2\sigma_{22})$$

Visually they look like

4np1pts.jpg

- There are two sets of points on each sigma contour. This is just to illustrate the difference between the points derived from principle square root and cholesky decomposition of P.
- Cholesky decomposition results in points that lie on the ellipse/ contour but not on the principle axis.

•The mean and all **higher order odd moments** are automatically satisfied as the points are chosen to be **symmetric**.

The covariance equations i.e. the second moments are

$$\begin{aligned}
E[x_1^2] &= \sum w_i x_{1i}^2 = (2w_1 r_1^2 + 2w_2 r_2^2)(\sigma_{11}^2 + \sigma_{12}^2) = P_{11} \\
E[x_1 x_2] &= \sum w_i x_{1i} x_{2i} = (2w_1 r_1^2 + 2w_2 r_2^2)(\sigma_{11}\sigma_{12} + \sigma_{12}\sigma_{22}) = P_{12} \\
E[x_2^2] &= \sum w_i x_{2i}^2 = (2w_1 r_1^2 + 2w_2 r_2^2)(\sigma_{12}^2 + \sigma_{22}^2) = P_{22}
\end{aligned}$$

• It can be seen that by choosing points on the sigma contours the 3 equations for a 2D case collapse into just solving one equations i.e.

$$2w_1 r_1^2 + 2w_2 r_2^2 = 1$$

Now looking at the 4th moment.

$$E[x_1^4] = \sum w_i x_{1i}^4 = (2w_1 r_1^4 + 2w_2 r_2^4)(\sigma_{11}^4 + \sigma_{12}^4) \equiv 3P_{11}^2$$

$$E[x_1^3 x_2] = \sum w_i x_{1i}^3 x_{2i}^2 = (2w_1 r_1^4 + 2w_2 r_2^4)(\sigma_{11}^3 \sigma_{12} + \sigma_{12}^3 \sigma_{22}) \equiv 3P_{11}P_{12}$$

$$E[x_1^2 x_2] = \sum w_i x_{1i}^2 x_{2i}^2 = (2w_1 r_1^4 + 2w_2 r_2^4)(\sigma_{11}^2 \sigma_{12}^2 + \sigma_{12}^2 \sigma_{22}^2) \equiv P_{11}P_{22} + 2P_{12}^2$$

$$E[x_1^1 x_2^3] = \sum w_i x_{1i} x_{2i}^3 = (2w_1 r_1^4 + 2w_2 r_2^4)(\sigma_{11} \sigma_{12}^3 + \sigma_{12} \sigma_{22}^3) \equiv 3P_{22}P_{12}$$

$$E[x_2^4] = \sum w_i x_{2i}^4 = (2w_1 r_1^4 + 2w_2 r_2^4)(\sigma_{12}^4 + \sigma_{22}^4) \equiv 3P_{22}^2$$

•Here we have 5 equations which seem to have the same left hand side of variables $2w_1 r_1^4 + 2w_2 r_2^4$ but different right hand side of constant values.

•These are like **parallel curves** that do not intersect and cannot be solved for. It does not matter how many points we take on the principle axis or in general sigma contour formed from $\sqrt{P}$, we always end up in the above structure.

Searching for Quadrature points on principle axis and conjugate axis that can satisfy the 4th moment "'exactly"

Again for a 2D case, the axis are chosen as

Principle axis:

$$\sigma_1 = \begin{bmatrix} \sigma_{11} \\ \sigma_{12} \end{bmatrix}$$

$$\sigma_2 = \begin{bmatrix} \sigma_{12} \\ \sigma_{22} \end{bmatrix}$$

Conjugate axis:

$$\sigma_3 = \sigma_1 + \sigma_2$$
$$\sigma_4 = \sigma_1 - \sigma_2$$

The points are therefore

$$
\begin{aligned}
&(0, 0) \\
&(r_1\sigma_{11}, r_1\sigma_{12}) \\
&(-r_1\sigma_{11}, -r_1\sigma_{12}) \\
&(r_1\sigma_{12}, r_1\sigma_{22}) \\
&(-r_1\sigma_{12}, -r_1\sigma_{22}) \\
&(r_2(\sigma_{11} + \sigma_{12}), r_2(\sigma_{12} + \sigma_{22})) \\
&(-r_2(\sigma_{11} + \sigma_{12}), -r_2(\sigma_{12} + \sigma_{22})) \\
&(r_2(\sigma_{11} - \sigma_{12}), r_2(\sigma_{12} - \sigma_{22})) \\
&(-r_2(\sigma_{11} - \sigma_{12}), -r_2(\sigma_{12} - \sigma_{22}))
\end{aligned}
$$

Visually they look like

eye2quadpts.jpg

P2Dquadpts.jpg

(a)Indentity Cov and (b)$[4, 2; 2, 3]$ Cov

Isserlis Theorem:

Any arbitrary moment of a Gaussian PDF is given by

$$E[x_1 x_2 x_3 .... x_{2n}] = \sum \prod E[x_i x_j]$$

for example the fourth moment is

$$E[x_1 x_2 x_3 x_4] = E[x_1 x_2]E[x_3 x_4] + E[x_1 x_3]E[x_4 x_5] + E[x_1 x_4]E[x_2 x_3]$$

Therefore any order moment higher than 2 is dependent on the entries of the covariance matrix

The 2nd moment equations are

$$
\begin{aligned}
E[x_1^2] = \sum w_i x_{1i}^2 &= (2w_1 r_1^2 + 4w_2 r_2^2)(\sigma_{11}^2 + \sigma_{12}^2) = P_{11} \\
E[x_1 x_2] = \sum w_i x_{1i} x_{2i} &= (2w_1 r_1^2 + 4w_2 r_2^2)(\sigma_{11}\sigma_{12} + \sigma_{12}\sigma_{22}) = P_{12} \\
E[x_2^2] = \sum w_i x_{2i}^2 &= (2w_1 r_1^2 + 4w_2 r_2^2)(\sigma_{12}^2 + \sigma_{22}^2) = P_{22}
\end{aligned}
$$

From these 3 equations we are only left to solve the equation

$$2w_1 r_1^2 + 4w_2 r_2^2 = 1$$

Now if we consider the 5- 4th order moment equations we are only left to solve the 2 equations

$$
\begin{aligned}
2w_1 r_1^4 + 4w_2 r_2^4 &= 3 \\
4w_2 r_2^4 &= 1
\end{aligned}
$$

Now if we apply the same procedure to n-dimensional system we can generalise the set of equations. In summary

$$\begin{aligned}
2w_1r_1^2 + 2^n w_2 r_2^2 &= 1 \\
2w_1r_1^4 + 2^n w_2 r_2^4 &= 3 \\
2^n w_2 r_2^4 &= 1 \\
w_0 = 1 - 2nw_1 - 2^n w_2
\end{aligned}$$

equations have to be solved for w_1, r_1, w_2, r_2. In total there are $2n + 2^n + 1$ points.1 point of weight w_0 on the mean, $2n$ points of weight w_1 lie on the principle axis on the $r1\sqrt{P}$ contour, rest 2^n points lie on the conjugate axis symmetrically.

Results of integration compared to Gauss Hermite integration for 3D system

Here in this section the new set of quadrature points for 3D case are put to test, **compared to gauss hermite integration**. As the new set of points are only **good till fourth moment** we would consider the integration of **polynomials till degree 4**.

$$P = \begin{bmatrix} 4 & 2 & 1 \\ 2 & 9 & 1 \\ 1 & 1 & 16 \end{bmatrix}$$

$$F = x_1^4 + x_2^4 + x_3^4 + x_1^3 x_2 + x_1^2 x_2^2 + \\ + x_3^2 x_2^2 + x_1^2 x_3^2 + x_1^3 x_3 + x_2^3 x_3 + x_3^3 x_2$$

No. of pts	$2^3 = 8$	$3^3 = 27$	$4^3 = 64$	$10^3 = 1000$ (Truth)
GH	1056.95571	1797.9999999	1798.00	1798.00
% error wrt Truth	41.2149211	4.299e-013	3.03502e-013	0

	No. of pts	Integration result	% error wrt Truth
NM	15	1797.99999998858	6.355208041935254e-010

Quadrature points to capture the 6th moment exactly

• To capture the 6th order moments exactly extra points are taken on different set of axis.

• For example consider the 3D case. There are 3 principle axis σ_1, σ_2 and σ_3. Additional axis that are **constructed from these axis** are

The principle axis have 6 points

$$(\sigma_1, \sigma_2, \sigma_3)$$

The conjugate axis have 8 points

$$\begin{aligned}\sigma_1 + \sigma_2 + \sigma_3 \\ \sigma_1 - \sigma_2 + \sigma_3 \\ \sigma_1 + \sigma_2 - \sigma_3 \\ \sigma_1 - \sigma_2 - \sigma_3\end{aligned}$$

The new axis lie in each of the orthogonal planes, pass through the origin and are 45^o to each axis in that plane. There are 12 pts on these axis

$$\sigma_1 + \sigma_2$$
$$\sigma_1 - \sigma_2$$
$$\sigma_1 + \sigma_3$$
$$\sigma_1 - \sigma_3$$
$$\sigma_2 + \sigma_3$$
$$\sigma_2 - \sigma_3$$

In total there are $6+8+12 = 26$ quadrature points to capture all the 6th order and lower moments.

3d6thmomentpts.jpg

2dcorr6thmomentpts.jpg

2d_weird_case2.jpg

For a general n- Dimensional system the equations to be solved a

$$
\begin{aligned}
2r_1^2 w_1 + 2^n r_2^2 w_2 + 4(n-1) r_3^2 w_3 &= 1 \\
2r_1^4 w_1 + 2^n r_2^4 w_2 + 4(n-1) r_3^4 w_3 &= 3 \\
2^n r_2^4 w_2 + 4 r_3^4 w_3 &= 1 \\
2r_1^6 w_1 + 2^n r_2^6 w_2 + 4(n-1) r_3^6 w_3 &= 15 \\
2^n r_2^6 w_2 + 4 r_3^6 w_3 &= 3 \\
2^n r_2^6 w_2 &= 1
\end{aligned}
$$

There are in total $2n + 2^n + 2n(n-1)$ points.

Results of integration compared to Gauss Hermite integration for 4D system

Here in this section the new set of quadrature points to capture the 6th moment also for **4D case** are put to test, compared to **gauss hermite quadrature points**. As the new set of points are only **good till sixth order moment** we would consider the integration of polynomials till **degree 6**.
The covariance of the gaussian Kernel

$$P = \begin{bmatrix} 4 & 1 & 2 & 1 \\ 1 & 9 & 2 & 3 \\ 2 & 2 & 16 & 4 \\ 1 & 3 & 4 & 25 \end{bmatrix}$$

$$F = x_1^6 + x_2^6 + x_3^6 + x_4^6 + x_1^4 + x_2^4 + x_3^4 + x_4^4 + x_1^2 x_2^2 x_3^2 +$$
$$+ x_1^3 x_3 + x_1^2 x_2^4 + x_3^4 x_2^2 + x_1^2 x_3^4 + x_1^3 x_3 + x_2^3 x_3 + x_3^3 x_2$$

No. of pts	$2^4 = 16$	$3^4 = 81$	$4^4 = 256$	$5^4 = 64$	$20^4 = 160000$
GH	84462.6354	293311.8446	375417.9999	375417.9999	375417.9999
% error	77.5017	21.8705	9.0702e-012	8.9152e-012	0

	No. of pts	Integration result	% error wrt Truth
NM	49	3.754206e+005	7.112483626542660e-004