

WIKIPEDIA

Unscented transform

The **unscented transform** (UT) is a mathematical function used to estimate the result of applying a given nonlinear transformation to a probability distribution that is characterized only in terms of a finite set of statistics. The most common use of the unscented transform is in the nonlinear projection of mean and covariance estimates in the context of nonlinear extensions of the Kalman filter. Its creator Jeffrey Uhlmann explained that "unscented" was an arbitrary name that he adopted to avoid it being referred to as the "Uhlmann filter."^[1]

Contents

Background

Motivation for the unscented transform

Sigma points

Example

Optimality of the unscented transform

Applications

Unscented Kalman Filter

See also

References

Background

Many filtering and control methods represent estimates of the state of a system in the form of a mean vector and an associated error covariance matrix. As an example, the estimated 2-dimensional position of an object of interest might be represented by a mean position vector, $[\boldsymbol{x}, \boldsymbol{y}]$, with an uncertainty given in the form of a 2x2 covariance matrix giving the variance in \boldsymbol{x} , the variance in \boldsymbol{y} , and the cross covariance between the two. A covariance that is zero implies that there is no uncertainty or error and that the position of the object is exactly what is specified by the mean vector.

The mean and covariance representation only gives the first two moments of an underlying, but otherwise unknown, probability distribution. In the case of a moving object, the unknown probability distribution might represent the uncertainty of the object's position at a given time. The mean and covariance representation of uncertainty is mathematically convenient because any linear transformation \boldsymbol{T} can be applied to a mean vector \boldsymbol{m} and covariance matrix \boldsymbol{M} as \boldsymbol{Tm} and \boldsymbol{TMT}^T . This linearity property does not hold for moments beyond the first raw moment (the mean) and the second central moment (the covariance), so it is not generally possible to determine the mean and covariance resulting from a nonlinear transformation because the result depends on all the moments, and only the first two are given.

Although the covariance matrix is often treated as being the expected squared error associated with the mean, in practice the matrix is maintained as an upper bound on the actual squared error. Specifically, a mean and covariance estimate (\mathbf{m}, \mathbf{M}) is conservatively maintained so that the covariance matrix \mathbf{M} is greater than or equal to the actual squared error associated with \mathbf{m} . Mathematically this means that the result of subtracting the expected squared error (which is not usually known) from \mathbf{M} is a semi-definite or positive-definite matrix. The reason for maintaining a conservative covariance estimate is that most filtering and control algorithms will tend to diverge (fail) if the covariance is underestimated. This is because a spuriously small covariance implies less uncertainty and leads the filter to place more weight (confidence) than is justified in the accuracy of the mean.

Returning to the example above, when the covariance is zero it is trivial to determine the location of the object after it moves according to an arbitrary nonlinear function $\mathbf{f}(\mathbf{x}, \mathbf{y})$: just apply the function to the mean vector. When the covariance is not zero the transformed mean will *not* generally be equal to $\mathbf{f}(\mathbf{x}, \mathbf{y})$ and it is not even possible to determine the mean of the transformed probability distribution from only its prior mean and covariance. Given this indeterminacy, the nonlinearly transformed mean and covariance can only be approximated. The earliest approximation was to linearize the nonlinear function and apply the resulting Jacobian matrix to the given mean and covariance. This is the basis of the extended Kalman Filter (EKF), and although it was known to yield poor results in many circumstances, there was no practical alternative for many decades.

Motivation for the unscented transform

In 1994 Jeffrey Uhlmann noted that the EKF takes a nonlinear function and partial distribution information (in the form of a mean and covariance estimate) of the state of a system but applies an approximation to the known function rather than to the imprecisely-known probability distribution. He suggested that a better approach would be to use the exact nonlinear function applied to an approximating probability distribution. The motivation for this approach is given in his doctoral dissertation, where the term *unscented transform* was first defined:^[2]

Consider the following intuition: *With a fixed number of parameters it should be easier to approximate a given distribution than it is to approximate an arbitrary nonlinear function/transformation.* Following this intuition, the goal is to find a parameterization that captures the mean and covariance information while at the same time permitting the direct propagation of the information through an arbitrary set of nonlinear equations. This can be accomplished by generating a discrete distribution having the same first and second (and possibly higher) moments, where each point in the discrete approximation can be directly transformed. The mean and covariance of the transformed ensemble can then be computed as the estimate of the nonlinear transformation of the original distribution. More generally, the application of a given nonlinear transformation to a discrete distribution of points, computed so as to capture a set of known statistics of an unknown distribution, is referred to as an *unscented transformation*.

In other words, the given mean and covariance information can be exactly encoded in a set of points, referred to as *sigma points*, which if treated as elements of a discrete probability distribution has mean and covariance equal to the given mean and covariance. This distribution can be propagated *exactly* by applying the nonlinear function to each point. The mean and covariance of the transformed set of points then represents the desired transformed estimate. The principal advantage of the approach is that the nonlinear function is fully exploited, as opposed to the EKF which replaces it with a linear one. Eliminating the need for linearization also provides

advantages independent of any improvement in estimation quality. One immediate advantage is that the UT can be applied with any given function whereas linearization may not be possible for functions that are not differentiable. A practical advantage is that the UT can be easier to implement because it avoids the need to derive and implement a linearizing Jacobian matrix.

Sigma points

To compute the unscented transform, one first has to choose a set of sigma points. Since the seminal work of Uhlmann, many different sets of sigma points have been proposed in the literature. A thoroughgoing review of these variants can be found in the work of Menegaz et. al.^[3] In general, $n + 1$ sigma points are necessary and sufficient to define a discrete distribution having a given mean and covariance in n dimensions.^[2]

A canonical set of sigma points is the symmetric set originally proposed by Uhlmann. Consider the following simplex of points in two dimensions:

$$\mathbf{s}_1 = \frac{1}{\sqrt{2}}[0, 2]^T, \quad \mathbf{s}_2 = -\frac{1}{\sqrt{2}}[\sqrt{3}, 1]^T, \quad \mathbf{s}_3 = -\frac{1}{\sqrt{2}}[-\sqrt{3}, 1]^T$$

It can be verified that the above set of points has mean $\mathbf{s} = [0, 0]^T$, and covariance $\mathbf{S} = \mathbf{I}$ (the identity matrix). Given any 2-dimensional mean and covariance, (\mathbf{x}, \mathbf{X}) , the desired sigma points can be obtained by multiplying each point by the matrix square root of \mathbf{X} and adding \mathbf{x} . A similar canonical set of sigma points can be generated in any number of dimensions n by taking the zero vector and the points comprising the rows of the identity matrix, computing the mean of the set of points, subtracting the mean from each point so that the resulting set has a mean of zero, then computing the covariance of the zero-mean set of points and applying its inverse to each point so that the covariance of the set will be equal to the identity.

Uhlmann showed that it is possible to conveniently generate a symmetric set of $2n + 1$ sigma points from the columns of $\pm\sqrt{n\mathbf{X}}$ and the zero vector, where \mathbf{X} is the given covariance matrix, without having to compute a matrix inverse. It is computationally efficient and, because the points form a symmetric distribution, captures the third central moment (the skew) whenever the underlying distribution of the state estimate is known or can be assumed to be symmetric.^[2] He also showed that weights, including negative weights, can be used to affect the statistics of the set. Julier also developed and examined techniques for generating sigma points to capture the third moment (the skew) of an arbitrary distribution and the fourth moment (the kurtosis) of a symmetric distribution.^{[4][5]}

Example

The unscented transform is defined for the application of a given function to any partial characterization of an otherwise unknown distribution, but its most common use is for the case in which only the mean and covariance is given. A common example is the conversion from one coordinate system to another, such as from a Cartesian coordinate frame to polar coordinates.^[4]

Suppose a 2-dimensional mean and covariance estimate, (\mathbf{m}, \mathbf{M}) , is given in Cartesian coordinates with:

$$\mathbf{m} = [12.3, 7.6]^T, \quad \mathbf{M} = \begin{bmatrix} 1.44 & 0 \\ 0 & 2.89 \end{bmatrix}$$

and the transformation function to polar coordinates, $f(\mathbf{x}, \mathbf{y}) \rightarrow [r, \theta]$, is:

$$r = \sqrt{x^2 + y^2}, \quad \theta = \arctan\left(\frac{y}{x}\right)$$

Multiplying each of the canonical simplex sigma points (given above) by $M^{\frac{1}{2}} = \begin{bmatrix} 1.2 & 0 \\ 0 & 1.7 \end{bmatrix}$ and adding the mean, m , gives:

$$\begin{aligned} m_1 &= [0, 2.40] + [12.3, 7.6] = [12.3, 10.0] \\ m_2 &= [-1.47, -1.20] + [12.3, 7.6] = [10.8, 6.40] \\ m_3 &= [1.47, -1.20] + [12.3, 7.6] = [13.8, 6.40] \end{aligned}$$

Applying the transformation function $f()$ to each of the above points gives:

$$\begin{aligned} m_1^+ &= f(12.3, 10.0) = [15.85, 0.68] \\ m_2^+ &= f(10.8, 6.40) = [12.58, 0.53] \\ m_3^+ &= f(13.8, 6.40) = [15.18, 0.44] \end{aligned}$$

The mean of these three transformed points, $m_{UT} = \frac{1}{3} \sum_{i=1}^3 m_i^+$, is the UT estimate of the mean in polar coordinates:

$$m_{UT} = [14.539, 0.551]$$

The UT estimate of the covariance is:

$$M_{UT} = \frac{1}{3} \sum_{i=1}^3 (m_i^+ - m_{UT})^2$$

where each squared term in the sum is a vector outer product. This gives:

$$M_{UT} = \begin{bmatrix} 2.00 & 0.0443 \\ 0.0443 & 0.0104 \end{bmatrix}$$

This can be compared to the linearized mean and covariance:

$$\begin{aligned} m_{\text{linear}} &= f(12.3, 7.6) = [14.46, 0.554]^T \\ M_{\text{linear}} &= \nabla_f M \nabla_f^T = \begin{bmatrix} 1.927 & 0.047 \\ 0.047 & 0.011 \end{bmatrix} \end{aligned}$$

The absolute difference between the UT and linearized estimates in this case is relatively small, but in filtering applications the cumulative effect of small errors can lead to unrecoverable divergence of the estimate. The effect of the errors are exacerbated when the covariance is underestimated because this causes the filter to be overconfident in the accuracy of the mean. In the above example it can be seen that the linearized covariance estimate is smaller than that of the UT estimate, suggesting that linearization has likely produced an underestimate of the actual error in its mean.

In this example there is no way to determine the absolute accuracy of the UT and linearized estimates without ground truth in the form of the actual probability distribution associated with the original estimate and the mean and covariance of that distribution after application of the nonlinear transformation (e.g., as determined analytically or through numerical integration). Such analyses have been performed for coordinate transformations under the assumption of Gaussianity for the underlying distributions, and the UT estimates tend to be significantly more accurate than those obtained from linearization.^{[6][7]}

Empirical analysis has shown that the use of the minimal simplex set of $n + 1$ sigma points is significantly less accurate than the use of the symmetric set of $2n$ points when the underlying distribution is Gaussian.^[7] This suggests that the use of the simplex set in the above example would not be the best choice if the underlying distribution associated with (\mathbf{m}, \mathbf{M}) is symmetric. Even if the underlying distribution is not symmetric, the simplex set is still likely to be less accurate than the symmetric set because the asymmetry of the simplex set is not matched to the asymmetry of the actual distribution.

Returning to the example, the minimal symmetric set of sigma points can be obtained from the covariance matrix $\mathbf{M} = \begin{bmatrix} 1.44 & 0 \\ 0 & 2.89 \end{bmatrix}$ simply as the mean vector, $\mathbf{m} = [12.3, 7.6]$ plus and minus the columns of $(2\mathbf{M})^{1/2} = \sqrt{2} * \begin{bmatrix} 1.2 & 0 \\ 0 & 1.7 \end{bmatrix} = \begin{bmatrix} 1.697 & 0 \\ 0 & 2.404 \end{bmatrix}$:

$$\mathbf{m}_1 = [12.3, 7.6] + [1.697, 0] = [13.997, 7.6]$$

$$\mathbf{m}_2 = [12.3, 7.6] - [1.697, 0] = [10.603, 7.6]$$

$$\mathbf{m}_3 = [12.3, 7.6] + [0, 2.404] = [12.3, 10.004]$$

$$\mathbf{m}_4 = [12.3, 7.6] - [0, 2.404] = [12.3, 5.196]$$

This construction guarantees that the mean and covariance of the above four sigma points is (\mathbf{m}, \mathbf{M}) , which is directly verifiable. Applying the nonlinear function $f()$ to each of the sigma points gives:

$$\mathbf{m}^+_1 = [15.927, 0.497]$$

$$\mathbf{m}^+_2 = [13.045, 0.622]$$

$$\mathbf{m}^+_3 = [15.854, 0.683]$$

$$\mathbf{m}^+_4 = [13.352, 0.400]$$

The mean of these four transformed sigma points, $\mathbf{m}_{UT} = \frac{1}{4} \sum_{i=1}^4 \mathbf{m}'_i$, is the UT estimate of the mean in polar coordinates:

$$\mathbf{m}_{UT} = [14.545, 0.550]$$

The UT estimate of the covariance is:

$$\mathbf{M}_{UT} = \frac{1}{4} \sum_{i=1}^4 (\mathbf{m}^+_i - \mathbf{m}_{UT})^2$$

where the each squared term in the sum is a vector outer product. This gives:

$$\mathbf{M}_{UT} = \begin{bmatrix} 1.823 & 0.043 \\ 0.043 & 0.012 \end{bmatrix}$$

The difference between the UT and linearized mean estimates gives a measure of the effect of the nonlinearity of the transformation. When the transformation is linear, for instance, the UT and linearized estimates will be identical. This motivates the use of the square of this difference to be added to the UT covariance to guard against underestimating of the actual error in the mean. This approach does not improve the accuracy of the mean but can significantly improve the accuracy of a filter over time by reducing the likelihood that the covariance is underestimated.^[2]

Optimality of the unscented transform

Uhlmann noted that given only the mean and covariance of an otherwise unknown probability distribution, the transformation problem is ill-defined because there is an infinite number of possible underlying distributions with the same first two moments. Without any a priori information or assumptions about the characteristics of the underlying distribution, any choice of distribution used to compute the transformed mean and covariance is as reasonable as any other. In other words, there is no choice of distribution with a given mean and covariance that is superior to that provided by the set of sigma points, therefore the unscented transform is trivially optimal.

This general statement of optimality is of course useless for making any quantitative statements about the performance of the UT, e.g., compared to linearization; consequently he, Julier and others have performed analyses under various assumptions about the characteristics of the distribution and/or the form of the nonlinear transformation function. For example, if the function is differentiable, which is essential for linearization, these analyses validate the expected and empirically-corroborated superiority of the unscented transform.^{[6][7]}

Applications

The unscented transform can be used to develop a non-linear generalization of the Kalman filter, known as the Unscented Kalman Filter (UKF). This filter has largely replaced the EKF in many nonlinear filtering and control applications, including for underwater,^[8] ground and air navigation,^[9] and spacecraft.^[10] The unscented transform has also been used as a computational framework for Riemann-Stieltjes optimal control.^[11] This computational approach is known as unscented optimal control.^{[12][13]}

Unscented Kalman Filter

Uhlmann and Simon Julier published several papers showing that the use of the unscented transformation in a Kalman filter, which is referred to as the unscented Kalman filter (UKF), provides significant performance improvements over the EKF in a variety of applications.^{[14][4][6]} Julier and Uhlmann published papers using a particular parameterized form of the unscented transform in the context of the UKF which used negative weights to capture assumed distribution information.^{[14][6]} That form of the UT is susceptible to a variety of numerical errors that the original formulations (the symmetric set originally proposed by Uhlmann) do not suffer. Julier has subsequently described parameterized forms which do not use negative weights and also are not subject to those issues.^[15]

See also

- Kalman filter
- Covariance intersection
- Ensemble Kalman filter
- Extended Kalman filter
- Non-linear filter
- Unscented optimal control

References

1. "First-Hand:The Unscented Transform – Engineering and Technology History Wiki" (http://www.ieeeeghn.org/wiki/index.php/First-Hand:The_Unscented_Transform).
2. Uhlmann, Jeffrey (1995). *Dynamic Map Building and Localization: New Theoretical Foundations* (Ph.D. thesis). University of Oxford.
3. Menegaz, Henrique M. T.; João, Y. Ishihara; Borges, Geovany A.; Vargas, Alessandro N. (16 February 2015). "A Systematization of the Unscented Kalman Filter Theory". *IEEE Transactions on Automatic Control*. **60** (10): 2583–2598. doi:10.1109/TAC.2015.2404511 (<https://doi.org/10.1109%2FTAC.2015.2404511>). hdl:20.500.11824/251 (<https://hdl.handle.net/20.500.11824%2F251>). S2CID 12606055 (<https://api.semanticscholar.org/CorpusID:12606055>).
4. Julier, S.; J. Uhlmann (1997). "Consistent Debiased Method for Converting Between Polar and Cartesian Coordinate Systems". *Proceedings of the 1997 SPIE Conference on Acquisition, Tracking, and Pointing*. **3086**. SPIE.
5. Julier, Simon (1998). "A Skewed Approach to Filtering". *The Proceedings of the 12th Intl. Symp. On Aerospace/Defense Sensing, Simulation and Controls*. **3373**. SPIE.
6. Julier, Simon; Uhlmann, Jeffrey (2000). "A New Method for the Nonlinear Transformation of Means and Covariances in Nonlinear Filters". *IEEE Transactions on Automatic Control*. **45** (3): 477–482. doi:10.1109/9.847726 (<https://doi.org/10.1109%2F9.847726>).
7. Zhang, W.; M. Liu; Z. Zhao (2009). "Accuracy Analysis of Unscented Transformation of Several Sampling Strategies". *Proc. of the 10th Intl. Conf. on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing*. ACIS.
8. Wu, L.; J. Ma; J. Tian (2010). "Self-Adaptive Unscented Kalman Filtering for Underwater Gravity Aided Navigation". *Proc. of IEEE/ION Plans*.
9. El-Sheimy, N; Shin, EH; Niu, X (2006). "Kalman Filter Face-Off: Extended vs. Unscented Kalman Filters for Integrated GPS and MEMS Inertial". *Inside GNSS: Engineering Solutions for the Global Navigation Satellite System Community*. **1** (2).
10. Crassidis, J.; Markley, F. (2003). "Unscented Filtering for Spacecraft Attitude Estimation". *Journal of Guidance, Control, and Dynamics*. **26** (4): 536–542. Bibcode:2003JGCD...26..536C (<https://ui.adsabs.harvard.edu/abs/2003JGCD...26..536C>). doi:10.2514/2.5102 (<https://doi.org/10.2514%2F2.5102>).
11. Ross, I. Michael; Proulx, Ronald J.; Karpenko, Mark; Gong, Qi (July 2015). "Riemann–Stieltjes Optimal Control Problems for Uncertain Dynamic Systems" (<https://escholarship.org/uc/item/4jq3s10x>). *Journal of Guidance, Control, and Dynamics*. **38** (7): 1251–1263. Bibcode:2015JGCD...38.1251R (<https://ui.adsabs.harvard.edu/abs/2015JGCD...38.1251R>). doi:10.2514/1.G000505 (<https://doi.org/10.2514%2F1.G000505>).
12. I. M. Ross, R. J. Proulx, and M. Karpenko, "Unscented Optimal Control for Space Flight," *Proceedings of the 24th International Symposium on Space Flight Dynamics (ISSFD)*, May 5–9, 2014, Laurel, MD. http://issfd.org/ISSFD_2014/ISSFD24_Paper_S12-5_Karpenko.pdf
13. Ross, I. Michael; Proulx, Ronald J.; Karpenko, Mark (July 2015). "Unscented guidance". *2015 American Control Conference (ACC)*. pp. 5605–5610. doi:10.1109/ACC.2015.7172217 (<https://doi.org/10.1109%2FACC.2015.7172217>). ISBN 978-1-4799-8684-2. S2CID 28136418 (<https://api.semanticscholar.org/CorpusID:28136418>).
14. Julier, S.; J. Uhlmann (1997). "New Extension of the Kalman Filter to Nonlinear Systems". *Proceedings of the 1997 SPIE Conference on Signal Processing, Sensor Fusion, and Target Recognition*. **3068**.
15. Julier, Simon (2002). "The Scaled Unscented Transformation". *Proceedings of the American Control Conference*. **6**. IEEE.

Retrieved from "https://en.wikipedia.org/w/index.php?title=Unscented_transform&oldid=1056192921"

This page was last edited on 20 November 2021, at 09:13 (UTC).

Text is available under the Creative Commons Attribution–ShareAlike License; additional terms may apply. By using this site, you agree to the Terms of Use and Privacy Policy. Wikipedia® is a registered trademark of the Wikimedia Foundation, Inc., a non–profit organization.