A New Set of Sigma points for a Gaussian PDF that can Capture all the 8th order and lower Moments 'Exactly'

August 1, 2011

Introduction

Consider a discrete dynamic system with noise of zero mean and Q_k covariance

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

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

$$P(x_{k+1}) = \int P(x_{k+1}|x_k)P(x_k)dx_k \tag{2}$$

$$P(x_{k+1}|x_k) = P(x_{k+1} - f(x_k, k)|Q_k) \tag{3}$$

In general the integration invoolved in the CKE is not simple and one has to resort to numerical techniques to evaluate this integral. But under some conditions there is an analytical soltuion to the CKE. When the system is linear, noise is gaussian and initial condition is gaussian the integral results in an analytical solution which is the famous Kalman filter prediction step

$$x_{k+1} = F_k x_k + v_k \tag{4}$$

$$P(x_{k+1}) = \int N(x_{k+1}, F_k x_k | Q_k) N(x_k, \mu_k | P_k) dx_k \tag{5}$$

$$= N(x_{k+1}, \mu_{k+1} | P_k) \tag{6}$$

$$\mu_{k+1} = F_k \mu_k \tag{7}$$

$$P_{k+1} = F_k P_k F_k^T + Q_k \tag{8}$$

In a slightly more general case, we could consider the system to be nonlinear and noise to be gaussian. Assumption of gaussian noise just makes the integration simpler and one could see the results of various approximations involved analytically. And above that the noise in many physical systems can be safely approximated as gaussian.
Incase the system in nonlinear and noise is gaussian, the CKE would be

$$P(x_{k+1}) = \int N(x_{k+1}, f(x_k, k) | x_k) P(x_k) dx_k \tag{9}$$

The rest of the report illustrates some of the approximate filters that essentially try to solve the integral in the CKE with the help of some approximations. In the final section a new method is proposed that is in line with the Unscented Transform. One could look at it as an upgrade to the UT or as a cubature method

to evalute multidimensional integrals with gaussian weighting function. The motivation for developing this method is the sheer reduction in the number of quadrature points required for the integration when compared to the traditional multidimensional gauss hermite points.

Extended Kalman Filter-EKF

As the state PDF $P(x_k)$ is not always gaussian there is no analytical solution to the CKE equation. Hence the EKF emerged that can provide us with an approximate but analytical solution to this CKE by considering the following approximations

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

- The dynamics of the system $f(x_k, k)$ is linearized .

The EKF works well for systems in which the linearized dynamics is a good approximation to the nonlinear system. If the nonlinearity is too strong the EKF would diverge. Above that during the linearization process the computation of the jacobian is computationally expensive. As a result the EKF complemetely disregards the actual state PDF and propagates only the first two moments of the PDF using the linearized dynamics.

$$\mu_{k+1} = f(\mu_k) \tag{10}$$

$$P_{k+1} = AP_k A^T + Q_k \tag{11}$$

$$A = \frac{\partial f}{\partial x}|_{\mu_k} \tag{12}$$

Where A is the jacobian of the system. For a system with many states or for a physical system the jacobian might not be analytically available or the computation of it might be time consuming. Computing the jacobian numerically is usually prone to error which would further undermine the process. To overcome the trouble of evaluating the jacobian required for the linearization process, a different approach for linearization is considered in the next method.

Linear Regression Kalman Filter - LRKF

The LRKF also uses linearized dynamics but the linearization is not done using the jacobian but by a method called statistical linearization. Firstly sample points are chosen about the current mean at time k such that the mean and covariance of the samples match the current mean and current covariance. Each point is propagated using the exact 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. This linearized dynamics of the system at time k are used to compute the mean and covariance at time step $k + 1$. The sample points X^i at time k are such that

$$\mu_k = \frac{1}{n} \sum_{i=1}^{N} X^i \tag{13}$$

$$P_k = \frac{1}{n} \sum_{i=1}^{N} (X^i - \mu_k)(X^i - \mu_k)^T \tag{14}$$

Each sample point is individually propagated using nonlinear dynamics

$$Y^i = f(X^i) \tag{15}$$

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

$$Y = AX + B \tag{16}$$

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

$$e_i = Y^i - AX^i - B \tag{17}$$

$$\frac{Min}{A, B} E = (e_i)^T (e_i) \tag{18}$$

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

$$\mu_{k+1} = A\mu_k \tag{19}$$
$$P_{k+1} = AP_k A^T + Q_k \tag{20}$$

This method again suffers the problems involved while the linearized dynamics are used. To perform the statistical linearization one would need a lot of samples to match the mean and covariance. In higher dimensions the number of points required could be very large.

Unscented Kalman Filter- UKF

The UKF works in the same ways as the LRKF but the samples are chosen in a deterministic way such that they always match the mean and covariance 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 only 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 sigma points of the gaussian weighting function. The UKF tends to perform better than EKF as it does not approximate the dynamics. But still the UKF only propagates the first two moments of the state PDF. One could look at the UKF in the following perspective, starting from CKE

$$P(x_{k+1}) = \int N(x_{k+1} - f(x_k, k), 0|x_k)P(x_k)dx_k \tag{21}$$

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

$$P(x_{k+1}) = \int N(x_{k+1} - f(x_k, k), 0|x_k)N(x_k)dx_k \tag{22}$$

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

$$\mu_{k+1} = \int \int x_{k+1}N(x_{k+1} - f(x_k), 0|Q_k)dx_{k+1}N(x_k)dx_k \tag{23}$$

$$= \int f(x_k)N(x_k)dx_k \tag{24}$$

The covariance can be calculated from the second raw moment

$$E[x_{k+1}x_{k+1}^T] = \int \int x_{k+1}x_{k+1}^T N(x_{k+1} - f(x_k)|Q_k)dx_{k+1}N(x_k)dx_k \tag{25}$$

$$= \int (f(x_k)f(x_k)^T + Q_k)N(x_k)dx_k \tag{26}$$

By parallel axis theorem for moments the covariace is calculated as

$$P_k = E[x_{k+1}x_{k+1}^T] - \mu_{k+1}\mu_{k+1}^T \tag{27}$$

Thus by evaluating two integrals in equations (24) and (26) 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 or may be even negative definite. The integrals are evaluated using sigma points or in the UKF case the sigma points. For example in a 1D case, the integrals would be

$$\mu_{k+1} = \int f(x_k)N(x_k)dx_k \tag{28}$$

$$= \sum_{i=1}^{N} w^i f(x_k^i) \tag{29}$$

$$E[x_{k+1}^2] = \int (f(x_k)^2 + Q_k)N(x_k)dx_k \tag{30}$$

$$= \sum_{i=1}^{N} w^i f(x_k^i)^2 + Q_k \sum_{i=1}^{N} w^i \tag{31}$$

$$= \sum_{i=1}^{N} w^i f(x_k^i)^2 + Q_k \tag{32}$$

The covariance is

$$P_{k+1} = E[x_{k+1}^2] - \mu_{k+1}^2 \tag{33}$$

$$= \sum_{i=1}^{N} w^i f(x_k^i)^2 + Q_k - \mu_{k+1}^2 \tag{34}$$

After some some algebraic rearrangement one could write the covariance equation (34) as

$$P_{k+1} = \sum_{i=1}^{N} w^i (f(x_k^i) - \mu_{k+1})^2 + Q_k \tag{35}$$

In the UKF perspective, evaluating the mean in equation (29) and covariance in equation (35) is like a monte carlo method but with deterministic samples. One could see that the underlying process is actually the integration process using quadrature points of a gaussian function. The integral can be evaluated by monte carlo method using random samples. In higher dimensions the number of samples required is very large. The Unscented Transform-UT does a very good job of reducing the number of samples while maintaining a good accuracy in the integral evaluation. The UT suggests the following $2n + 1$ sigma points and the

4

corresponding weights

$$X_0 = \mu \qquad\qquad W_0 = \frac{\kappa}{(n+\kappa)} \qquad\qquad (36)$$

$$X_i = \mu + (sqrt(n+\kappa)P)_i \qquad\qquad W_i = \frac{1}{2(n+\kappa)} \qquad\qquad (37)$$

$$X_{i+n} = \mu - (sqrt(n+\kappa)P)_i \qquad\qquad W_{i+n} = \frac{1}{2(n+\kappa)} \qquad\qquad (38)$$

The author suggests that for a gaussian distribution the κ value should be selected such that $n+\kappa = 3$. The $4n+1$ set of sigma points is a subset formed from neglecting the last two set of points of the $6n+1$ set of points. The sigma points and corresponding weights of $6n+1$ system are

$$X_0 = \mu \qquad\qquad W_0 = 1 - \frac{n(1+\alpha+\beta)}{1+4\alpha+9\beta}$$

$$X_i = \mu + (sqrtP)_i \qquad\qquad W_i = \frac{1}{2+8\alpha+18\beta}$$

$$X_{i+n} = \mu - (sqrtP)_i \qquad\qquad W_{i+n} = \frac{1}{2+8\alpha+18\beta}$$

$$X_{i+2n} = \mu + 2(sqrtP)_i \qquad\qquad W_{i+2n} = \frac{\alpha}{2+8\alpha+18\beta}$$

$$X_{i+3n} = \mu - 2(sqrtP)_i \qquad\qquad W_{i+3n} = \frac{\alpha}{2+8\alpha+18\beta}$$

$$X_{i+4n} = \mu + 3(sqrtP)_i \qquad\qquad W_{i+4n} = \frac{\beta}{2+8\alpha+18\beta}$$

$$X_{i+5n} = \mu - 3(sqrtP)_i \qquad\qquad W_{i+5n} = \frac{\beta}{2+8\alpha+18\beta}$$

the author suggests that for a gaussian distribution the values of α and β are set to be

$$\alpha = \frac{normpdf(2)}{normpdf(1)} \qquad\qquad (39)$$

$$\beta = \frac{normpdf(3)}{normpdf(1)} \qquad\qquad (40)$$

Unscented Transform-Integration using Sigma points for Gaussian Kernel

These Unscented Transform methods claim that they capture the second moment and all higher order odd moments. This is infact true because the points are chosen to be symmetric and thus odd moments are automatically satisfied. The prediction of mean and covariance is accurate till the second term in the taylor series expansion. This means that if the dynamics of our system is of polynomial nonlinearity of degree 2 or less, one could get accurate results of mean and covariance by these unscented transform methods. There is no gaurantee for convergence of the integral value for mean and covariance in equations (29) and (35) when the dynamics is a polynomial with degree greater than 2 or if it was a nonlinear function other than of polynomial type. 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. An example is presented here, evidently showing that these methods cannot capture the 4th moment and thus higher order moments of the gaussian weighting function.

A general 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 points on the principle axis of the σ contours. Points are chosen on these σ contours such that they are symmetric to the mean and hence all odd moments including the mean are captured. 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

$$P = AA \tag{41}$$

where A is the matrix square root of P. 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.

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

then

$$P_{11} = \sigma_{11}^2 + \sigma_{12}^2 \tag{42}$$
$$P_{12} = \sigma_{11}\sigma_{12} + \sigma_{12}\sigma_{22} \tag{43}$$
$$P_{22} = \sigma_{12}^2 + \sigma_{22}^2 \tag{44}$$

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 one σ contour and the rest 4 points on the principle axis of another σ contour. Lets consider the contours to be $r_1\sigma$ and $r_2\sigma$ contours. 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 (x_{1i}, x_{2i}) are

$$(0,0)$$
$$(r_1\sigma_{11}, r_1\sigma_{12})$$
$$(-r_1\sigma_{11}, -r_1\sigma_{12})$$
$$(r_1\sigma_{12}, r_1\sigma_{22})$$
$$(-r_1\sigma_{12}, -r_1\sigma_{22})$$
$$(r_2\sigma_{11}, r_2\sigma_{12})$$
$$(-r_2\sigma_{11}, -r_2\sigma_{12})$$
$$(r_2\sigma_{12}, r_2\sigma_{22})$$
$$(-r_2\sigma_{12}, -r_2\sigma_{22}) \tag{45}$$

Visually they look like
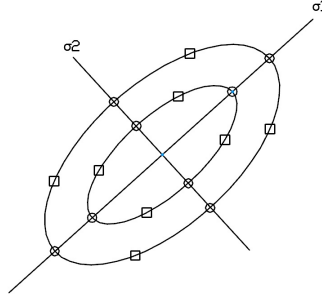

Figure 1: 4n+1 points on sigma contours

There are two sets of points on each sigma contour. This is just to illustrate the diffenrene 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 points derived from the principle square root lie on the axis of the ellipse and the contour. 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 computed from the set of points (45) are

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

Where we have used equations (42)-(44).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.

$$2w_1 r_1^2 + 2w_2 r_2^2 = 1 \tag{46}$$

Now looking at the 4th moment.

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

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

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

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

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

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. But one may make a very nice inference that by choosing points on the sigma contours we can exactly match the mean and covariance. This gives us a clue that we need to search in some other direction in addition to principle axis for these sigma points.

Searching for Sigma points on principle axis and "conjugate axis" that can satisfy the 4th moment "'exactly"
The term conjugate axis is just a name we chose to designate a particular axis of the ellipsoid and should not be confused with its meaning elsewhere. 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:

$$\sigma_3 = \sigma_1 + \sigma_2 \tag{47}$$

$$\sigma_4 = \sigma_1 - \sigma_2 \tag{48}$$

The points (x_{1i}, x_{2i}) are therefore

$$(0,0)$$
$$(r_1\sigma_{11}, r_1\sigma_{12})$$
$$(-r_1\sigma_{11}, -r_1\sigma_{12})$$
$$(r_1\sigma_{12}, r_1\sigma_{22})$$
$$(-r_1\sigma_{12}, -r_1\sigma_{22})$$
$$(r_2(\sigma_{11}+\sigma_{12}), r_2(\sigma_{12}+\sigma_{22}))$$
$$(-r_2(\sigma_{11}+\sigma_{12}), -r_2(\sigma_{12}+\sigma_{22}))$$
$$(r_2(\sigma_{11}-\sigma_{12}), r_2(\sigma_{12}-\sigma_{22}))$$
$$(-r_2(\sigma_{11}-\sigma_{12}), -r_2(\sigma_{12}-\sigma_{22})) \tag{49}$$
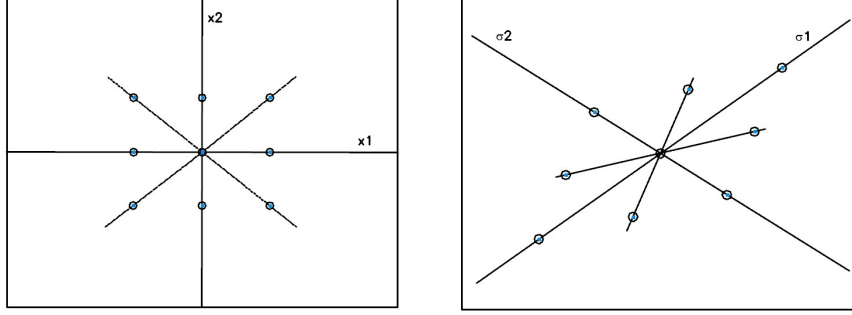
Visually they look like

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

Any arbitrary moment of a Gaussian PDF is given by **Isserlis Theorem**:

$$E[x_1 x_2 x_3 .... x_{2n}] = \sum \prod E[x_i x_j] \tag{50}$$

for example the fourth moment is

$$E[x_1 x_2 x_3 x_4] = E[x_1 x_2]E[x_3 x_4] + E[x_1 x_3]E[x_4 x_5] + E[x_1 x_4]E[x_2 x_3] \tag{51}$$

And the sixth moment is

$$
\begin{aligned}
E[x_1 x_2 x_3 x_4 x_5 x_6] = {} & E[x_1 x_2]E[x_3 x_4]E[x_5 x_6] + E[x_1 x_2]E[x_3 x_5]E[x_4 x_6] + E[x_1 x_2]E[x_3 x_6]E[x_4 x_5] \\
& + E[x_1 x_3]E[x_2 x_4]E[x_5 x_6] + E[x_1 x_3]E[x_2 x_5]E[x_4 x_6] + E[x_1 x_3]E[x_2 x_6]E[x_4 x_5] \\
& + E[x_1 x_4]E[x_2 x_3]E[x_5 x_6] + E[x_1 x_4]E[x_2 x_5]E[x_3 x_6] + E[x_1 x_4]E[x_2 x_6]E[x_3 x_5] \\
& + E[x_1 x_5]E[x_2 x_3]E[x_4 x_6] + E[x_1 x_5]E[x_2 x_4]E[x_3 x_6] + E[x_1 x_5]E[x_2 x_6]E[x_3 x_4] \\
& + E[x_1 x_6]E[x_2 x_3]E[x_4 x_5] + E[x_1 x_6]E[x_2 x_4]E[x_3 x_5] + E[x_1 x_6]E[x_2 x_5]E[x_3 x_4]
\end{aligned}
$$

The 2nd moment equations using the points set (49) are

$$E[x_1^2] = \sum w_i x_{1i}^2 = (2w_1 r_1^2 + 4w_2 r_2^2)(\sigma_{11}^2 + \sigma_{12}^2) = (2w_1 r_1^2 + 4w_2 r_2^2)P_{11} \tag{52}$$

$$E[x_1 x_2] = \sum w_i x_{1i} x_{2i} = (2w_1 r_1^2 + 4w_2 r_2^2)(\sigma_{11}\sigma_{12} + \sigma_{12}\sigma_{22}) = (2w_1 r_1^2 + 4w_2 r_2^2)P_{12} \tag{53}$$

$$E[x_2^2] = \sum w_i x_{2i}^2 = (2w_1 r_1^2 + 4w_2 r_2^2)(\sigma_{12}^2 + \sigma_{22}^2) = (2w_1 r_1^2 + 4w_2 r_2^2)P_{22} \tag{54}$$

Again by using equations (42)-(44),from these 3 equations we are only left to solve the equation

$$2w_1 r_1^2 + 4w_2 r_2^2 = 1 \tag{55}$$

Now if we consider the 5 - 4th order moments for the 2D case:

for continuous system

$$E[x_1^4] = 3P_{11}^2 = 3(\sigma_{11}^4 + \sigma_{12}^4) + 6(\sigma_{11}^2 \sigma_{12}^2)$$

$$E[x_1^3 x_2] = 3P_{11}P_{12} = 3(\sigma_{11}^3 \sigma_{12} + \sigma_{12}^3 \sigma_{22}) + 3(\sigma_{11}\sigma_{12}^3 + \sigma_{11}^2 \sigma_{12}\sigma_{22})$$

$$E[x_1^2 x_2^2] = P_{11}P_{22} + 2P_{12}^2 = 3(\sigma_{11}^2 \sigma_{12}^2 + \sigma_{12}^2 \sigma_{22}^2) + (\sigma_{12}^4 + 4\sigma_{11}\sigma_{12}^2 \sigma_{22} + \sigma_{11}^2 \sigma_{22}^2)$$

$$E[x_1^1 x_2^3] = 3P_{22}P_{12} = 3(\sigma_{11}\sigma_{12}^3 + \sigma_{12}\sigma_{22}^3) + 3(\sigma_{12}^3 \sigma_{22} + \sigma_{11}\sigma_{12}\sigma_{22}^2)$$

$$E[x_2^4] = 3P_{22}^2 = 3(\sigma_{12}^4 + \sigma_{22}^4) + 6(\sigma_{12}^2 \sigma_{22}^2)$$

For the equivalent discrete system the 4th order moments are

$$E[x_1^4] = \sum w_i x_{1i}^4 = (\sigma_{11}^4 + \sigma_{12}^4)(2r_1^4 w_1 + 4r_2^4 w_2) + 24w_2 r_2^4 (\sigma_{11}^2 \sigma_{12}^2)$$

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

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

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

$$E[x_2^4] = \sum w_i x_{2i}^4 = (\sigma_{12}^4 + \sigma_{22}^4)(2r_1^4 w_1 + 4r_2^4 w_2) + 24w_2 r_2^4 (\sigma_{12}^2 \sigma_{22}^2)$$

On comparing the set of equations for the continuous and discrete case, we are only left to solve the 2 equations

$$2w_1 r_1^4 + 4w_2 r_2^4 = 3 \tag{56}$$

$$4w_2 r_2^4 = 1 \tag{57}$$

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

$$2w_1 r_1^2 + 2^n w_2 r_2^2 = 1 \tag{58}$$

$$2w_1 r_1^4 + 2^n w_2 r_2^4 = 3 \tag{59}$$

$$2^n w_2 r_2^4 = 1 \tag{60}$$

The central weight at the mean is found by

$$w_0 = 1 - 2nw_1 - 2^n w_2 \tag{61}$$

These 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 sigma 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.

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

$$X = [x_1, x_2, x_3] \tag{62}$$
$$F(X) = x_1^4 + x_2^4 + x_3^4 + x_1^3 x_2 + x_1^2 x_2^2 + \tag{63}$$
$$+ x_3^2 x_2^2 + x_1^2 x_3^2 + x_1^3 x_3 + x_2^3 x_3 + x_3^3 x_2 \tag{64}$$

$$f = \int F(X) N(X, 0|P) dX \tag{65}$$

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

Transformation of an arbitrary Jointly distributed Gaussian PDF into an i.i.d Gaussian PDF with mean zero

Any jointly distributed gaussian PDF can be transformed into a Gaussian PDF of zero mean and identity covariance of same dimension. This way one could find the sigma points in the i.i.d space and then transform each point into the original jointly distributed space. Firstly translate the jointly distributed pdf so that the mean coincides with origin. Then the transformation or here a scaled rotation is performed. The inverse of covariance matrix is expanded using a singular value decomposition.

$$x^T P^{-1} x = x^T U^T \Sigma U x = x^T U^T \sqrt{\Sigma} \sqrt{\Sigma} U x \tag{66}$$

Let

$$y = \sqrt{\Sigma} U x \tag{67}$$

then

$$x^T P^{-1} x = y^T y \tag{68}$$

Thus by the transformation $y = \sqrt{\Sigma} U x$ we have the two Gaussian PDFs equivalent

$$N(x, 0|P) \equiv N(y, 0|I) \tag{69}$$

It is much easier to work with the i.i.d. gaussian variables. All the moments are preserved under this transformation. The principle axes just become

$$\sigma_i = I_i \tag{70}$$

where I is the identity matrix of appropriate dimension and I_i is the ith column of the I.

Sigma 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.Considering the mean to be the origin additional axis are constructed from these axis.

The 3 principle axis have 6 points. Each point is scaled with factor r_1 along the principle axis and has weight w_1

$$\pm\sigma_1$$
$$\pm\sigma_2$$
$$\pm\sigma_3 \tag{71}$$

The conjugate axis have 8 points. Each point on this axis is scaled by r_2 and has weight w_2

$$\pm\sigma_1 \pm \sigma_2 \pm \sigma_3 \tag{72}$$

A new set of axis are formed that lie in each of the orthogonal planes, they pass through the origin and are 45° to each principle axis in that plane. The points are scaled by r_3 and have weight w_3. There are 12 pts on these axis

$$\pm\sigma_1 \pm \sigma_2$$
$$\pm\sigma_1 \pm \sigma_3$$
$$\pm\sigma_2 \pm \sigma_3 \tag{73}$$

In total there are $6 + 8 + 12 = 26$ sigma points to capture all the 6th order and lower moments for this 3D case. To intuitively realize the complete distribution in Figure (3) for 3D system is difficult so the points only in one of the eight quadrants are shown in Figure (4). For 2D system there is more that one way where we can easily rearrange the points to capture the 6th moment. These two configurations are shown in figure (5).

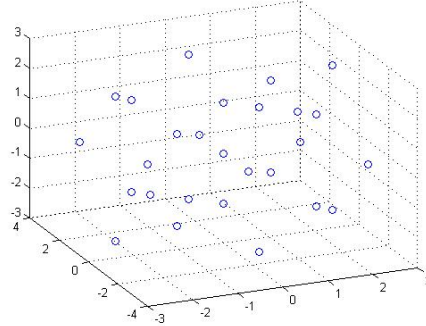For a general n- Dimensional system the equations to be solved a

Figure 3: Identity Cov 3D pts

$$2r_1^2 w_1 + 2^n r_2^2 w_2 + 4(n-1)r_3^2 w_3 = 1$$
$$2r_1^4 w_1 + 2^n r_2^4 w_2 + 4(n-1)r_3^4 w_3 = 3$$
$$2^n r_2^4 w_2 + 4r_3^4 w_3 = 1$$
$$2r_1^6 w_1 + 2^n r_2^6 w_2 + 4(n-1)r_3^6 w_3 = 15$$
$$2^n r_2^6 w_2 + 4r_3^6 w_3 = 3$$
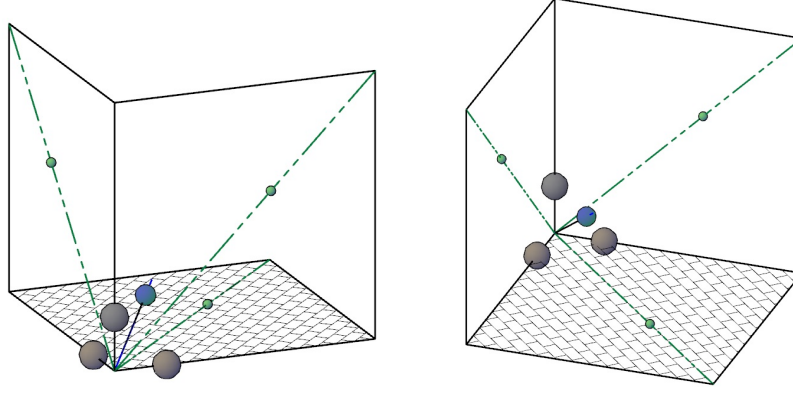$$2^n r_2^6 w_2 = 1$$

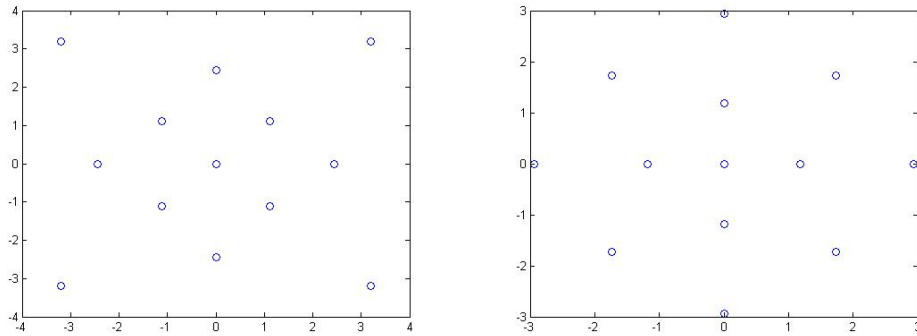Figure 4: Two perspective views of points in 1st quadrant of 3D space

Figure 5: 2 different configurations for 2D system

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

Here in this section the new set of sigma points to capture the 6th moment also for 4D case are put to test, compared to gauss hermite integration. 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

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

13

$$X = [x_1, x_2, x_3, x_4] \tag{74}$$

$$F(X) = x_1^6 + x_2^6 + x_3^6 + x_4^6 + x_1^4 + x_2^4 + x_3^4 + x_4^4 + x_1^2 x_2^2 x_3^2 +$$
$$+ x_1^3 x_3 + x_1^2 x_2^4 + x_3^4 x_2^2 + x_1^2 x_3^4 + x_1^3 x_3 + x_2^3 x_3 + x_3^3 x_2 \tag{75}$$

Evaluating the integral

$$f = \int F(X) N(X, 0|P) dX \tag{76}$$

No. of pts	$2^4 = 16$	$3^4 = 81$	$4^4 = 256$	$5^4 = 64$	$20^4 = 160000$ (Truth)
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

Sigma points to capture the 8th moment exactly To capture the 8th order moments exactly additional axis have to be considered. For example consider the 4D case. There are 4 principle axis σ_1, σ_2, σ_3 and σ_4. Additional axis constructed from these axis in the following way

The principle axis have 8 points with distance scaling factor r_1 and weight w_1

$$\sigma_1$$
$$\sigma_2$$
$$\sigma_3$$
$$\sigma_4 \tag{77}$$

The 2nd set of axis are formed from considering all combinations of the principle axis. The disnce of each point is scaled by r_2 and each has weight w_2.

$$\sigma_1 \pm \sigma_2 \pm \sigma_3 \pm \sigma_4 \tag{78}$$

The 3rd set of axis are constructed from considering 3 principle axis at a time. The corresponding distance scaling factor and weight are r_5 an w_5

$$\sigma_1 \pm \sigma_2 \pm \sigma_3$$
$$\sigma_1 \pm \sigma_2 \pm \sigma_4$$
$$\sigma_1 \pm \sigma_3 \pm \sigma_4$$
$$\sigma_2 \pm \sigma_3 \pm \sigma_4 \tag{79}$$

The 4th set of axis are constructed from considering 2 principle axis at a time. The corresponding distance scaling factor and weight are r_3 an w_3

$$\sigma_1 \pm \sigma_2$$
$$\sigma_1 \pm \sigma_3$$
$$\sigma_1 \pm \sigma_4$$
$$\sigma_2 \pm \sigma_3$$
$$\sigma_2 \pm \sigma_4$$
$$\sigma_3 \pm \sigma_4$$

The 5th set of axis are fromed from considering all combinations of principle axis with one axis in each combination scaled with a parameter. The corresponding distance scaling factor and weight are r_6 an w_6

$$h\sigma_1 \pm \sigma_2 \pm \sigma_3 \pm \sigma_4$$
$$\sigma_1 \pm h\sigma_2 \pm \sigma_3 \pm \sigma_4$$
$$\sigma_1 \pm \sigma_2 \pm h\sigma_3 \pm \sigma_4$$
$$\sigma_1 \pm \sigma_2 \pm \sigma_3 \pm h\sigma_4$$

where h is the scaling parameter. When the scaling paramter $h = 1$ then these set of axis match the 2nd set of axis . As h slowly changes from 1 the axis start to diverge away from the 2nd set of axis, but they are still arranged in space such that they are symmetrical. This additional set of axis can be visuallised for the 2D system in figure (6) and for 3D system in figure (7). Only one of the 8 quadrants of the 3D system is shown in different perspectives. All the quadrants are similar and symmetric abount mean.
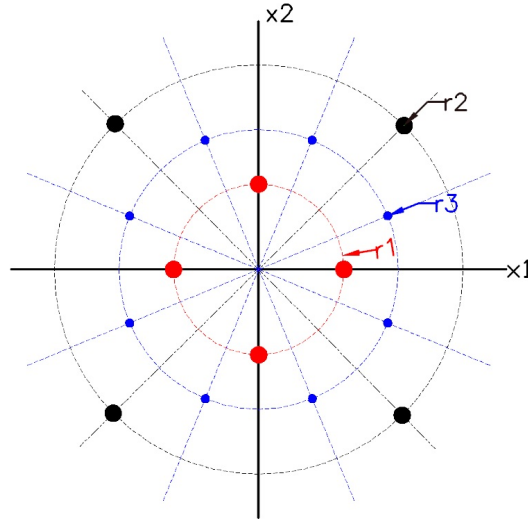

Figure 6: 2D system with points to capture 8th moment

This set of axis are very crucial to capture the various moments within the $8th$ set of moment. All the set of 8th moments are captured only when the points are situated in space, planar points cannot capture all the 8th moments. There are a total of 11 distinct equations in all from 2nd to 8th moment for sytems of dimension 4 or hihger. We thus need 11 variables to solve for. But by considering only the 5 set of axis above there will be only 10 variables in all $r_1, w_1, r_2, w_2, r_3, w_3, r_4, w_4, r_5, w_5$. Another set of points r_4 and weight w_4 are taken on the $2^n d$ set of axis : 12 variables and 11 equations to solve. We are free to choose the scaling parameter h and any one variable, or we could frame an optimization problem to minimize the
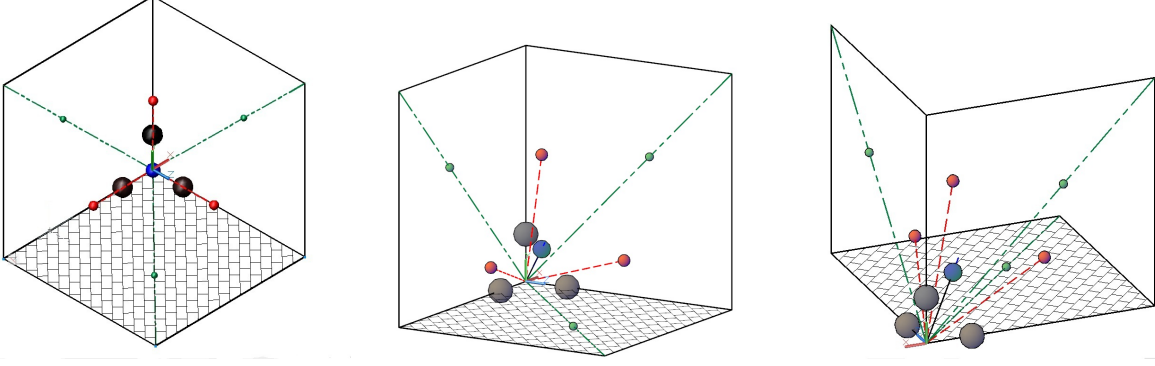
Figure 7: 1st quadrant of 3D system showing the points on axes

error in the 10th moment. The 11 equations to solve are

$$2r_1^2 w_1 + 16r_2^2 w_2 + 12r_3^2 w_3 + 16r_4^2 w_4 + 24r_5^2 w_5 + 48r_6^2 w_6 + 16h^2 r_6^2 w_6 = 1$$

$$2r_1^4 w_1 + 16r_2^4 w_2 + 12r_3^4 w_3 + 16r_4^4 w_4 + 24r_5^4 w_5 + 48r_6^4 w_6 + 16h^4 r_6^4 w_6 = 3$$

$$16r_2^4 w_2 + 4r_3^4 w_3 + 16r_4^4 w_4 + 16r_5^4 w_5 + 32r_6^4 w_6 + 32h^2 r_6^4 w_6 = 1$$

$$2r_1^6 w_1 + 16r_2^6 w_2 + 12r_3^6 w_3 + 16r_4^6 w_4 + 24r_5^6 w_5 + 48r_6^6 w_6 + 16h^6 r_6^6 w_6 = 15$$

$$16r_2^6 w_2 + 4r_3^6 w_3 + 16r_4^6 w_4 + 16r_5^6 w_5 + 32r_6^6 w_6 + 16h^2 r_6^6 w_6 + 16h^4 r_6^6 w_6 = 3$$

$$16r_2^6 w_2 + 16r_4^6 w_4 + 8r_5^6 w_5 + 16r_6^6 w_6 + 48h^2 r_6^6 w_6 = 1$$

$$2r_1^8 w_1 + 16r_2^8 w_2 + 12r_3^8 w_3 + 16r_4^8 w_4 + 24r_5^8 w_5 + 48r_6^8 w_6 + 16h^8 r_6^8 w_6 = 105$$

$$16r_2^8 w_2 + 4r_3^8 w_3 + 16r_4^8 w_4 + 16r_5^8 w_5 + 32r_6^8 w_6 + 16h^2 r_6^8 w_6 + 16h^6 r_6^8 w_6 = 15$$

$$16r_2^8 w_2 + 4r_3^8 w_3 + 16r_4^8 w_4 + 16r_5^8 w_5 + 32r_6^8 w_6 + 32h^4 r_6^8 w_6 = 9$$

$$16r_2^8 w_2 + 16r_4^8 w_4 + 8r_5^8 w_5 + 16r_6^8 w_6 + 32h^2 r_6^8 w_6 + 16h^4 r_6^8 w_6 = 3$$

$$16r_2^8 w_2 + 16r_4^8 w_4 + 64h^2 r_6^8 w_6 = 1$$

And similarly for a 5-D system the equations are

$$2r_1^2 w_1 + 32r_2^2 w_2 + 16r_3^2 w_3 + 32r_4^2 w_4 + 48r_5^2 w_5 + 128r_6^2 w_6 + 32h^2 r_6^2 w_6 = 1$$

$$2r_1^4 w_1 + 32r_2^4 w_2 + 16r_3^4 w_3 + 32r_4^4 w_4 + 48r_5^4 w_5 + 128r_6^4 w_6 + 32h^4 r_6^4 w_6 = 3$$

$$32r_2^4 w_2 + 4r_3^4 w_3 + 32r_4^4 w_4 + 24r_5^4 w_5 + 96r_6^4 w_6 + 64h^2 r_6^4 w_6 = 1$$

$$2r_1^6 w_1 + 32r_2^6 w_2 + 16r_3^6 w_3 + 32r_4^6 w_4 + 48r_5^6 w_5 + 128r_6^6 w_6 + 32h^6 r_6^6 w_6 = 15$$

$$32r_2^6 w_2 + 4r_3^6 w_3 + 32r_4^6 w_4 + 24r_5^6 w_5 + 96r_6^6 w_6 + 32h^2 r_6^6 w_6 + 32h^4 r_6^6 w_6 = 3$$

$$32r_2^6 w_2 + 32r_4^6 w_4 + 8r_5^6 w_5 + 64r_6^6 w_6 + 96h^2 r_6^6 w_6 = 1$$

$$2r_1^8 w_1 + 32r_2^8 w_2 + 16r_3^8 w_3 + 32r_4^8 w_4 + 48r_5^8 w_5 + 128r_6^8 w_6 + 32h^8 r_6^8 w_6 = 105$$

$$32r_2^8 w_2 + 4r_3^8 w_3 + 32r_4^8 w_4 + 24r_5^8 w_5 + 96r_6^8 w_6 + 32h^2 r_6^8 w_6 + 32h^6 r_6^8 w_6 = 15$$

$$32r_2^8 w_2 + 4r_3^8 w_3 + 32r_4^8 w_4 + 24r_5^8 w_5 + 96r_6^8 w_6 + 64h^4 r_6^8 w_6 = 9$$

$$32r_2^8 w_2 + 32r_4^8 w_4 + 8r_5^8 w_5 + 64r_6^8 w_6 + 64h^2 r_6^8 w_6 + 32h^4 r_6^8 w_6 = 3$$

$$32r_2^8 w_2 + 32r_4^8 w_4 + 32r_6^8 w_6 + 128h^2 r_6^8 w_6 = 1$$

For a N-D system the total number of points required to capture the 8th moment by this scheme are $1 + 2n^2 + \frac{4n(n-1)(n-2)}{3} + (n+2)2^n$. A simple process of solving these equations is to first consider 6 equations preferably the last 6 of the 11 equations and to solve for the weights $w_1, w_2, w_3, w_4, w_5, w_6$ symbolically using Mathematica. Substitute these 6 weight equations into the remaining 5 equations. These 5 equations are now in terms of the distances $r_1, r_2, r_3, r_4, r_5, r_6$. One of the distance could be assumed a suitable value and the rest of the 5 equations can be solved for the remaining 5 varibles. An instance of the solution for the 5D system followed by this procedure, where r_5 and h are assumed, is

$$h = 3$$
$$r_5 = 2$$
$$r_1 = 2.3143708172807447$$
$$r_2 = 0.8390942773980102$$
$$r_3 = 1.8307521253266494$$
$$r_4 = 1.3970397430644959$$
$$r_6 = 1.1134786327367021$$
$$w_1 = 0.010529034221546607$$
$$w_2 = 0.015144019639537572$$
$$w_3 = 0.0052828996967816825$$
$$w_4 = 0.0010671298950159158$$
$$w_5 = 0.00065104166666666666$$
$$w_6 = 0.00013776017592074394 \tag{80}$$

Results of integration compared to Gauss Hermite integration for 5D system

The new set of sigma points are evaluated to capture the 8th moment in addition to all previous moments and all higher order odd moments. For 5-D case, the points are created along their respective axis with the distances and weights given in solution (80). These points are used to evaluate the integral and this result is compared to gauss hermite quadrature integration. As the new set of points are only good till 8th order moments we would consider the integration of ap olynomial till degree 8.
The covariance of the gaussian Kernel

$$P = 1000 I_{5x5} \tag{81}$$

$$X = [x_1, x_2, x_3, x_4, x_5] \tag{82}$$

$$F(X) = x_1^8 + x_2^8 + x_3^6 + x_4^6 + x_1^4 x_5^4 + x_2^4 x_3^4 + x_4^2 + x_1^2 x_2^2 x_3^2 x_4^2 +$$
$$+ x_1^3 x_3 + x_1^4 x_2^2 x_5^2 + x_3^4 x_2^2 + x_1^2 x_3^4 + x_1^3 x_3 + x_2^3 x_3 + x_3^3 x_2 \tag{83}$$

Evaluating the integral

$$f = \int F(X) N(X, 0|P) dX \tag{84}$$

No. of pts	$2^5 = 32$	$3^5 = 243$	$4^5 = 1024$	$5^5 = 3125$	$15^5 = 759375$(Truth)
GH	6.0962801e+012	7.6616634e+013	1.84964e+014	2.329646e+014	2.3296463e+014
% error	97.38317	67.11233	20.6039857	5.184531e-011	0

	No. of pts	Integration result	% error wrt Truth
NM	355	2.3296463e+014	5.1509964e-011

The multi dimensional Gauss Hermite points amount to 3125 to capture the 8th moment and lower order moments . It is evident that by the new set of sigma points we need only 355 points to attain the same order of accuracy as that of traditional gauss hermite quadrature with 3125 points, while integrating polynomials of degree 8 or lower.