Generalizing the Unscented Ensemble Transform to Higher Moments

Deanna Easley

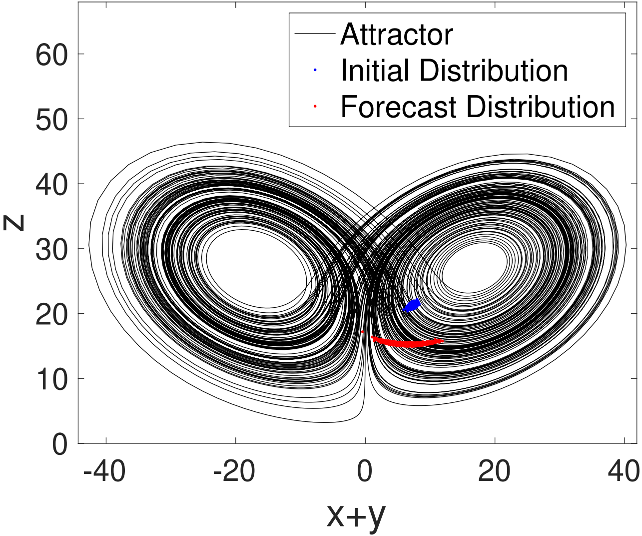
George Mason University

August 14, 2020

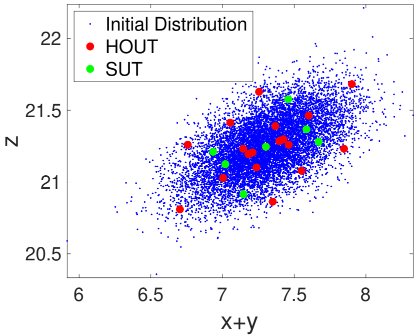
Outline

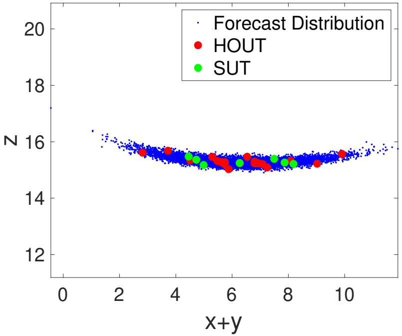
Outline

Motivation


Attractor
Initial Distribution
Forecast Distribution
60
50
40
30
20
10
0
z
-40
-20
0
20
40
x+y

Motivation


Initial Distribution
HOUT
SUT
22
21.5
21
20.5
z
6
6.5
7
7.5
8
x+y


Forecast Distribution
HOUT
SUT
20
18
16
14
12
z
0
2
4
6
8
10
x+y

Uncertainty Quantification (UQ)

Consider a random variable $X \in \mathbb{R}^n$ and a nonlinear function $f : \mathbb{R}^n \to \mathbb{R}^m$.

Fundamental question of UQ:
Given information about the distribution of X what can we say about the distribution of $f(X)$?

Unscented Transform

- Goal: Estimate $\mathbb{E}[f(X)] = \int f(x)p(x)\,dx$
- Idea: Generate quadrature points for the weighted integral
- Quadrature: $\mathbb{E}[f(X)] \approx \sum_{i=1}^{N} w_i f(x_i)$ where x_i are nodes and w_i are weights for $i = 1, ..., N$
- Degree-of-exactness: the largest value of m so that all polynomials of degree m and below are integrated exactly.

Unscented Transform

Julier's Idea: Suppose we choose the right nodes so that our quadrature has degree of exactness 2, i.e. matches the first two moments exactly.

The σ–points of the Unscented Transform

Suppose we are given the first two moments, the mean $\mu \in \mathbb{R}^d$ and the covariance $C \in \mathbb{R}^{d\times d}$. Then the σ–points are defined by

$$\sigma_i = \begin{cases} \mu + \sqrt{dC}_i & \text{if } i = 1, \ldots, d \\ \mu - \sqrt{dC}_{i-d} & \text{if } i = d+1, \ldots, 2d \end{cases}$$

Note: $\sum_{i=1}^{d} \sqrt{C}_i \sqrt{C}_i^\top = C$

Empirical mean and Empirical covariance [Julier & Uhmann, 1999]

We have that

$$\mu = \mathbb{E}[X] = \frac{1}{2d}\sum_{i=1}^{2d}\sigma_i$$

$$C = \mathbb{E}[(X-\mu)(X-\mu)^\top] = \frac{1}{2d}\sum_{i=1}^{2d}(\sigma_i-\mu)(\sigma_i-\mu)^\top$$

and if $q : \mathbb{R}^d \to \mathbb{R}$ is a polynomial of degree at most 2, we have,

$$\mathbb{E}[q(X)] = \sum_{i=1}^{2d} w_i q(\sigma_i)$$

where w_i are the corresponding weights of σ_i and $w_i = \dfrac{1}{2d}$ for the Unscented Transform.

Tensors

Tensors are basically multidimensional matrices.

$$\begin{bmatrix} T_{111} & T_{121} & T_{131} \\ T_{211} & T_{221} & T_{231} \\ T_{311} & T_{321} & T_{331} \end{bmatrix}$$

$$\begin{bmatrix} T_{112} & T_{122} & T_{132} \\ T_{212} & T_{222} & T_{232} \\ T_{312} & T_{322} & T_{332} \end{bmatrix}$$

$$\begin{bmatrix} T_{113} & T_{123} & T_{133} \\ T_{213} & T_{223} & T_{233} \\ T_{313} & T_{323} & T_{333} \end{bmatrix}$$

k-order tensor

For positive integers d and k, a tensor T belonging to $\mathbb{R}^{d^k}$ is called a <u>k-order tensor</u> or simply a <u>k-tensor</u>.

Tensors

We can represent the covariance as

$$C = \mathbb{E}[(X-\mu)^{\otimes 2}] = \int (x-\mu)^{\otimes 2} p(x)\, dx$$

and the skewness is defined as

$$S = \mathbb{E}[(X-\mu)^{\otimes 3}] = \int (x-\mu)^{\otimes 3} p(x)\, dx$$

where

$$S_{ijk} = \int (x-\mu)_i (x-\mu)_j (x-\mu)_k p(x)\, dx.$$

The kurtosis is defined as

$$K = \mathbb{E}[(X-\mu)^{\otimes 4}] = \int (x-\mu)^{\otimes 4} p(x)\, dx$$

where $(x-\mu)^{\otimes k} = \underbrace{(x-\mu)\otimes(x-\mu)\otimes\cdots\otimes(x-\mu)}_{k \text{ times}}$ is a k-tensor.

Outline

Eigendecomposition is Rank-1 Decomposition

Recall that a symmetric matrix $A \in \mathbb{R}^{d \times d}$ with d linearly independent eigenvectors u_i can be factored as

$$A = U \Lambda U^\top$$

where U is the square $d \times d$ matrix whose ith column is the eigenvector u_i of A, and Λ is the diagonal matrix whose diagonal elements are the corresponding eigenvalues λ_i.

$$\begin{aligned} A &= \sum \lambda_i u_i u_i^\top \\ &= \sum \lambda_i u_i^{\otimes 2} \end{aligned}$$

Rank-1 Decomposition for Higher Order Tensors

Our goal is to do the same thing for higher order tensors and give them a formula of what that might look like, i.e.

$$\begin{aligned} S &= \sum_i x_i^{\otimes 3} \\ K &= \sum_i x_i^{\otimes 4} \end{aligned}$$

Multiplying a 2-Tensor with a 1-Tensor

Recall that for a matrix $A \in \mathbb{R}^{d \times d}$ and $v \in \mathbb{R}^d$ matrix vector multiplication

$$(Av)_i = \sum_{j=1}^{d} A_{ij} v_j$$

So we define two natural products

$$\begin{aligned}(A \times_1 v)_i &= \sum_{j=1}^{d} A_{ji} v_j = (A^\top v)_i \\ (A \times_2 v)_i &= \sum_{j=1}^{d} A_{ij} v_j = (Av)_i\end{aligned}$$

Note that multiplying a tensor by a vector, the order decreases by 1.

Multiplying a 3-Tensor with a 1-Tensor

Applying the same line of thinking with tensors, for a tensor $S \in \mathbb{R}^{d \times d \times d}$ and vector $v \in \mathbb{R}^d$, tensor vector multiplication goes as follows

$$
\begin{aligned}
(S \times_1 v)_{ik} &= \sum_{j=1}^{d} S_{jik} v_j \\
(S \times_2 v)_{ik} &= \sum_{j=1}^{d} S_{ijk} v_j \\
(S \times_3 v)_{ik} &= \sum_{j=1}^{d} S_{ikj} v_j
\end{aligned}
$$

each case resulting in a $d \times d$ matrix.

Multiplying a 3-Tensor with a 1-Tensor

Let $S \in \mathbb{R}^{3\times3\times3}$ and $v \in \mathbb{R}^3$ such that

$$S = \begin{bmatrix} S_{111} & S_{121} & S_{131} \\ S_{211} & S_{221} & S_{231} \\ S_{311} & S_{321} & S_{331} \end{bmatrix} \quad \begin{bmatrix} S_{112} & S_{122} & S_{132} \\ S_{212} & S_{222} & S_{232} \\ S_{312} & S_{322} & S_{332} \end{bmatrix} \quad \begin{bmatrix} S_{113} & S_{123} & S_{133} \\ S_{213} & S_{223} & S_{233} \\ S_{313} & S_{323} & S_{333} \end{bmatrix}$$

$$S\times_1 v = \begin{bmatrix} S_{111}v_1 + S_{211}v_2 + S_{311}v_3 & S_{112}v_1 + S_{212}v_2 + S_{312}v_3 & S_{113}v_1 + S_{213}v_2 + S_{313}v_3 \\ S_{121}v_1 + S_{221}v_2 + S_{321}v_3 & S_{122}v_1 + S_{222}v_2 + S_{322}v_3 & S_{123}v_1 + S_{223}v_2 + S_{323}v_3 \\ S_{131}v_1 + S_{231}v_2 + S_{331}v_3 & S_{132}v_1 + S_{232}v_2 + S_{332}v_3 & S_{133}v_1 + S_{233}v_2 + S_{333}v_3 \end{bmatrix}$$

Eigenvectors of a 3-Tensor

Notice that all moments are symmetric, namely

$$M_{i_1 \cdots i_n} = M_{\sigma(i_1 \cdots i_n)}$$

for any permutation σ.

$$(S \times_1 v) \times_1 v = \lambda v$$

$$((S \times_1 v) \times_1 v)_j = \sum_{k,i=1}^{d} S_{kij} v_k v_i$$

We want to decompose our tensor, i.e. we ultimately want a rank-1 decomposition such that

$$S = \sum_{i=1}^{r} v_i \otimes v_i \otimes v_i$$

Tensor Eigenvectors

Theorem

The eigenvector v of a k-order tensor T with size d, i.e. $T \in \mathbb{R}^{d^k}$ such that

$$\underbrace{(((T \times_1 v) \times_1 v) \cdots \times_1 v)}_{k-1 \text{ times}} = \lambda v$$

with maximum $|\lambda|$ gives the best rank-1 *approximation* of T meaning

$$\|T - \lambda v^{\otimes k}\|$$

is minimized over all possible λ, $\|v\| = 1$. [Kofidis & Regalia, 2002]

Numerical methods such as HOPM and S-HOPM are available for finding tensor eigenvectors.

"Peeling Process"

Now that we found the best rank-1 *approximation*, we now want a rank-1 *decomposition*.

Theorem

Consider the process of finding an approximate rank-1 decomposition of T by starting from $T_0 = T$ and setting

$$T_{\ell+1} = T_\ell - \lambda_\ell v_\ell^{\otimes k}$$

where λ_ℓ is the largest eigenvalue in absolute value of T_ℓ and v_ℓ is the associated eigenvector. Assume also that there exists a universal constant $c \in (0, 1]$ such that $\lambda_\ell \geq c|(T_\ell)_{i_1 \ldots i_k}|$. Then $\|T_\ell\|_F \to 0$ and for $r = \sqrt{1 - \frac{c^2}{d^k}} \in (0, 1)$

$$\frac{\|T_{\ell+1}\|_F}{\|T_\ell\|_F} \leq r$$

$$T = \sum_{\ell=1}^{L} \lambda_\ell v_\ell^{\otimes k} + \mathcal{O}(r^L)$$

for all $L \in \mathbb{N}$.

Outline

Higher Order Unscented Ensemble

One issue we come across is once we find the rank-1 decompositions of the higher moments, $S = \sum_{i=1}^{J} \tilde{v}_i^{\otimes 3}$ and $K = \sum_{i=1}^{L} s_i \tilde{u}_i^{\otimes 4}$, where the numbers s_i denote the sign of the eigenvalues of K, then the moments of the eigenvectors

$$\tilde{\mu} = \sum_{i=1}^{J} \tilde{v}_i \neq \mu$$

and

$$\tilde{C} = \sum_{i=1}^{L} s_i \tilde{u}_i^{\otimes 2} \neq C$$

So we can't just tack on these decompositions to Julier's Unscented Ensemble. We have to create our own.

We have constructed our own set of σ-points and corresponding weights associated with μ, C, S, and K such that

$$
\begin{aligned}
\sum_{i=-2}^{N} w_i \sigma_i &= \mu \\
\sum_{i=-2}^{N} w_i (\sigma_i - \mu)^{\otimes 2} &= C \\
\sum_{i=-2}^{N} w_i (\sigma_i - \mu)^{\otimes 3} &\approx S \\
\sum_{i=-2}^{N} w_i (\sigma_i - \mu)^{\otimes 4} &\approx K
\end{aligned}
$$

Higher Order Unscented Ensemble

The 4 moment σ–points of the Higher Order Unscented Transform

Suppose we are given the first 4 moments: $\mu \in \mathbb{R}^d$, $C \in \mathbb{R}^{d\times d}$, $S \in \mathbb{R}^{d\times d\times d}$, and $K \in \mathbb{R}^{d\times d\times d\times d}$ such that C is positive definite and S and K have the rank-1 decompositions $S = \sum_{i=1}^{J} \tilde{v}_i^{\otimes 3}$ and $K = \sum_{i=1}^{L} s_i \tilde{u}_i^{\otimes 4}$ where the numbers s_i denote the sign of the eigenvalues of K. Now let $\alpha, \beta, \gamma, \delta \in \mathbb{R}$ and denote $\tilde{\mu} = \sum_{i=1}^{J} \tilde{v}_i$, $\tilde{C} = \sum_{i=1}^{L} s_i \tilde{u}_i^{\otimes 2}$, $\hat{\mu} = \left(1 - d\beta^{-2} - \hat{L}\delta^{-4}\right)\mu - \gamma^{-2}\tilde{\mu}$, where $\hat{L} = \sum_{i=1}^{L} s_i$ and $\hat{C} = C - \frac{1}{\delta^2}\tilde{C}$ with $\delta > \sqrt{\frac{\lambda_{\max}^{\tilde{C}}}{\lambda_{\min}^{C}}}$.

The 4 moment σ–points of the Higher Order Unscented Transform

Then we define the 4 moment σ-points by

$$\sigma_i = \begin{cases} \mu & \text{if } i = -2 \\ \mu + \alpha\hat{\mu} & \text{if } i = -1 \\ \mu - \alpha\hat{\mu} & \text{if } i = 0 \\ \mu + \beta\sqrt{\hat{C}}_i & \text{if } i = 1, \ldots, d \\ \mu - \beta\sqrt{\hat{C}}_{i-d} & \text{if } i = d+1, \ldots, 2d \\ \mu + \gamma\tilde{v}_{i-2d} & \text{if } i = 2d+1, \ldots, 2d+J \\ \mu - \gamma\tilde{v}_{i-2d-J} & \text{if } i = 2d+J+1, \ldots, 2d+2J \\ \mu + \delta\tilde{u}_{i-2d-2J} & \text{if } i = 2d+2J+1, \ldots, 2d+2J+L \\ \mu - \delta\tilde{u}_{i-2d-2J-L} & \text{if } i = 2d+2J+L+1, \ldots, N \end{cases}$$

and the corresponding weights are

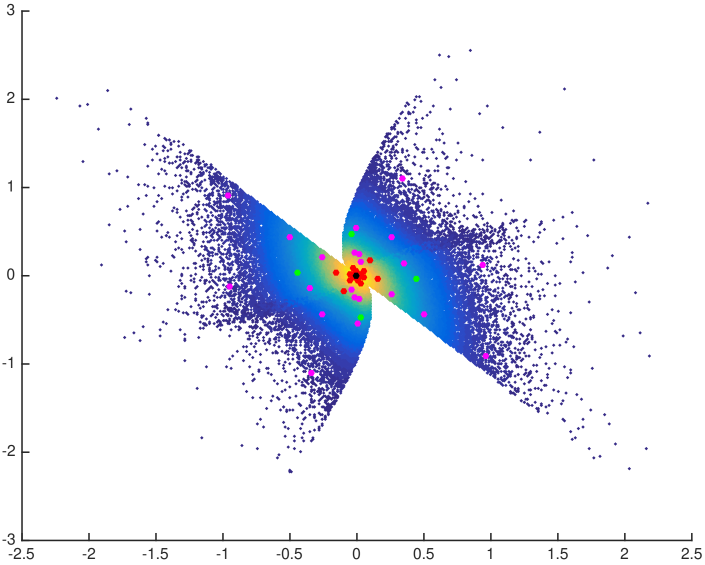
$$w_i = \begin{cases} 1 - d\beta^{-2} - \hat{L}\delta^{-4} & \text{if } i = -2 \\ \dfrac{1}{2\alpha} & \text{if } i = -1 \\ -\dfrac{1}{2\alpha} & \text{if } i = 0 \\ \dfrac{1}{2\beta^2} & \text{if } i = 1, \ldots, 2d \\ \dfrac{1}{2\gamma^3} & \text{if } i = 2d+1, \ldots, 2d+J \\ -\dfrac{1}{2\gamma^3} & \text{if } i = 2d+J+1, \ldots, 2d+2J \\ \dfrac{1}{2\delta^4}s_{i-2d-2J} & \text{if } i = 2d+2J+1, \ldots, 2d+2J+L \\ \dfrac{1}{2\delta^4}s_{i-2d-2J-L} & \text{if } i = 2d+2J+L+1, \ldots, N \end{cases}$$

For convenience, we denote $N = 2(d + J + L)$.

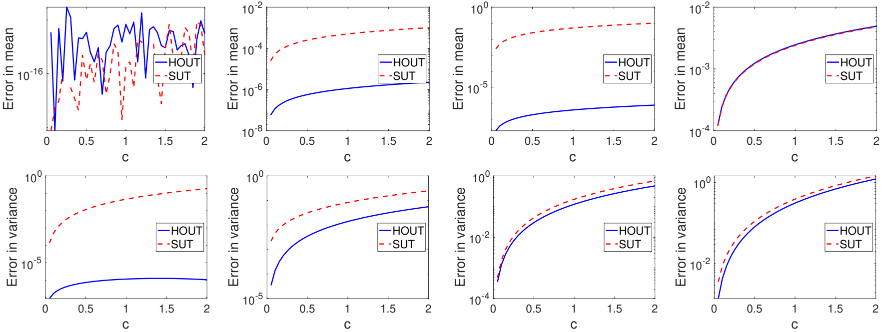
Theorem

Given the four moment σ-points associated with μ, C, S, and K, then for any $\delta > \sqrt{\frac{\lambda_{\max}^{\tilde{C}}}{\lambda_{\min}^{C}}}$ as defined above and α, β, γ we have

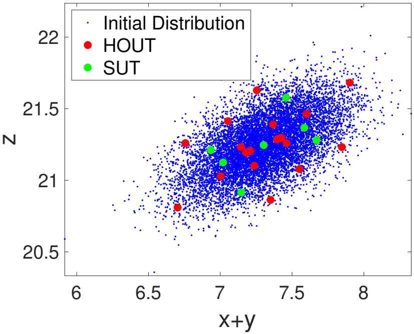
$$\begin{aligned}
\sum_{i=-1}^{N} w_i \sigma_i &= \mu \\
\sum_{i=-1}^{N} w_i (\sigma_i - \mu)^{\otimes 2} &= C \\
\sum_{i=-1}^{N} w_i (\sigma_i - \mu)^{\otimes 3} &= S + \alpha^2 \hat{\mu}^{\otimes 3} \\
\sum_{i=-1}^{N} w_i (\sigma_i - \mu)^{\otimes 4} &= K + \beta^2 \sum_{i=1}^{d} \sqrt{\hat{C}_i}^{\otimes 4}.
\end{aligned}$$

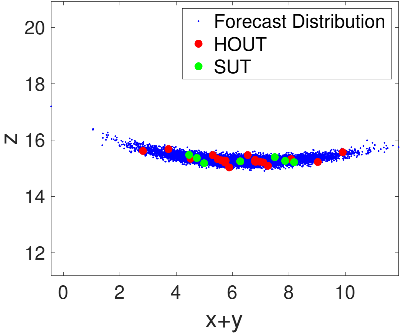

3
2
1
0
-1
-2
-3
-2.5 -2 -1.5 -1 -0.5 0 0.5 1 1.5 2 2.5

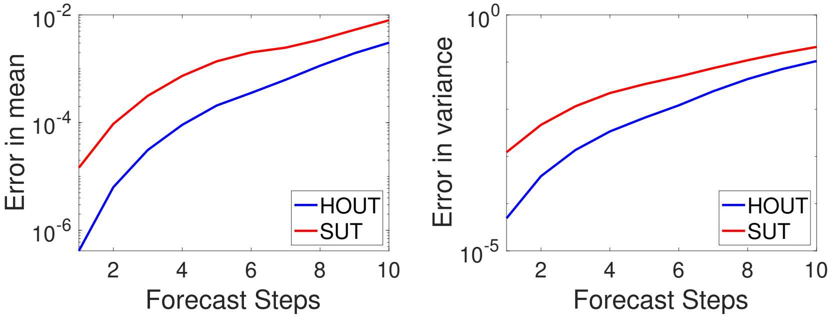
Comparing our transform to the Unscented Transform



$f(x) = ax + bcx^2$ $\quad$ $f(x) = ax + bcx^3$ $\quad$ $f(x) = ax + bcx^4$ $\quad$ $f(x) = ax + bcx^5$


Initial Distribution
HOUT
SUT
22
21.5
21
20.5
z
6
6.5
7
7.5
8
x+y


Forecast Distribution
HOUT
SUT
20
18
16
14
12
z
0
2
4
6
8
10
x+y


10^-2
10^-4
10^-6
Error in mean
2
4
6
8
10
Forecast Steps
HOUT
SUT
10^0
10^-5
Error in variance
2
4
6
8
10
Forecast Steps
HOUT
SUT

Bibliography

D. Easley and T. Berry.
A Higher Order Unscented Transform.
Preprint on arXiv.org. https://arxiv.org/abs/2006.13429

S. Julier and J. K. Uhmann,
A general method for approximating nonlinear transformations of probability distributions.
Research Gate, 1999.

E. Kofidis and P. Regalia.
On the Best Rank-1 Approximation of Higher-Order Supersymmetric Tensors.
SIAM J. MATRIX ANAL. APPL., Vol. 23, No. 3, pp. 863-884. 2002.