Generalizing the Unscented Ensemble Transform to Higher Moments

Deanna Easley

George Mason University

November 22, 2019

Outline

Outline

Uncertainty Quantification (UQ)

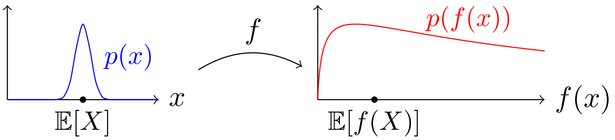
Consider a random variable $X \in \mathbb{R}^n$ and a nonlinear function $f : \mathbb{R}^n \to \mathbb{R}^m$.

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

Uncertainty Quantification (UQ)

Example:

- $X \sim \mathcal{N}\left(1, \frac{1}{100}\right)$, with distribution $p(x) = \frac{\exp(-50(x-1)^2)}{\sqrt{\pi/50}}$
- $f(x) = x^{10}$
- $\mathbb{E}[f(X)] = \int_{-\infty}^{\infty} f(x)p(x)\,dx = \int_{-\infty}^{\infty} x^{10} \frac{\exp(-50(x-1)^2)}{\sqrt{\pi/50}}\,dx \approx 1.5$



Uncertainty Quantification (UQ)

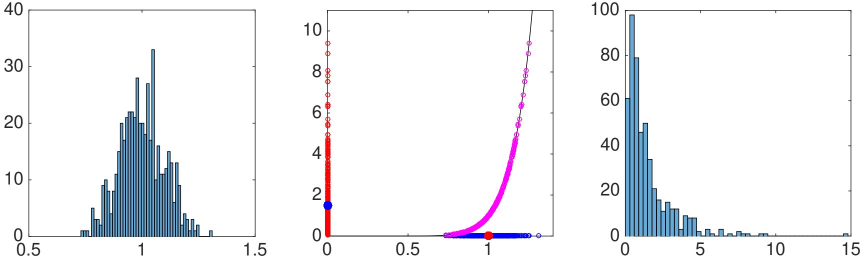
Consider a random variable $X \in \mathbb{R}^n$ and a nonlinear function $f : \mathbb{R}^n \to \mathbb{R}^m$.

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

- When distribution of X is fully known
- When we only know moments of X (mean, variance, etc.)
- When we have a finite collection of samples of X

Monte Carlo Simulation

- Sample $\{X_i\}_{i=1}^N$ from $p(x)$
- Estimate $\mathbb{E}[f(X)]$ by the average $\frac{1}{N}\sum_{i=1}^{N} f(X_i)$
- Problems: Need to know p and be able to sample it, need large N, f may be slow.



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

- One point quadrature: $\sigma_1 = \mu$, $w_1 = 1$

$$\int f(x)p(x)\,dx \approx w_1 f(\sigma_1) = f(\mu)$$

- Exact for $f(x) = ax + b$ (degree-of-exactness $= 1$)

$$\int (ax+b)p(x)\,dx = a\int xp(x)\,dx + b\int p(x)\,dx = a\mu + b$$

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

$$
\begin{aligned}
\mathbb{E}[X] &= \frac{1}{2d}\sum_{i=1}^{2d}\sigma_i \\
&= \frac{1}{2d}\sum_{i=1}^{d}(\mu + \sqrt{dC}_i) + \frac{1}{2d}\sum_{i=d+1}^{2d}(\mu - \sqrt{dC}_{i-d}) \\
&= \frac{1}{2d}\sum_{i=1}^{d}(\mu + \cancel{\sqrt{dC}_i} + \mu - \cancel{\sqrt{dC}_i}) \\
&= \frac{1}{2d}\sum_{i=1}^{d}2\mu \\
&= \frac{1}{2d}(2d\mu) \\
&= \mu
\end{aligned}
$$

Empirical covariance

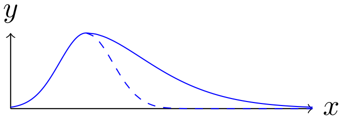
$$\begin{aligned}
\mathbb{E}[(X-\mu)(X-\mu)^\top] &= \frac{1}{2d}\sum_{i=1}^{2d}(\sigma_i-\mu)(\sigma_i-\mu)^\top \\
&= \frac{1}{2d}\left[\sum_{i=1}^{d}\sqrt{dC}_i\sqrt{dC}_i^\top + \sum_{i=d+1}^{2d}\sqrt{dC}_{i-d}\sqrt{dC}_{i-d}^\top\right] \\
&= \frac{1}{2d}\left[dC + dC\right] \\
&= \frac{1}{2d}(2dC\mu) \\
&= C
\end{aligned}$$

Unscented Transform

If $p(x)$ is Gaussian, then the Unscented Transform is good at approximating $\mathbb{E}[f(X)]$. But what if it isn't?

What about the next two moments?

Skewness: measure of the asymmetry of the probability distribution of a real-valued random variable about its mean



Kurtosis: measure of the "tailedness" of the probability distribution of a real-valued random variable

Tensors

Tensors are basically multidimensional matrices.

Let $x \in \mathbb{R}^d$. Then $xx^\top$ is a d-by-d matrix. Thus the ij-entry of $xx^\top$ can be represented as follows

$$(xx^\top)_{ij} = x_i x_j = (x \otimes x)_{ij} = (x^{\otimes 2})_{ij}$$

Thus we can represent the covariance as

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

More formally, the skewness is defined as

$$S = \int (x-\mu)^{\otimes 3} p(x)\, dx$$

where $(x-\mu)^{\otimes 3} = (x-\mu) \otimes (x-\mu) \otimes (x-\mu)$ is a 3-tensor so

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

The kurtosis is defined as

$$K = \int (x-\mu)^{\otimes 4} p(x)\, dx$$

where $(x-\mu)^{\otimes n} = \underbrace{(x-\mu) \otimes (x-\mu) \otimes \cdots \otimes (x-\mu)}_{n \text{ times}}$

Outline

Eigendecomposition

We are going to assume from now on that all moments are symmetric, namely

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

for any permutation σ.

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.

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

Eigendecomposition 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.

$$S = \sum_i x_i^{\otimes 3}$$

$$K = \sum_i x_i^{\otimes 4}$$

Finding the Eigendecomposition Numerically

Solving the characteristic polynomial is not an option for dimension $d \geq 5$ (no solution to general quintic).

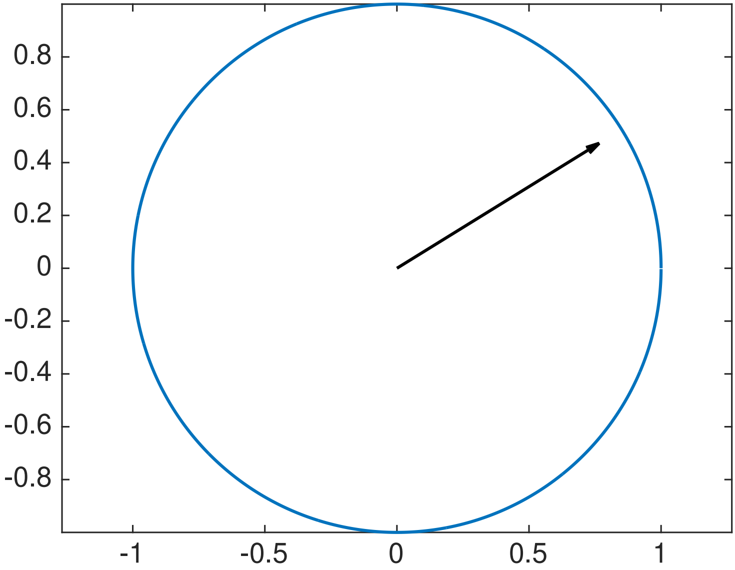
Power Iteration:

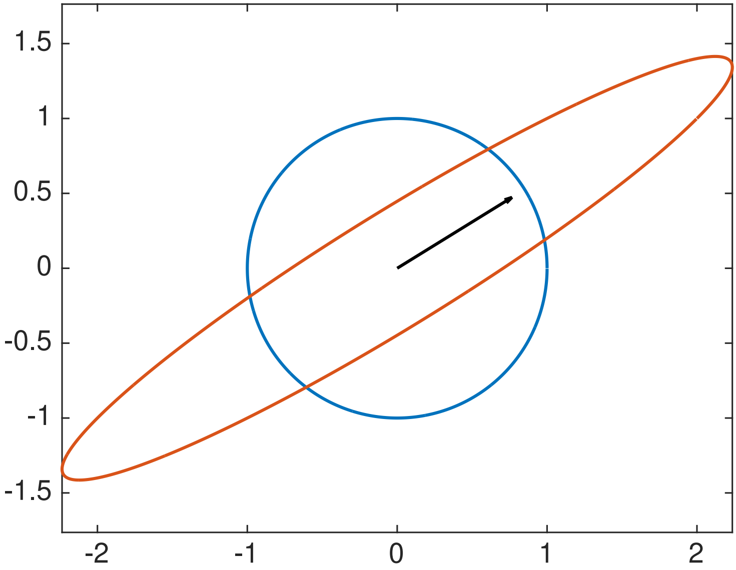
Random initial condition: $\vec{x} = \sum_{i=1}^{d} c_i \vec{u}_i$ (where $c_i = \langle \vec{x}, \vec{u}_i \rangle$)

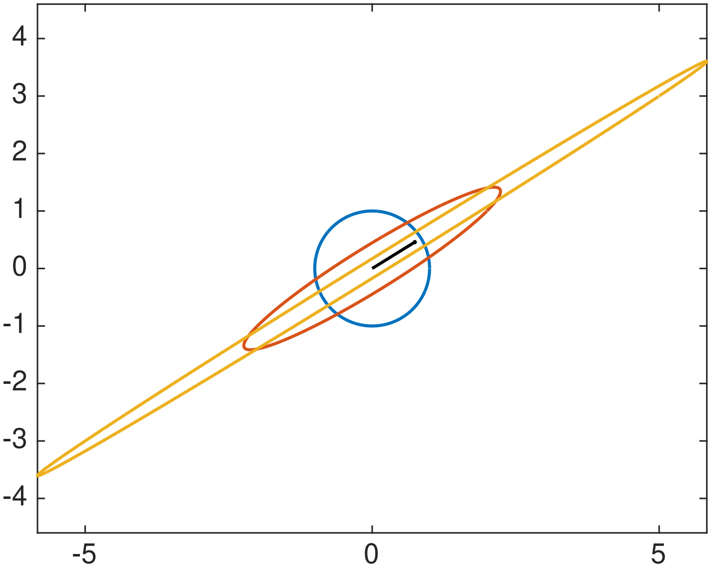
Multiply by A: $A\vec{x} = \sum_{i=1}^{d} c_i A\vec{u}_i = \sum_{i=1}^{d} c_i \lambda_i \vec{u}_i$

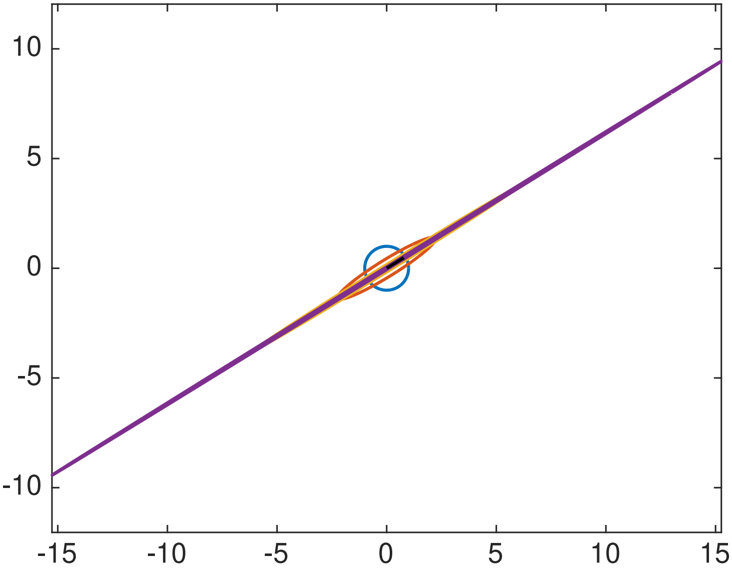
Repeat: $A^k \vec{x} = \sum_{i=1}^{d} c_i \lambda_i^k \vec{u}_i$

Largest eigenvalue wins.


0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
-0.5
0
0.5
1


1.5
1
0.5
0
-0.5
-1
-1.5
-2
-1
0
1
2


4
3
2
1
0
-1
-2
-3
-4
-5
0
5


10
5
0
-5
-10
-15
-10
-5
0
5
10
15

Power iteration blows up to ∞, so normalize

Normalized Power Iteration (NPI)

```
x = rand(d,1);
for k=1:10,
    x = A*x;
    x = x/norm(x);
end
```

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

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

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

Eigenvectors of a 3-Tensor

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

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

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

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

NPI for 3-Tensors

Tensor-Vector Product

```
repvec = size(S);
repvec(1) = 1;
Stimes1v = squeeze(sum(S.*repmat(v,repvec),1));
```

Symmetric Higher-Order Power Method (S-HOPM) [Kofidis & Regalia, 2002]

```
v = ones(size(S,1),1);
for iter=1:1000
    v = tensorXvector(S,v)*v;
    v = v/norm(v);
end

lambda = v'*(tensorXvector(S,v)*v)/(v'*v);
```

Symmetric Higher-Order Power Method (S-HOPM)

Theorem

The eigenvector u of a tensor T such that

$$(((T \times_1 u) \times_1 u) \cdots \times_1 u) = \lambda u$$

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

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

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

"Peeling Process"

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

After we've found the rank-1 approximation $u_1^{\otimes k}$, we subtract it from T and then recursively find the rank-1 approximation of the result and subtract it from T once again and repeat:

$$\begin{aligned} T_1 &= T - \lambda_1 u_1^{\otimes k} \\ T_2 &= T - \lambda_2 u_2^{\otimes k} \\ &\vdots \end{aligned}$$

The result will be our rank-1 decomposition: $T \approx \sum_i \lambda_i u_i^{\otimes k}$

Outline

Higher Order Unscented Ensemble

Recall our goal was when we are given the first four moments of the distribution of a random variable $X \in \mathbb{R}^n$ and we want to find the first four moments of the distribution of a nonlinear function $f : \mathbb{R}^n \to \mathbb{R}^m$.

We now have an effective algorithm for finding the rank-1 decomposition of tensors and thus have the ability to match multiple moments together.

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

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

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 $\alpha, \beta, \gamma, \delta, \zeta, \eta, \nu, \psi \in \mathbb{R}$ and denote $\tilde{\mu} = \sum_{i=1}^{J} \tilde{v}_i$, $\tilde{C} = \sum_{i=1}^{L} s_i \tilde{u}_i^{\otimes 2}$, $\hat{\mu} = \dfrac{(1 - 2d\eta - 2\hat{L}\psi)\mu - 2\nu\gamma\tilde{\mu}}{2\alpha\zeta}$, where $\hat{L} = \sum_{i=1}^{L} s_i$ and $\hat{C} = C - \dfrac{1}{\rho^2}\tilde{C}$ with $\rho > \sqrt{\dfrac{\lambda_{\max}^{\tilde{C}}}{\lambda_{\min}^{C}}}$.

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

Then we define the 4 moment σ-points by

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

and the corresponding weights are

$$w_i = \begin{cases} \zeta & \text{if } i = -1 \\ -\zeta & \text{if } i = 0 \\ \eta & \text{if } i = 1, \dots, 2d \\ \nu & \text{if } i = 2d+1, \dots, 2d+J \\ -\nu & \text{if } i = 2d+J+1, \dots, 2d+2J \\ \psi s_{i-2d-2J} & \text{if } i = 2d+2J+1, \dots, 2d+2J+L \\ \psi s_{i-2d-2J-L} & \text{if } i = 2d+2J+L+1, \dots, N \end{cases}$$

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

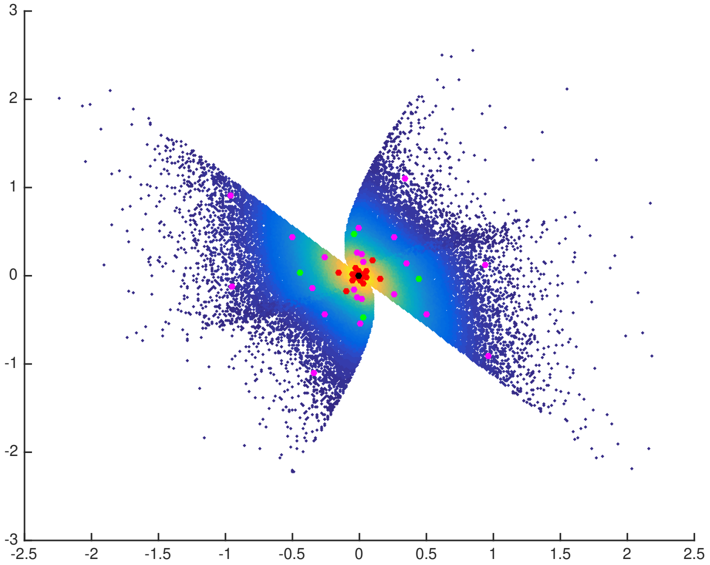
Theorem

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

$$\eta = \frac{1}{2\beta^2}, \quad \psi = \frac{1}{2\rho^4}, \quad \nu = \frac{1}{2\gamma^3}, \quad \text{and} \quad \delta^2 = \rho^2,$$

we have

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


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

Bibliography

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.