Research Statement

Deanna Easley

My research interests broadly live in the intersection of data driven analysis and applied dynamical systems. In particular, I am focused on inverse problems (especially Kalman filtering) and uncertainty quantification. Inverse problems can be roughly described as follows: given a set of observed noisy data, find the set of unknown variables, or hidden state, that generated them. When the hidden state comes from a dynamical system, we also have a system of equations that model the evolution of the hidden state over time. By finding the hidden state, we can then make predictions or forecasts based on the observed data. Key features of complex data are that they are noisy and the system may be evolving stochastically or randomly, which may prevent the determination of the true state of the system. Thus, the approach we use is to try to quantify uncertainty through examining error bounds or analyzing scenarios based on probability.

In addition to performing synthetic experiments, my research focus includes real-world applications. I have had the opportunity to become a guest researcher at National Institute of Standards and Technology (NIST) through a collaboration with Dr. Zeeshan Ahmed. For NIST, I am researching a unique defect in diamond called the nitrogen-vacancy (NV) center, where in place of two carbon atoms in the diamond carbon lattice, there is a nitrogen atom and a missing atom. The nitrogen-vacancy is sensitive to temperature, pressure, and electric and magnetic fields, which affect the quantum state defined by its Hamiltonian. However, there is a complex forward model that relates these parameters to spectroscopy curves that can be measured in a lab. This is an ideal example of an inverse problem, since NIST would like to use these NV diamonds as sensors (metrology). Specifically, I am looking into the promising applications of this quantum defect in quantum metrology and quantum information processing (QIP). My research group comprises of my advisor Dr. Tyrus Berry, my collaborator at NIST Dr. Zeeshan Ahmed, as well as second year graduate students Shraddha Rajpal and Jeanie Schreiber, whom I am mentoring. Together, we are working on the NV diamond sensor system where we plan to treat the estimated state as a probability distribution.

I consider uncertainty quantification to be a very exciting area of research due to the great potential impact it has for applications such as weather and climate predictions, and disease spread. For example, let us suppose we are given data of current weather patterns observed from satellites and past geological events observed from tree rings, ice cores and varves. To understand the evolution of the climate, we estimate the hidden states we are seeking, such as the mean or variance of the temperature of the planet over a long period of time. By analyzing the uncertainty of the observed data we try to estimate the uncertainty of the hidden state. Currently, an algorithm that achieves this is the Kalman filter. The Ensemble Kalman Filter (EnKF) was developed to account for nonlinear dynamical systems. However, the EnKF is not optimal, since it assumes that the set of data possesses characteristics that are generally unrealistic in nonlinear systems. My research deals with improving upon the EnKF to achieve a more accurate estimate of the hidden state. The principal objective is to develop a generalization of the EnKF, which I call the *Higher Order Kalman Filter* and I have made significant progress toward achieving this goal.

The Kalman Filter consists of two steps: forecast and assimilation. The forecast step is our prediction step in which the Kalman filter produces estimates of the current state variables, including their uncertainties. During the assimilation step, which takes place after the next measurement is observed, the estimates are updated using a weighted average. The forecast step of the EnKF is known as the *unscented transform*, which is a quadrature rule that estimates the expected value. The unscented transform was designed to match the mean and covariance of a distribution. The mean and covariance are known as the first two moments of the distribution.

In my research I have already finalized the forecast step of the Higher Order Kalman Filter by developing the *higher order unscented transform* (HOUT). I achieved this by expanding the method from the first two to the first four moments, where the third and the fourth moments are the skewness and the kurtosis, respectively. The *skewness* is a measure of the asymmetry of the probability distribution of a real-valued random variable about its mean, whereas the *kurtosis* is a measure of the "tailedness" of the probability distribution. The nodes of the HOUT are created using the CP decompositions of the first four moments. In addition, I developed a practical algorithm for computing a non-minimal CP decomposition of tensors and the first rigorous proof of convergence in linear time.

The methods I am developing are ideally suited for a wide variety of machine learning based applications. In fact, one of the initiatives I am planning to carry out at NIST is incorporating machine learning methods

into my algorithms. Specifically, we wish to establish whether machine learning would be well suited to learn the forward model in the case when working with a blackbox model. The ensemble based methods that I am working on for Uncertainty Quantification are the ideal tool for UQ in a machine learning model. This is because we do not need to compute or estimate any derivatives.

In what follows, a description of my research is provided. In Section 1, we generalize uncertainty quantification to higher moments for ensemble methods. In Section 2, we discuss methods for generalizing the Kalman equations to higher moments.

1 Uncertainty Quantification for Higher Order Moments

We are trying to estimate the expected value of some functional $f(x)$, $\mathbb{E}[f(X)] = \int f(x)p(x)\,dx$, based on some knowledge of the probability density function $p(x)$. This estimate can be achieved by a quadrature rule to approximate the expectation. A quadrature rule is a weighted sum of function values at specific points called *nodes*. The unscented transform is a quadrature rule that was introduced by Julier and Uhlmann in [4] and was further developed in [7, 5, 2, 3, 6]. The unscented transform uses the mean and covariance of a distribution to choose quadrature nodes and weights such that the quadrature rule has as degree of exactness two. Degree of exactness k means that a quadrature rule is exact for computing the expectation of polynomials up to degree k. (Equivalently, we match the first k moments of the distribution.) The nodes that are used in the Unscented Transform are called the *σ-points*.

Julier and Uhlmann used the rank decomposition of the covariance matrix for their σ-points. So when it came to developing the HOUT, it was natural to postulate that finding CP decompositions of the skewness and kurtosis would allow us to obtain new σ-points that could be used to match the higher order moments, namely, skewness and kurtosis.

To discuss higher moments, we need to make use of the notion of tensors. While the covariance of the distribution is represented by a matrix, the higher order moments, such as skewness and kurtosis, have to be represented by tensors, which can be viewed as multidimensional matrices. For positive integers d and k, a tensor T belonging to $\mathbb{R}^{d^k}$ is called a *k-order tensor* or simply a *k-tensor*. In particular, a vector can be thought of as a 1-tensor and a matrix as a 2-tensor. To represent the moments of a distribution in tensor notation, we make use of the following definition. Let $v \in \mathbb{R}^d$ and k a positive integer. Then the *kth-order tensor product* of v, denoted by

$$v^{\otimes k} = \underbrace{v \otimes v \otimes \cdots \otimes v}_{k \text{ times}},$$

is the k-tensor whose entries are defined by $(v^{\otimes k})_{i_1,\ldots,i_k} = v_{i_1} \cdots v_{i_k}$. Thus we may represent the covariance C, the skewness S, and the kurtosis K in tensor notation as follows

$$C = \int (x-\mu)^{\otimes 2}\, dp(x) \qquad S = \int (x-\mu)^{\otimes 3}\, dp(x) \qquad K = \int (x-\mu)^{\otimes 4}\, dp(x),$$

where μ is the mean and p is the probability density function.

Consider a symmetric k-tensor T of dimension d, i.e. $T \in \mathbb{R}^{d^k}$. By a theorem of [8], if v is the unit eigenvector of T associated to the largest eigenvalue λ (in absolute value), then $\lambda v^{\otimes k}$ is the best rank-1 approximation of T. It has been suggested in [8] that repeatedly subtracting the rank-1 approximations, namely, using $T_{\ell+1} = T_\ell - \lambda_\ell v_\ell^{\otimes k}$, where λ_ℓ is the largest eigenvalue in absolute value of T_ℓ and v_ℓ is an associated eigenvector, may result in an approximate CP decomposition. I recently formalized this and proved the following theorem.

Theorem 1.1 (Easley & Berry 2020, [1])**.** *Let T and T_ℓ be described as above. Then there exists an $r \in (0,1)$ such that the residuals, $\|T_\ell\|$, decay to 0, and we recover a linear upper bound on the decay rate. This yields an approximate CP decomposition of T:*

$$T = \sum_{\ell=1}^{L} \lambda_\ell v_\ell^{\otimes k} + \mathcal{O}(r^L) \quad \textit{for any} \;\; L \in \mathbb{N}.$$

Research Question 1: *The above theorem requires a bound on tensor eigenvalues, which we were able to prove for 3-tensors and 4-tensors, and determines the constant r. We wish to study how to extend these bounds to all tensors. Moreover, we look for ways to improve the bounds, since a sharper bound would result in a faster convergence rate.*

Given the following moments of the distribution of a random variable: the mean $\mu \in \mathbb{R}^d$, the covariance matrix $C \in \mathbb{R}^{d^2}$, the skewness tensor $S \in \mathbb{R}^{d^3}$, and kurtosis tensor $K \in \mathbb{R}^{d^4}$, we can use Theorem 1.1 to find the approximate CP decompositions of S and K. The key to forming an ensemble that matches all four moments simultaneously is to carefully balance the interactions between the moments. We were able to construct σ-points, σ_i, and weights w_i for $i = 0, \ldots, 2(d + J + L) + 3$, where J is the number of linearly independent eigenvectors of the skewness, and L is the number of linearly independent eigenvectors of the kurtosis. We call these the 4-*moment* σ-*points* of the Higher Order Unscented Transform. For details on the specific construction of the 4-moment σ-points, see [1].

With these 4-moment σ-points, the empirical mean and empirical covariance match exactly the mean and covariance, and the empirical skewness and empirical kurtosis approximate the skewness and kurtosis, respectively, up to an error term that can be made arbitrarily small.

Theorem 1.2 (Easley & Berry 2020, [1])**.** *Let τ be a specified tolerance for the absolute error of the skewness and kurtosis. Given the 4-moment σ-points associated with μ, C, S, and K, we have $\sum_{i=-2}^{N} w_i = 1$,*

$$\sum_{i=-2}^{N} w_i \sigma_i = \mu, \quad \sum_{i=-2}^{N} w_i (\sigma_i - \mu)^{\otimes 2} = C, \quad \left\| \sum_{i=-2}^{N} w_i (\sigma_i - \mu)^{\otimes 3} - S \right\|_F < \tau, \quad \text{and} \quad \left\| \sum_{i=-2}^{N} w_i (\sigma_i - \mu)^{\otimes 4} - K \right\|_F < \tau.$$

Having already developed a strategy for balancing four moments, I conjecture that this method could be generalized to build an ensemble for an arbitrary number of moments.

Research Question 2: *Is it possible to extend our method by building an ensemble for an arbitrary number of moments? If we succeed, we would try to analyze the limiting case as the number of moments grows to infinity.*

An application of HOUT is for forecasting dynamical systems. Being able to summarize a given distribution with a small set of points allows us to reduce the run time. In Figure 1 (B) and (C), we can see how the σ-points for the Unscented Transform are not capturing all the variation and are located mostly in the middle, missing a great deal, while the σ-points of the HOUT are capturing the full envelope of the distribution. The prediction errors comparing the two methods applied to the Lorenz dynamics are shown at the bottom.

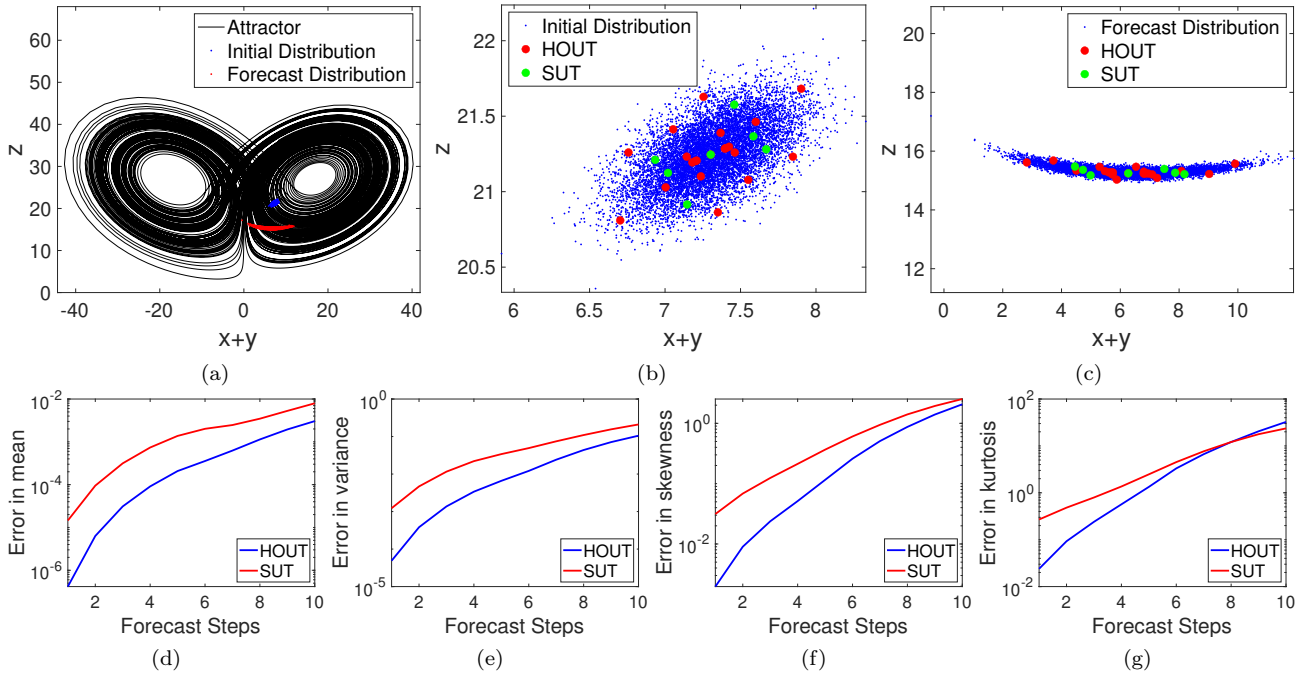

Figure 1: Comparison between the HOUT and the standard unscented transform (SUT) when estimating the mean $\mathbb{E}[f(X)]$ (top row) and higher moments of the Lorenz-63 model at various forecast horizons. In (A) we show the Lorenz-63 attractor (black) along with an example initial ensemble (blue) and forecast ensemble (red) used to compute the true statistics. In (B,C) we show the initial and forecast ensembles (blue) together with the HOUT (red) and SUT (green) ensembles. Results in (D-G) show the forecast accuracy versus the forecast horizon and are geometrically averaged over 500 different initial conditions on the attractor.

The prediction errors show that the Higher Order Unscented Ensemble provides improved estimates of the mean, variance, skewness and kurtosis.

2 Inverse Problems and Higher Order Kalman Filtering

The Kalman filter is a complicated inverse problem since it has a dynamically evolving state and there is noise in both the dynamics and observations. Given the dynamical system

$$\begin{aligned} x_k &= Fx_{k-1} + \omega_k \\ y_k &= Hx_k + \nu_k, \end{aligned}$$

where $\omega_k \sim \mathcal{N}(0, Q)$ and $\nu_k \sim \mathcal{N}(0, R)$, the forecast step of the Kalman filter consists of the equations

$$\begin{aligned} x_k^- &= Fx_{k-1}^+ \\ P_k^- &= FP_{k-1}^+F^\top + Q, \end{aligned}$$

where x_k^- is the mean and P_k^- is the covariance of the forecast distribution. The assimilation step of the Kalman filter consists of the equations

$$x_k^+ = x_k^- + K(y_k - Hx_k^-), \tag{1}$$
$$P_k^+ = (I - KH)P_k^-, \tag{2}$$

where x_k^+ is the mean and P_k^+ is the covariance of the posterior distribution, and $K = P^-H^\top(R + HP^-H^\top)^{-1}$ is the Kalman gain. Equalities (1) and (2) are known as the Kalman Equations. Given a nonlinear dynamical system

$$\begin{aligned} x_k &= f(x_{k-1}, \omega_k) \\ y_k &= h(x_k, \nu_k), \end{aligned}$$

in the forecast step, we can now use a higher order ensemble to forecast the first four moments, as was discussed in the previous section. What remains to be done is to generalize the assimilation step.

Moving forward, there are several ways the Higher Order Kalman filter could be derived. The original Kalman Filter was derived in three ways: the Bayesian approach, the Minimum Mean-Square Estimate (MMSE) approach and the Closure approach. Since each of these leads to the same Kalman filter, each one provides a different avenue for us to potentially generalize the Kalman Filter. We will now briefly discuss each of these in turn and how they might be generalized to higher order Kalman Filtering.

2.1 Bayesian approach

Assume we have the prior $p(x)$ (the probability of x) and the likelihood $p(y|x)$ (the probability of y given x). We wish to find the posterior $p(x|y)$ (probability of x given y). By Bayes' law,

$$p(x|y) \propto p(y|x)p(x).$$

Since both the likelihood and prior are Gaussian, the posterior is Gaussian as well. Writing the likelihood and the prior as $p(y|x) = e^{-\frac{1}{2}(y-Hx)^\top R^{-1}(y-Hx)}$ and $p(x) = e^{-\frac{1}{2}(x-x^-)^\top (P^+)^{-1}(x-x^-)}$, we get

$$p(x|y) \propto e^{-\frac{1}{2}(y-Hx)^\top R^{-1}(y-Hx)} e^{-\frac{1}{2}(x-x^-)^\top (P^+)^{-1}(x-x^-)}.$$

Thus, we derive the mean and covariance of the posterior, x^+ and P^+, respectively, where we define

$$P^+ = \left(H^\top R^{-1}H + (P^-)\right)^{-1} \quad \text{and} \quad x^+ = P^+\left(H^\top R^{-1}y + (P^-)^{-1}x^-\right).$$

Reformulating x^+, we get $x^+ = x^- + K(y - Hx^-)$ where $K = P^-H^\top(R + HP^-H^\top)^{-1}$ is the Kalman gain.

Research Question 3: *How can we generalize the Bayesian approach for the first four moments? A way to potentially achieve this is, instead of using a likelihood and prior that are exponentials with a quadratic function as the exponent, we use exponentials with a quartic function as the exponent. Thus instead of having Gaussian noise, we would allow a much more general class of noise.*

The nice feature about the Gaussian is that the mean and covariance can be immediately derived after completing the square in the exponent. This feature is lost when we deal with quartics. So there is an algebraic problem of how we complete the tesseract. We wish to use Bayes' law so we get

$$p(x|y) \propto e^{p(x)} e^{q(x)},$$

where $p(x)$ and $q(x)$ are quartic polynomials, and we wish to write it in the form $p(x|y) \propto e^{\hat{p}(x)}$, where $\hat{p}(x)$ is a quartic polynomial. The real challenge will be figuring out the connection between the moments and the coefficients to this polynomial.

2.2 Minimum Mean-Square Estimate (MMSE)

The Kalman filter can be derived by making the ansatz

$$\hat{x}_{k|k} = \begin{bmatrix} A_k & B_k \end{bmatrix} \begin{bmatrix} \hat{x}_{k|k-1} \\ \hat{y}_k \end{bmatrix} = A_k \hat{x}_{k|k-1} + B_k \hat{y}_k,$$

which is called a linear filter since the next estimate is a linear combination of the previous estimate and the observation. Constraining the filter to be unbiased (i.e. $\mathbb{E}[\hat{x}_{k|k}] = x_k$), we find that $A_k = I - B_k H_k$, so the filter becomes

$$\hat{x}_{k|k} = (I - B_k H_k)\hat{x}_{k|k-1} + B_k \hat{y}_k.$$

Minimizing over all possible B_k, we find that the minimum is achieved by the Kalman gain matrix, so $B_k = K_k$. This gives the MMSE filter

$$\hat{x}_{k|k} = (I - K_k H_k)\hat{x}_{k|k-1} + K_k \hat{y}_k,$$

which is identical to the Kalman filter.

Research Question 4: *How can we generalize the MMSE approach for the first four moments? One possibility is to look for optimal quadratic filters. We propose a quadratic ansatz as follows:*

$$\hat{x}_{k|k} = C_k \begin{bmatrix} \hat{x}_{k|k-1} \\ \hat{y}_k \end{bmatrix} + D_k \times_3 \begin{bmatrix} \hat{x}_{k|k-1} \\ \hat{y}_k \end{bmatrix} \times_2 \begin{bmatrix} \hat{x}_{k|k-1} \\ \hat{y}_k \end{bmatrix},$$

where C_k is a $n \times (n+m)$ matrix and D_k is a $n \times (n+m) \times (n+m)$ 3-tensor in which n is the dimension of the state and m is the dimension of the observation.

We still want an unbiased filter and to minimize the variance. We will probably have room for more constraints, and could even consider minimizing the kurtosis to help us look for the best possible C_k and D_k. There are some concerns, however. The linear filter has some nice properties that we are forced to give up by switching to this more strongly nonlinear form. Since stability might be an issue with a quadratic filter, we may need to use our constraints to enforce stability.

2.3 Closure approach

The evolution of the posterior probability density, $p(x,t)$, is described by the Kushner partial differential equation

$$dp = \mathcal{L}^* p \, dt + p(h - \overline{h}) R^{-1} \, dz.$$

The operator

$$\mathcal{L}^* p = -\sum_i \frac{\partial}{\partial x_i}(f_i p) + \sum_{i,j} \frac{\partial^2}{\partial x_j \partial x_i}(Q_{ij} p)$$

is the Kolmogorov Forward operator, where $Q_{ij} = \frac{1}{2}(qq^\top)_{ij} = \frac{1}{2}\sum_k q_{ik} q_{jk}$ is the system noise covariance matrix, h is the observation function, $\overline{h} = \mathbb{E}_p[h(x,t)] = \int h(x,t)p(x,t)\,dx$ is the expected observation, $R_{ij} = (rr^\top)_{ij} = \sum_k r_{ik} r_{jk}$ is the observation noise covariance matrix, $dz = dy - \overline{h} dt$ is the innovation process, and dy denotes the true observations.

Starting from the Kushner partial differential equation, we can compute ordinary differential equations for the mean and variance. Recalling the mean $\mu(t) = \mathbb{E}[x] = \int x p(x,t)\,dx$, its derivative is given by

$$\dot{\mu}(t) = \int x \frac{\partial}{\partial t} p(x,t)\,dx.$$

We then replace $\frac{\partial}{\partial t}p(x,t)$ with the right-hand side of the Kushner equation and integrate by parts, obtaining

$$\dot{\mu}(t) = (f(x_0) + Df(x_0)(\mu - x_0))\ dt + \sigma Dh(x_0)^\top R^{-1}\, dz.$$

Similarly, we can compute a dynamical system for the covariance

$$\dot{\sigma} = \left(\sigma F^\top + F\sigma + Q - \sigma H^\top R^{-1} H \sigma^\top\right)\ dt + S \cdot H^\top R^{-1}\, dz,$$

where S is the skewness tensor. If we assume the skewness equals zero (i.e. assume the distribution is Gaussian), this closure gives us the Kalman equations.

Research Question 5: *How can we generalize the closure approach for the first four moments? We could go further by deriving the ordinary differential equations for the skewness and kurtosis. Instead of assuming that the skewness is zero, we could find new closures for the first four moments rather than just the first two.*

One downside is that the equation for the kurtosis will involve the fifth moment, so we would have to make a closure there. The easiest approach would be to just assume that the fifth moment is zero. We will try to find more realistic closures in the future.

References

[1] Deanna C Easley and Tyrus Berry. A higher order unscented transform. *SIAM/ASA Journal on Uncertainty Quantification*, 9(3):1094–1131, 2021.

[2] Simon J Julier. Skewed approach to filtering. In *Signal and Data Processing of Small Targets 1998*, volume 3373, pages 271–282. International Society for Optics and Photonics, 1998.

[3] Simon J Julier. The scaled unscented transformation. In *Proceedings of the 2002 American Control Conference (IEEE Cat. No. CH37301)*, volume 6, pages 4555–4559. IEEE, 2002.

[4] Simon J Julier and Jeffrey K Uhlmann. A general method for approximating nonlinear transformations of probability distributions. Technical report, Technical report, Robotics Research Group, Department of Engineering Science, 1996.

[5] Simon J Julier and Jeffrey K Uhlmann. New extension of the kalman filter to nonlinear systems. In *Signal processing, sensor fusion, and target recognition VI*, volume 3068, pages 182–193. International Society for Optics and Photonics, 1997.

[6] Simon J Julier and Jeffrey K Uhlmann. Unscented filtering and nonlinear estimation. *Proceedings of the IEEE*, 92(3):401–422, 2004.

[7] Simon J Julier, Jeffrey K Uhlmann, and Hugh F Durrant-Whyte. A new approach for filtering nonlinear systems. In *Proceedings of 1995 American Control Conference-ACC'95*, volume 3, pages 1628–1632. IEEE, 1995.

[8] Eleftherios Kofidis and Phillip A Regalia. On the best rank-1 approximation of higher-order supersymmetric tensors. *SIAM Journal on Matrix Analysis and Applications*, 23(3):863–884, 2002.